DAO Office Note 1999-01

Office Note Series on
Global Modeling and Data Assimilation

Robert Atlas, Head
Data Assimilation Office
Goddard Space Flight Center
Greenbelt, Maryland

Estimation Theory and Foundations of Atmospheric Data Assimilation

Ricardo Todling

General Sciences Corporation
Data Assimilation Office, Goddard Laboratory for Atmospheres

Class Notes created for courses given at Instituto de Matemática Pura e Applicada in 1996 and at University of Maryland in 1997

Goddard Space Flight Center
Greenbelt, Maryland 20771
June 1999

The following is a collection of notes geared to provide an elementary introduction to the topic of data assimilation. The topic is presented with the point of view of estimation theory. As such, the first half of these notes is devoted to presenting basic concepts of probability theory, stochastic processes, estimation and filtering. The second half of these notes gives an introduction to atmospheric data assimilation and related problems. Illustrations of advanced assimilation procedures are given by discussing results from the application of Kalman filtering and smoothing to a linear shallow-water model.

Classes based on earlier versions of these notes have been presented at the Instituto de Matemática Pura e Aplicada, in Rio de Janeiro, at the Department of Meteorology of University of Maryland, and at Laboratorio Nacional de Computação Científica, Rio de Janeiro.

Contents

List of Figures

Chapter 1

Fundamental Concepts of Probability Theory

1.1 Probability Space

1.1.1 The Probability Triplet

The probability space is formally defined through the probability triplet $(\Omega, \mathcal{B}, P)$ where,

- Ω: is the sample space, which contains all possible outcomes of an experiment.
- $\mathcal{B}$: is a set of subsets of Ω (a Borel field — a closed set under operations of: union, intersection and complement)
- P: is a scalar function defined on $\mathcal{B}$, called the probability function or probability measure.

Each set $B \in \mathcal{B}$ is called an event, that is, B is a collection of specific possible outcomes. In what follows, the mathematical details corresponding to the field $\mathcal{B}$ will be ignored (e.g., see Chung [26], for a detailed treatment). The values $\omega \in \Omega$ are the realizations, and for each set $B \in \mathcal{B}$, the function $P(B)$ defines the probability that the realization ω is in B. The quantity P is a probability function if it satisfies the following axioms:

1. $0 \leq P(B) \leq 1$, for all $B \in \mathcal{B}$
2. $P(\Omega) = 1$
3. $P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i)$, for all disjoint sequences of $B_i \in \mathcal{B}$.

1.1.2 Conditional Probability

If A and B are two events and $P(B) \neq 0$, the conditional probability of A given B is defined as

$$P(A|B) \equiv P(A \cap B)/P(B) \tag{1.1}$$

The events A and B are statistically independent if $P(A|B) = P(A)$. Consequently, $P(A \cap B) = P(A)P(B)$.

Analogously,

$$P(B|A) = \frac{P(B \cap A)}{P(A)}, \tag{1.2}$$

for all $P(A) \neq 0$.

Combining the two relations above we have:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{1.3}$$

which is known as Bayes rule (or theorem) for probabilities. This relation is useful when we need to reverse the condition of events.

1.2 Random Variables

A scalar $\mathrm{x}(\omega)$ random variable (r.v.) is a function, whose value x is determined by the result ω of a random experiment. Note the typographical distinction between both quantities. In other words, an r.v. $\mathrm{x}(\omega)$ attributes a real number x to each point of the sample space. The particular value x assumed by the random variable is referred to as a *realization*. A random variable is defined in such a way that all sets $B \subset \Omega$ of the form

$$B = \{\omega : \mathrm{x}(\omega) \leq \xi\} \tag{1.4}$$

are in $\mathcal{B}$, for any value of $\xi \in R^1$.

1.2.1 Distribution and Density Functions

Each r.v. has a distribution function defined as

$$F_{\mathrm{x}}(x) \equiv P(\{\omega : \mathrm{x}(\omega) \leq x\}), \tag{1.5}$$

which represents the probability that x is less than or equal to x.

It follows, directly from the properties of the probability measure given above, that $F_{\mathrm{x}}(x)$ should be a non-decreasing function of x, with $F_{\mathrm{x}}(-\infty) = 0$ and $F_{\mathrm{x}}(\infty) = 1$. Under reasonable conditions, we can define a function called a probability density function, derived from the distribution function:

$$p_{\mathrm{x}}(x) \equiv \frac{dF_{\mathrm{x}}(x)}{dx}, \tag{1.6}$$

Table 1.1: Properties of probability density functions and distribution functions

$F_{\mathrm{X}}(-\infty) = 0$		(a)
$F_{\mathrm{X}}(+\infty) = 1$		(b)
$F_{\mathrm{X}}(x_1) \leq F_{\mathrm{X}}(x_2)$,	for all $x_1 \leq x_2$	(c)
$p_{\mathrm{X}}(x) \geq 0$,	for all x	(d)
$\int_{-\infty}^{\infty} p_{\mathrm{X}}(x)\, dx = 1$		(e)

Consequently, the inverse relation

$$F_{\mathrm{X}}(x) = \int_{-\infty}^{x} p_{\mathrm{X}}(s)\, ds\,, \tag{1.7}$$

provides the distribution function. The probability density function should be non-negative, and its integral over the real line should be unity. Table 1.1 presents a summary of the properties of probability density functions and distribution functions.

A few examples of continuous distribution functions are given below:

(I) Uniform:

$$p_{\mathrm{X}}(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{1.8}$$

$$F_{\mathrm{X}}(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases} \tag{1.9}$$

(ii) Exponential:

$$p_{\mathrm{X}}(x) = \begin{cases} \frac{1}{a} e^{-x/a} & 0 < x \\ 0 & \text{otherwise} \end{cases} \tag{1.10}$$

$$F_{\mathrm{X}}(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x/a} & x \geq 0 \end{cases} \tag{1.11}$$

(iii) Rayleigh:

$$p_{\mathrm{X}}(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{a^2} e^{-x^2/2a^2} & x \geq 0 \end{cases} \tag{1.12}$$

$$F_{\mathrm{X}}(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x^2/2a^2} & x \geq 0 \end{cases} \tag{1.13}$$

(iv) Gaussian:

$$p_{\mathrm{X}}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{1.14}$$

$$F_{\mathrm{X}}(x) = \operatorname{erf}\left(\frac{x-\mu}{\sigma}\right) \tag{1.15}$$

where $\operatorname{erf}(x)$ is the error function (Arfken [5], p. 568):

$$\operatorname{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy \tag{1.16}$$

Remark: An r.v. with Gaussian distribution is said to be normally distributed, with mean μ and variance σ^2 (see following section) and is represented symbolically by $\mathrm{x} \sim \mathcal{N}(\mu, \sigma^2)$.

(iv) χ^2 (Chi–Square):

$$p_{\mathrm{x}}(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{1.17}$$

$$F_{\mathrm{x}}(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^x u^{\nu/2-1} e^{-u/2}\, du \tag{1.18}$$

where $\Gamma(\nu)$ is the gamma function (Arfken [5], Chapter 10):

$$\Gamma(\nu) = \int_0^\infty t^{n-1} e^{-t}\, dt \tag{1.19}$$

Remarks: An r.v. that is χ^2 distributed has the form: $\chi^2 = \mathrm{x}_1^2 + \mathrm{x}_2^2 + \cdots + \mathrm{x}_\nu^2$, with the variables x_i, for $i = 1, 2, \cdots, \nu$, being normally distributed with mean zero and unity variance.

1.2.2 Expectations and Moments

The mean of an r.v. x is defined as

$$\mathcal{E}\{\mathrm{x}\} \equiv \int_{-\infty}^{\infty} x\, p_{\mathrm{x}}(x)\, dx\,. \tag{1.20}$$

In this course we will use interchangeably the expressions expected value and expectation as synonyms for mean. There are extensions of this definition for those cases in which the probability density $p_{\mathrm{x}}(x)$ does not exist; however in the context that interests us, the definition above is sufficient. Any measurable function of an r.v. is also an r.v. and its mean is given by:

$$\mathcal{E}\{f(\mathrm{x})\} = \int_{-\infty}^{\infty} f(x)\, p_{\mathrm{x}}(x)\, dx\,. \tag{1.21}$$

In particular, if $f(\mathrm{x}) = a = const.$, $\mathcal{E}\{a\} = a$, due to property (e) in Table 1.1.

If $f(\mathrm{x}) = a_1 g_1(\mathrm{x}) + a_2 g_2(\mathrm{x})$ then

$$\mathcal{E}\{a_1 g_1(\mathrm{x}) + a_2 g_2(\mathrm{x})\} = a_1 \mathcal{E}\{g_1(\mathrm{x})\} + a_2 \mathcal{E}\{g_2(\mathrm{x})\}. \tag{1.22}$$

A function of special interest is $f(\mathrm{x}) = \mathrm{x}^n$, where n is a positive integer. The means

$$\mathcal{E}\{\mathrm{x}^n\} \equiv \int_{-\infty}^{\infty} x^n p_{\mathrm{x}}(x) dx\,, \tag{1.23}$$

define the moments of order n of x. In particular, $\mathcal{E}\{\mathrm{x}^2\}$ is called the mean–square value. The expectations

$$\mathcal{E}\{(\mathrm{x} - \mathcal{E}\{\mathrm{x}\})^n\} \equiv \int_{-\infty}^{\infty} (x - \mathcal{E}\{\mathrm{x}\})^n p_{\mathrm{x}}(x) dx\,, \tag{1.24}$$

define the n–th moments of x about its mean (n–th central moment).

The second moment of x about its mean is called the variance of x, and is given by:

$$\begin{aligned} var(\mathrm{x}) \equiv \mathcal{E}\{(\mathrm{x}-\mathcal{E}\{\mathrm{x}\})^2\} &= \mathcal{E}\{\mathrm{x}^2\} - 2\mathcal{E}\{\mathrm{x}\mathcal{E}\{\mathrm{x}\}\} + (\mathcal{E}\{\mathrm{x}\})^2 \\ &= \mathcal{E}\{\mathrm{x}^2\} - (\mathcal{E}\{\mathrm{x}\})^2 . \end{aligned} \tag{1.25}$$

That is, the variance is the mean–square minus the square of the mean. Finally, the standard deviation is defined as the square–root of the variance:

$$\sigma(\mathrm{x}) \equiv \sqrt{[var(\mathrm{x})]}. \tag{1.26}$$

It is worth mentioning at this point that in many cases, the mean value of an r.v. is used as a guess (or estimate) for the true value of that variable. Other quantities of interest in this sense are the median, the mid–range, and the mode values. The median μ_1 is given by

$$\int_{-\infty}^{\mu_1} p_{\mathrm{X}}(x)\,dx = \int_{\mu_1}^{\infty} p_{\mathrm{X}}(x)\,dx = \frac{1}{2}, \tag{1.27}$$

the mid–range μ_∞ is given by

$$\mu_\infty = \frac{\max_x(x) + \min_x(x)}{2} \tag{1.28}$$

and the mode m is given by

$$\left.\frac{dp_{\mathrm{X}}(x)}{dx}\right|_{x=m} = 0 \tag{1.29}$$

The median divides the probability density function in two, each one covering the same area. The mode corresponds to values of the random variable for which the probability density function is maximum, that is, it corresponds to the most likely value. The importance, and more general meaning, of these quantities will become clear as we advance.

1.2.3 Characteristic Function

An r.v. can be represented, alternatively, by its characteristic function which is defined as

$$\phi_{\mathrm{X}}(u) \equiv \mathcal{E}\{\exp(iu\mathrm{x})\}, \tag{1.30}$$

where $i = \sqrt{-1}$.

According to the definition (1.20) of mean we see that the characteristic function is nothing more than the Fourier transform of the density function:

$$\phi_{\mathrm{X}}(u) = \int_{-\infty}^{\infty} \exp(iux) p_{\mathrm{X}}(x)\,dx , \tag{1.31}$$

from this it follows that the probability density is the inverse Fourier transform of the characteristic function, that is,

$$p_{\mathrm{X}}(x) = (1/2\pi)\int_{-\infty}^{\infty} \exp(-iux)\phi_{\mathrm{X}}(u)\,du . \tag{1.32}$$

Let us now take the derivative of the characteristic function (1.31) with respect to u:

$$\begin{aligned}\frac{d\phi_{\mathrm{x}}(u)}{du} &= \frac{d}{du}\int_{-\infty}^{\infty}\exp(iux)p_{\mathrm{x}}(x)\,dx\,,\\ &= \int_{-\infty}^{\infty}\frac{d\exp(iux)}{du}p_{\mathrm{x}}(x)\,dx\,,\\ &= i\mathcal{E}\{\mathrm{x}\exp(iu\mathrm{x})\}\,,\end{aligned} \tag{1.33}$$

where we used the definition of characteristic function to get the last equality. Notice that by choosing calculate the expression above for at $u = 0$ we have

$$\left.\frac{d\phi_{\mathrm{x}}(u)}{du}\right|_{u=0} = i\mathcal{E}\{\mathrm{x}\}\,, \tag{1.34}$$

or better yet,

$$\mathcal{E}\{\mathrm{x}\} = \frac{1}{i}\left.\frac{d\phi_{\mathrm{x}}(u)}{du}\right|_{u=0}\,, \tag{1.35}$$

which give an alternative way of calculating the first moment, if the characteristic function is given. As a matter of fact moments of order n can be calculated analogously, by taking n derivatives of the characteristic function and evaluating the result at $u = 0$. This procedure produces the equation

$$\mathcal{E}\{\mathrm{x}^n\} = \frac{1}{i^n}\left.\frac{d^n\phi_{\mathrm{x}}(u)}{du^n}\right|_{u=0}\,, \tag{1.36}$$

for the n–th moment.

1.3 Jointly Distributed Random Variables

1.3.1 Distribution, Density Function and Characteristic Function

The r.v.'s $\mathrm{x}_1, \cdots, \mathrm{x}_n$ are said to be jointly distributed if they are defined in the same probability space. They can be characterized by the joint distribution function

$$F_{\mathrm{x}_1\cdots\mathrm{x}_n} \equiv P\{\omega : \mathrm{x}_1 \leq x_1, \cdots, \mathrm{x}_n \leq x_n\} \tag{1.37}$$

where

$$\{\omega : \mathrm{x}_1 \leq x_1, \cdots, \mathrm{x}_n \leq x_n\} \equiv \{\mathrm{x}_1(\omega) \leq x_1\} \cap \cdots \cap \{\mathrm{x}_n(\omega) \leq x_n\} \tag{1.38}$$

or alternatively, by their joint density function:

$$F_{\mathrm{x}_1\cdots\mathrm{x}_n}(x_1, \cdots, x_n) \equiv \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} p_{\mathrm{x}_1\cdots\mathrm{x}_n}(x_1', \cdots, x_n')dx_1'\cdots dx_n'\,, \tag{1.39}$$

from which it follows that

$$p_{\mathrm{x}_1\cdots\mathrm{x}_n}(x_1, \cdots, x_n) = \frac{\partial^n}{\partial x_1\cdots\partial x_n}F_{\mathrm{x}_1\cdots\mathrm{x}_n}(x_1, \cdots, x_n) \tag{1.40}$$

assuming the existence of the derivatives.

The characteristic function of jointly distributed r.v.'s $\mathrm{x}_1, \cdots, \mathrm{x}_n$ is defined as:

$$\phi_{\mathrm{x}_1\cdots\mathrm{x}_n}(u_1, \cdots, u_n) \equiv \mathcal{E}\{\exp\left(i\sum_{j=1}^{n}u_j\mathrm{x}_j\right)\}\,. \tag{1.41}$$

1.3.2 Expectations and Moments

If f is a function of jointly distributed r.v.'s $\mathrm{x}_1, \cdots, \mathrm{x}_n$, and $\mathrm{y} = f(\mathrm{x}_1, \cdots, \mathrm{x}_n)$, then

$$\mathcal{E}\{\mathrm{y}\} \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) p_{\mathrm{x}_1 \cdots \mathrm{x}_n}(x_1, \cdots, x_n)\, dx_1 \cdots dx_n\,. \tag{1.42}$$

The expected value of x_k is given by

$$\mathcal{E}\{\mathrm{x}_k\} \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_k p_{\mathrm{x}_1 \cdots \mathrm{x}_n}(x_1, \cdots, x_n)\, dx_1 \cdots dx_n \tag{1.43}$$

and its second-order moment is given by

$$\mathcal{E}\{\mathrm{x}_k^2\} \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_k^2 p_{\mathrm{x}_1 \cdots \mathrm{x}_n}(x_1, \cdots, x_n)\, dx_1 \cdots dx_n\,. \tag{1.44}$$

Moments of higher order and central moments can be introduced in analogy to the definitions in Section 1.2.2. Joint moments and joint central moments can be defined as:

$$\mathcal{E}\{\mathrm{x}_k^\alpha \mathrm{x}_\ell^\beta\} \tag{1.45}$$

and

$$\mathcal{E}\{[\mathrm{x}_k - \mathcal{E}\{\mathrm{x}_k\}]^\alpha\, [\mathrm{x}_\ell - \mathcal{E}\{\mathrm{x}_\ell\}]^\beta\}\,, \tag{1.46}$$

respectively, where α and β are positive integers.

Notice that the characteristic function, of the jointly distributed r.v.'s, gives a convenient way of computing moments, just as it did in the scalar case. Taking the first derivative of the characteristic function (1.41) with respect to component u_k we have

$$\frac{\partial \phi_{\mathrm{x}_1 \cdots \mathrm{x}_n}(u_1, \cdots, u_n)}{\partial u_k} = i\mathcal{E}\{\mathrm{x}_k \exp\left(i \sum_{j=1}^{n} u_j \mathrm{x}_j\right)\}\,. \tag{1.47}$$

Evaluating this derivative at $(u_1, \cdots, u_n) = (0, \cdots, 0)$ provides a way to compute the first moment with respect to component x_k, that is,

$$\mathcal{E}\{\mathrm{x}_k\} = \frac{1}{i} \left.\frac{\partial \phi_{\mathrm{x}_1 \cdots \mathrm{x}_n}(u_1, \cdots, u_n)}{\partial u_k}\right|_{(u_1, \cdots, u_n) = (0, \cdots, 0)} \tag{1.48}$$

Successive n derivatives, with respect to arbitrary n components of $(u_1, \cdots, u_n)$ produce the n-th non-central moment

$$\mathcal{E}\{\mathrm{x}_k \mathrm{x}_l \cdots\} = \frac{1}{i^n} \left.\frac{\partial \phi_{\mathrm{x}_1 \cdots \mathrm{x}_n}(u_1, \cdots, u_n)}{\partial u_k \partial u_l \cdots}\right|_{(u_1, \cdots, u_n) = (0, \cdots, 0)}\,. \tag{1.49}$$

Of fundamental importance is the concept of covariance between x_k and x_ℓ, defined as:

$$cov(\mathrm{x}_k, \mathrm{x}_\ell) \equiv \mathcal{E}\{[\mathrm{x}_k - \mathcal{E}\{\mathrm{x}_k\}]\, [\mathrm{x}_\ell - \mathcal{E}\{\mathrm{x}_\ell\}]\}\,. \tag{1.50}$$

We have that

$$cov(\mathrm{x}_k, \mathrm{x}_\ell) = \mathcal{E}\{\mathrm{x}_k \mathrm{x}_\ell\} - \mathcal{E}\{\mathrm{x}_k\}\mathcal{E}\{\mathrm{x}_\ell\} \tag{1.51}$$

and also,

$$cov(\mathrm{x}_k, \mathrm{x}_k) = var(\mathrm{x}_k)\,. \tag{1.52}$$

The ratio

$$\rho(\mathrm{x}_k, \mathrm{x}_\ell) \equiv \frac{cov(\mathrm{x}_k, \mathrm{x}_\ell)}{\sigma(\mathrm{x}_k)\sigma(\mathrm{x}_\ell)} \tag{1.53}$$

defines the correlation coefficient between x_k and x_ℓ. Therefore, $\rho(\mathrm{x}_k, \mathrm{x}_k) = 1$.

It is of frequent interest to obtain the probability distribution or density function of a random variable, given its corresponding joint function. That is, consider two r.v.'s x_1 and x_2, jointly distributed, then

$$F_{\mathrm{x}_1}(x_1) = F_{\mathrm{x}_1\mathrm{x}_2}(x_1, \infty) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} p_{\mathrm{x}_1\mathrm{x}_2}(s_1, s_2)\, ds_1 ds_2\,, \tag{1.54}$$

and analogously, $F_{\mathrm{x}_2}(x_2) = F_{\mathrm{x}_1\mathrm{x}_2}(\infty, x_2)$, where $F_{\mathrm{x}_1}(x_1)$ and $F_{\mathrm{x}_2}(x_2)$, are referred to as marginal distribution functions. The marginal density function is then given by

$$p_{\mathrm{x}_1}(x_1) = \frac{\partial F_{\mathrm{x}_1\mathrm{x}_2}(x_1, \infty)}{\partial x_1} = \int_{-\infty}^{\infty} p_{\mathrm{x}_1\mathrm{x}_2}(x_1, x_2)\, dx_2\,. \tag{1.55}$$

It is convenient, at this point, to introduce a more compact notation utilizing vectors. Define the vector random variable (or simply the random vector) in n dimensions as:

$$\mathbf{x} = (\mathrm{x}_1 \mathrm{x}_2 \cdots \mathrm{x}_n)^T \tag{1.56}$$

where lower case bold letters refer to vectors, and T refers to the transposition operation. By analogy with the notation we have utilized up to here, we will refer to the value assumed by the random vector $\mathbf{x}$ as $\boldsymbol{x} = (x_1 x_2 \cdots x_n)^T$. In this manner,

$$p_{\mathbf{x}}(\boldsymbol{x}) \equiv p_{\mathrm{x}_1\mathrm{x}_2\cdots\mathrm{x}_n}(x_1, x_2, \cdots, x_n) \tag{1.57}$$

Likewise, the probability distribution can be written as

$$\begin{aligned} F_{\mathbf{x}}(\boldsymbol{x}) &\equiv \int_{-\infty}^{\boldsymbol{x}} p_{\mathbf{x}}(\boldsymbol{x}')d\boldsymbol{x}' \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_{\mathrm{x}_1\cdots\mathrm{x}_n}(x_1', \cdots, x_n')dx_1' \cdots dx_n'\,, \end{aligned} \tag{1.58}$$

where we call attention for the notation $d\boldsymbol{x} = d\mathrm{x}_1 \cdots d\mathrm{x}_n$, and similarly the probability density function becomes

$$p_{\mathbf{x}}(\boldsymbol{x}) = \frac{\partial^n F_{\mathbf{x}}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial^n F_{\mathbf{x}}(\boldsymbol{x})}{\partial x_1 \cdots \partial x_n} \tag{1.59}$$

The marginal probability density can be written as

$$p_{\mathrm{x}_k}(x_k) = \frac{\partial F_{\mathrm{x}_k}(x_k)}{\partial x_k} = \int_{-\infty}^{\infty} p_{\mathbf{x}}(\boldsymbol{x}')\, d\boldsymbol{x}'_{-k}\,, \tag{1.60}$$

where $d\boldsymbol{x}_{-k} = dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n$.

According to the definition of mean of a random variable, the mean of a random vector is given by the mean of its components:

$$\mathcal{E}\{\mathbf{x}\} = \begin{bmatrix} \mathcal{E}\{\mathrm{x}_1\} \\ \vdots \\ \mathcal{E}\{\mathrm{x}_n\} \end{bmatrix} = \begin{bmatrix} \int_{-\infty}^{\infty} x_1 p_{\mathbf{x}}(\boldsymbol{x}') d\boldsymbol{x}' \\ \vdots \\ \int_{-\infty}^{\infty} x_n p_{\mathbf{x}}(\boldsymbol{x}') d\boldsymbol{x}' \end{bmatrix} \tag{1.61}$$

Analogously, the mean of a random matrix is the mean of the matrix elements. The matrix formed by the mean of the outer product of the vector $\mathbf{x} - \mathcal{E}\{\mathbf{x}\}$ with itself is the $n \times n$ covariance matrix:

$$\begin{aligned} \mathbf{P}_{\mathbf{x}} &= \mathcal{E}\{(\mathbf{x} - \mathcal{E}\{\mathbf{x}\})(\mathbf{x} - \mathcal{E}\{\mathbf{x}\})^T\} \\ &= \begin{bmatrix} var(x_1) & cov(x_1, x_2) & \cdots & cov(x_1, x_n) \\ cov(x_2, x_1) & var(x_2) & \cdots & cov(x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_n, x_1) & cov(x_n, x_2) & \cdots & var(x_n) \end{bmatrix} . \end{aligned} \tag{1.62}$$

Notice that $\mathbf{P}_{\mathbf{x}}$ is a symmetric positive semi-definite matrix, that is, $\mathbf{y}\mathbf{P}_{\mathbf{x}}\mathbf{y}^T \geq \mathbf{0}$, for all $\mathbf{y} \in R^n$.

Two scalar r.v.'s x and y are said to be independent if any of the (equivalent) conditions are satisfied:

$$F_{\mathrm{xy}}(x, y) = F_{\mathrm{x}}(x) F_{\mathrm{y}}(y) \tag{1.63a}$$

$$p_{\mathrm{xy}}(x, y) = p_{\mathrm{x}}(x) p_{\mathrm{y}}(y) \tag{1.63b}$$

$$\mathcal{E}\{f(\mathrm{x}) g(\mathrm{y})\} = \mathcal{E}\{f(\mathrm{x})\} \mathcal{E}\{g(\mathrm{y})\} \tag{1.63c}$$

Analogously, two vector r.v.'s $\mathbf{x}$ and $\mathbf{y}$ are said to be jointly independent if

$$p_{\mathbf{xy}}(\boldsymbol{x}, \boldsymbol{y}) = p_{\mathbf{x}}(\boldsymbol{x}) p_{\mathbf{y}}(\boldsymbol{y}) \tag{1.64}$$

We say that two jointly distributed random vectors $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated if

$$cov(\mathbf{x}, \mathbf{y}) = \mathbf{0} \, , \tag{1.65}$$

since the correlation coefficient defined in (1.53) is null. As a matter of fact, two r.v.'s are said to be orthogonal when

$$\mathcal{E}\{\mathbf{x}\mathbf{y}^T\} = \mathbf{0} \, . \tag{1.66}$$

This equality is often referred to as the *orthogonality principle.*

The n r.v.'s $\{\mathrm{x}_1, \cdots, \mathrm{x}_n\}$ are said to be jointly Gaussian, or jointly normal, if their joint probability density function is given by

$$p_{\mathbf{x}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{P}|^{1/2}} \exp\left[-\frac{1}{2} (\boldsymbol{x} - \mu)^T \mathbf{P}^{-1} (\boldsymbol{x} - \mu) \right] , \tag{1.67}$$

where the notation $|\mathbf{P}|$ stands for the determinant of $\mathbf{P}$, and $\mathbf{P}^{-1}$ refers to the inverse of the matrix $\mathbf{P}$. The vector $\mathbf{x}$ is said to be normally distributed or Gaussian, with mean $\mu = \mathcal{E}\{\mathbf{x}\}$ and covariance $\mathbf{P}$, and is abbreviated by $\mathbf{x} \sim \mathcal{N}(\mu, \mathbf{P})$. Observe that, in order to simplify the notation, we temporarily eliminated the subscript $\mathbf{x}$ referring to the r.v. in question in $\boldsymbol{\mu}$ and $\mathbf{P}$.

Utilizing the vector notation, the joint characteristic function (1.41) can be written as:

$$\phi_{\mathbf{x}}(\mathbf{u}) = \mathcal{E}\{\ \exp(i\mathbf{u}^T\mathbf{x})\ \}\,. \tag{1.68}$$

In this way, the characteristic function of a normally distributed random vector can be calculated using the expression above and the transformation of variables $\boldsymbol{x} = \mathbf{P}^{1/2}\boldsymbol{y} + \boldsymbol{\mu}$, that is,

$$\begin{aligned} \phi_{\mathbf{x}}(\mathbf{u}) &\equiv \int_{-\infty}^{\infty} p_{\mathbf{x}}(\boldsymbol{x}) e^{i\mathbf{u}^T\boldsymbol{x}}\, d\boldsymbol{x} = \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}} \\ &\quad \times \int_{-\infty}^{\infty} \exp(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}) \exp\left[i\mathbf{u}^T(\mathbf{P}^{1/2}\boldsymbol{y} + \boldsymbol{\mu})\right] \, ||\mathrm{Jac}[\boldsymbol{x}(\boldsymbol{y})]||\, d\boldsymbol{y} \end{aligned} \tag{1.69}$$

where $||\mathrm{Jac}[\boldsymbol{x}(\boldsymbol{y})]||$ is the absolute value of the determinant of the Jacobian matrix, defined as

$$\mathrm{Jac}[\boldsymbol{x}(\boldsymbol{y})] \equiv \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}^T} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix} \tag{1.70}$$

of the transformation. Using the fact that

$$|\mathrm{Jac}[\boldsymbol{x}(\boldsymbol{y})]| = |\mathbf{P}^{1/2}| = |\mathbf{P}|^{1/2} \tag{1.71}$$

we can write

$$\phi_{\mathbf{x}}(\mathbf{u}) = \exp(i\mathbf{u}^T\boldsymbol{\mu}) \frac{1}{(2\pi)^{n/2}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\boldsymbol{y}^T\boldsymbol{y} - 2i\mathbf{u}^T\mathbf{P}^{1/2}\boldsymbol{y})\right] d\boldsymbol{y} \tag{1.72}$$

Adding and subtracting $(1/2)\mathbf{u}^T\mathbf{P}\mathbf{u}$ to complete the square in the integrand above, we obtain:

$$\begin{aligned} \phi_{\mathbf{x}}(\mathbf{u}) &= \exp(i\mathbf{u}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{u}^T\mathbf{P}\mathbf{u}) \\ &\quad \times \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}^T\boldsymbol{y} - 2i\mathbf{u}^T\mathbf{P}^{1/2}\boldsymbol{y} - \mathbf{u}^T\mathbf{P}\mathbf{u})\right] d\boldsymbol{y} \\ &= \exp(i\mathbf{u}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{u}^T\mathbf{P}\mathbf{u}) \\ &\quad \times \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\boldsymbol{y} - i\mathbf{P}^{1/2}\mathbf{u})^T(\boldsymbol{y} - i\mathbf{P}^{1/2}\mathbf{u})\right] d\boldsymbol{y} \end{aligned} \tag{1.73}$$

and making use of the integral

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right) d\boldsymbol{y} = \sqrt{(2\pi)^n} \tag{1.74}$$

we have that

$$\phi_{\mathrm{x}}(\mathbf{u}) = \exp(i\mathbf{u}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{u}^T\mathbf{P}\mathbf{u}), \tag{1.75}$$

is the characteristic function for a Gaussian distribution.

In the calculation of the integral above we defined the vector $\mathbf{y}$ as a function of the random vector $\mathbf{x}$ and transformed the integral in to a simpler integral. This gives an opportunity for us to mention a theorem relating functional transformation of random variable (vectors) and their respective probability distributions. Consider two n–dimensional random vectors $\mathbf{x}$ and $\mathbf{y}$ (not related to the characteristic function calculated above), that are related through a function $\mathbf{f}$ as $\mathbf{y} = \mathbf{f}(\mathbf{x})$, such that the inverse functional relation $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{y})$ exists. In this case, the probability density $p_{\mathrm{y}}(\boldsymbol{y})$ of $\mathbf{y}$ can be obtained given the probability density $p_{\mathrm{x}}(\boldsymbol{x})$ of $\mathbf{x}$ by the transformation:

$$p_{\mathrm{y}}(\boldsymbol{y}) \;=\; p_{\mathrm{x}}[\mathbf{f}^{-1}(\boldsymbol{y})]\,||\mathrm{Jac}([\mathbf{f}^{-1}(\boldsymbol{y})]|| \tag{1.76}$$

where $||\mathrm{Jac}([\mathbf{f}^{-1}(\boldsymbol{y})]||$ is the absolute value of the determinant of the Jacobian of the inverse transformation of $\mathbf{x}$ in to $\mathbf{y}$. A proof of this theorem is given in Jazwinski [84], pp. 34–35.

1.3.3 Conditional Expectations

Motivated by the conditional probability concept presented in Section 1.1.2, we now introduce the concept of conditional probability density. If $\mathbf{x}$ and $\mathbf{y}$ are random vectors, the probability density that the event $\mathbf{x}$ occurs given that the event $\mathbf{y}$ occurred is defined as

$$p_{\mathrm{x}|\mathrm{y}}(\boldsymbol{x}|\boldsymbol{y}) \equiv \frac{p_{\mathrm{xy}}(\boldsymbol{x},\boldsymbol{y})}{p_{\mathrm{y}}(\boldsymbol{y})}. \tag{1.77}$$

Analogously, reversing the meaning of $\mathbf{x}$ and $\mathbf{y}$,

$$p_{\mathrm{y}|\mathrm{x}}(\boldsymbol{y}|\boldsymbol{x}) \equiv \frac{p_{\mathrm{xy}}(\boldsymbol{x},\boldsymbol{y})}{p_{\mathrm{x}}(\boldsymbol{x})}, \tag{1.78}$$

and Bayes rule for probability densities immediately follows:

$$p_{\mathrm{x}|\mathrm{y}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{p_{\mathrm{y}|\mathrm{x}}(\boldsymbol{y}|\boldsymbol{x})p_{\mathrm{x}}(\boldsymbol{x})}{p_{\mathrm{y}}(\boldsymbol{y})}. \tag{1.79}$$

Based on the definition (1.77) we can define the conditional expectation (or mean) of an r.v. $\mathbf{x}$ given an r.v. $\mathbf{y}$ as:

$$\mathcal{E}\{\mathbf{x}|\mathbf{y}\} \equiv \int_{-\infty}^{\infty} \boldsymbol{x}\,p_{\mathrm{x}|\mathrm{y}}(\boldsymbol{x}|\boldsymbol{y})\,d\boldsymbol{x}. \tag{1.80}$$

Now remember that the unconditional mean is given by

$$\mathcal{E}\{\mathbf{x}\} = \int_{-\infty}^{\infty} \boldsymbol{x}\,p_{\mathrm{x}}(\boldsymbol{x})\,d\boldsymbol{x}, \tag{1.81}$$

and that the marginal probability density $p_{\mathbf{x}}(\boldsymbol{x})$ can be obtained from the joint probability density $p_{\mathbf{xy}}(\boldsymbol{x},\boldsymbol{y})$ as,

$$p_{\mathbf{x}}(\boldsymbol{x}) = \int_{-\infty}^{\infty} p_{\mathbf{xy}}(\boldsymbol{x},\boldsymbol{y})\, d\boldsymbol{y}\,. \tag{1.82}$$

Considering the definition (1.77) we can write,

$$p_{\mathbf{x}}(\boldsymbol{x}) = \int_{-\infty}^{\infty} p_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y}) p_{\mathbf{y}}(\boldsymbol{y})\, d\boldsymbol{y} \tag{1.83}$$

and substituting this result in (1.81) we have that

$$\begin{aligned} \mathcal{E}\{\mathbf{x}\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{x} p_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y}) p_{\mathbf{y}}(\boldsymbol{y})\, d\boldsymbol{y} d\boldsymbol{x} \\ &= \int_{-\infty}^{\infty} \mathcal{E}\{\mathbf{x}|\boldsymbol{y}\} p_{\mathbf{y}}(\boldsymbol{y})\, d\boldsymbol{y} \\ &= \mathcal{E}\{\mathcal{E}\{\mathbf{x}|\mathbf{y}\}\}\,, \end{aligned} \tag{1.84}$$

where we used definition (1.80) of conditional expectation. The expression above is sometimes referred to as the *chain rule* for conditional expectations. Analogously we can obtain:

$$\mathcal{E}\{f(\mathbf{x},\mathbf{y})\} = \mathcal{E}\{\mathcal{E}\{f(\mathbf{x},\mathbf{y})|\mathbf{y}\}\}\,. \tag{1.85}$$

We can also define the conditional covariance matrix as

$$\mathbf{P}_{\mathbf{x}|\mathbf{y}} \boxminus \mathcal{E}\{[\mathbf{x} - \mathcal{E}\{\mathbf{x}|\mathbf{y}\}][\mathbf{x} - \mathcal{E}\{\mathbf{x}|\mathbf{y}\}]^T|\mathbf{y}\}\,, \tag{1.86}$$

where we notice that $\mathbf{P}_{\mathbf{x}|\mathbf{y}}$ is a random matrix, contrary to what we encountered when we defined the unconditional covariance matrix (1.62).

We will now prove the following important result for normally distributed r.v.'s: the conditional probability of two normally distributed random vectors $\mathbf{x}$ and $\mathbf{y}$, with dimensions n and m respectively, is also normal and is given by:

$$p_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}_{\mathbf{x}|\mathbf{y}}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}})^T \mathbf{P}_{\mathbf{x}|\mathbf{y}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}})\right]\,, \tag{1.87}$$

where

$$\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} = \boldsymbol{\mu}_{\mathbf{x}} + \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathbf{y}})\,, \tag{1.88}$$

and

$$\mathbf{P}_{\mathbf{x}|\mathbf{y}} = \mathbf{P}_{\mathbf{x}} - \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{y}}^{-1}\mathbf{P}_{\mathbf{xy}}^T\,. \tag{1.89}$$

Now consider the following vector $\mathbf{z} = [\mathbf{x}^T\, \mathbf{y}^T]^T$ of dimension $(n+m)$. This vector has mean $\boldsymbol{\mu}_{\mathbf{z}}$ given by

$$\boldsymbol{\mu}_{\mathbf{z}} = \mathcal{E}\{\mathbf{z}\} = \begin{bmatrix} \mathcal{E}\{\mathbf{x}\} \\ \mathcal{E}\{\mathbf{y}\} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\mu}_{\mathbf{y}} \end{bmatrix} \tag{1.90}$$

and covariance $\mathbf{P}_{\mathbf{z}}$ that can be written as

$$\begin{aligned} \mathbf{P}_{\mathbf{z}} &= \mathcal{E}\{(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})^T\} \\ &= \begin{bmatrix} \mathcal{E}\{(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T\} & \mathcal{E}\{(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})^T\} \\ \mathcal{E}\{(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T\} & \mathcal{E}\{(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})^T\} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}_{\mathbf{x}} & \mathbf{P}_{\mathbf{xy}} \\ \mathbf{P}_{\mathbf{xy}}^T & \mathbf{P}_{\mathbf{y}} \end{bmatrix}\,. \end{aligned} \tag{1.91}$$

Let us make use of the following equality (simple to verify):

$$\begin{aligned}
\begin{bmatrix} \mathbf{I} & -\mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{P}_{\mathrm{z}} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T} & \mathbf{I} \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & -\mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\
&\quad \times \begin{bmatrix} \mathbf{P}_{\mathrm{x}} - \mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T} & \mathbf{P}_{\mathrm{xy}} \\ \mathbf{0} & \mathbf{P}_{\mathrm{y}} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{P}_{\mathrm{x|y}} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{\mathrm{y}} \end{bmatrix}, \qquad (1.92)
\end{aligned}$$

where $\mathbf{P}_{\mathrm{x|y}}$ is defined as in (1.89), and we are assuming that $\mathbf{P}_{\mathrm{y}}^{-1}$ exists. From this expression, it follows that the determinant of the covariance matrix $\mathbf{P}_{\mathrm{z}}$ is

$$\begin{aligned}
|\mathbf{P}_{\mathrm{z}}| &= |\mathbf{P}_{\mathrm{x|y}}|\,|\mathbf{P}_{\mathrm{y}}| \\
&= |\mathbf{P}_{\mathrm{x}} - \mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T}|\,|\mathbf{P}_{\mathrm{y}}| \qquad (1.93)
\end{aligned}$$

(Householder [83], p. 17). Moreover, we have that

$$\begin{aligned}
\mathbf{P}_{\mathrm{z}}^{-1} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{P}_{\mathrm{x|y}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{\mathrm{y}}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{P}_{\mathrm{x|y}}^{-1} & -\mathbf{P}_{\mathrm{x|y}}^{-1}\mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1} \\ -\mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T}\mathbf{P}_{\mathrm{x|y}}^{-1} & \mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T}\mathbf{P}_{\mathrm{x|y}}^{-1}\mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1} + \mathbf{P}_{\mathrm{y}}^{-1} \end{bmatrix} \qquad (1.94)
\end{aligned}$$

Therefore, multiplying $\mathbf{P}_{\mathrm{z}}^{-1}$ by $(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{z}})^T$ on the left and by $(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{z}})$ on the right, we have

$$\begin{aligned}
(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{z}})^T\mathbf{P}_{\mathrm{z}}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{z}}) &= (\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}})^T\mathbf{P}_{\mathrm{x|y}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}}) \\
&- (\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}})^T\mathbf{P}_{\mathrm{x|y}}^{-1}\mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}}) \\
&- (\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}})^T\mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T}\mathbf{P}_{\mathrm{x|y}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}}) \\
&+ (\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}})^T\mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T}\mathbf{P}_{\mathrm{x|y}}^{-1}\mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}}) \\
&+ (\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}})^T\mathbf{P}_{\mathrm{y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}}) \qquad (1.95)
\end{aligned}$$

and using the definition (1.88) we can write

$$\begin{aligned}
(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x|y}})^T\mathbf{P}_{\mathrm{x|y}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x|y}}) &= [(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}}) - \mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}})]^T\mathbf{P}_{\mathrm{x|y}}^{-1} \\
&\quad \times [(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}}) - \mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}})] \\
&= (\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}})^T\mathbf{P}_{\mathrm{x|y}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}}) \\
&- (\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}})^T\mathbf{P}_{\mathrm{x|y}}^{-1}\mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}}) \\
&- (\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}})^T\mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T}\mathbf{P}_{\mathrm{x|y}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x}}) \\
&+ (\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}})^T\mathbf{P}_{\mathrm{y}}^{-1}\mathbf{P}_{\mathrm{xy}}^{T}\mathbf{P}_{\mathrm{x|y}}^{-1}\mathbf{P}_{\mathrm{xy}}\mathbf{P}_{\mathrm{y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}}) \\
&\qquad (1.96)
\end{aligned}$$

so that (1.95) reduces to

$$(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{z}})^T\mathbf{P}_{\mathrm{z}}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{z}}) = (\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x|y}})^T\mathbf{P}_{\mathrm{x|y}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{x|y}}) + (\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}})^T\mathbf{P}_{\mathrm{y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{y}}) \qquad (1.97)$$

By the definition of conditional probability we have

$$\begin{aligned} p_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y}) &= \frac{p_{\mathbf{xy}}(\boldsymbol{x},\boldsymbol{y})}{p_{\mathbf{y}}(\boldsymbol{y})} = \frac{p_{\mathbf{z}}(\boldsymbol{z})}{p_{\mathbf{y}}(\boldsymbol{y})} \\ &= \frac{1}{(2\pi)^{n/2}} \frac{|\mathbf{P}_{\mathbf{y}}|^{1/2}}{|\mathbf{P}_{\mathbf{z}}|^{1/2}} \frac{\exp[-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_{\mathbf{z}})^T\mathbf{P}_{\mathbf{z}}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_{\mathbf{z}})]}{\exp[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_{\mathbf{y}})^T\mathbf{P}_{\mathbf{y}}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\mathbf{y}})]} \end{aligned} \tag{1.98}$$

and utilizing (1.93) and (1.97) we obtain

$$\begin{aligned} p_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y}) &= \frac{1}{(2\pi)^{n/2}|\mathbf{P}_{\mathbf{x}} - \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{y}}^{-1}\mathbf{P}_{\mathbf{xy}}^T|^{1/2}} \\ &\quad \times \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}})^T[\mathbf{P}_{\mathbf{x}} - \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{y}}^{-1}\mathbf{P}_{\mathbf{xy}}^T]^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}})] \end{aligned} \tag{1.99}$$

which is the desired result. The assumption made above that the inverse of $\mathbf{P}_{\mathbf{y}}$ exists is not necessary. When this inverse does not exist, it is possible to show (Kalman [89]) that the same result is still valid, but in place of the inverse of $\mathbf{P}_{\mathbf{y}}$, we should utilize the pseudo–inverse $\mathbf{P}_{\mathbf{y}}^{+}$.

The calculation above involved construction of the joint probability distribution $p_{\mathbf{z}}(\boldsymbol{z})$ of the random vector $\mathbf{z} = [\mathbf{x}^T\mathbf{y}^T]^T$. Let us assume for the moment that the two random vectors $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated, that is, $\mathbf{P}_{\mathbf{xy}} = \mathbf{0}$. Hence, referring back to (1.88) and (1.89) it follows that,

$$\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} = \boldsymbol{\mu}_{\mathbf{x}} \tag{1.100}$$

$$\mathbf{P}_{\mathbf{x}|\mathbf{y}} = \mathbf{P}_{\mathbf{x}} \tag{1.101}$$

which is intuitively in agreement with the notion of independence. Introducing these results in (1.97) we have

$$(\boldsymbol{z}-\boldsymbol{\mu}_{\mathbf{z}})^T\mathbf{P}_{\mathbf{z}}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_{\mathbf{z}}) = (\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{x}})^T\mathbf{P}_{\mathbf{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{x}}) + (\boldsymbol{y}-\boldsymbol{\mu}_{\mathbf{y}})^T\mathbf{P}_{\mathbf{y}}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\mathbf{y}}) \tag{1.102}$$

Moreover, it follows from (1.93) that for uncorrelated random vectors $\mathbf{x}$ and $\mathbf{y}$,

$$|\mathbf{P}_{\mathbf{z}}| = |\mathbf{P}_{\mathbf{x}}|\,|\mathbf{P}_{\mathbf{y}}| \tag{1.103}$$

Thus, the joint probability density function $p_{\mathbf{z}}(\boldsymbol{z})$ can then be written as

$$\begin{aligned} p_{\mathbf{z}}(\boldsymbol{z}) &= \frac{1}{(2\pi)^{(n+m)/2}}\frac{1}{|\mathbf{P}_{\mathbf{z}}|^{1/2}}\exp[-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_{\mathbf{z}})^T\mathbf{P}_{\mathbf{z}}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_{\mathbf{z}})] \\ &= \frac{1}{(2\pi)^{(n+m)/2}}\frac{1}{|\mathbf{P}_{\mathbf{x}}|^{1/2}|\mathbf{P}_{\mathbf{y}}|^{1/2}} \\ &\quad \times\exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{x}})^T\mathbf{P}_{\mathbf{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{x}}) - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_{\mathbf{y}})^T\mathbf{P}_{\mathbf{y}}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\mathbf{y}})] \\ &= \frac{1}{(2\pi)^{n/2}}\frac{1}{|\mathbf{P}_{\mathbf{x}}|^{1/2}}\exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{x}})^T\mathbf{P}_{\mathbf{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathbf{x}})] \\ &\quad \times\frac{1}{(2\pi)^{m/2}}\frac{1}{|\mathbf{P}_{\mathbf{y}}|^{1/2}}\exp[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_{\mathbf{y}})^T\mathbf{P}_{\mathbf{y}}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\mathbf{y}})] \\ &= p_{\mathbf{x}}(\boldsymbol{x})p_{\mathbf{y}}(\boldsymbol{y}) \end{aligned} \tag{1.104}$$

This shows that two normally distributed random vectors that are uncorrelated are also independent. We have seen earlier in this section that independence among random variables implied they are uncorrelated; the contrary was not necessarily true. However, as we have just shown, the contrary is true in the case of normally distributed random variables (vectors).

Exercises

1. Using the definition of Rayleigh probability density function given in (1.12): (a) calculate the mean and standard deviation for a r.v. with that distribution; (b) find the mode of the r.v., that is, is most likely value.

2. (Brown [19], Problem 1.40) A pair of random variables, x e y, have the following joint probability density function:

$$p_{\mathrm{xy}}(x,y) = \begin{cases} 1 & 0 \le y \le 2x \text{ e } 0 \le x \le 1 \\ 0 & \text{em everywhere else} \end{cases}$$

   Find $\mathcal{E}\{\mathrm{x}|\mathrm{y} = .5\}$. [Hint: Use (1.77) to find $p_{\mathrm{x}|\mathrm{y}}(x)$ for $y = 0.5$, and then integrate $xp_{\mathrm{x}|\mathrm{y}}(x)$ to find $\mathcal{E}\{\mathrm{x}|\mathrm{y} = .5\}$.]

3. Consider a zero–mean Gaussian random vector, with probability density and characteristic functions

$$f_{\mathbf{x}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}} \exp\left[-\frac{1}{2}\boldsymbol{x}^T\mathbf{P}^{-1}\boldsymbol{x}\right],$$

$$\phi_{\mathbf{x}}(\mathbf{u}) = \exp\left[-\frac{1}{2}\mathbf{u}^T\mathbf{P}\mathbf{u}\right],$$

   respectively. Show that the following holds for the first four moments of this distribution:

$$\begin{array}{ll} \mathcal{E}\{\mathrm{x}_k\} = 0 & \mathcal{E}\{\mathrm{x}_k\mathrm{x}_l\} = P_{kl} \\ \mathcal{E}\{\mathrm{x}_k\mathrm{x}_l\mathrm{x}_m\} = 0 & \mathcal{E}\{\mathrm{x}_k\mathrm{x}_l\mathrm{x}_m\mathrm{x}_n\} = P_{kl}P_{mn} + P_{km}P_{ln} + P_{kn}P_{lm} \end{array}$$

   where x_i, $i \in \{k, l, m, n\}$, are elements of the random vector x, and P_{ij}, $i, j \in \{k, l, m, n\}$, are elements of $\mathbf{P}$.

4. Show that the linear transformation of a normally distributed vector is also normally distributed. That is, show that for a given normally distributed vector $\mathbf{x}$, with mean $\mu_{\mathbf{x}}$ and covariance $\mathbf{R}_{\mathbf{x}}$, the linear transformation

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

   produces a normally distributed vector $\mathbf{y}$ with mean $\mu_{\mathbf{y}} = \mathbf{A}\mu_{\mathbf{x}} + \mathbf{b}$ and covariance $\mathbf{R}_{\mathbf{y}} = \mathbf{A}\mathbf{R}_{\mathbf{x}}\mathbf{A}^T$.

5. The log–normal distribution is defined by

$$p_{\mathrm{x}}(x) = \frac{1}{\sqrt{2\pi}s}\frac{1}{x}\exp\left[-\frac{1}{2s^2}\left(\ln\frac{x}{x_0}\right)^2\right]$$

   Show that:

(a) its mean and variance are

$$\mu = x_0 e^{s^2/2}$$
$$var(x) = x_0^2 e^{s^2}(e^{s^2} - 1)$$

respectively;

(b) introducing the variable

$$x^* = \beta \ln(\frac{x}{\gamma})$$

the probability density function $p_{\mathrm{X}}(x)$ above can be converted to a Gaussian probability density function $p_{\mathrm{X}}^*(x)$ of the form

$$p_{\mathrm{X}}^*(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x^* - x_0^*)^2}{2\sigma^2} \right]$$

where $\sigma = s\beta$, and

$$x_0^* = \beta \ln(\frac{x_0}{\gamma})$$

This justifies the name log–normal distribution for $p_{\mathrm{X}}(x)$, since the logarithm of its variable is normally distributed.

6. According to what we have seen in the previous exercise, let $\mathbf{v}$ an n–dimensional normally distributed r.v., defined as $\mathbf{v} \sim \mathcal{N}(\mu_{\mathbf{v}}, \mathbf{P})$. The vector with components w_j defined as $w_j = \exp(v_j)$ for $j = 1, \cdots, n$ is said to be distributed log–normally and it is represented by $\mathbf{w} \sim \mathcal{LN}(\mu_{\mathbf{w}}, \mathbf{R})$ where $\mu_{\mathbf{w}}$ is its mean and $\mathbf{R}$ its covariance. Show that

$$\mathcal{E}\{\mathrm{w}_j\} = \exp\left[\mathcal{E}\{\mathrm{v}_j\} + \frac{1}{2} P_{jj} \right] ,$$

and

$$R_{jk} = \mathcal{E}\{\mathrm{w}_j\}\mathcal{E}\{\mathrm{w}_k\}(e^{P_{jk}} - 1) .$$

(Hint: Utilize the concept of characteristic function.)

7. *Computer Assignment:* (Based on Tarantola [126]) Consider the experiment of measuring (estimating) the value of a constant quantity corrupted by "noise". To simulate this situation, let us use Matlab, to generate 101 measurements of the random variable y as follows:

   Enter: `y = 21 + rand(101,1)`; the intrinsic Matlab function *rand* generates a uniformly distributed r.v. in the interval $(0, 1)$

   Enter: `x=20:0.1:23`; to generate an array with 31 points in the neighborhood of 21

   Enter: `hist(y,x)`; this will show you a histogram corresponding to this experiment

   Now using the Matlab functions *median, mean, max* and *min*, calculate the median, mean and mid–range values for the experiment you have just performed. What did you get? Three distinct values! Can you tell which of these are the closest value to the true value?

8. *Computer Assignment:* Ok, you probably still can't answer the question above. So here is the real assignment:

(a) Construct a Matlab function that repeats the experiment of the previous exercise 20 times, for the given value of the scalar under noise. For each successive experiment, increase the number of samples used by 100, calculating and storing the values their corresponding median, mean, and mid–range. At the end of the 20 experiments, plot the values obtained for the median, mean and mid–range in each experiment. Can you guess now which one of these is the best estimate?

(b) To really confirm your guess, fix the number of samples at 100 and repeat the experiments 200 times, collecting the corresponding median, mean and mid–range values for each experiment. (It is a good idea, if you do it as another Matlab function.) In end of all 200 experiments, plot the histograms for each of these three quantities. Which one has the least scatter? Is this compatible with your guess from of the previous item?

(c) Repeat items (a) and (b) for the same constant, but now being disturbed by a normally distributed random variable with mean zero and unity variance. That is, replace the Matlab function *rand* by the function *randn*. *Caution: when construction the histogram in this item chose a relatively large interval for the outcome counting, e.g.,* $\boxed{x=18:0.1:24}$. *This is necessary because the Gaussian function has very long tails.*

Chapter 2

Stochastic Processes and Random Fields

A topic of intrinsic interest in this course is stochastic partial differential equations (SPDE). Although a rigorous treatment of this topic goes beyond our goals, in order to introduce the fundamental ideas of the basic theory of SPDE, it is necessary to discuss the basic concepts of stochastic processes and random fields. These two concepts are nothing more than extensions of the random variable concept, treated in the previous lecture, for cases in which these variables have temporal or spatial dependence, respectively. As a matter of fact there is much similarity between the two concepts at a fundamental level, with some particular nomenclature differences.

2.1 Definition and Probabilistic Concepts

In the previous lecture, the symbol $\boldsymbol{x}(\omega)$ referred to the value of a vector random variable $\mathbf{x}$, resulting from the realization of an experiment ω. Stochastic processes are those in which the random variable is also a function of time, that is, the random variables are defined on the product space $\Omega \times T$, where T represents the real time line. In this case, we denote by $\boldsymbol{x}(\omega, t)$ the result of a stochastic process $\mathbf{x}(t)$. In what follows, we utilize the most common abbreviation of a stochastic process, denoting the stochastic variables as $\boldsymbol{x}(t)$, where ω will be implicit in the notation. Stochastic processes are referred to as *discrete-time* or *continuous-time* depending whether the time domain is discrete or continuous, respectively.

In stochastic processes, an event B in the probability space is denoted by

$$B = \{\omega \in \Omega : \mathbf{x}(\omega, t) \leq \boldsymbol{x}\} . \qquad (2.1)$$

The distribution function for a discrete-time process with N random n-vectors $\mathbf{x}(t_1), \cdots, \mathbf{x}(t_N)$, is defined as:

$$F_{\mathbf{x}(t_1)\cdots\mathbf{x}(t_N)}[\boldsymbol{x}(t_1), \cdots, \boldsymbol{x}(t_N)] \equiv P(\{\omega \in \Omega : \mathbf{x}(\omega, t_1) \leq \boldsymbol{x}_1, \cdots, \mathbf{x}(\omega, t_N) \leq \boldsymbol{x}_N\}) . \qquad (2.2)$$

with probability density function (if it exits):

$$p_{\mathbf{x}(t_1),\cdots,\mathbf{x}(t_N)}[\boldsymbol{x}(t_1),\cdots,\boldsymbol{x}(t_N)] = \frac{\partial^{nN} F_{\mathbf{x}(t_1)\cdots\mathbf{x}(t_N)}[\boldsymbol{x}(t_1),\cdots,\boldsymbol{x}(t_N)]}{\partial\boldsymbol{x}(t_1)\cdots\partial\boldsymbol{x}(t_N)} \tag{2.3}$$

where we recall that $\partial^n/\partial\boldsymbol{x} = \partial^n/\partial x_1 \cdots \partial x_n$. Consequently, we can write

$$F_{\mathbf{x}(t_1)\cdots\mathbf{x}(t_N)}[\boldsymbol{x}(t_1),\cdots,\boldsymbol{x}(t_N)] = \\ \int_{-\infty}^{\boldsymbol{x}(t_1)} \cdots \int_{-\infty}^{\boldsymbol{x}(t_N)} p_{\mathbf{x}(t_1)\cdots\mathbf{x}(t_N)}(\boldsymbol{x}'_1,\cdots,\boldsymbol{x}'_N)\, d\boldsymbol{x}'_1 \cdots d\boldsymbol{x}'_N\,. \tag{2.4}$$

In case of a continuous–time process, the probability distribution and probability density functions are defined for all times t, and can be symbolically written as

$$\begin{aligned} F_{\mathbf{x}}(\boldsymbol{x},t) &\equiv P(\{\omega \in \Omega : \mathbf{x}(\omega,t) \leq \boldsymbol{x}\}) && (2.5a) \\ p_{\mathbf{x}}(\boldsymbol{x},t) &= \frac{\partial^n F_{\mathbf{x}}(\boldsymbol{x},t)}{\partial\boldsymbol{x}} && (2.5b) \end{aligned}$$

respectively.

The concepts of mean, variance and correlation introduced in the previous lecture can be extended directly to the case of stochastic processes. Therefore, we define concisely these quantities for this case:

- *Mean vector*:

$$\mu_{\mathbf{x}}(t) \equiv \mathcal{E}\{\mathbf{x}(t)\} = \int_{-\infty}^{\infty} \boldsymbol{x}\, p_{\mathbf{x}(t)}(\boldsymbol{x})\, d\boldsymbol{x}\,, \tag{2.6}$$

- *Stationary mean value vector*: defined when the mean is independent of time, that is

$$\bar{\mathbf{x}} \equiv \lim_{t_f\to\infty} \frac{1}{2t_f} \int_{-t_f}^{t_f} \boldsymbol{x}(t)\, dt\,. \tag{2.7}$$

  For the case in which the stationary mean value coincides with the ensemble mean μ, the process is called ergodic in the mean.

- *Mean for discrete–time processes*:

$$\bar{\mathbf{x}} \equiv \lim_{K\to\infty} \frac{1}{2K+1} \sum_{k=-K}^{K} \boldsymbol{x}(kT) \tag{2.8}$$

  where T is the sampling period.

- *Quadratic mean value matrix*:

$$\mathbf{\Gamma}_{\mathbf{x}}(t) \equiv \mathcal{E}\{\mathbf{x}(t)\mathbf{x}(t)^T\} = \int_{-\infty}^{\infty} \boldsymbol{x}\,\boldsymbol{x}^T\, p_{\mathbf{x}(t)}(\boldsymbol{x})\, d\boldsymbol{x}\,. \tag{2.9}$$

  We can still define the stationary quadratic mean value based on the definition of stationary mean value, as we can define the stationary quadratic mean value for a discrete–time process utilizing the corresponding definition given above.

- *Auto-correlation matrix*:
$$\mathbf{\Gamma}_{\mathbf{x}}(t,\tau) \equiv \mathcal{E}\{\mathbf{x}(t)\mathbf{x}^T(\tau)\}\,, \tag{2.10}$$
  or explicitly written,
$$\mathbf{\Gamma}_{\mathbf{x}}(t,\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{z}\,\boldsymbol{y}^T\, p_{\mathbf{x}(t)\mathbf{x}(\tau)}(\boldsymbol{z},\boldsymbol{y})\; d\boldsymbol{z}d\boldsymbol{y}\,. \tag{2.11}$$
  where the word *auto* refers to the stochastic process $\mathbf{x}(t)$.

- *Cross-correlation matrix*:
$$\mathbf{\Gamma}_{\mathbf{xy}}(t,\tau) \equiv \mathcal{E}\{\mathbf{x}(t)\mathbf{y}^T(\tau)\}\,. \tag{2.12}$$
  where the word *cross* refers to the two stochastic processes $\mathbf{x}(t)$ and $\mathbf{y}(t)$.

- *Auto-covariance matrix of a stochastic process*:
$$\mathbf{C}_{\mathbf{x}}(t,\tau) = cov\{\mathbf{x}(t),\mathbf{x}(\tau)\} \equiv \mathcal{E}\{[\mathbf{x}(t)-\boldsymbol{\mu}_{\mathbf{x}}(t)][\mathbf{x}(\tau)-\boldsymbol{\mu}_{\mathbf{x}}(\tau)]^T\}\,, \tag{2.13}$$
  where the designation *auto* refers to the stochastic process in question, in this case, only $\mathbf{x}(t)$. It is simple to show that
$$\mathbf{C}_{\mathbf{x}}(t,\tau) = \mathbf{\Gamma}_{\mathbf{x}}(t,\tau) - \mu_{\mathbf{x}}(t)\mu_{\mathbf{x}}^T(\tau) \tag{2.14}$$
  When $t=\tau$, we define the covariance matrix as:
$$\mathbf{P}_{\mathbf{x}}(t) \equiv \mathbf{C}_{\mathbf{x}}(t,t)\,, \tag{2.15}$$
  which is sometimes referred to as the variance matrix.

- *Cross-covariance matrix of a stochastic process*:
$$\mathbf{C}_{\mathbf{xy}}(t,\tau) \equiv \mathcal{E}\{[\mathbf{x}(t)-\boldsymbol{\mu}_{\mathbf{x}}(t)][\mathbf{y}(\tau)-\boldsymbol{\mu}_{\mathbf{y}}(\tau)]^T\}\,, \tag{2.16}$$
  where the designation *cross* refers to the stochastic processes $\mathbf{x}(t)$ and $\mathbf{y}(t)$. We can easily show that
$$\mathbf{C}_{\mathbf{xy}}(t,\tau) = \mathbf{\Gamma}_{\mathbf{xy}}(t,\tau) - \mu_{\mathbf{x}}(t)\mu_{\mathbf{y}}^T(\tau) \tag{2.17}$$

Correlation matrices can be defined analogously to the definitions given in the previous lecture.

2.2 Independent Process

We say that a stochastic process $\mathbf{x}(t)$ is independent when for all t and τ the probability density $p_{\mathbf{x}(t),\mathbf{x}(\tau)}(\boldsymbol{x}(t),\boldsymbol{x}(\tau)) = p_{\mathbf{x}(t)}p_{\mathbf{x}(\tau)}$. In this way, according to (2.11) it follows that

$$\begin{aligned}
\mathbf{\Gamma}_{\mathbf{x}}(t,\tau) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{z}\,\boldsymbol{y}^T\, p_{\mathbf{x}(t)}(\boldsymbol{z})p_{\mathbf{x}(\tau)}(\boldsymbol{y})\; d\boldsymbol{z}d\boldsymbol{y}\,. \\
&= \int_{-\infty}^{\infty} d\boldsymbol{z}\,\boldsymbol{z}\,p_{\mathbf{x}(t)}(\boldsymbol{z})\int_{-\infty}^{\infty} d\boldsymbol{y}\,\boldsymbol{y}^T\,p_{\mathbf{x}(\tau)}(\boldsymbol{y}) \\
&= \mathcal{E}\{\mathbf{x}(t)\}\,\mathcal{E}\{\mathbf{x}^T(\tau)\}
\end{aligned} \tag{2.18}$$

Therefore, from the definition of auto–covariance matrix it follows that $\mathbf{C}_{\mathbf{x}}(t, \tau) = \mathbf{0}$, for $t \neq \tau$, that is, an independent stochastic process is uncorrelated in time. In an entirely analogous way, we can show that if two stochastic processes $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are independent, they are also uncorrelated, that is, $\mathbf{C}_{\mathbf{xy}}(t, \tau) = \mathbf{0}$, for any t and τ. As in the case of random variables, the contrary of this relation is not necessarily true, that is, two uncorrelated processes are not necessarily independent.

2.3 Markov Process

As in stochastic processes in general, a Markov process can be continuous or discrete depending on whether the time parameter is continuous or discrete, respectively. A discrete stochastic process (i.e., stochastic sequence) $\{\mathbf{x}(t_k)\}$, for $t_k > t_0$, or a continuous stochastic process $\mathbf{x}(t)$, for $t > t_0$, is said to be a Markov process if, for all $\tau \leq t$,

$$p_{\mathbf{x}(t)|\Xi(\tau)}[\boldsymbol{x}(t)|\xi(\tau)] = p_{\mathbf{x}(t)|\mathbf{x}(\tau)}[\boldsymbol{x}(t)|\boldsymbol{x}(\tau)] \tag{2.19}$$

where $\Xi(\tau) = \{\mathbf{x}(s), t_0 \leq s \leq \tau \leq t\}$, and analogously $\xi(\tau) = \{\boldsymbol{x}(s), t_0 \leq s \leq \tau \leq t\}$. More specifically for the discrete case, a first–order Markov process, also referred to as a Markov–1 process, is one for which

$$p_{\mathbf{x}_k|\mathbf{x}_{k-1}\cdots\mathbf{x}_1\mathbf{x}_0}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \cdots, \boldsymbol{x}_1, \boldsymbol{x}_0) = p_{\mathbf{x}_k|\mathbf{x}_{k-1}}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) \tag{2.20}$$

That is to say, a Markov–1 process is one for which the probability density at time t, given all states up to t, in the interval $[t_0, \tau]$, depends only on the state at the final time, τ, of the interval. This is nothing more than a way of stating the causality principle: the state of a process at a particular moment in time is sufficient for us to determine the future states of the process, without us having to know its complete history.

In the discrete case, we can write for the joint probability density

$$\begin{aligned} p_{\Xi_k}(\xi_k) &= p_{\mathbf{x}_k\cdots\mathbf{x}_1\mathbf{x}_0}(\boldsymbol{x}_k, \cdots, \boldsymbol{x}_1, \boldsymbol{x}_0) \\ &= p_{\mathbf{x}_k|\mathbf{x}_{k-1}\cdots\mathbf{x}_1\mathbf{x}_0}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \cdots, \boldsymbol{x}_1, \boldsymbol{x}_0)\, p_{\mathbf{x}_{k-1}\cdots\mathbf{x}_1\mathbf{x}_0}(\boldsymbol{x}_{k-1}, \cdots, \boldsymbol{x}_1, \boldsymbol{x}_0) \end{aligned} \tag{2.21}$$

where we utilize the property (1.77). Assuming that the stochastic process is first–order Markov, according to (2.20) we have that

$$p_{\Xi_k}(\xi_k) = p_{\mathbf{x}_k|\mathbf{x}_{k-1}}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})\, p_{\mathbf{x}_{k-1}\cdots\mathbf{x}_1\mathbf{x}_0}(\boldsymbol{x}_{k-1}, \cdots, \boldsymbol{x}_1, \boldsymbol{x}_0) \tag{2.22}$$

and utilizing repeatedly the definition (2.20) we obtain

$$p_{\Xi_k}(\xi_k) = p_{\mathbf{x}_k|\mathbf{x}_{k-1}}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})\, p_{\mathbf{x}_{k-1}|\mathbf{x}_{k-2}}(\boldsymbol{x}_{k-1}|\boldsymbol{x}_{k-2}) \cdots p_{\mathbf{x}_1|\mathbf{x}_0}(\boldsymbol{x}_1|\boldsymbol{x}_0)\, p_{\mathbf{x}_0}(\boldsymbol{x}_0) \tag{2.23}$$

Therefore, the joint probability density of a Markov–1 process can be determined from the initial marginal probability density $p_{\mathbf{x}(0)}[\boldsymbol{x}(0)]$, and from the probability density $p_{\mathbf{x}(t)|\mathbf{x}(s)}[\boldsymbol{x}(t)|\boldsymbol{x}(s)]$, for $t \geq s \in [t_0, \tau)$, and $t < \tau$.The quantity $p_{\mathbf{x}(t)|\mathbf{x}(s)}[\boldsymbol{x}(t)|\boldsymbol{x}(s)]$ is known as the *transition probability density* of a Markov process.

The concept of Markovian process can be extended to define Markov processes of different orders. For example, a discrete stochastic process for which the probability density at time t_k depends on the process at times t_{k-1} and t_{k-2} can be defined as those for which we have:

$$p_{\mathbf{x}_k|\mathbf{x}_{k-1}\cdots\mathbf{x}_1\mathbf{x}_0}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}\cdots,\boldsymbol{x}_0) = p_{\mathbf{x}_k|\mathbf{x}_{k-1}\mathbf{x}_{k-2}}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1},\boldsymbol{x}_{k-2}) \tag{2.24}$$

which in this case is called a second–order Markov process, or Markov–2. Analogously, we can define k^{th}–order Markov processes. In those cases considered in this course, the definition given in (2.20) for first–order Markov processes is sufficient.

2.4 Gaussian Process

A stochastic process in n dimensions $\{\mathbf{x}(t), t \in T\}$, where T is an arbitrary interval of time, is said to be Gaussian if for any N instants of time $t_1, t_2, \cdots, t_N$ in T, its density function, distribution function, or characteristic function, is normal. In other words, the process is Gaussian if the vectors $\mathbf{x}(t_1), \mathbf{x}(t_2), \cdots, \mathbf{x}(t_N)$ are jointly Gaussian distributed. According to what was seen in the previous lecture we can write the density function of this process as:

$$p_{\mathbf{z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{Nn/2}|\mathbf{P}_{\mathbf{z}}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_{\mathbf{z}})^T \mathbf{P}_{\mathbf{z}}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_{\mathbf{z}})\right], \tag{2.25}$$

where the vector $\boldsymbol{z}$, of dimension $Nn = N \times n$, is defined as:

$$\boldsymbol{z} \equiv \begin{pmatrix} \boldsymbol{x}(t_1) \\ \boldsymbol{x}(t_2) \\ \vdots \\ \boldsymbol{x}(t_N) \end{pmatrix}, \tag{2.26}$$

the mean vectors $\boldsymbol{\mu}_{\mathbf{z}}(t_i)$, of dimension Nn are given by

$$\boldsymbol{\mu}_{\mathbf{z}}(t_i) \equiv \mathcal{E}\{\mathbf{z}(t_i)\}, \tag{2.27}$$

for $i = 1, 2, \cdots, N$, and the covariance $\mathbf{P}_{\mathbf{z}}$, of dimension $(N \times n)^2 = Nn \times Nn$, has elements which are the sub–matrices

$$\mathbf{P}_{\mathbf{z}} = [\mathbf{P}_{\mathbf{x}}]_{ij} \equiv \mathcal{E}\{[\mathbf{x}(t_i) - \boldsymbol{\mu}_{\mathbf{x}}(t_i)][\mathbf{x}(t_j) - \boldsymbol{\mu}_{\mathbf{x}}(t_j)]^T\} \tag{2.28}$$

for $i, j = 1, 2, \cdots, N$. In this way, a Gaussian process is completely determined by its mean and its autocovariance. A process which is simultaneously Gaussian and Markovian is said to be a Gauss–Markov process.

2.5 Stationary Process

A precise definition of the concept of stationary process can be given by returning to the concept of probability. However, for what interests us, it is sufficient to utilize wide–sense stationary processes, which only requires that the first two moments be time–independent. In this sense, a stationary process is one for which the mean is independent of time:

$$\boldsymbol{\mu}_{\mathbf{x}}(t) = \boldsymbol{\mu}_{\mathbf{x}}, \tag{2.29}$$

and for which the correlation only depends on the time interval τ between events:

$$\mathbf{\Gamma}_{\mathbf{x}}(t, \tau) = \mathbf{\Gamma}_{\mathbf{x}}(t - \tau), \tag{2.30}$$

which can be written as:

$$\mathbf{\Gamma}_{\mathbf{x}}(\tau) = \mathbf{\Gamma}_{\mathbf{x}}(t + \tau, t) = \mathcal{E}\{\mathbf{x}(t + \tau)\mathbf{x}^T(t)\}. \tag{2.31}$$

An even weaker concept of stationary process is defined when the covariance is stationary. In this case,

$$\mathbf{C}_{\mathbf{x}}(\tau) = \mathbf{C}_{\mathbf{x}}(t + \tau, t) = \mathbf{\Gamma}_{\mathbf{x}}(t + \tau, t) - \boldsymbol{\mu}_{\mathbf{x}}(t + \tau)\boldsymbol{\mu}_{\mathbf{x}}^T(t) \tag{2.32}$$

These concepts apply similarly to "cross–quantities", that is, the cross–correlation and cross–covariance

$$\mathbf{\Gamma}_{\mathbf{xy}}(\tau) = \mathbf{\Gamma}_{\mathbf{xy}}(t + \tau, t) = \mathcal{E}\{\mathbf{x}(t + \tau)\mathbf{y}^T(t)\}. \tag{2.33a}$$

$$\mathbf{C}_{\mathbf{xy}}(\tau) = \mathbf{C}_{\mathbf{xy}}(t + \tau, t) = \mathbf{\Gamma}_{\mathbf{xy}}(t + \tau, t) - \boldsymbol{\mu}_{\mathbf{x}}(t + \tau)\boldsymbol{\mu}_{\mathbf{y}}^T(t) \tag{2.33b}$$

respectively, are stationary. In this case, it is simple to show that

$$\mathbf{\Gamma}_{\mathbf{xy}}(\tau) = \mathbf{\Gamma}_{\mathbf{yx}}(-\tau) \tag{2.34a}$$

$$\mathbf{C}_{\mathbf{xy}}(\tau) = \mathbf{C}_{\mathbf{yx}}(-\tau) \tag{2.34b}$$

since stationary covariances and correlations are invariant under a time translation of $-\tau$.

2.6 Wiener–Khintchine Relation

A definition that follows from the concept of stationary process introduced above is given by the Wiener–Khintchine relation. This relation defines the spectral density of the stationary covariance as being the Fourier transform of the covariance. For a continuous stochastic process the power spectrum of the covariance can be written as:

$$\hat{\mathbf{C}}_{\mathbf{x}}(\omega) \equiv \int_{-\infty}^{\infty} \mathbf{C}_{\mathbf{x}}(\tau)e^{-i\omega\tau}\,d\tau, \tag{2.35}$$

and consequently, by the inverse Fourier transform we have that

$$\mathbf{C}_{\mathbf{x}}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\mathbf{C}}_{\mathbf{x}}(\omega)e^{i\omega\tau}\,d\omega. \tag{2.36}$$

For discrete stationary processes the discrete Fourier transform defines the corresponding power spectrum.

2.7 White Noise Process

The simplest power spectrum that we can think of is the one given by a constant, that is, one for which $\hat{\mathbf{C}}(\omega) \equiv \hat{\mathbf{Q}}_{\mathbf{w}}(\omega) = \mathbf{Q}_{\mathbf{w}}$, where the stochastic process $\mathbf{w}$ is called white noise. In this case, the covariance becomes a Dirac delta:

$$\begin{aligned} \mathbf{Q}_{\mathbf{w}}(\tau) &= \frac{1}{2\pi}\mathbf{Q}_{\mathbf{w}} \int_{-\infty}^{\infty} e^{i\omega\tau}\, d\omega \\ &= \mathbf{Q}_{\mathbf{w}}\delta(\tau)\,. \end{aligned} \tag{2.37}$$

Even if this noise is completely non–physical, because it is infinite at the origin, it is of great importance in the development of stochastic differential equations.

2.8 Wiener Process

A Wiener process, also called Brownian motion, denoted by $\mathbf{b}(t)$, is defined as the integral of a stationary, Gaussian white noise process $\mathbf{w}(t)$ with zero mean:

$$\mathbf{b}(t) = \int_0^t \mathbf{w}(t)\, dt\,, \tag{2.38}$$

where

$$cov\{\mathbf{w}(t), \mathbf{w}(\tau)\} = \mathbf{Q}_{\mathbf{w}}\delta(t-\tau)\,, \tag{2.39}$$

as we saw above. Some of the properties of this process are listed below:

1. $\mathbf{b}(t)$ is normally distributed.
2. $\mathcal{E}\{\mathbf{b}(t)\} = \mathbf{0}$, for all $t \leq 0$.
3. $P\{\mathbf{b}(0) = 0\} = 1$.
4. $\mathbf{b}(t)$ has independent and stationary increments, that is, independent of time. We refer to increments as being the differences $\mathbf{b}(t_1) - \mathbf{b}(t_2), \cdots, \mathbf{b}(t_{n-1}) - \mathbf{b}(t_n)$, where $t_{i+1} < t_i$, with $t_i \in T$.
5. $\mathbf{b}(t)$ is a Markov process.

Moreover the variance of a Wiener process increases linearly in time:

$$\begin{aligned} var\{\mathbf{b}(t)\} = \mathcal{E}\{\mathbf{b}(t)\mathbf{b}^T(t)\} &= \int_0^t \int_0^t \mathcal{E}\{\mathbf{w}(t_1)\mathbf{w}^T(t_2)\}\, dt_1\, dt_2 \\ &= \mathbf{Q}_{\mathbf{w}} \int_0^t \int_0^t \delta(t_1 - t_2)\, dt_1\, dt_2 \\ &= \mathbf{Q}_{\mathbf{w}} \int_0^t dt_1 = \mathbf{Q}_{\mathbf{w}}\, t\,, \end{aligned} \tag{2.40}$$

where we used the following definition of a delta function:

$$\int_0^t f(s)\delta(t-s)\,ds =, f(t)\,. \tag{2.41}$$

It is important to notice that the difficulty encountered in the description of a Gaussian white noise process, the problem of infinite variance, does not exist for the Wiener process. That means, the latter is a well-behaved process.

2.9 Spatial Random Fields

The literature on random stochastic fields is relatively smaller than that on stochastic processes. Still, there are several treatments, such as those of Vanmarcke [132] and Yaglom [140]. A more recent treatment, directed toward earth science applications is the one of Christakos [24]. In what follows, we will be as concise as possible, keeping in mind that the main purpose of this section is to introduce the concepts of homogeneity and isotropy for random fields.

The concept of random fields can be introduced similarly to the way we introduced stochastic processes. In this case, we associate with each random variable $\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n$ the points $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_n$ in the space R^n. A random spatial field can be considered a function of events $\omega \in \Omega$, where Ω is the sample space introduced in the previous lecture, and also a function of the spatial position $\mathbf{r} \in R^n$, that is, $\mathrm{x}(\mathbf{r}) = \mathrm{x}(\omega, \mathbf{r})$. When we write $\mathrm{x}(\mathbf{r})$ we are simplifying the notation in a manner entirely analogous to what we did in the previous section, when the variable was the time. This concept can be extended to several random variables depending on space in order to motivate the introduction to vector random spatial fields. We denote by $\mathbf{x}(\mathbf{r})$ the vector random field which represents the set of random spatial fields $\mathrm{x}_1(\mathbf{r}), \mathrm{x}_2(\mathbf{r}), \cdots, \mathrm{x}_m(\mathbf{r})$, that is,

$$\mathbf{x}(\mathbf{r}) = [\mathrm{x}_1(\mathbf{r}), \mathrm{x}_2(\mathbf{r}), \cdots, \mathrm{x}_m(\mathbf{r})]^T \tag{2.42}$$

The distribution function of a vector random spatial field is then defined as:

$$F_{\mathbf{x}}(\boldsymbol{x}, \mathbf{r}) = P(\{\omega : \mathbf{x}(\mathbf{r}) \leq \boldsymbol{x}; \mathbf{r} \in R^n\})\,. \tag{2.43}$$

We emphasize once more that the concept of random fields is an extension of the concept of stochastic process. A stochastic process is a random field for which the spatial argument $\mathbf{r} \in R^n$, is introduced for $n = 1$ and $\mathbf{r} \to r \to t$ so that the random variable becomes $\mathbf{x}(t)$, as before.

The distribution function is related to the probability density by means of the expression

$$p_{\mathbf{x}}(\boldsymbol{x}, \mathbf{r}) = \frac{\partial^n F_{\mathbf{x}}(\boldsymbol{x}, \mathbf{r})}{\partial \boldsymbol{x}} \tag{2.44}$$

and consequently

$$F_{\mathbf{x}}(\boldsymbol{x}, \mathbf{r}) = \int_{-\infty}^{\boldsymbol{x}} p_{\mathbf{x}}(\boldsymbol{x}', \mathbf{r})\, d\boldsymbol{x}'\,. \tag{2.45}$$

The concepts of mean, variance and correlation can be extended directly for the case of random spatial fields. Therefore we define concisely these quantities in this case:

- *Mean value of a random field*:

$$\mu_{\mathbf{x}}(\mathbf{r}) \equiv \mathcal{E}\{\mathbf{x}(\mathbf{r})\} = \int_{-\infty}^{\infty} \boldsymbol{x} p_{\mathbf{x}(\mathbf{r})}(\boldsymbol{x}, \mathbf{r})\, d\boldsymbol{x} , \tag{2.46}$$

- *Auto–covariance matrix of a random spatial field*:

$$\mathbf{C}_{\mathbf{x}}(\mathbf{r}_i, \mathbf{r}_j) = cov\{\mathbf{x}(\mathbf{r}_i), \mathbf{x}(\mathbf{r}_j)\} \equiv \mathcal{E}\{[\mathbf{x}(\mathbf{r}_i) - \boldsymbol{\mu}_{\mathbf{x}}(\mathbf{r}_i)][\mathbf{x}(\mathbf{r}_j) - \boldsymbol{\mu}_{\mathbf{x}}(\mathbf{r}_j)]^T\} , \tag{2.47}$$

  for two spatial points $\mathbf{r}_i$ and $\mathbf{r}_j$, where we made an analogy with what we saw in stochastic processes; *auto* refers to the random field in question, in this case $\mathbf{x}(\mathbf{r})$. Then, we have that

$$\mathbf{C}_{\mathbf{x}}(\mathbf{r}_i, \mathbf{r}_j) = \mathcal{E}\{\mathbf{x}(\mathbf{r}_i)\mathbf{x}^T(\mathbf{r}_j)\} - \mu_{\mathbf{x}}(\mathbf{r}_i)\mu_{\mathbf{x}}^T(\mathbf{r}_j) \tag{2.48}$$

  When $\mathbf{r}_i = \mathbf{r}_j$, we have the variance matrix which describes the local behavior of the random field.

In order to simplify and more easily demonstrate the notation, consider the case of a scalar random field $\mathrm{x}(\mathbf{r})$. The mean introduced above becomes a scalar quantity $\mu(\mathbf{r})$, that is, a function of "one" spatial variable $\mathbf{r} \in R^n$. The covariance becomes a function (no longer a matrix) of "two" spatial variables $\mathbf{r}_i, \mathbf{r}_j$. The variance is a function given by

$$\sigma_{\mathrm{X}}^2(\mathbf{r}) \;=\; C_{\mathrm{X}}(\mathbf{r}, \mathbf{r}_j = \mathbf{r}) \tag{2.49}$$

for $\mathbf{r} = \mathbf{r}_i$. We can still introduce the spatial correlation function $\rho_{\mathrm{X}}(\mathbf{r}_i, \mathbf{r}_j)$ between two points as:

$$\rho_{\mathrm{X}}(\mathbf{r}_i, \mathbf{r}_j) \;=\; \frac{C_{\mathrm{X}}(\mathbf{r}_i, \mathbf{r}_j)}{\sigma_{\mathrm{X}}(\mathbf{r}_i)\sigma_{\mathrm{X}}(\mathbf{r}_j)} \tag{2.50}$$

A scalar random spatial field is said to be uncorrelated when

$$C_{\mathrm{X}}(\mathbf{r}_i, \mathbf{r}_j) = \begin{cases} \sigma_{\mathrm{X}}^2(\mathbf{r}) , & \text{for } \mathbf{r}_i = \mathbf{r}_j = \mathbf{r} \\ 0 , & \text{otherwise} \end{cases} \tag{2.51}$$

and in fact, such a random field is said to be a white field (analogously to the white process seen previously).

Basically all the concepts defined for random processes can be generalized for spatial random fields:

- Markovian *process* → Markovian *field*
- Gaussian *process* → Gaussian *field*
- white *process* → white *field*

as for the concepts of characteristic function, conditional probability function, conditional mean, conditional covariance, etc.

A very important generalization is that of the concept of stationarity of a stochastic process, which for spatial random fields translates into the concept of *homogeneity.* In the wide sense, a spatial random field is said to be homogeneous when its mean value is independent of the spatial variable, and its covariance depends only on the distance between two points in space. For the scalar case, this can be written as:

$$\begin{aligned} \mu_X(\mathbf{r}) &= \mu && (2.52a) \\ C_X(\mathbf{r}_i, \mathbf{r}_j) &= C_X(\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j) && (2.52b) \end{aligned}$$

Another fundamental concept is that of the *isotropic* spatial random field, which is defined as a field for which

$$C_X(\mathbf{r}_i, \mathbf{r}_j) = C_X(r = |\mathbf{r}_i - \mathbf{r}_j|) \qquad (2.53)$$

is satisfied. That is, a spatial random field is said to be isotropic when its covariance depends only on the magnitude of the distance between two points in space.

It is possible to show (e.g., Christakos [24]) that for a homogeneous random field, not necessarily isotropic, we can write the covariance function as

$$C_X(\mathbf{r}) = \int_{R^n} \exp(i\mathbf{w}^T\mathbf{r})\, \hat{C}_X(\mathbf{w})\, d\mathbf{w} \qquad (2.54)$$

where $\hat{C}_X(\mathbf{w})$ is the spectral density function that, by the inverse Fourier transform, can be written as:

$$\hat{C}_X(\mathbf{w}) = \frac{1}{(2\pi)^n} \int_{R^n} \exp(-i\mathbf{w}^T\mathbf{r})\, C_X(\mathbf{r})\, d\mathbf{r} \qquad (2.55)$$

This result can be generalized for the case of vector random fields. Notice that for real random fields, the covariance and spectral density can in fact be expressed in terms of Fourier cosine integrals.

$$\begin{aligned} C_X(\mathbf{r}) &= \int_{R^n} \cos(\mathbf{w}^T\mathbf{r})\, \hat{C}_X(\mathbf{w})\, d\mathbf{w} && (2.56a) \\ \hat{C}_X(\mathbf{w}) &= \frac{1}{(2\pi)^n} \int_{R^n} \cos(\mathbf{w}^T\mathbf{r})\, C_X(\mathbf{r})\, d\mathbf{r} && (2.56b) \end{aligned}$$

The importance of these results lies in the fact that they provide a relatively simple criterion to determine whether a continuous and symmetric function in R^n can be a covariance function. In fact, the necessary and sufficient condition for a continuous function $C_X(\mathbf{r}_i, \mathbf{r}_j)$ in R^n to be a covariance function is that it be a positive–semidefinite function, that is,

$$\int_{R^n} \int_{R^n} C_X(\mathbf{r}_i, \mathbf{r}_j) f(\mathbf{r}_i) f(\mathbf{r}_j)\, d\mathbf{r}_i\, d\mathbf{r}_j \geq 0 \qquad (2.57)$$

for any function $f(\mathbf{r})$. This criterion is generally very difficult to verify, even for homogeneous random fields. However, utilizing the spectral representation above, Bochner's [15] theorem says that the criterion for a continuous and symmetric function in R^n to be a covariance function is that its spectral function be positive–semidefinite

$$\hat{C}_X(\mathbf{w}) \geq 0 \tag{2.58}$$

for $\mathbf{w} \in R^n$.

A relevant result that appears in atmospheric data assimilation concerns the isotropic case with $n = 2$, that is, in R^2. In this case, $C_X(\mathbf{r}) = C_X(r)$, where $r = |\mathbf{r}|$. Introducing polar coordinates: $\mathbf{r} = (x, y) = (r\cos\theta, r\sin\theta)$ and $\mathbf{w} = (w\cos\varphi, w\sin\varphi)$; and recalling the change of variables in integrals means that we should calculate the determinant of the Jacobian matrix that corresponds to the transformation, that is

$$|\mathrm{Jac}(r,\theta)| \equiv \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r \tag{2.59}$$

where the notation $|.|$ is used for the determinant. In this way, using the fact that the integral over R^2 for any function $f(x, y)$ is transformed into an integral over the circle C as

$$\int_0^\infty \int_0^\infty f(x,y)\, dx\, dy = \int_0^\infty \int_0^{2\pi} f(r\cos\theta, r\sin\theta)\, r\, dr\, d\theta \tag{2.60}$$

(e.g., Apostol [4], pp. 479 485), the integral in (2.56b) becomes

$$\begin{aligned} \hat{C}_X(\mathbf{w}) = \hat{C}_X(w) &= \frac{1}{(2\pi)^2} \int_0^\infty C_X(r)\, r \int_0^{2\pi} \cos(\mathbf{w}^T\mathbf{r})\, d\theta \\ &= \frac{1}{(2\pi)^2} \int_0^\infty C_X(r)\, r \int_0^{2\pi} \cos[wr\cos(\theta-\varphi)]\, d\theta \qquad (2.61) \end{aligned}$$

where the last equality is obtained by treating the inner product explicitly:

$$\begin{aligned} \mathbf{w}^T\mathbf{r} &= rw\cos\theta\cos\varphi + rw\sin\theta\sin\varphi \\ &= rw\cos(\theta-\varphi) \qquad (2.62) \end{aligned}$$

where $w = |\mathbf{w}|$. Now performing the transformation, $\theta \to \theta + \varphi + \pi/2$, we have that $\cos(\theta - \varphi) \to -\sin\theta$, and therefore the integral of the expression above is independent of φ. This means that the result of the integral is also independent of φ, as should be the case for isotropic covariances. Introducing the Bessel function of order zero:

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \cos(x\sin\theta)\, d\theta \tag{2.63}$$

(e.g., Arfken [5], pp. 579–580), we have that in two dimensions

$$\hat{C}_X(w) = \frac{1}{2\pi} \int_0^\infty J_0(wr) C_X(r)\, r\, dr \tag{2.64}$$

Utilizing the orthogonality of the Bessel function of order zero:

$$\int_0^\infty r J_0(wr) J_0(w'r)\, dr = \frac{1}{w}\delta(w - w') \tag{2.65}$$

(e.g., Arfken [5], p. 594), we obtain for the isotropic covariance function in two dimensions the formula:

$$C_{\mathrm{X}}(r) = 2\pi \int_0^\infty J_0(wr)\hat{C}_{\mathrm{X}}(w)w\, dw \tag{2.66}$$

It is interesting to mention that the concept of ergodicity can also be extended to spatial random fields. In an entirely analogous way to what can be done for stochastic processes, a spatial random field is said to be ergodic if its *spatial* mean and covariance coincide with its ensemble mean and covariance, respectively.

EXERCISES

1. (Problem 4.3, Meditch [103]) Assuming that three scalar stochastic process $\{\mathrm{x}(t), t \in T\}$, $\{\mathrm{y}(t), t \in T\}$ and $\{\mathrm{z}(t), t \in T\}$ are pairwise independent, show that they are not necessarily triplewise (simultaneously) independent.

2. Calculate the power spectrum for stationary processes having the following autocorrelation functions:

   (a) Gaussian pulse: $\Gamma(\tau) = \sigma^2 e^{-\tau^2/T^2}$

   (b) Damped cosine wave: $\Gamma(\tau) = \sigma^2 e^{-\beta|\tau|} \cos\omega_0\tau$

   (c) Triangular pulse:

   $$\Gamma(\tau) = \begin{cases} 1 - |\tau|, & \text{for } |\tau| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

3. (Problem 2.17, Brown [19]) The stationary process $x(t)$ has mean $\mu = const.$ and an autocorrelation function of the form

   $$\Gamma(t, t+\tau) \;=\; \Gamma(\tau) \;=\; \sigma^2 e^{-\tau^2/T^2}$$

   Another process $y(t)$ is related to $x(t)$ by the deterministic equation

   $$y(t) \;=\; a\,x(t) + b$$

   where a and b are known constants.

   (a) What is the auto–correlation function for $y(t)$?

   (b) What is the cross–correlation function $\Gamma_{xy}(\tau)$?

4. (Problem 2.20, Brown [19]) Two random processes are defined by

   $$\begin{aligned} x(t) &= a\,\sin(\omega t + \theta) \\ y(t) &= b\,\sin(\omega t + \theta) \end{aligned}$$

   where θ is a random variable with uniform distribution between 0 and 2π, and ω is a known constant. The coefficients a and b are both normal random variables $\mathcal{N}\{0, \sigma^2\}$ and are correlated with a correlation coefficient ρ. What is the cross–correlation function $\Gamma_{xy}(\tau)$? (Assume a and b are independent of θ.)

5. Show that the following are admissible candidates for a covariance function:

   (a) $C_X(\mathbf{r}) = a\delta(\mathbf{r})$, for $a > 0$ and $\mathbf{r} \in R^n$

   (b) $C_X(x,y) = C_X(r = |x-y|) = \pi \exp(-r^2)$, for $x, y \in R^1$. (Hint: In this case, the proof can be obtained by either showing that (2.57) is true, or showing that (2.58) is satisfied. Use (2.58) and expand $\exp(2xy)$ in Taylor series.)

6. Show that in R^3 the isotropic spectral density function can be expressed as

$$\hat{C}_X(w) = \frac{1}{2\pi^2}\int_0^\infty \frac{\sin(wr)}{w} C_X(r)\, r\, dr$$

and that consequently the corresponding covariance function is given by

$$C_X(r) = 4\pi \int_0^\infty \frac{\sin(wr)}{r}\hat{C}_X(w)\, w\, dw$$

7. (Problem 7.14, Maybeck [101]) In Monte Carlo analyses and other type of system simulations (e.g., non–identical twin experiments), it is often desired to generate samples of a discrete–time white Gaussian noïse vector process, described by mean zero and covariance

$$\mathcal{E}\{\mathbf{w}_k\mathbf{w}_k^T\} = \mathbf{Q}_k$$

with $\mathbf{Q}_k$ nondiagonal. Independent scalar white Gaussian noises can be simulated readily through use of pseudorandom codes (as we have seen in our first computer assignment), but the question remains, how does one properly provide for cross–covariances of the scalar noises?

   (a) Let $\mathbf{v}_k$ be a vector process composed of independent scalar white Gaussian noises of zero mean and unit variance:

$$\mathcal{E}\{v_{j,k}\} = 0 \quad \mathcal{E}\{v_{j,k}^2\} = 1 \quad \text{for } k = 1, 2, \ldots$$

   where $v_{j,k}$ is the j–th component of $\mathbf{v}_k$, at time t_k. Show that

$$\mathbf{w}_k = \mathbf{L}_k\mathbf{v}_k \quad \text{for all } k$$

   properly models the desired characteristics. The matrix $\mathbf{L}_k$ above corresponds to the Cholesky lower triangular square root of $\mathbf{Q}_k$, that is, $\mathbf{Q}_k = \mathbf{L}_k\mathbf{L}_k^T$. Notice, that if the Cholesky upper triangular square root had been used instead, the corresponding expression for generating $\mathbf{w}_k$ would be

$$\mathbf{w}_k = \mathbf{U}_k\mathbf{v}_k \quad \text{for all } k$$

   where, in this case, $\mathbf{Q}_k = \mathbf{U}_k\mathbf{U}_k^T$.

   (b) If $\mathbf{U}_k$ and $\mathbf{D}_k$ are the $\mathbf{U}$ $\mathbf{D}$ factors of $\mathbf{Q}_k$, that is, if $\mathbf{Q}_k = \mathbf{U}_k\mathbf{D}_k\mathbf{U}_k^T$, where $\mathbf{U}_k$ are upper triangular and unitary matrices and $\mathbf{D}_k$ are diagonal matrices, show that,

$$\mathbf{w}_k = \mathbf{U}_k\mathbf{u}_k \quad \text{for all } k$$

also provides the desired model if $\mathbf{u}_k$ is a vector process composed of independent scalar white Gaussian noises of mean zero and variance:

$$\mathcal{E}\{u_{j,k}^2\} = d_{j,j;k}\,,$$

where $u_{j,k}$ is the j–th component of $\mathbf{u}_k$, at time t_k, and $d_{j,j;k}$ is the (j,j) element of $\mathbf{D}_k$, at time t_k (i.e., the j–th element along its diagonal).

8. *Computer Assignment:* We want to use the results of the previous problem to perform Monte Carlo experiments for a given correlation and/or covariance structure. As a preparation for that, in this problem we are going to create a Matlab function that generates a homogeneous correlation on a grid defined over R^1 and examine some of its properties. Let us start by consider the interval $(-L_x, L_x]$, and let us divide it in a uniform grid of J points. Consider also the homogeneous and isotropic, Gaussian correlation function in $R^1 \times R^1$, that is,

$$Q(x,y) = Q(r = |x-y|) = \exp(-\frac{1}{2}(x-y)^2/L_d^2)$$

where $r = |x-y|$ is the distance between any two points in the domain, and L_d is the (de)correlation length. Therefore the points in the discrete domain can be defined as

$$x_j = j\Delta x$$

where $\Delta x = 2L_x/J$, for $j \in \{-J/2+1, J/2\}$, and the elements of the homogeneous, isotropic correlation matrix $\mathbf{Q}$ are given by

$$Q_{ij} = Q(x_i, y_j)$$

(a) Construct a Matlab function that returns the covariance matrix $\mathbf{Q}$, given the half–length of the domain L_x, the number of grid points J, and the (de)correlation length L_d. For $(L_x, L_d, J) = (1, 0.2, 32)$, compute $\mathbf{Q}$ using this function. Make a contour plot of the correlation array. (Note: A real convenient way of generating this matrix in Matlab is using the intrinsic function $\boxed{\text{meshgrid}}$.)

(b) For the parameters of the previous item, plot the correlation function at the following two specific locations: $x_j \in \{0, L_x\}$.

(c) Is the $\mathbf{Q}$ obtained above an *acceptable* correlation matrix? Explain it. (Hint: Check its eigenvalues.)

(d) From the figures constructed in the previous items, we see that the correlation decreases very quickly toward values that are nearly zero. (You can actually print out the values in $\mathbf{Q}$ to check it further). It could be computationally advantageous, particularly to reduce storage requirements, to approximate this correlation "function" (matrix) by one that neglects correlation values beyond a certain cut–off length L_c. In this way, only the elements of the matrix corresponding to $|r| \leq L_c$ would need to be stored. Without worrying about the storages savings, modify the function of item (a), to construct a new matrix $\mathbf{Q}_c$, by replacing the values of $\mathbf{Q}$ for which $|r| > L_c$ by zero. Using the same parameters as in item (a), and a cut–off value of $L_c = 3L_d$, make a contour plot of the resulting correlation structure. Also, repeat item (b).

(e) A visual comparison of the plots in items (a) and (b) with those of item (d) seem to indicate that our approximation is fairly reasonable. Is the correlation of the previous item, an acceptable correlation matrix?

9. *Computer Assignment*: The result obtained in the last item of the previous exercise makes us wonder what is the correct way of constructing a correlation field that has the structure that we want, but is zero beyond a certain correlation length. The procedure to generate what are called compact–support correlation functions is through the use of convolution of functions. (Note: see Gaspari and Cohn (1996) for the details on constructing these correlation functions in R^2 and R^3, that are of primary importance in modeling covariances for data assimilation). Another way of looking at convolution of functions is to think on the Hadamard product for the case of matrices. Without getting into the mathematical details, this problem has the intention to guide you though the steps of building an actual correlation matrix for the function of the previous problem. Consider then the compact–support triangular correlation function discussed earlier in the text:

$$T(x,y) == \begin{cases} 1-|x-y|/L_c\,, & \text{for } |x-y| \leq L_c \\ 0 & \text{otherwise} \end{cases}$$

Then perform the following tasks:

(a) Repeat items (a)–(c) of the previous exercise, but now for the compact–support function $T(r) = T(|x-y|)$. (Note: The matrix $\mathbf{T}$ has elements $T_{ij} = T(x_i, y_j)$.)

(b) Construct a matrix $\bar{\mathbf{Q}}$ as the Hadamard product of the matrix $\mathbf{Q}$, of item (a) in the previous exercise, and $\mathbf{T}$ from the previous item, corresponding to the function $T(r)$. That is, let $\bar{\mathbf{Q}}$ be given by

$$\bar{\mathbf{Q}} \equiv \mathbf{Q} \circ \mathbf{T} = [Q_{ij}T_{ij}]$$

(Note: Matlab does the Hadamard product trivially). Make a contour plot $\bar{\mathbf{Q}}$, and repeat item (c) of the previous exercise.

(c) To get yet another visual representation of what the correlations from $\mathbf{Q}$, $\mathbf{T}$ and $\bar{\mathbf{Q}}$ are like, plot the correlation functions obtained from these three matrices at point $x = 0$. (Please, have all three curves on the same frame).

10. *Computer Assignment:* Using the Matlab function created in item (a) of Exercise 8 (i.e., without a cut–off length), let us apply the results of Exercise 7 to understand better what correlated noise actually is.

(a) Create a Matlab function that performs a Monte Carlo experiment given the number of samples. We want the output of this function to be the sampled correlation matrix, obtained from a weighted sum of outer products of the vectors $\mathbf{w}_k$ of Exercise 7. To obtain the Cholesky decomposition of the correlation matrix $\mathbf{Q}$ of Exercise 8, use the Matlab function $\boxed{\text{chol}}$. Make contour plots for the three sampled correlation matrices obtained by using 100, 1000, 10000 samples.

(b) Now using the identity matrix, in the 32–dimensional space of Exercise 8, perform a Monte Carlo run, with 1000 samples, assuming this identity matrix is the correlation matrix of interest. Make a contour plot of the resulting sampled correlation matrix. Compare this result with those obtained in the previous item. In particular, explain the meaning of using an identity correlation matrix.

Chapter 3

Stochastic Differential Equations

In this lecture we will present a simple discussion of systems governed by stochastic differential equations. To maintain the most pleasant possible notation, we will not make explicit distinction between the random variable $\mathbf{u}$ and its corresponding value $\boldsymbol{u}$, the first being utilized to represent the stochastic process of interest. This notation will be utilized through the end of this course.

3.1 Linear Dynamical Systems

Linear transformations of an r.v. consist of one of the most fundamental transformations in stochastic processes. A linear transformation of special interest is found in dynamical systems. In this case, a certain initial condition evolves according to the dynamics of a *linear operator*. In case the distribution function of the r.v.'s in question is a Gaussian, a linear transformation of this variable produces an r.v. with the Gaussian distribution (see Exercise 1.4). This occurs in problems in linear dynamical systems: given an initial condition with the Gaussian distribution, the final result will also be Gaussian distributed. In this lecture, we concentrate in calculating mean and (co)variances of stochastic processes, since in the normally distributed case these quantities define the process completely. Higher moments are necessary when either the process is not Gaussian or the process is nonlinear. The treatment in this lecture is general and independent of the distribution under consideration, meaning that any moments can in principle be calculated according to the procedures given below. Both time–continuous and time–discrete stochastic processes are discussed here. A brief introduction to the case of systems governed by stochastic random fields is discussed in the end of this lecture.

3.1.1 Continuous Processes

Consider the following linear dynamics of first order in time for the random n–vector $\mathbf{u}$:

$$\dot{\mathbf{u}} = \frac{d\mathbf{u}(t)}{dt} = \mathbf{F}(t)\mathbf{u}(t) + \mathbf{G}(t)\mathbf{w}(t)\,, \tag{3.1}$$

where the mean and the (co)variance of the initial state $\mathbf{u}(t_0)$ and of the m–vector o noise $\mathbf{w}(t)$ are given by:

$$\boldsymbol{\mu}_{\mathbf{w}}(t) = \mathcal{E}\{\mathbf{w}(t)\} \qquad \mathbf{P}_{\mathbf{w}}(t_1, t_2) = cov\{\mathbf{w}(t_1), \mathbf{w}(t_2)\} \tag{3.2}$$

$$\boldsymbol{\mu}_{\mathbf{u}}(t_0) = \mathcal{E}\{\mathbf{u}(t_0)\} \qquad \mathbf{P}_{\mathbf{u}}(t_0) = var\{\mathbf{u}(t_0)\} \tag{3.3}$$

where the matrices $\mathbf{P}_{\mathbf{w}}$ and $\mathbf{P}_{\mathbf{u}}$ are of dimension $m \times m$ and $n \times n$, respectively. Moreover, we consider the process $\mathbf{w}(t)$ to be uncorrelated with the initial process $\mathbf{u}(t_0)$, that is,

$$cov\{\mathbf{u}(t_0), \mathbf{w}(t)\} = \mathbf{0}\,, \tag{3.4}$$

for $t \geq t_0$. The problem we want to approach is to find the mean $\boldsymbol{\mu}_{\mathbf{u}}(t)$ and the covariance $\mathbf{P}_{\mathbf{u}}(t_1, t_2)$, for any $t, t_1, t_2 > t_0$.

The general solution of (3.1) is given by

$$\mathbf{u}(t) = \boldsymbol{\Psi}(t, t_0)\mathbf{u}(t_0) + \int_{t_0}^{t} \boldsymbol{\Psi}(t, \tau)\mathbf{G}(\tau)\mathbf{w}(\tau)\, d\tau\,, \tag{3.5}$$

where $\boldsymbol{\Psi}(t, t_0)$ is the transition matrix of the system, the solution of the homogeneous linear differential equation:

$$\dot{\boldsymbol{\Psi}}(t, t_0) = \frac{d\boldsymbol{\Psi}(t, t_0)}{dt} = \mathbf{F}(t)\boldsymbol{\Psi}(t, t_0)\,, \tag{3.6}$$

with initial condition

$$\boldsymbol{\Psi}(t_0, t_0) = \mathbf{I}\,, \tag{3.7}$$

where $\mathbf{I}$ is the identity matrix.

We can determine the mean of $\mathbf{u}(t)$ by applying the ensemble mean operator to (3.5):

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{u}}(t) \equiv \mathcal{E}\{\mathbf{u}(t)\} &= \mathcal{E}\{\boldsymbol{\Psi}(t, t_0)\mathbf{u}(t_0)\} + \mathcal{E}\{\int_{t_0}^{t} \boldsymbol{\Psi}(t, \tau)\mathbf{G}(\tau)\mathbf{w}(\tau)\, d\tau\} \\ &= \boldsymbol{\Psi}(t, t_0)\boldsymbol{\mu}_{\mathbf{u}}(t_0) + \int_{t_0}^{t} \boldsymbol{\Psi}(t, \tau)\mathbf{G}(\tau)\boldsymbol{\mu}_{\mathbf{w}}(\tau)\, d\tau \end{aligned} \tag{3.8}$$

where the last equality is obtained by exchanging the ensemble mean operator with the integration operator, since they act on different variables. In this way, the mean of the process $\mathbf{u}(t)$ satisfies an expression analogous to the solution of the equation (3.1), except that the processes $\mathbf{u}$ and $\mathbf{w}$ are substituted by their respective means.

The integration in (3.8) is complicated in the majority of cases, therefore it is convenient that we determine an auxiliary expression to obtain the mean of $\mathbf{u}(t)$. This can be done by applying the ensemble mean operator directly to the equation (3.1). In this case we obtain:

$$\dot{\boldsymbol{\mu}}_{\mathbf{u}}(t) = \mathbf{F}(t)\boldsymbol{\mu}_{\mathbf{u}}(t) + \mathbf{G}(t)\boldsymbol{\mu}_{\mathbf{w}}(t)\,, \tag{3.9}$$

subject to the initial condition $\boldsymbol{\mu}_{\mathbf{u}}(t_0)$. The solution of this equation is obviously given by (3.8); however, computationally the equation above has a much simpler solution.

Before deriving the equation for the (co)variance $\mathbf{P}_{\mathbf{u}}(t_1, t_2)$ it is helpful to notice that from the definition of cross-covariance $\mathbf{P}_{\mathbf{uv}}(t_1, t_2) = cov\{\mathbf{u}(t_1), \mathbf{v}(t_2)\}$, for general n-vectors $\mathbf{u}(t_1)$ and m-vectors $\mathbf{v}(t_2)$, it follows that if we change $\mathbf{u}(t_1) \to \mathbf{A}(t)\mathbf{u}(t_1)$, for any non-stochastic $n \times n$ matrix $\mathbf{A}$(t), we have

$$\begin{aligned}
cov\{\mathbf{A}(t)\mathbf{u}(t_1), \mathbf{v}(t_2)\} &= \mathcal{E}\{[\mathbf{A}(t)\mathbf{u}(t_1) - \mathcal{E}\{\mathbf{A}(t)\mathbf{u}(t_1)\}][\mathbf{v}(t_2) - \mathcal{E}\{\mathbf{v}(t_2)\}]^T\} \\
&= \mathcal{E}\{[\mathbf{A}(t)\mathbf{u}(t_1) - \mathbf{A}(t)\mathcal{E}\{\mathbf{u}(t_1)\}][\mathbf{v}(t_2) - \mathcal{E}\{\mathbf{v}(t_2)\}]^T\} \\
&= \mathcal{E}\{[\mathbf{A}(t)[\mathbf{u}(t_1) - \mathcal{E}\{\mathbf{u}(t_1)\}]][\mathbf{v}(t_2) - \mathcal{E}\{\mathbf{v}(t_2)\}]^T\} \\
&= \mathbf{A}(t)\mathcal{E}\{[\mathbf{u}(t_1) - \mathcal{E}\{\mathbf{u}(t_1)\}][\mathbf{v}(t_2) - \mathcal{E}\{\mathbf{v}(t_2)\}]^T\} \\
&= \mathbf{A}(t) cov\{\mathbf{u}(t_1), \mathbf{v}(t_2)\} \\
&= \mathbf{A}(t)\mathbf{P}_{\mathbf{uv}}(t_1, t_2) \qquad (3.10)
\end{aligned}$$

Analogously, if instead of changing $\mathbf{u}(t_1)$, we had changed $\mathbf{v}(t_2) \to \mathbf{B}(\tau)\mathbf{v}(t_2)$, for an arbitrary non-stochastic $m \times m$ matrix $\mathbf{B}(\tau)$, we would have

$$\begin{aligned}
cov\{\mathbf{u}(t_1), \mathbf{B}(\tau)\mathbf{v}(t_2)\} &= \mathcal{E}\{[\mathbf{u}(t_1) - \mathcal{E}\{\mathbf{u}(t_1)\}][\mathbf{B}(\tau)\mathbf{v}(t_2) - \mathcal{E}\{\mathbf{B}(\tau)\mathbf{v}(t_2)\}]^T\} \\
&= \mathcal{E}\{[\mathbf{u}(t_1) - \mathcal{E}\{\mathbf{u}(t_1)\}][\mathbf{B}(\tau)[\mathbf{v}(t_2) - \mathcal{E}\{\mathbf{v}(t_2)\}]^T]\} \\
&= \mathcal{E}\{[\mathbf{u}(t_1) - \mathcal{E}\{\mathbf{u}(t_1)\}][\mathbf{v}(t_2) - \mathcal{E}\{\mathbf{v}(t_2)\}]^T\mathbf{B}^T(\tau)\} \\
&= \mathcal{E}\{[\mathbf{u}(t_1) - \mathcal{E}\{\mathbf{u}(t_1)\}][\mathbf{v}(t_2) - \mathcal{E}\{\mathbf{v}(t_2)\}]^T\}\mathbf{B}^T(\tau) \\
&= cov\{\mathbf{u}(t_1), \mathbf{v}(t_2)\}\mathbf{B}^T(\tau) \\
&= \mathbf{P}_{\mathbf{uv}}(t_1, t_2)\mathbf{B}^T(\tau) \qquad (3.11)
\end{aligned}$$

It is left for the reader to show that if both transformations on the vectors $\mathbf{u}(t_1)$ and $\mathbf{v}(t_2)$ are performed simultaneously, it follows that,

$$cov\{\mathbf{A}(t)\mathbf{u}(t_1), \mathbf{B}(\tau)\mathbf{v}(t_2)\} = \mathbf{A}(t)\mathbf{P}_{\mathbf{uv}}(t_1, t_2)\mathbf{B}^T(\tau) \qquad (3.12)$$

Using the relationships above and the solution (3.5) we can derive the expression for the covariance, that is,

$$\begin{aligned}
\mathbf{P}_{\mathbf{u}}(t_1, t_2) &\equiv cov\{\mathbf{u}(t_1), \mathbf{u}(t_2)\} \\
&= \boldsymbol{\Psi}(t_1, t_0) cov\{\mathbf{u}(t_0), \mathbf{u}(t_0)\}\boldsymbol{\Psi}^T(t_2, t_0) \\
&\quad + \boldsymbol{\Psi}(t_1, t_0) cov\left\{\mathbf{u}(t_0), \int_{t_0}^{t_2} \boldsymbol{\Psi}(t_2, \tau)\mathbf{G}(\tau)\mathbf{w}(\tau)\, d\tau\right\} \\
&\quad + cov\left\{\int_{t_0}^{t_1} \boldsymbol{\Psi}(t_1, \tau)\mathbf{G}(\tau)\mathbf{w}(\tau)\, d\tau, \boldsymbol{\Psi}(t_2, t_0)\mathbf{u}(t_0)\right\} \\
&\quad + cov\left\{\int_{t_0}^{t_1} \boldsymbol{\Psi}(t_1, \tau_1)\mathbf{G}(\tau_1)\mathbf{w}(\tau_1)\, d\tau_1,\right. \\
&\quad\quad \left.\int_{t_0}^{t_2} \boldsymbol{\Psi}(t_2, \tau_2)\mathbf{G}(\tau_2)\mathbf{w}(\tau_2)\, d\tau_2\right\} \qquad (3.13)
\end{aligned}$$

or also,

$$\mathbf{P}_{\mathbf{u}}(t_1, t_2) = \boldsymbol{\Psi}(t_1, t_0)\mathbf{P}_{\mathbf{u}}(t_0)\boldsymbol{\Psi}^T(t_2, t_0)$$

$$\begin{aligned}
&+\Psi(t_1,t_0)\int_{t_0}^{t_2} cov\{\mathbf{u}(t_0),\mathbf{w}(\tau)\}\mathbf{G}^T(\tau)\Psi^T(t_2,\tau)\,d\tau \\
&+\int_{t_0}^{t_1} \Psi(t_1,\tau)\mathbf{G}(\tau)cov\{\mathbf{w}(\tau),\mathbf{u}(t_0)\}\Psi^T(t_2,t_0)\,d\tau \\
&+\int_{t_0}^{t_1} d\tau_1 \int_{t_0}^{t_2} d\tau_2\Psi(t_1,\tau_1)\mathbf{G}(\tau_1)\mathbf{P}_{\mathbf{w}}(\tau_1,\tau_2)\mathbf{G}^T(\tau_2)\Psi^T(t_2,\tau_2)\,.
\end{aligned} \tag{3.14}$$

Utilizing the condition that $\mathbf{w}(t)$ and $\mathbf{u}(t_0)$ are uncorrelated (3.4), the expression above can be written as:

$$\begin{aligned}
\mathbf{P}_{\mathbf{u}}(t_1,t_2) &= \Psi(t_1,t_0)\mathbf{P}_{\mathbf{u}}(t_0)\Psi^T(t_2,t_0) \\
&+\int_{t_0}^{t_1} d\tau_1 \int_{t_0}^{t_2} d\tau_2\Psi(t_1,\tau_1)\mathbf{G}(\tau_1)\mathbf{P}_{\mathbf{w}}(\tau_1,\tau_2)\mathbf{G}^T(\tau_2)\Psi^T(t_2,\tau_2)\,,
\end{aligned} \tag{3.15}$$

for $t_1, t_2 > t_0$.This expression is of little utility in this form. Therefore, let us suppose that the process $\mathbf{w}(t)$ is a white noise, which means that the covariance $\mathbf{P}_{\mathbf{w}}(\tau_1,\tau_2)$ in this case is given by:

$$\mathbf{P}_{\mathbf{w}}(\tau_1,\tau_2) = \mathbf{Q}(\tau_1)\delta(\tau_1-\tau_2)\,, \tag{3.16}$$

where $\delta(\tau)$ is the symmetric Dirac delta (distribution) function, defined by:

$$\int_a^b f(\tau)\delta(\tau-t)\,d\tau = \begin{cases} 0\,, & \text{if } t<a \text{ or } t>b \\ \frac{f(a)}{2}\,, & \text{if } t=a \\ \frac{f(b)}{2}\,, & \text{if } t=b \\ f(t)\,, & \text{if } a<t<b \end{cases} \quad . \tag{3.17}$$

Supposing, for the moment, that $t_1 < t_2$, and using the fact that $\mathbf{w}(t)$ is a white noise, the expression (3.15) can be decomposed as:

$$\begin{aligned}
&\int_{t_0}^{t_1} d\tau_1 \int_{t_0}^{t_2} d\tau_2\Psi(t_1,\tau_1)\mathbf{G}(\tau_1)\mathbf{P}_{\mathbf{w}}(\tau_1,\tau_2)\mathbf{G}^T(\tau_2)\Psi^T(t_2,\tau_2) = \\
&\int_{t_0}^{t_1} d\tau_1 \left\{\int_{t_0}^{t_1} d\tau_2\Psi(t_1,\tau_1)\mathbf{G}(\tau_1)\mathbf{P}_{\mathbf{w}}(\tau_1,\tau_2)\mathbf{G}^T(\tau_2)\Psi^T(t_2,\tau_2)\right. \\
&\quad\left.+\int_{t_1}^{t_2} d\tau_2\Psi(t_1,\tau_1)\mathbf{G}(\tau_1)\mathbf{P}_{\mathbf{w}}(\tau_1,\tau_2)\mathbf{G}^T(\tau_2)\Psi^T(t_2,\tau_2)\right\} = \\
&\int_{t_0}^{t_1} d\tau_1\Psi(t_1,\tau_1)\mathbf{G}(\tau_1)\mathbf{Q}(\tau_1)\left\{\int_{t_0}^{t_1} d\tau_2\,\delta(\tau_1-\tau_2)\mathbf{G}^T(\tau_2)\Psi^T(t_2,\tau_2)\right. \\
&\quad\left.+\int_{t_1}^{t_2} d\tau_2\,\delta(\tau_1-\tau_2)\mathbf{G}^T(\tau_2)\Psi^T(t_2,\tau_2)\right\} = \\
&\int_{t_0}^{t_1} d\tau_1\Psi(t_1,\tau_1)\mathbf{G}(\tau_1)\mathbf{Q}(\tau_1)\mathbf{G}^T(\tau_1)\Psi^T(t_2,\tau_1)
\end{aligned} \tag{3.18}$$

where the last equality is obtained by using the definition of the symmetric Dirac function (3.17): the second integral within the brackets gives no contribution. An equivalent argument applies when we take $t_1 > t_2$, however, for this case the integration over the interval $[t_0, t_1$ should be carried over first.

Substituting this result in (3.15) we have that

$$\begin{aligned}\mathbf{P_u}(t_1,t_2) &= \mathbf{\Psi}(t_1,t_0)\mathbf{P_u}(t_0)\mathbf{\Psi}^T(t_2,t_0) \\ &+ \int_{t_0}^{\min(t_1,t_2)} \mathbf{\Psi}(t_1,\tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}^T(\tau)\mathbf{\Psi}^T(t_2,\tau)\,, \end{aligned} \tag{3.19}$$

which is a much simpler expression. The equation for the variance $\mathbf{P_u}(t)$ can be found by letting $t_1 = t_2 = t$:

$$\begin{aligned}\mathbf{P_u}(t) &= \mathbf{\Psi}(t,t_0)\mathbf{P_u}(t_0)\mathbf{\Psi}^T(t,t_0) \\ &+ \int_{t_0}^{t} \mathbf{\Psi}(t,\tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}^T(\tau)\mathbf{\Psi}^T(t,\tau)\,. \end{aligned} \tag{3.20}$$

An analogous comment to the one made for the solution of the equation for the mean is applicable for the variance and covariance expressions, that is, the integrations in (3.19) and (3.20) are very complicated and it is worthwhile to search for simpler expressions to work with. Direct differentiation of the solution (3.20) leads to a differential equation for the variance (see Exercise 3.1):

$$\dot{\mathbf{P}}_\mathbf{u}(t) = \mathbf{F}(t)\mathbf{P_u}(t) + \mathbf{P_u}(t)\mathbf{F}^T(t) + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}^T(t)\,, \tag{3.21}$$

which can be obtained in an even simpler way by means of applying the definition of variance to the quantity $\dot{\mathbf{P}}_\mathbf{u}(t)$. The equation (3.21) is known as the Lyapunov equation. Given the initial condition of the variance $\mathbf{P_u}(t_0)$, the Lyapunov equation determines the dynamic evolution of the variance for any $t > t_0$. Notice that the Lyapunov equation does not require knowledge of the transition matrix $\mathbf{\Psi}$.

It is possible to show that for the case of white noise the covariance $\mathbf{P_u}(t_1,t_2)$ can be obtained by means of the expressions:

$$\mathbf{P_u}(t_1,t_2) = \begin{cases} \mathbf{\Psi}(t_1,t_2)\mathbf{P_u}(t_2)\,, & \text{if } t_1 > t_2 \\ \mathbf{P_u}(t_1)\mathbf{\Psi}^T(t_2,t_1)\,, & \text{if } t_1 < t_2 \end{cases}\,, \tag{3.22}$$

(see Exercise 3.2).

3.1.2 Discrete Processes

Consider now the first-order discrete n–dimensional dynamical system:

$$\mathbf{u}(k+1) = \mathbf{\Psi}(k+1,k)\mathbf{u}(k) + \mathbf{\Gamma}(k)\mathbf{w}(k)\,, \tag{3.23}$$

where the notation $\mathbf{u}(k)$ stands for $\mathbf{u}(t_k)$, $T = t_{k+1} - t_k$ is the sampling interval, and the noise $\mathbf{w}(k)$ is a vector of dimension m.

Analogously to the continuous case, we define the mean and the covariance of the discrete processes $\mathbf{u}(0)$ and $\mathbf{w}(k)$ as:

$$\boldsymbol{\mu}_\mathbf{w}(k) = \mathcal{E}\{\mathbf{w}(k)\} \qquad \mathbf{P_w}(k,j) = cov\{\mathbf{w}(k),\mathbf{w}(j)\} \tag{3.24}$$

$$\boldsymbol{\mu}_\mathbf{u}(0) = \mathcal{E}\{\mathbf{u}(0)\} \qquad \mathbf{P_u}(0) = var\{\mathbf{u}(0)\} \tag{3.25}$$

also, the processes $\mathbf{u}(0)$ and $\mathbf{w}(k)$ are considered to be uncorrelated,

$$cov\{\mathbf{u}(0), \mathbf{w}(k)\} = \mathbf{0}\,, \tag{3.26}$$

for all $k \geq 0$, and the matrices $\mathbf{P_w}$ and $\mathbf{P_u}$ are $m \times m$ and $n \times n$ dimensional, respectively.

To obtain the general solution of (3.23) let us write its corresponding expression for $\mathbf{u}(1)$ and $\mathbf{u}(2)$, that is,

$$\mathbf{u}(1) = \mathbf{\Psi}(1,0)\mathbf{u}(0) + \mathbf{\Gamma}(0)\mathbf{w}(0) \tag{3.27}$$

and

$$\mathbf{u}(2) = \mathbf{\Psi}(2,1)\mathbf{u}(1) + \mathbf{\Gamma}(1)\mathbf{w}(1) \tag{3.28}$$

respectively. Substituting the first equation into the second we get

$$\begin{aligned} \mathbf{u}(2) &= \mathbf{\Psi}(2,1)[\mathbf{\Psi}(1,0)\mathbf{u}(0) + \mathbf{\Gamma}(0)\mathbf{w}(0)] + \mathbf{\Gamma}(1)\mathbf{w}(1) \\ &= \mathbf{\Psi}(2,0)\mathbf{u}(0) + \sum_{j=0}^{1} \mathbf{\Psi}(2,j+1)\mathbf{\Gamma}(j)\mathbf{w}(j)\,, \end{aligned} \tag{3.29}$$

where we noticed that $\mathbf{\Psi}(2,0) = \mathbf{\Psi}(2,1)\mathbf{\Psi}(1,0)$, and $\mathbf{\Psi}(2,2) = \mathbf{I}$. Continuing this procedure, we can show that the solution of the equation (3.23) can be written as

$$\mathbf{u}(k) = \mathbf{\Psi}(k,0)\mathbf{u}(0) + \sum_{j=0}^{k-1} \mathbf{\Psi}(k,j+1)\mathbf{\Gamma}(j)\mathbf{w}(j)\,, \tag{3.30}$$

for $k > 0$,

$$\mathbf{\Psi}(k,0) = \mathbf{\Psi}(k,k-1)\mathbf{\Psi}(k-1,k-2)\ldots\mathbf{\Psi}(2,1)\mathbf{\Psi}(1,0) \tag{3.31}$$

and $\mathbf{\Psi}(k,k) = \mathbf{I}$ for all k.

The expression for the mean is calculated by applying the ensemble mean operator to the solution written above, that is,

$$\begin{aligned} \boldsymbol{\mu}_\mathbf{u}(k) &= \mathcal{E}\{\mathbf{u}(k)\} \\ &= \mathcal{E}\{\mathbf{\Psi}(k,0)\mathbf{u}(0)\} + \mathcal{E}\{\sum_{j=0}^{k-1} \mathbf{\Psi}(k,j+1)\mathbf{\Gamma}(j)\mathbf{w}(j)\} \\ &= \mathbf{\Psi}(k,0)\boldsymbol{\mu}_\mathbf{u}(0) + \sum_{j=0}^{k-1} \mathbf{\Psi}(k,j+1)\mathbf{\Gamma}(j)\boldsymbol{\mu}_\mathbf{w}(j)\,, \end{aligned} \tag{3.32}$$

since the ensemble mean operator acts only on the stochastic quantities $\mathbf{u}(0)$ and $\mathbf{w}(j)$. An alternative, recursive equation for the mean can be obtained by applying the ensemble mean operator directly to (3.23), that is,

$$\boldsymbol{\mu}_\mathbf{u}(k+1) = \mathbf{\Psi}(k+1,k)\boldsymbol{\mu}_\mathbf{u}(k) + \mathbf{\Gamma}(k)\boldsymbol{\mu}_\mathbf{w}(k)\,. \tag{3.33}$$

Before we determine the covariance $\mathbf{P_u}(k,j) = cov\{\mathbf{u}(k), \mathbf{u}(j)\}$, it is useful to recognize that a general cross–covariance $\mathbf{P_{uv}}(k,j) = cov\{\mathbf{u}(k), \mathbf{v}(j)\}$, for arbitrary n–vectors $\mathbf{u}(k)$, and

m-vectors $\mathbf{v}(j)$, and arbitrary non-stochastic $n \times n$ matrices $\mathbf{A}(k')$, and $m \times m$ matrices $\mathbf{B}(j')$, we can write

$$\begin{aligned} cov\{\mathbf{A}(k')\mathbf{u}(k), \mathbf{v}(j)\} &= \mathbf{A}(k')\mathbf{P}_{\mathbf{uv}}(k,j) && (3.34a) \\ cov\{\mathbf{u}(k), \mathbf{B}(j')\mathbf{v}(j)\} &= \mathbf{P}_{\mathbf{uv}}(k,j)\mathbf{B}^T(j') && (3.34b) \\ cov\{\mathbf{A}(k')\mathbf{u}(k), \mathbf{B}(j')\mathbf{v}(j)\} &= \mathbf{A}(k')\mathbf{P}_{\mathbf{uv}}(k,j)\mathbf{B}^T(j') && (3.34c) \end{aligned}$$

which can be demonstrated in an analogous way as done for the continuous-time case.

Using the relations above and (3.30), the covariance $\mathbf{P}_{\mathbf{u}}(k,j)$ can be calculated by

$$\begin{aligned} \mathbf{P}_{\mathbf{u}}(k,j) &\equiv cov\{\mathbf{u}(k)\mathbf{u}(j)\} \\ &= \mathbf{\Psi}(k,0)\mathbf{P}_{\mathbf{u}}(0)\mathbf{\Psi}^T(j,0) \\ &\quad + \mathbf{\Psi}(k,0)\sum_{\ell=0}^{j-1} cov\{\mathbf{u}(0), \mathbf{w}(\ell)\}\mathbf{\Gamma}^T(\ell)\mathbf{\Psi}^T(j,\ell+1) \\ &\quad + \sum_{i=0}^{k-1} \mathbf{\Psi}(k,i+1)\mathbf{\Gamma}(i) cov\{\mathbf{w}(i), \mathbf{u}(0)\}\mathbf{\Psi}^T(j,0) \\ &\quad + \sum_{i=0}^{k-1}\sum_{\ell=0}^{j-1} \mathbf{\Psi}(k,i+1)\mathbf{\Gamma}(i)\mathbf{P}_{\mathbf{w}}(i,\ell)\mathbf{\Gamma}^T(\ell)\mathbf{\Psi}^T(j,\ell+1)\,, \qquad (3.35) \end{aligned}$$

where $\mathbf{P}_{\mathbf{w}}(k,j) = cov\{\mathbf{w}(k), \mathbf{w}(j)\}$. In this way, the assumption of decorrelation (3.26) leads us to write

$$\begin{aligned} \mathbf{P}_{\mathbf{u}}(k,j) &= \mathbf{\Psi}(k,0)\mathbf{P}_{\mathbf{u}}(0)\mathbf{\Psi}^T(j,0) \\ &\quad + \sum_{i=0}^{k-1}\sum_{\ell=0}^{j-1} \mathbf{\Psi}(k,i+1)\mathbf{\Gamma}(i)\mathbf{P}_{\mathbf{w}}(i,\ell)\mathbf{\Gamma}^T(\ell)\mathbf{\Psi}^T(j,\ell+1)\,, \qquad (3.36) \end{aligned}$$

which is the general expression for the covariance of the discrete process (3.23).

As for the continuous case, we can obtain a simpler expression for the covariance if we consider the sequence $\{\mathbf{w}(k)\}$ to be white. Therefore, for the case in which

$$\mathbf{P}_{\mathbf{w}}(k,j) = \mathbf{Q}_k\delta_{k,j}\,, \qquad (3.37)$$

where $\delta_{k,j}$ is the Kronecker delta, the equation for the covariance is reduced to:

$$\begin{aligned} \mathbf{P}_{\mathbf{u}}(k,j) &= \mathbf{\Psi}(k,0)\mathbf{P}_{\mathbf{u}}(0)\mathbf{\Psi}^T(j,0) \\ &\quad + \sum_{i=0}^{\min(j-1,k-1)} \mathbf{\Psi}(k,i+1)\mathbf{\Gamma}(i)\mathbf{Q}_i\mathbf{\Gamma}^T(i)\mathbf{\Psi}^T(j,i+1)\,. \qquad (3.38) \end{aligned}$$

A recursive expression for the variance $\mathbf{P}_{\mathbf{u}}(k+1)$ can be obtained directly from the definition of variance and from (3.23). That is, by forming the outer product of (3.23) with itself and by applying the ensemble mean operator it follows that

$$\begin{aligned} \mathbf{P}_{\mathbf{u}}(k+1) &= \mathbf{\Psi}(k+1,k)\mathbf{P}_{\mathbf{u}}(k)\mathbf{\Psi}^T(k+1,k) \\ &\quad + \mathbf{\Psi}(k+1,k) cov\{\mathbf{u}(k), \mathbf{w}(k)\}\mathbf{\Gamma}^T(k) \\ &\quad + \mathbf{\Gamma}(k) cov\{\mathbf{w}(k), \mathbf{u}(k)\}\mathbf{\Psi}^T(k+1,k) \\ &\quad + \mathbf{\Gamma}(k)\mathbf{P}_{\mathbf{w}}(k)\mathbf{\Gamma}^T(k)\,. \qquad (3.39) \end{aligned}$$

Since $\mathbf{u}(k)$ depends only on $\mathbf{w}(j)$, for $j < k$, the second and third terms of the expression above vanish, so that the variance can be written as

$$\mathbf{P}_{\mathbf{u}}(k+1) = \mathbf{\Psi}(k+1,k)\mathbf{P}_{\mathbf{u}}(k)\mathbf{\Psi}^T(k+1,k) + \mathbf{\Gamma}(k)\mathbf{P}_{\mathbf{w}}(k)\mathbf{\Gamma}^T(k)\,, \tag{3.40}$$

which the corresponding discrete Lyapunov equation.

For a white noise sequence $\{\mathbf{w}(k)\}$ we have $\mathbf{P}_{\mathbf{w}}(k) = \mathbf{P}_{\mathbf{w}}(k,k) = \mathbf{Q}_k$, and also

$$\mathbf{P}_{\mathbf{u}}(k,j) = \begin{cases} \mathbf{\Psi}(k,j)\mathbf{P}_{\mathbf{u}}(j) & k \geq j \\ \mathbf{P}_{\mathbf{u}}(k)\mathbf{\Psi}^T(j,k) & k \leq j \end{cases}\,, \tag{3.41}$$

which can be verified by substitution of (3.40) into (3.38). That is, consider the case $k \geq j$, then:

$$\begin{aligned}
\mathbf{P}_{\mathbf{u}}(k,j) &= \mathbf{\Psi}(k,0)\mathbf{P}_{\mathbf{u}}(0)\mathbf{\Psi}^T(j,0) \\
&\quad + \sum_{i=0}^{j-1} \mathbf{\Psi}(k,i+1)\,[\mathbf{P}_{\mathbf{u}}(i+1) \\
&\qquad - \mathbf{\Psi}(i+1,i)\mathbf{P}_{\mathbf{u}}(i)\mathbf{\Psi}^T(i+1,i)\big]\,\mathbf{\Psi}^T(j,i+1) \\
&= \mathbf{\Psi}(k,0)\mathbf{P}_{\mathbf{u}}(0)\mathbf{\Psi}^T(j,0) \\
&\quad + \sum_{i=0}^{j-1} \mathbf{\Psi}(k,i+1)\mathbf{P}_{\mathbf{u}}(i+1)\mathbf{\Psi}^T(j,i+1) \\
&\quad - \sum_{i=0}^{j-1} \mathbf{\Psi}(k,i)\mathbf{P}_{\mathbf{u}}(i)\mathbf{\Psi}^T(j,i) \\
&= \sum_{i=0}^{j-1} \mathbf{\Psi}(k,i+1)\mathbf{P}_{\mathbf{u}}(i+1)\mathbf{\Psi}^T(j,i+1) - \sum_{i=1}^{j-1} \mathbf{\Psi}(k,i)\mathbf{P}_{\mathbf{u}}(i)\mathbf{\Psi}^T(j,i) \\
&= \sum_{i=1}^{j-1} \mathbf{\Psi}(k,i)\mathbf{P}_{\mathbf{u}}(i)\mathbf{\Psi}^T(j,i) + \mathbf{\Psi}(k,j)\mathbf{P}_{\mathbf{u}}(j)\mathbf{\Psi}^T(j,j) \\
&\quad - \sum_{i=1}^{j-1} \mathbf{\Psi}(k,i)\mathbf{P}_{\mathbf{u}}(i)\mathbf{\Psi}^T(j,i)
\end{aligned} \tag{3.42}$$

where, the two sums in the last equality cancel, and we used $\mathbf{\Psi}(j,j) = \mathbf{I}$ to obtain the desired result. The case $k \leq j$ can be obtained in an analogous manner.

3.1.3 Relation between the Continuous and Discrete Cases

A fundamental relation between continuous white noise and discrete white noise is that, as the sample of the discrete stochastic process becomes dense, the covariance of the discrete white process:

$$cov\{\mathbf{w}(kT), \mathbf{w}(jT)\} = \mathbf{Q}_k\delta_{k,j} \tag{3.43}$$

becomes the covariance of the continuous process:

$$cov\{\mathbf{w}(t), \mathbf{w}(\tau)\} = \lim_{k,j\to\infty} cov\{\mathbf{w}(kT), \mathbf{w}(jT)\} = \mathbf{Q}(t)\delta(t-\tau) \tag{3.44}$$

where the limit is also taken for $kT \to t$, $jT \to \tau$, and also for $T \to 0$. The variances in the expressions above are related by

$$\mathbf{Q}(t = kT) = T\,\mathbf{Q}_k \tag{3.45}$$

where some care should be taken with the notation used here: in spite of the fact that the variance matrices for the discrete and continuous processes above are represented by the same letter, they are in fact distinct matrices; the distinction is made by using subscripts in the discrete case $\mathbf{Q}_k$, in contrast to the explicit functional time dependence $\mathbf{Q}(t)$ for the continuous case.

The transition matrix of the continuous system can be written formally as

$$\mathbf{\Psi}(t, t_0) = \exp\{\int_{t_0}^{t} \mathbf{F}(s)\, ds\} \tag{3.46}$$

so that, for $t = (k+1)T$ and $t_0 = kT$, we can write

$$\begin{aligned}\mathbf{\Psi}((k+1)T, kT) &= \mathbf{I} + \int_{kT}^{(k+1)T} \mathbf{F}(s)\, ds \\ &\approx \mathbf{I} + \mathbf{F}(kT)T\,, \end{aligned} \tag{3.47}$$

where we made a gross approximation of the integral — which becomes reasonable as the sample becomes dense.

Substituting (3.45) and (3.47) in (3.40) we have that

$$\mathbf{P}_{\mathbf{u}}((k+1)T) = [\mathbf{I} + T\mathbf{F}(kT)]\mathbf{P}_{\mathbf{u}}(kT)[\mathbf{I} + T\mathbf{F}(kT)]^T + T\mathbf{G}(kT)\mathbf{Q}(kT)\mathbf{G}^T(kT)\,, \tag{3.48}$$

where we made the correspondence: $\mathbf{G}(t = kT) = \mathbf{\Gamma}(kT)/T$. The expression above can be also written as

$$\begin{aligned}\mathbf{P}_{\mathbf{u}}((k+1)T) &= \mathbf{P}_{\mathbf{u}}(kT) + T\mathbf{F}(kT)\mathbf{P}_{\mathbf{u}}(kT) + T\mathbf{P}_{\mathbf{u}}(kT)\mathbf{F}^T(kT) \\ &\quad + T\mathbf{G}(kT)\mathbf{Q}(kT)\mathbf{G}^T(kT) + o(T^2)\,, \end{aligned} \tag{3.49}$$

so that in the limit $T \to 0$, and $kT \to t$, we have

$$\begin{aligned}\dot{\mathbf{P}}_{\mathbf{u}}(t) &= \lim_{T\to 0} \frac{\mathbf{P}_{\mathbf{u}}((k+1)T) - \mathbf{P}_{\mathbf{u}}(kT)}{T} \\ &= \mathbf{F}(t)\mathbf{P}_{\mathbf{u}}(t) + \mathbf{P}_{\mathbf{u}}(t)\mathbf{F}^T(t) + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}^T(t)\,, \end{aligned} \tag{3.50}$$

where we retained only the terms of lower order in T. This means that the limit for the discrete variance evolution equation as the sample time becomes dense is given by the Lyapunov equation.

3.2 Nonlinear Dynamical Systems

3.2.1 Continuous Processes

Consider now the system of nonlinear differential equations for the random n–vector $\mathbf{u}(t)$,

$$\frac{d\mathbf{u}(t)}{dt} = \mathbf{f}[\mathbf{u}(t), t] + \mathbf{G}[\mathbf{u}(t), t]\mathbf{w}(t)\,, \tag{3.51}$$

where $\mathbf{w}(t)$ is a random m–vector white in time, with mean zero and (co)variance $\mathbf{Q}(t)$, that is,

$$\begin{aligned} \mathcal{E}\{\mathbf{w}(t)\} &= \mathbf{0} & (3.52a) \\ cov\{\mathbf{w}(t), \mathbf{w}(\tau)\} &= \mathbf{Q}(t)\delta(t-\tau) & (3.52b) \end{aligned}$$

moreover, we also assume $\mathbf{w}(t)$ to be Gaussian. The function $\mathbf{G}[\mathbf{u}(t), t]$ is an $n \times m$ matrix.

Formally, the solution of the equation above can be written in the form

$$\mathbf{u}(t) = \mathbf{u}(t_0) + \int_{t_0}^{t} \mathbf{f}[\mathbf{u}(s), s]\, ds + \int_{t_0}^{t} \mathbf{G}[\mathbf{u}(s), s]\, d\mathbf{b}(s)\,, \qquad (3.53)$$

where $\{\mathbf{b}(t)\}$ is the Wiener process defined in Section 2.8, since we assumed $\{\mathbf{w}(t)\}$ to be white and Gaussian. The first integral in the solution (3.53) is an ordinary integral (in the sense of Riemann); however, the second integral is questionable, since it involves increments $d\mathbf{b}$ of a function which is not necessarily finite. This last integration can be accomplished by employing possible generalizations of the concept of Lebesgue–Stieltjes integrals to the stochastic realm. One of these generalizations is due to Ito, and it defines what is called stochastic integral calculus.

The treatment of stochastic integrals is beyond the scope of what we intend to cover in this course. It is worth saying that for the case in which the matrix function $\mathbf{G}$ is independent of the process $\mathbf{u}(t)$, there is no difference between the stochastic integral calculus and ordinary calculus, in reference to solving the last integral in (3.53). Therefore, from this point on let us consider a simplified version of (3.51) given by

$$\frac{d\mathbf{u}(t)}{dt} = \mathbf{f}[\mathbf{u}(t), t] + \mathbf{G}(t)\mathbf{w}(t)\,. \qquad (3.54)$$

As in the previous sections we are interested in determining the moments of the statistics of the process $\{\mathbf{u}(t)\}$. We know that in the *linear* case, the assumption of a Gaussian driving forcing $\{\mathbf{w}(t)\}$ implies that the process $\{\mathbf{u}(t)\}$ is Gaussian as well, for Gaussian $\{\mathbf{u}(0)\}$. For this reason we concentrated on deriving equations for the first two moments in the previous sections. In the *nonlinear* case, the Gaussian assumptions on $\{\mathbf{u}(0)\}$ and $\{\mathbf{w}(t)\}$ do not guarantee the process $\{\mathbf{u}(t)\}$ to be Gaussian, consequently, even for Gaussian initial condition and driving forcing all moments are required in principle to describe the statistics of the process $\{\mathbf{u}(t)\}$ completely. However, to keep the calculations simple, we are still only going to concentrate in deriving equations for the first two moments of $\{\mathbf{u}(t)\}$.

The easiest way to obtain an expression for the mean is to apply the ensemble mean operator to equation (3.54). Proceeding this way, it follows that

$$\frac{d\boldsymbol{\mu}_{\mathbf{u}}(t)}{dt} = \mathcal{E}\{\mathbf{f}[\mathbf{u}(t), t]\}\,, \qquad (3.55)$$

where we used the fact that the process $\{\mathbf{w}(t)\}$ has mean zero. To determine an explicit expression for the right-hand side of the equation above, we expand the function $\mathbf{f}$ about the mean $\boldsymbol{\mu}_{\mathbf{u}}$. In this way, a Taylor expansion up to the second order yields,

$$\begin{aligned} \mathbf{f}[\mathbf{u}(t), t] \approx\ & \mathbf{f}[\boldsymbol{\mu}_{\mathbf{u}}(t), t] + \mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}(t), t](\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \\ & + \frac{1}{2}\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(t), t]\,(\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \otimes (\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \qquad (3.56) \end{aligned}$$

where $\mathcal{F}'$ is the $n \times n$ gradient (Jacobian) matrix of $\mathbf{f}$ given by

$$\begin{aligned}\mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}(t), t] &= \left.\frac{\partial \mathbf{f}[\mathbf{u}(t), t]}{\partial \mathbf{u}^T(t)}\right|_{\mathbf{u}(t)=\boldsymbol{\mu}_{\mathbf{u}}(t)} \\ &= \left.\left[\frac{\partial \mathbf{f}[\mathbf{u}(t), t]}{\partial u_1} \vdots \frac{\partial \mathbf{f}[\mathbf{u}(t), t]}{\partial u_2} \vdots \cdots \vdots \frac{\partial \mathbf{f}[\mathbf{u}(t), t]}{\partial u_n}\right]\right|_{\mathbf{u}(t)=\boldsymbol{\mu}_{\mathbf{u}}(t)} \end{aligned} \tag{3.57}$$

and $\mathcal{F}''$ is the $n \times n^2$ Hessian matrix given by

$$\begin{aligned}\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(t), t] &= \left.\frac{\partial^2 \mathbf{f}[\mathbf{u}(t), t]}{\partial \mathbf{u}^T(t) \partial \mathbf{u}^T(t)}\right|_{\mathbf{u}(t)=\boldsymbol{\mu}_{\mathbf{u}}(t)} \\ &= \left.\frac{\partial \mathcal{F}'[\mathbf{u}(t), t]}{\partial \mathbf{u}^T(t)}\right|_{\mathbf{u}(t)=\boldsymbol{\mu}_{\mathbf{u}}(t)} \\ &= \left.\left[\frac{\partial \mathcal{F}'[\mathbf{u}(t), t]}{\partial u_1} \vdots \frac{\partial \mathcal{F}'[\mathbf{u}(t), t]}{\partial u_2} \vdots \cdots \vdots \frac{\partial \mathcal{F}'[\mathbf{u}(t), t]}{\partial u_n}\right]\right|_{\mathbf{u}(t)=\boldsymbol{\mu}_{\mathbf{u}}(t)} \end{aligned} \tag{3.58}$$

Here we are using Vetter's notation [136] and [137] for the calculus of matrices (see also Brewer [18] for an overview of matrix calculus). The operation $\otimes$ represents the Kronecker product for matrices, which for any $n \times m$ matrix $\mathbf{A}$ and $p \times q$ matrix $\mathbf{B}$ is defined as

$$\mathbf{A} \otimes \mathbf{B} \equiv \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \dots & a_{nm}\mathbf{B} \end{pmatrix} \tag{3.59}$$

where a_{ij} is the (i, j)–th element of $\mathbf{A}$. The result $\mathbf{A} \otimes \mathbf{B}$ is a matrix of dimension $np \times mq$.

Using this definition, the Kronecker product of a general n–vector $\mathbf{v}$ with itself can be written as

$$\mathbf{v} \otimes \mathbf{v} = \begin{pmatrix} v_1\mathbf{v} \\ v_2\mathbf{v} \\ \vdots \\ v_n\mathbf{v} \end{pmatrix} \tag{3.60}$$

which is a column vector of dimension $n^2 = n^2 \times 1$. Furthermore, let us introduce the notation $vec(.)$ to represent the vector (column string) constructed from the columns of a general $n \times m$ matrix $\mathbf{A}$ as

$$\begin{aligned} vec(\mathbf{A}) &\equiv \left[\mathbf{a}_1^T\, \mathbf{a}_2^T \cdots \mathbf{a}_m^T\right]^T \\ &= \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{pmatrix} \end{aligned} \tag{3.61}$$

where $\mathbf{a}_i$, for $i = 1, \cdots, m$, is the i–th n–dimensional column of the matrix $\mathbf{A}$. According to this definition, $vec(\mathbf{A})$ is a vector of dimension $nm = nm \times 1$. Using this notation, it follows that (3.60) we be written as

$$\mathbf{v} \otimes \mathbf{v} = vec[\mathbf{v}\mathbf{v}^T] \tag{3.62}$$

Hence, referring back to (3.56), we see that

$$[\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)] \otimes [\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)] = vec[(\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t))(\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t))^T] \tag{3.63}$$

It is relatively simple to verify that the second-order term in (3.56) can be written explicitly as

$$\begin{aligned} &\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(t), t][\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)] \otimes [\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)] \\ = &\sum_{i=1}^{n}\sum_{j=1}^{n} [u_i(t) - \mu_i(t)][u_j(t) - \mu_j(t)] \left.\frac{\partial^2 \mathbf{f}[\mathbf{u}(t), t]}{\partial u_i(t) \partial u_j(t)}\right|_{\mathbf{u}(t)=\boldsymbol{\mu}_{\mathbf{u}}(t)} \end{aligned} \tag{3.64}$$

where, to simplify the notation, the subscript $\mathbf{u}$ was neglected when we wrote the i–th element μ_i of the mean of $\mathbf{u}(t)$, in the expression above.

Consequently, the equation for the mean, after application of the ensemble mean operator, reduces to

$$\frac{d\boldsymbol{\mu}_{\mathbf{u}}(t)}{dt} = \mathbf{f}[\boldsymbol{\mu}_{\mathbf{u}}(t), t] + \frac{1}{2}\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(t), t]\, vec[\mathbf{P}_{\mathbf{u}}(t)]\,, \tag{3.65}$$

where we notice that the first-order term in the Taylor expansion of $\mathbf{f}[\mathbf{u}(t), t]$ is automatically canceled. Therefore, the evolution of the mean depends on the variance $\mathbf{P}_{\mathbf{u}}(t)$. This is an unpleasant property of nonlinear systems: the evolution of moments of a given order depends on moments of higher order. To solve the equation above, it is necessary to determine an expression for $\mathbf{P}_{\mathbf{u}}(t)$. As we will observe below, this expression also depends on still higher-order moments, and so on. Consequently, depending on the nonlinearities in $\mathbf{f}[\mathbf{u}(t), t]$, it may be impossible to obtain a closed system of equations that determines completely the statistics of the process. In practice, we seek approximations, known as closures, for the equations of the desired moments, in order to obtain a solvable (closed) system of equations.

A simple approximation for the mean equation is to use only up to the first-order term in the expansion of the function $\mathbf{f}[\mathbf{u}(t), t]$. This leads us to the following expression for the evolution of the mean:

$$\frac{d\boldsymbol{\mu}_{\mathbf{u}}(t)}{dt} = \mathbf{f}[\boldsymbol{\mu}_{\mathbf{u}}(t), t]\,, \tag{3.66}$$

which is a closed equation — it is not coupled to any other equation.

To find an equation for the variance we can differentiate its definition, that is,

$$\begin{aligned} \frac{d\mathbf{P}_{\mathbf{u}}(t)}{dt} &= \frac{d\mathcal{E}\{[\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)][\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]^T\}}{dt} \\ &= \mathcal{E}\left\{[\dot{\mathbf{u}}(t) - \dot{\boldsymbol{\mu}}_{\mathbf{u}}(t)][\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]^T + [\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)][\dot{\mathbf{u}}(t) - \dot{\boldsymbol{\mu}}_{\mathbf{u}}(t)]^T\right\} \\ &= \mathcal{E}\Big\{[\mathbf{f}[\mathbf{u}(t), t] - \dot{\boldsymbol{\mu}}_{\mathbf{u}}(t) + \mathbf{G}(t)\mathbf{w}(t)]\,[\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]^T \\ &\quad + [\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]\,[\mathbf{f}[\mathbf{u}(t), t] - \dot{\boldsymbol{\mu}}_{\mathbf{u}}(t) + \mathbf{G}(t)\mathbf{w}(t)]^T\Big\} \\ &= \mathcal{E}\Big\{[\mathbf{f}[\mathbf{u}(t), t] - \dot{\boldsymbol{\mu}}_{\mathbf{u}}(t)]\,[\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]^T + \mathbf{G}(t)\mathbf{w}(t)[\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]^T \\ &\quad + [\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]\,[\mathbf{f}[\mathbf{u}(t), t] - \dot{\boldsymbol{\mu}}_{\mathbf{u}}(t)]^T + [\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]\mathbf{w}^T(t)\mathbf{G}^T(t)\Big\} \end{aligned} \tag{3.67}$$

where again we use the notation $(\dot{\;})$ to indicate differentiation with respect to time. This expression becomes very complicated when taken to high order in the expansion of $\mathbf{f}$ about the mean $\boldsymbol{\mu}_{\mathbf{u}}(t)$. This would lead to an equation for the variance depending on moments of higher order, thus not being a closed equation, just as it happened for the mean equation.

For the sake of simplicity, let us consider the second-order approximation (3.56) for $\mathbf{f}[\mathbf{u}(t),t)]$, as well as the second-order approximation for the evolution of the mean (3.65), so that we can write

$$\begin{aligned}
\mathbf{f}[\mathbf{u}(t),t] - \dot{\boldsymbol{\mu}}_{\mathbf{u}}(t) &= \mathbf{f}[\boldsymbol{\mu}_{\mathbf{u}}(t),t] + \mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}(t),t](\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \\
&\quad + \frac{1}{2}\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(t),t]\,(\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \otimes (\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \\
&\quad - \mathbf{f}[\boldsymbol{\mu}_{\mathbf{u}}(t),t] - \frac{1}{2}\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(t),t]\,vec[\mathbf{P}_{\mathbf{u}}(t)] \\
&= \mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}(t),t](\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \\
&\quad + \frac{1}{2}\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(t),t]\,\{\,(\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \otimes (\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \\
&\qquad\qquad - vec[\mathbf{P}_{\mathbf{u}}(t)]\,\} \qquad (3.68)
\end{aligned}$$

Substituting this result into (3.67) yields

$$\begin{aligned}
\frac{d\mathbf{P}_{\mathbf{u}}(t)}{dt} &= \mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}(t),t]\mathbf{P}_{\mathbf{u}}(t) + \mathbf{P}_{\mathbf{u}}(t)\mathcal{F}'^{T}[\boldsymbol{\mu}_{\mathbf{u}}(t),t] \\
&\quad + \frac{1}{2}\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(t),t]\mathcal{E}\,\{[(\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \otimes (\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \\
&\qquad\qquad - vec[\mathbf{P}_{\mathbf{u}}(t)]]\,(\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t))^{T}\Big\} \\
&\quad + \frac{1}{2}\mathcal{E}\,\{(\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t))\,[(\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \otimes (\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)) \\
&\qquad\qquad - vec[\mathbf{P}_{\mathbf{u}}(t)]]^{T}\Big\}\,\mathcal{F}''^{T}[\boldsymbol{\mu}_{\mathbf{u}}(t),t] \\
&\quad + \mathbf{G}(t)\mathcal{E}\{\mathbf{w}(t)\mathbf{u}^{T}(t)\} + \mathcal{E}\{\mathbf{u}(t)\mathbf{w}^{T}(t)\}\mathbf{G}^{T}(t)\,. \qquad (3.69)
\end{aligned}$$

The third and fourth terms of this expression refer to the third-order moments. To determine a closed set of evolution equations for the mean and variance, we ignore moments of order higher than two. Therefore, it follows that

$$\begin{aligned}
\frac{d\mathbf{P}_{\mathbf{u}}(t)}{dt} &= \mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}(t),t]\mathbf{P}_{\mathbf{u}}(t) + \mathbf{P}_{\mathbf{u}}(t)\mathcal{F}'^{T}[\boldsymbol{\mu}_{\mathbf{u}}(t),t] \\
&\quad + \mathbf{G}(t)\mathcal{E}\{\mathbf{w}(t)\mathbf{u}^{T}(t)\} + \mathcal{E}\{\mathbf{u}(t)\mathbf{w}^{T}(t)\}\mathbf{G}^{T}(t)\,, \qquad (3.70)
\end{aligned}$$

where the terms containing explicitly the ensemble mean can be evaluated by means of the formal solution (3.53) corresponding to the equation (3.54). That is, by evaluating the term $\mathcal{E}\{\mathbf{w}(t)\mathbf{u}^{T}(t)\}$, we have:

$$\begin{aligned}
\mathcal{E}\{\mathbf{w}(t)\mathbf{u}^{T}(t)\} &= \mathcal{E}\{\mathbf{w}(t)\mathbf{u}^{T}(t_0)\} + \int_{t_0}^{t} \mathcal{E}\{\mathbf{w}(t)\mathbf{f}^{T}[\mathbf{u}(s),s]\}\,ds \\
&\quad + \int_{t_0}^{t} \mathcal{E}\{\mathbf{w}(t)\mathbf{w}^{T}(s)\}\mathbf{G}^{T}(s)\,ds \\
&= \int_{t_0}^{t} \mathcal{E}\{\mathbf{w}(t)\mathbf{w}^{T}(s)\}\mathbf{G}^{T}(s)\,ds\,, \qquad (3.71)
\end{aligned}$$

where the second equality is obtained by noticing that the first term at the right of the first equality vanishes since $\mathbf{u}(t_0)$ and $\mathbf{w}(t)$ are uncorrelated; and the second term at the right of the first equality vanishes when we use the expansion of $\mathbf{f}[\mathbf{u}(t), t]$ about the mean, because the process $\mathbf{w}(t)$ has mean zero and we disregard moments of order higher than two, that is,

$$\begin{aligned} \int_{t_0}^{t} \mathcal{E}\{\mathbf{w}(t)\mathbf{f}^T[\mathbf{u}(s), s]\}\, ds &= \int_{t_0}^{t} \mathcal{E}\{\mathbf{w}(t)\}\mathbf{f}^T[\boldsymbol{\mu}(s), s]\, ds \\ &\quad + \int_{t_0}^{t} \mathcal{E}\{\mathbf{w}(t)[\mathbf{u}(s) - \boldsymbol{\mu}_{\mathbf{u}}(s)]\}\mathcal{F}'^T[\boldsymbol{\mu}_{\mathbf{u}}(s), s]\, ds + \cdots \\ &= \int_{t_0}^{t} \mathcal{E}\{\mathbf{w}(t)\mathbf{u}(s)\}\mathcal{F}'^T[\boldsymbol{\mu}_{\mathbf{u}}(s), s]\, ds + \ldots \\ &\approx 0 \end{aligned} \tag{3.72}$$

where the last "equality" invoked second order moment closure.

Therefore, from the definition of the symmetric Dirac delta function in (3.17) we have that

$$\begin{aligned} \mathcal{E}\{\mathbf{w}(t)\mathbf{u}^T(t)\} &= \int_{t_0}^{t} \delta(t-s)\mathbf{Q}(s)\mathbf{G}^T(s)\, ds \\ &= \frac{1}{2}\mathbf{Q}(t)\mathbf{G}^T(t)\,. \end{aligned} \tag{3.73}$$

An analogous expression can be obtained for the transposed term so that the variance equation, to first order, becomes:

$$\frac{d\mathbf{P}_{\mathbf{u}}(t)}{dt} = \mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}(t), t]\mathbf{P}_{\mathbf{u}}(t) + \mathbf{P}_{\mathbf{u}}(t)\mathcal{F}'^T[\boldsymbol{\mu}_{\mathbf{u}}(t), t] + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}^T(t)\,. \tag{3.74}$$

It is relevant to stress that equation (3.74) for the evolution of the variance $\mathbf{P}_{\mathbf{u}}(t)$ is of second order in the difference $[\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]$. Therefore, it is more consistent to use the second-order expression (3.65), instead of the first-order expression in $[\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]$, given by (3.66), in order to calculate the evolution of both the mean and the variance. This is an important fact that is sometimes ignored in order to reduce the amount of calculation involved involved in solving the equation of the mean, since (3.66) requires less computational effort than (3.65).

3.2.2 Discrete Processes

The nonlinear discrete–time system equivalent to the nonlinear continuous–time system studied in the previous subsection is represented by the equation

$$\mathbf{u}(k+1) = \boldsymbol{\psi}[\mathbf{u}(k), k] + \boldsymbol{\Gamma}(k)\mathbf{w}(k)\,, \tag{3.75}$$

where the m–dimensional process $\{\mathbf{w}(k)\}$ has the same characteristics as that defined for the linear discrete–time case of Section 3.1.2; $\boldsymbol{\psi}[\mathbf{u}(k), k]$ is a nonlinear n–vector function of the n–vector state $\mathbf{u}(t)$. In the more general case, the $n \times m$ matrix $\boldsymbol{\Gamma}(k)$ can be a function of $\mathbf{u}(k)$; however, for reasons analogous to those stated in the continuous–time case, we only consider the simpler situation described by the system above.

By proceeding as in the continuous nonlinear case, it is relatively simple to show that second-order closure produces the following expressions for the evolution equations of the mean and the variance:

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{u}}(k+1) &= \boldsymbol{\psi}[\boldsymbol{\mu}_{\mathbf{u}}(k), k] + \frac{1}{2}\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(k), k]vec[\mathbf{P}_{\mathbf{u}}(k)] && (3.76a) \\ \mathbf{P}_{\mathbf{u}}(k+1) &= \mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}(k), k]\mathbf{P}_{\mathbf{u}}(k)\mathcal{F}'^T[\boldsymbol{\mu}_{\mathbf{u}}(k), k] + \boldsymbol{\Gamma}(k)\mathbf{Q}(k)\boldsymbol{\Gamma}^T(k) && (3.76b) \end{aligned}$$

where

$$\begin{aligned} \mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}(k), k] &= \left.\frac{\partial\boldsymbol{\psi}[\mathbf{u}(k), k]}{\partial\mathbf{u}^T(k)}\right|_{\mathbf{u}(k)=\boldsymbol{\mu}_{\mathbf{u}}(k)} && (3.77a) \\ \mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(k), k] &= \left.\frac{\partial^2\boldsymbol{\psi}[\mathbf{u}(k), k]}{\partial\mathbf{u}^T(k)\partial\mathbf{u}^T(k)}\right|_{\mathbf{u}(k)=\boldsymbol{\mu}_{\mathbf{u}}(k)} && (3.77b) \end{aligned}$$

are now the Jacobian and Hessian matrices, respectively. Higher-order equations can also be obtained, however, this goes beyond the scope of an introductory course.

3.3 Stochastic Nonlinear Partial Differential Equations

In this section we are interested in the case where the state variable $\mathbf{u}$ is a function not only of time, but also of space, that is, $\mathbf{u} = \mathbf{u}(\mathbf{r}, t)$, with $\mathbf{r} \in R^n$. In this case, the equations governing the state evolution are partial differential equations describing the behavior of a stochastic random field. Our goal here is to indicate concisely how to derive evolution equations for the mean and covariance of the random field. A rigorous treatment of stochastic partial differential equations is complicated, especially when boundary conditions are included in order to define the problem completely. In what follows, we only give a formal description of the problem, ignoring the mathematical details. Moreover, we consider only the scalar case, so that there is only one random field to refer to, that is, $u = u(\mathbf{r}, t)$. More complete, and mathematically precise descriptions are found in Omatu & Sienfeld [109] and in the collection of articles in Stavroulakis [125]. These treatments are geared toward estimation problems for systems governed by partial differential equations, known as distributed parameter systems. For our simple treatment, the scalar (or univariate) random field $u = u(\mathbf{r}, t)$ is continuous in both space and time. Recall that, since u is random we should have in mind that it also depends on a variable ω referring to the realizations of this field. The variable ω is kept implicit in order to maintain the notation as compact as possible, and compatible with our previous notation.

Consider the following system of governing equations:

$$\frac{\partial u(\mathbf{r}, t)}{\partial t} = f[u(\mathbf{r}, t)] + w(\mathbf{r}, t) \qquad (3.78)$$

where $f[u(\mathbf{r}, t)]$ is a scalar differential operator which involves spatial partial derivatives, possibly nonlinear in the variable u; the scalar function $w(\mathbf{r}, t)$ represents a stochastic forcing, which we assume to be white in time with mean $\mu_w(\mathbf{r}, t)$ and covariance $Q(\mathbf{r}, \mathbf{s}, t)$:

$$\begin{aligned} \mathcal{E}\{w(\mathbf{r}, t)\} &= \mu_w(\mathbf{r}, t) && (3.79a) \\ \mathcal{E}\{[w(\mathbf{r}, t) - \mu_w(\mathbf{r}, t)][w(\mathbf{s}, \tau) - \mu_w(\mathbf{s}, \tau)]\} &= Q(\mathbf{r}, \mathbf{s}, t)\delta(t - \tau) && (3.79b) \end{aligned}$$

with $\mathbf{s} \in R^n$. When referring to Q as covariance, we have in mind its spatial structure; we could refer to this quantity as a variance if we had in mind its temporal structure. Also, because we are dealing with a *scalar* stochastic random field, Q is *scalar* function, and not a matrix.

We assume further that the processes $u(\mathbf{r}, 0)$ and $w(\mathbf{r}, t)$ are uncorrelated, that is,

$$\mathcal{E}\{w(\mathbf{r}, t)u(\mathbf{s}, 0)\} = 0 \tag{3.80}$$

for all $\mathbf{r}, \mathbf{s} \in R^n$, and all times $t \geq 0$.

Proceeding as in the previous section, but now for the univariate case, an equation for the evolution of the mean, written here as $\mu(\mathbf{r}, t) \equiv \mathcal{E}\{u(\mathbf{r}, t)\}$, can be found by applying the ensemble mean operator directly to the governing equation (3.78). Therefore,

$$\frac{\partial \mu(\mathbf{r}, t)}{\partial t} = \mathcal{E}\{f[\mathbf{u}(\mathbf{r}, t)]\} + \mu_w(\mathbf{r}, t) \tag{3.81}$$

Expanding $f[u(\mathbf{r}, t)]$ in a Taylor series about its mean $\mu(\mathbf{r}, t)$ we have

$$\begin{aligned} f[u(\mathbf{r}, t), t] &\approx f[\mu(\mathbf{r}, t), t] + \mathcal{F}'[\mu(\mathbf{r}, t)](u(\mathbf{r}, t) - \mu(\mathbf{r}, t)) \\ &\quad + \frac{1}{2}\mathcal{F}''[\mu(\mathbf{r}, t)]\,(u(\mathbf{r}, t) - \mu(\mathbf{r}, t))^2 \end{aligned} \tag{3.82}$$

which is identical to the expansion (3.56), except for the fact that now the Jacobian $\mathcal{F}'$ and the Hessian $\mathcal{F}''$ are functions (differential operators), not matrices. These quantities are defined in an entirely analogous way as to the way we saw in the previous section, that is,

$$\mathcal{F}'[\mu(\mathbf{r}, t)] = \left.\frac{\partial f[u(\mathbf{r}, t)]}{\partial u(\mathbf{r}, t)}\right|_{u(\mathbf{r}, t) = \mu(\mathbf{r}, t)} \tag{3.83}$$

and

$$\mathcal{F}''[\mu(\mathbf{r}, t)] = \left.\frac{\partial^2 f[u(\mathbf{r}, t)]}{\partial u^2(\mathbf{r}, t)}\right|_{u(\mathbf{r}, t) = \mu(\mathbf{r}, t)} \tag{3.84}$$

Therefore, the equations for the mean and covariance, with second-order closure, are:

$$\frac{\partial \mu(\mathbf{r}, t)}{\partial t} = f[\mu(\mathbf{r}, t)] + \frac{1}{2}\mathcal{F}''[\mu(\mathbf{r}, t), t]\, P(\mathbf{r}, \mathbf{r}, t)\,, \tag{3.85}$$

and

$$\frac{\partial P(\mathbf{r}, \mathbf{s}, t)}{\partial t} = \mathcal{F}'[\mu(\mathbf{r}, t)]P(\mathbf{r}, \mathbf{s}, t) + \mathcal{F}'[\mu(\mathbf{s}, t)]P(\mathbf{r}, \mathbf{s}, t) + Q(\mathbf{r}, \mathbf{s}, t) \tag{3.86}$$

respectively. The equation for the mean involves the variance — covariance $P(\mathbf{r}, \mathbf{s}, t)$, for $\mathbf{s} = \mathbf{r}$. Details in obtaining these equations can be found in Cohn [28]. A simple case, taken from this work and that of Ménard [104] is given in exercises.

Exercises

1. Derive the continuous Lyapunov equation, for the linear system (3.1), in two distinct ways:

(a) Differentiating the solution (3.20) — use Leibnitz integration rule [1].

(b) Differentiating the definition of variance:

$$\mathbf{P}(t) \equiv \mathcal{E}\{[\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)][\mathbf{u}(t) - \boldsymbol{\mu}_{\mathbf{u}}(t)]^T\}$$

2. Show that the expressions (3.22) satisfy equation (3.19).

3. Consider general matrices **A**, **B** and **C** of dimensions $n \times m$, $m \times p$, and $p \times q$, respectively. Furthermore, notice that the product of two matrices **A** and **B** can be written as a column operation according to

$$(\mathbf{AB})_{\bullet j} = \mathbf{A}\mathbf{b}_j$$

where $\mathbf{b}_j$ is the j–th column of **B**, and the notation $(\bullet j)$ on the left–hand–side stands for the j–th column of the product matrix (**AB**). Representing the (i, j)-th element of **B** as b_{ij}, we can write the product of two matrices in the alternative form

$$(\mathbf{AB})_{\bullet j} = \sum_i (\mathbf{A})_{\bullet i} b_{ij}$$

With that in mind, engage in the following proofs:

(a) Show that:

$$vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B})$$

(b) Using the previous result, show that

$$vec(\mathbf{AB}) = (\mathbf{I}_p \otimes \mathbf{A})vec(\mathbf{B})$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix.

(c) By noticing that for matrices **A** and $\tilde{\mathbf{A}}$, of dimension $n \times m$,

$$vec(\mathbf{A} + \tilde{\mathbf{A}}) = vec(\mathbf{A}) + vec(\tilde{\mathbf{A}})\,,$$

show that the continuous–time Lyapunov equation (3.21) can be written as

$$vec(\dot{\mathbf{P}}_{\mathbf{u}}) = [\mathbf{F}(t) \otimes \mathbf{I}_n\; + \;\mathbf{I}_n \otimes \mathbf{F}(t)]\, vec(\mathbf{P}_{\mathbf{u}}) \;+\; [\mathbf{G}(t) \otimes \mathbf{G}(t)]\, vec(\mathbf{Q})$$

(equivalent to Problem 3.1–1 in Lewis [94].)

(d) Analogously, show that the discrete–time Lyapunov equation (3.40) can be written as

$$\begin{aligned} vec(\mathbf{P}_{\mathbf{u}}(k+1)) \;&=\; [\boldsymbol{\Psi}(k+1,k) \otimes \boldsymbol{\Psi}(k+1,k)]\, vec(\mathbf{P}_{\mathbf{u}}(k)) \\ &\quad + [\boldsymbol{\Gamma}(k) \otimes \boldsymbol{\Gamma}(k)]\, vec(\mathbf{Q}_k) \end{aligned}$$

(equivalent to Problem 2.2–1 in Lewis [94].)

[1] Leibnitz integration rule is

$$\frac{d}{dt}\int_{h(t)}^{g(t)} f(t,\tau)\,d\tau = \int_{h(t)}^{g(t)} \frac{\partial f(t,\tau)}{\partial t}\,d\tau + f[t, g(t)]\frac{dg(t)}{dt} - f[t, h(t)]\frac{dh(t)}{dt}$$

4. (Maybeck [101], Problem 2.15)

   (a) Show that, for all t_0, t_1, and t,

$$\mathbf{\Phi}(t, t_0) = \mathbf{\Phi}(t, t_1)\mathbf{\Phi}(t_1, t_0)$$

   by showing that both quantities satisfy the same linear differential equation and "initial condition" at time t_1. Thus, the solution of $\dot{\mathbf{u}}(t) = \mathbf{F}(t)\mathbf{u}(t)$ with $\mathbf{u}(t_0) = \mathbf{u}_0$ [i.e., $\mathbf{\Phi}(t, t_0)\mathbf{u}(t_0)$] at any time t_2 can be obtained by forming $\mathbf{u}(t_1) = \mathbf{\Phi}(t_1, t_0)\mathbf{u}(t_0)$ and using it to generate $\mathbf{u}(t_2) = \mathbf{\Phi}(t_2, t_1)\mathbf{u}(t_1)$.

   (b) Since it can be shown that $\mathbf{\Phi}(t, t_0)$ is non–singular, show that the above property implies that

$$\mathbf{\Phi}^{-1}(t, t_0) = \mathbf{\Phi}(t_0, t)$$

5. (Mostly from Maybeck [101], Problem 2.18) Given a homogeneous linear differential equation $\dot{\mathbf{u}}(t) = \mathbf{F}(t)\mathbf{u}(t)$, for the n–vector $\mathbf{u}(t)$, the associated "adjoint" differential equation is the differential equation for the n–vector $\mathbf{v}(t)$ such that the inner product of $\mathbf{v}(t)$ with $\mathbf{u}(t)$ is constant for all time:

$$\mathbf{u}^T(t)\mathbf{v}(t) = \text{const}$$

   (a) Take the derivative of this expression to show that the adjoint equation associated with $\dot{\mathbf{u}}(t) = \mathbf{F}(t)\mathbf{u}(t)$ is

$$\dot{\mathbf{v}}(t) = -\mathbf{F}^T(t)\mathbf{v}(t)$$

   (b) If $\mathbf{\Phi}_{\mathbf{u}}(t, t_0)$ is the state transition matrix associated with $\mathbf{F}(t)$ and $\mathbf{\Phi}_{\mathbf{v}}(t, t_0)$ is the state transition matrix associated with $[-\mathbf{F}^T(t)]$, then show that

$$\mathbf{\Phi}_{\mathbf{v}}(t, t_0) = \mathbf{\Phi}_{\mathbf{u}}^T(t_0, t) = [\mathbf{\Phi}_{\mathbf{u}}^T(t, t_0)]^{-1}$$

   To do this, show that $[\mathbf{\Phi}_{\mathbf{v}}^T(t, t_0)\mathbf{\Phi}_{\mathbf{u}}(t, t_0)]$ and $\mathbf{I}$ satisfy the same differential equation and initial condition.

   (c) Show that, as a function of its second argument, $\mathbf{\Phi}_{\mathbf{u}}(t, \tau)$ must satisfy

$$\frac{\partial \mathbf{\Phi}_{\mathbf{u}}(t, \tau)}{\partial \tau} = -\mathbf{\Phi}_{\mathbf{u}}(t, \tau)\mathbf{F}(\tau)$$

   or, in other words,

$$\frac{\partial \mathbf{\Phi}_{\mathbf{u}}^T(t, \tau)}{\partial \tau} = [-\mathbf{F}^T(\tau)]\mathbf{\Phi}_{\mathbf{u}}^T(t, \tau)$$

   (d) If the inner product to be preserved in time is modified to be

$$\mathbf{u}^T(t)\mathbf{E}\mathbf{v}(t) = \text{const}$$

   where the $n \times n$ matrix $\mathbf{E}$ is assumed to be invertible and independent of time, derive the corresponding modification to the adjoint equation in (a).

6. (Mostly from Maybeck [101], Problem 2.17) Let the $n \times n$ matrix $\mathbf{F}$ be constant. Then the evaluation of $\mathbf{\Phi}(t, t_0) = \mathbf{\Phi}(t - t_0)$ can be obtained by

(a) approximating through truncation of series definition of matrix exponential, $e^{\mathbf{F}(t-t_0)}$:

$$e^{\mathbf{F}(t-t_0)} = \mathbf{I} + \mathbf{F}(t-t_0) + \frac{1}{2!}\mathbf{F}^2(t-t_0)^2 + \ldots$$

(b) Laplace methods of solving $\dot{\mathbf{\Phi}}(t-t_0) = \mathbf{F}\mathbf{\Phi}(t-t_0)$, $\mathbf{\Phi}(0) = \mathbf{I}$:

$$\mathbf{\Phi}(t-t_0) = \mathcal{L}^{-1}\{[s\mathbf{I} - \mathbf{F}]^{-1}\}\Big|_{t-t_0}$$

where $\mathcal{L}^{-1}\{.\}|_{(t-t_0)}$ denotes inverse Laplace transform evaluated with time argument equal to $(t-t_0)$.

(c) Cayley–Hamilton theorem (for $\mathbf{F}$ with nonrepeated eigenvalues)

$$\mathbf{\Phi}(t-t_0) = \alpha_0\mathbf{I} + \alpha_1\mathbf{F} + \alpha_2\mathbf{F}^2 + \ldots + \alpha_{n-1}\mathbf{F}^{n-1}$$

To solve for the n functions of $(t-t_0)$, $\alpha_0, \alpha_1, \cdots, \alpha_{n-1}$, the n eigenvalues of $\mathbf{F}$ are determined as $\lambda_1, \cdots, \lambda_n$. Then

$$e^{\lambda_i(t-t_0)} = \alpha_0 + \alpha_1\lambda_i + \alpha_2\lambda_i^2 + \ldots + \alpha_{n-1}\lambda_i^{n-1}$$

must be satisfied for each eigenvalue λ_i, for $i = 1, \cdots, n$, yielding n equations for the n unknown α_i's.

(d) Sylvester expansion theorem (for $\mathbf{F}$ with nonrepeated eigenvalues)

$$\mathbf{\Phi}(t-t_0) = \mathbf{F}_1 e^{\lambda_1(t-t_0)} + \mathbf{F}_2 e^{\lambda_2(t-t_0)} + \ldots + \mathbf{F}_n e^{\lambda_n(t-t_0)}$$

where λ_i is the i–th eigenvalue of $\mathbf{F}$ and $\mathbf{F}_i$ is given as the following product of $(n-1)$ factors:

$$\mathbf{F}_i = \left[\frac{\mathbf{F} - \lambda_1\mathbf{I}}{\lambda_i - \lambda_1}\right] \cdots \left[\frac{\mathbf{F} - \lambda_{i-1}\mathbf{I}}{\lambda_i - \lambda_{i-1}}\right] \left[\frac{\mathbf{F} - \lambda_{i+1}\mathbf{I}}{\lambda_i - \lambda_{i+1}}\right] \cdots \left[\frac{\mathbf{F} - \lambda_n\mathbf{I}}{\lambda_i - \lambda_n}\right]$$

The matrix $\mathbf{F}_i$ is a projector onto the direction of the i th eigenvector of $\mathbf{F}$.

(e) If the eigendecompostion of $\mathbf{F}$ is given by

$$\mathbf{F} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$$

where $\mathbf{U}$ is the matrix whose columns are the eigenvectors of $\mathbf{F}$, and $\mathbf{D}$ is a diagonal matrix with the eigenvalues λ_i of $\mathbf{F}$ along the diagonal, that is, $\mathbf{D} = diag(\lambda_1, \ldots, \lambda_n)$, then, the Maclaurin expansion of item (a) above can be used to show that $\mathbf{\Phi}(t, t_0)$ has the same eigenvectors of $\mathbf{F}$ with eigenvalues $e^{\lambda_i(t-t_0)}$, that is,

$$\mathbf{\Phi}(t, t_0) = \mathbf{U}\begin{pmatrix} e^{\lambda_1(t-t_0)} & 0 & \ldots & 0 \\ 0 & e^{\lambda_2(t-t_0)} & \ldots & 0 \\ \ldots & \ldots & \ddots & \ldots \\ 0 & 0 & \ldots & e^{\lambda_n(t-t_0)} \end{pmatrix}\mathbf{U}^{-1}$$

Use the five methods[2] above to evaluate $\mathbf{\Phi}(t,0)$ if $\mathbf{F}$ is given by

$$\mathbf{F} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

7. Consider the Lorenz (1960; [99]) system of equations

$$\begin{aligned} \dot{X} &= aYZ \\ \dot{Y} &= bXZ \\ \dot{Z} &= cXY \end{aligned}$$

where a, b and c are constants to be specified later. Under certain conditions, Lorenz showed that the solution of this system is periodic with predictable period $2K$ for X,'and period $4K$ for Y and Z. The expression for K is a function of the initial condition $[X_0\, Y_0\, Z_0]$, and as a consequence, even though the system is deterministic, i.e., non–chaotic, the period of oscillation may change considerably due a small change in the initial condition. The amplitude of the oscillations may change as well.

(a) Assuming there is uncertain knowledge of the initial state, we can think on these equations as a set of stochastic differential equations. Thus, defining the 3–vector $\mathbf{u}(t) = [X\, Y\, Z]^T$ and writing the system of equations as

$$\dot{\mathbf{u}}(t) = \mathbf{f}(\mathbf{u})$$

derive approximate equations for the mean $\boldsymbol{\mu}_{\mathbf{u}}(t)$,

$$\boldsymbol{\mu}_{\mathbf{u}}(t) = \mathcal{E}\{\mathbf{u}(t)\} = [\mu_x\, \mu_y\, \mu_z]^T$$

and for the (co)variance $\mathbf{P}(t)$,

$$\begin{aligned} \mathbf{P}(t) &= \mathcal{E}\{(\mathbf{u}(t) - \boldsymbol{\mu}(t))(\mathbf{u}(t) - \boldsymbol{\mu}(t))^T\} \\ &= \begin{pmatrix} p_x & p_{xy} & p_{xz} \\ p_{xy} & p_y & p_{yz} \\ p_{xz} & p_{yz} & p_z \end{pmatrix} \end{aligned}$$

to second order. Notice that in this exercise we are taking $\mathbf{Q} = \mathbf{0}$. (Hint: We have already derived these equations for a general stochastic system of ordinary differential equations, so this is just an exercise of calculating the appropriate Jacobian and Hessian matrices.)

(b) *Computer Assignment:* [Ehrendorfer (1994a,b; [47, 48]), and Epstein (1969; [50])]

[2] For method (b) the following inverse Laplace transforms are useful:

$$\mathcal{L}^{-1}\left\{\frac{1}{s+a}\right\} = e^{-at}$$

$$\mathcal{L}^{-1}\left\{\frac{1}{(s+a)(s+b)}\right\} = \frac{e^{-at} - e^{-bt}}{b-a}$$

for $t > 0$, and $a \neq b$.

i. The solution of a system of ordinary differential equations of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t)$$

can be approximated by a fourth–order Runge–Kutta method for $\mathbf{x}_{n+1} \approx \mathbf{y}(t_{n+1})$ according to

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \frac{\Delta t}{3}(\frac{\mathbf{k}_1}{2} + \mathbf{k}_2 + \mathbf{k}_3 + \frac{\mathbf{k}_4}{2})$$

where

$$\begin{aligned} \mathbf{k}_1 &= \mathbf{f}(\mathbf{y}_n, t_n) \\ \mathbf{k}_2 &= \mathbf{f}(\mathbf{y}_n + \tfrac{\Delta t}{2}\mathbf{k}_1, t_n + \tfrac{\Delta t}{2}) \\ \mathbf{k}_3 &= \mathbf{f}(\mathbf{y}_n + \tfrac{\Delta t}{2}\mathbf{k}_2, t_n + \tfrac{\Delta t}{2}) \\ \mathbf{k}_4 &= \mathbf{f}(\mathbf{y}_n + \Delta t\mathbf{k}_3, t_n + \Delta t) \end{aligned}$$

for $t_{n+1} = t_n + \Delta t$, with Δt being the time step (e.g., Press et al. [115], pp. 550–554). Write a Matlab function that, given the initial time t_0, the final time t_f, the time step Δt, and the initial condition $\mathbf{y}_0$, finds the approximate solution of the system of ordinary differential equation for $\mathbf{x}(t)$, according to this differencing method[3].

ii. The solution of the Lorenz system for $\mathbf{u}(t)$ above, with $a = -0.1$, $b = 1.6$, and $c = -0.75$, and initial condition

$$\mathbf{u}_0 = \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix} = \begin{pmatrix} 0.12 \\ 0.24 \\ 0.10 \end{pmatrix}$$

has approximate period of 23.12 time units in $X(t)$, and approximate period of 46.24 time units in $Y(t)$ and $Z(t)$. Using the Matlab function you created in the previous item, solve the Lorenz system, for the parameters a, b and c, and initial condition given above, from time $t_0 = 0$ to $t_f = 250$, with a time step $\Delta t = 0.5$. Plot the $X(t)$, $Y(t)$ and $Z(t)$, as a function of time.

iii. Now, choose three distinct initial conditions generated as

$$\mathbf{u}(0) = \mathbf{u}_0 + \mathbf{w}$$

where the 3–vector $\mathbf{w}$ is normally distributed with mean zero and variance $0.01^2\mathbf{I}$, that is, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, 0.01^2\mathbf{I})$, with $\mathbf{I}$ being the 3×3 identity matrix. Plot the result of the three corresponding solutions versus time. Comment on the results you obtain here and those obtained in the previous item.

iv. Using the vector notation introduced in this lecture, and exploited in Exercise 3, we can write the mean and (co)variance equations of item (a) simultaneously as

$$\frac{d}{dt}\begin{pmatrix} \boldsymbol{\mu}_{\mathbf{u}}(t) \\ vec(\mathbf{P}_{\mathbf{u}}(t)) \end{pmatrix} = \mathbf{g}\begin{pmatrix} \boldsymbol{\mu}_{\mathbf{u}}(t) \\ vec(\mathbf{P}_{\mathbf{u}}(t)) \end{pmatrix}$$

[3] The Matlab functions $\boxed{\text{ode23}}$ and $\boxed{\text{ode45}}$ are "adaptive mesh" Runge–Kutta solvers for ordinary differential equations. The adaptive nature of these functions provide great accuracy, but do not allow for control of the number of time steps for a given integration interval. Since, in this exercise, we want to have the same number of time steps in all experiments, it is best to write our own Runge–Kutta solver.

where the function $\mathbf{g}$ is defined as

$$\mathbf{g}\begin{pmatrix} \boldsymbol{\mu}_{\mathbf{u}}(t) \\ vec(\mathbf{P}_{\mathbf{u}}(t)) \end{pmatrix} \equiv \begin{pmatrix} \mathbf{f}[\boldsymbol{\mu}_{\mathbf{u}}(t)] + \frac{1}{2}\mathcal{F}''[\boldsymbol{\mu}_{\mathbf{u}}(t)]vec(\mathbf{P}_{\mathbf{u}}(t)) \\ \{\mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}] \otimes \mathbf{I}_3 + \mathbf{I}_3 \otimes \mathcal{F}'[\boldsymbol{\mu}_{\mathbf{u}}]\}\, vec(\mathbf{P}_{\mathbf{u}}(t)) \end{pmatrix}$$

where $\mathbf{I}_3$ is the 3×3 identity matrix, and we noticed that since $\mathbf{f}$ does not depend explicitly on t, its Jacobian and Hessian matrices also do not depend on t explicitly. Identifying the vector $\mathbf{x}(t)$ as the 12–vector $[\boldsymbol{\mu}_{\mathbf{u}}\, vec(\mathbf{P}_{\mathbf{u}})]^T$, solve the equations for the evolution of the mean and (co)variance above, for the Lorenz system. Plot the results of the evolution of the mean $\boldsymbol{\mu}_{\mathbf{u}} = [\mu_u\, \mu_y\, \mu_z]$ versus time, as well as those for the variances p_x, p_y, p_z, and the cross–covariances p_{xy}, p_{xz} and p_{yz}.

v. Repeat the previous item, when the second–order correction term in the mean equation is neglected. Explain the difference from the results of the previous item.

(c) To really assess to correctness of the means and (co)variances obtained in the previous exercise it is necessary to perform Monte Carlo experiments, with extremely large samples or, alternatively, to solve the Liouville equation — Fokker–Planck equation when $\mathbf{Q} = \mathbf{0}$ — which is an equation for the time evolution of the probability density function related to the stochastic process under consideration. Not surprisingly, this latest approach has been shown by Ehrendorfer (1994a,b; [47, 48]), to provide the most reliable, and efficient, estimate of the moments of the probability distribution. This approach, however, is beyond the scope of our course, and for this reason we will rely on Monte Carlo experiments to assess reliability of the means and (co)variances obtained in the previous exercise.

i. Generating sample initial conditions in the same manner you generated the three distinct initial conditions in (*b.iii*), perform three Monte Carlo experiments which integrate the Lorenz equations for three distinct total number of ensemble members: 50, 100, and 200. Using the Matlab function $\boxed{\text{mean}}$ calculate the means $\mu_x^{MC=i}$, $\mu_y^{MC=i}$, and $\mu_z^{MC=i}$, where $i = 50, 100, 200$. Plot the results as a function of time. How do they compare with the evolution of the mean obtained in the previous exercise when the second order correction term was present in the mean equation?

ii. Using the Matlab function $\boxed{\text{cov}}$, calculate the variances $(\sigma_x^{MC=i})^2$, $(\sigma_y^{MC=i})^2$, and $(\sigma_z^{MC=i})^2$, for each sample size $i = 50, 100, 200$. Compare with the results obtained for p_x, p_y, and p_z of the previous exercise.

iii. Still using the same function $\boxed{\text{cov}}$, calculate the cross–covariances $cov(x, y)_{MC=i}$, $cov(x, z)_{MC=i}$, and $cov(y, z)_{MC=i}$ for each sample size $i = 50, 100, 200$. Compare with the results for p_{xy}, p_{xz}, and p_{yz} obtained in the previous exercise.

[Beware: We should really use a sample size of 10^4, or larger, to have a converged Monte Carlo run. However, this would only be feasible if we computed the means, variances, and cross–covariances on–line, that is, while running the time evolution. This would be the way to avoid the memory overload caused when storing the complete time history for each ensemble member, as we are doing in our experiments.]

8. Consider a system governed by the linear differential equation

$$\dot{\mathbf{u}}(t) = \mathbf{F}(t)\mathbf{u}(t)$$

and assume that stochasticity comes from the fact that we only know the initial condition to a certain degree. That is, the initial condition is

$$\mathbf{u}(0) = \mathbf{u}_0 + \boldsymbol{\epsilon}(0)$$

where $\boldsymbol{\epsilon}(0)$ has mean $\boldsymbol{\mu}_0$, and variance $\mathbf{P}_0$. We refer to $\boldsymbol{\epsilon}(0)$ as the initial error.

(a) Show that the error $\boldsymbol{\epsilon}(t)$, at time t, can be determined by

$$\boldsymbol{\epsilon}(t) = \boldsymbol{\Phi}(t, t_0)\boldsymbol{\epsilon}(0)$$

where $\boldsymbol{\Phi}(t, t_0)$ is the transition matrix related to the governing equation for $\mathbf{u}(t)$. Therefore, for linear dynamics $\mathbf{F}(t)$, the error evolves according to the same "law" as the state vector.

(b) Show that the ensemble average of the error $\boldsymbol{\mu}_{\boldsymbol{\epsilon}}(t) \equiv \mathcal{E}\{\boldsymbol{\epsilon}(t)\}$ evolves according to

$$\boldsymbol{\mu}_{\boldsymbol{\epsilon}}(t) = \boldsymbol{\Phi}(t, t_0)\boldsymbol{\mu}_{\boldsymbol{\epsilon}}(0)$$

and that the error variance $\mathbf{P}(t) \equiv \mathcal{E}\{[\boldsymbol{\epsilon}(t) - \boldsymbol{\mu}_{\boldsymbol{\epsilon}}(t)][\boldsymbol{\epsilon}(t) - \boldsymbol{\mu}_{\boldsymbol{\epsilon}}(t)]^T\}$ evolves according to

$$\mathbf{P}(t) = \boldsymbol{\Phi}(t, 0)\mathbf{P}_0\, \boldsymbol{\Phi}^T(t, 0)$$

This expression is the solution of the Lyapunov equation in the absence of $\mathbf{Q}$, and $\mathbf{P}(t)$ in this case is sometimes referred to as the predictability error (co)variance.

(c) In some applications it is important to determine which perturbations grow fastest within a given period of time. A measure of the growth of initial perturbations can be obtained by defining an amplification factor coefficient $A(t)$ as

$$\begin{aligned} A(t) &\equiv \frac{||\boldsymbol{\epsilon}(t)||^2}{||\boldsymbol{\epsilon}(0)||^2} \\ &= \frac{\boldsymbol{\epsilon}^T(t)\boldsymbol{\epsilon}(t)}{\boldsymbol{\epsilon}^T(0)\boldsymbol{\epsilon}(0)} \end{aligned}$$

(d) (Lacarra & Talagrand [91]) Going back to the time–independent matrix $\mathbf{F}$ of the previous problem, for which you have calculated the corresponding transition matrix $\boldsymbol{\Phi}(t, 0)$, perform the following tasks:

i. Calculate the amplification factor, at time $t = T$, for an initial vector $\boldsymbol{\epsilon}(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}^T$.

ii. What is the amplification factor corresponding to the largest eigenvalue of $\mathbf{F}$?

iii. Show that the eigenvectors of $\mathbf{F}$ are not orthogonal.

9. Consider the linear advection equation in R^1 for a univariate random field $u(x, t)$:

$$\frac{\partial u}{\partial t} + U\frac{\partial u}{\partial x} = 0$$

with initial condition

$$u(x, t = 0) = u_0(x)$$

where $U = const.$ represents the advection speed. Thus determine:

(a) The evolution equation for the mean $\mu(x,t) = \mathcal{E}\{u(x,t)\}$.

(b) The evolution equation for the covariance function between two points x e y, that is, $P(x,y,t) = \mathcal{E}\{[u(x,t) - \mu(x,t)][u(y,t) - \mu(y,t)]\}$.

10. (Cohn [28] and Ménard [104]) Consider Burger's equation in one spatial dimension, for $u = u(x,t)$:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \alpha\frac{\partial^2 u}{\partial x^2} + w$$

where $\alpha = const.$ and $w = w(x,t)$ is a stochastic forcing term white with zero mean and covariance $Q(x,y,t)$. Obtain the evolution equation for the mean $\mu(x,t)$ and for the covariance $P(x,y,t)$, up to second order. From the covariance equation, obtain the evolution equation for the variance field. (Hint: There is no need to recalculate all the equations as if nothing was known. The intention here is to apply directly the results obtained in the end of this chapter).

Chapter 4

Introduction to Estimation Theory

4.1 Concepts of Probabilistic Estimation

The problem we are interest in this lecture is that of estimating the value of an n-dimensional vector of parameters $\mathbf{w}$, of a given system, on the basis of p observations taken on these parameters, and stacked in a p dimensional observation vector $\mathbf{z}$. We refer to $\hat{\mathbf{w}}$ as the estimate of the vector of parameters $\mathbf{w}$ under investigation, and we refer to the quantity $\tilde{\mathbf{w}} = \hat{\mathbf{w}} - \mathbf{w}$ as the estimation error. Based on the statistical formulation of the problem, we assume that the observational process is imperfect, and therefore the observations can be considered realizations of a random variable. Analogously, the vector of parameters $\mathbf{w}$ is seen as a quantity belonging to realizations of another random vector.

4.1.1 Bayesian Approach

In Bayesian estimation theory we introduce a functional $\mathcal{J}$ which corresponds to a measure of the "risk" involved in the estimate obtained for the parameter $\mathbf{w}$. That is, we define

$$\begin{aligned} \mathcal{J}(\hat{\mathbf{w}}) &\equiv \mathcal{E}\{J(\tilde{\mathbf{w}})\} \\ &= \int_{-\infty}^{\infty} J(\tilde{\mathbf{w}})\, p_{\mathbf{w}}(\mathbf{w})\, d\mathbf{w} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(\tilde{\mathbf{w}})\, p_{\mathbf{wz}}(\mathbf{w},\mathbf{z})\, d\mathbf{z}\, d\mathbf{w} \end{aligned} \tag{4.1}$$

where $p_{\mathbf{w}}(\mathbf{w})$ is the marginal probability density of $\mathbf{w}$, $p_{\mathbf{wz}}(\mathbf{w},\mathbf{z})$ is the joint probability density of the random variables $\mathbf{w}$ and $\mathbf{z}$, and the function $J(\tilde{\mathbf{w}})$ is the one that provides the risk evaluation criteria, many times referred to as the cost function. The problem of determining an estimate $\hat{\mathbf{w}}$ gets reduced to that of minimizing the risk, or expected cost value, by means of an appropriate choice of the functional $J(\tilde{\mathbf{w}})$. We refer to the value of $\hat{\mathbf{w}}$ providing the minimum as the *optimal estimate.*

In general, the optimal estimate depends on the cost function being employed. Example of two common cost functions are the quadratic cost function,

$$J = ||\tilde{\mathbf{w}}||^2_{\mathbf{E}} = \tilde{\mathbf{w}}^T \mathbf{E} \tilde{\mathbf{w}} \,, \tag{4.2}$$

where the $n \times n$ matrix $\mathbf{E}$ is assumed to be non-negative and symmetric; and the uniform cost function,

$$J = \begin{cases} 0\,, & ||\tilde{\mathbf{w}}|| < \epsilon \\ 1/2\epsilon\,, & ||\tilde{\mathbf{w}}|| \geq \epsilon \end{cases} \,. \tag{4.3}$$

However, for a large class of estimation problems, the resulting estimate is independent of the choice of the cost function.

A desirable property of an estimate is that it be *unbiased*, that is, that its ensemble average equals the ensemble average of the variable of interest. This is expressed mathematically as

$$\mathcal{E}\{\hat{\mathbf{w}}\} = \mathcal{E}\{\mathbf{w}\} \tag{4.4}$$

or in other words, the estimation error is zero: $\mathcal{E}\{\tilde{\mathbf{w}}\} = \mathbf{0}$. Estimates satisfying the equality above are said to be *unconditionally unbiased*, which is more general than being a *conditionally unbiased* estimate, that is obeying

$$\mathcal{E}\{\hat{\mathbf{w}}|\mathbf{w}\} = \mathbf{w} \,. \tag{4.5}$$

4.1.2 Minimum Variance Estimation

The minimum variance estimate, denoted $\hat{\mathbf{w}}_{\mathrm{MV}}$, minimizes the risk function with the cost function given by (4.2). Therefore, the risk function to be minimized is written explicitly as

$$\mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\mathbf{w} - \hat{\mathbf{w}})^T \mathbf{E} (\mathbf{w} - \hat{\mathbf{w}})\, p_{\mathbf{wz}}(\mathbf{w}, \mathbf{z})\, d\mathbf{z}\, d\mathbf{w} \tag{4.6}$$

which, using the definition of conditional probability distribution (1.77), can also be written as

$$\mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}) = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} (\mathbf{w} - \hat{\mathbf{w}})^T \mathbf{E} (\mathbf{w} - \hat{\mathbf{w}})\, p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \right\} p_{\mathbf{z}}(\mathbf{z})\, d\mathbf{z} \,. \tag{4.7}$$

The outer integral does not involve $\hat{\mathbf{w}}$, and since the marginal probability density $p_{\mathbf{z}}(\mathbf{z})$ is always positive, we see that to search for the minimum of $\mathcal{J}_{\mathrm{MV}}$ is equivalent to minimizing the integral in the kernel of the expression above. The kernel can be identified as an expression for the conditional Bayes risk, that is,

$$\mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}|\mathbf{z}) \equiv \int_{-\infty}^{\infty} (\mathbf{w} - \hat{\mathbf{w}})^T \mathbf{E} (\mathbf{w} - \hat{\mathbf{w}})\, p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \tag{4.8}$$

which is what we want to minimize with respect to $\hat{\mathbf{w}}$.

Using the definition of differentiation of a *scalar* function $f = f(\mathbf{x})$ of an n-dimensional *vector* $\mathbf{x}$, that is,

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \equiv \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix} \tag{4.9}$$

we can show that for a constant n–vector $\mathbf{a}$ we have

$$\frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}\,. \tag{4.10}$$

Moreover, for an $n \times n$ symmetric matrix $\mathbf{A}$ we have

$$\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}\,. \tag{4.11}$$

Applying these rules of differentiation to the minimization of $\mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}|\mathbf{z})$ it follows that

$$\mathbf{0} = \left.\frac{\partial \mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}|\mathbf{z})}{\partial \hat{\mathbf{w}}}\right|_{\hat{\mathbf{w}}=\hat{\mathbf{w}}_{\mathrm{MV}}} = -\ 2\,\mathbf{E}\left.\int_{-\infty}^{\infty} (\mathbf{w} - \hat{\mathbf{w}})\, p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w}\right|_{\hat{\mathbf{w}}=\hat{\mathbf{w}}_{\mathrm{MV}}} \tag{4.12}$$

and for any $\mathbf{E}$,

$$\hat{\mathbf{w}}_{\mathrm{MV}} \int_{-\infty}^{\infty} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} = \int_{-\infty}^{\infty} \mathbf{w} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \tag{4.13}$$

since the integral of $p_{\mathbf{w}|\mathbf{z}}$ is unity (because p is a probability density), hence

$$\begin{aligned} \hat{\mathbf{w}}_{\mathrm{MV}}(\mathbf{z}) &= \int_{-\infty}^{\infty} \mathbf{w} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \\ &= \mathcal{E}\{\mathbf{w}|\mathbf{z}\} \end{aligned} \tag{4.14}$$

This estimate has the desirable property of being unbiased. This can be shown simply as

$$\begin{aligned} \mathcal{E}\{\tilde{\mathbf{w}}\} &= \mathcal{E}\{\mathbf{w} - \hat{\mathbf{w}}_{\mathrm{MV}}\} \\ &= \mathcal{E}\{\mathbf{w} - \mathcal{E}\{\mathbf{w}|\mathbf{z}\}\} \\ &= \mathcal{E}\{\mathbf{w}\} - \mathcal{E}\{\mathcal{E}\{\mathbf{w}|\mathbf{z}\}\} \\ &= \mathcal{E}\{\mathbf{w}\} - \mathcal{E}\{\mathbf{w}\} \\ &= \mathbf{0} \end{aligned} \tag{4.15}$$

where the fourth equality follows from the chain rule for expectation operators in (1.84).

That the solution (4.14) is in fact a minimum of $\mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}|\mathbf{z})$ can be seen by calculating the second derivative of this quantity with respect to $\hat{\mathbf{w}}$, that is,

$$\frac{\partial^2 \mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}|\mathbf{z})}{\partial \hat{\mathbf{w}}^2} = 2\,\mathbf{E} \tag{4.16}$$

and since $\mathbf{E}$ is a non-negative matrix, the second derivative is non-negative, therefore the solution represents a minimum. Notice the extremely important fact that the estimate with minimum error variance (4.14) corresponds to the conditional mean. Substitution of (4.14) in expression (4.6) provides the Bayes risk with minimum error variance.

4.1.3 Maximum *a posteriori* Probability Estimation

Another estimator is defined through the risk function for the uniform cost function (4.3), and can be written explicitly as

$$\mathcal{J}_U(\hat{\mathbf{w}}) \;=\; \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} J(\tilde{\mathbf{w}})\, p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \right\} p_{\mathbf{z}}(\mathbf{z})\, d\mathbf{z}$$

$$= \int_{-\infty}^{\infty} \left\{ \frac{1}{2\epsilon} \int_{-\infty}^{\hat{\mathbf{w}}-\epsilon} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} + \frac{1}{2\epsilon} \int_{\hat{\mathbf{w}}+\epsilon}^{\infty} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \right\} p_{\mathbf{z}}(\mathbf{z})\, d\mathbf{z} \tag{4.17}$$

where, some caution is needed in reading the integrals inside the brackets: these are multiple integrals and the notation $\hat{\mathbf{w}} \pm \epsilon$ should be interpreted as $\hat{w}_1 \pm \epsilon$, $\hat{w}_2 \pm \epsilon$, and so on, for each one of the n components of the vector $\hat{\mathbf{w}}$. Since $p_{\mathbf{w}|\mathbf{z}}$ is a probability density function its integral over the whole R^n domain is unity, consequently the Bayes risk function can be written as

$$\mathcal{J}_U(\hat{\mathbf{w}}) = \int_{-\infty}^{\infty} \frac{1}{2\epsilon} \left\{ 1 - \int_{\hat{\mathbf{w}}-\epsilon}^{\hat{\mathbf{w}}+\epsilon} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \right\} p_{\mathbf{z}}(\mathbf{z})\, d\mathbf{z}\,. \tag{4.18}$$

For the problem of minimizing $\mathcal{J}_U$ with respect to $\hat{\mathbf{w}}$, the first term gives no relevant contribution, thus we can think of minimizing

$$\mathcal{J}_U(\hat{\mathbf{w}}) \sim -(1/2\epsilon) \int_{-\infty}^{\infty} \left\{ \int_{\hat{\mathbf{w}}-\epsilon}^{\hat{\mathbf{w}}+\epsilon} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \right\} p_{\mathbf{z}}(\mathbf{z})\, d\mathbf{z}\,. \tag{4.19}$$

or yet, we can minimize the conditional Bayes risk

$$\mathcal{J}_U(\hat{\mathbf{w}}|\mathbf{z}) \equiv -(1/2\epsilon) \int_{\hat{\mathbf{w}}-\epsilon}^{\hat{\mathbf{w}}+\epsilon} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \tag{4.20}$$

since $p_{\mathbf{z}}(\mathbf{z})$ is positive. As $\epsilon \geq 0$ approaches 0, the mean value theorem for integrals[1] can be employed to produce

$$\mathcal{J}_U(\hat{\mathbf{w}}|\mathbf{z}) = -\, p_{\mathbf{w}|\mathbf{z}}(\hat{\mathbf{w}}|\mathbf{z}) \tag{4.21}$$

which can also be obtained by noticing that as ϵ approaches zero the cost function $J(\tilde{\mathbf{w}})$ turns into a common representation for the negative of the delta function, in an n–dimensional space, that is, the cost function becomes

$$J(\tilde{\mathbf{w}}) \to -\prod_{i=1}^{n} \delta(w_i - \hat{w}_i)\,. \tag{4.22}$$

Minimization of $\mathcal{J}_U(\hat{\mathbf{w}}|\mathbf{z})$ is equivalent to maximization of the conditional probability density function $p_{\mathbf{w}|\mathbf{z}}(\hat{\mathbf{w}}|\mathbf{z})$. The value $\hat{\mathbf{w}} = \hat{\mathbf{w}}_{\mathrm{MAP}}$ that maximizes this quantity is known as the maximum *a posteriori* probability (MAP) estimate, and is determined by means of

$$\left. \frac{\partial p_{\mathbf{w}|\mathbf{z}}(\hat{\mathbf{w}}|\mathbf{z})}{\partial \hat{\mathbf{w}}} \right|_{\hat{\mathbf{w}}=\hat{\mathbf{w}}_{\mathrm{MAP}}} = \mathbf{0}\,, \tag{4.23}$$

which is the same as

$$\left. \frac{\partial p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})}{\partial \mathbf{w}} \right|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{MAP}}} = \mathbf{0}\,, \tag{4.24}$$

[1] The mean value theorem for integrals (e.g., Butkov [22]) can be stated as:

$$(1/2\epsilon) \int_{-\epsilon}^{\epsilon} f(x)\, dx = (1/2\epsilon)(2\epsilon) f(\xi) = f(\xi)$$

for $-\epsilon \leq \xi \leq \epsilon$.

since that the variables $\mathbf{w}$ and $\hat{\mathbf{w}}$ play the role of "dummy" derivation variables. Knowing that $p_{\mathbf{w}|\mathbf{z}}$ is really a function of $\mathbf{w}$, we prefer to use (4.24) rather than (4.23) to avoid confusion. The designation *a posteriori* refers to the fact that the estimate is obtained *after* the observations have been collected, that is, probability of $\mathbf{w}$ *given* $\mathbf{z}$. An estimate of this type is briefly described in (1.29), consequently we can identify maximum *a posteriori* probability estimation with *mode* estimation.

To maximize the probability density above is also equivalent to maximize its natural logarithm, $\ln p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})$, with respect to $\mathbf{w}$. Using Bayes rule (1.79) we can write

$$\ln p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z}) = \ln[p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w})] - \ln p_{\mathbf{z}}(\mathbf{z}) \tag{4.25}$$

and since $p_{\mathbf{z}}(\mathbf{z})$ does not depend on $\mathbf{w}$ the maximum *a posteriori* probability estimate can be obtained by solving either

$$\left.\frac{\partial \ln[p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w})]}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{MAP}}} = \mathbf{0}\,, \tag{4.26}$$

or

$$\left.\frac{\partial p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{MAP}}} = \mathbf{0}\,. \tag{4.27}$$

In general, the unbiasedness of the estimate is not necessarily guaranteed in this case.

4.1.4 Maximum Likelihood Estimation

In maximum *a posteriori* probability estimation it is necessary to know the probability density of the process of interest, that is $p_{\mathbf{w}}(\mathbf{w})$. In maximum likelihood (ML) estimation, we assume this *a priori* information is unknown. Assuming for the moment that the *a priori* probability distribution is Gaussian, with mean $\boldsymbol{\mu}_{\mathbf{w}}$ and covariance $\mathbf{P}_{\mathbf{w}}$, we have

$$p_{\mathbf{w}}(\mathbf{w}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}_{\mathbf{w}}|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{w}})^T\mathbf{P}_{\mathbf{w}}^{-1}(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{w}})\right] \tag{4.28}$$

or yet

$$\ln p_{\mathbf{w}}(\mathbf{w}) = -\ln[(2\pi)^{n/2}|\mathbf{P}_{\mathbf{w}}|^{1/2}] - \frac{1}{2}\left[(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{w}})^T\mathbf{P}_{\mathbf{w}}^{-1}(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{w}})\right]\,. \tag{4.29}$$

Hence,

$$\frac{\partial \ln p_{\mathbf{w}}(\mathbf{w})}{\partial \mathbf{w}} = -\mathbf{P}_{\mathbf{w}}^{-1}(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{w}}) \tag{4.30}$$

which indicates that lack of information about the random variable $\mathbf{w}$ implies infinite variance, $\mathbf{P}_{\mathbf{w}} \to \infty$, or yet $\mathbf{P}_{\mathbf{w}}^{-1} \to \mathbf{0}$. Thus, without a priori knowledge on $\mathbf{w}$ we have

$$\frac{\partial \ln p_{\mathbf{w}}(\mathbf{w})}{\partial \mathbf{w}} = \mathbf{0}\,. \tag{4.31}$$

This is also assumed to be the case even when the probability distribution of $\mathbf{w}$ is not Gaussian.

From (4.24) and (4.25), the maximum likelihood estimate of $\mathbf{w}$ can be obtained by

$$\mathbf{0} = \left[\frac{\partial \ln p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})}{\partial \mathbf{w}} + \frac{\partial \ln p_{\mathbf{w}}(\mathbf{w})}{\partial \mathbf{w}}\right]\Bigg|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{MAP}}} = \frac{\partial \ln p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})}{\partial \mathbf{w}}\Bigg|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{ML}}}, \tag{4.32}$$

or equivalently,

$$\frac{\partial p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})}{\partial \mathbf{w}}\Bigg|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{ML}}} = \mathbf{0}\,. \tag{4.33}$$

The estimate $\hat{\mathbf{w}}_{\mathrm{ML}}$ is sometimes referred to as the most likely estimate. However, because of the assumptions used in obtaining (4.33), this estimate is only reliable under certain conditions (see Jazwinski [84], p. 157). Just as in the case of the MAP estimate, the ML estimate is also a mode estimation, in analogy to (1.29). When we choose to refer to mode estimation, we should always make explicit which conditional probability is being maximized to avoid confusion, this defines whether we are performing MAP or ML estimation. As in MAP estimation, the estimate from ML is not guaranteed to be unbiased.

4.2 Example: Estimation of a Constant Vector

In this section we exemplify the problem of estimation by treating the case of estimating a constant (time independent) vector $\mathbf{w}$ by means of an observational process corrupted by noise, represented by the vector $\mathbf{v}$. We assume that $\mathbf{w}$ and $\mathbf{v}$ are independent and Gaussian distributed: $\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{P})$, and $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. Moreover, the observational process is taken to be a linear transformation

$$\mathbf{z} = \mathbf{H}\mathbf{w} + \mathbf{v} \tag{4.34}$$

where $\mathbf{w}$ is an n–vector, $\mathbf{z}$ and $\mathbf{v}$ are m–vectors, and $\mathbf{H}$ is an $m \times n$ matrix, referred to as the observation matrix which accounts, for example, for linear combinations among elements of the vector $\mathbf{w}$. To obtain an estimate based on the methods described in the previous section, we investigate the probability densities of the random variables involved in the observational process.

For the minimum variance estimate we need to determined the *a posteriori* probability density $p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})$, so that we can solve the integral in (4.14). From Bayes rule we have

$$p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z}) = \frac{p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w})}{p_{\mathbf{z}}(\mathbf{z})} \tag{4.35}$$

and consequently we need to determine each one of the probability densities in this expression.

Since $\mathbf{w}$ is Gaussian, we can readily write

$$p_{\mathbf{w}}(\mathbf{w}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{w}-\boldsymbol{\mu})\right]. \tag{4.36}$$

Linear transformations of Gaussian distributed variables result in Gaussian distributed variables (e.g., Sage & Melsa [121], pp. 71–72; see also Exercise 4, here). Therefore, the probability distribution for the observations is given by

$$p_{\mathbf{z}}(\mathbf{z}) = \frac{1}{(2\pi)^{m/2}|\mathbf{P}_{\mathbf{z}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})^T\mathbf{P}_{\mathbf{z}}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})\right] \tag{4.37}$$

where $\boldsymbol{\mu}_{\mathbf{z}}$ and $\mathbf{P}_{\mathbf{z}}$ correspond to the mean and covariance of the random variable $\mathbf{z}$, respectively. These quantities can be determined by applying the ensemble average operator to (4.34), and using the definition of covariance. Thus,

$$\boldsymbol{\mu}_{\mathbf{z}} = \mathcal{E}\{\mathbf{Hw}\} + \mathcal{E}\{\mathbf{v}\} = \mathbf{H}\boldsymbol{\mu} \tag{4.38}$$

and also,

$$\begin{aligned}
\mathbf{P}_{\mathbf{z}} &= \mathcal{E}\{(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})^T\} \\
&= \mathcal{E}\{[(\mathbf{Hw}+\mathbf{v})-\mathbf{H}\boldsymbol{\mu})][(\mathbf{Hw}+\mathbf{v})-\mathbf{H}\boldsymbol{\mu})]^T\} \\
&= \mathcal{E}\{[(\mathbf{Hw}-\mathbf{H}\boldsymbol{\mu})-\mathbf{v}][(\mathbf{Hw}-\mathbf{H}\boldsymbol{\mu})-\mathbf{v}]^T\} \\
&= \mathbf{H}\mathcal{E}\{(\mathbf{w}-\boldsymbol{\mu})(\mathbf{w}-\boldsymbol{\mu})^T\}\mathbf{H}^T + \mathcal{E}\{\mathbf{v}\mathbf{v}^T\} \\
&\quad + \mathbf{H}\mathcal{E}\{(\mathbf{w}-\boldsymbol{\mu})\mathbf{v}^T\} + \mathcal{E}\{\mathbf{v}(\mathbf{w}-\boldsymbol{\mu})^T\}\mathbf{H}^T .
\end{aligned} \tag{4.39}$$

Noticing that $\mathbf{w}$ and $\mathbf{v}$ are independent $\mathcal{E}\{\mathbf{w}\mathbf{v}^T\} = \mathbf{0}$, and that $\mathbf{v}$ has zero mean, it follows that

$$\mathbf{P}_{\mathbf{z}} = \mathbf{HPH}^T + \mathbf{R} \tag{4.40}$$

and consequently, the probability distribution of $\mathbf{z}$ becomes

$$\begin{aligned}
p_{\mathbf{z}}(\mathbf{z}) &= \frac{1}{(2\pi)^{m/2}|(\mathbf{HPH}^T + \mathbf{R})|^{1/2}} \\
&\quad \times \exp\left[-\frac{1}{2}(\mathbf{z}-\mathbf{H}\boldsymbol{\mu})^T(\mathbf{HPH}^T + \mathbf{R})^{-1}(\mathbf{z}-\mathbf{H}\boldsymbol{\mu})\right] .
\end{aligned} \tag{4.41}$$

It remains for us to determine the conditional probability density $p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})$ explicitly. This distribution is also Gaussian (e.g., Sage & Melsa [121] pp. 73–74), and can be written as

$$p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w}) = \frac{1}{(2\pi)^{m/2}|\mathbf{P}_{\mathbf{z}|\mathbf{w}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}})^T\mathbf{P}_{\mathbf{z}|\mathbf{w}}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}})\right] \tag{4.42}$$

Analogously to what we have just done to determine $p_{\mathbf{z}}(\mathbf{z})$, we have

$$\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}} = \mathcal{E}\{\mathbf{Hw}|\mathbf{w}\} + \mathcal{E}\{\mathbf{v}|\mathbf{w}\} = \mathbf{Hw} \tag{4.43}$$

and

$$\begin{aligned}
\mathbf{P}_{\mathbf{z}|\mathbf{w}} &= \mathcal{E}\{(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}})(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}})^T|\mathbf{w}\} \\
&= \mathcal{E}\{[(\mathbf{Hw}+\mathbf{v})-\mathbf{Hw})][(\mathbf{Hw}+\mathbf{v})-\mathbf{Hw})]^T|\mathbf{w}\} \\
&= \mathcal{E}\{\mathbf{v}\mathbf{v}^T|\mathbf{w}\} \\
&= \mathcal{E}\{\mathbf{v}\mathbf{v}^T\} \\
&= \mathbf{R} .
\end{aligned} \tag{4.44}$$

Therefore,

$$p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w}) = \frac{1}{(2\pi)^{m/2}|\mathbf{R}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{z}-\mathbf{Hw})^T\mathbf{R}^{-1}(\mathbf{z}-\mathbf{Hw})\right] \tag{4.45}$$

which is the conditional probability of $\mathbf{z}$ given $\mathbf{w}$.

Combining the results (4.36), (4.41), and (4.45) in Bayes rule (4.35) it follows that the *a posteriori* probability distribution we are interested in takes the form

$$p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z}) = \frac{|\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}|^{1/2}}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}|\mathbf{R}|^{1/2}} \exp[-\frac{1}{2}J] \tag{4.46}$$

where J is defined as,

$$\begin{aligned} J(\mathbf{w}) \;\equiv\; & (\mathbf{z} - \mathbf{H}\mathbf{w})^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{w}) + (\mathbf{w} - \boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{w} - \boldsymbol{\mu}) \\ & - (\mathbf{z} - \mathbf{H}\boldsymbol{\mu})^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{z} - \mathbf{H}\boldsymbol{\mu}) \end{aligned} \tag{4.47}$$

This quantity J can also be written in the following more compact form:

$$J(\mathbf{w}) = (\mathbf{w} - \hat{\mathbf{w}})^T\mathbf{P}_{\tilde{\mathbf{w}}}^{-1}(\mathbf{w} - \hat{\mathbf{w}}) \tag{4.48}$$

where $\mathbf{P}_{\tilde{\mathbf{w}}}^{-1}$ is given by

$$\mathbf{P}_{\tilde{\mathbf{w}}}^{-1} = \mathbf{P}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\,, \tag{4.49}$$

the vector $\hat{\mathbf{w}}$ is given by

$$\hat{\mathbf{w}} = \mathbf{P}_{\tilde{\mathbf{w}}}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} + \mathbf{P}^{-1}\boldsymbol{\mu}) \tag{4.50}$$

and the reason for using the subscript $\tilde{\mathbf{w}}$ for the matrix $\mathbf{P}_{\tilde{\mathbf{w}}}$, indicating a relationship with the estimation error, will soon become clear.

According to (4.14), the minimum variance estimate is given by the conditional mean of the *a posteriori* probability density, that is,

$$\hat{\mathbf{w}}_{\text{MV}} = \int_{-\infty}^{\infty} \mathbf{w}\,p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\,d\mathbf{w} = \hat{\mathbf{w}} \tag{4.51}$$

where the integration can be performed using the approach of moments calculation of the Gaussian distribution (e.g., Maybeck [101]; see also Exercise 3, here).

The maximum *a posteriori* probability estimate (4.24) is the one that maximizes $p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})$ in (4.46), and is easily identified to be

$$\hat{\mathbf{w}}_{\text{MAP}} = \hat{\mathbf{w}}\,. \tag{4.52}$$

Thus we see that the minimum variance estimate coincides with the maximum *a posteriori* probability density estimate.

Let us now return to the reason for using the subscript $\tilde{\mathbf{w}}$ in $\mathbf{P}_{\tilde{\mathbf{w}}}$. For that, remember that we defined the estimation error $\tilde{\mathbf{w}}$ as the difference between the estimate and the actual value taken by the variable of interest, that is,

$$\tilde{\mathbf{w}} \equiv \hat{\mathbf{w}} - \mathbf{w}\,. \tag{4.53}$$

We want to show that $\mathbf{P}_{\tilde{\mathbf{w}}}$ is indeed the estimate *error covariance* matrix. To verify this, let us show first that $\boldsymbol{\mu}_{\tilde{\mathbf{w}}} = \mathbf{0}$, that is, that the ensemble mean error estimate is zero for the minimum variance and MAP estimates. In other words, we want to show that these estimates are *unbiased*. Using (4.50) we have

$$\begin{aligned}\boldsymbol{\mu}_{\tilde{\mathbf{w}}} &= \mathcal{E}\{(\hat{\mathbf{w}} - \mathbf{w})\} \\ &= \mathbf{P}_{\tilde{\mathbf{w}}}(\mathbf{H}^T\mathbf{R}^{-1}\mathcal{E}\{\mathbf{z}\} + \mathbf{P}^{-1}\boldsymbol{\mu}) - \boldsymbol{\mu} \\ &= \mathbf{P}_{\tilde{\mathbf{w}}}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})\boldsymbol{\mu} - \boldsymbol{\mu} \qquad (4.54)\end{aligned}$$

where we replaced $\mathbf{z}$ from (4.34), and we recall that $\mathbf{v}$ has zero mean. Therefore, using the definition of $\mathbf{P}_{\tilde{\mathbf{w}}}$ in (4.49), it follows that $\boldsymbol{\mu}_{\tilde{\mathbf{w}}} = \mathbf{0}$. Given what we know from (4.15), this result comes as no surprise in the case of the minimum variance estimate (4.51); in case of the MAP estimate this proves that (4.52) does provide an unbiased estimate.

To show that $\mathbf{P}_{\tilde{\mathbf{w}}}$ is the error covariance matrix of the estimate, we observe that $\tilde{\mathbf{w}}$ can be decomposed as

$$\begin{aligned}\mathbf{w} - \hat{\mathbf{w}} &= \mathbf{w} - \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{w} - \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{v} - \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{P}^{-1}\boldsymbol{\mu} \\ &= \mathbf{w} - \mathbf{P}_{\tilde{\mathbf{w}}}(\mathbf{P}_{\tilde{\mathbf{w}}}^{-1} - \mathbf{P}^{-1})\mathbf{w} - \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{v} - \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{P}^{-1}\boldsymbol{\mu} \\ &= \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{P}^{-1}(\mathbf{w} - \boldsymbol{\mu}) - \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{v}\,. \qquad (4.55)\end{aligned}$$

Therefore,

$$\begin{aligned}var\{\tilde{\mathbf{w}}\} = cov\{\tilde{\mathbf{w}}, \tilde{\mathbf{w}}\} &= \mathcal{E}\{(\hat{\mathbf{w}} - \mathbf{w})(\hat{\mathbf{w}} - \mathbf{w})^T\} \\ &= \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{P}^{-1}\mathcal{E}\{(\mathbf{w} - \boldsymbol{\mu})(\mathbf{w} - \boldsymbol{\mu})^T\}\mathbf{P}^{-1}\mathbf{P}_{\tilde{\mathbf{w}}} \\ &\quad + \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{H}^T\mathbf{R}^{-1}\mathcal{E}\{\mathbf{v}\mathbf{v}^T\}\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_{\tilde{\mathbf{w}}} \qquad (4.56)\end{aligned}$$

where the cross-terms give no contribution since $\mathbf{w}$ and $\mathbf{v}$ are independent, and because $\mathbf{v}$ has zero mean. Using the definition of $\mathbf{P}$ it follows that

$$\begin{aligned}var\{\tilde{\mathbf{w}}\} &= \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{P}^{-1}\mathbf{P}_{\tilde{\mathbf{w}}} + \mathbf{P}_{\tilde{\mathbf{w}}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_{\tilde{\mathbf{w}}} \\ &= \mathbf{P}_{\tilde{\mathbf{w}}}(\mathbf{P}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{P}_{\tilde{\mathbf{w}}} \\ &= \mathbf{P}_{\tilde{\mathbf{w}}} \qquad (4.57)\end{aligned}$$

where (4.49) was used. This shows that $\mathbf{P}_{\tilde{\mathbf{w}}}$ defined in (4.49) is indeed the estimation error covariance matrix, thus justifying its subscript $\tilde{\mathbf{w}}$. Moreover, it is simple to see that

$$|\mathbf{P}_{\tilde{\mathbf{w}}}| = |\mathbf{H}\mathbf{P}^{-1}\mathbf{H}^T + \mathbf{R}||\mathbf{P}||\mathbf{R}| \qquad (4.58)$$

and therefore (4.46) can be written as

$$p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}_{\tilde{\mathbf{w}}}|^{1/2}}\exp[-\frac{1}{2}(\mathbf{w} - \hat{\mathbf{w}})^T\mathbf{P}_{\tilde{\mathbf{w}}}^{-1}(\mathbf{w} - \hat{\mathbf{w}})] \qquad (4.59)$$

justifying the rewriting of J from (4.47) to (4.48).

It is now left for us to determine the maximum likelihood estimate (4.33). This can be done by maximizing the probability density $p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})$ in (4.45). Hence,

$$\begin{aligned}\mathbf{0} &= \left.\frac{\partial p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{ML}}} \\ &= \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\hat{\mathbf{w}}_{\mathrm{ML}}) \qquad (4.60)\end{aligned}$$

that is,

$$\hat{\mathbf{w}}_{\text{ML}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} \tag{4.61}$$

which is, in principle, distinct from the estimates obtained above, following the minimum variance and maximum *a posteriori* probability estimation approaches. Remembering now that in maximum likelihood estimation we assume lack of statistical information regarding the process $\mathbf{w}$, and observing that this means $\mathbf{P}^{-1} = \mathbf{0}$, we see from (4.50) and (4.49) that, in this case,

$$\hat{\mathbf{w}}_{\text{MV}}|_{\mathbf{P}^{-1}=\mathbf{0}} = \hat{\mathbf{w}}_{\text{MAP}}|_{\mathbf{P}^{-1}=\mathbf{0}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} = \hat{\mathbf{w}}_{\text{ML}} \tag{4.62}$$

and therefore all three estimation approaches produce the same result.

Applying the average operator to (4.61) we have

$$\begin{aligned}
\mathcal{E}\{\hat{\mathbf{w}}_{\text{ML}}\} &= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathcal{E}\{\mathbf{z}\} \\
&= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathcal{E}\{\mathbf{w}\} + \mathcal{E}\{\mathbf{v}\}) \\
&= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathcal{E}\{\mathbf{w}\} \\
&= \mathcal{E}\{\mathbf{w}\}
\end{aligned} \tag{4.63}$$

where we used the fact that $\mathbf{v}$ has mean zero. This shows that the ML estimate is also unbiased.

It is simple to show that the maximum likelihood estimate error covariance is given by

$$var\{\tilde{\mathbf{w}}_{\text{ML}}\} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{4.64}$$

which is always greater than the error covariance obtained with the minimum variance estimation approach. This makes sense since the minimum variance estimate is that corresponding to the *minimum* of the Bayes risk.

Notice that all estimates above result in a *linear combination* of the observations. Moreover, although in this example all three estimation procedures studied above provide the same estimate this is not always the case. An example in which these estimates do not coincide is given in Exercise 2.

Another remark can be made by noticing that in the maximum *a posteriori* probability estimation context the denominator in (4.35) is not relevant for the maximization of the *a posteriori* probability distribution, as indicated in equations (4.26) and (4.27). This implies that we can derive the result for in (4.52) by minimizing the part of the functional J in (4.47) corresponding only to the probability density functions in the numerator of (4.35). That is, we can define the functional corresponding to these probability densities as

$$J_{\text{MAP}}(\mathbf{w}) \equiv (\mathbf{z} - \mathbf{H}\mathbf{w})^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{w}) + (\mathbf{w} - \boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{w} - \boldsymbol{\mu}) \tag{4.65}$$

and its minimization can be shown to produces the same result as in (4.50) with error variance as in (4.49) — see Exercise 3. Analogously, we can define a cost function related to the *a priori* probability distribution associated with the maximum likelihood estimate, that is,

$$J_{\text{ML}}(\mathbf{w}) \boxminus (\mathbf{z} - \mathbf{H}\mathbf{w})^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{w})\,. \tag{4.66}$$

The minimization of J_{ML} gives the estimate in (4.61) with error variance (4.64).

4.3 Least Squares Estimation

All of the estimation methods seen so far, i.e., minimum variance, maximum *a posteriori* probability, and maximum likelihood, require statistical knowledge of part or all the random variables in question. However, when going from minimum variance and MAP to ML we relaxed the statistical assumptions by considering we knew nothing about the statistics of the variable(s) of interest ($\mathbf{w}$, in that case). Relaxing even further the statistical assumptions for the estimation problem takes us in to the situation where we have no statistical information about any of the variables involved in problem. In this extreme case, estimation reduces to the method of finding the least squares fit among the observations.

Let us consider again, as an example, the observational process in the previous section for an n–vector constant $\mathbf{w}$. Let us assume further that several observations are taken about the variable of interest, and that the i–th observation can be written as

$$\mathbf{z}_i = \mathbf{H}_i\mathbf{w} + \mathbf{v}_i \tag{4.67}$$

where $\mathbf{z}_i$, $\mathbf{H}_i$ and $\mathbf{v}_i$ represent an m_i–observation vector, a linear transformation matrix $m_i \times n$ and a m_i–noise vector, respectively. It is important to recognize now that we are assuming we do not know the statistics of the noise $\mathbf{v}_i$, and also that due to lack of statistical information we are not interpreting $\mathbf{w}$ as a random vector.

By collecting the result of k experiments in a long vector, we can write the expression above in the following compact form:

$$\tilde{\mathbf{z}}_k = \tilde{\mathbf{H}}_k\mathbf{w} + \tilde{\mathbf{v}}_k \tag{4.68}$$

where the $\tilde{m}_k$–vector $\tilde{\mathbf{z}}_k$ is defined as:

$$\tilde{\mathbf{z}}_k \equiv [\mathbf{z}_1^T\,\mathbf{z}_2^T\,\cdots\,\mathbf{z}_k^T]^T \tag{4.69}$$

for $\tilde{m}_k = \sum_{i=1}^k m_i$, and where

$$\tilde{\mathbf{v}}_k \equiv [\mathbf{v}_1^T\,\mathbf{v}_2^T\,\cdots\,\mathbf{v}_k^T]^T \tag{4.70}$$

and the matrix $\tilde{\mathbf{H}}_k$, of dimension $\tilde{m}_k \times n$, is defined as

$$\tilde{\mathbf{H}}_k \equiv [\mathbf{H}_1^T\,\mathbf{H}_2^T\,\cdots\,\mathbf{H}_k^T]^T\,. \tag{4.71}$$

The problem we want to consider is that of finding an estimate $\hat{\mathbf{w}}_k$ which minimizes the quadratic function $\mathcal{J}$,

$$\mathcal{J}(\hat{\mathbf{w}}_k) = \frac{1}{2}(\tilde{\mathbf{z}}_k - \tilde{\mathbf{H}}_k\hat{\mathbf{w}}_k)^T\tilde{\mathbf{O}}_k^{-1}(\tilde{\mathbf{z}}_k - \tilde{\mathbf{H}}_k\hat{\mathbf{w}}_k) \tag{4.72}$$

which measures the distance between the observations and the estimate. The value that minimizes this function is called the least squares estimate and is denoted by $\hat{\mathbf{w}}_k^{\mathrm{LS}}$. The positive definite and symmetric matrix $\tilde{\mathbf{O}}_k^{-1}$ represents weights attributed to each experiment, and convey a certain degree of confidence regarding the experiment in question.

The estimator function $\mathcal{J}$ is deterministic, therefore the problem of minimizing $\mathcal{J}$ is a common optimization problem, where the solution $\hat{\mathbf{w}}_k^{\mathrm{LS}}$ can be determined by means solving,

$$\left.\frac{\partial\mathcal{J}}{\partial\hat{\mathbf{w}}_k}\right|_{\hat{\mathbf{w}}_k=\hat{\mathbf{w}}_k^{\mathrm{LS}}} = \mathbf{0}\,. \tag{4.73}$$

Then, the differentiation of (4.72) yields

$$\tilde{\mathbf{H}}_k^T\tilde{\mathbf{O}}_k^{-1}(\mathbf{z}_k - \tilde{\mathbf{H}}_k\hat{\mathbf{w}}_k^{\text{LS}}) = \mathbf{0} \tag{4.74}$$

from where it follows that

$$\hat{\mathbf{w}}_k^{\text{LS}} = \mathbf{P}_k\tilde{\mathbf{H}}_k^T\tilde{\mathbf{O}}_k^{-1}\mathbf{z}_k\,, \tag{4.75}$$

which is the estimate for the value of $\mathbf{w}$. For convenience we define a matrix $\mathbf{P}_k$ of dimension $n \times n$ as

$$\mathbf{P}_k \equiv (\tilde{\mathbf{H}}_k^T\tilde{\mathbf{O}}_k^{-1}\tilde{\mathbf{H}}_k)^{-1}\,, \tag{4.76}$$

and assume that the inverse exists. The matrix $\mathbf{P}_k^{-1}$ is sometimes referred to as the Gram matrix. A comparison with the estimate provided by the ML (4.61) method shows certain resemblance, however, since $\mathbf{R}$ and $\mathbf{O}_k$ are not related in any way, this resemblance is purely formal.

Suppose now that an additional experiment was made and it produced a new observation $\mathbf{z}_{k+1}$:

$$\mathbf{z}_{k+1} = \mathbf{H}_{k+1}\mathbf{w} + \mathbf{v}_{k+1}. \tag{4.77}$$

Then, by means of the notation introduced above, we can write

$$\tilde{\mathbf{z}}_{k+1} = \tilde{\mathbf{H}}_{k+1}\mathbf{w} + \tilde{\mathbf{v}}_{k+1}\,, \tag{4.78}$$

where

$$\tilde{\mathbf{z}}_{k+1} = [\tilde{\mathbf{z}}_k^T\,\mathbf{z}_{k+1}^T]^T, \quad \tilde{\mathbf{H}}_{k+1} = [\tilde{\mathbf{H}}_k^T\,\mathbf{H}_{k+1}^T]^T, \quad \tilde{\mathbf{v}}_{k+1} = [\tilde{\mathbf{v}}_k^T\,\mathbf{v}_{k+1}^T]^T. \tag{4.79}$$

Direct use of the minimization procedure just described leads to an estimate including the new observation $\mathbf{z}_{k+1}$, and given by

$$\hat{\mathbf{w}}_{k+1}^{\text{LS}} = \mathbf{P}_{k+1}\tilde{\mathbf{H}}_{k+1}^T\tilde{\mathbf{O}}_{k+1}^{-1}\tilde{\mathbf{z}}_{k+1}\,, \tag{4.80}$$

where $\mathbf{P}_{k+1}$ is defined, in analogy to $\mathbf{P}_k$, as

$$\mathbf{P}_{k+1} \equiv (\tilde{\mathbf{H}}_{k+1}^T\tilde{\mathbf{O}}_{k+1}^{-1}\tilde{\mathbf{H}}_{k+1})^{-1}\,, \tag{4.81}$$

and $\tilde{\mathbf{O}}_{k+1}^{-1}$ is a new weight matrix that takes into account the observation $\mathbf{z}_{k+1}$.

The processing of an extra observation forces us to have to solve the minimization problem completely again. In particular, we have to calculate the inverse of an $n \times n$ matrix for each new observation made. This computational burden can be avoided if we assume that the matrix $\tilde{\mathbf{O}}_{k+1}^{-1}$ can be partitioned in the following manner:

$$\tilde{\mathbf{O}}_{k+1}^{-1} = \begin{bmatrix} \tilde{\mathbf{O}}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{O}_{k+1}^{-1} \end{bmatrix}. \tag{4.82}$$

that is, $\tilde{\mathbf{O}}_{k+1}^{-1}$ is assumed to be a block–diagonal matrix.

With this assumption, we can write the product of the matrices in $\mathbf{P}_{k+1}$ as

$$\begin{aligned} \tilde{\mathbf{H}}_{k+1}^T\tilde{\mathbf{O}}_{k+1}^{-1}\tilde{\mathbf{H}}_{k+1} &= [\tilde{\mathbf{H}}_k^T\,\mathbf{H}_{k+1}^T]\begin{bmatrix} \tilde{\mathbf{O}}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{O}_{k+1}^{-1} \end{bmatrix}\begin{bmatrix} \tilde{\mathbf{H}}_k \\ \mathbf{H}_{k+1} \end{bmatrix} \\ &= \tilde{\mathbf{H}}_k^T\tilde{\mathbf{O}}_k^{-1}\tilde{\mathbf{H}}_k + \mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1}\mathbf{H}_{k+1}\,. \end{aligned} \tag{4.83}$$

Furthermore, using the definitions of the matrices $\mathbf{P}$ given above, we have that

$$\mathbf{P}_{k+1}^{-1} = \mathbf{P}_k^{-1} + \mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1}\mathbf{H}_{k+1} \tag{4.84}$$

or yet, using the Sherman–Morrison–Woodbury formula (e.g., Golub & Van Loan [67], p. 51).

$$\begin{aligned}\mathbf{P}_{k+1} &= (\mathbf{P}_k^{-1} + \mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1}\mathbf{H}_{k+1})^{-1} \\ &= \mathbf{P}_k - \mathbf{P}_k\mathbf{H}_{k+1}^T(\mathbf{H}_{k+1}\mathbf{P}_k\mathbf{H}_{k+1}^T + \mathbf{O}_{k+1})^{-1}\mathbf{H}_{k+1}\mathbf{P}_k\,.\end{aligned} \tag{4.85}$$

Defining a matrix $\mathbf{G}_{k+1}$ as

$$\mathbf{G}_{k+1} \equiv \mathbf{P}_k\mathbf{H}_{k+1}^T(\mathbf{H}_{k+1}\mathbf{P}_k\mathbf{H}_{k+1}^T + \mathbf{O}_{k+1})^{-1}\,, \tag{4.86}$$

we can compactly write

$$\mathbf{P}_{k+1} = (\mathbf{I} - \mathbf{G}_{k+1}\mathbf{H}_{k+1})\mathbf{P}_k\,. \tag{4.87}$$

Therefore the estimate $\hat{\mathbf{w}}_{k+1}^{\mathrm{LS}}$, which includes the new observation can be re–written as

$$\hat{\mathbf{w}}_{k+1}^{\mathrm{LS}} = (\mathbf{I} - \mathbf{G}_{k+1}\mathbf{H}_{k+1})\mathbf{P}_k\,\tilde{\mathbf{H}}_{k+1}^T\tilde{\mathbf{O}}_{k+1}^{-1}\tilde{\mathbf{z}}_{k+1}\,. \tag{4.88}$$

Using the matrix partition for $\tilde{\mathbf{O}}_{k+1}^{-1}$, introduced above, we can decompose the expression for the estimate in two terms,

$$\tilde{\mathbf{H}}_{k+1}^T\tilde{\mathbf{O}}_{k+1}^{-1}\tilde{\mathbf{z}}_{k+1} = \tilde{\mathbf{H}}_k^T\tilde{\mathbf{O}}_k^{-1}\tilde{\mathbf{z}}_k + \mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1}\mathbf{z}_{k+1}\,, \tag{4.89}$$

and consequently (4.88) is transformed in

$$\begin{aligned}\hat{\mathbf{w}}_{k+1}^{\mathrm{LS}} &= [\mathbf{I} - \mathbf{G}_{k+1}\mathbf{H}_{k+1}]\mathbf{P}_k(\tilde{\mathbf{H}}_k^T\tilde{\mathbf{O}}_k^{-1}\tilde{\mathbf{z}}_k + \mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1}\mathbf{z}_{k+1})\,, \\ &= [\mathbf{I} - \mathbf{G}_{k+1}\mathbf{H}_{k+1}]\hat{\mathbf{w}}_k^{\mathrm{LS}} + [\mathbf{I} - \mathbf{G}_{k+1}\mathbf{H}_{k+1}]\mathbf{P}_k\mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1}\mathbf{z}_{k+1}\end{aligned} \tag{4.90}$$

where we used (4.75) to obtain the second equality.

A even better expression for the estimate can be derived if we use the definition for the matrix $\mathbf{G}_{k+1}$. In this case, the coefficient of the last term in the previous expression can be re–written as

$$\begin{aligned}[\mathbf{I} - \mathbf{G}_{k+1}\mathbf{H}_{k+1}]\mathbf{P}_k\mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1} &= [\mathbf{I} - \mathbf{G}_{k+1}\mathbf{H}_{k+1}] \\ &\quad\times \mathbf{G}_{k+1}(\mathbf{H}_{k+1}\mathbf{P}_k\mathbf{H}_{k+1}^T + \mathbf{O}_{k+1})\mathbf{O}_{k+1}^{-1} \\ &= [\mathbf{I} - \mathbf{G}_{k+1}\mathbf{H}_{k+1}]\mathbf{G}_{k+1} \\ &\quad\times (\mathbf{I} + \mathbf{H}_{k+1}\mathbf{P}_k\mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1}) \\ &= \mathbf{G}_{k+1}[\mathbf{I} + \mathbf{H}_{k+1}\mathbf{P}_k\mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1} \\ &\quad - \mathbf{H}_{k+1}\mathbf{G}_{k+1}(\mathbf{I} + \mathbf{H}_{k+1}\mathbf{P}_k\mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1})] \\ &= \mathbf{G}_{k+1}[\mathbf{I} + \mathbf{H}_{k+1}\mathbf{P}_k\mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1} \\ &\quad - \mathbf{H}_{k+1}\mathbf{P}_k\mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1}] \\ &= \mathbf{G}_{k+1}\,.\end{aligned} \tag{4.91}$$

Thus, the estimate can be placed finally in the form

$$\hat{\mathbf{w}}_{k+1}^{\text{LS}} = \hat{\mathbf{w}}_k^{\text{LS}} + \mathbf{G}_{k+1}(\mathbf{z}_{k+1} - \mathbf{H}_{k+1}\hat{\mathbf{w}}_k^{\text{LS}}), \tag{4.92}$$

where $\mathbf{G}_{k+1}$ is given by (4.86). This expression provides a *recursive* manner of updating the estimate, given a new observation of the variable of interest and the estimate obtained before the new observation had been made. This recursive expression requires inverting an $m_{k+1} \times m_{k+1}$ matrix embedded in the definition of $\mathbf{G}_{k+1}$ in (4.86), rather than the $n \times n$ matrix (4.81), for each new observation becoming available. This represents an enormous computational savings especially for $n \gg m_k$, for all k.

4.4 Relationship between Least Squares and Minimum Variance

The estimates produced by the minimum variance and least squares methods are of fundamental importance in many studies in estimation theory. Consequently, in this section, we explore the relationship between these two estimates.

To simplify this notation let us omit the index k from the previous section, so that the observational process can be written just as in (4.34),

$$\mathbf{z} = \mathbf{H}\mathbf{w} + \mathbf{v}, \tag{4.93}$$

Moreover, the estimate of $\mathbf{w}$ provided by the least squares method is written as

$$\hat{\mathbf{w}}_{\text{LS}} = \mathbf{M}\mathbf{z}, \tag{4.94}$$

where for convenience we define the $n \times m$ matrix $\mathbf{M}$ as

$$\mathbf{M} = (\mathbf{H}^T\mathbf{O}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{O}^{-1}. \tag{4.95}$$

Notice that $\mathbf{MH} = \mathbf{I}$ which, assuming the noise $\mathbf{v}$ has zero mean is a way of expressing the fact that the estimate $\hat{\mathbf{w}}_{\text{LS}}$ is unbiased. To see this, we define the error associated to the least squares estimate as

$$\tilde{\mathbf{w}}_{\text{LS}} \equiv \mathbf{w} - \hat{\mathbf{w}}_{\text{LS}}, \tag{4.96}$$

where once again we use a *tilde* to indicate an error vector. Application the ensemble average operator, and using (4.93) and (4.94), it follows that

$$\begin{aligned} \mathcal{E}\{\tilde{\mathbf{w}}_{\text{LS}}\} &= \mathcal{E}\{[\mathbf{w} - \mathbf{M}(\mathbf{H}\mathbf{w} + \mathbf{v})]\} \\ &= -\mathbf{M}\mathcal{E}\{\mathbf{v}\} \\ &= \mathbf{0}, \end{aligned} \tag{4.97}$$

which justifies the assertion above that the least squares estimate is unbiased.

The least squares estimate error variance can be calculated according to

$$\mathbf{P}_{\tilde{\mathbf{w}}_{\text{LS}}} = \mathcal{E}\{\tilde{\mathbf{w}}_{\text{LS}}\tilde{\mathbf{w}}_{\text{LS}}^T\} = \mathbf{M}\mathcal{E}\{\mathbf{v}\mathbf{v}^T\}\mathbf{M}^T = \mathbf{M}\mathbf{R}\mathbf{M}^T \tag{4.98}$$

where $\mathbf{R}$ is the (co)variance matrix of the noise $\mathbf{v}$, as defined in Section 4.3. Substituting the value of M as defined above we have

$$\mathbf{P}_{\tilde{\mathbf{w}}_{\mathrm{LS}}} = (\mathbf{H}^T\mathbf{O}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{O}^{-1}\mathbf{R}\mathbf{O}^{-1}\mathbf{H}(\mathbf{H}^T\mathbf{O}^{-1}\mathbf{H})^{-1}\,. \tag{4.99}$$

Now remember that, by the procedure of Section 4.3, the linear estimate of minimum variance, with zero mean $\boldsymbol{\mu}_{\mathbf{w}} = \mathbf{0}$ and for which $\mathbf{P}_{\mathbf{w}}^{-1} = \mathbf{0}$, is given by

$$\tilde{\mathbf{w}}_{\mathrm{MV}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}\,. \tag{4.100}$$

which is the same as that obtained when using the approach of maximum likelihood estimation. As we know, this estimate is also unbiased, and with associated error (co)variance

$$\mathbf{P}_{\tilde{\mathbf{w}}_{\mathrm{MV}}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\,, \tag{4.101}$$

as it can be seen in (4.50) and (4.51), and also (4.61) and (4.64), respectively. Therefore, we notice by comparison that the estimate obtained by the least squares method is the same as the one obtained by linear minimum variance when the matrix of weight $\mathbf{O}$ used by the first method is substituted by the noise (co)variance matrix, that is, $\mathbf{O} = \mathbf{R}$.

In general, the weight matrix used in the least squares method is a general positive definite and symmetric matrix, without any statistical meaning; since the estimate provided by the minimum variance approach is that with *minimum variance*, for the linear case, it follows that in general

$$\mathbf{P}_{\tilde{\mathbf{w}}_{\mathrm{LS}}} \geq \mathbf{P}_{\tilde{\mathbf{w}}_{\mathrm{MV}}}\,, \tag{4.102}$$

where the equality holds when $\mathbf{O} = \mathbf{R}$. This inequality is valid even if we do not use the fact that the estimate $\hat{\mathbf{w}}_{\mathrm{MV}}$ is that of minimum variance. To derive this inequality, we can use the following matrix inequality

$$\mathbf{A}^T\mathbf{A} \geq (\mathbf{B}^T\mathbf{A})^T(\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{B}^T\mathbf{A})\,, \tag{4.103}$$

for $\mathbf{A}$ and $\mathbf{B}$, of dimensions $n \times m$, with $n \geq m$, and $\mathbf{B}$ of full rank. This derivation is left as an exercise.

Exercises

1. (Sage & Melsa [121], Problem 6.1) Another example of cost function, aside from those given in the main text, is that defined by the absolute value of the error: $J(\tilde{w}) = |\tilde{w}| = |w - \hat{w}|$, considering the scalar case. Show that in this case, the estimate $\hat{w}_{\mathrm{ABS}}$ that minimizes the Bayes risk is the one for which we have:

$$\int_{-\infty}^{\hat{w}_{\mathrm{ABS}}} p_{w|z}(w|z)\,dw = \int_{\hat{w}_{\mathrm{ABS}}}^{\infty} p_{w|z}(w|z)\,dw$$

and that consequently, the estimate with minimum absolute value can be determined by solving:

$$\int_{\hat{w}_{\mathrm{ABS}}}^{\infty} p_{w|z}(w|z)\,dw = \frac{1}{2}$$

for $\hat{w} = \hat{w}_{\text{ABS}}$. In other words, the estimate with minimum absolute value $\hat{w}_{\text{ABS}}$ is the median, as introduced in (1.27). Derive the corresponding modification of the result above for the vector case, if we define the cost function to be

$$J(\tilde{\mathbf{w}}) = \sum_i |\tilde{w}_i|$$

2. Consider the observational process of a binary variable (binary signal), subject to noise (measurement errors). This scalar observation process can be written as

$$z = w + v$$

where w and v are independent, and v is a gaussian noise, represented by $\mathcal{N}(0, \sigma_v^2)$. The signal w follows the binary distribution defined as

$$p_w(w) = 0.5\delta(w-1) + 0.5\delta(w+1)$$

where δ represents the Dirac delta. Then,

(a) Determine the *a priori* probability density $p_{z|w}(z|w)$.

(b) Show that the probability density $p_z(z)$ is given by[2]

$$p_z(z) = \frac{1}{2\sqrt{2\pi}\sigma_v}\left\{\exp\left[-\frac{(z-1)^2}{2\sigma_v^2}\right] + \exp\left[-\frac{(z+1)^2}{2\sigma_v^2}\right]\right\}$$

(c) Show that the maximum *a posteriori* probability estimate is $\hat{w}_{\text{MAP}} = \text{sign}(z)$.

(d) Show that the minimum variance error estimate is $\hat{w}_{\text{MV}} = \tanh\left(\frac{z}{\sigma_v^2}\right)$.

In the minimum variance estimation case, what happens when the observations become more accurate?

3. Show that the solution of the minimization of J_{MAP} in (4.65) is given by (4.50) with error estimate (4.49).

4. Writing a few terms for the traces in the expressions below, verify that:

(a) $\frac{d[\text{Tr}(\mathbf{AB})]}{d\mathbf{A}} = \mathbf{B}^T$, where $\mathbf{AB}$ is symmetric

(b) $\frac{d[\text{Tr}(\mathbf{ACA}^T)]}{d\mathbf{A}} = 2\mathbf{AC}$, where $\mathbf{C}$ is also symmetric

Notice that is x is a scalar, we define its derivative with respect to a matrix $\mathbf{A}$ according to:

$$\frac{dx}{d\mathbf{A}} \equiv \begin{pmatrix} \frac{dx}{da_{11}} & \frac{dx}{da_{12}} & \cdots \\ \frac{dx}{da_{21}} & \frac{dx}{da_{22}} & \cdots \\ \cdot & & \cdot \\ \cdot & & \cdot \; \cdot \\ \cdot & & \cdot \quad \cdot \end{pmatrix}$$

where a_{ij} is the (i,j)-th element of matrix $\mathbf{A}$.

[2] If a random variable z is defined as the summation of two independent random variables w and v the probability of z can be obtained via the convolution integral:

$$p_z(z) = \int_{-\infty}^{\infty} p_w(z-v)p_v(v)\,dv$$

5. Show that

$$\mathbf{G}_{k+1} = \mathbf{P}_{k+1}\mathbf{H}_{k+1}^T\mathbf{O}_{k+1}^{-1}$$

is an alternative expression for the gain matrix $\mathbf{G}_{k+1}$ found in the least squares estimation method.

6. Let $\mathbf{A}$ and $\mathbf{B}$ be to $n \times m$ matrices, with $n \geq m$, and with $\mathbf{B}$ full rank (m). Show that

$$\mathbf{A}^T\mathbf{A} \geq (\mathbf{B}^T\mathbf{A})^T\,(\mathbf{B}^T\mathbf{B})^{-1}\,(\mathbf{B}^T\mathbf{A})\,.$$

(Hint: Use the following inequality:

$$(\mathbf{Ax} + \mathbf{By})^T(\mathbf{Ax} + \mathbf{By}) \geq 0$$

valid for any two m–vectors $\mathbf{x}$ e $\mathbf{y}$.) Now, to show the inequality in (4.102), without making use of the fact that $\hat{\mathbf{w}}_{MV}$ is a minimum variance estimate for the linear case, make the following choice:

$$\mathbf{A} = \mathbf{R}^{1/2}\mathbf{M}^T\,, \quad \mathbf{B} = \mathbf{R}^{-1/2}\mathbf{H}$$

and complete the proof as suggested in the end of section 4.5.

Chapter 5

The Linear Kalman Filter

In this lecture we derive and study the Kalman filter and its properties for the case of time–discrete dynamics and time–discrete observations. The case of time–continuous dynamics with time–continuous observations is mentioned without many details, and the case of time–continuous dynamics with time–discrete observations is not considered is this course. The content of this lecture can be found in classic books of stochastic processes and estimation theory, such as, Anderson & Moore [1], Gelb [60], Jazwinski [84], Meditch [103], and Sage & Melsa [121]. In this lecture, we also introduce a convenient notation to treat the assimilation problem of meteorological and oceanographic data, to be discussed in lectures that follow.

5.1 Derivation of the Linear Kalman Filter

5.1.1 Estimation Problem in Linear Systems

We derive the Kalman filter using the estimation approach of minimum variance, following the derivation of Todling & Cohn [129], which deals with the problem of atmospheric data assimilation to be studied later.

Consider a time–discrete, linear stochastic dynamical system written in matrix–vector notation as

$$\mathbf{w}_k^t = \mathbf{\Psi}_{k-1}\mathbf{w}_{k-1}^t + \mathbf{b}_{k-1}^t \,, \tag{5.1}$$

for the discrete times t_k, with $k = 1, 2, \ldots$, and where $\mathbf{w}_k^t$ is an n–vector representing the *true* state of the system at time t_k, $\mathbf{\Psi}_k$ is an $n \times n$ matrix that represents the dynamics, and the n–vector $\mathbf{b}_k^t$ is an additive random noise, which we refer to as the model error. The process $\mathbf{b}_k^t$ is assumed to be white in time, with mean zero and (co)variance $\mathbf{Q}_k$, that is,

$$\mathcal{E}\{\mathbf{b}_k^t\} = \mathbf{0} \,, \quad \mathcal{E}\{\mathbf{b}_k^t(\mathbf{b}_{k'}^t)^T\} = \mathbf{Q}_k\delta_{kk'} \,. \tag{5.2}$$

Consider also a linear observation process described by

$$\mathbf{w}_k^o = \mathbf{H}_k\mathbf{w}_k^t + \mathbf{b}_k^o \,, \tag{5.3}$$

where k now is a multiple of ℓ, the number of time steps of between two consecutive observations in time. The m_k-vector $\mathbf{w}_k^o$ is the vector of observations, the matrix $m_k \times n$ represents a linear transformation between the true variables into the observed ones, and the m_k-vector $\mathbf{b}_k^o$ is an additive noise, representing error in the observational process, as for example, error due to instrument accuracy. We assume that the random noise $\mathbf{v}_k$ is white in time, with mean zero and (co)variance $\mathbf{R}_k$, that is,

$$\mathcal{E}\{\mathbf{b}_k^o\} = \mathbf{0}\,, \;\; \mathcal{E}\{\mathbf{b}_k^o)(\mathbf{b}_{k'}^o)^T\} = \mathbf{R}_k\delta_{kk'}, \tag{5.4}$$

We also assume that the observation noise $\mathbf{v}_k$ and the model error are uncorrelated, that is,

$$\mathcal{E}\{\mathbf{b}_k^t(\mathbf{b}_{k'}^o)^T\} = \mathbf{0}\,. \tag{5.5}$$

The problem treated in the previous lecture was that of estimating $\mathbf{w}_k^t$ given the observation process (5.3) alone. In this lecture, we add to the estimation problem the constraint that the variable of interest comes from the linear stochastic dynamical system (5.1). However, since the dynamical system in (5.1) involves the stochastic noise $\mathbf{b}_k^t$ and an unknown initial state, we replace that model by what we refer to as a forecast model that we write as

$$\mathbf{w}_k^f = \mathbf{\Psi}_{k,k-\ell}\mathbf{w}_{k-\ell}^a\,, \tag{5.6}$$

where the symbol f stands for forecast and the symbol a stands for the "initial" condition at time $t_{k-\ell}$, from which we start a forecast, and referred to as the analysis. The forecast model represents another way we have of estimating the state of the system at a particular time. The matrix $\mathbf{\Psi}_{k,k-\ell}$ is the propagator, or transition matrix, between times $t_{k-\ell}$ and t_k, and is given by

$$\mathbf{\Psi}_{k,k-\ell} \equiv \mathbf{\Psi}_{k-1}\mathbf{\Psi}_{k-2}\cdots\mathbf{\Psi}_{k-\ell}\,, \tag{5.7}$$

where here we make a distinction between the propagator and the one–time step dynamics through the double subscripts to indicate the propagator.

An estimate of the state of the system at time t_k can be obtained by means of a linear combination between the observation at time t_k and the forecast at the same time. Therefore, we can write for the estimate $\mathbf{w}_k^a$ at time t_k,

$$\mathbf{w}_k^a = \tilde{\mathbf{L}}_k\mathbf{w}_k^f + \tilde{\mathbf{K}}_k\mathbf{w}_k^o\,, \tag{5.8}$$

where $\tilde{\mathbf{L}}_k$ and $\tilde{\mathbf{K}}_k$ are weighting matrices still to be determined.

Let us define the forecast and (estimate) analysis errors as

$$\mathbf{e}_k^f \equiv \mathbf{w}_k^f - \mathbf{w}_k^t\,, \tag{5.9a}$$
$$\mathbf{e}_k^a \equiv \mathbf{w}_k^a - \mathbf{w}_k^t\,. \tag{5.9b}$$

In analogy to what we saw in Lecture 4, we would like to have an estimate that is unbiased. In this way, subtracting $\mathbf{w}_k^t$ from both sides of (5.8), as well as from $\mathbf{w}_k^f$ in that expression, and using (5.3) it follows that

$$\mathbf{e}_k^a = \tilde{\mathbf{L}}_k\mathbf{e}_k^f + \tilde{\mathbf{K}}_k\mathbf{b}_k^o + (\tilde{\mathbf{L}}_k + \tilde{\mathbf{K}}_k\mathbf{H}_k - \mathbf{I})\mathbf{w}_k^t \tag{5.10}$$

Now assuming that the forecast error, at time t_k, is unbiased, that is, $\mathcal{E}\{\mathbf{e}_k^f\} = \mathbf{0}$, we should satisfy

$$(\tilde{\mathbf{L}}_k + \tilde{\mathbf{K}}_k \mathbf{H}_k - \mathbf{I})\mathcal{E}\{\mathbf{w}_k^t\} = \mathbf{0} \tag{5.11}$$

to obtain an unbiased estimate (analysis), i.e., $\mathcal{E}\{\mathbf{e}_k^a\} = \mathbf{0}$. As in general $\mathcal{E}\{\mathbf{w}_k^t\} \neq \mathbf{0}$, we have that

$$\tilde{\mathbf{L}}_k = \mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k \tag{5.12}$$

is the condition for an unbiased $\mathbf{w}_k^a$.

Substituting result (5.12) in (5.8) we can write for the estimate of the state of the system

$$\mathbf{w}_k^a = \mathbf{w}_k^f + \tilde{\mathbf{K}}_k(\mathbf{w}_k^o - \mathbf{H}_k \mathbf{w}_k^f), \tag{5.13}$$

and for the estimate error

$$\mathbf{e}_k^a = \left(\mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k\right) \mathbf{e}_k^f + \tilde{\mathbf{K}}_k \mathbf{b}_k^o. \tag{5.14}$$

The weight matrix $\tilde{\mathbf{K}}_k$, or gain matrix as it is more commonly known, represents the weights given to the difference between the observation vector and the forecast transformed by the observation matrix $\mathbf{H}_k$. We have seen in Lecture 4, that different procedures come up with a formula for the estimate that resembles (5.13), however they use distinct gain matrices, e.g., recall the comparison between minimum variance estimation and least squares estimation.

Using (5.1) and (5.6) it follows that

$$\mathbf{e}_k^f = \mathbf{\Psi}_{k,k-l}\mathbf{e}_{k-l}^a - \sum_{j=0}^{l-1} \mathbf{\Psi}_{k,k-j}\mathbf{b}_{k-j-1}^t, \tag{5.15}$$

which is an equation for the evolution of forecast error.

Introducing the forecast and analysis error covariance matrices

$$\mathbf{P}_k^f \equiv \mathcal{E}\{(\mathbf{w}_k^t - \mathbf{w}_k^f)(\mathbf{w}_k^t - \mathbf{w}_k^f)^T\} \tag{5.16a}$$

$$\mathbf{P}_k^a \equiv \mathcal{E}\{(\mathbf{w}_k^t - \mathbf{w}_k^a)(\mathbf{w}_k^t - \mathbf{w}_k^a)^T\}, \tag{5.16b}$$

we can proceed as in Section 3.2.2 to obtain an expression for the evolution of the forecast error covariance:

$$\mathbf{P}_k^f = \mathbf{\Psi}_{k,k-l}\mathbf{P}_{k-l}^a\mathbf{\Psi}_{k,k-l}^T + \sum_{j=0}^{l-1} \mathbf{\Psi}_{k,k-j}\mathbf{Q}_{k-j-1}\mathbf{\Psi}_{k,k-j}^T, \tag{5.17}$$

which is a form equivalent (iterated) to the discrete Lyapunov equation (3.40).

An expression for the (estimated) analysis error covariance $\mathbf{P}_k^a$ can be determined by multiplying (5.14) by its transpose and applying the ensemble average operator to the resulting expression. Therefore, we have

$$\mathbf{P}_k^a = (\mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k)\mathbf{P}_k^f(\mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k)^T + \tilde{\mathbf{K}}_k \mathbf{R}_k \tilde{\mathbf{K}}_k^T, \tag{5.18}$$

which is referred to as Joseph's formula. Equations (5.17) and (5.18) completely describe the evolution of errors in the forecast and analysis. An interesting property of the equations for the error (co)variances is that they are independent of the estimates (analysis and forecast), and also from the observations. The only necessary quantities to predict the evolution of the error (co)variances are the noise (co)variance matrices, $\mathbf{Q}_k$ and $\mathbf{R}_k$, the initial error (co)variance matrix $\mathbf{P}_0^a$, and matrices $\mathbf{H}_k$ and $\tilde{\mathbf{K}}_k$, at each time t_k. In principle, all these matrices are known, except for the gain matrix $\tilde{\mathbf{K}}_k$ which is to be determined by means of an optimization procedure that requires minimum error variance.

5.1.2 The Kalman Filter

To treat the problem stated above in the lights of minimum variance estimation we introduce an estimator that serves as a measure of reliability of the analysis. That is, a quantity measuring the distance between the estimate and the true value of the state of the system at time t_k,

$$\begin{aligned}\mathcal{J}_k^a &\equiv \mathcal{E}\{||\mathbf{e}_k^a||^2_{\mathbf{E}_k}\} \\ &= \mathcal{E}\left\{(\mathbf{e}_k^a)^T\mathbf{E}_k\,\mathbf{e}_k^a\right\} \\ &= \mathcal{E}\left\{\mathrm{Tr}\left[\mathbf{E}_k\,\mathbf{e}_k^a(\mathbf{e}_k^a)^T\right]\right\} \\ &= \mathrm{Tr}\,(\mathbf{E}_k\mathbf{P}_k^a)\ . \end{aligned} \tag{5.19}$$

As in Lecture 4, we want this measure of error to be minimum with respect to the elements of the gain matrix $\tilde{\mathbf{K}}_k$. The matrix $n \times n$ matrix $\mathbf{E}_k$ introduced in the functional above is a scaling matrix, which we assume to be positive definite and deterministic, which in many cases can be substituted by the identity matrix. As we will see below, the solution of the minimization $\mathcal{J}_k^a$ is in fact independent of $\mathbf{E}_k$.

Substituting the expression (5.18) for $\mathbf{P}_k^a$ in (5.19), differentiating with respect to $\tilde{\mathbf{K}}_k$, (using the differentiation rules of Exercise 4.4), and equating the result to zero we obtain

$$\mathbf{E}_k\left\{\mathbf{H}_k\mathbf{P}_k^f(\mathbf{I}-\tilde{\mathbf{K}}_k\mathbf{H}_k)^T\ -\ \mathbf{R}_k\tilde{\mathbf{K}}_k^T\right\} = \mathbf{0} \tag{5.20}$$

Therefore, independently of $\mathbf{E}_k$, the quantity between curly brackets becomes zero for

$$\tilde{\mathbf{K}}_k = \mathbf{K}_k \equiv \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1}\ , \tag{5.21}$$

which corresponds to the minimum of $\mathcal{J}_k^a$. The matrix $\mathbf{K}_k$ is the optimal weighting matrix, known as the Kalman gain matrix, since this estimation problem was solved by Kalman [87]. Although estimation problems date back from the times of Gauss [57], it was Kalman who solved the problem in the dynamical systems context, using the state–space approach. As a matter of fact, Kalman derived the result obtained above in a much more elegant way based on the orthogonal projections theorem. The solution obtained by Kalman has practical consequences that go much beyond previous results in estimation theory. Kalman & Bucy [90] extended the Kalman filter to the case of time–continuous dynamics and observation process. An excellent review of filtering theory can be found in Kailath [86], and the influence of Kalman's work in several theoretical and applied areas is collected in Antoulas [3].

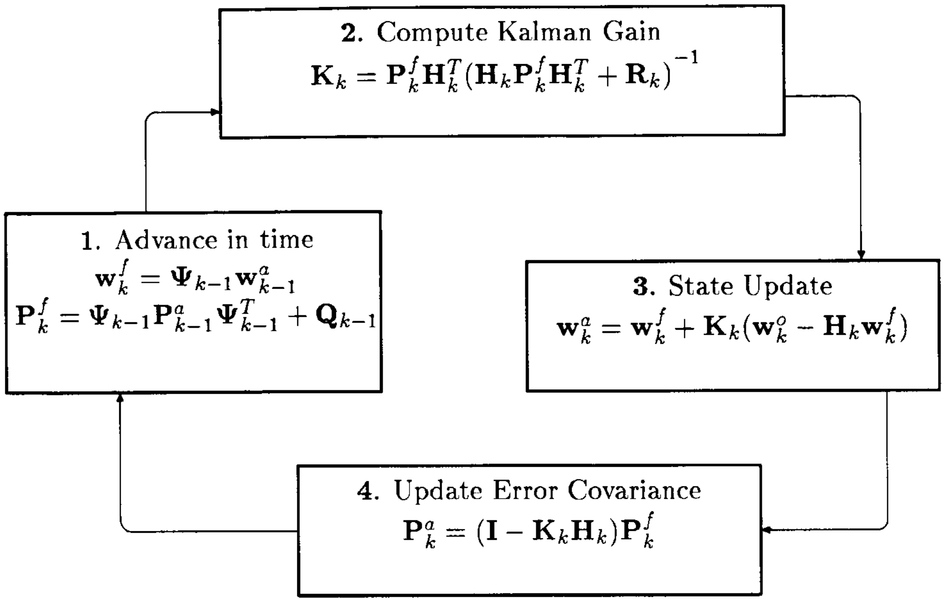
2. Compute Kalman Gain
$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1}$
1. Advance in time
$\mathbf{w}_k^f = \mathbf{\Psi}_{k-1}\mathbf{w}_{k-1}^a$
$\mathbf{P}_k^f = \mathbf{\Psi}_{k-1}\mathbf{P}_{k-1}^a\mathbf{\Psi}_{k-1}^T + \mathbf{Q}_{k-1}$
3. State Update
$\mathbf{w}_k^a = \mathbf{w}_k^f + \mathbf{K}_k(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f)$
4. Update Error Covariance
$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f$

Figure 5.1: Schematic diagram of the linear Kalman filter.

Substituting the Kalman gain matrix in the expression for the analysis error covariance (5.18), it is simple to show that this equation reduces to

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f\,, \tag{5.22}$$

which is a simpler expression. The optimal estimate of the state of the system at time t_k is given by (5.13) with a general gain matrix $\tilde{\mathbf{K}}_k$ replaced by its optimal value, that is,

$$\mathbf{w}_k^a = \mathbf{w}_k^f + \mathbf{K}_k(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f)\,. \tag{5.23}$$

Fig. 5.1 shows schematically the steps involved in the execution of the linear Kalman filter for the case $\ell = 1$, that is, when the observations are available at each time step. The case $\ell = 1$ will be considered from this point on to keep the notation simple.

5.1.3 Comments: Minimum Variance and Conditional Mean

We saw in the previous lecture that in Bayes estimation theory the estimate of minimum variance is given by the conditional mean. Let us now establish the connection between the derivation of the Kalman filter given above and the example discussed in the previous lecture, of estimation of a constant vector. What we will see is that, that the example corresponds to the analysis step of the Kalman filter.

Let us indicate by $\mathbf{W}_k^o = \{\mathbf{w}_k^o, \mathbf{w}_{k-1}^o, \cdots, \mathbf{w}_1^o\}$, the set of all observations (5.3) up and including time t_k. Similarly to the case in the previous lecture, the problem of estimating the state of the system at time t_k, based on the observations $\mathbf{W}_k^o$ can be placed as the problem of determining the conditional probability density $p(\mathbf{w}_k^t|\mathbf{W}_k^o)$, where, to simplify notation we omit the subscript in p referring to the stochastic process in question. By the result of Section 4.2.1, we know that

$$\begin{aligned}
\mathbf{w}_k^a &= \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_k^o\} \\
&= \int_{-\infty}^{+\infty} \mathbf{w}_k^t p(\mathbf{w}_k^t|\mathbf{W}_k^o)\, d\mathbf{w}_k^t
\end{aligned} \tag{5.24}$$

and therefore, knowledge of $p(\mathbf{w}_k^t|\mathbf{W}_k^o)$ is fundamental to determine the estimate.

In fact, using repeatedly the definition of conditional probability density we can write

$$\begin{aligned}
p(\mathbf{w}_k^t|\mathbf{W}_k^o) &= p(\mathbf{w}_k^t|\mathbf{w}_k^o, \mathbf{W}_{k-1}^o) \\
&= \frac{p(\mathbf{w}_k^t, \mathbf{w}_k^o, \mathbf{W}_{k-1}^o)}{p(\mathbf{w}_k^o, \mathbf{W}_{k-1}^o)} \\
&= \frac{p(\mathbf{w}_k^o|\mathbf{w}_k^t, \mathbf{W}_{k-1}^o)p(\mathbf{w}_k^t, \mathbf{W}_{k-1}^o)}{p(\mathbf{w}_k^o, \mathbf{W}_{k-1}^o)} \\
&= \frac{p(\mathbf{w}_k^o|\mathbf{w}_k^t, \mathbf{W}_{k-1}^o)p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o)p(\mathbf{W}_{k-1}^o)}{p(\mathbf{w}_k^o|\mathbf{W}_{k-1}^o)p(\mathbf{W}_{k-1}^o)} \\
&= \frac{p(\mathbf{w}_k^o|\mathbf{w}_k^t, \mathbf{W}_{k-1}^o)p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o)}{p(\mathbf{w}_k^o|\mathbf{W}_{k-1}^o)},
\end{aligned} \tag{5.25}$$

which related the transition probability of interest with transition probabilities that can be calculated more promptly.

Since the sequence of observational noise $\{\mathbf{b}_k^o\}$ is white, the following simplification applies:

$$p(\mathbf{w}_k^o|\mathbf{w}_k^t, \mathbf{W}_{k-1}^o) = p(\mathbf{w}_k^o|\mathbf{w}_k^t) \tag{5.26}$$

and therefore,

$$p(\mathbf{w}_k^t|\mathbf{W}_k^o) = \frac{p(\mathbf{w}_k^o|\mathbf{w}_k^t)p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o)}{p(\mathbf{w}_k^o|\mathbf{W}_{k-1}^o)} \tag{5.27}$$

It remains for us to determine each one of the transition probability densities in this expression.

Assuming the probability distributions of $\mathbf{w}_0^t$, $\mathbf{b}_k^t$ and $\mathbf{v}_k$ are Gaussian, we can draw a straight relationship among the variables here and those in Section 4.3. Specifically, we can identify $\mathbf{z}$ with $\mathbf{w}_k^o$ and $\mathbf{w}$ with $\mathbf{w}_k^t$, therefore, the probability densities $p_{\mathbf{z}}(\mathbf{z})$ and $p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})$ can be identified with the probability densities $p(\mathbf{w}_k^o)$ and $p(\mathbf{w}_k^o|\mathbf{w}_k^t)$, respectively. Consequently, we can write for $p(\mathbf{w}_k^o|\mathbf{w}_k^t)$,

$$p(\mathbf{w}_k^o|\mathbf{w}_k^t) = \frac{1}{(2\pi)^{m_k/2}|\mathbf{R}_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^t)^T\mathbf{R}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^t)\right] \tag{5.28}$$

where we notice that

$$\mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_k^t\} = \mathcal{E}\{(\mathbf{H}_k\mathbf{w}_k^t + \mathbf{b}_k^o)|\mathbf{w}_k^t\} = \mathbf{H}_k\mathbf{w}_k^t \tag{5.29}$$

and

$$\begin{aligned} cov\{\mathbf{w}_k^o, \mathbf{w}_k^o|\mathbf{w}_k^t\} &\equiv \mathcal{E}\{[\mathbf{w}_k^o - \mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_k^t\}][\mathbf{w}_k^o - \mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_k^t\}]^T|\mathbf{w}_k^t\} \\ &= \mathbf{R}_k . \end{aligned} \tag{5.30}$$

Analogously, we have

$$p(\mathbf{w}_k^o|\mathbf{W}_{k-1}^o) = \frac{1}{(2\pi)^{m_k/2}|\mathbf{\Gamma}_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f)^T\mathbf{\Gamma}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f)\right] \tag{5.31}$$

where we define $\mathbf{w}_k^f$ as

$$\mathbf{w}_k^f \equiv \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}, \tag{5.32}$$

the matrix $m_k \times m_k$ matrix $\mathbf{\Gamma}_k$ as

$$\mathbf{\Gamma}_k \equiv \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k , \tag{5.33}$$

and the $n \times n$ matrix $\mathbf{P}_k^f$ as

$$\begin{aligned} \mathbf{P}_k^f &\equiv \mathcal{E}\{[\mathbf{w}_k^t - \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}][\mathbf{w}_k^t - \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}]^T|\mathbf{W}_{k-1}^o\} \\ &= \mathcal{E}\{[\mathbf{w}_k^t - \mathbf{w}_k^f][\mathbf{w}_k^t - \mathbf{w}_k^f]^T|\mathbf{W}_{k-1}^o\} . \end{aligned} \tag{5.34}$$

To fully determine the *a posteriori* conditional probability density $p(\mathbf{w}_k^t|\mathbf{W}_k^o)$, it remains for us to find the *a priori* conditional probability density $p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o)$. Since we are assuming that $\mathbf{w}_0^t$ and $\mathbf{b}_k^t$ are Gaussian distributed, $p(\mathbf{w}_{k-1}^t|\mathbf{W}_{k-1}^o)$ is Gaussian, and it follows from the linearity of (5.1) that $p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o)$ is also Gaussian. Therefore, all that remains for us to determine are the mean $\mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}$ and the (co)variance $cov\{\mathbf{w}_k^t, \mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}$.

From the definition (5.32) of $\mathbf{w}_k^f$ we have

$$\begin{aligned} \mathbf{w}_k^f &= \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_{k-1}^o\} \\ &= \mathbf{\Psi}_{k-1}\mathcal{E}\{\mathbf{w}_{k-1}^t|\mathbf{W}_{k-1}^o\} + \mathcal{E}\{\mathbf{b}_{k-1}^t|\mathbf{W}_{k-1}^o\} \\ &= \mathbf{\Psi}_{k-1}\mathcal{E}\{\mathbf{w}_{k-1}^t|\mathbf{W}_{k-1}^o\} + \mathcal{E}\{\mathbf{b}_{k-1}^t\} \\ &= \mathbf{\Psi}_{k-1}\mathbf{w}_{k-1}^a \end{aligned} \tag{5.35}$$

where the last equality is obtained by observing that $\mathbf{b}_{k-1}^t$ has mean zero, and by using the definition of the estimate $\mathbf{w}_{k-1}^a$, as the conditional mean at time t_{k-1}. This expression represents the time evolution of the estimate, and it justifies the somewhat *ad hoc* forecast model that appeared in (5.6). The expression above is also identical to that found in (3.33) for the evolution of the mean.

The expression for the (co)variance matrix $cov\{\mathbf{w}_k^t, \mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}$ can be easily shown to be

$$\begin{aligned} cov\{\mathbf{w}_k^t, \mathbf{w}_k^t|\mathbf{W}_{k-1}^o\} &= \mathbf{\Psi}_{k-1}\mathbf{P}_{k-1}^a\mathbf{\Psi}_{k-1}^T + \mathbf{Q}_{k-1} \\ &= \mathbf{P}_k^f , \end{aligned} \tag{5.36}$$

where we recall that to simplify notation we are assuming that observations are available at all times, that is, the expression above corresponds to that in (5.17) with $\ell = 1$. Furthermore, (5.36) is identical to the time–discrete Lyapunov equation (3.40).

From the result (5.36) and the definition (5.32), we can write

$$p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}_k^f|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}_k^t - \mathbf{w}_k^f)^T(\mathbf{P}_k^f)^{-1}(\mathbf{w}_k^t - \mathbf{w}_k^f)\right] \tag{5.37}$$

so that, proceeding as in Section 4.3, the conditional probability density (5.27) of interest becomes

$$p(\mathbf{w}_k^t|\mathbf{W}_k^o) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}_k^a|^{1/2}} \exp\left[-\frac{1}{2}J_k^a\right] \tag{5.38}$$

where J_k^a is cost function defined as

$$J_k^a \equiv (\mathbf{e}_k^a)^T(\mathbf{P}_k^a)^{-1}\mathbf{e}_k^a \tag{5.39}$$

where $\mathbf{e}_k^a \equiv (\mathbf{w}_k^a - \mathbf{w}_k^t)$ as in (5.9). We can now identify the quantities $\hat{\mathbf{w}}_{\mathrm{MV}}$ and $\mathbf{P}_{\tilde{\mathbf{w}}}$ of Section 4.3 with $\mathbf{w}_k^a$ and $\mathbf{P}_k^a$, respectively. Consequently, it follows from this correspondence that

$$(\mathbf{P}_k^a)^{-1} = (\mathbf{P}_k^f)^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\,. \tag{5.40}$$

Since in Section 4.3 we showed that $\hat{\mathbf{w}}_{\mathrm{MV}}$ was the minimum variance estimate (4.51) for the problem dealt in that section, it follows immediately that $\mathbf{w}_k^a$ is the minimum variance estimate of the problem we are studying in this section.

To complete the correspondence between the treatment of this section and that of the previous section, we notice that the most remarkable difference between these two treatments is that the ensemble average operator of the previous section was the *unconditional* ensemble average. On the other hand, in this section, the ensemble average operators are the *conditional* ones, that is, conditioned on the observations. As a matter of fact, during the derivation performed in the previous section we advanced the result obtained in this section that the forecast and analysis error covariance matrices $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$ are in fact independent from the observations, see (5.36) and (5.40), that is,

$$\begin{aligned}
\mathbf{P}_k^f &= \mathcal{E}\{[\mathbf{w}_k^t - \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}][\mathbf{w}_k^t - \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}]^T|\mathbf{W}_{k-1}^o\} \\
&= \mathcal{E}\{[\mathbf{w}_k^t - \mathbf{w}_k^f][\mathbf{w}_k^t - \mathbf{w}_k^f]^T|\mathbf{W}_{k-1}^o\} \\
&= \mathcal{E}\{[\mathbf{w}_k^t - \mathbf{w}_k^f][\mathbf{w}_k^t - \mathbf{w}_k^f]^T\}\,,
\end{aligned} \tag{5.41}$$

and

$$\begin{aligned}
\mathbf{P}_k^a &= \mathcal{E}\{[\mathbf{w}_k^t - \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_k^o\}][\mathbf{w}_k^t - \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_k^o\}]^T|\mathbf{W}_k^o\} \\
&= \mathcal{E}\{[\mathbf{w}_k^t - \mathbf{w}_k^a][\mathbf{w}_k^t - \mathbf{w}_k^a]^T|\mathbf{W}_k^o\} \\
&= \mathcal{E}\{[\mathbf{w}_k^t - \mathbf{w}_k^a][\mathbf{w}_k^t - \mathbf{w}_k^a]^T\}\,.
\end{aligned} \tag{5.42}$$

Consequently we can replace the conditional error (co)variances by the unconditional error (co)variances.

Following some remarks in the previous chapter, we see that an equivalent cost function to that in (5.39), associated to the maximum *a posteriori* estimate, is

$$J_{\mathrm{3dVar}}(\mathbf{w}_k^t) \equiv (\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^t)^T\mathbf{R}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^t) + (\mathbf{w}_k^t - \mathbf{w}_k^f)^T(\mathbf{P}_k^f)^{-1}(\mathbf{w}_k^t - \mathbf{w}_k^f)\,. \tag{5.43}$$

This cost function can also be written in its 3–dimensional variational form (e.g., Courtier [36]), as

$$J_{\mathrm{3dVar}}(\delta\mathbf{w}_k) \equiv (\mathbf{v}_k - \mathbf{H}_k\delta\mathbf{w}_k)^T\mathbf{R}_k^{-1}(\mathbf{v}_k - \mathbf{H}_k\delta\mathbf{w}_k) + \delta\mathbf{w}_k^T(\mathbf{P}_k^f)^{-1}\delta\mathbf{w}_k \tag{5.44}$$

where $\delta\mathbf{w}_k \equiv \mathbf{w}_k^t - \mathbf{w}_k^f = -\mathbf{e}_k^f$, and we notice that

$$\begin{aligned} \mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^t &= \mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f + \mathbf{H}_k\mathbf{w}_k^f - \mathbf{H}_k\mathbf{w}_k^t \\ &= \mathbf{v}_k - \mathbf{H}_k\delta\mathbf{w}_k \end{aligned} \tag{5.45}$$

where $\mathbf{v}_k$ is the innovation vector, $\mathbf{v}_k \equiv \mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f$. And from the same discussion presented before, the minimization of (5.44) produces to the same solution as that found from the minimum variance approach.

5.2 Properties of the Kalman Filter

5.2.1 Whiteness of the Innovation Process

The behavior, or more adequately the performance of the Kalman filter is reflected in the statistical properties of the so called innovation sequence, where the innovation vector at time t_k is defined as

$$\mathbf{v}_k \equiv \mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f\,. \tag{5.46}$$

Adding and subtracting $\mathbf{w}_k^t$ on the right hand side of this expression, and using the equation for the observation process (5.3), we can re–write the innovation vector as

$$\mathbf{v}_k = \mathbf{v}_k - \mathbf{H}_k\mathbf{e}_k^f \tag{5.47}$$

from where it follows that $\mathcal{E}\{\mathbf{v}_k\} = \mathbf{0}$, that is, the innovation sequence has mean zero.

In this section we are interested in investigate the behavior of the cross–, or lagged–innovation covariance matrix, between times t_k and t_{k-j}, defined as

$$\begin{aligned} \mathbf{\Gamma}_{k,k-j} &\equiv \mathcal{E}\{(\mathbf{v}_k - \mathcal{E}\{\mathbf{v}_k\})(\mathbf{v}_{k-j} - \mathcal{E}\{\mathbf{v}_{k-j}\})^T\} \\ &= \mathcal{E}\{\mathbf{v}_k\mathbf{v}_{k-j}^T\} \end{aligned} \tag{5.48}$$

using that the innovation sequence has mean zero. From (5.47) we can write

$$\begin{aligned} \mathbf{\Gamma}_{k,k-j} &= \mathcal{E}\{[\mathbf{v}_k - \mathbf{H}_k\mathbf{e}_k^f][\mathbf{v}_{k-j} - \mathbf{H}_{k-j}\mathbf{e}_{k-j}^f]^T\} \\ &= \mathbf{H}_k\mathcal{E}\{\mathbf{e}_k^f(\mathbf{e}_{k-j}^f)^T\}\mathbf{H}_{k-j}^T + \mathcal{E}\{\mathbf{v}_k(\mathbf{v}_{k-j})^T\} \\ &\quad - \mathbf{H}_k\mathcal{E}\{\mathbf{e}_k^f(\mathbf{v}_{k-j})^T\} - \mathcal{E}\{\mathbf{v}_k(\mathbf{e}_{k-j}^f)^T\}\mathbf{H}_{k-j}^T \end{aligned} \tag{5.49}$$

For the particular case of $j = 0$, the innovation covariance takes the form:

$$\mathbf{\Gamma}_k = \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k \tag{5.50}$$

where we used (5.4) and (5.16).

To investigate the case with $j \geq 1$, it helps to derive a general expression for the forecast error $\mathbf{e}_k^f$. In this regard, let us combine (5.14) and (5.15) to get

$$\mathbf{e}_k^f = \mathbf{\Psi}_{k-1}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-1}\mathbf{H}_{k-1}\right]\mathbf{e}_{k-1}^f + \mathbf{\Psi}_{k-1}\tilde{\mathbf{K}}_{k-1}\mathbf{b}_{k-1}^o - \mathbf{b}_{k-1}^t \tag{5.51}$$

for any gain matrix $\tilde{\mathbf{K}}_{k-1}$, and reminding the reader that we are considering the case $\ell = 1$. Making the transformation $k \rightarrow k-1$ in the expression above, we have

$$\mathbf{e}_{k-1}^f = \mathbf{\Psi}_{k-2}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-2}\mathbf{H}_{k-2}\right]\mathbf{e}_{k-2}^f + \mathbf{\Psi}_{k-2}\tilde{\mathbf{K}}_{k-2}\mathbf{b}_{k-2}^o - \mathbf{b}_{k-2}^t \tag{5.52}$$

and substituting this back in (5.51) it follows that

$$\begin{aligned}
\mathbf{e}_k^f &= \mathbf{\Psi}_{k-1}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-1}\mathbf{H}_{k-1}\right]\mathbf{\Psi}_{k-2}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-2}\mathbf{H}_{k-2}\right]\mathbf{e}_{k-2}^f \\
&\quad + \mathbf{\Psi}_{k-1}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-1}\mathbf{H}_{k-1}\right]\mathbf{\Psi}_{k-2}\tilde{\mathbf{K}}_{k-2}\mathbf{b}_{k-2}^o + \mathbf{\Psi}_{k-1}\tilde{\mathbf{K}}_{k-1}\mathbf{b}_{k-1}^o \\
&\quad - \mathbf{\Psi}_{k-1}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-1}\mathbf{H}_{k-1}\right]\mathbf{b}_{k-2}^t - \mathbf{b}_{k-1}^t \, .
\end{aligned} \tag{5.53}$$

We can continue this iterative procedure by making the transformation $k \rightarrow k-2$ in (5.51), substitute the result back in the expression above, and so on, so that after j iterations we get

$$\mathbf{e}_k^f = \mathbf{\Phi}_{k,k-j}\mathbf{e}_{k-j}^f + \sum_{i=k-j}^{k-1} \mathbf{\Phi}_{k,i+1}\left[\mathbf{\Psi}_i\tilde{\mathbf{K}}_i\mathbf{b}_i^o - \mathbf{b}_i^t\right] \tag{5.54}$$

where we define the transition matrix $\mathbf{\Phi}_{k,k-j}$ as

$$\begin{aligned}
\mathbf{\Phi}_{k,k-j} &\equiv \mathbf{\Psi}_{k-1}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-1}\mathbf{H}_{k-1}\right]\mathbf{\Psi}_{k-2}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-2}\mathbf{H}_{k-2}\right] \\
&\quad \ldots \mathbf{\Psi}_{k-j}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-j}\mathbf{H}_{k-j}\right]
\end{aligned} \tag{5.55}$$

and also $\mathbf{\Phi}_{k,k} \equiv \mathbf{I}$.

Substituting the result (5.54) in the general expression for the innovation covariance matrix (5.49) we have

$$\begin{aligned}
\mathbf{\Gamma}_{k,k-j} &= \mathbf{H}_k\mathbf{\Phi}_{k,k-j}\mathbf{P}_{k-j}^f\mathbf{H}_{k-j}^T + \mathcal{E}\{\mathbf{v}_k(\mathbf{v}_{k-j})^T\} \\
&- \mathbf{H}_k\sum_{i=k-j}^{k-1}\mathbf{\Phi}_{k,i+1}\left[\mathbf{\Psi}_i\tilde{\mathbf{K}}_i\mathcal{E}\{\mathbf{b}_i^o(\mathbf{v}_{k-j})^T\} - \mathcal{E}\{\mathbf{b}_i^t(\mathbf{v}_{k-j})^T\}\right]\mathbf{H}_{k-j}^T
\end{aligned} \tag{5.56}$$

where we notice that, by causality, the term containing $\mathcal{E}\{\mathbf{v}_k(\mathbf{e}_{k-j}^f)^T\}$ in (5.49) is zero. Using the fact that the sequence of observation noise is white (5.4), and also that the model error $\mathbf{b}_k^t$ are uncorrelated with the observation error $\mathbf{v}_{k'}$ (5.5), for all k and k', it follows that

$$\begin{aligned}
\mathbf{\Gamma}_{k,k-j} &= \mathbf{H}_k\mathbf{\Phi}_{k,k-j}\mathbf{P}_{k-j}^f\mathbf{H}_{k-j}^T - \mathbf{H}_k\mathbf{\Phi}_{k,k-j+1}\mathbf{\Psi}_{k-j}\tilde{\mathbf{K}}_{k-j}\mathcal{E}\{\mathbf{b}_{k-j}^o(\mathbf{v}_{k-j})^T\} \\
&= \mathbf{H}_k\mathbf{\Phi}_{k,k-j}\mathbf{P}_{k-j}^f\mathbf{H}_{k-j}^T - \mathbf{H}_k\mathbf{\Phi}_{k,k-j+1}\mathbf{\Psi}_{k-j}\tilde{\mathbf{K}}_{k-j}\mathbf{R}_{k-j}
\end{aligned} \tag{5.57}$$

We can write this expression in a more convenient form, by noticing that

$$\mathbf{\Phi}_{k,k-j} = \mathbf{\Phi}_{k,k-j+1}\mathbf{\Psi}_{k-j}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-j}\mathbf{H}_{k-j}\right] \tag{5.58}$$

and making use of the optimal Kalman gain matrix $\mathbf{K}_k$, that is,

$$\begin{aligned}
\mathbf{\Gamma}_{k,k-j} &= \mathbf{H}_k\mathbf{\Phi}_{k,k-j+1}\mathbf{\Psi}_{k-j}\left[\mathbf{I} - \tilde{\mathbf{K}}_{k-j}\mathbf{H}_{k-j}\right]\mathbf{P}^f_{k-j}\mathbf{H}^T_{k-j} \\
&\quad - \mathbf{H}_k\mathbf{\Phi}_{k,k-j+1}\mathbf{\Psi}_{k-j}\tilde{\mathbf{K}}_{k-j}\mathbf{R}_{k-j} \\
&= \mathbf{H}_k\mathbf{\Phi}_{k,k-j+1}\mathbf{\Psi}_{k-j}\left[\mathbf{P}^f_{k-j}\mathbf{H}^T_{k-j} - \tilde{\mathbf{K}}_{k-j}\mathbf{H}_{k-j}\mathbf{P}^f_{k-j}\mathbf{H}^T_{k-j} - \tilde{\mathbf{K}}_{k-j}\mathbf{R}_{k-j}\right] \\
&= \mathbf{H}_k\mathbf{\Phi}_{k,k-j+1}\mathbf{\Psi}_{k-j}\left[\mathbf{P}^f_{k-j}\mathbf{H}^T_{k-j} - \tilde{\mathbf{K}}_{k-j}(\mathbf{H}_{k-j}\mathbf{P}^f_{k-j}\mathbf{H}^T_{k-j} + \mathbf{R}_{k-j})\right] \\
&= \mathbf{H}_k\mathbf{\Phi}_{k,k-j+1}\mathbf{\Psi}_{k-j}\left[\mathbf{K}_{k-j} - \tilde{\mathbf{K}}_{k-j}\right]\left(\mathbf{H}_{k-j}\mathbf{P}^f_{k-j}\mathbf{H}^T_{k-j} + \mathbf{R}_{k-j}\right) \\
&= \mathbf{H}_k\mathbf{\Phi}_{k,k-j+1}\mathbf{\Psi}_{k-j}\left[\mathbf{K}_{k-j} - \tilde{\mathbf{K}}_{k-j}\right]\mathbf{\Gamma}_{k-j}
\end{aligned} \tag{5.59}$$

where the second to last equality is obtained by noticing that (5.21) can be written as

$$\mathbf{P}^f_{k-j}\mathbf{H}^T_{k-j} = \mathbf{K}_{k-j}\left(\mathbf{H}_{k-j}\mathbf{P}^f_{k-j}\mathbf{H}^T_{k-j} + \mathbf{R}_{k-j}\right) \tag{5.60}$$

making $k \to k-j$. Consequently, for the optimal filter, when $\tilde{\mathbf{K}}_{k-j} = \mathbf{K}_{k-j}$, we see that the innovation covariance is zero, that is,

$$\mathbf{\Gamma}_{k,k-j} = \mathbf{0} \quad \text{for all } k \text{, and for all } j > 0 \ . \tag{5.61}$$

In other words, the innovation sequence is white in time when filter is optimal. This property stimulates the monitoring of the innovation sequence to determine the performance of a general sub–optimal filter.

5.2.2 Orthogonality between the Estimate and the Estimation Error

The estimate produced by the Kalman filter, $\mathbf{w}^a_k$, at any given time t_k, and its correspondent error $\mathbf{e}^a_k$ are orthogonal. Mathematically, this is expressed as

$$\mathcal{E}\{\mathbf{w}^a_k(\mathbf{e}^a_k)^T\} = \mathbf{0}\,, \tag{5.62}$$

which is only true in the optimal case, that is, when $\tilde{\mathbf{K}}_k = \mathbf{K}_k$. A path to demonstrate this property is indicated in Exercise 5.4.

5.2.3 Observability and Controllability

The concepts of observability and controllability are independent of the Kalman filter theory being considered in this lecture. These concepts are related to dynamic systems in general. However, they are of fundamental importance when studying stability properties of the Kalman filter, and for that reason we introduce these concepts in what follows.

Observability is a concept introduced to express our ability to construct the states $\mathbf{w}^t_0, \mathbf{w}^t_1, \cdots, \mathbf{w}^t_k$ of a system, given a sequence of observations $\mathbf{w}^o_0, \mathbf{w}^o_1, \cdots, \mathbf{w}^o_k$. To exemplify observability (cf. Gelb [60]), consider the evolution equation for the true state of the system for the case in which there is no stochastic forcing and in which the dynamics is independent of time, that is,

$$\mathbf{w}_k^t = \mathbf{\Psi}^\ell \mathbf{w}_{k-\ell}^t \tag{5.63}$$

represented the n–vector of the state of the system at time t_k obtained from the state at time $t_{k-\ell}$. Furthermore, consider a perfect observation process, for which the observation matrix $\mathbf{H}$ is a vector $\mathbf{h}^T$ of dimension $1 \times n$, and independent of time. In this way, we can write

$$\begin{aligned} w_0^o &= \mathbf{h}^T \mathbf{w}_0^t \\ w_1^o &= \mathbf{h}^T \mathbf{\Psi} \mathbf{w}_0^t \\ w_2^o &= \mathbf{h}^T \mathbf{\Psi}^2 \mathbf{w}_0^t \\ &\vdots \\ w_{n-1}^o &= \mathbf{h}^T \mathbf{\Psi}^{n-1} \mathbf{w}_0^t \end{aligned} \tag{5.64}$$

or yet, using vector notation,

$$\begin{pmatrix} w_0^o \\ w_1^o \\ \vdots \\ w_{k-1}^o \end{pmatrix} = \mathbf{Z}\mathbf{w}_0^t \tag{5.65}$$

Therefore the question of observability reduces to the ability of reconstructing the initial state of the system by means of the observations $w_0^o, w_1^o, \cdots, w_{n-1}^o$. Whether we can recovering the initial condition $\mathbf{w}_0^t$ of the system from the observations or not, can be assessed by considering the matrix $\mathbf{Z} = \mathbf{Z}_n$, of dimension $n \times n$, defined as

$$\mathbf{Z}_n \equiv \begin{pmatrix} \mathbf{h} & \mathbf{\Psi}^T\mathbf{h} & \cdots & (\mathbf{\Psi}^T)^{n-1}\mathbf{h} \end{pmatrix}^T \tag{5.66}$$

and whether this matrix is invertible or not. The matrix $\mathbf{Z}_n$ is invertible if it is of rank n. We say that a system is *observational* in a time $t_k > t_0$, if it is possible to construct an initial state $\mathbf{w}_0^t$ from observations $\mathbf{w}_k^o$ in the time interval (t_0, t_k). The system is said to be *completely observational* if the states $\mathbf{w}_k^t$ can be obtained from all of the observations $\mathbf{w}_k^o$.

In the general case, when the matrix $\mathbf{H}$ is of dimension $m \times n$, where m is the number of available observations, the observability matrix $\mathbf{Z}_n$ is redefined as:

$$\mathbf{Z}_n \equiv \begin{pmatrix} \mathbf{H}^T & \mathbf{\Psi}^T\mathbf{H}^T & \cdots & (\mathbf{\Psi}^T)^{n-1}\mathbf{H}^T \end{pmatrix}^T \tag{5.67}$$

and it is a matrix of dimension $nm \times n$, which should be of rank n for the system to be completely observable.

The concept of observability can be made more precise by introducing the so called *information matrix* $\mathcal{I}$,

$$\mathcal{I}(k, k-N) \equiv \sum_{i=k-N}^{k} \mathbf{\Psi}_{i,k}^T \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{\Psi}_{i,k} \tag{5.68}$$

which occurs in several recursive forms in least squares problems (or in the Kalman filter; see Jazwinski [84] pp. 205–207). According to Kalman [88] the dynamic system (5.1) and (5.3) is said to be *completely observable* if, and only if,

$$\mathcal{I}(k, 0) > \mathbf{0} \tag{5.69}$$

for all $k > 0$. Moreover, the system is said to be *uniformly completely observable* if there is an integer N, and positive constants α and β, such that

$$\mathbf{0} < \alpha\mathbf{I} \leq \mathcal{I}(k, k-N) \leq \beta\mathbf{I} \tag{5.70}$$

for all $k \geq N$. It is interesting to notice that observability depends on the properties of the dynamics $\mathbf{\Psi}_{k,k-1}$ and the observation matrix $\mathbf{H}_k$, but not explicitly on the observations $\mathbf{w}_k^o$.

Analogously, we can introduce the concept of *controllability*. This concept comes from the idea of introducing a deterministic forcing term in the evolution equation to drive the system toward a pre-specified state, within a certain period of time. This subject is, in itself, the motivation for the development of a theory called optimal control. Analogously to what is done in estimation theory, in optimal control a performance index [similar to the cost function $\mathcal{J}$ in (5.19)] serves as a measure of the proximity of the solution to the specified state. The minimization of the performance index determines the optimal forcing term, in the least squares sense, necessary to drive the state of the system to the specified state. The problem of *linear* optimal control is said to be the *dual* of the *linear*, estimation problem, in the sense that results from estimation theory have equivalent counterparts in control theory. In particular, the concept of observability, briefly introduced above, is the dual of the concept of controllability. As a consequence, we can study controllability by means of the *controllability matrix*, defined as

$$\mathcal{C}(k, k-N) \equiv \sum_{i=k-N}^{k} \mathbf{\Psi}_{i,k}\mathbf{Q}_i^{-1}\mathbf{\Psi}_{i,k}^T \tag{5.71}$$

which is the dual analogous of the observability matrix. Consequently, we say that the dynamic system (5.1) and (5.3) is *completely controllable* if, and only if,

$$\mathcal{C}(k, 0) > \mathbf{0} \tag{5.72}$$

for all k. Furthermore, we say that the system is *uniformly completely controllable* if there exists an integer N, and positive constants α and β such that

$$\mathbf{0} < \alpha\mathbf{I} \leq \mathcal{C}(k, k-N) \leq \beta\mathbf{I} \tag{5.73}$$

for all $k \geq N$. More details about this duality can be found in Kalman's original work [87], as well as in textbooks such as Gelb [60], Bryson & Ho [20], and also in the atmospheric sciences literature Ghil & Malanotte–Rizzoli [64].

The concepts of observability and controllability mentioned above are fundamental to establish *stability* results for the Kalman filter. In what follows, we summarize these results, following Dee's summary [44], which is based on the discussion Jazwinski's Section 7.6 [84].

When we inquire about system stability in the context of the Kalman filter, we are referring to the stability of the stochastic system described by the analysis equation

$$\mathbf{w}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{\Psi}_{k,k-1}\mathbf{w}_{k-1}^a + \mathbf{K}_k\mathbf{w}_k^o \tag{5.74}$$

where $\mathbf{K}_k$ is a Kalman gain matrix (5.21). The dynamics $\mathbf{\Psi}_{k,k-1}$ is assumed to be *stable*, that is ,

$$||\mathbf{\Psi}_{k,0}|| \leq c_1 \tag{5.75}$$

for all $k \geq 0$. Here $||.||$ is an appropriate matrix norm, such as the spectral norm. In fact, the homogeneous system corresponding to (5.1) is said to be *asymptotically stable* if

$$||\mathbf{\Psi}_{k,0}|| \rightarrow 0 \tag{5.76}$$

for $k \rightarrow \infty$. Furthermore, the homogeneous system corresponding to (5.1) is said to be *uniformly asymptotically stable* if

$$||\mathbf{\Psi}_{k,0}|| \rightarrow c_2 \exp(-c_3 k) \tag{5.77}$$

for all $k \geq 0$.

For this stable dynamics, the following results can be obtained, for the system governed by (5.74):

1. The analysis error covariance matrix $\mathbf{P}_k^a$ is uniformly bounded from above and below:

   $$[\mathcal{I}(k, k-N) + \mathcal{C}^{-1}(k, k-N)]^{-1} \leq \mathbf{P}_k^a \leq [\mathcal{I}^{-1}(k, k-N) + \mathcal{C}(k, k-N)]^{-1} \tag{5.78}$$

   for all $k \geq N$.

2. If $\mathbf{P}_0^a \geq \mathbf{0}$, the Kalman filter is *uniformly asymptotically stable*, that is, there are constants c_4 and c_5 such that

   $$||\mathbf{\Phi}_{k,0}|| \leq c_4 \exp(-c_5 k) \tag{5.79}$$

   for all $k \geq 0$, where $\mathbf{\Phi}_{k,0}$ is the transition matrix introduced in (5.55).

3. If $\mathbf{P}_k^a$ and $\mathbf{S}_k^a$ are two solutions of the Kalman filter equations for two initial conditions $\mathbf{P}_0^a \geq \mathbf{0}$ and $\mathbf{S}_0^a \geq \mathbf{0}$, then

   $$||\mathbf{P}_k^a - \mathbf{S}_k^a|| \leq c_4 \exp(-2c_5 k)||\mathbf{P}_0^a - \mathbf{S}_0^a|| \tag{5.80}$$

   which means that the error estimates of the Kalman filter are stable with respect to the errors of the initial state. In other words, the linear Kalman filter eventually — as data is processed in time —"forgets" about the uncertainty in the initial error covariance.

The notions of observability and controllability were initially introduced for systems governed by ordinary differential equations (see Ghil & Ide [63] for an application of interest to atmospheric sciences). These concepts can be extended to the case of systems governed by partial differential equations. A series of articles on this subject can be found in the Stavroulakis [125]. The problem of observability for discrete partial differential equations was investigated by Cohn & Dee [31].

Table 5.1: Computational requirements of the Kalman filter ($m_k = m$).

"Brute–force" implementation of the Kalman filter				
Ref.	Variable	Equation	Calculation	Flops
F1	$\mathbf{w}_k^f$	$\mathbf{\Psi}_{k-1}\mathbf{w}_{k-1}^a$	$\mathbf{\Psi w}$	$2n^2 - n$
F2	$\mathbf{P}_k^f$	$\mathbf{\Psi}_{k-1}\mathbf{P}_{k-1}^a\mathbf{\Psi}_{k-1}^T + \mathbf{Q}_{k-1}$	$\mathbf{P\Psi}^T$	$2n^3 - n^2$
			$\mathbf{\Psi}(\mathbf{P\Psi}^T)$	$2n^3 - n^2$
			$(\mathbf{\Psi P\Psi}^T) + \mathbf{Q}$	n^2
F3	$\mathbf{K}_k$	$\mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1}$	$\mathbf{HP}$	$2n^2m - nm$
			$(\mathbf{HP})\mathbf{H}^T$	$2nm^2 - m^2$
			$(\mathbf{HPH}^T) + \mathbf{R}$	m^2
			$(\mathbf{HPH}^T + \mathbf{R})^{-1}$	$2m^3$
			$(\mathbf{PH}^T)(\mathbf{HPH}^T + \mathbf{R})^{-1}$	$2nm^2 - nm$
F4	$\mathbf{w}_k^a$	$\mathbf{w}_k^f + \mathbf{K}_k(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f)$	$\mathbf{Hw}^f$	$2nm - m$
			$\mathbf{w}^o - \mathbf{Hw}^f$	m
			$\mathbf{K}(\mathbf{w}^o - \mathbf{Hw}^f)$	$2nm - n$
			$\mathbf{w}^f + [\mathbf{K}(\mathbf{w}^o - \mathbf{Hw}^f)]$	n
F5	$\mathbf{P}_k^a$	$(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f$	$\mathbf{K}(\mathbf{HP})$	$2nm^2 - m^2$
			$\mathbf{P} - \mathbf{K}(\mathbf{HP})$	n^2

5.3 Computational Aspects of the Kalman Filter

5.3.1 Generalities

We show in Table 5.3.1 the equations involved in the implementation of the Kalman filter. Although these equations are used for the case of linear systems, many approximations for the nonlinear case involve similar equations with equivalent computational cost — some computational burden is added to nonlinear systems due to the calculation of the Jacobian matrices (see the following lecture). The table displays computational cost measured in units of flops – floating point operations (multiplications and additions) — related to "brute-force" implementation of these equations. By "brute-force" we mean implementations following the operations in the table neither taking into account storage savings of certain quantities nor preventing repetitive calculations of other quantities. A detailed treatment of various implementations of the Kalman filter equations is given in Mendel [106], however for atmospheric data assimilation applications the description here should suffice. In these applications, many of the matrices in the formulas in Table 5.3.1 are not explicitly invoked due to their complexity, and are rather treated in operator form.

The following are factors that may be exploited to render computation costs more acceptable:

- the symmetry of the error covariance matrices can be used to reduce storage requirement.
- the analysis $\mathbf{P}_k^a$ and forecast $\mathbf{P}_k^f$ error covariance matrices can share the same space in memory.
- in applications to atmospheric data assimilation, the dynamics $\mathbf{\Psi}_k$ is a sparse matrix due relatively small finite–difference stencils, and only its non–zero elements need to be stored in memory. As a matter of fact, in this case, the operations corresponding to the application of $\mathbf{\Psi}_k$ to an n–vector is of order n, instead of n^2, as indicated in the table for the general case. Moreover, $\mathbf{\Psi}$ never really exists as a matrix, but rather as an operator.

The Kalman filter is subject to computational instabilities due to different possible ways to program its equations. A simple case is discussed below showing that Joseph's formula (5.18) for calculating the analysis error covariance matrix is computationally more stable than the expression (5.22), with respect to errors in calculating the gain matrix $\mathbf{K}_k$ (see next section). Even the ordering of the factors in the multiplication among matrices in the algorithm is relevant and may be responsible for numerical instability as discussed in details by Verhaegen & Van Dooren [134].

Assuming that $n \gg m$, or else, that the number of degrees of freedom n of the system is much greater than the number of observations $m_k = m$, at any given time, it is clear from Table 5.3.1 that equation F2 is responsible for the major part of the computational cost in the Kalman filter algorithm. In general, the cost of propagating the analysis error covariance matrix, to get the forecast error covariance matrix, is of the order of n^3; in the particular case of sparse dynamics, the cost gets reduced to n^2. For problems governed by *partial* differential equations, as in the case of atmospheric data assimilation, the number of degrees of freedom n reaches levels as high as 10^6–10^7, with great potential for increase as resolution of atmospheric models increase. This large number of degrees of freedom for problems in assimilation data assimilation prohibits "brute–force" implementation of the Kalman filter, even when the factors for cost reduction mentioned above are taken into account. Consequently, we are required to develop approximations to equation F2, and in some cases even to the analysis error covariance update equation F5. A lot of the research in applying the Kalman filter to atmospheric data assimilation has been done with relation to this topic (see Todling & Cohn [129], and references therein).

5.3.2 Sensitivity of the Filter to the Gains

The asymptotic stability concept for the Kalman filter discussed previously in this lecture is relatively strong, and not always the conditions for uniform asymptotic stability are satisfied. In practice, however, instability in the Kalman filter algorithm, or in suboptimal implementations of the algorithm, can be associated to lack of knowledge of model errors, observation errors, and even to specific problems due to numerical implementation of the algorithm. In this section, we look at more closely to this last aspect of instability, that is, that due to numerical implementation. We show that certain formulas are in fact more prone to numerical errors and can be, sometimes, the cause of eventual divergence of the filter.

In order to simplify notation, we momentarily omit the index referring time in the filter equations. In this manner the error covariance matrix update equation can be written using Joseph's formula as

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T \tag{5.81}$$

where the Kalman gain matrix is given by

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{5.82}$$

Alternatively, as we have seen above, the simpler formula for the analysis error covariance matrix can be obtained by substituting the optimal gain (5.82) in (5.81), that is,

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f \tag{5.83}$$

Numerical implementation of the Kalman filter generates numerical errors, even when the optimal filter is utilized — e.g., due to randoff error. In this regard, we want to investigate the effect in $\mathbf{P}^a$ caused by small errors in calculating $\mathbf{K}$ numerically. For that, assume that the gain $\mathbf{K}$ undergoes a modification $\delta\mathbf{K}$ after numerically solving (5.82), so that from (5.83) it follows that,

$$\mathbf{P}^a + \delta\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f + \delta\mathbf{KHP}^f \tag{5.84}$$

and therefore, the instantaneous error in $\mathbf{P}^a$ is given by

$$\delta\mathbf{P}^a = \delta\mathbf{KHP}^f \tag{5.85}$$

which is of first order in $\delta\mathbf{K}$.

Instead, using Joseph's formula (5.81) for the modified gain we have

$$\begin{aligned}
\mathbf{P}^a + \delta\mathbf{P}^a_{Joe} &= (\mathbf{I} - \mathbf{KH} - \delta\mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH} - \delta\mathbf{KH})^T \\
&\quad + (\mathbf{K} + \delta\mathbf{K})\mathbf{R}(\mathbf{K} + \delta\mathbf{K})^T \\
&= (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T \\
&\quad - (\mathbf{I} - \mathbf{KH})\mathbf{P}^f\mathbf{H}^T\delta\mathbf{K}^T - \delta\mathbf{KHP}^f(\mathbf{I} - \mathbf{KH})^T \\
&\quad + \delta\mathbf{KHP}^f\mathbf{H}^T\delta\mathbf{K}^T + \mathbf{KR}\delta\mathbf{K}^T + \delta\mathbf{KRK}^T + \delta\mathbf{KR}\delta\mathbf{K}^T \\
&= (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T + \delta\mathbf{K}(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})\delta\mathbf{K}^T \\
&\quad + [\mathbf{K}(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}) - \mathbf{P}^f\mathbf{H}^T]\delta\mathbf{K}^T \\
&\quad + \delta\mathbf{K}[\mathbf{K}(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}) - \mathbf{P}^f\mathbf{H}^T]^T
\end{aligned} \tag{5.86}$$

and therefore, using (5.82) and (5.81) it follows that

$$\delta\mathbf{P}^a_{Joe} = \delta\mathbf{K}(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})\delta\mathbf{K}^T \tag{5.87}$$

This shows that Joseph's formula is of second order in errors made when calculating the gain matrix, and therefore it is numerically more stable. Consequently, in many engineering implementations of the Kalman filter Joseph's formula is preferably used.

5.3.3 Serial Processing of Observations

Serial processing of observations was introduced in the literature by Bierman [12], and discussed in Parrish & Cohn [113] in the context of atmospheric data assimilation. In this section, we assume for simplicity that all the available observations are uncorrelated at all times t_k. We have in mind the uncorrelatedness not only in time, but also among variables at a fixed time.

When m observations are available at time t_k, to say these observations are uncorrelated among themselves is to say that the matrix $\mathbf{R}_k$ is diagonal, for all k, that is

$$\mathbf{R}_k = diag(\sigma_1^2, ..., \sigma_p^2) \tag{5.88}$$

where σ_i, $i = 1, 2, ..., m$, are the observation error standard deviations. Following the treatment of Parrish & Cohn [113], let us omit the index k in this section to simplify notation.

In this case, the observation process in (5.3) can be decomposed as

$$w_j^o = \mathbf{h}_j^T \mathbf{w}^t + b_j^o \tag{5.89}$$

for $j = 1, 2, ..., p$, where w_j^o is a single scalar observation, the vector $\mathbf{h}_j^T$ is the j-th row of the observation matrix $\mathbf{H}$, and b_j^o is a random number that satisfies

$$\mathcal{E}\{(b_j^o)^2\} = \sigma_j^2 \,, \tag{5.90}$$

for each j.

The assumption that the m observations, available at any given time, are uncorrelated of each another means that these observations can be processed (or assimilated) as if they became available at infinitesimally small time intervals apart. Consequently, we can iterate the equations (5.21), (5.18) and (5.23) over the observations so that we get, for each observation j:

$$\begin{aligned}
\mathbf{k}_j &= \mathbf{P}_{j-1}\mathbf{h}_j(\mathbf{h}_j^T\mathbf{P}_{j-1}\mathbf{h}_j + \sigma_j^2)^{-1} && \text{(5.91a)}\\
\mathbf{P}_j &= (\mathbf{I} - \mathbf{k}_j\mathbf{h}_j^T)\mathbf{P}_{j-1} && \text{(5.91b)}\\
\mathbf{w}_j &= \mathbf{w}_{j-1} + \mathbf{k}_j(w_j^o - \mathbf{h}_j^T\mathbf{w}_j) && \text{(5.91c)}
\end{aligned}$$

which resembles the algorithm derived in Section 4.4 for processing a newly available observation vector with the least squares algorithm. In that case, we have also assumed uncorrelatedness among observations, which was explicitly seen when writing (4.82).

Since the quantities in parenthesis in (5.91a) and in (5.91c) are scalars, and the vector $\mathbf{P}_{j-1}\mathbf{h}_j$ is used many times in different places, we can introduce an auxiliary vector $\mathbf{v}_j$ (which should not be confused with the innovation vector introduced in earlier in this lecture),

$$\mathbf{v}_j = \mathbf{P}_{j-1}\mathbf{h}_j \,, \tag{5.92}$$

so that the observation process gets reduced to the following algorithm: initialize with the forecast error covariance matrix and the forecast state vector,

$$\mathbf{P}_0 = \mathbf{P}^f \,, \tag{5.93}$$

and

$$\mathbf{w}_0 = \mathbf{w}^f \; ; \tag{5.94}$$

respectively, and iterate the following set of equations:

$$\begin{aligned}
\alpha_j &= \mathbf{h}_j^T \mathbf{v}_j + \sigma_j^2 \, , && (5.95\text{a}) \\
\mathbf{k}_j &= \frac{1}{\alpha_j} \mathbf{v}_j \, , && (5.95\text{b}) \\
\bar{\mathbf{P}}_j &= \mathbf{P}_{j-1} - \mathbf{k}_j \mathbf{v}_j^T \, , && (5.95\text{c}) \\
\bar{\mathbf{v}}_j &= \bar{\mathbf{P}}_j \mathbf{h}_j \, , && (5.95\text{d}) \\
\mathbf{P}_j &= \bar{\mathbf{P}}_j - \bar{\mathbf{v}}_j \mathbf{k}_j^T + \sigma_j^2 \mathbf{k}_j \mathbf{k}_j^T \, , && (5.95\text{e}) \\
\beta_j &= w_j^o - \mathbf{h}_j^T \mathbf{w}_{j-1} \, , && (5.95\text{f}) \\
\mathbf{w}_j &= \mathbf{w}_{j-1} + \beta_j \mathbf{k}_j \, , && (5.95\text{g})
\end{aligned}$$

for each $j = 1, 2, ..., m$, so that at the last iteration we have

$$\mathbf{P}_m = \mathbf{P}^a \, , \tag{5.96}$$

for the analysis error covariance matrix, and

$$\mathbf{w}_m = \mathbf{w}^a \, , \tag{5.97}$$

for the analysis state vector. The computational advantage of this algorithm is that it avoids the need to invert the $m \times m$ innovation error covariance matrix in (5.21), to calculate Kalman gain matrix $\mathbf{K}_k$. In the serial processing procedure, the inversion of this matrix is replaced by the inversion of the m scalar quantities in (5.95a). The demonstration of consistence between the serial algorithm above and the standard algorithm can be done by following an analogous procedure to that of Section 4.4, to process a newly available observation with *the least squares* algorithm.

The use of Joseph's formula and the consequent use of $\bar{\mathbf{P}}_j$ may suggest the need to define an auxiliary matrix of the size of the forecast error covariance matrix. However, this is only apparent, due to the notation used in writing the algorithm above. When programming these equations, the matrix $\mathbf{P}_j$ is the only one required, that is, matrices $\bar{\mathbf{P}}_j$ and $\mathbf{P}_j$ can share the same storage space. Also notice that when the elements of the state vector are directly observed, that is, when there are no linear combinations between the elements of the state vector in order to produce the observations, the elements of the vector $\mathbf{h}_j$ are all zeros except for one of them, which is in fact the unity. Consequently, the operations in (5.92) and (5.95d) are equivalent to extracting a column of the matrices $\mathbf{P}_j$.

One disadvantage of the serial processing is that we do not have access to the complete gain matrix $\mathbf{K}$, but rather only to the arrays $\mathbf{k}_j$. If we are only interested in the final result of the analysis, there is no need to obtain $\mathbf{K}$ explicitly; however, if we are particularly interested in investigating the influence of a certain observation on to distinct elements of the state vector (e.g., Ghil et al. [66]), it is necessary to calculate the complete gain matrix. The simplest way to recover the gain matrix, when using serial processing, is to do so after having obtained the analysis error covariance matrix by making use of the alternative expression for the gain matrix,

$$\mathbf{K}_k = \mathbf{P}_k^a \mathbf{H}_k^T \mathbf{R}_k^{-1} \, , \tag{5.98}$$

where, in writing the expression above we restored the time subscript k, to emphasize the fact that this should be done at the end of each analysis time t_k.

Exercises

1. Show that (5.18) reduces to (5.22) for the optimal Kalman filter gain.

2. (Gelb [60], Problem 4.8). Consider the following continuous–time dynamical system, and corresponding continuous–time observation process:

$$\dot{\mathbf{x}} = \mathbf{F}(t)\mathbf{x} + \mathbf{G}(t)\mathbf{w}$$

$$\mathbf{z} = \mathbf{H}(t)\mathbf{x} + \mathbf{v}$$

where the noises $\mathbf{w}$ e $\mathbf{v}$ are considered $\mathcal{N}(\mathbf{0}, \mathbf{Q}(t))$ and $\mathcal{N}(\mathbf{0}, \mathbf{R}(t))$, respectively, and are also decorrelated. Assume that the state estimate evolves according to the following expression:

$$\dot{\hat{\mathbf{x}}} = \tilde{\mathbf{L}}\hat{\mathbf{x}} + \tilde{\mathbf{K}}\mathbf{z}$$

where the matrices $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{K}}$ are to be determined following estimation and optimization arguments. Imposing the restriction that the estimate be unbiased, show that $\tilde{\mathbf{L}} = \mathbf{F} - \tilde{\mathbf{K}}\mathbf{H}$, and obtain the following simplified form for the estimate evolution equation:

$$\dot{\hat{\mathbf{x}}} = \mathbf{F}(t)\hat{\mathbf{x}} + \tilde{\mathbf{K}}(\mathbf{z} - \mathbf{H}\hat{\mathbf{x}})$$

Next, show that the error estimate covariance matrix evolves according to the following expression:

$$\dot{\mathbf{P}} = (\mathbf{F} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{P} + \mathbf{P}(\mathbf{F} - \tilde{\mathbf{K}}\mathbf{H})^T + \mathbf{G}\mathbf{Q}\mathbf{G}^T + \tilde{\mathbf{K}}\mathbf{R}\tilde{\mathbf{K}}^T;$$

notice that this is a general expression, in the sense that it is valid for any matrix $\tilde{\mathbf{K}}$. This expression is continuum equivalent of the Joseph formula (5.18) for the discrete–time case. As a matter of fact, we can show through a limiting procedure equivalent to that of Section 3.1.3, that the expression for the discrete case reduces to the expression above as time approaches zero (e.g., see Gelb [60]). Defining a cost function as a measure of the ratio of error change, that is, $J = \mathrm{Tr}(\dot{\mathbf{P}})$, show that its minimization leads to the following expression for the optimal gain matrix $\tilde{\mathbf{K}} = \mathbf{K}$:

$$\mathbf{K} = \mathbf{P}\mathbf{H}^T\mathbf{R}^{-1}$$

Using this formula for $\mathbf{K}$, show that the evolution equation for the error covariance is transformed to

$$\dot{\mathbf{P}} = \mathbf{F}\mathbf{P} + \mathbf{P}\mathbf{F}^T - \mathbf{P}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{P} + \mathbf{G}\mathbf{Q}\mathbf{G}^T,$$

which is known as the Riccati equation (e.g., Bittanti et al. [13])

3. (Gelb [60], Problem 4.11). Consider the following dynamical system and measurement processes:

$$\dot{x} = ax + w$$

$$z = bx + v$$

where the noises w and v are white in time, and normal, with mean zero and variances $q = const.$ and $r = const.$, respectively, for constants a and b. Assuming the initial error variance is p_0, show that the optimal filter error variance is given by

$$p(t) = \frac{(ap_0 + q)\sinh\beta t \; + \; \beta p_0 \cosh\beta t}{\left(\frac{b^2}{r}p_0 - a\right)\sinh\beta t \; + \; \beta\cosh\beta t}$$

where

$$\beta = a\sqrt{1 + \frac{b^2 q}{a^2 r}}$$

Furthermore, show that the steady–state ($t \to \infty$) variance is given by

$$p_\infty = \frac{ar}{b^2}(1 + \frac{\beta}{a})$$

which is independent of the initial variance p_0. Obtain p_∞ for a perfect model, that is, when $q = 0$. Give an interpretation to this result.

4. Show that the Kalman filter estimate $\mathbf{w}_k^a$ is orthogonal to its error $\mathbf{e}_k^a$, for all k. Using finite induction, start by showing that

$$\mathcal{E}\{\mathbf{w}_1^a(\mathbf{e}_1^a)^T\} \; = \mathbf{0}$$

and that

$$\mathcal{E}\{\mathbf{w}_2^a(\mathbf{e}_2^a)^T\} \; = \mathbf{0}$$

Then, assume that $\mathcal{E}\{\mathbf{w}_k^a(\mathbf{e}_k^a)^T\} \; = \mathbf{0}$ is true, and show that

$$\mathcal{E}\{\mathbf{w}_{k+1}^a(\mathbf{e}_{k+1}^a)^T\} \; = \mathbf{0}$$

is satisfied.

5. (Ghil et al. [66]) Consider the Kalman filter applied to the scalar, discrete–time system:

$$x_k = ax_{k-1} \; + \; w_k$$
$$z_k = x_k \; + \; v_k$$

where the noises w_k and v_k are white, normal with mean zero and variances $q = const.$ and $r = const.$, respectively. In this case, the Kalman filter reduces to the following system of equations:

$$p_k^f = Ap_{k-\ell}^a \; + \; Bq$$

$$p_k^a = \begin{cases} rp_k^f/(p_k^f + r)\,, & \text{para } k = j\ell,\, j = 1, 2, \cdots \\ p_k^f & \text{de outro modo} \end{cases}$$

where

$$A = a^{2\ell}\,, \quad B = \textstyle\sum_{m=1}^{\ell-1} a^{2m}$$

Defining $s_j = p_{j\ell}^a$, for $j = 0, 1, , 2, \cdots$ show that

$$s_j = \frac{(As_{j-1} + Bq)r}{As_{j-1} + Bq + r}$$

Consider now the perfect model case, that is, when $q = 0$, with initial error variance $p_0 = s_0$. Show that for $|a| \neq 1$,

$$s_j = \frac{A^j(A-1)s_0 r}{A(A^j-1)s_0 + (A-1)r}$$

and that for $|a| = 1$,

$$s_j = \frac{s_0 r}{j s_0 + r}$$

Finally, show that when $j \to \infty$ we have

$$\begin{array}{ll} s_j \to 0 & \text{para } |a| \leq 1 \quad , \\ s_j \to (1 - \frac{1}{A})r \, , & \text{para } |a| > 1 \, . \end{array}$$

Interpret the asymptotic results obtained above.

6. (Chui & Chen [25], Problema 2.14) Some typical engineering applications are classified under the designation ARMA (autoregressive moving–average), and can be written as:

$$\mathbf{v}_k = \sum_{i=1}^{N} \mathbf{B}_i \mathbf{v}_{k-i} + \sum_{i=0}^{M} \mathbf{A}_i \mathbf{u}_{k-i} \, ,$$

where the matrices $\mathbf{B}_1, \ldots, \mathbf{B}_N$ are $n \times n$ dimensional, and the matrices $\mathbf{A}_0, \ldots, \mathbf{A}_M$, are $n \times q$, and are independent of the time variable k. Considering $M \leq N$, show that this process can be written in the following vector form:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k$$

$$\mathbf{v}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k$$

for a vector $\mathbf{x}_k$ of dimension nN, with $\mathbf{x}_0 = \mathbf{0}$, and where

$$\mathbf{A} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{B}_2 & \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & & & \vdots \\ \mathbf{B}_{N-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \\ \mathbf{B}_N & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} , \qquad \mathbf{B}^T = \begin{pmatrix} \mathbf{A}_1 + \mathbf{B}_1\mathbf{A}_0 \\ \mathbf{A}_2 + \mathbf{B}_2\mathbf{A}_0 \\ \vdots \\ \mathbf{A}_M + \mathbf{B}_M\mathbf{A}_0 \\ \mathbf{B}_{M+1}\mathbf{A}_0 \\ \vdots \\ \mathbf{B}_N\mathbf{A}_0 \end{pmatrix} ,$$

$$\mathbf{C} = [\mathbf{I}\,\mathbf{0} \cdots \mathbf{0}] \quad \text{e} \quad \mathbf{D} = \mathbf{A}_0 \, .$$

7. *Multiple choice.* (from Bryson & Ho [20]) Consider the scalar estimation problem

$$x_{i+1} = x_i + w_i$$

$$z_i = x_i + v_i$$

where $w_i \sim \mathcal{N}(0, q)$ and white; $v_i \sim \mathcal{N}(0, 1)$ and white; w_i and v_j are uncorrelated for all i and j; and there is no initial knowledge of x_0. If $0 < q < \infty$, then the optimal estimate $\hat{x}_i$ is given by

(a) $\hat{x}_i = \frac{1}{i}\sum_{j=1}^{i} z_j$

(b) $\hat{x}_i = z_i$

(c) $\hat{x}_{i+1} = \hat{x}_i + k_i(z_{i+1} - \hat{x}_i)$, $1/(i+1) < k_i < 1$

(d) $\hat{x}_{i+1} = \hat{x}_i + k_i(z_{i+1} - \hat{x}_i)$, $1 < k_i < \infty$

Justify your answer.

8. *Multiple choice.* (from Bryson & Ho [20]) A static estimate of $\mathbf{x}$ is made from a measurement $\mathbf{z}$:

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

with $\mathbf{v} \sim \mathcal{N}(\bar{\mathbf{v}}, \mathbf{R})$ and $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P})$. The estimate is

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{K}(\mathbf{z} - \mathbf{H}\bar{\mathbf{x}})$$

where $\mathbf{K}$ is some constant matrix. The estimate is

(a) unbiased

(b) biased with a bias of $(\mathbf{K}\mathbf{H}\bar{\mathbf{x}})$

(c) biased with a bias of $(\mathbf{K}\bar{\mathbf{v}})$

(d) biased with a bias of $[\mathbf{K}(\bar{\mathbf{v}} - \mathbf{H}\bar{\mathbf{x}})]$

Justify your answer.

9. *Computer Assignment.* (Partially based on Lewis (1986) [1], Example 2.5-2.) *Computer Assignment.* Consider the following linear dynamical process[2]

$$\mathbf{x}_k \equiv \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} = \begin{pmatrix} 1 & T \\ -\omega^2 T & 1 - 2\alpha T \end{pmatrix} \begin{pmatrix} x_1(k-1) \\ x_2(k-1) \end{pmatrix} + \begin{pmatrix} w_1(k-1) \\ w_2(k-1) \end{pmatrix}$$

and the following observation process

$$z(k) = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + v(k)$$

for $\mathbf{w}(k) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, $v(k) \sim \mathcal{N}(0, r/T)$ and both uncorrelated from each other at all times. Here the (co)variance $\mathbf{Q}$ is given by

$$\mathbf{Q} = \begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix}$$

For the choice of parameters: $\omega = 0$, $\alpha = -0.1$,$r = 0.02$, and $T = 0.02$, address the following questions:

(a) Is the dynamical system stable or unstable?

[1] Lewis, F.L., 1986: *Optimal Estimation with an Introduction to Stochastic Control Theory.* John Wiley & Sons, 376 pp.

[2] This dynamical system arises from an Euler discretization of "damped" harmonic oscillator given by

$$\ddot{y}(t) + 2\alpha\dot{y} + \omega^2 = 0$$

where stochastic forcing is applied after discretization.

(b) Using Matlab, simulate the stochastic dynamical system from $k = 0$ to $k = 500$, starting from $\mathbf{x}_0 = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}$. Plot the state $\mathbf{x}_k$ against k.

(c) Using the linear Kalman filter, simulate the evolution of the error (co)variance matrix, starting from the initial condition $\mathbf{P}_0^a = \mathbf{I}$, where $\mathbf{I}$ is the 2×2 identity matrix. Plot the analysis error variance, in both variables, for the same time interval as in the previous item.

(d) Is the filter stable or unstable? Explain.

(e) Are your answers to questions (a) and (d) incompatible? Explain.

(f) Plot the true state evolution together with the analysis estimate[3] for both variables and for the time interval in item (b).

(g) *Suboptimal filters:* Let us now study the behavior of two suboptimal filters. Before starting, however, replace the analysis error (co)variance equation in your Matlab program by Joseph formula (if you are now already using it). We mentioned in this lecture that Joseph formula is valid for any gain matrix $\tilde{\mathbf{K}}_k$, thus we can use it to evaluate the performance of suboptimal filters.

i. Assuming the calculation of the forecast error (co)variance is computationally too costly for the present problem, we want to construct a suboptimal filter that somehow replaces the calculation of $\mathbf{P}_k^f$ by a simpler equation. Let us think on replacing the equation for $\mathbf{P}_k^f$ by the simple expression $\mathbf{P}_k^f = \mathbf{I}$. With this choice of forecast error (co)variance, it is simple to see that the gain matrix becomes

$$\begin{aligned} \tilde{\mathbf{K}}_k &= \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + r/T)^{-1} \\ &= \tfrac{1}{1+r/T}\mathbf{H}^T \end{aligned}$$

where we used explicitly that $\mathbf{H} = (1 \;\; 0)$ for the system under consideration. Keeping the equation for $\mathbf{P}_k^f$, in your Matlab code as dictated by the Kalman filter, replace the expression for the optimal gain by the one given above. This turns the state estimate in a suboptimal estimate. Also, since you have kept the original expression for the forecast error (co)variance evolution, and your are using Joseph formula for the analysis error (co)variance, these two quantities provide now filter performance information due to suboptimal choices of gains. With the "approximate" gain matrix above, is the resulting filter stable or unstable? Explain. If this is not a successful choice of gain matrix, can you explain why that is?

ii. Let us know build another suboptimal filter that replaces the gain by the asymptotic gain obtained from the optimal run in item (b). To obtain the optimal asymptotic gain, you need to run the experiment in item (b) again, output the gain matrix at the last time step from that run, and use it as a suboptimal choice for the gain matrix in this item. You should actually

[3] Remember that your initial estimate should be sampled from the initial state where the initial error is $\mathcal{N}(0, \mathbf{P}_0^a)$, that is,

$$\mathbf{x}_0^a = \mathbf{x}_0 + \boxed{\text{chol}(\mathbf{P}_0^a)} * \boxed{\text{randn}(:)}$$

writing is a very symbolic manner.

make sure that the gain has asymptote by looking at its value for a few time steps before the last time step, and verifying that these values are indeed the same. Now run a similar experiment than that of the previous item, but using the asymptotic gain for the suboptimal gains at all time steps. Is the resulting filter stable or unstable? (Note: This choice of gain corresponds to using the so called Wiener filter.)

Chapter 6

Basic Concepts in Nonlinear Filtering

6.1 Introduction

In this lecture, we are interested in treating the estimation problem for nonlinear systems. Due to the difficulties involved in problems of this type, and the various possible methods to approach them, we will confine this lecture to the study of the case of systems governed by stochastic *ordinary* differential equations. In fact, we assume that the case of atmospheric data assimilation, which the governing equations are a set of *partial* differential equations, can be formulated in terms of a system of ordinary differential equations. In general, this can be done by treating the space variables on a discrete grid, restricting in this manner the infinite dimensional problem to the case of finite dimension. Although this type of argument is in general a good starting point for dealing with the problem of data assimilation in earth sciences, it appears that the best way would be to study the system of governing equations and observational process in the continuum, at least to a very good extent. Discretization leads to modeling errors which have not been treated appropriately so far in that field. For further details related to this point of view the reader is referred to the excellent discussion in Cohn [27].

We saw in Lecture 4, that in many cases of Bayes estimation an estimate of the variable of interest reduces to the conditional mean. As a matter of fact, the estimate of minimum variance is the conditional mean. In case of Gaussian processes, other optimization criteria produces similar estimates to the one given by the minimum variance. Furthermore, we saw in Lecture 5, that when a linear observation process is combined with a linear dynamical process the Kalman filter provides the best linear unbiased estimate (BLUE). When the statistics of errors are Gaussian the Kalman filter estimates correspond to the conditional mean. In the nonlinear case, even with Gaussian error statistics, the resulting estimates are not Gaussian distributed and consequently different Bayes optimization criteria lead to distinct estimates, in particular not necessarily coinciding with the conditional mean. One of the consequences of Gaussian error statistics in the linear case is that only the first two moments are enough to describe the process completely; in the nonlinear case, on the other hand, moments of higher order may play an important role in describing the process. Ideally, the transition probabilities related to the processes under consideration should be the quantities being calculated, however, in most practical applications calculating these quantities requires computations well beyond available resources.

A precise treatment of the estimation problem for nonlinear systems can be made following statistical arguments. Since the probability density of the variables of the system contain all the necessary information to describe the system, the probabilistic method studies the evolution of the probability density in time, as well as the way by which this quantity is modified as observations become available. In Lecture 3, we obtained evolution equations only for the first two moments of the probability density for nonlinear dynamical systems. In fact, it is possible to show that the probability density evolves according to the Fokker–Planck equation, and once its evolution is determined we can determine any desired moment.

For the continuous–discrete system case, the conditional probability density evolves through the Fokker–Planck equation during the intervals of time in which there are no observations. At observation times the transition undertaken by the probability density due to the observations can be evaluated through Bayes rule. The rigorous mathematical treatment following this procedure can be found, for example, in the classic text books of Jazwinski [84] and Sage & Melsa [121], or in more modern texts such as that of Øksendal's [108]. The precise treatment of nonlinear estimation problems is beyond the scope of our introductory course.

6.2 The Extended Kalman Filter

In this section, we follow the simple treatment of Gelb [60] to derive the so called extended Kalman filter. We consider the continuous–discrete system problem, that is, the case in which the dynamics evolves continuously in time whereas the observations are available at discrete times $t_1, t_2, \ldots$. The modification for the case in which the dynamical system is discrete in time can be derived using the results from Section 3.2.2.

The continuous–time dynamical process, corresponding to the evolution of the n–vector $\mathbf{w}^t(t)$ — the variable of interest — is written here as

$$\frac{d\mathbf{w}^t(t)}{dt} = \mathbf{f}(\mathbf{w}^t, t) + \mathbf{b}^t(t) \tag{6.1}$$

and the discrete–time observation process, at times $t_{k-1} \leq t < t_k$, is written as

$$\mathbf{w}_k^o = \mathbf{h}(\mathbf{w}_k^t) + \mathbf{b}_k^o \tag{6.2}$$

where the $\mathbf{w}_k^t \equiv \mathbf{w}^t(t = t_k)$, and the m–vector $\mathbf{w}_k^o$ corresponds to the observation vector. We assume the n–vector process noise $\{\mathbf{b}^t(t)\}$ is white in time, Gaussian, with mean zero and (co)variance $\mathbf{Q}(t)$. Similarly, we assume the m–vector observation noise $\{\mathbf{b}_k^o\}$ is white in time, Gaussian, with mean zero and (co)variance $\mathbf{R}_k$. Moreover, the processes $\{\mathbf{b}^t(t)\}$ and $\{\mathbf{b}_k^o\}$ are assumed to be uncorrelated at all times. Analogously to what we have done in the previous lecture, let us indicate by $\mathbf{W}_k^o = \{\mathbf{w}_k^o, \mathbf{w}_{k-1}^o, \cdots, \mathbf{w}_1^o\}$ the set of all observations up to and including time t_k. The n–vector function $\mathbf{f}$ corresponds to the dynamics of the system and the m–vector function $\mathbf{h}$ corresponds to the observation operator.

The most common procedure to deal with estimation problems for nonlinear systems is that of minimum variance. Since the estimate with minimum variance corresponds to the conditional mean, we choose to calculate the conditional mean during the interval of time in which there are no observations. In this way, we want to calculate $\mathcal{E}\{\mathbf{w}^t(t)|\mathbf{W}^o_{k-1}\}$ during the interval of time $t_{k-1} \leq t < t_k$. According to to (6.1) it follows that

$$\frac{d\mathcal{E}\{\mathbf{w}^t(t)|\mathbf{W}^o_{k-1}\}}{dt} = \mathcal{E}\{\mathbf{f}[\mathbf{w}^t(t), t]|\mathbf{W}^o_{k-1}\} \tag{6.3}$$

where we used the fact that the process $\{\mathbf{b}(t)\}$ is white and has mean zero.

A measure of the error in the estimate can be obtained by means of the conditional error (co)variance matrix $\mathbf{P}(t)$, defined as

$$\mathbf{P}(t) \equiv \mathcal{E}\{[\mathbf{w}^t(t) - \mathcal{E}\{\mathbf{w}^t(t)|\mathbf{W}^o_{k-1}\}]\,[\mathbf{w}^t(t) - \mathcal{E}\{\mathbf{w}^t(t)|\mathbf{W}^o_{k-1}\}]^T\,|\mathbf{W}^o_{k-1}\} \tag{6.4}$$

for $t_{k-1} \leq t < t_k$. The evolution equation of this matrix between two consecutive observation times can be determined as in Lecture 3. Integrating (6.3) between t_{k-1} and t_k, substituting the result in the definition of $\mathbf{P}$, differentiating the resulting expression and using the properties of the processes $\{\mathbf{w}^t(t)\}$ and $\{\mathbf{b}^t(t)\}$ we obtain (see Exercise 6.1):

$$\begin{aligned}\dot{\mathbf{P}}(t) &= \mathcal{E}\{\mathbf{w}^t(t)\mathbf{f}^T[\mathbf{w}^t(t), t]|\mathbf{W}^o_{k-1}\} \;- \mathcal{E}\{\mathbf{w}^t(t)|\mathbf{W}^o_{k-1}\}\mathcal{E}\{\mathbf{f}[\mathbf{w}^t(t), t]|\mathbf{W}^o_{k-1}\}^T \\ &+ \mathcal{E}\{\mathbf{f}[\mathbf{w}^t(t), t]\mathbf{w}^{tT}(t)|\mathbf{W}^o_{k-1}\} - \mathcal{E}\{\mathbf{f}[\mathbf{w}^t(t), t]|\mathbf{W}^o_{k-1}\}\mathcal{E}\{\mathbf{w}^t(t)|\mathbf{W}^o_{k-1}\}^T \\ &+ \mathbf{Q}(t)\end{aligned} \tag{6.5}$$

which is often written in the more compact form,

$$\begin{aligned}\dot{\mathbf{P}}(t) &= \mathcal{E}\{\mathbf{w}^t\mathbf{f}^T\}_{k-1} - \mathcal{E}\{\mathbf{w}^t\}_{k-1}\mathcal{E}\{\mathbf{f}\}^T_{k-1} \\ &+ \mathcal{E}\{\mathbf{f}\mathbf{w}^{tT}\}_{k-1} - \mathcal{E}\{\mathbf{f}\}_{k-1}\mathcal{E}\{\mathbf{w}^t\}^T_{k-1} + \mathbf{Q}(t)\end{aligned} \tag{6.6}$$

where we wrote the conditional ensemble mean operator in the compact form: $\mathcal{E}\{(.)\}_{k-1} = \mathcal{E}\{.|\mathbf{W}^o_{k-1}\}$, and we omitted the explicit functional dependencies of $\mathbf{w}^t$ and $\mathbf{f}$. The equations for the mean and error (co)variance are not ordinary differential equations in the usual sense because they depend on the ensemble mean. To solve these equations it is necessary to know the probability density of the process $\{\mathbf{w}^t(t)\}$, which in general is not known. Moreover, we should calculate the corresponding moments depending on the function $\mathbf{f}[\mathbf{w}^t(t)]$.

The simplest approximation to the equation for the evolution of the mean (6.3) and to the equation for the evolution of the second moment (6.6), follows what we have seen in Lecture 3. That is, let us introduce the forecast vector $\mathbf{w}^f_k$ as a suitable approximation for the conditional mean, that is,

$$\mathbf{w}^f(t) \approx \mathcal{E}\{\mathbf{w}^t(t)|\mathbf{W}^o_{k-1}\} \tag{6.7}$$

In the extended Kalman filter, we expand the function $\mathbf{f}[\mathbf{w}^t(t)]$ as a Taylor series about the approximate mean $\mathbf{w}^f(t)$, and retain only up to the the first order term. Thus, in the time interval $t_{k-1} \leq t < t_k$, between two consecutive observations, we write

$$\mathbf{f}[\mathbf{w}^t(t), t] \approx \mathbf{f}[\mathbf{w}^f(t), t] + \mathcal{F}'[\mathbf{w}^f(t), t](\mathbf{w}^t(t) - \mathbf{w}^f(t)) \tag{6.8}$$

where, as in Lecture 3, $\mathcal{F}$ is the $n \times n$ Jacobian matrix defined as

$$\mathcal{F}'[\mathbf{w}^f(t), t] \equiv \left. \frac{\partial \mathbf{f}[\mathbf{w}^t(t), t]}{\partial [\mathbf{w}^t(t)]^T} \right|_{\mathbf{w}^t(t) = \mathbf{w}^f(t)} \tag{6.9}$$

Consequently, using the expansion and (6.7) the forecast equation becomes

$$\dot{\mathbf{w}}^f(t) = \mathbf{f}[\mathbf{w}^f(t), t] \tag{6.10}$$

valid for the times t in the interval between t_{k-1} and t_k: Using the expansion (6.8) in (6.6) we obtain that

$$\dot{\mathbf{P}}^f(t) = \mathcal{F}[\mathbf{w}^f(t), t]\mathbf{P}^f(t) + \mathbf{P}(t)\mathcal{F}^T[\mathbf{w}^f(t), t] + \mathbf{Q}(t) \tag{6.11}$$

which is identical to what we saw in Lecture 3, with the additional restriction that this expression applies only when $t \in [t_{k-1}, t_k)$. Notice that here,

$$\mathbf{P}^f(t) \equiv \mathcal{E}\{[\mathbf{w}^t(t) - \mathbf{w}^f(t)\}]\,[\mathbf{w}^t(t) - \mathbf{w}^f(t)\}]^T\} \approx \mathbf{P}(t) \tag{6.12}$$

that is, $\mathbf{P}^f(t)$ is an approximation to the conditional error (co)variance matrix $\mathbf{P}(t)$.

The problem of producing an estimate $\mathbf{w}_k^a$ in t_k of the state of the system, using the observation $\mathbf{w}_k^o$, is what we want to solve in order to obtain a filtered estimate. Motivated by the results obtained in the linear case, we assume that such an estimate can be obtained as a linear combination among the observations. Hence, we write

$$\mathbf{w}_k^a = \mathbf{u}_k + \tilde{\mathbf{K}}_k \mathbf{w}_k^o \tag{6.13}$$

where the n–vector $\mathbf{u}_k$ and the $n \times m$ matrix $\tilde{\mathbf{K}}_k$ are deterministic (non-stochastic) quantities to be determined from statistical and optimization arguments, just as we did in the linear case.

Introduce the analysis and forecast estimation errors, that is, $\mathbf{e}_k^a$ is the error in the estimate at time t_k which includes the current observation, while $\mathbf{e}_k^f$ is the error in the estimate at time t_k which includes observations only up to time t_{k-1}:

$$\mathbf{e}_k^a \equiv \mathbf{w}_k^a - \mathbf{w}_k^t \tag{6.14a}$$

$$\mathbf{e}_k^f \equiv \mathbf{w}_k^f - \mathbf{w}_k^t \tag{6.14b}$$

By adding and subtracting $\mathbf{w}_k^t$ from the left–hand–side of expression (6.13), and using (6.2) we can write

$$\begin{aligned} \mathbf{e}_k^a &= \mathbf{w}_k^a - \mathbf{w}_k^t \\ &= \mathbf{u}_k + \tilde{\mathbf{K}}_k \left[\mathbf{h}(\mathbf{w}_k^t) + \mathbf{b}_k^o\right] - \mathbf{w}_k^t \end{aligned} \tag{6.15}$$

Now, adding and subtracting $\mathbf{w}_k^f$ from the right–hand–side of last equality above we get

$$\mathbf{e}_k^a = \mathbf{u}_k + \tilde{\mathbf{K}}_k \left[\mathbf{h}(\mathbf{w}_k^t) + \mathbf{b}_k^o\right] + \mathbf{e}_k^f - \mathbf{w}_k^f \tag{6.16}$$

According to Bayes estimation, one of the desired properties of an estimate is that it be unconditionally unbiased. This means that we want $\mathcal{E}\{\mathbf{e}_k^a\} = \mathbf{0}$. Therefore, applying the ensemble mean operator to the expression above it follows that

$$\mathcal{E}\{\mathbf{e}_k^a\} = \mathcal{E}\{\left[\mathbf{u}_k + \tilde{\mathbf{K}}_k\mathbf{h}(\mathbf{w}_k^t) - \mathbf{w}_k^f\right]\} + \mathcal{E}\{\mathbf{b}_k^o\} + \mathcal{E}\{\mathbf{e}_k^f\} \tag{6.17}$$

Recall that the sequence $\{\mathbf{b}_k^o\}$ has mean zero, thus $\mathcal{E}\{\mathbf{b}_k^o\} = \mathbf{0}$. Moreover, inspired by the linear case, we assume that the forecast error is unbiased. A word of caution is appropriate here: it is important to recognize that this is an assumption we know can only be approximately correct since the forecast is only an approximation to the conditional mean, that is,

$$\begin{aligned}\mathcal{E}\{\mathbf{e}_k^f\} &= \mathcal{E}\{\mathbf{w}_k^f - \mathbf{w}_k^t\} \\ &\approx \mathcal{E}\{\mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}\} - \mathcal{E}\{\mathbf{w}_k^t\} \\ &= \mathcal{E}\{\mathbf{w}_k^t\} - \mathcal{E}\{\mathbf{w}_k^t\} \\ &= \mathbf{0}\end{aligned} \tag{6.18}$$

thus $\mathcal{E}\{\mathbf{e}_k^f\} \approx \mathbf{0}$. With that in mind, from (6.17) we see that for the estimate to be (approximately) unbiased we should satisfy:

$$\mathcal{E}\{\left[\mathbf{u}_k + \tilde{\mathbf{K}}_k\mathbf{h}(\mathbf{w}_k^t) - \mathbf{w}_k^f\right]\} = \mathbf{0} \tag{6.19}$$

Since $\mathbf{u}_k$ was assumed to be deterministic we have

$$\begin{aligned}\mathbf{u}_k &= \mathcal{E}\{\left[\mathbf{w}_k^f - \tilde{\mathbf{K}}_k\mathbf{h}(\mathbf{w}_k^t)\right]\} \\ &= \mathbf{w}_k^f - \tilde{\mathbf{K}}_k\mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\}\end{aligned} \tag{6.20}$$

Substituting this result in (6.13) we obtain

$$\mathbf{w}_k^a = \mathbf{w}_k^f + \tilde{\mathbf{K}}_k\left[\mathbf{w}_k^o - \mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\}\right] \tag{6.21}$$

Moreover, the analysis error can be re-written as

$$\mathbf{e}_k^a = \mathbf{e}_k^f + \tilde{\mathbf{K}}_k\left[\mathbf{h}(\mathbf{w}_k^t) - \mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\} + \mathbf{b}_k^o\right] \tag{6.22}$$

As in the linear case, we want to minimize the analysis error variance (6.22). We introduce the analysis error (co)variance matrix

$$\mathbf{P}_k^a \equiv \mathcal{E}\{\mathbf{e}_k^a\,(\mathbf{e}_k^a)^T\} \tag{6.23}$$

and the problem of minimization is reduced to the problem of minimizing the trace of this matrix,

$$\mathcal{J}_k^a \equiv \mathrm{Tr}(\mathbf{P}_k^a) \tag{6.24}$$

as in (5.19), but using $\mathbf{E}_k = \mathbf{I}$ in that expression, without loss of generality.

From (6.22) it follows that

$$\begin{aligned}\mathbf{P}_k^a &= \mathbf{P}_k^f + \tilde{\mathbf{K}}_k\mathcal{E}\left\{\left[\mathbf{h}(\mathbf{w}_k^t) - \mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\}\right]\left[\mathbf{h}(\mathbf{w}_k^t) - \mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\}\right]^T\right\}\tilde{\mathbf{K}}_k^T \\ &+ \mathcal{E}\left\{\mathbf{e}_k^f\left[\mathbf{h}(\mathbf{w}_k^t) - \mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\}\right]^T\right\}\tilde{\mathbf{K}}_k^T \\ &+ \tilde{\mathbf{K}}_k\mathcal{E}\left\{\left[\mathbf{h}(\mathbf{w}_k^t) - \mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\}\right](\mathbf{e}_k^f)^T\right\} + \tilde{\mathbf{K}}_k\mathbf{R}_k\tilde{\mathbf{K}}_k^T\end{aligned} \tag{6.25}$$

Applying the trace operator to this expression and using the rules from matrix calculus introduced earlier in this lecture, we can solve

$$\frac{\partial \mathcal{J}_k^a}{\partial \tilde{\mathbf{K}}_k} = 0 \tag{6.26}$$

to find that the gain matrix minimizing $\mathcal{J}_k^a$ is given by

$$\begin{aligned}\tilde{\mathbf{K}}_k \equiv \mathbf{K}_k \ &= \ \mathcal{E}\left\{\mathbf{e}_k^f \left[\mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\} - \mathbf{h}(\mathbf{w}_k^t)\right]^T\right\} \\ &\quad \times \left\{\mathcal{E}\{\left[\mathbf{h}(\mathbf{w}_k^t) - \mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\}\right]\left[\mathbf{h}(\mathbf{w}_k^t) - \mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t)\}\right]^T\} + \mathbf{R}_k\right\}^{-1}\end{aligned} \tag{6.27}$$

Substituting this gain matrix in the general equation for the analysis error (co)variance at time t_k we have

$$\mathbf{P}_k^a = \mathbf{P}_k^f - \mathbf{K}_k \mathcal{E}\left\{\left[\mathcal{E}\{\mathbf{h}(\mathbf{w}_k^t) - \mathbf{h}(\mathbf{w}_k^t)\}\right](\mathbf{e}_k^f)^T\right\} \tag{6.28}$$

(see Exercise 6.2). Equations (6.22), (6.27) and (6.28) provide the minimum variance estimate, optimal gain and corresponding error (co)variance at time t_k. These expressions involve the ensemble average operator and consequently cannot be used directly. Analogously to the Taylor expansion used for $\mathbf{f}[\mathbf{w}^t(t), t]$, when deriving a closed form for the evolution of the first and second moments in the interval of time between two consecutive observations, we expand the function $\mathbf{h}[\mathbf{w}^t(t)]$ about the estimate of the state of the system available at time t_k before the observations are processed, that is, $\mathbf{w}_k^f$. Therefore,

$$\begin{aligned}\mathbf{h}[\mathbf{w}^t(t)] \ &\approx \ \mathbf{h}[\mathbf{w}^f(t)] - \mathcal{H}'[\mathbf{w}^f(t), t](\mathbf{w}^f(t) - \mathbf{w}^t(t)) \\ &= \ \mathbf{h}[\mathbf{w}^f(t)] - \mathcal{H}'[\mathbf{w}^f(t), t]\mathbf{e}^f(t)\end{aligned} \tag{6.29}$$

where $\mathcal{H}'$ is the $m \times m$ Jacobian matrix defined as

$$\mathcal{H}'[\mathbf{w}^f(t), t] \equiv \left.\frac{\partial \mathbf{h}[\mathbf{w}^t(t), t]}{\partial [\mathbf{w}^t(t)]^T}\right|_{\mathbf{w}^t(t) = \mathbf{w}^f(t)} \tag{6.30}$$

of the m–vector function $\mathbf{h}$. The substitution of this approximation in expressions (6.22), (6.27) and (6.28) produces

$$\mathbf{w}_k^a = \mathbf{w}_k^f + \mathbf{K}_k\left[\mathbf{w}_k^o - \mathbf{h}(\mathbf{w}_k^f)\right] \tag{6.31a}$$

$$\mathbf{K}_k = \mathbf{P}_k^f \mathcal{H}'^T(\mathbf{w}_k^f)\left[\mathcal{H}'(\mathbf{w}_k^f)\mathbf{P}_k^f \mathcal{H}'^T(\mathbf{w}_k^f) + \mathbf{R}_k\right]^{-1} \tag{6.31b}$$

$$\mathbf{P}_k^a = \left[\mathbf{I} - \mathbf{K}_k \mathcal{H}'(\mathbf{w}_k^f)\right]\mathbf{P}_k^f \tag{6.31c}$$

which correspond to a closed set of equations for the update of the state estimate, gain matrix and corresponding error (co)variance. The equations above, together with (6.10) and (6.11), form the set of equations constituting the *extended* Kalman filter. In case the functions $\mathbf{f}$ and $\mathbf{h}$ are linear, these equations reduced to those of the standard Kalman filter, derived in the previous lecture. The results obtained above can be extended to the case of continuous–time observation processes. Also, higher order expressions can be derived by considering higher order terms in the Taylor expansions for the functions $\mathbf{f}$ and $\mathbf{h}$ (see Gelb [60], Jazwinski [84], and Sage & Melsa [121]). Equations for the case of discrete–time

Table 6.1: Extended Kalman filter: discrete–discrete systems.

Dynamical Process	$\mathbf{w}_k^t = \psi(\mathbf{w}_{k-1}^t) + \mathbf{b}_{k-1}^t$
Observational Process	$\mathbf{w}_k^o = \mathbf{h}(\mathbf{w}_k^t) + \mathbf{b}_k^o$
Estimate Propagation	$\mathbf{w}_k^f = \psi(\mathbf{w}_{k-1}^a)$
Error Covariance Propagation	$\mathbf{P}_k^f = \mathcal{F}'(\mathbf{w}_{k-1}^a)\mathbf{P}_{k-1}^a\mathcal{F}'^T(\mathbf{w}_{k-1}^a) + \mathbf{Q}_{k-1}$
Gain Matrix	$\mathbf{K}_k = \mathbf{P}_k^f\mathcal{H}'^T(\mathbf{w}_k^f)\left[\mathcal{H}'(\mathbf{w}_k^f)\mathbf{P}_k^f\mathcal{H}'^T(\mathbf{w}_k^f) + \mathbf{R}_k\right]^{-1}$
Estimate Update	$\mathbf{w}_k^a = \mathbf{w}_k^f + \mathbf{K}_k\left[\mathbf{w}_k^o - \mathbf{h}(\mathbf{w}_k^f)\right]$
Estimate Error Covariance Update	$\mathbf{P}_k^a = \left[\mathbf{I} - \mathbf{K}_k\mathcal{H}'(\mathbf{w}_k^f)\right]\mathbf{P}_k^f$
Definitions	$\mathcal{F}'(\mathbf{w}_k^a) \equiv \left.\frac{\partial\psi(\mathbf{w})}{\partial\mathbf{w}^T}\right\vert_{\mathbf{w}=\mathbf{w}_k^a}$ $\mathcal{H}'(\mathbf{w}_k^f) \equiv \left.\frac{\partial\mathbf{h}(\mathbf{w})}{\partial\mathbf{w}^T}\right\vert_{\mathbf{w}=\mathbf{w}_k^f}$

dynamics and discrete–time observations which can be derived simply by using the equations for mean and (co)variance evolution given in Section 3.2.2, are displayed in Table 6.2, which is an adaptation of Table 9.4-3 of Sage & Melsa [121]. Applications of the extended Kalman filter in the contexts of atmospheric and oceanic data assimilation are those of Bürger & Cane [21], Daley [38], Evensen [52], Ménard [104], Miller et al. [107], to cite just a few.

It is important to notice that, contrary to what we saw in Section 5.1.3, The analysis and forecast error (co)variance matrices now depend on the observations – therefore expressions (5.41)–(5.42) is not valid in the nonlinear case. The error (co)variance matrices are functions of the Jacobian matrices $\mathcal{F}'$ and $\mathcal{H}'$, which are functions of the current estimate — which in turn depends on the observations themselves. Thus, the gain matrix $\mathbf{K}_k$ and the error (co)variances $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$ are random, due to the fact that they depend on the set of observations $\mathbf{W}^o$. But most importantly is the fact that neither one of these covariance matrices correspond to the conditional error (co)variance matrices, but are rather approximations to these quantities. The same is true about the estimates $\mathbf{w}_k^f$ and $\mathbf{w}_k^a$ provided by the extended Kalman filter, that is, they represent only approximations to the conditional mean, in particular these estimates are only approximately unbiased. Therefore, precisely putting it, the extended Kalman filter provides biased state estimates.

6.3 An Approach to Parameter Estimation

In this section we want to briefly point out that extensions of the Kalman filter to nonlinear systems, such as the extended Kalman filter discussed in last section, can be used to estimate unknown system parameters, even when the dynamics and observation process are linear. These parameters can be either related to the dynamics, that is, to the vector function $\mathbf{f}$ or to the observation operator $\mathbf{h}$. It is also possible to estimate parameters related to the statistics of the errors involved in the problem.

As a simple illustration, consider the discrete–discrete system described by

$$\begin{aligned} \mathbf{w}_k^t &= \Psi(\boldsymbol{\theta})\mathbf{w}_{k-1}^t + \mathbf{b}_{k-1}^t && (6.32a) \\ \mathbf{w}_k^o &= \mathbf{H}\mathbf{w}_k^t + \mathbf{b}_k^o && (6.32b) \end{aligned}$$

where the sequence of the noises $\{\mathbf{b}_k^t\}$ and $\{\mathbf{b}_k^o\}$ are as in the previous section Gaussian with mean zero and given (co)variances, that is, $\mathbf{b}_k^t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ and $\mathbf{b}_k^o \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$, and are mutually uncorrelated. The system (6.32) does not represent the most general form for problems of parameter estimation, since we are assuming that the equations are linear in the state variable. Another simplification in the system studied here is that the observation function is taken as known, with no parameters to be determined to describe it. Also, the noises error (co)variances are assumed to be completely known. Even with all these simplifications, the system above is sufficient to exemplify the main idea of the approach of parameter estimation based on the extended Kalman filter.

In system (6.32) the variable $\boldsymbol{\theta}$ represents an r–vector of constant, but unknown, coefficients that we intend to estimate. Notice from the beginning that the problem of parameter estimation is always nonlinear (expect in the case of additive unknown parameters — see Jazwinski [84], Section 8.4), making essential the use of nonlinear filter procedures. If we imagine that the parameters $\boldsymbol{\theta}$ are functions of time, the fact that they are in reality constant can be expressed as $\boldsymbol{\theta} = \boldsymbol{\theta}_k^t = \boldsymbol{\theta}_{k-1}^t$. This equality produces an extra equation that we can append to the system above, to augment the state vector, that is

$$\mathbf{u}_k^t \equiv \begin{pmatrix} \mathbf{w}_k^t \\ \boldsymbol{\theta}_k^t \end{pmatrix} \qquad (6.33)$$

where $n + r$–vector $\mathbf{u}^t$ is now the re–defined state variable. Unfortunately, this procedure does not lead to anything (see Exercise 6.3), in terms of estimating $\boldsymbol{\theta}$.

To be able to actually estimate $\boldsymbol{\theta}$ through, say, the extended Kalman filter it is necessary to treat the vector of deterministic, constant and unknown parameters as if it were a random vector. Thus, we write the equation for the parameters to be estimated as

$$\boldsymbol{\theta}_k^t = \boldsymbol{\theta}_{k-1}^t + \boldsymbol{\epsilon}_k \qquad (6.34)$$

where $\boldsymbol{\epsilon}_k$ is an r–random vector with assumed known statistics, $\boldsymbol{\epsilon}_k \sim \mathcal{N}(0, \mathbf{S}_k)$ — taken to be uncorrelated from the errors $\mathbf{b}_k^t$ and $\mathbf{b}_k^o$ — system (6.32) can be re–written in the form

$$\begin{aligned} \mathbf{u}_k^t &= \mathbf{f}(\mathbf{u}_{k-1}^t) + \begin{pmatrix} \mathbf{b}_k^t \\ \boldsymbol{\epsilon}_k \end{pmatrix} && (6.35a) \\ \mathbf{w}_k^o &= \begin{pmatrix} \mathbf{H} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{w}_k^t \\ \boldsymbol{\theta}_k^t \end{pmatrix} + \mathbf{b}_k^o && (6.35b) \end{aligned}$$

where $\mathbf{f}(\mathbf{u}_{k-1}^t)$ is defined as

$$\mathbf{f}(\mathbf{u}_{k-1}^t) \equiv \begin{pmatrix} \Psi(\boldsymbol{\theta}_{k-1}^t) & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \mathbf{u}_{k-1}^t = \begin{pmatrix} \Psi(\boldsymbol{\theta}_{k-1}^t)\mathbf{w}_{k-1}^t \\ \boldsymbol{\theta}_{k-1}^t \end{pmatrix} \tag{6.36}$$

Let us assume that initially, at $t = t_0$, the estimates $\mathbf{w}_0^t$ and $\boldsymbol{\theta}_0^t$ are

$$\mathbf{u}_0^a \equiv \begin{pmatrix} \mathbf{w}_0^a \\ \boldsymbol{\theta}_0^a \end{pmatrix} = \begin{pmatrix} \mathcal{E}\{\mathbf{w}_0^t\} \\ \bar{\boldsymbol{\theta}}_0 \end{pmatrix} \tag{6.37}$$

with error (co)variance

$$\mathbf{P}_0^a = \begin{pmatrix} cov\{\mathbf{w}_0^t, \mathbf{w}_0^t\} & \mathbf{0} \\ \mathbf{0} & \Theta_0 \end{pmatrix} \tag{6.38}$$

Following the extended Kalman filter equations listed in Table 6.2, for the discrete–discrete case, we need to calculate the Jacobian of the modified dynamics in (6.35a). That is,

$$\begin{aligned} \mathcal{F}'(\mathbf{u}_k^a) &\equiv \left.\frac{\partial \mathbf{f}(\mathbf{u}_k^t)}{\partial [\mathbf{u}_k^t]^T}\right|_{\mathbf{u}_k^t=\mathbf{u}_k^a} \\ &= \begin{pmatrix} \left.\frac{\partial \Psi(\boldsymbol{\theta}_k^t)\mathbf{w}_k^t}{\partial [\mathbf{w}_k^t]^T}\right|_{\mathbf{u}_k^t=\mathbf{u}_k^a} & \left.\frac{\partial \Psi(\boldsymbol{\theta}_k^t)\mathbf{w}_k^t}{\partial [\boldsymbol{\theta}_k^t]^T}\right|_{\mathbf{u}_k^t=\mathbf{u}_k^a} \\ \left.\frac{\partial \boldsymbol{\theta}_k^t}{\partial [\mathbf{w}_k^t]^T}\right|_{\mathbf{u}_k^t=\mathbf{u}_k^a} & \left.\frac{\partial \boldsymbol{\theta}_k^t}{\partial [\boldsymbol{\theta}_k^t]^T}\right|_{\mathbf{u}_k^t=\mathbf{u}_k^a} \end{pmatrix} \\ &= \begin{pmatrix} \Psi(\boldsymbol{\theta}_k^a) & \left.\frac{\partial \Psi(\boldsymbol{\theta}_k^t)}{\partial [\boldsymbol{\theta}_k^t]^T}\right|_{\boldsymbol{\theta}_k^t=\boldsymbol{\theta}_k^a} \mathbf{w}_k^a \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \end{aligned} \tag{6.39}$$

Then, the forecast step of the extended Kalman filter becomes

$$\begin{pmatrix} \mathbf{w}_k^f \\ \boldsymbol{\theta}_k^f \end{pmatrix} = \begin{pmatrix} \Psi(\boldsymbol{\theta}_{k-1}^a)\mathbf{w}_{k-1}^a \\ \boldsymbol{\theta}_{k-1}^a \end{pmatrix} \tag{6.40a}$$

$$\begin{aligned} \mathbf{P}_k^f &= \begin{pmatrix} \Psi(\boldsymbol{\theta}_{k-1}^a) & \left.\frac{\partial \Psi(\boldsymbol{\theta}_k^t)}{\partial [\boldsymbol{\theta}_k^t]^T}\mathbf{w}_{k-1}^a\right|_{\boldsymbol{\theta}_k^t=\boldsymbol{\theta}_{k-1}^a} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \mathbf{P}_{k-1}^a \\ &\quad \times \begin{pmatrix} \Psi(\boldsymbol{\theta}_{k-1}^a) & \left.\frac{\partial \Psi(\boldsymbol{\theta}_k^t)}{\partial [\boldsymbol{\theta}_k^t]^T}\mathbf{w}_{k-1}^a\right|_{\boldsymbol{\theta}_k^t=\boldsymbol{\theta}_{k-1}^a} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}^T \\ &\quad + \begin{pmatrix} \mathbf{Q}_{k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{k-1} \end{pmatrix} \end{aligned} \tag{6.40b}$$

and analysis step becomes

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_k^f \begin{pmatrix} \mathbf{H} & \mathbf{0} \end{pmatrix}^T \\ &\quad \left[\begin{pmatrix} \mathbf{H} & \mathbf{0} \end{pmatrix} \mathbf{P}_k^f \begin{pmatrix} \mathbf{H} & \mathbf{0} \end{pmatrix}^T + \mathbf{R}_k\right]^{-1} \end{aligned} \tag{6.41a}$$

$$\mathbf{P}_k^a = \Big[\mathbf{I} - \mathbf{K}_k \begin{pmatrix} \mathbf{H} & \mathbf{0} \end{pmatrix}\Big] \mathbf{P}_k^f \tag{6.41b}$$

$$\begin{pmatrix} \mathbf{w}_k^a \\ \boldsymbol{\theta}_k^a \end{pmatrix} = \begin{pmatrix} \mathbf{w}_k^f \\ \boldsymbol{\theta}_k^f \end{pmatrix} + \mathbf{K}_k \Big[\mathbf{w}_k^o - \mathbf{H}\mathbf{w}_k^f\Big] \tag{6.41c}$$

A few comments are pertinent:

- It is interesting to mention that this type of application of the extended Kalman filter converts the filter into an *adaptive* filter. This means that, at each time step the filter described above improves upon the knowledge of the parameter vector $\boldsymbol{\theta}$. In other words, the system "learns" about itself.

- The technique used here to construct the extended Kalman filter for the parameter estimation problem, that is, that of incorporating the parameter vector into the state vector, is known as *state augmentation* technique. This nomenclature is relatively evident, since in the case studied above, the augmented state $\mathbf{u}_k^t$ contains the system state vector $\mathbf{w}_k^t$ as well as the vector of parameters $\boldsymbol{\theta}_k^t$. State augmentation is a very common and powerful technique in estimation theory. Examples in which this technique is used are in problems of smoothing (e.g., Anderson & Moore [1]); colored noises, that is, noises that are not white (Anderson & Moore [1]); more general parameter estimation problems, as that of estimating parameters related to the noise statistics. Application of these ideas to atmospheric and oceanic data assimilation are those of Hao [71] and Hao & Ghil [72].

- The problem of parameter estimation belongs to a wider class of problems more commonly referred to as *system identification* (e.g., Sage & Melsa [120]). Many alternative methods, which do not use concepts related to the Kalman filter can be found in the literature; some identification methods are based on statistics, but not all (see comments in Gelb [60], pp. 350).

Exercises

1. Following the procedure indicated in Section 6.2, derive equation (6.5) for the forecast (prediction) error covariance evolution.

2. Following the procedure indicated in Section 6.2, derive expression (6.28) for the analysis error covariancia.

3. Consider the following scalar system:

$$\begin{aligned} w_k^t &= \theta w_{k-1}^t \\ w_k^o &= w^t + b^o \end{aligned}$$

where $b_k^o \sim \mathcal{N}(0, \sigma^2)$, and θ is an unknown parameter, with an initial estimate of $\hat{\theta}_0$. Show that if the parameter θ is modeled as a deterministic quantity the state augmentation technique, together with the Kalman filter, give no further information on the unknown parameter θ, as the filter gets iterated in time. In other words, show that in this case, the extended Kalman filter can be partitioned as

$$\mathbf{K}_k = \begin{pmatrix} g_k \\ 0 \end{pmatrix}$$

and explain why this partition implies nothing is learned about the parameter θ, by the filtering procedure. This is an example of a problem known as identifiability, i.e., θ is non–identifiable.

4. *Computer Assignment.* Consider again the Lorenz (1960) [1] model of Exercise 3.7. We want to implement an assimilation system based on the extended Kalman filter, and a simple modification of it, for this model. Because this is just a simulation, we have to "define the true model". We take for that the exact same model, that is,

$$\frac{d\mathbf{w}^t}{dt} = \mathbf{f}(\mathbf{w}^t)$$

where $\mathbf{f}$ is the Lorenz model of Exercise 3.7, but we choose a different initial condition that is taken from a realization of $\mathbf{w}_0^t = \bar{\mathbf{w}}_0 + \mathbf{b}_0^t$, where

$$\bar{\mathbf{w}}_0 = \begin{pmatrix} 0.12 \\ 0.24 \\ 0.10 \end{pmatrix}$$

and $\mathbf{b}_0^t \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_0^a)$, with $\mathbf{P}_0^a = (\sigma_0^a)^2\mathbf{I}$. Moreover, we make the perfect model assumption by saying that $\mathbf{b}_k^t = \mathbf{0}$, for all $k = 1, 2, \ldots$. Consequently, $\mathbf{Q}_k = 0$, for all $k = 1, 2, \ldots$.

Let us choose the initial standard deviation error $\sigma_0^a = 0.1$, which is ten times larger than the value we used in our Monte Carlo experiments before. To facilitate your evaluation of different results to be obtained below, fix a seed, in the very beginning of the code, for the random number generator of Matlab.

All experiments that follow are to be performed in the time interval from $t_0 = 0$ to $t_f = 250$, and time step $\Delta t = 0.5$, of Exercise 3.7. With the choice of initial error given above:

True state and approximate mean state evolution: Plot the evolution, of all three variables, of the true state and those produced by a prediction model based on the mean equation, that is,

$$\frac{d\boldsymbol{\mu}}{dt} = \mathbf{f}(\boldsymbol{\mu})$$

where $\mathbf{f}$ is given by the Lorenz model in Exercise 3.7, for all three variables. Take for the initial mean state the value $\boldsymbol{\mu}(0) = \bar{\mathbf{w}}_0$, given above. Does the "predicted" state have any resemblance with the true state?

To construct an assimilation system we need an observation process, which for this problem is taken to be simply

$$\mathbf{w}_k^o = \mathbf{w}_k^t + \mathbf{b}_k^o$$

[1] Lorenz, E.N., 1960: Maximum simplification of dynamical equations. *Tellus*, **12**, 243–254.

for $k = 1, 2, \ldots$. That is, the observation process is linear, with all three variables of the model being observed under noise $\mathbf{b}_k^o \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, with $\mathbf{R} = (\sigma^o)^2\mathbf{I}$.

The extended Kalman filter: In all cases below plot the true state evolution against the estimate evolution. Also, separately, plot the evolution of the variances for all three variables.

(a) *Low frequency update.* Taking the observation noise level to be $\sigma^o = 0.2$, and the observation interval to be $\Delta t_{obs} = 50$ time units, insert the appropriate equations in your Matlab code to perform the analysis step of the extended Kalman filter. Notice that, between two consecutive observations your program should evolve the mean and (co)variance just as it did in Exercise 3.7. Observe also, that in the extended Kalman filter the mean equation does not include the bias correction term involving the Hessian of the dynamical model. Plot the evolution of the mean on the same frame as that for the true state, for all three variables of the model. Does assimilation improve the prediction you had in the previous item? What do the variance plots tell you?

(b) *More frequent observations.* Reduce the assimilation (observation) interval in the experiment of the previous item to half of what it was. How do the estimates chance? What happens if the observation interval is reduces further to $\Delta t_{obs} = 10$ time units?

(c) *More accurate observations.* The observations considered above are quite lousy — the observation error level is about 100% of the value of the amplitude of the variables of the system — a more sensible observation error level would be considerably smaller. In this way, taking $\sigma^o = 0.05$, repeat the filtering experiment of item (a). Comment on how this changes the estimate, and what the variance plots tell you.

The bias correction term: We saw in the experiments in Exercise 3.7 that the bias correction term can have a considerable influence on the evolution of the mean and (co)variance. In particular, its presence may allow for error variance saturation, avoiding indefinite growth of error. Here, we want to examine the effect of this term in the context of assimilation (filtering). The inclusion of the bias correction term provides a second order filter, which is in principle more accurate than the extended Kalman filter. Repeat items (a) and (b) above when the bias correction term is included in the equation for the evolution of the mean. Compare the results with those found previously in (a) and (b).

Now that you have constructed a small data assimilation system, you might what to change your dynamical model to be a more interesting one, such as the Lorenz (1963)[2] chaotic model. You can use as a guide for some experiments the work of Miller et al. (1994)[3], where the extended Kalman filter was first applied to that model. Have fun.

[2] Lorenz, E.N., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.*, **20**, 130–141.

[3] Miller, R.N., M. Ghil, & F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037–1056.

Chapter 7

Basic Concepts of Atmospheric Dynamics

In this lecture the basic equations that govern atmospheric dynamics are introduced. The main goal here is to outline the necessary concepts for a better understanding of the problem of atmospheric data assimilation to be treated in the following lectures. Much of the content of this lecture can be found in meteorology text books such as: Daley [39], Ghil & Childress [62], Haltiner & Williams [70], Holton [82], and Pedlosky [114].

In Section 7.1, we introduce the governing equations. In Section 7.2, we make a scale analysis of the governing equations for synoptic scale problems, which leads us to introduce the notions of hydrostatic and geostrophic approximations, which are discussed in Section 7.3. Notions on vertical stratification notion are introduced in Section 7.4. In Section 7.5 we solve the equations of motion for the simple in which the atmosphere is approximated by the linearized shallow water equations. Finally in Section 7.6, we discretize the shallow–water equations using a relatively simple finite difference scheme.

7.1 Governing Equations

(I) Momentum Equation:

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\frac{1}{\rho}\nabla p + \mathbf{g} + \mathbf{f} \tag{7.1}$$

where $\mathbf{v}$ is the velocity vector of the atmospheric "fluid" in three dimensions, in a rotating frame of reference; $\boldsymbol{\Omega}$ is the three–dimensional angular velocity vector (velocity with which the rotating frame of reference moves); ρ is the density of the atmospheric "fluid"; p is its pressure; $\mathbf{g}$ is the gravitational acceleration vector in three dimensions; $\mathbf{f}$ represents the three–dimensional friction force (e.g. between the atmosphere and the earth surface); and ∇ is the gradient vector in three spatial dimensions.

(ii) Continuity Equation:

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0 \tag{7.2}$$

This equation means that the rate of change of the local density is equal to the negative of the (mass) density divergence. It is common to re–write this equation as:

$$\frac{1}{\alpha}\frac{D\alpha}{Dt} + \nabla\cdot\mathbf{v} = 0\,, \tag{7.3}$$

where $\alpha = 1/\rho$ is the specific volume, and the operator D/Dt is formally defined as:

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\,. \tag{7.4}$$

(iii) First Law of Thermodynamics:

This law states that the change in internal energy of the system is equal to the difference between the heat added to the system and the work done by the system. For the atmosphere, the first law of thermodynamics translates into:

$$\frac{De}{Dt} = -p\frac{D\alpha}{Dt} + Q\,, \tag{7.5}$$

where e is the specific internal energy, which is only a function of the temperature T of the system, and Q is the heat per unit of mass. It is worth noticing that the temperature is a function of the space variables, as well as of time.

Introducing the specific heat at constant volume for the dry air $c_v \equiv e/T$, we have that

$$c_v\frac{DT}{Dt} = -p\frac{D\alpha}{Dt} + Q\,. \tag{7.6}$$

where T is the temperature of the system.

In this lecture, we refer very little to the thermodynamics equation and therefore the equations above are sufficient.

7.2 Scales of the Equations of Motion

In spherical coordinates (λ, φ, z) the three components of the momentum equation (Newton's equations) can be written as (e.g., Washington & Parkinson [138]):

$$\frac{du}{dt} - \frac{uv\tan\varphi}{r} + \frac{uw}{r} = -\frac{1}{\rho r\cos\varphi}\frac{\partial p}{\partial\lambda} + fv - \hat{f}w + f_\lambda \tag{7.7}$$

$$\frac{dv}{dt} - \frac{u^2\tan\varphi}{r} + \frac{vw}{r} = -\frac{1}{\rho r}\frac{\partial p}{\partial\varphi} - fu + f_\varphi \tag{7.8}$$

$$\frac{dw}{dt} - \frac{u^2+v^2}{r} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g + \hat{f}u + f_z \tag{7.9}$$

Table 7.1: Definition of the scale parameters

$U \sim 10$ m/s	horizontal velocities
$W \sim 1$ cm/s	vertical velocities
$L \sim 10^6$ m	length (horizontal; $1/2\pi$ wave length)
$H \sim 10^4$ m	depth (vertical)
$\Delta P/\rho \sim 10^3$ m^2/s^2	horizontal pressure fluctuations
$L/U \sim 10^5$ s	time

where we use the definitions $\mathbf{v} \equiv (u, v, w)^T$ and $\mathbf{f} \equiv (f_\lambda, f_\varphi, f_z)^T$, and we notice that

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \tag{7.10}$$

$$= \frac{\partial}{\partial t} + \frac{u}{r\cos\varphi}\frac{\partial}{\partial \lambda} + \frac{v}{r}\frac{\partial}{\partial \varphi} + w\frac{\partial}{\partial z}. \tag{7.11}$$

The parameters f and $\hat{f}$ are defined as:

$$f = 2\Omega \sin\varphi \tag{7.12}$$

$$\hat{f} = 2\Omega \cos\varphi \tag{7.13}$$

where f is known as the Coriolis parameter, and Ω is the magnitude of the vector $\mathbf{\Omega}$. Moreover, $r = a + z$, with a representing the radius of the earth and z the height, starting from the surface.

In this lecture, we are interested in synoptic scale dynamics, and therefore we introduce scale variables that refer to synoptic atmospheric systems as in Table 7.2 (see Holton 1979, p. 36 for more details). In particular, notice that the time scale is on the order of days; this scale is called advective time scale, where pressure systems move approximately with the horizontal winds.

Disregarding the friction force $\mathbf{f}$ from this point on, we can proceed with the scale analysis of the equations (7.7)–(7.9), noticing that an estimate of the scale of the Coriolis parameter can be obtained for the mid-latitude $\varphi = \varphi_0 = 45^o$ as:

$$[f] = [\hat{f}] = f_0 = 2\Omega \sin\varphi_0 = 2\Omega\cos\varphi_0 \tag{7.14}$$

$$= 2\left(\frac{2\pi}{86164}\right)\cos(45^o) \sim 10^{-4}, \tag{7.15}$$

where the notation [.] is used to indicate the scale of the quantity between the curly brackets.

The requirement of synoptic dynamics imposes a restriction in the horizontal direction. To define scales in the vertical direction it is necessary to establish at what height we are interested in describing the atmosphere. For tropospheric dynamics, the pressure gradient can be represented by the scale defined by P_0/H, where P_0 ($\sim$ 1000 mb = 1 atm) is the pressure at the surface and H is the troposphere depth introduced above.

Table 7.2 shows the results of the scale analysis, where the magnitude of each term in the equations (7.7)–(7.9) is indicated. We see directly from the table that the horizontal and

Table 7.2: Scale analysis of the components of the momentum equation

Horizontal scale analysis						
zonal	$[\frac{du}{dt}]$	$[fv]$	$[\hat{f}w]$	$[\frac{uw}{r}]$	$[\frac{uv\tan\varphi}{r}]$	$[\frac{1}{\rho r\cos\varphi}\frac{\partial p}{\partial\lambda}]$
meridional	$[\frac{dv}{dt}]$	$[fu]$		$[\frac{vw}{r}]$	$[\frac{u^2\tan\varphi}{r}]$	$[\frac{1}{\rho r}\frac{\partial p}{\partial\varphi}]$
scales	$\frac{U^2}{L}$	f_0U	f_0U	$\frac{UW}{[r]}$	$\frac{U^2}{[r]}$	$\frac{\Delta P}{\rho L}$
magnitude ($\mathrm{m/s^2}$)	10^{-4}	10^{-3}	10^{-6}	10^{-8}	10^{-5}	10^{-3}
Vertical scale analysis						
vertical	$[\frac{dw}{dt}]$	$[\hat{f}u]$			$[\frac{u^2+v^2}{r}]$	$[\frac{1}{\rho}\frac{\partial p}{\partial z} - g]$
scales	$\frac{UW}{L}$	f_0U			$\frac{U^2}{[r]}$	$\frac{P_0}{\rho H} + [g]$
magnitude ($\mathrm{m/s^2}$)	10^{-7}	10^{-3}			10^{-5}	10

vertical scales are independent. This fact is exactly what allows us to distinguish between horizontal and vertical motion as approximately separate entities.

7.3 Geostrophic and Hydrostatic Approximations

The scale analysis of the momentum equations in the horizontal direction shows that synoptic dynamics are dominated by the Coriolis term and by the pressure gradient term. In this way, to first order the horizontal equations can be approximated by

$$fv \approx \frac{1}{\rho}\frac{\partial p}{\partial\lambda} \qquad (7.16)$$

$$-fu \approx \frac{1}{\rho}\frac{\partial p}{\partial\varphi} \qquad (7.17)$$

This approximation motivates us to define the so–called geostrophic winds as those satisfying exactly the relation:

$$\mathbf{v}_g \equiv \mathbf{k} \times \frac{1}{f\rho}\nabla p \qquad (7.18)$$

where **k** is the unit vector in vertical direction.

The table 7.2 also indicates that a reasonable simplification of the vertical component of the momentum equation is

$$\frac{1}{\rho}\frac{\partial p}{\partial z} \approx -g\,, \qquad (7.19)$$

meaning that the pressure field is nearly in hydrostatic balance. In other words, the pressure at a point is approximately equal to the weight of the air column above the point. In the same way that we were motivated to introduce geostrophic winds, we can define a standard pressure $\bar{p}$ as the one that follows exactly the hydrostatic relation:

$$\frac{d\bar{p}}{dz} = \bar{\rho}g \tag{7.20}$$

where $\bar{\rho}$ is a standard density. Notice that, by simplifying the vertical component of the momentum equation the vertical winds disappear. This means that at synoptic scales these winds are negligible.

7.4 Vertical Stratification

Let us examine the hydrostatic approximation introduced in the previous section in more detail. Since $g\rho > 0$, the pressure p monotonically decreases with the height z. Moreover, within the tropospheric layer $g \approx$ *const.*, which means that given a density function $\rho = \rho(z, p)$, the hydrostatic approximation

$$\frac{dp}{dz} = -g\rho\,, \tag{7.21}$$

when satisfied exactly, provides a model for the vertical atmosphere.

A simple atmospheric model, one called homogeneous, is that for which the density $\rho = \bar{\rho}$ is constant (independent of height and pressure). In this case,

$$p = \bar{p} - g\bar{\rho}(z - \bar{z})\,, \tag{7.22}$$

where the quantities with a bar are standard quantities, defined generally at sea level.

A more realistic model, not homogeneous, is found when we consider the atmosphere as an ideal gas. In this case, the pressure and density are related by the ideal gas law.

$$\frac{p}{\rho} = RT\,, \tag{7.23}$$

where T is the temperature and R is the gas constant for the dry air.

In this way, the hydrostatic balance can be written as:

$$\frac{dp}{p} = -\frac{g}{R}\frac{dz}{T_0 - \Gamma z} \tag{7.24}$$

where we use the fact that in the troposphere the rate of temperature decrease is approximately constant: $dT/dz = -\Gamma$, for Γ being the lapse rate, and T_0 the temperature of an isothermal atmosphere.

One of the conventional ways of taking measurements of the atmosphere is by means of balloons. They usually measure the temperature, pressure and wind. That is, the temperature and wind are functions of pressure, in particular $T = T(p)$. The hydrostatic relation, written as:

$$\frac{dz}{dp} = -\frac{R}{g}\frac{T(p)}{p} \tag{7.25}$$

can be used to obtain the temperature, pressure and density profiles as functions of the height. This information can then be used in the solution of the governing equations.

In fact, we can simplify this transformation procedure by introducing pressure as the vertical coordinate, instead of the height z. By defining the geopotential function $\phi \equiv gz$, the hydrostatic equation becomes:

$$\frac{d\phi}{dp} = -\frac{RT}{p}. \tag{7.26}$$

The governing equations can be written using pressure as the vertical coordinate (e.g., Haltiner & Williams [70], Section 1.9).

7.5 Linearized Shallow–Water Equations

The system of shallow–water equations, in cartesian coordinates, can be written as:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + g\frac{\partial h}{\partial x} = 0 \tag{7.27a}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + g\frac{\partial h}{\partial y} = 0 \tag{7.27b}$$

$$\frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y} + h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0 \tag{7.27c}$$

where x and y indicate the zonal and meridional directions, respectively, and we consider the Coriolis parameter $f = f_0$ to be constant. The boundary conditions that interest us at this moment are periodic in both directions and the extent of the domain is taken as $2\pi a$, where a is the radius of the earth.

A simple linearization that we can use for the system above, with relevant meaning, is when the reference state (or basic state) consists of a null winds, i.e., state of rest, and of a free surface height, i.e., independent of space and time. That means, the basic state is defined as:

$$u = 0 + u' \tag{7.28a}$$

$$v = 0 + v' \tag{7.28b}$$

$$h = H + h' \tag{7.28c}$$

where $H = const.$ e $(.)'$ is used to indicate perturbations. In this case, the equations (7.27a)–(7.27c) are reduced to the equations

$$\frac{\partial u}{\partial t} - f_0 v + \frac{\partial \phi}{\partial x} = 0 \tag{7.29a}$$

$$\frac{\partial v}{\partial t} + f_0 u + \frac{\partial \phi}{\partial y} = 0 \tag{7.29b}$$

$$\frac{\partial \phi}{\partial t} + \Phi(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) = 0 \tag{7.29c}$$

where we eliminate the notation $(\cdot)'$ so that u, v and ϕ in the system of equations above refer to perturbations; moreover, we introduce the basic geopotential height $\Phi = gH$ and its correspondent perturbation $\phi = gh$.

One of the ways to solve the equations above is to introduce the stream function ψ and the potential velocity χ by means of the Helmholtz theorem:

$$u = -\frac{\partial \psi}{\partial y} + \frac{\partial \chi}{\partial x} \tag{7.30a}$$

$$v = \frac{\partial \psi}{\partial x} + \frac{\partial \chi}{\partial y} \tag{7.30b}$$

from where it follows that

$$\nabla^2 \psi = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{7.31a}$$

$$\nabla^2 \chi = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \tag{7.31b}$$

Then, the equations (7.27a) and (7.27b) can be transformed into equations for the relative vorticity and divergence:

$$\frac{\partial \nabla^2 \psi}{\partial t} + f_0 \nabla^2 \chi = 0 \tag{7.32a}$$

$$\frac{\partial \nabla^2 \chi}{\partial t} - f_0 \nabla^2 \psi + \nabla^2 \phi = 0, \tag{7.32b}$$

where $\nabla^2\psi$ is the vertical component of relative vorticity, and $\nabla^2\chi$ is the divergence (see Holton [82] pp. 73, 83–86 for more details). Using (7.30a) and (7.30b) we can re–write the equation for the perturbation geopotential height as:

$$\frac{\partial \phi}{\partial t} + \Phi \nabla^2 \chi = 0 \tag{7.33}$$

The expressions (7.32)–(7.33) form a system of coupled, constant–coefficient, linear partial differential equations, which can be solved by normal mode expansion (i.e., Fourier series).

In this way we write:

$$\begin{pmatrix} \psi(x,y,t) \\ \chi(x,y,t) \\ \phi(x,y,t) \end{pmatrix} = \begin{pmatrix} \hat{\psi}(t) \\ i\hat{\chi}(t) \\ f_0\sqrt{k}\hat{\phi}(t) \end{pmatrix} \exp\left\{ i \left[\frac{(mx + ny)}{a} \right] \right\} \tag{7.34}$$

where m is the zonal wave number, n is the meridional wave number, $i = \sqrt{-1}$, and the constant k is given by

$$k = \frac{(m^2 + n^2)\Phi}{a^2 f_0^2}. \tag{7.35}$$

Then, we see that

$$\nabla^2 \psi = -\frac{f_0^2}{\Phi} k \psi \tag{7.36a}$$

$$\nabla^2 \chi = -\frac{f_0^2}{\Phi} k \chi \tag{7.36b}$$

$$\nabla^2 \phi = -\frac{f_0^2}{\Phi} k \phi \tag{7.36c}$$

Therefore, substituting (7.34) and (7.36) in equations (7.32)–(7.33) we obtain:

$$\frac{d\hat{\psi}}{dt} + if_0\hat{\chi} = 0 \tag{7.37a}$$

$$i\frac{d\hat{\chi}}{dt} - f_0\hat{\psi} + f_0\sqrt{k}\hat{\phi} = 0 \tag{7.37b}$$

$$f_0\sqrt{k}\frac{d\hat{\phi}}{dt} - if_0^2k\hat{\chi} = 0 \tag{7.37c}$$

These equations can be written in compact form,

$$\frac{d\hat{\mathbf{w}}(t)}{dt} = -if_0\hat{\mathbf{L}}\hat{\mathbf{w}}(t)\,, \tag{7.38}$$

where the vector $\hat{\mathbf{w}} \equiv (\hat{\psi}, \hat{\chi}, \hat{\phi})^T$, and the matrix $\hat{\mathbf{L}}$ is given by

$$\hat{\mathbf{L}} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -\sqrt{k} \\ 0 & -\sqrt{k} & 0 \end{pmatrix}. \tag{7.39}$$

The solution of equation (7.38) is

$$\hat{\mathbf{w}}(t) = e^{-if_0\hat{\mathbf{L}}t}\hat{\mathbf{w}}(0)\,, \tag{7.40}$$

where $\hat{\mathbf{w}}(0)$ represents the initial condition vector. This expression can be written in a more convenient way if we expand the vector $\hat{\mathbf{w}}(0)$ in terms of the eigenvectors of $\hat{\mathbf{L}}$. These eigenvectors can be determined by solving the equation:

$$(\hat{\mathbf{L}} - \sigma_\ell\mathbf{I})\hat{\mathbf{v}}_\ell = \mathbf{0} \tag{7.41}$$

where σ_ℓ refers to the eigenvalue corresponding to the eigenvector $\hat{\mathbf{v}}_\ell$, and $\mathbf{I}$ represents the 3×3 identity matrix. Notice that writing the solution as in (7.34) produces a matrix $\hat{\mathbf{L}}$ which is real and symmetric.

It is simple to show that the eigenvalues of the matrix $\hat{\mathbf{L}}$ are determined by solving the characteristic equation,

$$\sigma_\ell^3 - (1 + k)\sigma_\ell = 0 \tag{7.42}$$

whose solutions can be written as: $\sigma_\ell \in \{\sigma_G^- = -\sqrt{1+k}, \sigma_R = 0, \sigma_G^+ = +\sqrt{1+k}\}$, and the subscripts R and G indicate frequencies of rational and gravity waves, respectively. The eigenvectors corresponding to the eigenvalues above can be obtained by substituting each value of σ in (7.41). In this way, we can build a matrix $\hat{\mathbf{V}}$ whose columns correspond to the eigenvectors $\hat{\mathbf{v}}_\ell$ of the problem of $\hat{\mathbf{L}}$. That is,

$$\hat{\mathbf{V}} = \frac{1}{\sqrt{2(1+k)}}\begin{pmatrix} 1 & \sqrt{2k} & 1 \\ -\sqrt{1+k} & 0 & \sqrt{1+k} \\ -\sqrt{k} & \sqrt{2} & -\sqrt{k} \end{pmatrix}, \tag{7.43}$$

where the first and third columns of $\hat{\mathbf{V}}$ correspond to the eigenvalues $\sigma_G^\pm$ and the middle column corresponds to the eigenvalue σ_R. It is easy to verify that the column vectors form a complete orthonormal set of eigenvectors and therefore the matrix $\hat{\mathbf{V}}$ is unitary: $\hat{\mathbf{V}}^T = \hat{\mathbf{V}}^{-1}$. Moreover, the matrix $\hat{\mathbf{V}}$ is the one that diagonalizes the matrix $\hat{\mathbf{L}}$:

$$\hat{\mathbf{V}}^{-1}\hat{\mathbf{L}}\hat{\mathbf{V}} = \mathbf{\Lambda}\,, \tag{7.44}$$

where Λ is a diagonal matrix whose elements are the respective eigenvalues: $\mathbf{\Lambda} = diag(\sigma_G^-, \sigma_R, \sigma_G^+)$.

Returning to the solution, (7.40), we write the expansion of the initial vector by utilizing the eigenvectors of $\hat{\mathbf{L}}$ as

$$\hat{\mathbf{w}}(0) = \sum_\ell \hat{c}_\ell \hat{\mathbf{v}}_\ell \tag{7.45}$$

and noticing that the eigenvectors of $e^{-if_0\hat{\mathbf{L}}t}$ are the same as those of $\hat{\mathbf{L}}$, with eigenvalues $e^{i\sigma_\ell t}$, we have

$$\hat{\mathbf{w}}(t) = \sum_\ell \hat{c}_\ell \hat{\mathbf{v}}_\ell e^{i\sigma_\ell t}\,. \tag{7.46}$$

Once the initial condition is known $\hat{\mathbf{w}}(0) = (\psi(0), \xi(0), \phi(0))^T$, the expression (7.45) can be inverted to obtain the expansion coefficients $\hat{c}_\ell$. Therefore, using the fact that the matrix $\hat{\mathbf{V}}$ is unitary we have

$$\hat{\mathbf{c}} = \hat{\mathbf{V}}^T\hat{\mathbf{w}}(0) \tag{7.47}$$

where we define $\hat{\mathbf{c}} \equiv (\hat{c}_-, \hat{c}_0, \hat{c}_+)^T$.

7.6 Numerical Solution: A Finite Difference Method

In general, there are no analytic solutions for the system of governing equations, including the thermodynamic processes. Therefore, these equations are solved numerically in some way (e.g., Haltiner & Williams [70]) with computer assistance. In this section we will exemplify a practical way of solving the governing equations by means of applying a finite difference scheme to a simple set of equations.

Consider the system of two-dimensional shallow-water equations, on a β-plane, linearized about a basic state with zero meridional wind $v \equiv 0$, and constant zonal wind $u = U$ (see Exercise 6.2, for a guide to the derivation of these equations):

$$\frac{\partial u}{\partial t} + U\frac{\partial u}{\partial x} + \frac{\partial \phi}{\partial x} - (f - U_y)v \;=\; 0\,, \tag{7.48a}$$

$$\frac{\partial v}{\partial t} + U\frac{\partial v}{\partial x} + \frac{\partial \phi}{\partial y} + fu \;=\; 0\,, \tag{7.48b}$$

$$\frac{\partial \phi}{\partial t} + U\frac{\partial \phi}{\partial x} + \Phi(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) + \Phi_y v \;=\; 0\,, \tag{7.48c}$$

where u, v are the perturbations in the velocity field, ϕ is the perturbation in the geopotential field, $f = f_0 + \beta y$ is the Coriolis parameter, and the basic state satisfies:

$$fU + \frac{d\Phi}{dy} = 0\,, \tag{7.49}$$

and the equations are applied to a doubly periodic domain.

The system of equations (7.48) can be written in vector form as

$$\frac{\partial \mathbf{w}}{\partial t} + \frac{\partial}{\partial x}(\mathbf{A}\,\mathbf{w}) + \frac{\partial}{\partial y}(\mathbf{B}\,\mathbf{w}) + \mathbf{C}\,\mathbf{w} = \mathbf{0}\,, \tag{7.50}$$

where $\mathbf{w} = (u, v, \phi)^T$, as in the previous section, and the matrices $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ are given by

$$\mathbf{A} = \begin{pmatrix} U & 0 & 1 \\ 0 & U & 0 \\ \Phi & 0 & U \end{pmatrix}, \tag{7.51a}$$

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & \Phi & 0 \end{pmatrix}, \tag{7.51b}$$

$$\mathbf{C} = \begin{pmatrix} 0 & -f & 0 \\ f & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{7.51c}$$

Notice that these matrices depend on the variable y, since the geopotential function of the basic state Φ and the Coriolis parameter f are functions of the latitude.

Let us apply the Richtmyer two step–version of the Lax–Wendroff finite difference scheme (see Richtmyer & Morton [117]; Ghil et al. [66]; and Parrish & Cohn [113]). For that we consider the 3–vector $\mathbf{w}(x, y, t)$ to be in a two–dimensional uniform grid $I \times J$ whose approximate value at a point is given by

$$\mathbf{w}(x_i, y_j, t_k) = \mathbf{w}_{ij}^k \approx \mathbf{w}[(i-1)\Delta x, (j-1)\Delta y, k\Delta t]\,, \tag{7.52}$$

with $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 0, 1, \ldots$, and $\Delta x = L_x/I$, $\Delta y = L_y/(J-1)$, for L_x, L_y representing the extension of the rectangular domain in the zonal and meridional directions, respectively.

The Richtmyer version of the Lax–Wendroff scheme has the form of a predictor–corrector scheme for which the first step, the predictor, can be written as:

$$\mathbf{w}_{i+1/2,j+1/2}^{k+1/2} = \mu_x \mu_y \mathbf{w}_{i+1/2,j+1/2}^{k} - \frac{1}{2}\mathbf{L}_{j+1/2}\mathbf{w}_{i+1/2,j+1/2}^{k} \tag{7.53}$$

for $i = 1, 2, \ldots, I$ e $j = 1, 2, \ldots, J-1$, where the operator $\mathbf{L}_j$ is defined by

$$\mathbf{L}_j \equiv \lambda_x \mu_y \delta_x \mathbf{A}_j + \lambda_y \mu_x \delta_y \mathbf{B}_j + \Delta t \mu_x \mu_y \mathbf{C}_j, \tag{7.54}$$

with $\lambda_x \equiv \Delta t/\Delta x$, $\lambda_y \equiv \Delta t/\Delta y$, and we notice that $\mathbf{C}$ does not depend on y. The second order operators of spatial mean and difference are defined as:

$$\delta_x \mathbf{w}_{ij} \equiv \mathbf{w}_{i+1/2,j} - \mathbf{w}_{i-1/2,j}\,, \tag{7.55a}$$

$$\mu_x \mathbf{w}_{ij} \equiv \frac{1}{2}(\mathbf{w}_{i+1/2,j} + \mathbf{w}_{i-1/2,j})\,, \tag{7.55b}$$

and analogously for δ_y and μ_y:

$$\delta_y \mathbf{w}_{ij} \equiv \mathbf{w}_{i,j+1/2} - \mathbf{w}_{i,j-1/2}\,, \tag{7.56a}$$

$$\mu_y \mathbf{w}_{ij} \equiv \frac{1}{2}(\mathbf{w}_{i,j+1/2} + \mathbf{w}_{i,j-1/2})\,. \tag{7.56b}$$

The second step of this finite difference scheme, the corrector step, serves to propagate the state from the half-grid intermediate points to the full-grid points, that is,

$$\mathbf{w}_{ij}^{k+1} = \mathbf{w}_{ij}^{k} - \mathbf{L}_j \mathbf{w}_{ij}^{k+1/2} \tag{7.57}$$

for $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$.

The periodic boundary conditions in the East-West direction as well as in the North-South direction can be taken into consideration by doing:

$$\mathbf{w}_{i,J+1} = \mathbf{w}_{i,1} \tag{7.58a}$$

$$\mathbf{w}_{i,0} = \mathbf{w}_{i,J} \tag{7.58b}$$

for $i = 1, 2, \cdots, I$, and

$$\mathbf{w}_{I+1,j} = \mathbf{w}_{1,j} \tag{7.59a}$$

$$\mathbf{w}_{0,j} = \mathbf{w}_{I,j} \tag{7.59b}$$

for $j = 1, 2, \cdots, J$.

Although it is not necessary, and in general cannot be done in implementation of numeric methods for practical problems in meteorology, we can combine the expressions (7.53) and (7.57) to write the finite difference system of equations in the following, more compact, form

$$\mathbf{w}^{k+1} = \mathbf{\Psi} \mathbf{w}^{k}, \tag{7.60}$$

where $\mathbf{\Psi}$ is the transition matrix of the system, also called the dynamics matrix. By writing the system of equations in this form it becomes easy to understand the connection between the problems studied in the previous lectures and the problem of assimilation of meteorological data to be studied below.

In any event, we can illustrate the morphology of the transition matrix by considering an idealized grid with resolution 4×5. The two stages (7.53) and (7.57) of the finite difference scheme can be combined as

$$\begin{aligned} \mathbf{w}_{ij}^{k+1} \;=\; & \mathbf{Q}_j^1 \mathbf{w}_{i-1,j-1}^{k} + \mathbf{Q}_j^2 \mathbf{w}_{i,j-1}^{k} + \mathbf{Q}_j^3 \mathbf{w}_{i+1,j-1}^{k} + \\ & \mathbf{Q}_j^4 \mathbf{w}_{i-1,j}^{k} \;\; + \mathbf{Q}_j^5 \mathbf{w}_{i,j}^{k} \;\; + \mathbf{Q}_j^6 \mathbf{w}_{i+1,j}^{k} \;\; + \\ & \mathbf{Q}_j^7 \mathbf{w}_{i-1,j+1}^{k} + \mathbf{Q}_j^8 \mathbf{w}_{i,j+1}^{k} + \mathbf{Q}_j^9 \mathbf{w}_{i+1,j+1}^{k}\,, \end{aligned} \tag{7.61}$$

where the matrices $\mathbf{Q}$ have dimension 3×3, and consist of linear combinations of the matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$, calculated at a specific grid points. Explicit form for the auxiliary matrices $\mathbf{Q}$ can be found in Parrish & Cohn [113], with an appropriate modification due to different boundary conditions. A simplified version of (7.61) is treated here in the exercises.

$$\mathbf{\Psi} = \begin{pmatrix}
\mathbf{Q}_1^5 & \mathbf{Q}_1^6 & 0 & \mathbf{Q}_1^4 & \mathbf{Q}_1^8 & \mathbf{Q}_1^9 & 0 & \mathbf{Q}_1^7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_1^2 & \mathbf{Q}_1^3 & 0 & \mathbf{Q}_1^1 \\
\mathbf{Q}_1^4 & \mathbf{Q}_1^5 & \mathbf{Q}_1^6 & 0 & \mathbf{Q}_1^7 & \mathbf{Q}_1^8 & \mathbf{Q}_1^9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_1^1 & \mathbf{Q}_1^2 & \mathbf{Q}_1^3 & 0 \\
0 & \mathbf{Q}_1^4 & \mathbf{Q}_1^5 & \mathbf{Q}_1^6 & 0 & \mathbf{Q}_1^7 & \mathbf{Q}_1^8 & \mathbf{Q}_1^9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_1^1 & \mathbf{Q}_1^2 & \mathbf{Q}_1^3 \\
\mathbf{Q}_1^6 & 0 & \mathbf{Q}_1^4 & \mathbf{Q}_1^5 & \mathbf{Q}_1^9 & 0 & \mathbf{Q}_1^7 & \mathbf{Q}_1^8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_1^3 & 0 & \mathbf{Q}_1^1 & \mathbf{Q}_1^2 \\
\mathbf{Q}_2^2 & \mathbf{Q}_2^3 & 0 & \mathbf{Q}_2^1 & \mathbf{Q}_2^5 & \mathbf{Q}_2^6 & 0 & \mathbf{Q}_2^4 & \mathbf{Q}_2^8 & \mathbf{Q}_2^9 & 0 & \mathbf{Q}_2^7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{Q}_2^1 & \mathbf{Q}_2^2 & \mathbf{Q}_3^2 & 0 & \mathbf{Q}_2^4 & \mathbf{Q}_2^5 & \mathbf{Q}_2^6 & 0 & \mathbf{Q}_2^7 & \mathbf{Q}_2^8 & \mathbf{Q}_2^9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \mathbf{Q}_2^1 & \mathbf{Q}_2^2 & \mathbf{Q}_3^2 & 0 & \mathbf{Q}_2^4 & \mathbf{Q}_2^5 & \mathbf{Q}_2^6 & 0 & \mathbf{Q}_2^7 & \mathbf{Q}_2^8 & \mathbf{Q}_2^9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{Q}_2^3 & 0 & \mathbf{Q}_2^1 & \mathbf{Q}_2^2 & \mathbf{Q}_2^6 & 0 & \mathbf{Q}_2^4 & \mathbf{Q}_2^5 & \mathbf{Q}_2^9 & 0 & \mathbf{Q}_2^7 & \mathbf{Q}_2^8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{Q}_3^2 & \mathbf{Q}_3^3 & 0 & \mathbf{Q}_3^1 & \mathbf{Q}_3^5 & \mathbf{Q}_3^6 & 0 & \mathbf{Q}_3^4 & \mathbf{Q}_3^8 & \mathbf{Q}_3^9 & 0 & \mathbf{Q}_3^7 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{Q}_3^1 & \mathbf{Q}_3^2 & \mathbf{Q}_3^3 & 0 & \mathbf{Q}_3^4 & \mathbf{Q}_3^5 & \mathbf{Q}_3^6 & 0 & \mathbf{Q}_3^7 & \mathbf{Q}_3^8 & \mathbf{Q}_3^9 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \mathbf{Q}_3^1 & \mathbf{Q}_3^2 & \mathbf{Q}_3^3 & 0 & \mathbf{Q}_3^4 & \mathbf{Q}_3^5 & \mathbf{Q}_3^6 & 0 & \mathbf{Q}_3^7 & \mathbf{Q}_3^8 & \mathbf{Q}_3^9 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{Q}_3^3 & 0 & \mathbf{Q}_3^1 & \mathbf{Q}_3^2 & \mathbf{Q}_3^6 & 0 & \mathbf{Q}_3^4 & \mathbf{Q}_3^5 & \mathbf{Q}_3^9 & 0 & \mathbf{Q}_3^7 & \mathbf{Q}_3^8 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_4^2 & \mathbf{Q}_4^3 & 0 & \mathbf{Q}_4^1 & \mathbf{Q}_4^5 & \mathbf{Q}_4^6 & 0 & \mathbf{Q}_4^4 & \mathbf{Q}_4^8 & \mathbf{Q}_4^9 & 0 & \mathbf{Q}_4^7 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_4^1 & \mathbf{Q}_4^2 & \mathbf{Q}_4^3 & 0 & \mathbf{Q}_4^4 & \mathbf{Q}_4^5 & \mathbf{Q}_4^6 & 0 & \mathbf{Q}_4^7 & \mathbf{Q}_4^8 & \mathbf{Q}_4^9 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_4^1 & \mathbf{Q}_4^2 & \mathbf{Q}_3^4 & 0 & \mathbf{Q}_4^4 & \mathbf{Q}_4^5 & \mathbf{Q}_4^6 & 0 & \mathbf{Q}_4^7 & \mathbf{Q}_4^8 & \mathbf{Q}_4^9 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_4^3 & 0 & \mathbf{Q}_4^1 & \mathbf{Q}_4^2 & \mathbf{Q}_4^6 & 0 & \mathbf{Q}_4^4 & \mathbf{Q}_4^5 & \mathbf{Q}_4^9 & 0 & \mathbf{Q}_4^7 & \mathbf{Q}_4^8 \\
\mathbf{Q}_5^8 & \mathbf{Q}_5^9 & 0 & \mathbf{Q}_5^7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_5^2 & \mathbf{Q}_5^3 & 0 & \mathbf{Q}_5^1 & \mathbf{Q}_5^5 & \mathbf{Q}_5^6 & 0 & \mathbf{Q}_5^4 \\
\mathbf{Q}_5^7 & \mathbf{Q}_5^8 & \mathbf{Q}_5^9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_5^1 & \mathbf{Q}_5^2 & \mathbf{Q}_3^5 & 0 & \mathbf{Q}_5^4 & \mathbf{Q}_5^5 & \mathbf{Q}_5^6 & 0 \\
0 & \mathbf{Q}_5^7 & \mathbf{Q}_5^8 & \mathbf{Q}_5^9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_5^1 & \mathbf{Q}_5^2 & \mathbf{Q}_3^5 & 0 & \mathbf{Q}_5^4 & \mathbf{Q}_5^5 & \mathbf{Q}_5^6 \\
\mathbf{Q}_5^9 & 0 & \mathbf{Q}_5^7 & \mathbf{Q}_5^8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_5^3 & 0 & \mathbf{Q}_5^1 & \mathbf{Q}_5^2 & \mathbf{Q}_5^6 & 0 & \mathbf{Q}_5^4 & \mathbf{Q}_5^5
\end{pmatrix}$$

Figure 7.1: Morphology of the dynamics for the discretized system of o shallow–water equations, with the Richtmyer version of the Lax–Wendroff finite difference scheme on a 4×5 resolution grid.

From the expression (7.61) we see that the state vector at a grid point is determined by a combination of the values at 8 adjacent grid points and the value at the same point in the previous step. The matrices $\mathbf{Q}$ are then blocks within the dynamics matrix $\mathbf{\Psi}$ which, for a 4×5 grid has the form displayed in Fig. 7.1 .

Exercises

1. (Daley [39], Problem 6.1) Show that the eigenvectors of the operator $\hat{\mathbf{L}}$ in Section 6.5, given by the columns of the matrix $\hat{\mathbf{V}}$, are orthogonal. Show also that the matrix $\hat{\mathbf{V}}$ is unitary, that is, it satisfies $\hat{\mathbf{V}}^T = \hat{\mathbf{V}}^{-1}$.

2. Derive the shallow–water systems of equations. Starting from Newton's equations (7.1) without external forcing, that is, for $\mathbf{f} = \mathbf{0}$, and considering a cartesian coordinate system, show that the explicit form of (7.1) is

$$\begin{aligned}
\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} - fv &= -\frac{1}{\rho}\frac{\partial p}{\partial x} \\
\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} + fu &= -\frac{1}{\rho}\frac{\partial p}{\partial y} \\
\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} &= -\frac{1}{\rho}\frac{\partial p}{\partial y} - g
\end{aligned}$$

where x, y and z indicate the zonal, meridional, and vertical directions, respectively, $\mathbf{v} = (u, v, w)^T$, and $\mathbf{g} = g\mathbf{k}$ for $\mathbf{k}$ representing a unit vector in the vertical direction.

   (a) Assuming hydrostatic balance, and a homogeneous atmosphere, in which $\bar{p} = p_0 = const.$, $\bar{\rho} = \rho_0 = const.$, and $\bar{z} = h(x, y, t)$, show that the horizontal pressure gradient is independent of the vertical coordinate z.

   (b) Performing a scale analysis in the equations for u and v above, show that these can be reduced to

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv = -g\frac{\partial h}{\partial x}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu = -g\frac{\partial h}{\partial y} \qquad (7.62)$$

(c) Considering now the equation for the vertical velocity w, and remembering that we are assuming hydrostatic balance, show by scale analysis considerations that this equation can be reduced to

$$u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} = 0$$

(d) Noticing that, due to the results of item (b), the horizontal quantities do not depend on z, integrate the equation for w, obtained in the previous item, imposing the following boundary conditions:

$$\begin{array}{ll} w(x,y,z,t) = 0 & \text{na superficie, onde } z = h_t(x,y) \\ w(x,y,z,t) = 0 & \text{no topo da atmosfera, onde } z = h(x,y,t) \end{array}$$

Hence, show that the vertical equation reduces to

$$\frac{\partial h}{\partial t} + \frac{\partial[u(h-h_t)]}{\partial x} + \frac{\partial[v(h-h_t)]}{\partial y} = 0$$

for the height of the atmosphere.

3. Assuming the absence of topography, show that the shallow-water system of equations obtained in the previous problem, linearized about the following basic state:

$$\begin{aligned} u &= U(y) + u' \\ v &= 0 + v' \\ h &= H(y) + h' \end{aligned}$$

and with $f = f_0 + \beta y$, reduces to:

$$\begin{aligned} \frac{\partial u'}{\partial t} + U\frac{\partial u'}{\partial x} + \frac{\partial \phi'}{\partial x} - (f - U_y)v' &= 0 \\ \frac{\partial v'}{\partial t} + U\frac{\partial v'}{\partial x} + \frac{\partial \phi'}{\partial y} + fu' &= 0 \\ \frac{\partial \phi'}{\partial t} + +U\frac{\partial \phi'}{\partial x} + \Phi(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}) + \Phi_y v' &= 0 \end{aligned}$$

where we introduced the geopotential height for the basic state as $\Phi = gH$, its corresponding perturbation as $\phi' = gh'$ and

$$fU + \Phi_y = 0$$

Here, the subscript y indicates derivation with respect to the variable y.

4. Defining the total energy of the system governed by the linear shallow-water equations, obtained in the previous problem, as

$$E = \frac{1}{2}\int\int \Phi(u^2 + v^2) + \phi^2]\,dx\,dy$$

where u, v and ϕ refer to perturbation fields, show that

$$\frac{dE}{dt} = -\int\int \Phi U_y uv \, dx \, dy$$

where the integrals extend through the whole (x, y) plane. Interpret the case $U(y) = U_0 = const.$.

5. (Cohn [30], Ghil et al. [66]) Let us apply the finite difference scheme of Section 6.6 to the one dimensional shallow-water system of equations:

$$\begin{aligned} \frac{\partial u'}{\partial t} + U\frac{\partial u'}{\partial x} + \frac{\partial \phi'}{\partial x} - fv' &= 0 \\ \frac{\partial v'}{\partial t} + U\frac{\partial v'}{\partial x} + fu' &= 0 \\ \frac{\partial \phi'}{\partial t} + U\frac{\partial \phi'}{\partial x} + \Phi_0\frac{\partial u'}{\partial x} - f_0 U v' &= 0 \end{aligned}$$

where U, f_0, e Φ_0 are constants. In this case:

(a) Write the system's equations in flux form, that is,

$$\frac{\partial \mathbf{w}}{\partial t} + \mathbf{A}\frac{\partial \mathbf{w}}{\partial x} + \mathbf{C}\mathbf{w} = 0$$

obtaining explicit expressions for $\mathbf{A}$ and $\mathbf{C}$.

(b) Show that the second step of the Lax–Wendroff scheme can be written as:

$$\mathbf{w}_i^{k+1} = \mathbf{w}_i^k - \lambda\mathbf{A}(\mathbf{w}_{i+1/2}^{k+1/2} - \mathbf{w}_{i-1/2}^{k+1/2}) - \frac{\Delta t}{2}\mathbf{C}(\mathbf{w}_{i-1/2}^{k+1/2} + \mathbf{w}_{i+1/2}^{k+1/2})$$

for $i = 1, 2, \cdots, I$, and $\lambda = \Delta t/\Delta x$.

(c) Show that the first step of the Lax–Wendroff scheme can be written as:

$$\mathbf{w}_{i+1/2}^{k+1/2} = \frac{1}{2}(\mathbf{I} - \frac{\Delta t}{2}\mathbf{C})(\mathbf{w}_i^k + \mathbf{w}_{i+1}^k) - \frac{\lambda}{2}\mathbf{A}(\mathbf{w}_{i+1}^k - \mathbf{w}_i^k)$$

for $i = 1, 2, \cdots, I$.

(d) Substituting this result into item (b), as well as the result obtained via the transformation $i \rightarrow i-1$ in item (c), show that

$$\mathbf{w}_i^{k+1} = \mathbf{Q}_{-1}\mathbf{w}_{i-1}^{k+1} + \mathbf{Q}_0\mathbf{w}_i^{k+1} + \mathbf{Q}_{+1}\mathbf{w}_{i+1}^{k+1}$$

where

$$\mathbf{Q}_0 = \mathbf{I} - \lambda^2\mathbf{A}^2 - \frac{\Delta t}{2}\mathbf{C}(\mathbf{I} - \frac{\Delta t}{2}\mathbf{C})$$

and

$$\mathbf{Q}_{\pm 1} = \mp\frac{\lambda}{2}\mathbf{A} + \frac{\lambda^2}{2}\mathbf{A}^2 \pm \frac{\lambda\Delta t}{4}(\mathbf{AC} + \mathbf{CA}) - \frac{\Delta t}{4}\mathbf{C}(\mathbf{I} - \frac{\Delta t}{2}\mathbf{C})$$

(e) Indicate the morphology of the one-time step transition matrix.

Chapter 8

Atmospheric Data Assimilation: Conventional Methods

8.1 History

In this section we give a short review of the history of conventional atmospheric data assimilation methods. More detailed reviews can be found in Daley [39], in Ghil & Malanotte-Rizzoli [64], and in Harms et al. [73]. We call conventional methods the assimilation schemes currently used in operational centers, as for example at the National Centers for Environmental Prediction (NCEP[1]); Parrish & Derber [112]) in the United States and the European Centre for Medium-Range Weather Forecasts (ECMWF; Heckley et al. [74]), in England.

Weather forecasting started after the invention of synoptic charts (meteorological maps). Nowadays, data analysis by synoptic charts is known as *subjective analysis*, since these were designed by hand, and consequently were subjected to empiricism and skill. These charts are built by marking the magnitude of an observed quantity, in the locations where observations are made, on an ordinary geographic map, and by tracing contours between the marked points. Visual extrapolation and interpolation are made in order to allow for contours drawing. From these charts, many atmospheric conditions can be inferred, including conditions about variables not directly observed, by applying a series of rules based on geometric arguments (Bjerkenes [14]; see also Saucier [122] for a detailed explanation of these procedures). A meteorologist with great experience can then issue forecasts for one or two days based on these charts.

The advent of computers and the evolution of numerical analysis methods introduced a more rapid and consistent way of generating synoptic charts. The first *objective analysis*, as it was called, was produced by Panofsky [111]. He used a technique of fit by least squares in two dimensions. This technique consists basically in expanding the fields (variables), which are to be analyzed, in a series of polynomials about the observation point, minimizing the

[1] Old National Meteorological Center

square of their differences with the observed values. The expansion coefficients are then determined by inverting a matrix.

Furthermore, the availability of computers contributed to the arrival of a new area of research in the beginning of the 40's. Led by the great John von Neumann at the Institute of Advanced Studies in Princeton, a prominent group of meteorologists initiated what was later called numerical weather prediction (NWP). Initiated in the end of the 40's, in a rich collaborative environment that continued until the mid–50's, as described by Wiin–Neilsen [139], the first weather prediction was performed by Charney, Fjørtoft & Neumann [23]. This prediction was based on the numerical integration of the barotropic vorticity equation, as proposed by Charney.

The work of Panofsky [111] in objective analysis was motivated by the Princeton project and executed there. This procedure was perfected by ideas such as those of Bergthórsson & Döös [11], who used a numerical forecasts as the first guess to the analyzed field. This led to the successive corrections method that represents considerable computational improvement over the polynomial fit method. As described in Lorenc et al. [96], an improved version of this method is used today, and is called *analysis correction*, at the British forecast center (U.K. Meteorological Office).

The work of Eliassen [49], Gandin [57], Eddy [45, 46], and Phillips [116] introduced statistical interpolation ideas to atmospheric science problems. This procedure is analogous to the successive corrections procedure in the sense that the analyzed variable, w_j^a, at a point j, is still obtained by means of a linear combination between the forecast field w_j^f (here we use f for forecast), at the same point, and the increment (*innovation*) due to the observation at this point, according to the expression

$$w_j^a = w_j^f + \mathbf{k}_j^T(\mathbf{w}^o - \mathbf{H}\mathbf{w}^f) \tag{8.1}$$

where $\mathbf{w}^o$ is a m–vector representing m observations and the $m \times n$ matrix $\mathbf{H}$ corresponds to necessary interpolations in order to transfer information from the forecast n–vector $\mathbf{w}^f$, usually obtained on a regular grid, to the observation places. Analysis methods based on (8.1) are said to be univariate, since observations of a certain quantity corrects only equivalent quantities. That is, we can identify the vector w as being the temperature, so that the use of (8.1) does not affect the winds, but the temperature field alone.

The quantities $\mathbf{k}_j$ above are the weights given to the observation increments. These can be determined based on what we saw in the initial lectures using statistical concepts. For example, they can be determined by imposing the condition that the ensemble mean of the difference between the analysis and the true value of the analyzed quantity be minimum. In the method of least squares, the quantity to be minimized is a measure of Euclidean distance between the exact value and its estimate, whose solution produces an expression for the weights in (8.1),

$$[\mathbf{H}\mathbf{S}^f\mathbf{H}^T + \mathbf{R}]\mathbf{k}_j = \mathbf{s}_j^f \tag{8.2}$$

where $\mathbf{S}^f$ and $\mathbf{R}$ correspond to the forecast and observation error covariance matrices; $\mathbf{s}_j^f$ is a j^th column of the forecast error covariance matrix.

Physical constraints were incorporated into statistical analysis by Gandin [57] and applied to operational data assimilation systems by McPherson et al. [102] and Lorenc [95]. These constraints are taken into consideration by means of an extension of the univariate form, briefly described above, for the multivariate case. In this last case, the analyzed state at a point is built as a combination of information about various variables in distinct points. For example, the forecast and analysis vectors above can be redefined to include winds and the mass fields, as $\mathbf{w}_j^{f,a} \equiv (u_j^{a,f}, v_j^{f,a}, h_j^{f,a})^T$, where u and v represent the two components of the wind in the zonal and meriodional directions, and, h represents the mass field. The calculation of the weights $\mathbf{k}_j$ is accomplished by following an expression similar to (8.2), except that now, the forecast error covariance matrix $\mathbf{S}^f$ contains terms of cross–covariances related to the cross–correlations among different variables. Therefore, the matrices in (8.2) are of larger dimension than in that of the univariate case. Therefore, the inclusion of physical constraints results in a computationally more intensive system of equations to be solved to produce an analysis.

To solve the system of equations (8.2), we assume the statistical properties of the forecast errors is known, that is, that the matrix $\mathbf{S}^f$ is known. In fact, this matrix is prescribed based on assumptions such as vertical separability and horizontal homogeneity and isotropy of error correlations. Observational studies such as those of Rutherford [119] and Schlatter [123], based on differences of observed and climatological forecast fields are designed to determine the statistics of forecast errors. Analytical expressions can be derived to parameterize error statistics, and the study of Balgovind et al. [6] (see Exercise 8.3) justifies some of these parameterizations by using a univariate model based on the quasi–geostrophic potential vorticity equation.

Both methods of univariate and multivariate statistical interpolation or, as more commonly known, *optimal interpolation*, are used nowadays in forecast centers such as NCEP and ECMWF. In some of its applications, forecast errors variances are taken as a function of a linear growth parameter, while the correlation fields are prescribed in an analytical form, by imposing geostrophic balance for the error fields (Schlatter et al. [124]), as the physical constraint.

As we mentioned previously, the major part of the computational load in optimal interpolation is primarily due to the matrix inversion in (8.2); for regions with high data density, this matrix can have a very large dimension. Defining a region of influence, to produce an analyzed variable at a certain point taking into consideration observations within a radius on influence only is one attempt to reduce the computational cost (Lorenc [95]). This technique is known as *data selection*, and in some cases it has unpleasant consequences (da Silva et al. [37]). Another technique, known as superobing, substitutes various observations occurring in nearby locations by a single observation, as for example, the mean of all the observations, with modified standard deviation. Superobing reduces the number of observations included in (8.2) (see Lorenc [95]).

Another factor responsible for high computational cost in operational analysis systems is the need for *quality control* of the data. Quality control systems are developed in order to eliminate and/or correct observations with gross errors, which are introduced by artificial means, distinct from the measurement process itself, such as data corruption due to transmission via the telecommunication network. The study of Hollingsworth et al. [80] shows that the final analysis result is very sensitive to quality control procedures. Other approaches to the quality control problem are found in the work of Lorenc & Hammon [97] and Gandin [58].

The greatest disadvantage of the optimal interpolation method is that the forecast error covariance is *prescribed* in a relatively arbitrary manner. Although the errors possess dynamic balance, the dynamic balance is not exactly satisfied by the governing equations, moreover these errors do not propagate in any way. Explicit use of the governing equations to derive error statistics is the main context of advanced methods for data assimilation, to be discussed in the following lecture. Among other things, the advanced methods produce error statistics with appropriate balance.

8.2 Initialization

Intialization is the procedure by which gravity waves are filtered out from the initial conditions to allow for an evolution *practically* free of fast components. We say practically because in the general nonlinear case, the evolution of the initialized state (after initialization has been performed) still contains allows for fast components of the system to be excited, since the all modes interact due to nonlinearity. For the linear case, however, various initialization procedures exist and consist of the motivation for extensions to the nonlinear case.

Initialization is not a data assimilation method, however, it is a fundamental ingredient for atmospheric forecast obtained through computer models. The majority of assimilation methods used operationally use some type of initialization method for the fields to be used as initial condition in the forecast model. It is worth saying that the need to initialize the analyzed fields, that is, fields obtained after after the assimilation of observations, could be, in some cases, eliminated if the assimilation procedure was done "correctly". In other words, referring to the existing balance due to dynamical processes, it is possible, in some cases, to produce analyzed fields which are automatically balanced, without us having to use an explicit initialization procedure. The word "correct" used above, refers to assimilation schemes in which the initialization procedure is embedded in its structure. This will become more clear as we progress. It is good to underline that we are referring only to balances due to dynamic processes, such as geostrophic balance; balances due to physical processes, such as heat exchange in the atmospheric system, are much more complicated to incorporate automatically in data assimilation procedure. The major part of this topic goes beyond what we intend to cover in these lectures, and we will see in the following sections only superficially what we have just mentioned.

One way of describing the initialization problem is to imagine that in the solution space of the equations that govern atmospheric motion there exists a subspace of slow solutions called $\mathcal{S}$, which is free of high frequency waves (gravity waves — with potential to destroy possible weather forecasts). In fact, there are various definitions for what it is understood for the slow subspace (slow manifold; see Boyd [17]) but we will not go into these details in what follows. Initialization can be seen as the process of projecting, in some way, a general state represented by the n–vector $\mathbf{w}^a$, at a specific time, onto this slow manifold. Through concepts of linear algebra, we know that there is a matrix (operator) of projection $\mathbf{\Pi}$, of dimension $n \times n$, which satisfies the following conditions

$$\begin{aligned} Range\,\mathbf{\Pi} &= \mathcal{S}\,, && (8.3a)\\ \mathbf{\Pi}^2 &= \mathbf{\Pi}\,, && (8.3b)\\ (\mathbf{E\Pi})^T &= \mathbf{E\Pi}\,, && (8.3c) \end{aligned}$$

(e.g., Halmos [69], Section 75). This matrix is known as the **E**–orthogonal projection matrix onto $\mathcal{S}$, or simply the orthogonal projection in when **E** is the identity matrix. Here, **E** is assumed to be positive definite and symmetric.

Let $\mathbf{V}_S$ be an $n \times n_S$ matrix, with columns built from the slow eigenvectors of a dynamical system, that is, the slow normal modes which total n_S. Hence, the **E**–orthogonal projector is given by:

$$\mathbf{\Pi} = \mathbf{V}_S \left(\mathbf{V}_S^T \mathbf{E} \mathbf{V}_S\right)^{-1} \mathbf{V}_S^T \mathbf{E}\,, \qquad (8.4)$$

which can be verified by substitution in the conditions (8.3).

Making use of this projector we can describe the initialization problem as a least squares problem: find the n–vector $\mathbf{w}^i$ in $\mathcal{S}$ that is as close as possible to the general vector $\mathbf{w}^a$. In other words, we want to minimize the following functional

$$\eta = (\mathbf{w}^i - \mathbf{w}^a)^T \mathbf{E} (\mathbf{w}^i - \mathbf{w}^a) \qquad (8.5)$$

where **E** is a symmetric and positive definite weight, or re–scaling, matrix. As we will show, the solution of this problem is, as we can expect, given uniquely by

$$\mathbf{w}^i = \mathbf{\Pi} \mathbf{w}^a\,. \qquad (8.6)$$

Posed in this way, the initialization problem is know as *linear variational normal mode initialization* (e.g., Daley [40], Temperton [127]).

To prove the result above, notice that certainly $\mathbf{w}^i \equiv \mathbf{\Pi}\mathbf{w}^a \in \mathcal{S}$. Moreover, a general element of $\mathcal{S}$ can be written as

$$\mathbf{w} = \mathbf{w}^i + \epsilon = \mathbf{\Pi}\mathbf{w}^a + \epsilon \qquad (8.7)$$

where

$$\epsilon = \mathbf{\Pi}\epsilon\,, \qquad (8.8)$$

since $\epsilon = \mathbf{w} - \mathbf{w}^i$ should be in $\mathcal{S}$. Now,

$$\begin{aligned} \eta = \eta(\mathbf{w}) &= (\mathbf{w} - \mathbf{w}^a)^T \mathbf{E} (\mathbf{w} - \mathbf{w}^a) \\ &= \left[(\mathbf{\Pi} - \mathbf{I})\mathbf{w}^a + \epsilon\right]^T \mathbf{E} \left[(\mathbf{\Pi} - \mathbf{I})\mathbf{w}^a + \epsilon\right] \\ &= \left[(\mathbf{\Pi} - \mathbf{I})\mathbf{w}^a\right]^T \mathbf{E} \left[(\mathbf{\Pi} - \mathbf{I})\mathbf{w}^a\right] + \epsilon^T \mathbf{E} \epsilon + 2\delta\,, && (8.9) \end{aligned}$$

where

$$\begin{aligned} \delta &= \epsilon^T \mathbf{E} (\mathbf{\Pi} - \mathbf{I}) \mathbf{w}^a \\ &= (\mathbf{\Pi}\epsilon)^T \mathbf{E} (\mathbf{\Pi} - \mathbf{I}) \mathbf{w}^a\,, && (8.10) \end{aligned}$$

according to (8.8). Based on (8.3c) and on the fact that $\mathbf{E}^T = \mathbf{E}$ it follows that

$$\begin{aligned}\delta &= \epsilon^T \mathbf{\Pi}^T \mathbf{E}(\mathbf{\Pi} - \mathbf{I})\mathbf{w}^a \\ &= \epsilon^T (\mathbf{E}\mathbf{\Pi})^T (\mathbf{\Pi} - \mathbf{I})\mathbf{w}^a \\ &= \epsilon^T \mathbf{E}\mathbf{\Pi}(\mathbf{\Pi} - \mathbf{I})\mathbf{w}^a \\ &= \epsilon^T \mathbf{E}(\mathbf{\Pi}^2 - \mathbf{\Pi})\mathbf{w}^a \\ &= 0\,, \end{aligned} \tag{8.11}$$

where (8.3b) was used to obtain the last equality. By (8.9) we have

$$\begin{aligned}\eta(\mathbf{w}) &= (\mathbf{w}^i - \mathbf{w}^a)^T \mathbf{E}(\mathbf{w}^i - \mathbf{w}^a) + \epsilon^T \mathbf{E}\epsilon \\ &= \eta(\mathbf{w}^i) + \epsilon^T \mathbf{E}\epsilon \\ &\geq \eta(\mathbf{w}^i)\,, \end{aligned} \tag{8.12}$$

where the equality prevails if and only if $\epsilon = \mathbf{0}$, since $\mathbf{E}$ is positive defined. Hence, $\mathbf{w}^i = \mathbf{\Pi}\mathbf{w}^a$ is the unique minimizer.

As an example of the initialization procedure, we follow Daley [39] and consider the system of shallow water equations linearized about the state of rest and dealt with here earlier in Section 6.5. For that, we introduce first an operator $\mathbf{F}$ corresponding to the Fourier transform in two spatial dimensions. Thus, formally a vector $\mathbf{w}(x, y, t)$ can be transformed according to

$$\hat{\mathbf{w}}(k, \ell, t) = \mathbf{F}[\mathbf{w}(x, y, t)] \tag{8.13}$$

The Fourier operator $\mathbf{F}^{-1}$ "=" $\mathbf{F}^*$ which correspond to the inverse Fourier transform is such that

$$\mathbf{w}(x, y, t) = \mathbf{F}^*[\mathbf{w}(k, \ell, t)] \tag{8.14}$$

where $*$ indicates the conjugate transpose of an operator. Consequently, we can write the matrix $\hat{\mathbf{V}}_S$ from the previous lecture as,

$$\hat{\mathbf{V}}_S \equiv \mathbf{F}[\mathbf{V}_S] \tag{8.15}$$

Furthermore, we define the Fourier component $\hat{\mathbf{E}}$ of the matrix $\mathbf{E}$ above as:

$$\hat{\mathbf{E}} \equiv \mathbf{F}(\mathbf{F}^*\mathbf{E})^T \tag{8.16}$$

Then, formally again, by inserting the appropriate identity in expression (8.4) for the projector $\mathbf{\Pi}$ we have

$$\begin{aligned}\mathbf{\Pi} &= (\mathbf{F}^*\mathbf{F})\mathbf{V}_S \left[\mathbf{V}_S^T(\mathbf{F}\mathbf{F}^*)\mathbf{E}(\mathbf{F}^*\mathbf{F})\mathbf{V}_S\right]^{-1} \mathbf{V}_S^T(\mathbf{F}^*\mathbf{F})\mathbf{E}(\mathbf{F}^*\mathbf{F})\,, \\ &= \mathbf{F}^*\hat{\mathbf{V}}_S \left(\hat{\mathbf{V}}_S^T \hat{\mathbf{E}} \hat{\mathbf{V}}_S\right)^{-1} \hat{\mathbf{V}}_S^T \hat{\mathbf{E}}\mathbf{F}\,, \\ &= \mathbf{F}^*\hat{\mathbf{\Pi}}\mathbf{F} \end{aligned} \tag{8.17}$$

where

$$\hat{\mathbf{\Pi}} \equiv \hat{\mathbf{V}}_S \left(\hat{\mathbf{V}}_S^T \hat{\mathbf{E}} \hat{\mathbf{V}}_S\right)^{-1} \hat{\mathbf{V}}_S^T \hat{\mathbf{E}}\,, \tag{8.18}$$

Explicitly, we recall that, in the Fourier space, the solution of the system of shallow water equations in time t is given by

$$\hat{\mathbf{w}}(t) = \hat{\mathbf{V}}\boldsymbol{\Lambda}(t) \; = \; \hat{\mathbf{V}} \begin{pmatrix} c_{-}e^{i\sigma^{-}t}0 & 0 & \\ 0 & c_0 & 0 \\ 0 & 0 & c_{+}e^{i\sigma^{+}t} \end{pmatrix} \tag{8.19}$$

where $\hat{\mathbf{V}}$ is the eigenvector matrix in (7.43), which for convenience we write as in Daley [39]:

$$\hat{\mathbf{V}} = \begin{pmatrix} G_{\psi}^{-} & R_{\psi} & G_{\psi}^{+} \\ G_{\chi}^{-} & R_{\chi} & G_{\chi}^{+} \\ G_{\phi}^{-} & R_{\phi} & G_{\phi}^{+} \end{pmatrix}, \tag{8.20}$$

and where the coefficients $\hat{\mathbf{c}} = (\hat{c}_{-}, \hat{c}_0, \hat{c}_{+})^T$ can be determined from the initial condition, as in (7.47), and which we repeat for ease of reference once again:

$$\hat{\mathbf{c}} = \hat{\mathbf{V}}^T \begin{pmatrix} \hat{\psi}^a(0) \\ \hat{\chi}^a(0) \\ \hat{\phi}^a(0) \end{pmatrix} \tag{8.21}$$

for $\hat{\mathbf{w}}^a(0) = [\hat{\psi}^a(0), \hat{\chi}^a(0), \hat{\phi}^a(0)]^T$, a general initial state, meaning an analysis.

Therefore, according to (8.21) and (8.19), if a general initial condition has fast components, the solution for all times t_k will also have fast components unless the coefficients $\hat{c}_{-}$ and $\hat{c}_{+}$ are zero. The goal of the initialization procedure is to reconstruct the initial state $\mathbf{w}^a(0)$ as a (initialized) state $\hat{\mathbf{w}}^i(0) \equiv [\hat{\psi}^i(0), \hat{\chi}^i(0), \hat{\phi}^i(0)]^T$, free of fast components, so that its evolution (8.19) is also free of fast (gravity) waves.

Moreover, in the case of simple linear systems, the slow subspace coincides with the geostrophic space (rotational), and the "matrix" $\mathbf{V}_S$ of slow eigenvectors is given by

$$\hat{\mathbf{V}}_S = \begin{pmatrix} R_{\psi} \\ R_{\chi} \\ R_{\phi} \end{pmatrix} \tag{8.22}$$

for a number $n_S = 1$ of slow vectors.

Choosing the weighting matrix $\hat{\mathbf{E}}$ as a diagonal matrix (in Fourier space) and representing it by $\hat{\mathbf{E}} = diag(w_{\psi}, w_{\chi}, w_{\phi})$ the kernel $\left(\hat{\mathbf{V}}_S^T\hat{\mathbf{E}}\hat{\mathbf{V}}_S\right)^{-1}$ of the $\mathbf{E}$–orthogonal projector can then be calculated according to

$$\begin{aligned} \left(\hat{\mathbf{V}}_S^T\hat{\mathbf{E}}\hat{\mathbf{V}}_S\right)^{-1} &= \left[\begin{pmatrix} R_{\psi} & R_{\chi} & R_{\phi} \end{pmatrix} \begin{pmatrix} w_{\psi} & 0 & 0 \\ 0 & w_{\chi} & 0 \\ 0 & 0 & w_{\phi} \end{pmatrix} \begin{pmatrix} R_{\psi} \\ R_{\chi} \\ R_{\phi} \end{pmatrix}\right]^{-1} \\ &= \frac{1}{w_{\psi}R_{\psi}^2 + w_{\chi}R_{\chi}^2 + w_{\phi}R_{\phi}^2} \\ &= \frac{k+1}{kw_{\psi} + w_{\phi}} \end{aligned} \tag{8.23}$$

where we make explicit use of the elements of the matrix $\hat{\mathbf{V}}_S$. Therefore, the projector $\hat{\mathbf{\Pi}}$ takes the form

$$\begin{aligned}
\hat{\mathbf{\Pi}} &= \frac{k+1}{kw_\psi + w_\phi} \begin{pmatrix} R_\psi \\ R_\chi \\ R_\phi \end{pmatrix} \begin{pmatrix} R_\psi & R_\chi & R_\phi \end{pmatrix} \begin{pmatrix} w_\psi & 0 & 0 \\ 0 & w_\chi & 0 \\ 0 & 0 & w_\phi \end{pmatrix} \\
&= \frac{1}{kw_\psi + w_\phi} \begin{pmatrix} kw_\psi & 0 & \sqrt{k}w_\phi \\ 0 & 0 & 0 \\ \sqrt{k}w_\psi & 0 & w_\phi \end{pmatrix} \qquad (8.24)
\end{aligned}$$

Applying this projector to the general, non–initialized, vector $\mathbf{w}^a$ as in (8.6), we have

$$\hat{\mathbf{w}}^i \equiv \begin{pmatrix} \hat{\psi}^i \\ \hat{\chi}^i \\ \hat{\phi}^i \end{pmatrix} = \frac{1}{kw_\psi + w_\phi} \begin{pmatrix} kw_\psi\hat{\psi}^a + \sqrt{k}w_\phi\hat{\phi}^a \\ 0 \\ \sqrt{k}w_\psi\hat{\psi}^a + w_\phi\hat{\phi}^a \end{pmatrix} \qquad (8.25)$$

The projector for which the diagonal elements of the matrix $\hat{\mathbf{E}}$ are unity, that is, $w_\phi = w_\chi = w_\phi = 1$ is called ***slow orthogonal projector.*** A slow state generated by means of this projector is one corresponding to zero (divergence) velocity potential $\hat{\chi} = 0$, and stream and geopotential functions given by:

$$\begin{aligned}
\hat{\psi}^i &= \frac{k\hat{\psi}^a + \sqrt{k}\hat{\phi}^a}{k+1} \qquad (8.26a) \\
\hat{\phi}^i &= \frac{\sqrt{k}\hat{\psi}^a + \hat{\phi}^a}{k+1} \qquad (8.26b)
\end{aligned}$$

respectively.

Notice that this operation leaves the geostrophic modes unaltered. That is, for the case in which the general initial state is geostrophically balanced, we have that $\hat{\psi}^a = \sqrt{k}\hat{\phi}^a$, and from the expressions above it follows that the initialized state is given by,

$$\begin{aligned}
\hat{\psi}^i &= \sqrt{k}\hat{\phi}^a = \hat{\psi}^a \qquad (8.27a) \\
\hat{\phi}^i &= \hat{\phi}^a \qquad (8.27b)
\end{aligned}$$

Therefore, the initialized state $\hat{\psi}^i = \sqrt{k}\hat{\phi}^i$, is also geostrophically balanced.

In fact, from (8.26) it follows that

$$k\hat{\psi}^i + \sqrt{k}\hat{\phi}^i = k\hat{\psi}^a + \sqrt{k}\hat{\phi}^a \qquad (8.28)$$

Thus, observing that $\hat{q} = k\hat{\psi} + \sqrt{k}\hat{\phi}$ is the expression for the quasi–geostrophical potential vorticity (see Exercise 7.1), we see that $\hat{q}^i = \hat{q}^a$ means that the slow orthogonal projector keeps this quantity conserved.

The slow orthogonal projector used above can be written as

$$\hat{\mathbf{\Pi}}_{||} = \frac{1}{k+1} \begin{pmatrix} k & 0 & \sqrt{k} \\ 0 & 0 & 0 \\ \sqrt{k} & 0 & 1 \end{pmatrix} \qquad (8.29)$$

Other possible choices for the weights w_ψ and w_ϕ produce the following projectors:

$$\hat{\mathbf{\Pi}}_g = \begin{pmatrix} 0 & 0 & \sqrt{k} \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{8.30}$$

for $w_\psi = 0$ e $w_\phi = 1$, and

$$\hat{\mathbf{\Pi}}_r = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1/\sqrt{k} & 0 & 0 \end{pmatrix} \tag{8.31}$$

for $w_\psi = 1$ and $w_\phi = 0$. An initialized state generated by means of $\hat{\mathbf{\Pi}}_g$ is referred to as a state with *geopotential constraint*, since this projector maintains the geopotential component $\hat{\phi}^a$ of a general state $\mathbf{w}^a$ unaltered; analogously, a state generated by the projector $\hat{\mathbf{\Pi}}_r$ is said to be a state of *rotational wind constraint*, since in this case, it is stream function $\hat{\psi}^a$ is unaltered.

The projector $\hat{\mathbf{\Pi}}_g$ is adequate when the geopotential component of the general state $\mathbf{w}^a$ is the one in which we have greater dependability (less error) than the other state components; the projector $\hat{\mathbf{\Pi}}_r$, however, is convenient when we have great reliability in the component of the stream function. In the general case, in which the reliability of the components of an arbitrary state $\mathbf{w}^a$ is not well defined, the orthogonal projector $\hat{\mathbf{\Pi}}_{||}$ is the most adequate. Obviously, when the component of divergent winds $\hat{\chi}$ is of great reliability, we have to choose one of the projectors above. In fact, if we do not know anything about the geopotential and stream function components, we can do nothing in terms of initialization for the case of the shallow water equations linearized about the state of rest.

It is important to notice that, in the example we are considering in this section, the normal modes of the system are orthogonal. In the more general case in which these modes are not orthogonal, basically due to the fact that the matrix $\hat{\mathbf{L}}$ of Section 6.5 is not symmetric due to general basic flows, more general projectors can be obtained. In these cases we should make use of the a bi–orthonormal set of modes that can be obtained through the use of eigenvectors of the adjoint matrix of $\hat{\mathbf{L}}$ (e.g., Ghil [61], for more details in the context of shallow water model in one dimension, linearized about a constant jet, see also Exercise 7.2).

8.3 Dynamic Relaxation

Dynamic relaxation, also called nudging or Newtonian relaxation, is a data assimilation procedure continuous in time. The observations are introduced in the governing equations by means of forcing terms added to the equations, in order to "pull" (relax) the fields in the direction of the observations. Dynamic relaxation is employed during a period of time known as the *pre–forecast*, so that at the end of this period the solution is as close as we want to the observations; from that time on, it is possible to produce regular forecasts. One of the advantages of this procedure is that the initial fields, from which the forecasts are issued, are automatically in dynamic balance at the end of the assimilation (relaxation) period.

More clearly, following the treatment of Haltiner & Williams [70], the data assimilation procedure by dynamic relaxation consists of the following steps:

1. Specify the initial condition, at time $t_0 - T$, when the evolution of the pre-forecast begins, where T represents the period of forecast time and t_0 the instant of time when the forecasts are to be issued from.
2. Solve the governing equations during the interval of time $[t_0 - T, t_0]$, including the forcing terms to relax the solution in the direction of the observations (or analysis).
3. Arriving at time t_0, evolve the governing equations, without the forcing terms, up to the time t of the desired forecast.

In general, the evolution of any prognostic quantity, at a mesh point, where an observation supposedly exists, can be represented by the equation:

$$\frac{\partial w}{\partial t} = f(\mathbf{w}) + \gamma(w^o - w) \tag{8.32}$$

where w is the scalar quantity of interest, f is a function of the vector state $\mathbf{w}$ of the system, which includes the terms of the governing dynamics, the last term is a component of the forcing term, added to the governing equations during the pre-forecast period, and includes the observation w^o, with relaxation parameter γ. Written in this form, the equation above presumes the availability of the observation at the mesh point of interest, and eventually at all grid points. Since observations are rarely available at grid points, it is best to replace the observation w^o by the analyzed value w^a. In this way the relaxation expression can be written as

$$\frac{\partial w}{\partial t} = f(\mathbf{w}) + \gamma(w^a - w) \tag{8.33}$$

The intention of the method can be understood from a simple example, by considering $f = 0$ in the equation above. Assuming that the observation w^o is independent of time, and integrating (8.32) from $t_0 - T$ to t_0 we have:

$$\begin{aligned} w &= w_0 e^{-\gamma T} + \gamma w^a e^{-\gamma t_0} \int_{t_0 - T}^{t_0} e^{\gamma s}\, ds \\ &= w_0 e^{-\gamma T} + (1 - e^{-\gamma T}) w^a \end{aligned} \tag{8.34}$$

where w_0 is solution at time $t_0 - T$. Therefore, as the relaxation interval T increases, the solution approaches the value of the analysis w^a (observation w^o). In practice, the interval T is fixed and the relaxation parameter γ is chosen in order to relax the solution more rapidly, or more slowly, in the direction of the analyses (observations).

Another example can be presented by returning to the system of equations in Section 6.5. When we introduce the dynamic relaxation terms referring to the analyzed fields u^a, v^a and ϕ^a, of winds and geopotential, respectively, we have:

$$\frac{\partial u}{\partial t} - f_0 v + \frac{\partial \phi}{\partial x} - \gamma_u(u^a - u) = 0 \tag{8.35a}$$

$$\frac{\partial v}{\partial t} + f_0 u + \frac{\partial \phi}{\partial y} - \gamma_v(v^a - v) = 0 \tag{8.35b}$$

$$\frac{\partial \phi}{\partial t} + \Phi(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) - \gamma_\phi(\phi^a - \phi) = 0 \tag{8.35c}$$

where, in this case, the relaxation parameters γ_u, γ_v, and γ_ϕ are in principle distinct. The analytic study of this problem can be found in Hoke & Anthes [81], for the unidimensional case.

To simplify the problem mathematically, let us follow Daley [39] where only the geopotential variable is relaxed (and observed), that is, $\gamma_u = \gamma_v = 0$. In this way, the equations for u and v are identical to those in Section 6.5, without the forcing term, thus we can use in their places the corresponding equations for vorticity and divergence. So, the system of equations to be solved becomes:

$$\frac{\partial \nabla^2 \psi}{\partial t} + f_0 \nabla^2 \chi = 0 \tag{8.36a}$$

$$\frac{\partial \nabla^2 \chi}{\partial t} - f_0 \nabla^2 \psi + \nabla^2 \phi = 0, \tag{8.36b}$$

$$\frac{\partial \phi}{\partial t} + \Phi \nabla^2 \chi + \gamma_\phi \phi = \gamma_\phi \phi^a \tag{8.36c}$$

Let us assume that the exact solution of the evolution of the fields $\psi = \psi^t, \chi = \chi^t$ and $\phi = \phi^t$ follows the system of original equations (7.29), without the forcing term, and that the state to be analyzed evolves according to the forced system of equations so that $\psi = \psi^i, \chi = \chi^i$ and $\phi = \phi^i$, are now the quantities which satisfy (8.36). Notice that both systems of equations (7.29) and (8.36) are linear. Defining the initialization errors in an usual way:

$$\tilde{\psi} = \psi^i - \psi^t \tag{8.37a}$$

$$\tilde{\chi} = \chi^i - \chi^t \tag{8.37b}$$

$$\tilde{\phi} = \phi^i - \phi^t \tag{8.37c}$$

we can identify the system of equations (8.36) as describing the evolution of errors. That is, the initialization errors evolve according to

$$\frac{\partial \nabla^2 \tilde{\psi}}{\partial t} + f_0 \nabla^2 \tilde{\chi} = 0 \tag{8.38a}$$

$$\frac{\partial \nabla^2 \tilde{\chi}}{\partial t} - f_0 \nabla^2 \tilde{\psi} + \nabla^2 \tilde{\phi} = 0, \tag{8.38b}$$

$$\frac{\partial \tilde{\phi}}{\partial t} + \Phi \nabla^2 \tilde{\chi} + \gamma_\phi \tilde{\phi} = \gamma_\phi \tilde{\phi}^a \tag{8.38c}$$

where $\tilde{\phi}^a$ is the analysis error in the geopotential field..

The system above has constant coefficients, as the system in Section 6.5, and we can once again solve it by normal modes. We introduce a transformation equivalent to that in (7.34), but now for the error fields:

$$\begin{pmatrix} \tilde{\psi}(x,y,t) \\ \tilde{\chi}(x,y,t) \\ \tilde{\phi}(x,y,t) \end{pmatrix} = \begin{pmatrix} \hat{\psi}(t) \\ i\hat{\chi}(t) \\ f_0\sqrt{k}\hat{\phi}(t) \end{pmatrix} \exp\left\{ i \left[\frac{(mx + ny)}{a} \right] \right\} \tag{8.39}$$

where m, n and a have the same meaning as in previously, and where the constant k is defined as in (7.35). Notice that the amplitudes $\hat{\psi}(t), \hat{\chi}(t)$ and $\hat{\phi}(t)$ in this case are not necessarily the same as those in Section 6.5, although they are represented by the same symbols. Furthermore, we decompose the error in the analysis ϕ^a in a similar manner:

$$\tilde{\phi}^a(x,y,t) = f_0\sqrt{k}\hat{\phi}^a(t)\exp\left\{i\left[\frac{(mx+ny)}{a}\right]\right\} \tag{8.40}$$

Therefore, the system of equations (8.38) is reduced to an ordinary non-homogeneous differential equation:

$$\frac{d\hat{\mathbf{w}}(t)}{dt} + if_0\hat{\mathbf{L}}'\hat{\mathbf{w}}(t) = \gamma\hat{\mathbf{w}}^a, \tag{8.41}$$

where the vector $\hat{\mathbf{w}} \boxminus (\hat{\psi}, \hat{\chi}, \hat{\phi})^T$, the matrix $\hat{\mathbf{L}}'$ is given by

$$\hat{\mathbf{L}}' = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -\sqrt{k} \\ 0 & -\sqrt{k} & -i\gamma_\phi/f_0 \end{pmatrix}. \tag{8.42}$$

and the non–homogeneous part of the system of equations is given by the the analysis error vector $\hat{\mathbf{w}}^a \equiv (0, 0, \hat{\phi}^a)^T$. The solution of (8.41) consists of the solution of the homogeneous part of the system of equations plus the particular solution, that is,

$$\hat{\mathbf{w}}(t_0) = e^{-if_0\hat{\mathbf{L}}'T}\hat{\mathbf{w}}(t_0 - T) + \hat{\mathbf{w}}_p(t_0) \tag{8.43}$$

where $\hat{\mathbf{w}}(t_0 - T)$ is the initial condition vector and $\hat{\mathbf{w}}_p(t_0)$ represents a particular solution.

The interest here is to study of the behavior of the wave frequencies of this modified system. These frequencies are given by the eigenvalues of the matrix $\hat{\mathbf{L}}'$. However, now this matrix is imaginary, so its eigenvalues are also imaginary. It is simple to verify that the secular equation is:

$$\sigma_\ell^3 + i\frac{\gamma_\phi}{f_0}\sigma^2 - (1+k)\sigma_\ell - i\frac{\gamma_\phi}{f_0} = 0 \tag{8.44}$$

whose solutions can be written as,

$$\sigma = Re(\sigma) + iIm(\sigma) \tag{8.45}$$

where $Re(\sigma)$ and $Im(\sigma)$ are the real and imaginary parts of the eigenvalues. The real part represents, as before, the frequencies of oscillation, while the imaginary part represents decaying, or growing, modes.

For $k = 0$, the squares of (8.44) are

$$\sigma_R = -i\frac{\gamma_p hi}{f_0} \tag{8.46a}$$

$$\sigma_G^\pm = \pm 1 \tag{8.46b}$$

So that the errors in the rotational mode for $k = 0$ decrease with time, while the errors in the gravitational modes are oscillatory. For $k > 0$, it is possible to show that, when $\gamma_\phi/f_0 > 0$, we have $Im(\sigma) < 0$. This means that the procedure of dynamic relaxation introduces decaying modes, except in the case of the two inertial–gravity modes for $k = 0$. Also, $Re(\sigma_R) = 0$, for $k \geq 0$, meaning that the frequencies of the rotational modes are not modified by relaxation procedure. (see Daley [39], pp. 359–360, for an approximate calculation of the frequencies for the case $k > 0$.)

8.4 Optimal Interpolation

As discussed in the introduction of this lecture, data assimilation by the method of optimal interpolation (OI) uses an expression as in (8.1) to update instantaneously the values of the variables of the system at a mesh point (analysis point). Contrary to the relaxation method seen in the previous section, OI is an intermittent assimilation method, since it is used only at synoptic times, that is, instants of time considered standard in meteorology, such as 00, 06, 12, and 18 GMT (Greenwich Mean Time). During the 6 hours period, between the synoptic times, the state of the atmosphere evolves by means of a system of equations discretized in space and time, representing a model of general circulation of the atmosphere. These general circulation models produce a base state (do not confuse it with basic state), or forecast state $\mathbf{w}_k^f$, at time t_k. The correction due to the availability of observations can be obtained on the basis of the methods discussed in Lecture 4, through the formula

$$\mathbf{w}_k^a = \mathbf{w}_k^f + \tilde{\mathbf{K}}_k(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f) \tag{8.47}$$

where the notation used here is the same as in previous lectures. As seen in the introduction of this lecture, the weights $\tilde{\mathbf{K}}_k$ in the OI method are obtained by means of the expression

$$\tilde{\mathbf{K}}_k = \mathbf{S}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{S}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{8.48}$$

where we use a "tilde" over the weighting matrix to highlight the fact that these is not the Kalman gains. The reason for this being that the forecast error covariance matrix $\mathbf{S}_k^f$ is *specified*, rather than predicted according to the Kalman filer equations. As we have seen in Lecture 5, for the case of linear systems, the calculation of the forecast error covariance matrix involves an enormous computational cost. As a matter of fact, the reasons to avoid explicit calculation of $\mathbf{S}_k^f$ go beyond the computational issue. They are also attributed to the nonlinearity of governing equations, as well as to lack of knowledge of quantities such as the modeling and observation error covariance matrices. In OI the elements of the matrix $\mathbf{S}^f$ are specified based on statistical evaluations and dynamic constraints, as described below.

In a relatively general way, the state vector at a point $\mathbf{r} = (\lambda, \varphi, p)$, at a certain instant of time, encompasses the wind vector and the geopotential function $\mathbf{w}(\mathbf{r}) = (u, v, \phi)^T(\mathbf{r})$, where for the moment we omit the time index. Thus, the error covariance matrix $\mathbf{S}^f$ between two points $\mathbf{r}_i$ e $\mathbf{r}_j$ is given by

$$\mathbf{S}^f(\mathbf{r}_i, \mathbf{r}_j) = \mathcal{E}\{\tilde{\mathbf{w}}(\mathbf{r}_i)\tilde{\mathbf{w}}^T(\mathbf{r}_j)\} \tag{8.49}$$

where $\tilde{\mathbf{w}}(\mathbf{r}) = \mathbf{w}^f(\mathbf{r}) - \mathbf{w}^t(\mathbf{r})$ is the forecast error, for $\mathbf{w}^t(\mathbf{r})$ representing the real value of the state of the atmosphere. Therefore, we can decompose $\mathbf{S}^f$ as

$$\mathbf{S}^f(\mathbf{r}_i, \mathbf{r}_j) \equiv \begin{pmatrix} S^{f|uu}(\mathbf{r}_i, \mathbf{r}_j) & S^{f|uv}(\mathbf{r}_i, \mathbf{r}_j) & S^{f|u\phi}(\mathbf{r}_i, \mathbf{r}_j) \\ S^{f|vu}(\mathbf{r}_i, \mathbf{r}_j) & S^{f|vv}(\mathbf{r}_i, \mathbf{r}_j) & S^{f|v\phi}(\mathbf{r}_i, \mathbf{r}_j) \\ S^{f|\phi u}(\mathbf{r}_i, \mathbf{r}_j) & S^{f|\phi v}(\mathbf{r}_i, \mathbf{r}_j) & S^{f|\phi\phi}(\mathbf{r}_i, \mathbf{r}_j) \end{pmatrix} \tag{8.50}$$

where $S^{f|\cdot\cdot}(\mathbf{r}_i, \mathbf{r}_j)$ are the cross–covariance functions defined in analogy to (8.49), that is,

$$S^{f|uu}(\mathbf{r}_i, \mathbf{r}_j) = \mathcal{E}\{\tilde{u}(\mathbf{r}_i)\tilde{u}(\mathbf{r}_j)\} \tag{8.51}$$

where $\tilde{u}(\mathbf{r}) = u^f(\mathbf{r}) - u^t(\mathbf{r})$ represents the forecast error in the variable u, at point $\mathbf{r}$, and similarly for the rest of the error cross–covariances in (8.50).

The geopotential–geopotential forecast error covariance function $S^{f|\phi\phi}$, can be written in terms of the correlation and variance fields as

$$S_{ij}^{f|\phi\phi} \equiv S^{f|\phi\phi}(\mathbf{r}_i, \mathbf{r}_j) = \sigma_i^\phi \sigma_j^\phi C_{ij}^{\phi\phi}, \tag{8.52}$$

using a compact notation for the error standard deviation $\sigma_i^\phi = \sigma^\phi(\mathbf{r}_i)$ and the correlation, $C_{ij}^{\phi\phi} = C^{\phi\phi}(\mathbf{r}_i, \mathbf{r}_j)$. In conventional OI, the variance and correlation fields for the geopotential error are specified empirically as seen below, and the remainder of the cross correlations is specified by imposing the geostrophic constraint; that is, by assuming that the prediction error fields are geostrophically balanced.

In conventional applications of OI, we assume the field of standard deviation of geopotential errors to be independent of the coordinates in the horizontal $\mathbf{s} \boxminus (\lambda, \varphi)$. Hence, the standard deviations of geopotential errors

$$\sigma_i^\phi = \sigma^\phi(\mathbf{r}_i) = \sigma^\phi(p_i) \tag{8.53}$$

are taken as a function of the pressure (height) levels alone. Moreover, the geopotential–geopotential correlation field is considered to be horizontally homogeneous, and separable from the vertical components, that is,

$$C_{ij}^{f|\phi\phi} = C^{\phi\phi}(\mathbf{s}_i - \mathbf{s}_j) V^{\phi\phi}(p_i, p_j) \tag{8.54}$$

Here we recall the notion of homogeneous random fields introduced in Lecture 2. Finally, we impose the hypothesis of horizontal isotropy, so that we can write

$$C_{ij}^{\phi\phi} = C^{\phi\phi}(s_{ij} = |\mathbf{s}_i - \mathbf{s}_j|)\,. \tag{8.55}$$

The effects of the homogeneity hypothesis were carefully studied in Cohn & Morone [32], for the case of spherical geometry. In particular, these authors observed that, in certain cases, the hypothesis that the standard deviations are independent of the horizontal coordinate is responsible for up to 30% errors in the real value of the standard deviations. The separability hypothesis of the correlations field in the vertical is currently seen as one of the main barriers to accurately forecast dramatic atmospheric events such as strongly baroclinic systems. Recent research has concentrated in eliminating, or at least, relaxing some of these hypothesis. Examples of these efforts are the work of Bartello & Mitchell [7] in non–separable covariance fields, and those of Gaspari & Cohn [59] in specification of non–homogeneous correlation fields.

Focusing our attention on the conventional procedure of OI, consider the geostrophic balance relations among the variables u_i, v_i and ϕ_i:

$$u_i = \alpha_i \frac{\partial \phi_i}{\partial \varphi_i} \tag{8.56a}$$

$$v_i = \beta_i \frac{\partial \phi_i}{\partial \lambda_i}\,, \tag{8.56b}$$

for,

$$\alpha_i = -\frac{1}{f_i a} \tag{8.57a}$$

$$\beta_i = \frac{1}{f_i a \cos\varphi_i}\,, \tag{8.57b}$$

where a is the earth radius and $f_i = 2\Omega \sin \varphi_i$ is the Coriolis parameter. Notice that the expressions above apply only to mid-latitudes.

Assuming that the ensemble mean represents an approximation of the real variables $\mathbf{w}^t$, after applying the ensemble mean operator to (8.56), we can write

$$u_i^t = \mathcal{E}\{u_i\} = \alpha_i \frac{\partial \mathcal{E}\{\phi_i\}}{\partial \varphi_i} \tag{8.58a}$$

$$v_i^t = \mathcal{E}\{v_i\} = \beta_i \frac{\partial \mathcal{E}\{\phi_i\}}{\partial \lambda_i} \tag{8.58b}$$

and subtracting (8.58) from (8.56) we obtain the geostrophic relation among the errors in the variables u, v and ϕ:

$$\tilde{u}_i = \alpha_i \frac{\partial \tilde{\phi}_i}{\partial \varphi_i} \tag{8.59a}$$

$$\tilde{v}_i = \beta_i \frac{\partial \tilde{\phi}_i}{\partial \lambda_i} . \tag{8.59b}$$

From these relations it follows that

$$S_{ij}^{f|u\phi} = \mathcal{E}\{\tilde{u}_i \tilde{\phi}_j\} = \alpha_i \frac{\partial}{\partial \varphi_i} \mathcal{E}\{\tilde{\phi}_i \tilde{\phi}_j\} , \tag{8.60a}$$

$$S_{ij}^{f|\phi u} = \mathcal{E}\{\tilde{\phi}_i \tilde{u}_j\} = \alpha_j \frac{\partial}{\partial \varphi_j} \mathcal{E}\{\tilde{\phi}_i \tilde{\phi}_j\} , \tag{8.60b}$$

$$S_{ij}^{f|v\phi} = \mathcal{E}\{\tilde{v}_i \tilde{\phi}_j\} = \beta_i \frac{\partial}{\partial \lambda_i} \mathcal{E}\{\tilde{\phi}_i \tilde{\phi}_j\} , \tag{8.60c}$$

$$S_{ij}^{f|\phi v} = \mathcal{E}\{\tilde{\phi}_i \tilde{v}_j\} = \beta_j \frac{\partial}{\partial \lambda_j} \mathcal{E}\{\tilde{\phi}_i \tilde{\phi}_j\} , \tag{8.60d}$$

$$S_{ij}^{f|uv} = \mathcal{E}\{\tilde{u}_i \tilde{v}_j\} = \alpha_i \beta_j \frac{\partial}{\partial \varphi_i \partial \lambda_j} \mathcal{E}\{\tilde{\phi}_i \tilde{\phi}_j\} , \tag{8.60e}$$

$$S_{ij}^{f|vu} = \mathcal{E}\{\tilde{v}_i \tilde{u}_j\} = \alpha_j \beta_i \frac{\partial}{\partial \lambda_i \partial \varphi_j} \mathcal{E}\{\tilde{\phi}_i \tilde{\phi}_j\} , \tag{8.60f}$$

$$S_{ij}^{f|uu} = \mathcal{E}\{\tilde{u}_i \tilde{u}_j\} = \alpha_i \alpha_j \frac{\partial}{\partial \varphi_i \partial \varphi_j} \mathcal{E}\{\tilde{\phi}_i \tilde{\phi}_j\} , \tag{8.60g}$$

$$S_{ij}^{f|vv} = \mathcal{E}\{\tilde{v}_i \tilde{v}_j\} = \beta_i \beta_j \frac{\partial}{\partial \lambda_i \partial \lambda_j} \mathcal{E}\{\tilde{\phi}_i \tilde{\phi}_j\} , \tag{8.60h}$$

where all the covariances are written as a function of the error covariance function ϕ–ϕ given in (8.52). The complete forecast error covariance matrix can be written symbolically as

$$\mathbf{S}_{ij}^f = \begin{pmatrix} G_i^u \\ G_i^v \\ 1 \end{pmatrix} \left[\begin{pmatrix} G_j^u \\ G_j^v \\ 1 \end{pmatrix} S_{ij}^{f|\phi\phi} \right]^T \tag{8.61}$$

where G_m^u and G_m^v are differential operators defined as

$$G_m^u \equiv \alpha_m \frac{\partial}{\partial \varphi_m} \tag{8.62a}$$

$$G_m^v \equiv \beta_m \frac{\partial}{\partial \lambda_m} \tag{8.62b}$$

for $m = i, j$.

By means of a limiting procedure, it is possible to show (see Cohn & Morone [32]) that the forecast error standard deviations for the winds are given by

$$\sigma_i^u = \sigma_i^\phi \, |\alpha_i| \left[\lim_{p_j \to p_i} \frac{\partial \log C^{\phi\phi}}{\partial \varphi_i \partial \varphi_j} \right]^{1/2}, \qquad (8.63a)$$

$$\sigma_i^v = \sigma_i^\phi \, |\beta_i| \left[\lim_{p_j \to p_i} \frac{\partial \log C^{\phi\phi}}{\partial \lambda_i \partial \lambda_j} \right]^{1/2}, \qquad (8.63b)$$

as a function of the geopotential height (co)variance at a point.

Consider a simple example the case of specifying the matrix $\mathbf{S}_k^f$ for on a plane atmosphere. This eliminates the need to specify the vertical correlations $V^{\phi\phi}$ in (8.54). Therefore, the state vector $\mathbf{w}^{f,a}$ in consideration is similar to that considered in the previous section, and in the end of the previous lecture, for the shallow water system of equations. To simplify the problem further, we treat here the case of a β–plane, with latitude and longitude represented by the dependent variables x and y, respectively. In this case, we notice that the constants α_i and β_i defined above are substituted by

$$\alpha_i = -\frac{1}{f_i}, \qquad (8.64a)$$

$$\beta_i = \frac{1}{f_i}, \qquad (8.64b)$$

where now $f_i = f_0 + \beta y_i$ is the Coriolis parameter. In practice, the state vector is treated on a grid, and therefore the derivatives seen above should be interpreted as finite differences.

A common model for the geopotential–geopotential correlation function is the Gaussian model, that is,

$$C_{ij}^{\phi\phi} = \exp\left(-\frac{b}{2} s_{ij}^2\right), \qquad (8.65)$$

where, s_{ij} now is the distance between two points, (x_i, y_i) and (x_j, y_j) on the plane,

$$s_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2, \qquad (8.66)$$

and b is an empirical constant proportional to the inverse of the decorrelation distance. Therefore, the derivative of $C^{\phi\phi}$ with respect to the variable ξ_k can be written as:

$$\frac{\partial C_{ij}^{\phi\phi}}{\partial \xi_k} = -\frac{b}{2} C_{ij}^{\phi\phi} \frac{\partial s_{ij}^2}{\partial \xi_k} \qquad (8.67)$$

where ξ represents either x or y, and k represents either i or j. It is clear that, according to the definition (2.52), the covariance $\mathrm{S}^{f|\phi\phi}$ represents an isotropic field (consequently, homogeneous).

Substituting the expression for the distance s_{ij}^2 in (8.67), and expressions (8.60) we have

$$S_{ij}^{f|u\phi} \,/\, S_{ij}^{f|\phi\phi} = -\alpha_i \, (y_i - y_j) \, b, \qquad (8.68a)$$

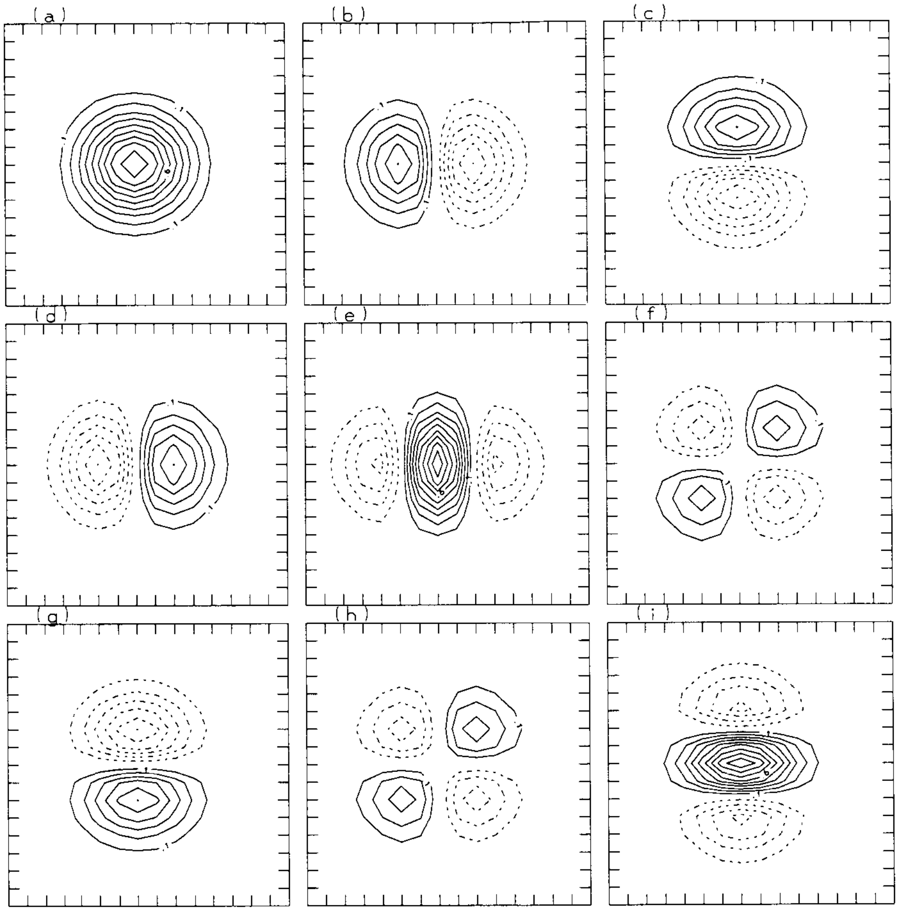

Figure 8.1: Point–correlations at the center of a square domain using the geostrophic balance relation; the equations obtained in this section: (a) ϕ–ϕ; (b) ϕ–v; (c) ϕ–u; (d) v–ϕ; (e) v–v; (f) v–u; (g) u–ϕ; (h) u–v; and (i) u–u.

$$S_{ij}^{f|\phi u} \,/\, S_{ij}^{f|\phi\phi} = \alpha_j \,(y_i - y_j)\, b\,, \tag{8.68b}$$

$$S_{ij}^{f|v\phi} \,/\, S_{ij}^{f|\phi\phi} = -\beta_i \,(x_i - x_j)\, b\,, \tag{8.68c}$$

$$S_{ij}^{f|\phi v} \,/\, S_{ij}^{f|\phi\phi} = \beta_j \,(x_i - x_j)\, b\,, \tag{8.68d}$$

$$S_{ij}^{f|uv} \,/\, S_{ij}^{f|\phi\phi} = \alpha_i \,\beta_j \,(x_i - x_j)\,(y_i - y_j)\, b^2\,, \tag{8.68e}$$

$$S_{ij}^{f|vu} \,/\, S_{ij}^{f|\phi\phi} = \alpha_j \,\beta_i \,(x_i - x_j)\,(y_i - y_j)\, b^2\,, \tag{8.68f}$$

$$S_{ij}^{f|uu} \,/\, S_{ij}^{f|\phi\phi} = \alpha_i \,\alpha_j \left[1 - b(x_i - x_j)^2\right] b\,, \tag{8.68g}$$

$$S_{ij}^{f|vv} \,/\, S_{ij}^{f|\phi\phi} = \beta_i \,\beta_j \left[1 - b(y_i - y_j)^2\right] b\,. \tag{8.68h}$$

Also, substituting (8.65) in the expressions (8.63), the standard deviation for the forecast errors in the winds becomes

$$\sigma_i^u = \sigma_i^\phi \sqrt{b}|\alpha_i|\,, \tag{8.69a}$$

$$\sigma_i^v = \sigma_i^\phi \sqrt{b}|\alpha_i|\,. \tag{8.69b}$$

All the quantities are now written as functions of the standard deviation of the forecast errors in geopotential heights. The correlation formulas obtained above are shown on Fig. 8.1 for a square cartesian plane, at its mid-point.

Exercises

1. Based on the shallow-water system considered in this chapter, while we studied the initialization problem, show that the quantity corresponding to the quasi-geostrophic potential vorticity:

$$q = \nabla^2\psi + f_0 - \frac{f_0}{\Phi}\phi$$

is conserved. Furthermore, using the Fourier transform introduced in (7.34), show that we can write

$$\hat{q} = k\hat{\psi} + \sqrt{k}\hat{\phi}$$

(Note: The quantity q above refers to the complete field, that is, basic state plus perturbations.)

2. Consider the shallow-water equations in Exercise 7.2. Assume these equations are applied to a doubling periodic channel; using Fourier transform for u', v' e ϕ', obtain the matrix $\hat{\mathbf{L}}$ for the corresponding system.

3. Balgovind et al. [6] proposed a simple model to describe the spatial structure of forecast geopotential error fields for time scales of one to two days. Based on barotropic potential vorticity conservation equation, these researchers arrived to the following stochastic equation for the errors $\tilde{\phi} = \tilde{\phi}(x, y)$ in the geopotential field, in a tangent plane:

$$L\tilde{\phi} = F(x, y)$$

where the operator L is defined as

$$L(x, y) \equiv (\nabla^2 - \alpha^2)$$

and ∇ is the Laplacian in two Cartesian dimensions x–y, x and y representing longitude and latitude, respectively, and $\alpha > 0$ being a constant related with the atmosphere depth. The term F is the stochastic forcing, with known statistics:

$$\begin{aligned} \mathcal{E}\{F(\mathbf{s})\} &= 0 \\ \mathcal{E}\{F(\mathbf{s}_1, \mathbf{s}_2)\} &= \sigma^2 \delta(\mathbf{s}_1 - \mathbf{s}_2) \end{aligned}$$

and represents the uncertainties in the model. Here, $\mathbf{s} = (x, y)$, $\mathbf{s}_1$, $\mathbf{s}_2$ are two points in the x-y plane, and σ is the noise F variance, assumed known. Moreover, define the error covariance function $P(\mathbf{s}_1, \mathbf{s}_2)$ as

$$P(\mathbf{s}_1, \mathbf{s}_2) \equiv \mathcal{E}\{\tilde{\phi}(\mathbf{s}_1)\tilde{\phi}(\mathbf{s}_2)\}\,.$$

Hence, show that:

(a) The equation for the spatial correlation structure ρ is

$$L_1 L_2 \rho(\mathbf{s}_1, \mathbf{s}_2) = \sigma \delta(\mathbf{s}_1 - \mathbf{s}_2)$$

For that, assume that the geopotential error variance field is identical to the forcing variance field F.

(b) Considering the unidimensional case, that is, when $\mathbf{s} \to x$, and assuming the correlation field is homogeneous, the correlation function ρ satisfies the following equation:

$$\left(\frac{d^2}{ds^2} - \alpha^2\right)^2 \rho(s) = \delta(s)$$

where $s \equiv |x_1 - x_2|$.

(c) The solution to the equation in the previous item can be written as:

$$\rho(x_1, x_2) = \rho(x_1 - x_2) = \frac{1}{4\alpha^3}\,(1 + \alpha|x_1 - x_2|)\; e^{-\alpha|x_1 - x_2|}$$

(Hint: Use symmetric Fourier transform.) This model for error correlations is sometimes called second order autoregressive, a more complicated version of which is utilized in some operational OI systems.

Chapter 9

Atmospheric Data Assimilation: Advanced Methods

9.1 Developments toward Advanced Methods

The tendency in operational data assimilation centers these days such as NCEP and ECMWF, is to evolve in the direction of eliminating certain hypothesis about approximating the evolution of forecast error covariances. As discussed previously, one of the most fundamental ingredients of the methods based on estimation theory is the propagation of error covariances by means of the dynamics of the system. If it were possible to calculate this evolution completely, the separability hypothesis between the vertical and horizontal correlations, geostrophic assumption, homogeneity, and isotropy hypothesis would not be necessary, because the dynamical properties would be present in the corresponding error covariance matrix; no artificial properties would have to be imposed by ad hoc constraints. Moreover, the error covariances would evolve instead of being stationary. However, we know that the calculation of the forecast error covariance is impractical due to the large computational burden. The computing progress of the past few years has been very promising. As a consequence, it is becoming possible to develop methods that allow for a slow relaxation of many of these conventional hypothesis. Some data assimilation systems today are designed with the goal of allowing easy progress and implementation of improvements on the error covariance structure. An example of such a versatile system is the Physical-Space Statistical Analysis System (PSAS; da Silva et al. [37] e Guo & da Silva [68], of the Goddard Space Flight Center. The procedure to relax the hypothesis in the construction of error covariances is referred to as construction of *error covariance models*, as each premise that gets eliminated, or substituted by less restrictive premises, generates a new error covariance model.

In this section we describe two ideas for relaxing some of the conventional constraints imposed on modeling error covariances. Initially, we consider a way of eliminating the *geostrophic* dynamical constraint, by presenting a way of building error covariance with coherent balances from the governing dynamics, for simple dynamics. Later, we consider a model that eliminates the *separability* hypothesis between vertical and horizontal correlations. Also in this Lecture, we present the basic ideas of what is called *parameterization* of error covariance and how to calibrate the quantities involved in these parameterizations. Finally, we discuss an application of the Kalman filter for a simple dynamical system, and some approximations to the Kalman filter with possible practical potential.

9.1.1 Generation of Balanced Cross–Covariances

The imposition of the geostrophic balance dynamic constraint used in OI is certainly problematic since the equations of motion described in Lecture 7 do not satisfy this constraint exactly, but only approximately. Consequently, the forecast error covariance matrix which corresponds to the complete system of governing equations is only approximately geostrophic. When the geostrophic approximation is used to construct a forecast error covariance $\mathbf{S}_k^f$, the corresponding gains obtained through (8.48) have components in the rapid modes of the system, consequently generating a analyses $\mathbf{w}_k^a$ initialized incorrectly, that is, containing rapid waves, that degrade the quality of the forecasts. One of the ways to avoid this type of inconsistency is to develop a procedure that combines the real intrinsic balances of the governing dynamics with balances imposed in the forecast error covariance matrix by means of the slow modes of the system, without them necessarily being geostrophic modes. The generation of balanced covariances in this form suggests the possibility of eliminating the initialization stage of the analysis produced at a given instant of time. This fact was initially observed in the implementation of the analysis system of NCEP, the so called "Spectral Statistical–Interpolation" (SSI) developed by Parrish & Derber [112], using the linear balance relation, instead of the geostrophic relation. Similarly, the European system at ECMWF substitutes the geostrophic relation by a more general balance obtained through the Hough modes (see Heckley et al. [74]). What we describe below is a simplified procedure, analogous to this latter one.

To describe this type of procedure we will follow the treatment of Todling & Cohn [129], remembering that this procedure can be extended to the nonlinear case in which the dynamics is that of a general circulation model. In the simple cases to be considered here, the governing dynamics that we have in mind is linear with n degrees of freedom and consequently n normal modes; n_S modes classified as slow. Moreover, we take a simple system, such as that from the discretized shallow water equations on the plane. Therefore, the system variables are the zonal and meridional winds u and v, respectively, and the heights h, at each grid point.

In this case, as in OI, the idea is to base the construction the error covariance matrix $\mathbf{S}^f$, omitting the index k referring to the time t_k, by specifying only the error forecast error covariance matrix for the height fields, designated by $\mathbf{S}^{f|hh}$. Here we use the same notation as in the previous lecture, with the following difference: now the height–height forecast error covariance is a matrix of dimension $n/3 \times n/3$, instead of a function. This is simply due to the fact that we are now assuming that the governing equations have been discretized in some way. Therefore, not only the height–height forecast error covariance is a matrix, but also all the other covariances and cross–covariances in $\mathbf{S}^f$ are matrices. That is, the decomposition (8.50) is redefined as

$$\mathbf{S}^f = \begin{pmatrix} \mathbf{S}^{f|uu} & \mathbf{S}^{f|uv} & \mathbf{S}^{f|uh} \\ \mathbf{S}^{f|vu} & \mathbf{S}^{f|vv} & \mathbf{S}^{f|vh} \\ \mathbf{S}^{f|hu} & \mathbf{S}^{f|hv} & \mathbf{S}^{f|hh} \end{pmatrix} \tag{9.1}$$

where now all the elements of $\mathbf{S}^f$ are written in bold face because they are matrices of dimension $n/3 \times n/3$. Notice here we use height instead of the geopotential function, without loss of generality.

We should recognize the fact that not every error covariance matrix hh (referring to height–height), $\mathbf{S}^{f|hh}$, can be a block corresponding to a "slow" multivariate error covariance matrix $\mathbf{S}^f$. This is easily understood if we notice that in general $\mathbf{S}^{f|hh}$ has dimension $n/3 \times n/3$, but n_S can be less than $n/3$. If $\mathbf{S}^{f|hh}$ should be part of a slow matrix $\mathbf{S}^f$, then, it can be at most rank n_S. Thus, we have to establish from the start the "slowness" of $\mathbf{S}^{f|hh}$, and later build the rest of the error covariance matrix, that is , the cross covariance matrices so that the resulting $\mathbf{S}^f$ is slow. This leads to an expression similar to the one introduced in (8.61), but in which the operators that act on $\mathbf{S}^{f|hh}$ carry the fact that $\mathbf{S}^f$ is "slow" instead of geostrophic.

As in the shallow water problem considered in Lectures 6 and 7, we assume that the normal modes of the system in question can be collected as the columns of a matrix $\mathbf{V}$, of dimension $n \times n$. The matrix formed by n_S slow vectors is represented by $\mathbf{V}_S$ and has dimension $\times n_S$. It is necessary to partition this last matrix as

$$\mathbf{V}_S = \begin{pmatrix} \mathbf{V}_u \\ \mathbf{V}_v \\ \mathbf{V}_h \end{pmatrix} \tag{9.2}$$

where each block has dimension $n/3 \times n_S$, and where for example, $\mathbf{V}_h$ represents the matrix of the height components of slow normal modes. Furthermore, we designate by $\mathcal{S}_h$ the subspace that spanned by the n_S columns of the matrix $\mathbf{V}_h$.

First, we want to modify any covariance matrix $\mathbf{S}^{f|hh}$ to make it "slow". This problem can be arranged as follows. Find a matrix $\mathbf{X}$ that satisfies the following conditions:

$$\mathit{Range}\, \mathbf{X} \in \mathcal{S}_h\,, \tag{9.3}$$

and

$$\mathbf{X}^T = \mathbf{X}\,, \tag{9.4}$$

minimizing at the same time the scalar functional

$$\eta \equiv ||\mathbf{X} - \mathbf{S}^{f|hh}||_F^2 \tag{9.5}$$

The condition (9.3) imposes that the columns of $\mathbf{X}$ belong to the subspace $\mathcal{S}_h$, i.e., that the height components are a linear combination of slow modes. The second condition imposes that $\mathbf{X}$ be symmetric, since it has to represent a covariance matrix. We do not demand above that this matrix be positive semi–definite, but we will see that the solution to the problem (9.3)–(9.5) is positive semi–definite, and therefore it can represent a covariance matrix. The last condition above demands that $\mathbf{X}$ be as close as possible to $\mathbf{S}^{f|hh}$ in the Frobenius norm $||.||_F$, which is defined as the sum of the squares of the elements of a matrix (Sec Golub & Van Loan [67]). This is a norm that penalizes equally each element of the difference between $\mathbf{S}^{f|hh}$ and $\mathbf{X}$. Another norm could have been chosen, as for example a norm with weights, but the Frobenius norm leads to a very simple solution.

As we seen below, the solution of the problem (9.3)–(9.5) is unique and given by

$$\mathbf{X} = \mathbf{\Pi}_h \mathbf{S}^{f|hh} \mathbf{\Pi}_h^T \,, \tag{9.6}$$

where $\mathbf{\Pi}_h$ is the orthogonal projector for the subspace $\mathcal{S}_h$, that is,

$$\mathbf{\Pi}_h = \mathbf{V}_h (\mathbf{V}_h^T \mathbf{V}_h)^{-1} \mathbf{V}_h^T \,. \tag{9.7}$$

The solution, $\mathbf{X}$ in (9.6) is positive semi–definite, since it corresponds to a congruence transformation of a positive semi–definite matrix $\mathbf{S}_k^{f|hh}$. Also, notice that it is not possible for the solution to be positive definite, because $\mathbf{\Pi}_h$ has rank n_S, since the matrix is invertible $\mathbf{V}_h^T \mathbf{V}_h$ in (9.7) has dimension $n_S \times n_S$; consequently $\mathbf{X}$ is at most rank n_S.

We want to show that $\mathbf{X}$ given in (9.6) is the unique solution of (9.5). For this, notice that as $\mathbf{\Pi}_h$ is the orthogonal projector on $\mathcal{S}_h$, it satisfies the following condition:

$$\mathit{Range}\, \mathbf{\Pi}_h = \mathcal{S}_h \,, \tag{9.8a}$$
$$\mathbf{\Pi}_h^2 = \mathbf{\Pi}_h \,, \tag{9.8b}$$
$$\mathbf{\Pi}_h^T = \mathbf{\Pi}_h \,. \tag{9.8c}$$

Equation (9.3) is satisfied by $\mathbf{X}$, given the condition (9.8a); moreover (9.4) it is obviously satisfied. Now we show that $\mathbf{X}$ given in (9.6) minimizes (9.5), in a unique way.

For the moment, let us denote by $\mathbf{X}_*$ the solution given in (9.6), i.e.,

$$\mathbf{X}_* \equiv \mathbf{\Pi}_h \mathbf{S}^{f|hh} \mathbf{\Pi}_h^T \,. \tag{9.9}$$

Then, any $\mathbf{X}$ which satisfies (9.3) and (9.4) should be of form

$$\mathbf{X} = \mathbf{X}_* + \tilde{\mathbf{X}} \,, \tag{9.10}$$

where

$$\mathit{Range}\, \tilde{\mathbf{X}} \in \mathcal{S}_h \,, \tag{9.11}$$

and

$$\tilde{\mathbf{X}}^T = \tilde{\mathbf{X}} \,. \tag{9.12}$$

To show that $\mathbf{X}_*$ is the only minimizer, we need to show that $\tilde{\mathbf{X}} = \mathbf{0}$ at the minimum.

Substituting (9.10) in (9.5) we have

$$\begin{aligned} \eta = \eta(\mathbf{X}) &\equiv ||\mathbf{X} - \mathbf{S}^{f|hh}||_F^2 \\ &= \mathrm{Tr}\,(\mathbf{X} - \mathbf{S}^{f|hh})^T (\mathbf{X} - \mathbf{S}^{f|hh}) \\ &= \mathrm{Tr}\,(\mathbf{X}_* - \mathbf{S}^{f|hh} + \tilde{\mathbf{X}})^T (\mathbf{X}_* - \mathbf{S}^{f|hh} + \tilde{\mathbf{X}}) \\ &= \eta(\mathbf{X}_*) + ||\tilde{\mathbf{X}}||_F^2 + 2\delta \,, \end{aligned} \tag{9.13}$$

where we use the definition of the Frobenius norm, as well as the definition of the *trace* operator (see Golub & Van Loan [67], pp. 56 and 332), and where we write

$$\delta = \mathrm{Tr}\,\tilde{\mathbf{X}}^T(\mathbf{X}_* - \mathbf{S}^{f|hh}) . \tag{9.14}$$

From expressions (9.11) and (9.12) it follows that $\tilde{\mathbf{X}}$ should be of form

$$\tilde{\mathbf{X}} = \mathbf{\Pi}_h \mathbf{Y} \mathbf{\Pi}_h^T , \tag{9.15}$$

for any symmetric matrix $\mathbf{Y}$. Substituting the expression above and (9.9) in (9.14) we obtain

$$\begin{aligned} \delta &= \mathrm{Tr}\,\mathbf{\Pi}_h^T \mathbf{Y} \mathbf{\Pi}_h (\mathbf{\Pi}_h \mathbf{S}^{f|hh} \mathbf{\Pi}_h^T - \mathbf{S}^{f|hh}) \\ &= \mathrm{Tr}\,\mathbf{Y} \mathbf{\Pi}_h (\mathbf{\Pi}_h \mathbf{S}^{f|hh} \mathbf{\Pi}_h^T - \mathbf{S}^{f|hh}) \mathbf{\Pi}_h^T , \end{aligned} \tag{9.16}$$

where we use (9.8c) and the fact that

$$\mathrm{Tr}\,\mathbf{A}^T\mathbf{B} = \mathrm{Tr}\mathbf{B}\,\mathbf{A}^T , \tag{9.17}$$

for any two matrices $\mathbf{A}$ and $\mathbf{B}$ with same dimensions. Using (9.8b) in (9.16) we have

$$\begin{aligned} \delta &= \mathrm{Tr}\,\mathbf{Y} (\mathbf{\Pi}_h \mathbf{S}^{f|hh} \mathbf{\Pi}_h^T - \mathbf{\Pi}_h \mathbf{S}^{f|hh} \mathbf{\Pi}_h^T) \\ &= 0 , \end{aligned} \tag{9.18}$$

and therefore (9.13) can be written as:

$$\begin{aligned} \eta(\mathbf{X}) &= \eta(\mathbf{X}_*) + ||\tilde{\mathbf{X}}||_F^2 \\ &\geq \eta(\mathbf{X}_*) , \end{aligned} \tag{9.19}$$

where the sign of equality prevails if, and only if, $\tilde{\mathbf{X}} = \mathbf{0}$, since $||\tilde{\mathbf{X}}||_F$ is canceled, and if and only if, $\tilde{\mathbf{X}} = \mathbf{0}$. Then, $\mathbf{X}_*$ minimizes η, in a unique way, completing the demonstration.

Combining the expressions (9.6) and (9.7) we have that

$$\mathbf{X} = \mathbf{V}_h \hat{\mathbf{X}} \mathbf{V}_h^T , \tag{9.20}$$

where

$$\hat{\mathbf{X}} \equiv (\mathbf{V}_h^T\mathbf{V}_h)^{-1}\mathbf{V}_h^T\mathbf{S}^{f|hh}\mathbf{V}_h(\mathbf{V}_h^T\mathbf{V}_h)^{-1} \tag{9.21}$$

is a symmetric matrix of dimension $n_S \times n_S$. It is not difficult to observe that any slow covariance matrix $\mathbf{S}$ should be of form

$$\mathbf{S} = \mathbf{V}_S \hat{\mathbf{S}} \mathbf{V}_S^T , \tag{9.22}$$

for any symmetric matrix $\hat{\mathbf{S}}$, of dimension $n_S \times n_S$. In this case, $\hat{\mathbf{S}}$ is the representation of $\mathbf{S}$ in the space of normal modes. Comparing (9.2), (9.20) and (9.22), it follows that

$$\mathbf{S}^f \equiv \mathbf{V}_S \hat{\mathbf{X}} \mathbf{V}_S^T \tag{9.23}$$

is a unique slow covariance matrix for which the covariance block *hh* coincides with the matrix of slow *hh* covariance $\mathbf{X}$. In this way, the formulas (9.21) and (9.23) provide a way of building a dynamically balanced slow error covariance matrix $\mathbf{S}^f$, given general height error covariance matrix $\mathbf{S}^{f|hh}$. The matrix $\hat{\mathbf{X}}$ is a representation in the space of normal modes of $\mathbf{S}^f$. Equations (9.21) and (9.23) replace the construction of geostrophically balanced covariances through (8.61).

9.1.2 A Non–Separable Covariance Model

Let us consider now the case of abandoning the vertical separability hypothesis that has been mentioned in the previous lecture. This hypothesis is known to be responsible for the failure of data assimilation system in producing analysis capable of forecast regions of rapid vertical atmospheric motion. The baroclinic instability involves such motions and it is one of the main atmospheric instabilities. Assimilation systems currently in operations still largely underestimate these atmospheric instabilities due to their poorly prescribed forecast error covariance models. Since these models account for no correlation among fields at different vertical levels, what happens in those cases is that the information provided by observations at a certain levels of the atmosphere is not correctly transferred to other levels because of the vertical separability assumptions. The treatment of this section is due to Bartello & Mitchell [7], and it has the goal of building error covariance functions for which the vertical and horizontal relations are entirely determined by the dynamics, and therefore being non-separable. What these authors proposed is based on a system of simplified equations, analogously to what its done in OI, where the covariance structure is build on the basis of the geostrophic balance relation. As simple as it might be, Bartello & Mitchell's model is an extremely promising one.

Following Bartello & Mitchell's description, we consider the system of primitive equations linearized about a basic state with buoyancy fluctuation $\nu = \sqrt{g^2/(c_pT)}$, independent of height, where c_p is the the constant of specific heat to constant pressure for the dry air. Moreover, the vertical coordinate is taken as the pressure, so that $Z = -H\ln(p/p_s)$, where $H = RT/g$ is the height scale, $p_s = 1000$ mb is the pressure at the surface, and R, T, and g has the same meanings as in Lecture 7. The basic state is that of rest, and the Coriolis parameter $f = f_0$ is taken to be constant.

The equations of motion in this case (see Holton [82], Section 11.3) are given by

$$\frac{\partial u}{\partial t} - f_0 v + \frac{\partial \phi}{\partial x} = 0 \tag{9.24a}$$

$$\frac{\partial v}{\partial t} + f_0 u + \frac{\partial \phi}{\partial y} = 0 \tag{9.24b}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial W}{\partial Z} - \frac{W}{H} = 0 \tag{9.24c}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \phi}{\partial Z}\right) + \nu^2 W = 0 \tag{9.24d}$$

The boundary conditions on top and at the surface of the atmosphere can be written as:

$$W = \frac{dZ}{dt} = -\frac{\delta}{g}\frac{\partial \phi}{\partial t} \tag{9.25}$$

for $Z = 0$ e $Z = \bar{Z}$, where $\bar{Z}$ represents the top of the atmosphere. Here, $\delta = 0$ corresponds to a zero vertical speed.

We can eliminate W, partially, from the equations above by substituting (9.24d) in (9.24c), that is,

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = \frac{1}{\nu^2}\left(\frac{\partial}{\partial Z} - \frac{1}{H}\right)\frac{\partial}{\partial t}\left(\frac{\partial \phi}{\partial Z}\right)$$

$$= \frac{1}{\nu^2}\frac{\partial}{\partial t}\left(\frac{\partial^2}{\partial Z^2} - \frac{1}{H}\frac{\partial}{\partial Z}\right)\phi \tag{9.26}$$

where we say partially since the boundary conditions are still given as a function of W.

Let us look for solutions of the form

$$
\begin{aligned}
u &= U(x,y,t)A(Z) && (9.27a)\\
v &= V(x,y,t)A(Z) && (9.27b)\\
\phi &= \Phi(x,y,t)A(Z) && (9.27c)\\
W &= \Omega(x,y,t)B(Z) && (9.27d)
\end{aligned}
$$

Substituting these functions in the equations for u and v (9.24a) and (9.24b), respectively, it is simple to see that due to the fact that these equations do not involve the coordinate Z, they involve U, V and Φ only, that is,

$$
\begin{aligned}
\frac{\partial U}{\partial t} - f_0 V + \frac{\partial \Phi}{\partial x} &= 0 && (9.28a)\\
\frac{\partial V}{\partial t} + f_0 U + \frac{\partial \Phi}{\partial y} &= 0 && (9.28b)
\end{aligned}
$$

Now substituting (9.27) in (9.26) we have that

$$A\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right) = \frac{1}{\nu^2}\frac{\partial \Phi}{\partial t}\left(\frac{d^2 A}{dZ^2} - \frac{1}{H}\frac{dA}{dZ}\right) \tag{9.29}$$

or yet,

$$\left(\frac{\partial \Phi}{\partial t}\right)^{-1}\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right) = \frac{1}{\nu^2 A}\left(\frac{d^2 A}{dZ^2} - \frac{1}{H}\frac{dA}{dZ}\right) \tag{9.30}$$

Noticing that the left hand side of this equality is independent of Z, while the right hand side is independent of (x,y,t), we can separate this equation in two:

$$
\begin{aligned}
\frac{\partial \Phi}{\partial t} + c^2\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right) &= 0 && (9.31)\\
\left(\frac{d^2 A}{dZ^2} - \frac{1}{H}\frac{dA}{dZ}\right) + \frac{\nu^2}{c^2}A &= 0 && (9.32)
\end{aligned}
$$

where c^2 is the separability constant. The expression (9.32) is the vertical structure equation.

To re–write the boundary conditions (9.25) with solutions given by (9.27) we substitute (9.25) in (9.24d)

$$\frac{\partial}{\partial t}\left(\frac{\partial \phi}{\partial Z}\right) = \nu^2 \frac{\delta}{g}\frac{\partial \phi}{\partial t} \tag{9.33}$$

for $Z = 0$ e $Z = \bar{Z}$. And therefore, using (9.27) we have that

$$\frac{dA}{dZ} = \frac{\delta \nu^2}{g} A \tag{9.34}$$

for $Z = 0$ e $Z = \bar{Z}$.

The vertical structure equation (9.32), with boundary conditions (9.34), is a Sturm–Liouville problem (e.g., Arfken [5]) which can be solved without difficulty. Due to the boundary conditions, the constant c^2 can assume only certain discrete values c_n^2, with corresponding solutions $A(Z) = A_n(Z)$. Through the orthogonality properties of the solutions of the Sturm–Liouville problem we can write

$$\int_0^{\bar{Z}} A_n(Z) A_m(Z) e^{-Z/H}\, dZ = \delta_{nm}, \tag{9.35}$$

where δ_{nm} is a Kronecker delta.

Furthermore, the equations (9.28a), (9.28b), (9.31) for U, V, and Φ, respectively, representing the horizontal structure can be solved by the applying Fourier transform, in analogy to what we did in Lecture 7. Assuming that we are treating the case for the infinite plane, U, V and Φ can be written as

$$\begin{pmatrix} U \\ V \\ \Phi \end{pmatrix} (\mathbf{r}, t) = \int_{R^2} \begin{pmatrix} \hat{U} \\ \hat{V} \\ \hat{\Phi} \end{pmatrix} (\mathbf{s}, t)\, e^{-i\mathbf{s}^T\mathbf{r}}\, d\mathbf{s} \tag{9.36}$$

where $\mathbf{r} = (x, y)$ and $\mathbf{s} = (k, \ell)$, and the Fourier transform is given as in (2.54). The coefficients $\hat{U}$, $\hat{V}$, $\hat{\Phi}$ can be determined by substituting the expression above in equations (9.28a), (9.28b), (9.31). So that the solution for u, v and ϕ can be put in form

$$\mathbf{w}(\mathbf{r}, Z, t) = \sum_{n=0}^{\infty} A_n(Z) \int_{R^2} \hat{\mathbf{w}}(\mathbf{s}, t)\, e^{-i\mathbf{s}^T\mathbf{r}}\, d\mathbf{s} \tag{9.37}$$

where $\mathbf{w} \equiv (u, v, \phi)^T$ and $\hat{\mathbf{w}} \equiv (\hat{u}, \hat{v}, \hat{\phi})^T$.

To build a covariance model we can follow a similar path to that in Section 8.4. Assume that the real state $\mathbf{w}^t(\mathbf{r}, Z, t)$, as well as the forecast state $\mathbf{w}^f(\mathbf{r}, Z, t)$, obey the same equations of motion (9.24a)–(9.24d). Therefore, due to the linearity we have that

$$\mathbf{e}^f(\mathbf{r}, Z, t) = \sum_{n=0}^{\infty} A_n(Z) \int_{R^2} \hat{\mathbf{e}}^f(\mathbf{s}, t)\, e^{-i\mathbf{s}^T\mathbf{r}}\, d\mathbf{s} \tag{9.38}$$

where $\mathbf{e}^f = \mathbf{w}^f - \mathbf{w}^t$ is the forecast error. The vertical structure functions are the same, for the errors as well as for the fields, since the errors follow the same separation of variables (9.27).

The error covariance matrix between two spatial points $(\mathbf{r}_i, Z_i)$ and $(\mathbf{r}_j, Z_j)$ can be found by calculating the outer product between error vectors $\mathbf{e}^f$ at two spatial points i and j, and by using the ensemble mean operator. That is,

$$\begin{aligned} \mathbf{S}^f(\mathbf{r}_i, Z_i, \mathbf{r}_j, Z_j, t) &\equiv \mathcal{E}\{\mathbf{e}^f(\mathbf{r}_i, Z_i, t)\, (\mathbf{e}^f(\mathbf{r}_i, Z_i, t))^*\} \\ &= \sum_{n,m=0}^{\infty} A_n(Z_i)\, A_m(Z_j) \\ &\quad \int_{R^2} \int_{R^2} \mathcal{E}\{\hat{\mathbf{e}}^f(\mathbf{s}_i, t)(\hat{\mathbf{e}}^f(\mathbf{s}_j, t))^*\}\, e^{i(\mathbf{s}_i^T\mathbf{r}_i - \mathbf{s}_j^T\mathbf{r}_j)}\, d\mathbf{s}_i\, d\mathbf{s}_j \end{aligned} \tag{9.39}$$

where $*$ represents the transpose conjugated. This expression can still be written as

$$\mathbf{S}^f(\mathbf{r}_i, Z_i, \mathbf{r}_j, Z_j, t) = \sum_{n,m=0}^{\infty} A_n(Z_i)\, A_m(Z_j)\, \bar{\mathbf{S}}^f(\mathbf{r}_i, \mathbf{r}_j, t) \tag{9.40}$$

where $\bar{\mathbf{S}}^f(\mathbf{r}_i, \mathbf{r}_j, t)$ is a horizontal covariance

$$\bar{\mathbf{S}}^f(\mathbf{r}_i, \mathbf{r}_j, t) = \int_{R^2} \int_{R^2} \mathcal{E}\{\hat{\mathbf{e}}^f(\mathbf{s}_i, t)(\hat{\mathbf{e}}^f(\mathbf{s}_j, t))^*\}\, e^{i(\mathbf{s}_i^T \mathbf{r}_i - \mathbf{s}_j^T \mathbf{r}_j)}\, d\mathbf{s}_i\, d\mathbf{s}_j \tag{9.41}$$

depending on time t. From the expression (9.40) we see that the covariance is non-separability for the horizontal and vertical components.

At this point Bartello & Mitchell [7] comment that this covariance model can be used to determine the complete covariance matrix, that is, the covariance functions $S^{f|uu}$, $S^{f|uv}$, etc., where homogeneity and isotropy hypothesis can be employed. Another possibility is to use only the block of the covariance matrix $\mathbf{S}^f$ corresponding to the height-height covariance function $S^{f|hh}$. The remainder of the covariances and cross-covariances can be determined by means of the geostrophic balance relation. This, in fact, simplifies calibration procedures that have to be used so that we obtain error covariances with relevant (physical) meaning for assimilation systems.

Imposition of the homogeneity and isotropy assumptions for $S^{f|hh}$ lead to

$$\bar{S}^{f|hh}(r, t) = 2\pi \int_0^{\infty} \hat{S}^{f|hh}(\omega)\, J_0(\omega r)\, \omega\, d\omega \tag{9.42}$$

where J_0 is the order-zero Bessel function, and $r = |\mathbf{r}_i - \mathbf{r}_j|$. This is identical to the result that we obtained in (2.66) when we discussed isotropic covariances in R^2.

Finally we can write

$$S^{f|hh}(r, Z_i, Z_j, t) = 2\pi \sum_{n,m=0}^{\infty} A_n(Z_i)\, A_m(Z_j) \int_0^{\infty} \hat{S}^{f|hh}(\omega)\, J_0(\omega r)\, \omega\, d\omega \tag{9.43}$$

is the complete height-height error covariance function. In practice, this function needs to be transformed into the matrix $\mathbf{S}^{f|hh}$), and adjusted to real data; for example, by means of fitting techniques, such as least squares. Details on practical implementations are discussed in the original work of Bartello & Mitchell [7].

9.1.3 Covariance Tuning

At this point it should be clear that modeling error covariances is fundamental in atmospheric data assimilation. Once a covariance model is constructed, for example for the forecast error covariance $\mathbf{S}^f$, we need to make the analytical model correspond to reality in some way. This is done in general by comparison with the data provided by the observational network and the model forecasts provided by the general circulation models. As mentioned in the previous section, one of the consistent ways of making the adjustment is by means of least squares methods. For schemes such as conventional OI, or models described in the previous section, the height–height error covariance matrix $\mathbf{S}^{f|hh}$ is modeled in some way and the remainder of the multivariate error covariance is obtained by some type of balance constraint. In the example of OI given in Section 8.4 the correlation distance is in general the parameter to be calibrated, or determined, by means of the comparison against data. For the model of the previous section, the basic state variables, such as the temperature T, and also the order of truncation of the sum and integral in (9.43), are parameters to be estimated in order to calibrate the error covariance statistics.

In this section we describe a way of estimating parameters in an error covariance model based on the ideas of maximum likelihood seen in Lecture 4. This procedure was suggested by Dee [41, 42] with the main intention of calibrating parameters in models for the model error covariance $\mathbf{Q}_k$ and of observation $\mathbf{R}_k$, in the context of advanced assimilation schemes like the Kalman filter, since the statistics of these errors is in general not well known. A particular application of this method is when we need to estimate parameters in the forecast error covariance $\mathbf{S}^f$ (see Dee [41, 42]).

Consider the case in which a error field is represented by the m_k–vector $\mathbf{v}_k$, in time t_k. We want to approximate the covariance matrix of these errors by a matrix $\mathbf{S}_k(\theta)$, where θ is an r–vector of parameters to be determined. We can write

$$\mathcal{E}\{\mathbf{v}_k\mathbf{v}_k^T\} \approx \mathbf{S}_k(\theta)\,. \tag{9.44}$$

In what follows we refer to the error vectors $\mathbf{v}_k$ as pseudo–innovations.

To calibrate $\mathbf{S}_k(\theta)$ based on samples (or realizations) of the pseudo–innovations vector $\mathbf{v}_k$, we assume that the errors represented by the error covariance matrix are normally distributed, with mean zero and covariance $\mathbf{S}_k(\theta)$,

$$\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_k(\theta_*))\ , \tag{9.45}$$

at least for some choice of the parameters in (9.44) so that $\theta = \theta_*$.

As we have seen in previous lectures, the assumption made above, together with the assumption that the pseudo–innovations $\{\mathbf{v}_k\}$, for $k = 1, 2, \cdots, K$, is an independent sequence, says that the conditional probability density $p_{\{\mathbf{v}_k\}|\theta}(\{\mathbf{v}_k\}|\theta) = p(\{\mathbf{v}_k\}|\theta)$ is given by the product of Gaussian densities

$$\begin{aligned} p(\{\mathbf{v}_k\}|\theta) &= \prod_{k=1}^{K} p(\mathbf{v}_k|\theta) \\ &= \prod_{k=1}^{K} \frac{1}{(2\pi)^{p_k/2}|\mathbf{S}_k(\theta)|^{1/2}} \\ &\quad \exp\left[-\frac{1}{2}[(\mathbf{v}_k - \boldsymbol{\mu}_k)^T\,\mathbf{S}_k^{-1}(\theta)\,(\mathbf{v}_k - \boldsymbol{\mu}_k)]\right]\,, \end{aligned} \tag{9.46}$$

where $\boldsymbol{\mu}_k = \mathcal{E}\{\mathbf{v}_k\}$.

Following the methodology of maximum likelihood estimation we can obtain an estimate θ_{ML} for θ_* as being the value that maximizes the conditional probability above, that is,

$$\theta_{\mathrm{ML}} = \arg\max_{\theta}\, p(\{\mathbf{v}_k\}|\theta) = \arg\min_{\theta}\, f(\theta) \tag{9.47}$$

where

$$f(\theta) \equiv \sum_{k=1}^{K} \left[\ln |\mathbf{S}_k(\theta)| + (\mathbf{v}_k - \boldsymbol{\mu}_k)^T \mathbf{S}_k^{-1}(\theta)\,(\mathbf{v}_k - \boldsymbol{\mu}_k)\right] \tag{9.48}$$

which is obtained by taking the natural logarithm of (9.46) and ignoring the constant term.

Assuming that the covariance model is stationary, that is,

$$\mathbf{S}_k(\theta) = \mathbf{S}(\theta) \tag{9.49}$$

the likelihood function takes the form

$$\begin{aligned} f(\theta) &= \sum_{k=1}^{K} \left[\ln |\mathbf{S}(\theta)| + (\mathbf{v}_k - \boldsymbol{\mu}_k)^T \mathbf{S}^{-1}(\theta)\,(\mathbf{v}_k - \boldsymbol{\mu}_k)\right] \\ &= K \ln |\mathbf{S}(\theta)| + \sum_{k=1}^{K} \mathrm{Tr}\left[(\mathbf{v}_k - \boldsymbol{\mu}_k)^T \mathbf{S}^{-1}(\theta)\,(\mathbf{v}_k - \boldsymbol{\mu}_k)\right] \end{aligned} \tag{9.50}$$

where we introduce the *trace* operator in the last equality, for convenience, after observing that the function f is scalar. Now, notice that

$$\begin{aligned} \sum_{k=1}^{K} \mathrm{Tr}\left[(\mathbf{v}_k - \boldsymbol{\mu}_k)^T \mathbf{S}^{-1}(\theta)\,(\mathbf{v}_k - \boldsymbol{\mu}_k)\right] &= \sum_{k=1}^{K} \mathrm{Tr}\left[\mathbf{S}^{-1}(\theta)\,(\mathbf{v}_k - \boldsymbol{\mu}_k)\,(\mathbf{v}_k - \boldsymbol{\mu}_k)^T\right] \\ &= \mathrm{Tr}\left[\mathbf{S}^{-1}(\theta) \sum_{k=1}^{K}(\mathbf{v}_k - \boldsymbol{\mu}_k)\,(\mathbf{v}_k - \boldsymbol{\mu}_k)^T\right] \end{aligned} \tag{9.51}$$

where we use the property of the trace that $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$, and the last equality is obtained by exchanging the order between the sum under k with the trace operator, since this last one is also a summation operation.

By defining $\bar{f} = f/K$ we have that

$$\bar{f}(\theta) \equiv \ln |\mathbf{S}(\theta)| + \mathrm{Tr}(\mathbf{S}^{-1}(\theta)\,\bar{\mathbf{S}}) \tag{9.52}$$

where $\bar{\mathbf{S}}$ is the sampling covariance matrix, or sample covariance:

$$\bar{\mathbf{S}} = \frac{1}{K} \sum_{k=1}^{K}(\mathbf{v}_k - \boldsymbol{\mu}_k)\,(\mathbf{v}_k - \boldsymbol{\mu}_k)^T \tag{9.53}$$

The function, defined in (9.52) is the one to be minimized so that we can determine the parameters θ. This can be done in practice by means of function minimization methods. Many of these methods need the gradient function (9.52), as discussed in Dee [41]. For a relatively small number of parameters θ, that is, when $r \sim o(1)$, and with a reasonable quantity of data $m \sim o(10^2)$, it is possible to obtain a good estimate of parameters, as indicated recently by recent work (Dee 1996, pers. communic.).

9.2 The Kalman Filter for a Simple Model

The Kalman filter was implemented by Cohn & Parrish [33] for a simple model of the atmosphere and we consider this case as an example for a data assimilation system in what follows. We consider the shallow–water equations linearized about a basic state with constant zonal velocity U and zero meridional velocity, and apply it to a β–plane. These equations are discretized with the finite difference scheme discussed in Lecture 7. The distinction from what we saw in Lecture 7, and the application now, is that the boundary conditions here, and in Cohn & Parrish [33], are only periodic in the East–West direction, with "solid walls" in the North–South direction, that is, the perturbations in meridional velocity are zero for all time along the North–South boundaries. This makes the morphology of the finite difference scheme somewhat different from that shown in Fig. 7.1. The extent of the domain of interest in this case is shown here in Fig. 9.1, and encompasses a region with a size equivalent to that of the contiguous United States. The necessary parameters to fully define the system and finite–difference are listed in Table 9.2.

Table 9.1: Shallow–water model parameters as in Cohn & Parrish [33].

Parameters	Values
I grid points in the zonal directions	25
J grid points in the meridional direction	16
East–West extent of channel, L_x	5000 km
North–South extent of channel, L_y	3000 km
Grid size, $\Delta x = \Delta y$	200 km
Time step, Δt	400 s
Coriolis parameter, f_0	6.15×10^{-5} s^{-1}
β–plane parameter, β	1.82×10^{-11} m^{-1}s^{-1}
Basic state geopotential height Φ_0	3×10^4 m^{2}s^{-2}
Basic state zonal speed U	25 ms^{-1}

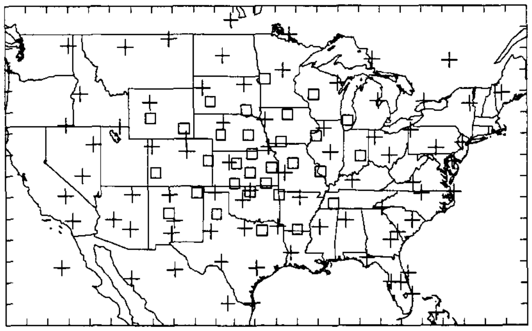

Figure 9.1: Domain of the model of Cohn & Parrish [33] encompassing the contiguous United State. The tick–marks indicate grid points; the "plus" signs indicate radiosonde observations; and the "squares" indicate the wind profilers.

To mimic a real data assimilation system, Cohn & Parrish [33] considered two observational networks existing in the region of study, just as indicated in Fig. 9.1. The observational network referred as the A–network is composed of 77 radiosondes that observe the winds and the mass field (heights) every 12 hours; and the observational system referred to a the B–network is composed of the radiosondes of the A–network plus 31 wind profilers, which observe the only winds every hour. The error standard deviation for each of these observing systems are shown in Table 9.2.

The equations corresponding to the Kalman filter are implemented for this system with the intention of studying the error evolution in an assimilation period of 2.5 days. Since, in this case, the system of governing equations is linear, the error covariance evolution is decoupled from the state estimate evolution. Therefore, as in Cohn & Parrish [33], we focus discussion only in the behavior of the error evolution, and ignore what happens with the states.

Table 9.2: Observational error standard deviations.

Observations	σ_u (ms^{-1})	σ_v (ms^{-1})	σ_h (m)	No. Obs.
Radiosondes	2.8	2.8	11	77
Wind Profilers	1.5	1.5	——	31

Fig. 9.2 shows the result of assimilation experiments using the observational networks A and B introduced above. The figure shows the time evolution of the error standard deviation, averaged over the domain, for all three variables of the system u, v and h. The plotted quantities correspond to the square root of the sum of the elements of the main diagonal of the forecast and/or analysis error covariance matrices $\mathbf{P}^{f,a}$, divided by the total number of grid points, for each one of the variables. The curves indicated by A refer to the results obtained when only the radiosonde data is assimilated. In this case, we see that at every 12 hours the curves display a jump, resulting in an instantaneous reduction of errors. These jumps correspond to the analysis times, when the radiosonde observations are processed by the filter. Between two consecutive observation intervals, the errors grow due to the presence of the model error, represented by the matrix $\mathbf{Q}_k$ of Lecture 5 (see Cohn & Parrish [33] for more details on this quantity). Notice further in Fig. 9.2 that the errors in the meridional velocity and heights are below the radiosonde observational error levels (indicated by the curves marked OLV; numbers listed in Table 9.2). The errors in the zonal winds do not fall below the observational error level, which is a particular property of the solution of the shallow–water equations.

When the wind profilers are present (curves indicated as B), we see that during two consecutive A–network observation periods, errors decrease every hour due to the assimilation of these wind profilers. The presence of these extra wind observations produce an overall reduction in the errors in all variables, including heights which are not directly observed by the B–network. The contribution of the wind profilers to reducing the height errors is a consequence of the fact that the analysis procedure of the Kalman filter is multivariate, and moreover, that the Kalman filter transfers the information content in the wind profilers observations appropriately to the height fields.

Fig. 9.3 shows the spatial distribution of the forecast error standard deviations after 2.5 days in the assimilation cycle. The contour maps are built from the square root of the

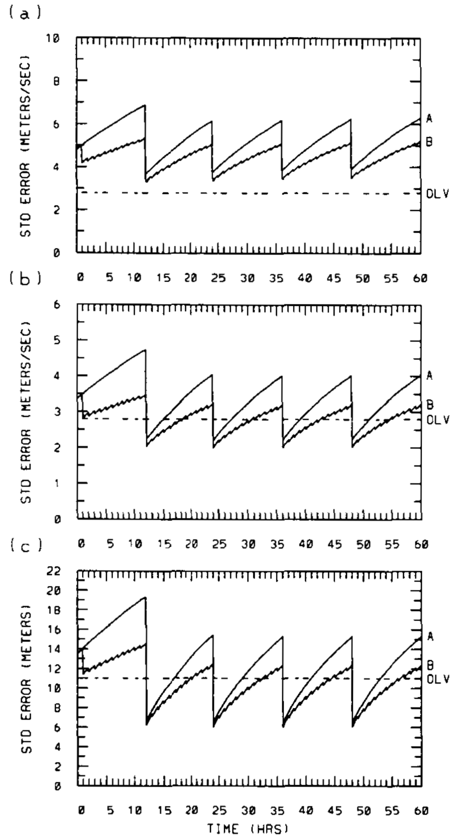

Figure 9.2: Results obtained with the Kalman filter: forecast and analysis expected error standard deviations averaged over the domain, for each of the three variables of the model, as a time function. The panels show the result for the errors in the (a) zonal velocity, (b) meridional velocity, and (c) height. Each curve is indicate by the observational system in question, that is, A for radiosondes and B for the radiosondes and wind profilers. The height errors are given in meters while the wind errors are in meters per second. The dotted lines, indicated by OLV, refer to the radiosondes observational error levels for each of the variables, according to Table 9.2.

elements of the main diagonal part of $\mathbf{P}_k^f$ for the winds and the height fields. The panels (a) refer to the experiment with the A–network, including only the radiosondes, while panels (b) refer to the experiment using both the A– and B–networks. In the case of the radiosonde–only assimilation [panels (a)] we see that the forecast error standard deviation practically uniformly distributed over the domain. This results from the fact that the radiosonde network is relatively uniformly distributed over the domain. The forecast error standard deviation in v, panel (a.2) has pronounced gradients near the North and South boundaries due to the boundary condition $v = 0$ along these boundaries. This boundary condition is equivalent to observing the variable v along of the North and South boundaries, without observation error. The presence of the dynamics allow for the consequent appearance of gradients along the boundaries in the zonal wind and height forecast error fields as seen in panels (a.1) e (a.2), respectively.

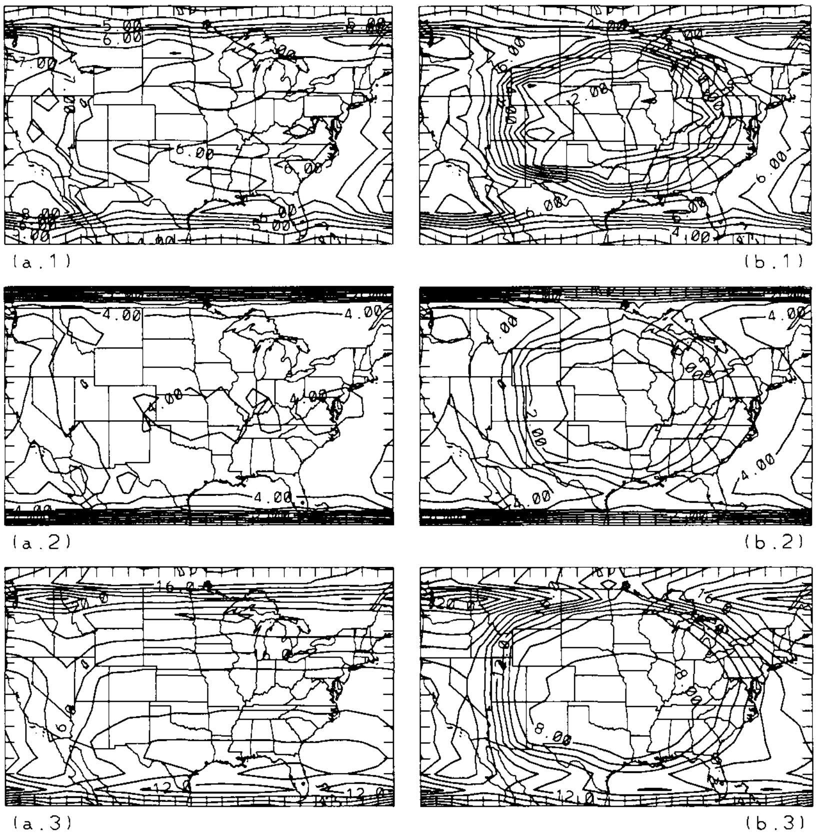

Figure 9.3: Spatial distribution of the forecast error standard deviation. Panels (a) refer to experiment A, while panels (b) refer to experiment (B): (a.1) and (b.1) for σ_u, (a.2) and (b.2) for σ_v, and (a.3) and (b.3) for σ_h. Contour interval is of 1 m for height errors, and of 0.5 m s^{-1} for wind errors.

The introduction of wind profilers, experiment with B–network, produces large gradients in the East–West direction, as seen in panels (b.1)–(b.3) of Fig. 9.3. These gradients result from the fact that the wind profilers are mainly located in the central region of the domain, contrary to the radiosondes, and therefore reflect the differences between regions of dense observation density and those of sparse observation density. We also see that in the panels referring to the experiment in question, the Western error gradients are more pronounced than the Eastern ones. In other words, the contours close to the Eastern boundary are more separated among themselves than those close to the Western boundary [more markedly noticeable in panels (b.1) and (b.3), for the forecast error standard deviations of u and h, respectively]. This is a sole consequence of the error propagation induced by the Kalman filter equations, which makes the wind profilers information in the central part of the domain be advected in the direction of the flow. Therefore, the forecast error standard deviation is smaller in the East than in the West side of the wind profilers region. Specifically, this error propagation is due to the calculation of the first term in the expression (5.17), for $\mathbf{P}_k^f$.

It is important to mention that current operational data assimilation systems do not possess the ability to propagate information in the pronounced way seen in the results of the experiment using the B–network. The reason is, as discussed previously in this and in the previous lecture, that the forecast error covariance matrix in operational systems is assumed stationary and prescribed in some way to cope with computational feasibility. The incorporation of a dynamic flavor in the forecast error covariance matrix for operational systems has been the object of a great number of basic research. We can imagine, by what we studied so far, that an intensive line of research in atmospheric data assimilation is the search for alternative ways to propagate error covariance that are computationally feasible and do not involve as many computations as those following the Kalman filter, or its nonlinear extensions.

In what follows here we concentrate on alternatives to simplify expression F2 in Table 5.3.1, for the case of linear filter. Once the more viable alternatives for the linear case are determined, it is possible to make extensions to the nonlinear case, which represent in fact the cases of practical interest in meteorology. The majority of existing approximations in the literature belong to one, or more, or the following categories (see Todling & Cohn [129] for more references and explanations):

- covariance error modeling (e.g., OI: Bergman [10], Gandin [57], Jiang & Ghil [85], Lorenc [95], McPherson et al. [102]; SSI: Parrish & Derber [112]; three–dimensional variational analysis (3D–Var): Andersson et al. [2], Heckley et al. [74], Pailleux [110], Vasiljević et al. [133]; PSAS: da Silva et al. [37])
- dynamic simplification (e.g., Dee [43], Todling & Cohn [129])
- reduced resolution (order resolution; e.g., Cohn & Todling [34], Fukumori [54], Fukumori & Malanotte–Rizzoli [55], Le Moyne & Alvarez [92], Verlaan & Heemink [135])
- local representation (e.g., Boggs et al. [16], Cohn [28, 29], Parrish & Cohn [113], Riedel [118])
- limiting filtering (e.g., Fu et al. [53], Fukumori et al. [56], Heemink [75], Heemink & Kloosterhuis [76], Hoang et al. [77])
- Monte Carlo approach (e.g., Leith [93], Evensen [51])

Some of these possibilities have been tested in the context of the the Kalman filter applied to the linear shallow–water equations with the goal of investigating its behaviors in comparison with the exact result provided by the Kalman filter. We describe briefly below the approximations considered in Todling & Cohn [129] for stable dynamics, and in Cohn & Todling [34] for stable and unstable dynamics. These approximations range from a simplified representation of the assimilation scheme by optimal interpolation [OI; item (a)], a somewhat improved version of OI which allows for advection of the height error covariance field by an advection operator **A** [HVA; section item (b)], to an even more sophisticated scheme that allows for the propagation of all height–height error covariance field by means of a simplified dynamics **A** [SKF; item (c)]. Since these schemes specify the height–height error covariance, it is necessary to use an algorithm to generate the missing covariances and cross–covariances. At this point we can impose the geostrophic balance constraint, however, as we have mentioned before, this does not generate good results, except in some cases. Alternatively, we can use the cross–variance generation algorithm studied previously in this lecture. Some of the schemes cited below use this procedure.

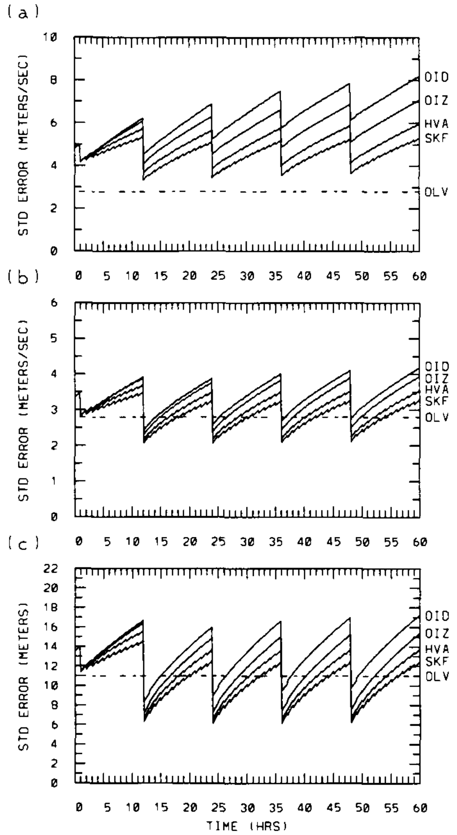

Figure 9.4: Analogous to Fig. 9.2, but only for the case of the B–network. The curves refer to the evaluation result of the performance of different approximations for the system of Cohn & Parrish. The approximations being evaluated are: balanced optimal interpolation with domain averaged error growth (OID); balanced optimal interpolation with latitudinal dependent error growth (OIZ); advection of balanced height error variance field (HVA); and advection of balanced height–height error covariance field (SKF). The Kalman filter results, that serve as the basis of comparison to these approximations are those indicated by the curves B in Fig. 9.2.

A summary of the results of evaluation of these approximations is presented in Fig. 9.4.

Considering the system of Cohn & Parrish [33], for the B–network case, the curves in the figure show the performance of two types of OI: OID, which uses a single constant value for the height error variance growth rate (elements along the diagonal of matrix $\mathbf{D}^h$ mentioned in item (a) below); and OIZ, which uses a height variance growth rate with latitudinal dependency. Also, the performance of the two other schemes, the HVA and the SKF, is concisely described in items (b) and (c) below. The figure shows the gradual improvement occurring when we increase the sophistication of the assimilation scheme, due to the gradual incorporation of dynamic information. In particular, the performance of the SKF scheme is practically indistinguishable from the exact result shown in Fig. 9.2 (curves B).

The way in which the schemes OI, HVA and SKF incorporate some dynamics (or not, in the case of OI), in the generation of forecast error covariances is very simplistic. To test these simplifications further an unstable dynamical system has been considered. Although the Kalman filter for unstable linear systems produces reliable results (see Ghil & Todling [65] and Todling & Ghil [130]), approximate assimilation schemes, based on the ideas described above, do not produce results equally reliable. In this way, alternative data assimilation schemes are necessary; some possibilities are briefly described in the last items (d) and (e) below. Both schemes described in these items are iterative. An Lanczos type algorithm (e.g, Golub & Van Loan [67]) is necessary to implement either schemes. The partial singular value decomposition filter (PSF) [item (d)] proposes to use the L leading singular modes of the propagator (tangent linear model) to propagate the analysis error covariance matrix; the partial eigendecomposition filter (PEF) [item (e)] proposes to generate only L leading eigenvalues/vector of the forecast error covariance matrix, $\mathbf{S}^p$. These two schemes are low–rank and information referring to the trailing part of the error covariances, which we designate by $\mathbf{S}^p_T$, should be provided in some way. That is, in these two cases there is a need to model trailing error covariances. These approximations are adequate also for stable dynamics.

(a) Optimal Interpolation

Category: Error Covariance Modeling

Forecast Error Covariance Matrix Partition $\mathbf{S}^f_k$:

$$\mathbf{S}^f_k \equiv \begin{bmatrix} \mathbf{S}^{f|uu} & \mathbf{S}^{f|uv} & \mathbf{S}^{f|uh} \\ \mathbf{S}^{f|vu} & \mathbf{S}^{f|vv} & \mathbf{S}^{f|vh} \\ \mathbf{S}^{f|hu} & \mathbf{S}^{f|hv} & \mathbf{S}^{f|hh} \end{bmatrix}_k$$

Block corresponding to the forecast error covariance (hh):

$$\mathbf{S}^{f|hh}_k = (\mathbf{D}^{f|h}_k)^{1/2}\mathbf{C}^{hh}(\mathbf{D}^{f|h}_k)^{1/2}$$

$\mathbf{D}^{f|h}_k$ is a diagonal matrix $n/3 \times n/3$ corresponding to the height error variance; $\mathbf{C}^{hh}$ is the height–height error correlation matrix $n/3\times n/3$, which is prespecified.

$\mathbf{D}^{f|h}_k$ is in general modeled to account for a linear variance error growth in time according to

$$\mathbf{D}^{f|h}_k = \mathbf{D}^{a|h}_{k-\ell} + \mathbf{D}^h$$

where ℓ is the number of model time steps between two consecutive analyses; $\mathbf{D}^h$ is a diagonal matrix corresponding to the error growth.

$\mathbf{C}^{hh}$ is prespecified assuming homogeneity and isotropy of the mass error field, and it is usually considered to be Gaussian. The rest of the error covariance matrix is obtained by the balanced covariance generation procedure described previously.

(b) Variance Evolution

Category: Local representation

Height error variance propagation:

$$\mathbf{D}_k^{p|h} = \mathbf{A}_{k,k-\ell}\mathbf{D}_{k-\ell}^{a|h}$$

where $\mathbf{A}_{k,k-\ell}$ represents the operator of an advection scheme.

Construction of the height–height error covariance matrix:

$$\mathbf{S}_k^{p|hh} = (\mathbf{D}_k^{p|h})^{1/2}\mathbf{C}^{hh}(\mathbf{D}_k^{p|h})^{1/2}.$$

The covariances remaining are calculated by means of the balanced covariance generation procedure discussed above .

(c) Simplified Kalman Filter

Category: dynamic simplification

Propagation of Height–height error covariance:

$$\mathbf{S}_k^{p|hh} = \mathbf{A}_{k,k-\ell}\,\mathbf{S}_{k-\ell}^{a|hh}\,\mathbf{A}_{k,k-\ell}^T$$

where $\mathbf{A}_{k,k-\ell}$ represents the operator of an advection scheme; the balanced covariances generation procedure is used for the remainder of the covariances.

(d) Partial Singular Value Decomposition Filter

Category: Local representation/reduced resolution

Forecast error covariance:

$$\mathbf{S}_k^f = \mathbf{S}_{k,k-\ell}^p + \tilde{\mathbf{Q}}_{k,k-\ell}$$

where $\mathbf{S}^p_{k,k-\ell}$ is the dynamically propagated part — predictability error covariance.

Consider the following singular value decomposition of the propagator $\mathbf{\Psi}_{k,k-\ell}$:

$$\mathbf{\Psi}_{k,k-\ell} = \left(\mathbf{U}\,\mathbf{D}\,\mathbf{V}^T\right)_{k,k-\ell}$$

and partition the matrices above in leading (L) and trailing (T) parts so that:

$$\mathbf{U}_{k,k-\ell} = [\mathbf{U}_L\,\mathbf{U}_T]_{k,k-\ell}\,, \qquad \mathbf{V}_{k,k-\ell} = [\mathbf{V}_L\,\mathbf{V}_T]_{k,k-\ell}$$

$$\mathbf{D}_{k,k-\ell} = \begin{bmatrix} \mathbf{D}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_T \end{bmatrix}_{k,k-\ell}$$

Consider the following model for $\mathbf{S}^p_{k,k-\ell}$:

$$\mathbf{S}^p_{k,k-\ell} = (\mathbf{S}^p_L + \mathbf{S}^p_T)_{k,k-\ell} = \left(\tilde{\mathbf{\Psi}}\,\mathbf{S}^a\,\tilde{\mathbf{\Psi}}^T + \mathbf{S}^p_T\right)_{k,k-\ell}$$

where

$$\tilde{\mathbf{\Psi}}_{k,k-\ell} = \left(\mathbf{U}_L\,\mathbf{D}_L\,\mathbf{V}^T_L\right)_{k,k-\ell}$$

so that,

$$\mathbf{S}^p_{L,k,k-\ell} = \left(\sum_{i=1}^{N_L}\sum_{j=1}^{N_L} d_i d_j \left(\mathbf{v}_i^T \mathbf{S}^a \mathbf{v}_j\right)\mathbf{u}_i \mathbf{u}_j^T\right)_{k,k-\ell}$$

where $d_i = diag(\mathbf{D}_L)_i$, $\mathbf{u}_i = col(\mathbf{U}_L)_i$, $\mathbf{v}_i = col(\mathbf{V}_L)_i$,
e $N_L = n^{\underline{o}}\,cols(\mathbf{V}_L)$.

$\mathbf{S}^p_T$ is specified by an adaptively tuned covariance model based on the innovations. Computational cost $\sim o(10L)$ model integrations.

(e) Partial Eigendecomposition Filter

Category: Local representation/reduced resolution

Forecast covariance error:

$$\mathbf{S}^f_k = \mathbf{S}^p_{k,k-\ell} + \tilde{\mathbf{Q}}_{k,k-\ell}$$

where $\mathbf{S}^p_{k,k-\ell}$ is the dynamically propagated part — predictability error covariance.

Consider the following eigendecomposition for the forecast error covariance $\mathbf{S}^p$:

$$\mathbf{S}^p_{k,k-\ell} = \left(\mathbf{\Psi}\,\mathbf{S}^a\,\mathbf{\Psi}^T\right)_{k,k-\ell} = \left(\mathbf{U}\mathbf{D}\mathbf{U}^T\right)_{k,k-\ell}$$

and partition the factors above in leading (L) and trailing (T) parts:

$$\mathbf{U}_{k,k-\ell} = [\mathbf{U}_L\,\mathbf{U}_T]_{k,k-\ell}\,,$$

$$\mathbf{D}_{k,k-\ell} = \begin{bmatrix} \mathbf{D}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_T \end{bmatrix}_{k,k-\ell}$$

Assume the following approximation for $\mathbf{S}^p_{k,k-\ell}$:

$$\mathbf{S}^p_{k,k-\ell} = (\mathbf{S}^p_L + \mathbf{S}^p_T)_{k,k-\ell} = \left(\mathbf{U}_L \, \mathbf{D}_L \, \mathbf{U}^T_L + \mathbf{S}^p_T\right)_{k,k-\ell}$$

$\mathbf{S}^p_T$ is specified for an adaptively tuned covariance model based on the innovations. Computational cost $\sim o(10L)$ model integrations.

The motivation to search for approximate schemes for covariance propagation, instead of ways of implementing an algorithm for complete error covariance propagation, goes beyond the fact that the latter is computationally infeasible. In fact, even if calculating the complete covariance evolution were (or comes to be) feasible, it would be a wast of computational resources, since:

- The governing equations are nonlinear, consequently only approximate schemes are possible — in general, we cannot calculate moments of all orders.

- Many observational systems involve nonlinear relations among the forecast and observed variables. Therefore, the same problem mentioned in the item above applies, i.e., we cannot calculate moments of all orders.

- Lack of knowledge of model and observation error statistics. This means that, at best we have to approximate, model and parameterize these quantities, forcing the assimilation system in to a sub–optimal situation.

- The number of available observations in a given time is relatively close to the number of degrees of freedom of the system, and it tends to become even greater. This means that the equations for the analysis step, that is, equations F3, F5 of Table 5.3.1, or their nonlinear equivalents, should be approximated in some way, due to the excess of computational effort required to solve them exactly.

Even the promising developments in parallel computation [78, 79, 100], and the possibility of an increase in computational capability, will not be sufficient to solve the covariance propagation equation completely. This because the tendency in meteorology has been to increase the resolution of general circulation models, whenever there is an increase in computer power. Thus, pushing computers to their limit, just to produce a forecast, leaving little room for forecasting the statistics . It will always be necessary to have approximate models for error covariance propagation, to make operational systems for atmospheric data assimilation computational feasible and practical.

9.3 The Fixed–Lag Kalman Smoother

9.3.1 Theory

Let us address the issue of improving an estimate provided by the Kalman filter by making use of observations *past* the analysis time. Let us use the Bayesian approach for this purpose, but to keep the derivation simple we consider only the problem of improving the estimate at time t_{k-1} given the observations at just one time step ahead that time, that is, at time t_k. This constitutes the lag–1 smoother problem (this can be identified with the fixed–point smoother). We could extend the problem to that of improving the filter state estimate at time $t_{k-\ell}$ using observations up to and including time t_k. This is the general, lag–ℓ[1], fixed–lag smoother problem. The solution to the general problem can be found in a variety of texts, as for example, Anderson & Moore [1] and Meditch [103].

In complete analogy to the previous section we can seek for the minimum variance estimate, which written in terms of the conditional mean is

$$\begin{aligned}
\mathbf{w}^a_{k-1|k} &= \mathcal{E}\{\mathbf{w}^t_{k-1}|\mathbf{W}^o_k\} \\
&= \int_{-\infty}^{+\infty} \mathbf{w}^t_{k-1} p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_k)\, d\mathbf{w}^t_{k-1} \qquad (9.54)
\end{aligned}$$

where we use the notation $j|k$ to indicate the estimate at time t_j conditioned on observations up to and including time t_k. Notice that in this notation the analysis estimate provided by the Kalman filter can be indicated as $\mathbf{w}^a_{k|k}$, and the forecast estimate can be indicated as $\mathbf{w}^f_{k|k-1}$. Similarly, the analysis and forecast error covariances can be indicated by $\mathbf{P}^a_{k|k}$ and $\mathbf{P}^f_{k|k-1}$, respectively. Once again, the fundamental quantity to determine is the conditional probability density $p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_k)$ in the expression above.

Repeated use of the definition of conditional probability gives

$$\begin{aligned}
p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_k) &= p(\mathbf{w}^t_{k-1}|\mathbf{w}^o_k, \mathbf{W}^o_{k-1}) \\
&= \frac{p(\mathbf{w}^t_{k-1}, \mathbf{w}^o_k, \mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k, \mathbf{W}^o_{k-1})} \\
&= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1}, \mathbf{W}^o_{k-1})p(\mathbf{w}^t_{k-1}, \mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k, \mathbf{W}^o_{k-1})} \\
&= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1}, \mathbf{W}^o_{k-1})p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1})p(\mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k|\mathbf{W}^o_{k-1})p(\mathbf{W}^o_{k-1})} \\
&= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1}, \mathbf{W}^o_{k-1})p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k|\mathbf{W}^o_{k-1})}, \\
&= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1})p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k|\mathbf{W}^o_{k-1})}, \qquad (9.55)
\end{aligned}$$

[1] This ℓ here is not to be confused with the ℓ used earlier in this notes to denote the number of time steps between two consecutive model time steps. Remember that the number of model time steps between consecutive observations has been fixed to one from Fig. 5.1 on.

where the last equality is obtained after noticing that the observation sequence is white in time. In this equation we recognize the denominator as being the same denominator as that in (5.27) corresponding to the probability density function of the innovation vector. Moreover, changing k into $k-1$ in (5.38) we can identify $p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1})$ as the probability density function of the filter analysis at time t_{k-1}. Thus, the only quantity remaining to be calculated in the expression above is the first term in the numerator.

Because the statistics of all errors and initial state are Gaussian, the probability density $p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1})$ is also Gaussian and can be written as

$$p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1}) = \frac{1}{(2\pi)^{m_k/2}|\tilde{\mathbf{R}}_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}^o_k - \mathcal{E}\{\mathbf{w}^o_k|\mathbf{w}^t_{k-1}\})^T(\tilde{\mathbf{R}}_k)^{-1}(\mathbf{w}^o_k - \mathcal{E}\{\mathbf{w}^o_k|\mathbf{w}^t_{k-1}\})\right] \tag{9.56}$$

where $\tilde{\mathbf{R}}_k$ denotes the following conditional error covariance

$$\tilde{\mathbf{R}}_k \equiv \mathcal{E}\{[\mathbf{w}^o_k - \mathcal{E}\{\mathbf{w}^o_k|\mathbf{w}^t_{k-1}\}][\mathbf{w}^o_k - \mathcal{E}\{\mathbf{w}^o_k|\mathbf{w}^t_{k-1}\}]^T|\mathbf{w}^t_{k-1}\}\,. \tag{9.57}$$

Using (5.1) and (5.3) we can calculate the conditional mean above as

$$\begin{aligned} \mathcal{E}\{\mathbf{w}^o_k|\mathbf{w}^t_{k-1}\} &= \mathcal{E}\{(\mathbf{H}_k\mathbf{w}^t_k + \mathbf{v}_k)|\mathbf{w}^t_{k-1}\} \\ &= \mathcal{E}\{(\mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}^t_{k-1} + \mathbf{H}_k\mathbf{b}^t_{k-1} + \mathbf{v}_k)|\mathbf{w}^t_{k-1}\} \\ &= \mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}^t_{k-1} \end{aligned} \tag{9.58}$$

where we noticed that the noise sequences $\{\mathbf{b}^t_k\}$ and $\{\mathbf{v}_k\}$ have mean zero. Similarly, it follows that the conditional error covariance $\tilde{\mathbf{R}}_k$ is given by

$$\tilde{\mathbf{R}}_k = \mathbf{H}_k\mathbf{Q}_{k-1}\mathbf{H}^T_k + \mathbf{R}_k \tag{9.59}$$

This completes the amount of information required to fully determine the probability density (9.56) and consequently the probability density (9.55).

Substituting (5.27) with $k \to k-1$, (5.31) and (9.56) in (9.55) it follows that

$$p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_k) = \frac{|\mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{P}^a_{k-1}\mathbf{\Psi}^T_{k-1}\mathbf{H}^T_k + \tilde{\mathbf{R}}_k|^{1/2}}{(2\pi)^{n/2}|\mathbf{P}^a_{k-1}|^{1/2}|\tilde{\mathbf{R}}_k|^{1/2}} \exp[-\frac{1}{2}J] \tag{9.60}$$

where J, in this case, is defined as

$$\begin{aligned} J(\mathbf{w}^t_{k-1}) \equiv\ & (\mathbf{w}^o_k - \mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}^t_{k-1})^T\tilde{\mathbf{R}}^{-1}_k(\mathbf{w}^o_k - \mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}^t_{k-1}) \\ & + (\mathbf{w}^t_{k-1} - \mathbf{w}^a_{k-1})^T(\mathbf{P}^a_{k-1})^{-1}(\mathbf{w}^t_{k-1} - \mathbf{w}^a_{k-1}) \\ & - (\mathbf{w}^o_k - \mathbf{H}_k\mathbf{w}^f_k)^T(\mathbf{H}_k\mathbf{P}^f_k\mathbf{H}^T_k + \mathbf{R}_k)^{-1}(\mathbf{w}^o_k - \mathbf{H}_k\mathbf{w}^f_k) \\ =\ & (\mathbf{w}^o_k - \mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}^t_{k-1})^T\tilde{\mathbf{R}}^{-1}_k(\mathbf{w}^o_k - \mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}^t_{k-1}) \\ & + (\mathbf{w}^t_{k-1} - \mathbf{w}^a_{k-1})^T(\mathbf{P}^a_{k-1})^{-1}(\mathbf{w}^t_{k-1} - \mathbf{w}^a_{k-1}) \\ & - (\mathbf{w}^o_k - \mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}^a_{k-1})^T(\mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{P}^a_{k-1}\mathbf{\Psi}^T_{k-1}\mathbf{H}^T_k + \tilde{\mathbf{R}}_k)^{-1}(\mathbf{w}^o_k - \mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}^a_{k-1}) \end{aligned} \tag{9.61}$$

where the first two terms in J come from the two terms in the numerator of (9.55), respectively, while the last term in J comes from the denominator of (9.55).

Hence, the the probability density function in (9.54) can be written as

$$p(\mathbf{w}_{k-1}^t|\mathbf{W}_k^o) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}_{k-1|k}^a|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1|k}^a)^T(\mathbf{P}_{k-1|k}^a)^{-1}(\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1|k}^a)\right] \tag{9.62}$$

where its maximum $\mathbf{w}_{k-1|k}^a$ corresponds to the estimate we are seeking and is given by

$$\mathbf{w}_{k-1|k}^a = \mathbf{w}_{k-1}^a + \mathbf{P}_{k-1}^a\mathbf{\Psi}_{k-1}^T\mathbf{H}_k^T\mathbf{\Gamma}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f)\,, \tag{9.63}$$

with corresponding error covariance

$$(\mathbf{P}_{k-1|k}^a)^{-1} = (\mathbf{P}_{k-1}^a)^{-1} + \mathbf{\Psi}_{k-1}^T\mathbf{H}_k^T\tilde{\mathbf{R}}_k^{-1}\mathbf{H}_k\mathbf{\Psi}_{k-1}\,. \tag{9.64}$$

These quantities are sometimes referred to as the retrospective analysis and the retrospective analysis error covariance matrix, respectively (see Todling et al. [131]).

In complete analogy to the remark made when introducing the maximum *a posteriori* functional J_{MAP} in (4.65), we could derive the same results in (9.63) and (9.64) by minimizing the cost function J_{MAP} for this case, that is,

$$\begin{aligned} J_{\text{MAP}}(\mathbf{w}_{k-1}^t) \;\equiv\; & (\mathbf{w}_k^o - \mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}_{k-1}^t)^T\tilde{\mathbf{R}}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{w}_{k-1}^t) \\ & + (\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1}^a)^T(\mathbf{P}_{k-1}^a)^{-1}(\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1}^a)\,. \end{aligned} \tag{9.65}$$

As before, this corresponds to considering only the probability distributions in the numerator of (9.60).

An alternative expression to (9.64) can be obtained using the Sherman–Morrison–Woodbury formula (c.f., Golub & Van Loan [67], p. 51), which gives

$$\begin{aligned} \mathbf{P}_{k-1|k}^a &= \mathbf{P}_{k-1}^a - \mathbf{P}_{k-1}^a\mathbf{\Psi}_{k-1}^T\mathbf{H}_k^T\mathbf{\Gamma}_k^{-1}\mathbf{H}_k\mathbf{\Psi}_{k-1}\mathbf{P}_{k-1}^a \\ &= \mathbf{P}_{k-1}^a - \mathbf{K}_{k-1|k}\mathbf{H}_k\mathbf{P}_{k,k-1|k-1}^{fa} \end{aligned} \tag{9.66}$$

where we introduced the following definitions, to write the last equality:

$$\mathbf{P}_{k|k-1}^{fa} \equiv \mathbf{\Psi}_{k-1}\mathbf{P}_{k-1}^a \tag{9.67a}$$

$$\mathbf{K}_{k-1|k} \equiv (\mathbf{P}_{k|k-1}^{fa})^T\mathbf{\Gamma}_k^{-1} \tag{9.67b}$$

Observer that the advantage of expression (9.66) over (9.64) for $\mathbf{P}_{k-1|k}^a$ is that (9.66 does not involve the inverse of usually large error covariance matrices. Also, with the definitions in (9.67), the expression for the lag–1 smoother estimate (9.63) becomes

$$\mathbf{w}_{k-1|k}^a = \mathbf{w}_{k-1}^a + \mathbf{K}_{k-1|k}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f) \tag{9.68}$$

Therefore, the smoother analysis at time t_{k-1} using data up to an including t_k corresponds to an update of the filter analysis $\mathbf{w}_{k-1}^a$ at time t_{k-1}, based on the same innovation vector used to calculated the filter analysis at the time of the latest observation, t_k.

The procedure above can be generalized to any number of lags larger than one, but the probabilistic framework above does not provide the simplest method for this generalization. A much simpler way is to use the approach of state augmentation combined with the minimum variance approach we described earlier in this notes. This procedure is outlined in Anderson & Moore [1], and it is explicitly invoked in Todling & Cohn [128] to derived the nonlinear extended fixed–lag smoother. Here, we only list the equations that needed to be supplemented to the linear Kalman filter so it becomes the linear fixed–lag Kalman smoother:

$$\begin{aligned}
\mathbf{w}^a_{k-\ell|k} &= \mathbf{w}^a_{k-\ell|k-1} + \mathbf{K}_{k-\ell|k}(\mathbf{w}^o_k - \mathbf{H}_k\mathbf{w}^f_k) && (9.69a)\\
\mathbf{P}^{aa}_{k,k-\ell|k} &= \left(\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_k\right)\mathbf{P}^{fa}_{k,k-\ell|k-1} && (9.69b)\\
\mathbf{P}^a_{k-\ell|k} &= \mathbf{P}^a_{k-\ell|k-1} - \mathbf{K}_{k-\ell|k}\mathbf{H}_k\mathbf{P}^{fa}_{k,k-\ell|k-1} && (9.69c)
\end{aligned}$$

where the generalization of definitions (9.67) are

$$\begin{aligned}
\mathbf{P}^{fa}_{k,k-\ell|k-1} &= \mathbf{\Psi}_{k-1}\mathbf{P}^{aa}_{k-1,k-\ell|k-1} && (9.70a)\\
\mathbf{K}_{k-\ell|k} &= \left(\mathbf{P}^{fa}_{k,k-\ell|k-1}\right)^T\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k && (9.70b)
\end{aligned}$$

respectively. In applying these expressions to the case we have just derived of $\ell = 1$ we must recognize the fact that we used a compact notation for the filtering problem, since in that case it avoided unnecessary complexity. Explicitly, when using $\ell = 1$ in (9.69a) we are confronted with the estimate $\mathbf{w}^a_{k-1|k-1}$, on the right–hand–side of the equation, which ultimately arises from the filtering problem. The filter estimate at time t_{k-1} is conditioned on all observations up to and including time t_{k-1}, which is just (5.24),

$$\mathbf{w}^a_{k-1|k-1} \equiv \mathcal{E}\{\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1}\} = \mathbf{w}^a_{k-1} \qquad (9.71)$$

which reduces to the simpler notation used in the filtering problem. A similar argument applies to the analysis error covariance $\mathbf{P}^a_{k-1|k-1}$ we encounter when substituting $\ell = 1$ in (9.69b) and (9.69c), i.e., we must realize that $\mathbf{P}^a_{k-1|k-1} = \mathbf{P}^a_{k-1}$. In this case, however, not using the notation with the conditioning explicitly written is more than just notational simplification for the filtering problem. It represents the fact that the filter error covariances are not conditioned on the data, as we observed in (5.41) and (5.42).

9.3.2 Application to a Linear Shallow-Water Model

Let us now examine the results of the fixed–lag Kalman smoother applied to the linear shallow–water model of the previous section. The interest here is to improve up on previously calculated filter analysis by using data past the analysis time. Following Cohn et al. [35], we consider the case of the A–network introduced above, but to show the more stringent results from that work we consider the case in which only the western half of the radiosondes are used in the assimilation experiments.

Figure 9.5 displays the time evolution of the domain–averaged root–mean–square errors for a period of 10 days of assimilation. The figure depicts only to the analysis errors, for all three variables of the model, in contrast to Fig. 9.2 where both the analysis and forecast errors are displayed, for these variables. This means that whenever comparing both figures, we should only care about the lower envelope of the curves in Fig. 9.2, corresponding to the analysis errors. The results in Fig. 9.5 are for both the Kalman filter and the Kalman

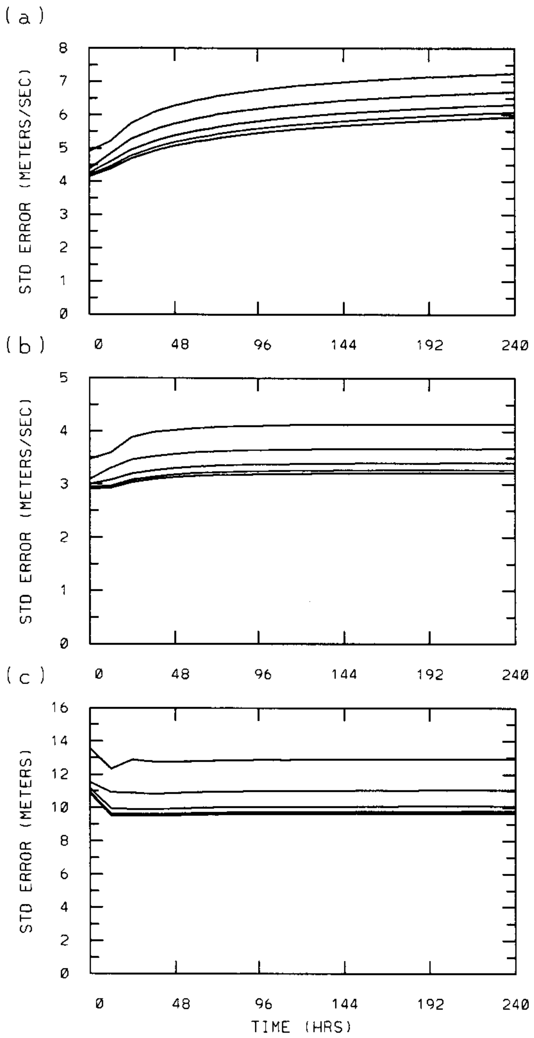

Figure 9.5: Expected analysis error standard deviations averaged over the domain, for each variable of the model, as a function of time. Panels are ordered as in Fig. 9.2, and results are for the fixed–lag Kalman smoother using only the western half of the radiosonde observations displayed in Fig. 9.1. The uppermost curve in each panel corresponds to the Kalman filter analysis (analogous to the lower envelope of the curves in Fig. 9.2), and successively lower curves are for the retrospective analysis results for 12–hour lags $\ell = 1, 2, 3$ and 4.

smoother for up to lag $\ell = 4$. The highest curve on each panel correspond to the filter errors. When compared against the corresponding curves (labeled A) in Fig. 9.2, we see that the errors are now larger, for all variables, than they were. This is a consequence of the fact that the number of observations is roughly one–half of what it is in the experiment with the A–network of Fig. 9.2. Successive lower curves, than the filter curve, in each of the panels in Fig. 9.5 refer to successive retrospective analyses obtained using data 12, 24, 36, and 48 hours ahead of the analysis time. These correspond to the fixed–lag Kalman smoother results for lags $\ell = 1, 2, 3$, and 4, respectively. The improvement achieved by the consecutive retrospective analyses is clearly seen by the decrease in the errors with the increase of the lags.

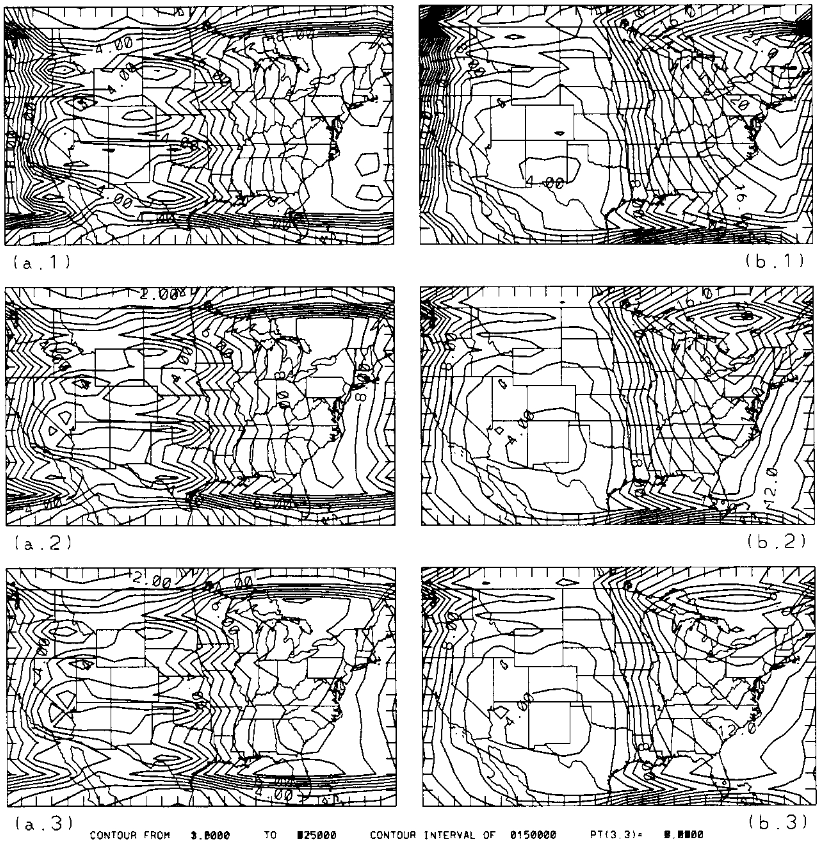
4.00
8.00
(a.1)
4.00
(b.1)
2.00
(a.2)
12.0
(b.2)
2.00
4.00
(a.3)
12.0
(b.3)
CONTOUR FROM 3.0000 TO ▮25000 CONTOUR INTERVAL OF 0150000 PT(3,3)= ▮.▮▮00

Figure 9.6: Maps of analysis error standard deviations at day 2 of the assimilation period for the fixed–lag Kalman smoother. Panels (a.1)–(a.3) refer to errors in u for the Kalman filter and Kalman smoother for lags $\ell = 1$ and 4, respectively; panels (b.1)–(b.3) refer to errors in h for the Kalman filter and Kalman smoother for lags $\ell = 1$ and 4, respectively.

To further illustrate the analysis improvement due to smoothing we show in Fig. 9.6 maps of the analysis errors in the u (left panels) and h (right panels) fields, at day 2. The top–two panels are for the filter analysis, the central–two panels for lag $\ell = 1$, i.e., when the smoother uses data at day 2.5 to further correct the filter analysis at day 2, and the bottom–two panels are for lag $\ell = 4$, that is, when data between days 2.5 and 4 have been used to correct the filter analysis at day 2. Looking at the filter results [panels (a.1) and (b.1)], we see that the large errors are located over the eastern part of the domain, where there are no radiosondes in this case. This is more dramatically seen from the height error fields, but the same is true for all fields, including the errors in v (not shown). Just as we encountered before for the case of Fig. 9.3, the contours in the top–two maps reflect the change in the observational data density. Moreover, the larger separation among the contours as we look from West to East reflect the advection of errors in the direction of the flow. Looking at the retrospective analyses results in the central–two and bottom–two panels, we observe an overall reduction of errors, especially for the lag $\ell = 4$ case, when compared against the top–two panels in the figure. However, the most striking feature of all in the figure is the propagation of the "region" of maximum analysis errors over the data void from East to West, as we look down from the top panels in the figure to the bottom panels, in each column. The magnitude of maximum errors not only gets reduced but also propagates against the flow. This illustrates the ability of the fixed–lag smoother to propagate information upstream (see also Ménard & Daley [105]).

Exercise

The Kalman filter applied to a linear advection equation. Consider the one–dimensional advection equation

$$\frac{\partial u}{\partial t} + U\frac{\partial u}{\partial x} = 0$$

where $U = const.$ represents the advection speed, applied to a periodic domain defined by the interval $[-2, 2]$ over the line. Take for initial condition a "rectangular" wave of the form

$$u(x, t=0) = \begin{cases} 1, & \text{for } -1 \leq x \leq 0 \\ 0, & \text{otherwise} \end{cases}$$

Using an up–wind finite difference scheme we can write an approximate solution to the advection equation as

$$v_j = Cv_{j-1} + (1 - C)v_j$$

where v_j represents the numeric solution for $u(x = j\Delta x)$ with Δx as the spatial interval, and where $C = U\Delta t/\Delta x$ is the Courant number, with time step Δt.

Simulation experiments: Using the parameters in the table below, obtain plots of the state evolution at the initial and final times for an integration taken from $T_0 = 0$ to $t_{final} = 1$ using the following Courant numbers: (i) $C = 1$, $C = 0.95$, and $C = 0.95$. Explain the difference in the results.

Table 9.3: Parameters for the finite–difference.

Domain	for $-2 \leq x \leq 2$
Mesh size	$\Delta x = 0.2$
Time step	$\Delta t 0.05$

Let us now slowly build the components to have a Kalman filter assimilation experiment constructed. We assume that all error statistics necessary for the filter are Gaussian and white. What we have to do next is to construct error covariance matrices for all stochastic processes involved in the problem. For simplicity, we take the perfect model assumption, so that we do not have to worry about the model error $\mathbf{b}_k^t$ and its error covariance $\mathbf{Q}_k$, i.e., $\mathbf{b}_k^t = 0$ at all times. Moreover, we assume there are no correlations among observation errors, so that the observation error covariance matrix $\mathbf{R}_k$ is diagonal. Furthermore, this matrix is assumed to be time independent with elements alone the diagonal to be specified depending of the experiments to be performed below under these conditions. The only error covariance left to specify it that of the initial estimate, $\mathbf{P}_0^a$. Constructing spatial error covariances that really satisfy the requirements of being an error covariance can be a pretty delicate issue. Instead of getting into these problems, we use a Matlab function gcorr provided with this exercise to construct an appropriate error correlation field that can be used to generate the require covariance matrix. Use the help of this function to see its usage and perform the following tasks for three different choices of the (de)correlation length parameter $L_d = 0.5, 1$ and 1.5 the following plots:

1. Is the matrix you constructed an acceptable correlation matrix?
2. Plot the two–point correlation function at two distinct arbitrary locations. Comment on what you see.
3. Make a contour plot of the correlation matrix. What you will see corresponds to the "shape" of a homogeneous and isotropic correlation matrix.

Now write a Matlab program with the Kalman filter equations for the state estimates $\mathbf{w}_k^f$ and $\mathbf{w}_k^a$, and their corresponding covariances $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$, respectively.

In this problem we use what is called *simulated observations*, where we take the solution of the advection equation at specific spatial locations and time intervals Δt_{obs} and add a Gaussian distributed error to it. This can be done in Matlab using the random number generator randn for normally distributed variables.

The following experiments fall in the category of what is referred in the literature as *wave generation*, where we take the initial guess (initial analysis) to be zero, i.e., $\mathbf{w}_0^a = \mathbf{0}$ and we try to reconstruct the true state processing the observations with the Kalman filter.

In what follows, you are asked to make plots for the true state and its estimate at the final time of the assimilation as well as plots of the time evolution of the domain–averaged forecast and analysis error standard deviation.

Half observation coverage case: The first case we consider is one for which the observation are all located over the left–half of the domain.

1. Following the choice of parameters in the table below, obtain the output for the true state and its estimate at the final time of the assimilation experiment.
2. What happens if the observation error level is increased to 0.1?
3. What happens if in the previous case, the assimilation period is 5 time units?
4. Comment on the results you just obtained.

Table 9.4: Observational error standard deviations.

Assimilation time period	1 time unit
Courant number	0.95
Obs frequency	$4 \times \Delta t$
Obs error std dev	0.02
Obs sparsity	left–half of grid points

You can try lots of other combinations and possibilities with this little program. There is a lot you can learn from just a small example such as this. Changing at least one of the parameters in the table above, here are a couple of other possible scenarious to investigate:

- Take observations at every grid point.
- Change the Courant number to make the dynamics more (numerically) dissipative.

What happens to the filter results in these cases?

Acknowledgments

The present notes are part of a larger set of notes initially prepared for a two–week course presented at the Instituto de Matemática Pura e Aplicada, Rio de Janeiro, in January 1996. I express my gratitude to Professor Dan Marchesin for his invitation and encouragement. I am also thankful to Beata Gundelach for her careful editing of the original Portuguese version of these notes. The Portuguese version became obsolete when these notes were used as the basis for a course thaught at the Meteorology Department of the University of Maryland in the Winter semester of 1997, after invitation by Peter Lyster, who I thank for the support and interest. I also thank Robert Rosenberg and the management of General Sciences Corporation for their understanding during that semester. Since then, shorter versions of these notes have been used at different occasions for short introductory courses in the subject of atmospheric data assimilation: Laboratório Nacional Computation Científica, Rio de Janeiro; George Mason Univerity, Virgina; and at the Workshop on Inverse Methods in Global Biogeochemical Cycles, Crete. Finally, I thank all the participants of all these classes who in many ways contributed to my constant motivation to improve the quality of these notes.

Bibliography

[1] Anderson, B.D.O., & J.B. Moore, 1979: *Optimal Filtering.* Prentice–Hall, 357 pp.

[2] Andersson, E., J. Pailleux, J.–N. Thépaut, J.R. Eyre, A.P. McNally, G.A. Kelly, & P. Courtier, 1993: Use of radiances in 3D/4D variational data assimilation. *Proc. ECMWF Workshop on Variational Assimilation*, Reading, U.K., 123–156.

[3] Antoulas, A.C. (Ed.), 1991: *Mathematical System Theory. The Influence of R.E. Kalman.* Springer–Verlag, 605 pp.

[4] Apostol, T.M., 1978: *Calculus, Vol. 2*, 2nd ed. Reverté, 813 pp.

[5] Arfken, G., 1985: *Mathematical Methods for Physicists*, 3rd ed. Academic Press, 985 pp.

[6] Balgovind, R., A. Dalcher, M. Ghil, & E. Kalnay, 1983: A stochastic–dynamic model for the spatial structure of forecast error statistics. *Mon. Wea. Rev.,* **111**, 701–722.

[7] Bartello, P., & H.L. Mitchell, 1992: A continuous three–dimensional model for short–range forecast error covariances. *Tellus,* **44A**, 217–235.

[8] Bengtsson, L., M. Ghil, & E. Källén (eds.), 1981: *Dynamic Meteorology: Data Assimilation Methods.* Springer–Verlag, 330 pp.

[9] Bennett, A.F, 1992: *Inverse Methods in Physical Oceanography.* Cambridge University Press., 346 pp.

[10] Bergman, K.H., 1979: Multivariate analysis of temperatures and winds using optimum interpolation. *Mon. Wea. Rev.,* **107**, 1423–1444.

[11] Bergthórsson, P., & B.R. Döös, 1955: Numerical weather map analysis. *Tellus,* **7**, 329–340.

[12] Bierman, G.J., 1977: *Factorization Methods for Discrete Sequential Estimation.* Academic Press, 241 pp.

[13] Bittanti, S. A.J. Laub, & J.C. Willems (Eds.), 1991: *The Riccati Equation.* Springer–Verlag, 338 pp.

[14] Bjerknes, V., 1911: *Dynamic Meteorology and Hydrography. Part II: Kinematics.* Carnegie Institute, Gibson Bros., 175 pp.

[15] Bochner, S., 1959: *Lectures on Fourier Integrals.* Princeton University Press, 333 pp.

[16] Boggs, D., M. Ghil, & C. Keppenne, 1995: A stabilized sparse–matrix U–D square–root implementation of a large–state extended Kalman filter. *Proc. Intl. Symp. on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan, WMO, Vol. 1, 219–224.

[17] Boyd, J.P., 1995: Eight definitions of the slow manifold: seiches, pseudoseiches and exponential smallness. *Dyn. Atmos. Oceans,* **22**, 49–75.

[18] Brewer, J.W., 1978: Kronecker products and matrix calculus in system theory. *IEEE Trans. Circuits and Systems*, **25**, 772–781.

[19] Brown, R.G., 1983: *Random Signal Analysis and Kalman Filtering.* John Wiley & Sons, 347 pp.

[20] Bryson, A.E. & Y. Ho, 1975: *Optimization, Estimation and Control.* Wiley, 481 pp.

[21] Bürger, G., & M.A. Cane, 1994: Interactive Kalman filtering. *J. Geophys. Res.,* **99**, 8015–8031.

[22] Butkov, E., 1968: *Mathematical Physics.* Addison–Wesley Publishing Co., XXX pp.

[23] Charney, J., R. Fjørtoft, & J. von Neumann, 1950: Numerical integration of the barotropic vorticity equation. *Tellus,* **2**, 237–254.

[24] Christakos, G., 1992: *Random Field Models in Earth Sciences.* Academic Press, 474 pp.

[25] Chui, C.K., & G. Chen, 1991: *Kalman Filtering with Real Time Applications*, 2nd ed. Springer–Verlag, Vol. *17*, 191 pp.

[26] Chung, K.L. 1974: *A Course in Probability Theory*, 2nd ed. Academic Press, 365 pp.

[27] Cohn, S.E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, in press.

[28] ——, 1993: Dynamics of short–term univariate forecast error covariances. *Mon. Wea. Rev.,* **121**, 3123–3149.

[29] ——, 1992: Short–term dynamics of forecast error covariances. *Proc. ECMWF Workshop on Variational Assimilation*, Reading, U.K., pp. 157–170.

[30] ——, 1982: *Methods of Sequential Estimation for Determining Initial Data in Numerical Weather Prediction.* Ph.D. Thesis, Courant Institute of Mathematical Sciences, New York University, 183 pp.

[31] ——, & D.P. Dee, 1988: Observability of discretized partial diferential equations. *SIAM J. Numer. Anal.*, **25**, 586–617.

[32] ——, & L.L. Morone, 1984: *The effect of horizontal gradients of height-height forecast error variances upon OI forecast error statistics.* Office Note 296, National Meteorological Center, Washington DC 20233, 37 pp.

[33] ——, & D.F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.,* **119**, 1757–1785.

[34] ———., & R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Japan,* **74**, 63–75.

[35] ———, N. S. Sivakumaran, & R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.,* **122**, 2838–2867.

[36] Courtier, P., 1997: Variational methods. M. Ghil, K. Ide, A. Bennett, P. Courtier, M. Kimoto, N. Nagata, & N. Sato (Eds.): *Data Assimilation in Meteorology and Oceanography: Theory and Practice*, Universal Academic Press, 211–218.

[37] da Silva, A., J. Pfaendtner, J. Guo, M. Sienkiewicz, & S. E. Cohn, 1995: Assessing the effects of data selection with DAO's physical–space statistical analysis system. *Proc. Intl. Symp. on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan, WMO, Vol. 1, 273–278.

[38] Daley, R., 1995: Estimating the wind field from chemical constituent observations: Experiments with a one-dimensional extended Kalman filter. *Mon. Wea. Rev.,* **123**, 181–198.

[39] ———, 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

[40] ———, 1978: Variational non–linear normal mode initialization. *Tellus,* **30**, 201–218.

[41] Dee, D.P., 1995: On–line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123**, 1128–1145.

[42] ———, 1992: A simple scheme for tuning forecast error covariance parameters. *Workshop on Variational Assimilation*, Reading, England, ECMWF, 191–206.

[43] ———, 1991: Simplification of the Kalman filter for meteorological data assimilation. *Quart. J. Roy. Meteor. Soc.,* **117**, 365–384.

[44] ———, 1983: *Computational Aspects of Adaptive Filtering and Applications to Numerical Weather Prediction.* Ph.D. Thesis, Courant Inst. Math. Sci., New York University, 150 pp.

[45] Eddy, A., 1964: The objective analysis of horizontal wind divergence fields. *Quart. J. Roy. Meteor. Soc.,* **90**, 424–440.

[46] ———, 1967: The statistical objective analysis of scalar fields. *J. Appl. Meteor.,* **6**, 597–609.

[47] Ehrendorfer, M., 1994a: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Mon. Wea. Rev.,* **122**, 703–713.

[48] ———, 1994b: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: Applications. *Mon. Wea. Rev.,* **122**, 714–728.

[49] Eliassen, A., 1954: *Provisional Report of Calculation of Spatial Covariance and Autocorrelation of Pressure Field.* Rept. No. **5**, Institute of Weather and Climate Res., Academy of Science Oslo, 11 pp. (reprinted in Bengtsson, L., M. Ghil, & E. Källén (Eds.), pp. 319–330).

[50] Epstein, E.S., 1969: Stochastic dynamic prediction. *Tellus,* **21**, 739–759.

[51] Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans,* **99**, 10143–10162.

[52] _____, 1992: Using the extended Kalman filter with a multilayer quasigeostrophic ocean model. *J. Geophys. Res.,* **97**, 17905–17924.

[53] Fu, L.-L., I. Fukumori, & R.N. Miller, 1993: Fitting dynamic models to the Geosat sea level observations in the tropical Pacific Ocean. Part II: A linear, wind-driven model. *J. Phys. Oceanogr.,* **23**, 2162-2181.

[54] Fukumori, I., 1995: Assimilation of TOPEX sea level measurements with a reduced-gravity, shallow water model of the tropical Pacific Ocean. *J. Geophys. Res. Oceans,* **100**, 25027–25039.

[55] _____, & P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation; an example with an idealized Gulf Stream model. *J. Geophys. Res. Oceans,* **100**, 6777–6793.

[56] _____, J. Benveniste, C. Wunsch, & D.B. Haidvogel, 1993: Assimilation of sea surface topography into an ocean circulation model using a steady-state smoother. *J. Phys. Oceanogr.,* **23**, 1831–1855.

[57] Gandin, L.S., 1963: *Objective Analysis of Meteorological Fields.* Gidrometeor. Izdat., Leningrad. (English translation: Israel Program for Scientific Translation, Jerusalem, 1965, 242 pp.)

[58] _____, 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.,* **116**, 1137–1156.

[59] Gaspari, G., & S.E. Cohn, 1996: Theory and applications of correlation function modeling on the sphere. *Math. Geology*, submitted.

[60] Gelb, A. (Ed.), 1974: *Applied Optimal Estimation.* MIT Press, 374 pp.

[61] Ghil, M., 1989: Meteorological data assimilation for oceanographers. Part I: description and theoretical framework. *Dyn. Atmos. Oceans,* **13**, 171–218.

[62] _____, & S. Childress, 1987: *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics.* Academic Press, 485 pp.

[63] _____, & K. Ide, 1994: Extended Kalman filtering for vortex systems: An example of observing system design. P. Brasseur & J.C.H. Nihoul (Eds.): *Data Assimilation for Modelling the Ocean in a Global Change Perspective*, Springer-Verlag, 167–193.

[64] _____, & P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances in Geophysics,* Vol. 33, Academic Press, 141–266.

[65] _____, & R. Todling, 1996: Tracking atmospheric instabilities with the Kalman filter. Part II: Two-Layer Results. *Mon. Wea. Rev.,* in press.

[66] _____, S. Cohn, J. Tavantzis, K. Bube & E. Isaacson, 1981: Applications of estimation theory to numerical weather prediction. Bengtsson, L., M. Ghil, & E. Källén (Eds.): *Dynamic Meteorology: Data Assimilation Methods.* Springer-Verlag, 139–224.

[67] Golub, G.H., & C.F. Van Loan, 1989: *Matrix Computations*, 2nd ed. The Johns Hopkins University Press, 642 pp.

[68] Guo, J., & A. da Silva, 1997: Computational Aspects of Goddard's Physical-Space Statistical Analysis System (PSAS). In *Numerical Simulations in the Environmental and Earth Sciences: Proceedings of the 2nd UNAM-CRAY Supercomputing Conference*, F. García-García, G. Cisneros, A. Fernández-Eguiarte, and R. Álvarez, Eds., Cambridge University Press, 203–209.

[69] Halmos, P.R., 1958: *Finite-Dimensional Vector Spaces*, 2nd ed. D. Van Nostrand Company, Princeton, NJ, 200 pp.

[70] Haltiner, G.J., & R.T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology.* John Wiley & Sons, 477 pp.

[71] Hao, Z., 1994: *Data Assimilation for Interannual Climate-Change Prediction.* Ph.D. Thesis, University of California, Los Angeles, 224 pp.

[72] ——, & M. Ghil, 1995: Sequential parameter estimation for a coupled ocean–atmosphere model. *Proc. Intl. Symp. on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan, WMO, Vol. 1, 181–186.

[73] Harms, D.E., S. Raman, & R.V. Madala, 1992: An examination of four–dimensional data–assimilation techniques for numerical weather prediction. *Bull. Amer. Meteor. Soc.,* **73**, 425–440.

[74] Heckley, W.A., P. Courtier, J. Pailleux, & E. Andersson, 1993: The ECMWF variational analysis: general formulation and use of background information. *Proc. ECMWF Workshop on Variational Assimilation*, Reading, U.K., 49–94.

[75] Heemink, A.W., 1988: Two–dimensional shallow water flow identification. *Appl. Math. Modelling,* **12**, 109–118.

[76] ——, & H. Kloosterhuis, 1990: Data assimilation for non–linear tidal models. *Intl. J. Numer. Methods Fluids,* **11**, 1097–1112.

[77] Hoang, H.S., P. De Mey, O. Talagrand, & R. Baraille, 1995: Assimilation of altimeter data in a multilayer quasi–geostrophic ocean model by simple nonlinear adaptive filter. *Proc. Intl. Symp. on Assimilation of Observations in Meteorology and Oceanography,* Tokyo, Japan, WMO, 521–526.

[78] Hoffmann, G.-R., & D.F. Snelling (Eds.) 1988: *Multiprocessing in Meteorological Models.* Springer–Verlag, 438 pp.

[79] ——, & D.K. Maretis (Eds.) 1990: *The Dawn of Massively Parallel Processing in Meteorology.* Springer–Verlag, 376 pp.

[80] Hollingsworth, A., A.C. Lorenc, M.S. Tracton, K. Arpe, G. Cats, S. Uppala, & P. Kallberg, 1985: The response of numerical weather prediction systems to FGGE level IIb data. Part I: Analyses. *Quart. J. Roy. Meteor. Soc.,* **111**, 1–66.

[81] Hoke, J.E., & R.A. Anthes, 1976: The initialization of numerical models by a dynamic–initialization technique. *Mon. Wea. Rev.,* **104**, 1551–1556.

[82] Holton, J.R., 1979: *An Introduction to Dynamic Meteorology*, 2nd ed. Academic Press, 391 pp.

[83] Householder, A.S., 1964: *The Theory of Matrices in Numerical Analysis.* Dover, 257 pp.

[84] Jazwinski, A.H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 376 pp.

[85] Jiang, S., & M. Ghil, 1995: Toward monitoring the nonlinear variability of Western boundary currents — assimilation of simulated altimeter data into a wind–driven, double–gyre, shallow–water model. *Proc. Intl. Symp. on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan, WMO, Vol. 2, 533–538.

[86] Kailath, T., 1974: A view of three decades of linear filtering theory. *IEEE Trans. Inform. Theory,* **20**, 146–181.

[87] Kalman, R.E., 1960a: A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic. Eng.,* **82**, 35–45.

[88] ______, 1960b: Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana,* **5**, 102–119.

[89] ______, 1963: New methods in Wiener filtering theory. *Proc. 1st Symp. Eng. Appl. Random Function Theory and Probability*, J.L. Bogdanoff & F. Kozin, Eds., John Wiley & Sons, 270–388.

[90] ______, & R.S. Bucy, 1961: New Results in Linear Filtering and Prediction Theory. *Trans. ASME, Ser. D, J. Basic. Eng.,* **83**, 95–108.

[91] Lacarra, J.–F., and O. Talagrand, 1988: Short–range evolution of small perturbations in a barotropic model. *Tellus,* **40A**, 81–95.

[92] Le Moyne, L., & J. Alvarez, 1991: Analysis of dynamic data assimilation for atmospheric phenomena. Effect of the model order. *Atmósfera*, **4**, 145–164.

[93] Leith, C.E., 1983: Predictability in theory and practice. In *Large–Scale Dynamical Processes in the Atmosphere*, B. Hoskins & R. Pearce, Eds. Academic Press, 391 pp.

[94] Lewis, F.L., 1986: *Optimal Estimation with an Introduction to Stochastic Control Theory.* John Wiley & Sons, 376 pp.

[95] Lorenc, A.C., 1981: A global three–dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.,* **109**, 701–721.

[96] ______, R.S. Bell, & B. MacPherson, 1991: The Meteorological Office analysis correction data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **117**, 59–89.

[97] ______, & O. Hammon, 1988: Objective quality control of observations using Bayesian methods. Theory and practical implementation. *Quart. J. Roy. Meteor. Soc.,* **114**, 515–543.

[98] Lorenz, E.N., 1963: Deterministic non–periodic flow. *J. Atmos. Sci.,* **20**, 130–141.

[99] Lorenz, E.N., 1960: Maximum simplification of dynamical equations. *Tellus,* **12**, 243–254.

[100] Lyster, P.M., S.E. Cohn, R. Ménard, L.-P. Chang, S.-J. Lin, & R. Olsen, 1995: An Implementation of a Two-Dimensional Kalman Filter for Atmospheric Chemical Constituent Assimilation on Massively Parallel Computers. *Mon. Wea. Rev.,* submitted.

[101] Maybeck, P.S., 1979: *Stochastic Models, Estimation, and Control,* Vol. 1. Academic Press, 423 pp.

[102] McPherson, R.D., K.H. Bergman, R.E. Kristler, G.E. Rasch, & D.S. Gordon, 1979: The NMC operational global data assimilation system. *Mon. Wea. Rev.,* **107**, 1445–1461.

[103] Meditch, J.S., 1969: *Stochastic Linear Estimation and Control.* McGraw Hill Book Co., 394 pp.

[104] Ménard, R., 1994: *Kalman Filtering of Burger's Equation and its Application to Atmospheric Data Assimilation.* Ph.D. Thesis, McGill University, 211 pp.

[105] ——, and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus,* **48A**, 221–237.

[106] Mendel, J.M., 1971: Computational requirements for a discrete Kalman filter. *IEEE Trans. Auto. Control,* **16**, 748–758.

[107] Miller, R.N., M. Ghil, & F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51**, 1037–1056.

[108] Øksendal, B., 1992: *Stochastic Differential Equations: An Introduction with Applications,* 3rd ed. Springer–Verlag, 224 pp.

[109] Omatu, S., & J.H. Seinfeld, 1989: *Distributed Parameter Systems: Theory and Applications.* Oxford University Press, 430 pp.

[110] Pailleux, J., 1990: A global variational assimilation scheme and its application for using TOVS radiances. *Proc. Intl. Symp. Assim. Obsrv. Meteorol. Oceanogr.,* Clermont–Ferrand, France, WMO, pp. 325–328.

[111] Panofsky, H., 1949: Objective weather map analysis. *J. Meteor.,* **6**, 386–392.

[112] Parrish, D.F., & J.C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.,* **120**, 1747–1763.

[113] ——, & S.E. Cohn, 1985: *A Kalman Filter for a Two-Dimensional Shallow-Water Model: Formulation and Preliminary Experiments.* Office Note 304, National Meteorological Center, Washington DC 20233, 64 pp.

[114] Pedlosky, J., 1987: *Geophysical Fluid Dynamics,* Springer–Verlag, 624 pp.

[115] Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, 1989: *Numerical Recipes: The Art of Scientific Computing (FORTRAN Version)* Cambridge University Press, 702 pp.

[116] Phillips, N.A., 1976: The impact of synoptic observing and analysis system on flow pattern forecasts. *Bull. Amer. Meteor. Soc.,* **57**, 1225–1240.

[117] Richtmyer, R.D., & K.W. Morton, 1967: *Differential Methods for Initial Value Problems*, 2nd ed. Wiley–Interscience, 405 pp.

[118] Riedel, K.S., 1993: Block diagonally dominant positive definite approximate filters and smoothers. *Automatica,* **29**, 779–783.

[119] Rutherford, I.D., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.,* **29**, 809–815.

[120] Sage, A.P, & J.L. Melsa, 1971: *System Identification.* Academic Press, 221 pp.

[121] ——, & ——, 1970: *Estimation Theory with Applications to Communications and Control.* McGraw–Hill Book Co., 529 pp.

[122] Saucier, W.J., 1989: *Principles of Meteorological Analysis.* Dover Publications, 438 pp.

[123] Schlatter, T.W., 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.,* **103**, 246–257.

[124] ——, G.W. Branstator, L.G. Thiel, 1976: Testing a global multivariate statistical objective analysis scheme with observed data. *Mon. Wea. Rev.,* **104**, 765–783.

[125] Stavroulakis, P. (Ed.), 1983: *Distributed Parameter Systems Theory, Part II: Estimation.* Hutchinson Ross Publ. Co., 391 pp.

[126] Tarantola, A., 1987: *Inverse Problem Theory,* Elsevier, Amsterdam, 613 pp.

[127] Temperton, C., 1984: Variational normal mode initialization for a multi–level model. *Mon. Wea. Rev.,* **112**, 2303–2316.

Todling, R., 1997: Computational aspects of Kalman filtering and smoothing for atmospheric data assimilation. F. García–García, G. Cisneros, A. Fernández–Eguiarte, and R. Álvarez, Eds., Cambridge University Press, 191–202.

[128] ——, & S.E. Cohn, 1996: Some strategies for Kalman filtering and smoothing. *Proc. ECMWF Seminar on data assimilation,* 91–111.

[129] ——, & S.E. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.,* **122**, 2530–2557.

[130] ——, & M. Ghil, 1994: Tracking atmospheric instabilities with the Kalman filter. Part I: Methodology and one–layer results. *Mon. Wea. Rev.,* **122**, 183–204.

[131] Todling, R., S.E. Cohn, and N.S. Sivakumaran, 1998: Suboptimal schemes for retrospective data assimilation based on the fixed–lag Kalman smoother. *Mon. Wea. Rev.,* **126**, 2274-2286.

[132] Vanmarcke, E., 1983: *Random Fields: Analysis and Synthesis.* The MIT Press, 382 pp.

[133] Vasiljević, D., C. Cardinali, & P. Undén, 1993: ECMWF 3D variational data assimilation of conventional observations. *Proc. ECMWF Workshop on Variational Assimilation*, Reading, U.K., pp. 389–436.

[134] Verhaegen, M. & P. Van Dooren, 1986: Numerical aspects of different Kalman filter implementations. *IEEE Trans. Auto. Control,* **31**, 907–917.

[135] Verlaan, M, & A. W. Heemink, 1995: Reduced rank square root filters for large scale data assimilation problems. *Proc. Intl. Symp. on Assimilation of Observations in Meteorology and Oceanography,* Tokyo, Japan, WMO, Vol. 1, 247–252.

[136] Vetter, W.J., 1970: Derivative operations on matrices. *IEEE Trans. Auto. Control,* **15**, 241–244.

[137] ———, 1973: Matrix calculus operations and Taylor expansions. *SIAM Rev.,* **15**, 352–369.

[138] Washington, W.M., & C.L. Parkinson, 1986: *An Introduction to Three–Dimensional Climate Modeling.* University Science Books, 422 pp.

[139] Wiin–Nielsen, A., 1991: The birth of numerical weather prediction. *Tellus,* **43AB**, Special Issue, 36–52.

[140] Yaglom, A.M., 1987: *Correlation Theory of Stationary and Related Random Functions, Vol 1: Basic Results.* Springer–Verlag, 526 pp.