Inverse Modeling of the Ocean and Atmosphere

Andrew F. Bennett
College of Oceanic and Atmospheric Sciences
Oregon State University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 10.25/13.5 pt. *System* LaTeX 2ε [TB]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Bennett, Andrew F., 1945–
Inverse modeling the ocean and atmosphere / Andrew F. Bennett.
p. cm.
Includes bibliographical references and index.
ISBN 0 521 81373 5
1. Oceanography – Mathematical models. 2. Meteorology – Mathematical models.
3. Inverse problems (Differential equations) I. Title.
GC10.4.M36 B47 2002
551.46'001'5118 – dc21 2001052655

ISBN 0 521 81373 5 hardback

(1.5.17)

(1.5.18)

□

)–(1.5.16),
ions devel-
.1)–(4.2.6).

Chapter 2

Interpretation

The calculus of variations uses Green's functions and representers to express the best fit to a linear model and data. Mathematical construction of the representers is devious, and the meaning of the representer solution to the "control problem" of Chapter 1 is not obvious. There is a geometrical interpretation, in terms of observable and unobservable degrees of freedom. Unobservability defines an orthogonality, and the representers span a finite-dimensional subspace of the space of all model solutions or "circulations". The representers are in fact the observable degrees of freedom.

A statistical interpretation is also available: if the unknown errors in the model are regarded as random fields having prescribed means and covariances, then the representers are related, via the measurement processes, to the covariances of the circulations. Thus the representer solution to the variational problem is also the optimal linear interpolation, in time and space, of data from multivariate, inhomogeneous and nonstationary random fields. The minimal value of the penalty functional that defines the generalized inverse or control problem is a random number. It is the χ^2 variable, if the prescribed error means or covariances are correct, and has one degree of freedom per datum. Measurements need not be pointwise values of the circulation; representers along with their geometrical and statistical interpretations may be constructed for all bounded linear measurement functionals.

Analysis of the conditioning of the determination of the representer amplitudes reveals those degrees of freedom which are the most stable with respect to the observations. This characterization also indicates the efficiency of the observing system – the fewer unstable degrees of freedom, the better.

Interpreting the variational formulation is completed by demonstrating the relationship between weights, covariances and roughness penalties.

2.1 Geometrical interpretation

2.1.1 Alternatives to the calculus of variations

After formulating the penalty functional that defines the best fit to our model and our data, we found a local extremum using the theory of the calculus of the first variation. Specifically, we derived the Euler–Lagrange equations, and explicitly expressed their solution with representers. These functions were defined as special and directly calculable solutions of Euler–Lagrange-like equations. We shall now construct the same extremum for the penalty functional using Hilbert Space theory (Yoshida, 1980). This geometrical construction reveals the efficiency of minimization algorithms based on the Euler–Lagrange equations.

Exercise 2.1.1
How do we know that we shall find the same extremum? □

2.1.2 Inner products

We begin by defining an inner product for two "ocean circulations" $u = u(x,t)$ and $v = v(x,t)$:

$$\begin{aligned}\langle u, v\rangle \equiv & \int_0^T dt \int_0^T ds \int_0^L dx \int_0^L dy \left\{\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right) u(x,t)\right\} \\ & \times W_f(x,t,y,s)\left\{\left(\frac{\partial}{\partial s} + c\frac{\partial}{\partial y}\right) v(y,s)\right\} \\ & + \int_0^L dx \int_0^L dy\, u(x,0) W_i(x,y) v(y,0) \\ & + \int_0^T dt \int_0^T ds\, u(0,t) W_b(t,s) v(0,s) \\ = & f_u \bullet W_f \bullet f_v + u \circ W_i \circ v + u * W_b * v, \end{aligned} \tag{2.1.1}$$

where f_u is the residual for u (Bennett, 1992).

Exercise 2.1.2
Verify that $\langle\ ,\ \rangle$ is an inner product, that is:

(i) $\langle u, v\rangle = \langle v, u\rangle$ (assume the W are symmetric),
(ii) $\langle cu + dw, v\rangle = c\langle u, v\rangle + d\langle w, v\rangle$ for all real numbers c and d,
(iii) $\langle u, u\rangle \geq 0$,
(iv) $\langle u, u\rangle = 0 \quad \Leftrightarrow \quad u \equiv 0$ (nontrivial).

In terms of the inner product, our penalty functional is

$$\mathcal{J}[u] = \langle u - u_F, u - u_F \rangle + (\mathbf{d} - \mathbf{u})^{\mathrm{T}} \mathbf{W} (\mathbf{d} - \mathbf{u}). \tag{2.1.2}$$

□

2.1.3 Linear functionals and their representers; unobservables

Consider the linear mapping

$$u \to u(\xi, \tau), \tag{2.1.3}$$

where the lhs is a field, while the rhs is a particular value of the field. This mapping is a *linear functional*: it linearly maps a function to a single number.

Theorem 2.1.1
If *the vector space of admissible fields u, with the inner product* $\langle\ ,\ \rangle$, *is complete (that is, if it is a Hilbert Space),* **then** *there is a function* $\rho(x, t, \xi, \tau)$ *such that*

$$\langle \rho, u \rangle = u(\xi, \tau). \tag{2.1.4}$$

So ρ "represents" the measurement process. This is the Riesz representation theorem. Given ρ, we may express $\mathcal{J}$ entirely in terms of inner products (Wahba and Wendelberger, 1980):

$$\mathcal{J}[u] = \langle u - u_F, u - u_F \rangle + (\mathbf{d} - \langle \boldsymbol{\rho}, u \rangle)^{\mathrm{T}} \mathbf{W} (\mathbf{d} - \langle \boldsymbol{\rho}, u \rangle), \tag{2.1.5}$$

where $\rho_m = \rho(x, t, x_m, t_m)$, $1 \le m \le M$.

Now, *any* field $u = u(x, t)$ may be expressed as

$$u(x, t) = u_F(x, t) + \sum_{m=1}^{M} \nu_m \rho(x, t, x_m, t_m) + g(x, t), \tag{2.1.6}$$

where u_F is again the solution of (1.2.2)–(1.2.4), and where the ν_m are *any* coefficients, since we may always choose

$$g \equiv u - u_F - \boldsymbol{\nu}^{\mathrm{T}} \boldsymbol{\rho}. \tag{2.1.7}$$

Let us now impose the condition that g is "unobservable":

$$\langle \rho_m, g \rangle = g(x_m, t_m) = 0 \tag{2.1.8}$$

for $1 \le m \le M$. That is, g is orthogonal to each ρ_m. For a *given* u and a *given* u_F, we may use (2.1.8) to derive M equations for the ν_m; then g is uniquely defined by (2.1.7).

2.1.4 Geometric minimization with representers

But we're *not* given u; we're only given u_F. Thus ν and g are arbitrary. We wish to find the u that minimizes $\mathcal{J}[u]$. Let us evaluate $\mathcal{J}$ using (2.1.5) and (2.1.6):

$$\begin{aligned}\mathcal{J}[u] &= \langle \nu^{\mathrm{T}}\rho + g, \nu^{\mathrm{T}}\rho + g\rangle \\ &\quad + (\mathbf{d} - \langle \rho, u_F + \rho^{\mathrm{T}}\nu + g\rangle)^{\mathrm{T}}\mathbf{W}(\mathbf{d} - \langle \rho, u_F + \rho^{\mathrm{T}}\nu + g\rangle) \\ &= \nu^{\mathrm{T}}\langle \rho, \rho^{\mathrm{T}}\rangle\nu + \nu^{\mathrm{T}}\langle \rho, g\rangle + \langle g, \rho^{\mathrm{T}}\rangle\nu + \langle g, g\rangle \\ &\quad + (\mathbf{d} - \langle \rho, u_F\rangle - \langle \rho, \rho^{\mathrm{T}}\rangle\nu - \langle \rho, g\rangle)^{\mathrm{T}}\mathbf{W}(\mathbf{d} - \langle \rho, u_F\rangle - \langle \rho, \rho^{\mathrm{T}}\rangle\nu - \langle \rho, g\rangle).\end{aligned} \tag{2.1.9}$$

Next impose the M orthogonality conditions (2.1.8), and use the representing property of ρ to obtain

$$\begin{aligned}\mathcal{J}[u] = \mathcal{J}[\nu, g] &= \nu^{\mathrm{T}}\langle \rho, \rho^{\mathrm{T}}\rangle\nu + \langle g, g\rangle \\ &\quad + (\mathbf{d} - \mathbf{u}_F - \langle \rho, \rho^{\mathrm{T}}\rangle\nu)^{\mathrm{T}}\mathbf{W}(\mathbf{d} - \mathbf{u}_F - \langle \rho, \rho^{\mathrm{T}}\rangle\nu).\end{aligned} \tag{2.1.10}$$

The penalty functional $\mathcal{J}$ is now expressed explicitly in terms of ν and g. Note that g only appears once on the rhs of (2.1.10). Clearly $\mathcal{J}$ is least with respect to the choice of g if $\langle g, g\rangle = 0$, that is

$$g = \hat{g} \equiv 0. \tag{2.1.11}$$

We discard the field g orthogonal to all the representers. It remains to select the ν_m, $1 \le m \le M$. But first note that

$$\begin{aligned}\sigma_{lm} \equiv \langle \rho, \rho^{\mathrm{T}}\rangle_{lm} &= \langle \rho_l, \rho_m\rangle = \langle \rho_m, \rho_l\rangle \\ &= \rho_m(x_l, t_l) = \rho_l(x_m, t_m) \\ &= \sigma_{ml},\end{aligned} \tag{2.1.12}$$

so $\sigma = \sigma^{\mathrm{T}}$ and $\mathcal{J}$, which now only depends upon ν, may be expressed as

$$\mathcal{J}[u] = \mathcal{J}[\nu] = \nu^{\mathrm{T}}\sigma\nu + (\mathbf{h} - \sigma\nu)^{\mathrm{T}}\mathbf{W}(\mathbf{h} - \sigma\nu), \tag{2.1.13}$$

where $\mathbf{h} \equiv \mathbf{d} - \mathbf{u}_F$. Completing the square,

$$\mathcal{J}[\nu] = (\nu - \hat{\nu})^{\mathrm{T}}\mathbf{S}(\nu - \hat{\nu}) + \mathbf{h}^{\mathrm{T}}\mathbf{W}\mathbf{h} - \hat{\nu}^{\mathrm{T}}\mathbf{S}\hat{\nu}, \tag{2.1.14}$$

where $\mathbf{S} = \sigma + \sigma\mathbf{W}\sigma$ and $\mathbf{S}\hat{\nu} = \sigma\mathbf{W}\mathbf{h}$, both of which are given. We finally minimize $\mathcal{J}$ by choosing $\nu = \hat{\nu}$, and then $\hat{\mathcal{J}} \equiv \mathcal{J}[\hat{\nu}] = \mathbf{h}^{\mathrm{T}}\mathbf{W}\mathbf{h} - \hat{\nu}^{\mathrm{T}}\mathbf{S}\hat{\nu}$. Provided σ is nonsingular, we may untangle these results to find

$$(\sigma + \mathbf{W}^{-1})\hat{\nu} = \mathbf{h}, \tag{2.1.15}$$

which looks familiar.

Recall that our minimizer is

$$\hat{u} = u_F + \hat{\nu}^{\mathrm{T}}\rho + \hat{g}, \tag{2.1.16}$$

where $\hat{g}$ satisfies (2.1.11), and $\hat{\nu}$ satisfies (2.1.15).

2.1.5 Equivalence of variational and geometric minimization: the data space

Surely the representers ρ defined by the representing property (2.1.4) are the same as the representer functions $\mathbf{r}$ that satisfy the Euler–Lagrange-like system (1.3.8)–(1.3.13), in which case

$$\sigma = \mathbf{R}, \quad \hat{\nu} = \hat{\beta}\,? \tag{2.1.17}$$

Exercise 2.1.3

Show that

$$\rho_m(x,t) = r_m(x,t) \tag{2.1.18}$$

for $0 \le x \le L$, $0 \le t \le T$, and $1 \le m \le M$.

Hint

Consider (2.1.4):

$$\begin{aligned} u(x_m,t_m) &\equiv \langle \rho_m, u\rangle \\ &\equiv \int_0^T dt \int_0^T ds \int_0^L dx \int_0^L dy \left\{ \left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x} \right) \rho_m(x,t) \right\} \\ &\quad \times W_f(x,t,y,s) \left\{ \left(\frac{\partial}{\partial s} + c\frac{\partial}{\partial y} \right) \right\} u(y,s) \\ &\quad + \int_0^L dx \int_0^L dy\, \rho_m(x,0) W_i(x,y) u(y,0) \\ &\quad + \int_0^T dt \int_0^T ds\, \rho_m(0,t) W_b(t,s) u(0,s). \end{aligned} \tag{2.1.19}$$

Integrate the first integral by parts, and then compare (2.1.19) to

$$u(x_m,t_m) = \int_0^T dt \int_0^L dx\, u(x,t)\delta(x-x_m)\delta(t-t_m). \tag{2.1.20}$$

Since the partially-integrated (2.1.19) must agree with (2.1.20) for all fields $u(x,t)$ having initial values $u(x,0)$ and boundary values $u(0,t)$, we may equate their respective coefficients, arriving at (1.3.8)–(1.3.13). We have proved that for any field u,

$$\boxed{\langle \mathbf{r}, u\rangle = \mathbf{u}.} \tag{2.1.21}$$

□

Note 1. We have established that

$$\hat{u}(x,t) - u_F(x,t) = \hat{\beta}^{\mathrm{T}}\mathbf{r}(x,t). \tag{2.1.22}$$

That is, the difference between the inverse estimate $\hat{u}$ and the prior estimate u_F is a linear combination of the M representers $r_1, \ldots, r_M$. The difference lies in the *observable* space, that is, we reject any additional difference g that is *unobservable*: $\langle \mathbf{r}, g \rangle = 0$. We began with a search for the optimal or best-fit *field* $\hat{u}(x,t)$, where $0 \le x \le L$ and $0 \le t \le T$. This would be a search amongst an infinite number of degrees of freedom (the "state space"). We have exactly reduced the task to a search for the M optimal representer coefficients $\hat{\beta}_1, \ldots, \hat{\beta}_M$: see Fig. 2.1.1. These are the *observable* degrees of freedom (the "data space").

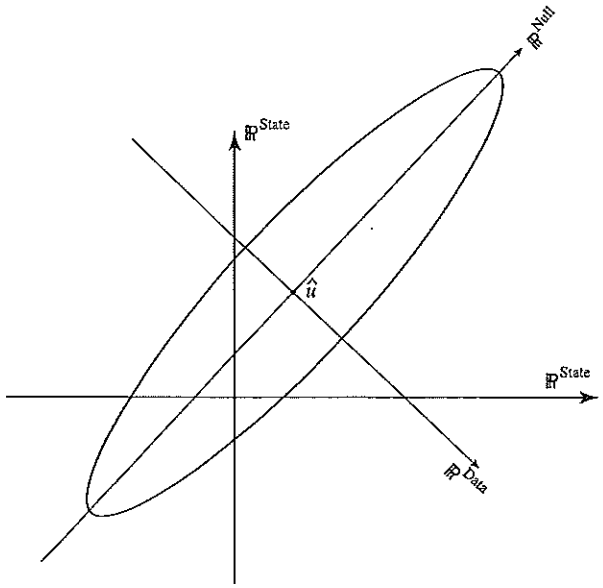

Figure 2.1.1 The plane represents the state space. It has an axis $u(x,t)$ for each (x,t) in the intervals $0 \le x \le L$, $0 \le t \le T$. In principle this is an infinite dimensional space. In practice, when we replace continuous intervals and partial differential equations with grids and partial difference equations, the state space usually has a very large but finite dimension. The contour is defined by a constant value for the penalty functional $\mathcal{J}[u]$, and has principal axes $\mathbb{R}^{\text{Data}}$ (for $\beta^{\mathrm{T}}\mathbf{r}(x,t)$) and $\mathbb{R}^{\text{Null}}$ (for $g(x,t)$). Note that the representers $r_1(x,t), \ldots, r_M(x,t)$ are known, and *span* the data space, so only the unknown $\beta_1, \ldots, \beta_M$ vary in the data space. The unobservable field $g(x,t)$ is unknown and variable for $0 \le x \le L$, $0 \le t \le T$. Realizing that $\hat{g}$ is zero greatly reduces the size of the search for $\hat{u}$, as we need only search in the data space.

Note 2. Recall from §1.3.3 the definition of Γ as the representer for point measurements. Hence for any field u,

$$\langle \Gamma, u \rangle = u, \tag{2.1.23}$$

and consequently Γ is known as a "reproducing kernel."

2.2 Statistical interpretation: the relationship to "optimal interpolation"

2.2.1 Random errors

Viewed as a generalized inverse or as a control problem, the ocean circulation u is estimated by adjusting the forcing f, initial value i, and boundary value b in order to obtain a better fit to the data, given that there are weights or "costs" W_f, W_i, W_b and $\mathbf{W}$ for these control variables and for the data misfit. Alternatively, the fields f, b and i for $0 \le x \le L$ and $0 \le t \le T$, and the data error vector ϵ, may be viewed as members of an *ensemble* of such quantities. That is, they are *random*. We shall attempt to estimate which f, i, b and ϵ were present in the "ocean" in $0 \le x \le L$ during our "cruise" for $0 \le t \le T$. We shall recover an interpretation of the "variational assimilation" of §1.2 in terms of the "optimal interpolation" routinely used in meteorology and oceanography.

2.2.2 Null hypotheses

In order to make these estimates, we shall have to make some assumptions about the ensemble. These assumptions compose a *null hypothesis* $\mathcal{H}_0$:

$$Ef(x,t) = Ei(x) = Eb(t) = E\epsilon_m = 0 \tag{2.2.1}$$

for $0 \le x \le L$, $0 \le t \le T$ and $1 \le m \le M$, where $E(\,)$ denotes the ensemble average or mean;

$$\begin{aligned} E(f(x,t)f(y,s)) &= C_f(x,t,y,s), \\ E(i(x)i(y)) &= C_i(x,y), \\ E(b(t)b(s)) &= C_b(t,s), \end{aligned} \tag{2.2.2}$$

and

$$E(\epsilon\epsilon^{\mathrm{T}}) = \mathbf{C}_\epsilon,$$

while

$$E(fi) = E(fb) = E(f\epsilon_m) = E(ib) = E(i\epsilon_m) = E(b\epsilon_m) = 0. \tag{2.2.3}$$

Note 1. The covariances C_f, C_i, C_b and $\mathbf{C}_\epsilon$ are explicit functional or tabular forms.
Note 2. Only first and second moments are given in $\mathcal{H}_0 = \{(2.2.1), (2.2.2), (2.2.3)\}$; if the random variables f, i, b and ϵ are Gaussian, these moments determine the probability distribution function (pdf).
Note 3. The *alternative hypothesis* is that either (2.2.1), (2.2.2), or (2.2.3) is *not* true.

2.2.3 The reproducing kernel is a covariance

Let us now define $v = v(x,t)$ by

$$u(x,t) = u_F(x,t) + v(x,t). \tag{2.2.4}$$

That is, v is the random error in our prior estimate or forward solution u_F. The latter corresponds to our prior estimates F, I and B for the forcing, etc. These estimates are made *prior* to knowing the data $\mathbf{d}$. Clearly

$$\frac{\partial v}{\partial t} + c\frac{\partial v}{\partial x} = f, \tag{2.2.5}$$

subject to $v = i$ at $t = 0$, and $v = b$ at $x = 0$. We may use the Green's function γ to write

$$v = \gamma \bullet f + \gamma \circ i + c\gamma * b. \tag{2.2.6}$$

Hence

$$Ev = 0. \tag{2.2.7}$$

Exercise 2.2.1
Derive in detail the covariance for v:

$$\begin{aligned} C_v \equiv E(vv) &= \gamma \bullet E(ff) \bullet \gamma + \gamma \circ E(ii) \circ \gamma + c^2\gamma * E(bb) * \gamma \\ &= \gamma \bullet C_f \bullet \gamma + \gamma \circ C_i \circ \gamma + c^2\gamma * C_b * \gamma \qquad (2.2.8) \\ &= \Gamma. \qquad (2.2.9) \end{aligned}$$

That is, the covariance of the errors in the prior estimate is just the reproducing kernel (Weinert, 1982; Bennett, 1992). □

2.2.4 "Optimal Interpolation", or best linear unbiased estimation; equivalence of generalized inversion and OI

We shall now outline the method of "optimal interpolation" (OI) for estimating a field u, given a first-guess u_F and data $\mathbf{d}$ (Bretherton *et al.*, 1976; Daley, 1991; Thiébaux and Pedder, 1987). The first guess need not be a model solution.

Suppose the true field is

$$u(x,t) = u_F(x,t) + q(x,t), \tag{2.2.10}$$

and suppose as before that

$$\mathbf{d} = \mathbf{u} + \epsilon, \tag{2.2.11}$$

where

$$Eq(x,t) = E\epsilon_m = 0 \tag{2.2.12}$$

for $0 \le x \le L$, $0 \le t \le T$ and $1 \le m \le M$; and suppose that

$$\left.\begin{aligned} &E(q(x,t)q(y,s)) = C_q(x,t,y,s), \\ &E(\epsilon\epsilon^{\mathrm{T}}) \quad\quad\quad = \mathbf{C}_\epsilon, \\ &\text{and} \\ &E(q(x,t)\epsilon) \quad\quad = \mathbf{0} \end{aligned}\right\} \tag{2.2.13}$$

for $0 \le x, y \le L$, $0 \le t, s \le T$.

Note 1. $E(\,)$ denotes an ensemble average, given the prior estimate u_F.
Note 2. $Eu = Eu_F + Eq = u_F + 0 = u_F$.
Note 3. $E\mathbf{d} = E\mathbf{u} + E\epsilon = E\mathbf{u}_F + E\mathbf{q} + E\epsilon = \mathbf{u}_F + \mathbf{0} + \mathbf{0} = \mathbf{u}_F$.

We seek the best linear unbiased estimate of u, that is

$$u_L(x,t) = u_F(x,t) + (\mathbf{d} - \mathbf{u}_F)^{\mathrm{T}}\mathbf{s}(x,t), \tag{2.2.14}$$

where $s_1(x,t), \ldots, s_M(x,t)$ are M as yet unchosen non-random interpolants.

Note 1. u_L is linear in u_F and $\mathbf{d}$.
Note 2. u_L is unbiased: $Eu_L = u_F = Eu$.
Note 3. u_L is best if

$$Ee_L^2(x,t) \equiv E\{u(x,t) - u_L(x,t)\}^2 \tag{2.2.15}$$

is least for *each* (x,t).

Now

$$\begin{aligned} Ee_L^2 &= E\{u_F + q - u_F - (\mathbf{u}_F + \mathbf{q} + \epsilon - \mathbf{u}_F)^{\mathrm{T}}\mathbf{s}\}^2 \qquad (2.2.16)\\ &= E\{q - (\mathbf{q}+\epsilon)^{\mathrm{T}}\mathbf{s}\}^2 \\ &= Eq^2 - E\{q(\mathbf{q}+\epsilon)^{\mathrm{T}}\}\,\mathbf{s} - \mathbf{s}^{\mathrm{T}}E\{(\mathbf{q}+\epsilon)q\} \\ &\quad + \mathbf{s}^T E\{(\mathbf{q}+\epsilon)(\mathbf{q}+\epsilon)^{\mathrm{T}}\}\mathbf{s} \qquad (2.2.17) \end{aligned}$$

for each (x,t). We want

$$\frac{\partial Ee_L^2}{\partial s_m} = 0 \tag{2.2.18}$$

for $1 \le m \le M$, at each (x, t). That is,

$$-E\{(\mathbf{q}+\epsilon)q\} + E\{(\mathbf{q}+\epsilon)(\mathbf{q}+\epsilon)^{\mathrm{T}}\}\,\mathbf{s} = \mathbf{0},$$

or

$$-E(\mathbf{q}q) + \{E(\mathbf{q}\mathbf{q}^{\mathrm{T}}) + E(\epsilon\epsilon^{\mathrm{T}})\}\,\mathbf{s} = \mathbf{0} \tag{2.2.19}$$

since $E(q\epsilon) = \mathbf{0}$ by assumption. In detail, (2.2.19) is

$$-C_q(x, t, x_n, t_n) + \sum_{m=1}^{M} \{C_q(x_n, t_n, x_m, t_m) + C_{\epsilon_{n,m}}\} s_m(x, t) = 0 \tag{2.2.20}$$

for $0 \le x \le L$, $0 \le t \le T$ and $1 \le n \le M$. Solving (2.2.20) for s_n yields

$$s_n(x, t) = \sum_{m=1}^{M} \{\mathbf{C}_q + \mathbf{C}_\epsilon\}^{-1}_{n,m} C_q(x, t, x_m, t_m), \tag{2.2.21}$$

where the superscript "-1" indicates a matrix inverse, and $\{\mathbf{C}_q\}_{n,m} = C_q(x_n, t_n, x_m, t_m)$. These $\mathbf{s}(x, t)$ are the optimal interpolants. They do not depend upon u_F or $\mathbf{d}$, but do depend upon the prior covariances C_q and $\mathbf{C}_\epsilon$. In conclusion our best linear unbiased estimate or BLUE is

$$u_L(x, t) = u_F(x, t) + (\mathbf{d} - \mathbf{u}_F)^{\mathrm{T}} \{\mathbf{C}_q + \mathbf{C}_\epsilon\}^{-1} \mathbf{C}_q(x, t). \tag{2.2.22}$$

Now compare (2.2.22) with (1.3.24), and recall the first line in (1.3.32).

We have proved:

> Generalized inversion (the minimization of the integral penalty functional $\mathcal{J}[u]$) is the same as optimal interpolation (the minimization of the local error variance $Ee_L^2(x, t)$) when the solution of the forward model u_F is the mean field, when the data weight matrix $\mathbf{w}$ is the inverse of the data error covariance matrix $\mathbf{C}_\epsilon$, and when the reproducing kernel Γ is the covariance C_q. In particular (see (2.2.8), (2.2.9)),
>
> $$r_m(x, t) = \Gamma(x, t, x_m, t_m) = C_v(x, t, x_m, t_m) = C_q(x, t, x_m, t_m),$$
> $$R_{nm} = \Gamma(x_n, t_n, x_m, t_m) = C_v(x_n, t_n, x_m, t_m) = C_q(x_n, t_n, x_m, t_m).$$

Generalized Inversion is Optimal Interpolation

Note 1. Our model is linear, and the data are pointwise.

Note 2. OI is widely used in meteorology and oceanography, for the "analysis" or "mapping" of scalar data when the statistical properties of the fields are plausibly independent of coordinate origins or orientations, both in space and time. That is, when

$$C_q(x, t, y, s) = C_q(|x - y|,\ |t - s|), \tag{2.2.23}$$

for example. Such covariances need only involve a few parameters, which should be reliably estimable from reasonably large sets of data. However, we are increasingly obliged to admit that different fields are dependent, on dynamical or chemical or biological grounds, so we should use multivariate or vector forms of OI. Moreover, planetary-scale and coastal circulation are obviously statistically inhomogeneous, while the endless emergence of trends suggests statistical nonstationarity. That is, (2.2.23) is false. OI may be generalized to the multivariate, inhomogeneous and nonstationary case provided that there are credible prior estimates for all the parameters in the covariances of the fields being mapped. We hope that our dynamical models are getting so faithful to the larger scales that model errors like f must be limited to the smaller scales at which (2.2.23) may be plausible. Thus, we should only need to estimate $C_f(x, t, y, s) = C_f(|x - y|, |t - s|)$. We may then use generalized inversion to generate, in effect, the inhomogeneous and nonstationary multivariate equivalents of $C_v = C_v(x, t, y, s)$ and then perform, in effect, an OI of the data.

A serious caution must now be offered. It is misleadingly easy to declare that the dynamical error f, initial error i, etc., are random variables belonging to some ensemble, and to manipulate their ensemble moments Ef, Ei, $E(ff)$, $E(fi)$, etc. It is much harder to devise a credible method for estimating these moments. The fields must clearly be statistically homogeneous at least in one spatial direction or in time, but the presence of spatial or climatological trends makes such homogeneity far from clear. Worse, our dynamical models have already been Reynolds-averaged or subgridscale-averaged, so f in particular is already an average of a certain kind. The statistical interpretation of variational assimilation requires, therefore, a *second* randomization. This difficult issue will be discussed in greater detail in §5.3.7.

2.3 The reduced penalty functional

2.3.1 Inversion as hypothesis testing

Inverse methods enable us to smooth data using a dynamical model as a constraint. Equally, the methods enable us to test the model using the data. The concept of a model is extended here to include not only equations of motion, initial conditions and boundary conditions, but also an hypothesis concerning the errors in each such piece of information. If the model fails the test for a given data set, then the interpolated data or "analysis field" is suspect. If the test is failed repeatedly for many data sets, then the hypothesis is suspect. This would be an unsatisfactory state of affairs from the point of view of the ocean analyst or ocean forecaster, but should please the ocean modeler: something new would have been learned about the ocean, namely, that the errors in the dynamics, initial conditions or boundary conditions had been underestimated. Lagrange multipliers make it possible (exercise!) to distinguish between forcing errors and additive components of parameterization errors.

The hypothesis test is based on the statistical interpretation developed in the previous section. The derivation of the test is very short, but the realization that inverse methods enable model testing is so important that a separate section is warranted.

2.3.2 Explicit expression for the reduced penalty functional

First, recall from (1.3.26) and §2.2.3 that the minimum value of the penalty functional $\mathcal{J}$ is

$$\begin{aligned}\hat{\mathcal{J}} \equiv \mathcal{J}[\hat{u}] &= \mathbf{h}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{h} \\ &= (\mathbf{d}-\mathbf{u}_F)^{\mathrm{T}}(\mathbf{R}+\mathbf{C}_\epsilon)^{-1}(\mathbf{d}-\mathbf{u}_F),\end{aligned} \tag{2.3.1}$$

where the actual or true "ocean circulation" is

$$u = u_F + v, \tag{2.3.2}$$

where v is the model response to the random inputs f, i and b, and the data are

$$\mathbf{d} = \mathbf{u} + \epsilon, \tag{2.3.3}$$

where ϵ is random measurement error.

Hence

$$\mathbf{h} \equiv \mathbf{d} - \mathbf{u}_F = \mathbf{u} + \epsilon - (\mathbf{u} - \mathbf{v}) = \epsilon + \mathbf{v}, \tag{2.3.4}$$

$$E\mathbf{h} = E\epsilon + E\mathbf{v} = \mathbf{0}, \tag{2.3.5}$$

$$\begin{aligned}E(\mathbf{h}\mathbf{h}^{\mathrm{T}}) &= E((\epsilon+\mathbf{v})(\epsilon+\mathbf{v})^{\mathrm{T}}) \\ &= E(\epsilon\epsilon^{\mathrm{T}}) + E(\epsilon\mathbf{v}^{\mathrm{T}}) + E(\mathbf{v}\epsilon^{\mathrm{T}}) + (\mathbf{v}\mathbf{v}^{\mathrm{T}}) \\ &= \mathbf{C}_\epsilon + \mathbf{0} + \mathbf{0} + \mathbf{R} = \mathbf{P},\end{aligned} \tag{2.3.6}$$

which statements are parts of, or consequences of, our null hypothesis $\mathcal{H}_0$ defined by (2.2.1)–(2.2.3).

Now define $\mathbf{P}^{\frac{1}{2}}$, which is meaningful since $\mathbf{P}$ is positive-definite and symmetric, and hence define

$$\mathbf{k} \equiv \mathbf{P}^{-\frac{1}{2}}\mathbf{h}. \tag{2.3.7}$$

Then

$$E\mathbf{k} = \mathbf{P}^{-\frac{1}{2}}\,E\mathbf{h} = \mathbf{0}, \tag{2.3.8}$$

$$\begin{aligned}E(\mathbf{k}\mathbf{k}^{\mathrm{T}}) &= \mathbf{P}^{-\frac{1}{2}}\,E(\mathbf{h}\mathbf{h}^{\mathrm{T}})\,\mathbf{P}^{-\frac{1}{2}} \\ &= \mathbf{P}^{-\frac{1}{2}}\mathbf{P}\mathbf{P}^{-\frac{1}{2}} \\ &= \mathbf{I}.\end{aligned} \tag{2.3.9}$$

That is,

$$E(k_n k_m) = \delta_{nm}. \tag{2.3.10}$$

2.3.3 Statistics of the reduced penalty: χ^2 testing

The scaled, prior data misfits $k_1, \ldots, k_M$ are zero-mean, uncorrelated, unit-variance random variables and

$$\begin{aligned} \hat{\mathcal{J}} = \mathbf{h}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{h} &= \mathbf{k}^{\mathrm{T}}\mathbf{P}^{\frac{1}{2}}\,\mathbf{P}^{-1}\mathbf{P}^{\frac{1}{2}}\,\mathbf{k} \\ &= \mathbf{k}^{\mathrm{T}}\mathbf{k} \\ &= k_1^2 + \cdots + k_M^2. \end{aligned} \tag{2.3.11}$$

Therefore $\hat{\mathcal{J}} = \chi^2_M$, the chi-squared random variable with M degrees of freedom. Or is it? To be precise,

$$\chi^2_M = x_1^2 + \cdots + \cdots + x_M^2, \tag{2.3.12}$$

where the pdf for x_m is

$$p(x_m) = (2\pi)^{-\frac{1}{2}} \exp\left(-x_m^2/2\right), \tag{2.3.13}$$

$1 \le m \le M$. That is, each x_m is a Gaussian random variable having zero mean and unit variance: $x_m \sim N(0, 1)$ (Press *et al.*, 1986). If we had included in $\mathcal{H}_0$ the assumption that f, i, b and ϵ were Gaussian, then by linearity $\mathbf{h}$ and hence $\mathbf{k}$ would also be Gaussian. If we do not make that assumption, we may invoke the *central limit theorem* when M is large, to infer that

$$k_n = \sum_{m=1}^{M} (\mathbf{P}^{-\frac{1}{2}})_{nm} h_m \sim N(0, 1) \tag{2.3.14}$$

as $M \to \infty$.

But roughly, if $\mathcal{H}_0$ is true then

$$\hat{\mathcal{J}} = \chi^2_M. \tag{2.3.15}$$

So we have a chi-squared test for our null hypothesis. Now

$$E\left(\chi^2_M\right) = M, \ \mathrm{var}\left(\chi^2_M\right) \equiv E\left(\left(\chi^2_M\right)^2\right) - \left(E\left(\chi^2_M\right)\right)^2 = 2M. \tag{2.3.16}$$

If we perform the inversion a number of times with different data, and find that our sample distribution has significantly bigger first or second moments than those of χ^2_M, then we should *reject* $\mathcal{H}_0$. *We would have learned something about the ocean, from the data.* Specifically, we would have learned that the ocean differs from the model, by *more* than we had hypothesized. Recall that $\mathcal{J}$ is inversely proportional to C_f, etc.

Exercise 2.3.1

Show that

(i) $\mathcal{J}_F \equiv \mathcal{J}[u_F] = \mathbf{h}^{\mathrm{T}}\mathbf{C}_\epsilon^{-1}\,\mathbf{h}$, (2.3.17)

(ii) $\hat{\mathcal{J}}_{\text{mod}} \equiv (\hat{u} - u_F, \hat{u} - u_F) = \hat{\beta}^{\mathrm{T}} \mathbf{R} \hat{\beta}$, (2.3.18)

(iii) $\hat{\mathcal{J}}_{\text{data}} \equiv (\hat{\mathbf{u}} - \mathbf{d})^{\mathrm{T}} \mathbf{C}_\epsilon^{-1} (\hat{\mathbf{u}} - \mathbf{d}) = \hat{\beta}^{\mathrm{T}} \mathbf{C}_\epsilon \hat{\beta}$, (2.3.19)

where $\mathbf{P}\hat{\beta} = \mathbf{h}$.

Note 1. $\mathcal{J}_F$ is "only data misfit".
Note 2. In general, $\hat{\mathcal{J}}_{\text{mod}} \neq \hat{\mathcal{J}}_{\text{data}}$. □

Exercise 2.3.2
Show that

(i) $E\mathcal{J}_F = \text{Tr}\left(\mathbf{C}_\epsilon^{-\frac{1}{2}} \mathbf{R} \mathbf{C}_\epsilon^{-\frac{1}{2}}\right) + M$, (2.3.20)

(ii) $E\hat{\mathcal{J}}_{\text{mod}} = \text{Tr}\left(\mathbf{R}^{\frac{1}{2}} \mathbf{P}^{-1} \mathbf{R}^{\frac{1}{2}}\right)$, (2.3.21)

(iii) $E\hat{\mathcal{J}}_{\text{data}} = \text{Tr}\left(\mathbf{C}_\epsilon^{\frac{1}{2}} \mathbf{P}^{-1} \mathbf{C}_\epsilon^{\frac{1}{2}}\right)$. (2.3.22)

Note 1. Usually $E\mathcal{J}_F \gg E\hat{\mathcal{J}} = M$.
Note 2. In general, $E\hat{\mathcal{J}}_{\text{mod}} \neq E\hat{\mathcal{J}}_{\text{data}}$.
Note 3. In order to devise a rigorous and objective test for an ocean model, we have extended the definition of a model to include an hypothesis about the statistics of the errors in the dynamics, in the initial conditions and in the boundary conditions as well as in the data. □

Exercise 2.3.3
Give meanings to the left-hand sides of (2.3.17)–(2.3.22). □

Exercise 2.3.4 (Bennett *et al.*, 2000)
Show that

(i) $\text{var}(\mathcal{J}_F) \sim 2\,\text{Tr}\,(\mathbf{C}_\epsilon^{-1} \mathbf{P}^2 \mathbf{C}_\epsilon^{-1})$, (2.3.23)

(ii) $\text{var}(\hat{\mathcal{J}}_{\text{data}}) \sim 2\,\text{Tr}\,(\mathbf{C}_\epsilon \mathbf{P}^{-2} \mathbf{C}_\epsilon)$, (2.3.24)

(iii) $\text{var}(\hat{\mathcal{J}}_{\text{mod}}) \sim 2\,\text{Tr}\,[(\mathbf{I} - \mathbf{P}^{-\frac{1}{2}} \mathbf{C}_\epsilon \mathbf{P}^{-\frac{1}{2}})^2]$ (2.3.25)

as $M \to \infty$. □

Remark
It is difficult to develop a credible null hypothesis $\mathcal{H}_0$. In particular it is difficult to develop the covariances C_f, etc. It follows that $\hat{u}$, the resulting inverse estimate or analysis of the circulation, also lacks credibility. It is a misconception, however, to view inversion as "garbage in, garbage out". Rather, inversion puts the hypothesis to the test. Forward modeling is no less exposed to the charge of "garbage in, garbage

out": it tests the nullest of null hypotheses, namely, that the dynamical errors, initial errors and boundary errors are all zero, which is the rankest of garbage.

2.4 General measurement

2.4.1 Point measurements

Our data thus far have been direct measurements of the circulation field u at isolated points in space and time:

$$d_m = u(x_m, t_m) + \epsilon_m \tag{2.4.1}$$

for $1 \leq m \leq M$, where d_m is the datum and ϵ_m the measurement error. We shall now consider more general measurements (Bennett, 1985, 1990).

2.4.2 Measurement functionals

First note that a map sending a *field* u into a *single real number* $u(z, w)$ is an example of a *functional*

$$u \to \mathcal{L}[u] = u(z, w). \tag{2.4.2}$$

In general u may be a *vector field* of velocity components, pressure, temperature etc., but a measurement of u produces a *single number*. If the field is a streamfunction ψ, and the datum is the meridional component of velocity collected from a current meter, then the appropriate functional is

$$\psi \to \frac{\partial \psi}{\partial x}(z, w); \tag{2.4.3}$$

if the field is sea-level elevation h, and the datum is the vertical acceleration of a wave-rider buoy, then

$$h \to \frac{\partial^2 h}{\partial t^2}(z, w); \tag{2.4.4}$$

if the field is fluid velocity u along a zonal acoustic path, and the datum derives from reciprocal-shooting tomography, then

$$u \to \int_{z_1}^{z_2} u(x, w)\, dx; \tag{2.4.5}$$

if the field is sea-level elevation, and the datum is collected by a radar beam, then

$$h \to \int_0^T dt \int_0^L dx\, K(x, t) h(x, t); \tag{2.4.6}$$

finally, if the field is stratospheric temperature θ and the datum is a radiative energy flux, then by Stefan's law,

$$\theta \to \theta^4(z, w). \tag{2.4.7}$$

Note 1. Examples (2.4.2)–(2.4.6) are *linear*:

$$\mathcal{L}[au + bv] = a\mathcal{L}[u] + b\mathcal{L}[v] \tag{2.4.8}$$

for any fields u, v and real numbers a, b.

Note 2. Example (2.4.7) is *nonlinear*.

Note 3. Each of (2.4.2)–(2.4.5) can be expressed as (2.4.6):

$$K(x, t) = \delta(x - z)\delta(t - w) \quad \text{for} \quad (2.4.2),$$
$$K(x, t) = -\delta'(x - z)\delta(t - w) \quad \text{for} \quad (2.4.3),$$
$$K(x, t) = \delta(x - z)\delta''(t - w) \quad \text{for} \quad (2.4.4).$$

Exercise 2.4.1

Find K for (2.4.5). □

2.4.3 Representers for linear measurement functionals

The penalty functional may be now expressed as

$$\mathcal{J}[u] = (u - u_F, u - u_F) + (\mathbf{d} - \mathcal{L}[u])^{\mathrm{T}}\mathbf{C}_\epsilon^{-1}(\mathbf{d} - \mathcal{L}[u]), \tag{2.4.9}$$

where $\mathcal{L}^{\mathrm{T}} = (\mathcal{L}_1, \ldots, \mathcal{L}_M)$ indicates M linear measurement functionals. Note that in earlier sections,

$$\mathcal{L}_m[u] \equiv u(x_m, t_m). \tag{2.4.10}$$

Furthermore, the Riesz representation theorem establishes that if $u \to \mathcal{L}_m[u]$ is a *bounded* ($\sup |\mathcal{L}_m[u]| / \| u \| < \infty$) linear functional acting on a Hilbert space, then there is an element (a field) r_m in the space such that

$$(u, r_m) = \mathcal{L}_m[u] \tag{2.4.11}$$

for any field u.

Exercise 2.4.2

Verify that

$$r_m(x, t) = \mathcal{L}_{m_{(y,s)}}[\Gamma(x, t, y, s)], \tag{2.4.12}$$

where Γ is the reproducing kernel, and the subscripts (y, s) indicate that $\mathcal{L}_m$ acts on Γ as a field over (y, s), for each (x, t). Recall that Γ is the representer for evaluation of u at (y, s). In fact, show that r_m satisfies

$$\frac{\partial r_m}{\partial t} + c\frac{\partial r_m}{\partial x} = C_f \bullet \alpha_m, \tag{2.4.13}$$

subject to $r_m = C_i \circ \alpha_m$ at $t = 0$, and $r_m = cC_b * \alpha_m$ at $x = 0$, where

$$-\frac{\partial \alpha_m}{\partial t} - c\frac{\partial \alpha_m}{\partial x} = \mathcal{L}_{m_{(y,s)}}[\delta(x - y)\delta(t - s)], \tag{2.4.14}$$

subject to $\alpha_m = 0$ at $t = T$, and $\alpha_m = 0$ at $x = L$. □

We may now write

$$\mathcal{J}[u] = \langle u - u_F, u - u_F \rangle + (\mathbf{d} - \langle u, \mathbf{r} \rangle)^{\mathrm{T}} \mathbf{C}_\epsilon^{-1} (\mathbf{d} - \langle u, \mathbf{r} \rangle) \tag{2.4.15}$$

and, as before, the minimizer is

$$\hat{u} = u_F + \mathbf{h}^{\mathrm{T}} \mathbf{P}^{-1} \mathbf{r}, \tag{2.4.16}$$

where

$$\mathbf{h} \equiv \mathbf{d} - \mathcal{L}[u_F] \tag{2.4.17}$$

and $\mathbf{P} = \mathbf{R} + \mathbf{C}_\epsilon$, where

$$\mathbf{R} = \langle \mathbf{r}, \mathbf{r}^{\mathrm{T}} \rangle = \mathcal{L}[\mathbf{r}^{\mathrm{T}}] = \text{“}\mathcal{L}[\Gamma]\mathcal{L}^{\mathrm{T}}\text{”}. \tag{2.4.18}$$

From now on we shall assume that $\mathbf{r}$, with its adjoint field α, represents a general linear measurement functional. We shall reserve the notation and nomenclature of the rk $\Gamma = \Gamma(x, t, z, w)$, with its adjoint field the Green's function $\gamma = \gamma(x, t, z, w)$, for the *evaluation functional* (2.4.2).

Exercise 2.4.3

Derive the Euler–Lagrange equations for extrema of (2.4.9). In particular, show that the generalization of (1.3.1) is

$$-\frac{\partial \lambda}{\partial t} - c\frac{\partial \lambda}{\partial x} = \mathcal{L}^{\mathrm{T}}[\delta\delta] \mathbf{C}_\epsilon^{-1} (\mathbf{d} - \mathcal{L}[\hat{u}]). \tag{2.4.19}$$

□

2.5 Array modes

2.5.1 Stable combinations of representers

We have seen that, amongst all free and forced solutions of the forward model, the observing system or "array" only detects the representers. We now ask: are some combinations of representers more stably detected than others?

2.5.2 Spectral decomposition, rotated representers

Assume general linear measurement functionals $\mathcal{L}=(\mathcal{L}_1, \ldots, \mathcal{L}_M)^{\mathrm{T}}: u \to \mathcal{L}[u] \in \mathbb{R}^M$. That is, $\mathcal{L}$ maps the field u linearly into the M real numbers $\mathcal{L}[u]$. The data $\mathbf{d}$ are of the form $\mathbf{d} = \mathcal{L}[u] + \epsilon$, where ϵ is the vector of measurement errors. The representer matrix is

$$R_{nm} = \mathcal{L}_{n_{(x,t)}} \mathcal{L}_{m_{(y,s)}} [\Gamma(x, t, y, s)] \tag{2.5.1}$$

for $1 \leq n, m \leq M$, where $\mathcal{L}_{n_{(x,t)}}$ acts on $\Gamma(x, t, \ , \ )$, etc. In vector notation,

$$\mathbf{R} = \mathcal{L}\Gamma\mathcal{L}^{\mathrm{T}}. \tag{2.5.2}$$

Recall again that the reproducing kernel Γ is also the covariance C_v: see §2.2, Exercise 2.2.1. The minimization of the penalty functional $\mathcal{J}$, defined by (1.5.7), reduces to the solution of the M-dimensional linear system

$$\mathbf{P}\hat{\beta} \equiv (\mathbf{R} + \mathbf{C}_\epsilon)\hat{\beta} = \mathbf{h} \equiv \mathbf{d} - \mathcal{L}[u_F], \tag{2.5.3}$$

where u_F is the solution of the forward model. The representer matrix $\mathbf{R}$ depends upon the dynamics, the prior covariances C_f, C_i and C_b for dynamical, initial and boundary residuals, and upon the array $\mathcal{L}$, while $\mathbf{C}_\epsilon$ is the covariance of measurement errors. Thus $\mathbf{P}$ encapsulates all of our prior knowledge of the ocean in general but does not depend upon the prior estimates of forcing, initial and boundary values F, B and I, provided the dynamics and measurement functionals are linear. The symmetry and positive definiteness of $\mathbf{P}$ implies the spectral decomposition

$$\mathbf{P} = \mathbf{Z}\Phi\mathbf{Z}^{\mathrm{T}}, \tag{2.5.4}$$

where $\mathbf{Z}$ is orthogonal: $\mathbf{Z}\mathbf{Z}^{\mathrm{T}} = \mathbf{Z}^{\mathrm{T}}\mathbf{Z} = \mathbf{I}$, and Φ is diagonal: $\Phi = \mathrm{diag}\,(\phi_1, \ldots, \phi_M)$, where $\phi_1 \geq \cdots \geq \phi_M > 0$.

Let $\mathcal{L}'$ be a rotated vector of measurement functionals:

$$\mathcal{L}' \equiv \mathbf{Z}^{\mathrm{T}}\mathcal{L}, \tag{2.5.5}$$

and define the rotated representers $\mathbf{r}' = \mathbf{r}'(x, t)$ by

$$\mathbf{r}' \equiv \mathbf{Z}^{\mathrm{T}}\mathbf{r} = \mathbf{Z}^{\mathrm{T}}\mathcal{L}[\Gamma]. \tag{2.5.6}$$

These are the *array modes* (Bennett, 1985, 1992). In particular,

$$\begin{aligned}\mathbf{R}' &= \mathcal{L}'\Gamma\mathcal{L}'^{\mathrm{T}} = \mathbf{Z}^{\mathrm{T}}\mathcal{L}\Gamma\mathcal{L}^{\mathrm{T}}\mathbf{Z} \\ &= \mathbf{Z}^{\mathrm{T}}\mathbf{R}\mathbf{Z},\end{aligned} \tag{2.5.7}$$

while

$$\mathbf{C}'_\epsilon = \mathbf{Z}^{\mathrm{T}}\mathbf{C}_\epsilon\mathbf{Z}. \tag{2.5.8}$$

Hence

$$\mathbf{P}' = \mathbf{R}' + \mathbf{C}'_\epsilon = \mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{Z} = \boldsymbol{\Phi}, \tag{2.5.9}$$

which is diagonal. The rotated representer coefficients $\hat{\boldsymbol{\beta}}'$ then obey

$$\mathbf{P}'\hat{\boldsymbol{\beta}}' = \mathbf{h}', \tag{2.5.10}$$

where

$$\mathbf{h}' = \mathbf{Z}^{\mathrm{T}}\mathbf{h}. \tag{2.5.11}$$

2.5.3 Statistical stability, clipping the spectrum

The solution for $\hat{\boldsymbol{\beta}}'$ is trivial, since $\mathbf{P}' = \boldsymbol{\Phi}$ is diagonal:

$$\hat{\beta}'_m = \frac{h'_m}{\phi_m} \tag{2.5.12}$$

for $1 \le m \le M$. We may deduce from (2.3.5) and (2.3.6) that

$$E\mathbf{h}' = \mathbf{0}, \;\; E(\mathbf{h}'\mathbf{h}'^{\mathrm{T}}) = \mathbf{P}' = \boldsymbol{\Phi}, \tag{2.5.13}$$

so

$$E\hat{\beta}'_m = 0, \quad E((\hat{\beta}'_m)^2) = \frac{E((h'_m)^2)}{\phi_m^2} = \frac{\phi_m}{\phi_m^2} = \phi_m^{-1}. \tag{2.5.14}$$

That is, the estimated array mode coefficients $\hat{\beta}'_m$ have greater variance if the corresponding eigenvalue ϕ_m is smaller; (2.5.12) shows the inverse to be unstable if the prior data misfit $\mathbf{h}$ projects significantly onto eigenvectors of $\mathbf{P}$ having very small eigenvalues. Such projections should be discarded for $m > m_c$, where m_c is some cut-off. The exact inverse is

$$\hat{u} = u_F + \mathbf{r}^{\mathrm{T}}\hat{\boldsymbol{\beta}} = u_F + \mathbf{r}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\hat{\boldsymbol{\beta}} = u_F + \mathbf{r}'^{\mathrm{T}}\hat{\boldsymbol{\beta}}' = u_F + \mathbf{r}'^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{h}', \tag{2.5.15}$$

or

$$\hat{u}(x,t) = u_F(x,t) + \sum_{m=1}^{M} r'_m(x,t)\phi_m^{-1}h'_m, \tag{2.5.16}$$

and so the stabilized approximation is

$$\hat{u}(x,t) \cong u_F(x,t) + \sum_{m=1}^{m_c} r'_m(x,t)\phi_m^{-1}h'_m. \tag{2.5.17}$$

Array modes $r'_{m_c+1}, \ldots, r'_M$ have been made redundant.

Note 1. The components of the vector of rotated measurement functionals need not correspond to individual elements in the array. They correspond to linear combinations of the elements.

Note 2. If we arbitrarily make $\mathbf{P}$ more diagonally dominant:

$$\mathbf{P} \to \mathbf{P} + \sigma^2\mathbf{I}, \tag{2.5.18}$$

where σ^2 is additional, independent measurement error variance, then the eigenvalues of $\mathbf{P}$ become $\phi_1 + \sigma^2, \ldots, \phi_M + \sigma^2$, which all exceed σ^2. Thus (2.5.18) would seem to stabilize the inverse. Spectral decomposition (2.5.4), rotation (2.5.6) or clipping (2.5.17) would not be required. However, $\mathbf{P}$ and $\mathbf{P} + \sigma^2\mathbf{I}$ have the same eigenvectors, so the array modes are unaffected by (2.5.18). The modes $r'_{m_c}(x,t), \ldots, r'_M(x,t)$ usually have very fine structure, and retaining them at almost any level yields a "noisy" inverse $\hat{u}(x,t)$. It is better to clip the spectrum of $\mathbf{P}$ than to make $\mathbf{P}$ more diagonally dominant.

Note 3. The construction of array modes is essentially an analysis of the condition or stability of the generalized inverse of the model plus array, that is, the stability of the minimization of the penalty functional denoted by (1.5.7), (1.5.9) or (2.1.2). There are two major steps in constructing the inverse. The first is the discard (2.1.11) of all the unobservable fields (2.1.8); it is effected by admitting only solutions of the Euler–Lagrange equations. The second step is the solution of the finite-dimensional linear system denoted as (2.1.15) or (2.5.3). Once this system is solved, the coupling in the Euler–Lagrange equations is resolved and the generalized inverse is finally obtained by the explicit assembly of (1.3.24), or equivalently by a backward integration followed by a forward integration (see §3.1.2). The dimension of the algebraic system is M, the total number of data. The condition of the system is determined by the M eigenvalues of the coefficient matrix $\mathbf{P} = \mathbf{R} + \mathbf{C}_\epsilon$. The essential point is that the condition of the inverse is determined without first making a numerical approximation to the model using, say, finite differences; the condition is determined at the continuum level. That is, the condition is set by the partial differential equations, initial conditions, and boundary conditions of the model, by the measurement functionals for the observing system or array, and by the form and weighting of the penalty functional (the actual inputs to the model: internal forcing, initial values, boundary values and data values, have no influence; the stability of the inverse is its sensitivity to them as a class). The inevitable numerical approximation will indeed modify the null space of unobservable

fields somewhat, and will also alter the eigenvalues, especially the smallest, but these effects are spurious and are suppressed in practice by physical diffusion in the dynamics, by convolution with the covariances in the Euler–Lagrange equations, and by the measurement error variance which has a stabilizing influence in general. Nevertheless, the continuum and discrete analyses of condition make for an interesting comparison. They may be found in Bennett (1985) and Courtier *et al.* (1993), respectively.

To end with a caution, it is imperative to realize that the array modes and assessment of conditioning depend not only upon the dynamics of the ocean model and the structure of the observing system or array, but also upon the hypothesized or prior covariances of the errors in the model and observing system. If subsequent testing of the hypothesis, using data collected by the array, leads to a rejection of the hypothesis, then the array assessment must also be rejected. Model testing and array assessment are inextricably intertwined. Examples will be presented in Chapter 5. For another approach to array design, see Hackert *et al.* (1998).

2.6 Smoothing norms, covariances and convolutions

2.6.1 Interpolation theory

The mathematical theory of interpolation is very old. It attracted the attention of the founders of analysis, including Newton, Lagrange and Gauss. The subject was in an advanced state of development by 1940; it then experienced a major reinvigoration with the advent of electronic computers. See Press *et al.* (1986; Section 2) for a neat outline of common methods, and Daley (1991; Chapter 2) for an authoritative account of methods widely used in meteorology and oceanography. What follows here is a brief outline of the theory attributed to E. Parzen, linking analytical and statistical interpolation. Aside from offering deeper insight into penalty functionals, the theory enables us to design and "tune" roughness penalties essentially equivalent to prescribed covariances (and vice versa). This is of critical importance if one intends, either out of taste or necessity, to minimize a penalty functional by searching in the control subspace rather than in the data subspace. The former search requires roughness penalties or weighting operators; the latter search exploits the Euler–Lagrange equations which incorporate covariances.

It has been argued in §2.1 that the data-subspace search is in principle highly efficient, but this efficiency will be wasted if the convolution-like integrals of the covariances and adjoint variables appearing in the Euler–Lagrange equations cannot be computed quickly. Fast convolution methods for standard covariances are given here; the methods are critical to the feasibility of data-subspace searches and hence generalized inversion itself. The section ends with some technical notes on rigorous inferences from penalty functionals, and on compounding covariances.

2.6.2 Least-squares smoothing of data; penalties for roughness

Let us set aside dynamics for now. Just consider interpolating some simple data $d_1, \ldots, d_M$, which are erroneous measurements of the scalar field $u = u(\mathbf{x})$ at the points $\mathbf{x}_1, \ldots, \mathbf{x}_M$. For simplicity, assume that $\mathbf{x}$ is planar: $\mathbf{x} = (x, y)$. We may define a quadratic penalty functional by

$$\mathcal{J}_0 = \mathcal{J}_0[u] = W_0 \iint_{\mathcal{D}} u^2 d\mathbf{x} + w|\,\mathbf{u} - \mathbf{d}\,|^2, \tag{2.6.1}$$

where $\mathcal{D}$ is some planar domain, W_0 and w are positive weights, and $\mathbf{u} = (u(\mathbf{x}_1), \ldots, u(\mathbf{x}_M))^{\mathrm{T}}$. If $\hat{u} = \hat{u}(\mathbf{x})$ is an extremum of $\mathcal{J}$, then the calculus of variations implies that

$$W_0\hat{u}(\mathbf{x}) = -w\delta^{\mathrm{T}}(\hat{\mathbf{u}} - \mathbf{d}), \tag{2.6.2}$$

where $\delta^{\mathrm{T}} = \delta^{\mathrm{T}}(\mathbf{x}) = (\delta(x - x_1)\delta(y - y_1), \ldots, \delta(x - x_M)\delta(y - y_M))$. So the "smallest" field that "nearly" fits the data is a crop of delta-functions. This is hardly useful. We would prefer a smoother field, so we should penalize the roughness of u, using

$$\mathcal{J}_1[u] = W_1 \iint_{\mathcal{D}} |\,\nabla u\,|^2\, d\mathbf{x} + w|\,\mathbf{u} - \mathbf{d}\,|^2. \tag{2.6.3}$$

Extrema of $\mathcal{J}_1$ satisfy

$$W_1\nabla^2\hat{u} = w\delta^{\mathrm{T}}(\hat{\mathbf{u}} - \mathbf{d}). \tag{2.6.4}$$

So the field of least gradient which nearly fits the data is a crop of logarithms: recall that $\nabla^2 \ln|\mathbf{x}| = -(2\pi)^{-1}\delta(x)\delta(y)$. What's more, the solution of (2.6.4) is undefined up to harmonic functions ($\nabla^2 v = 0$) such as bilinear functions, which may or may not be fixed by boundary conditions. Logarithmic singularities are most likely undesirable, so we are led to consider

$$\mathcal{J}_2[u] = \iint \left[W_0 u^2 + W_1|\nabla u|^2 + W_2 \left\{ \left(\frac{\partial^2 u}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 u}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 u}{\partial y^2}\right)^2 \right\} \right] d\mathbf{x},$$
$$+ w\,|\,\mathbf{u} - \mathbf{d}\,|^2, \tag{2.6.5}$$

which has extrema satisfying

$$W_2\nabla^4\hat{u} - W_1\nabla^2\hat{u} + W_0\hat{u} = -w\delta^{\mathrm{T}}(\hat{\mathbf{u}} - \mathbf{d}). \tag{2.6.6}$$

Solutions of (2.6.6) behave like $|\,\mathbf{x} - \mathbf{x}_m\,|^2 \ln|\,\mathbf{x} - \mathbf{x}_m\,|$ for $\mathbf{x}$ near $\mathbf{x}_m$. This is usually acceptable. Evidently, any desired degree of smoothness may be achieved by imposing a sufficiently severe penalty for roughness. Note that the homogeneous equation corresponding to (2.6.6), subject to suitable boundary conditions, has only the trivial solution: $\hat{u} \equiv 0$.

2.6.3 Equivalent covariances

Now consider v, the Fourier transform of u:

$$v(\mathbf{k}) = \iint u(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{x}, \tag{2.6.7a}$$

where the range of integration is the entire plane and $\mathbf{k} = (k, l)$. The inverse transform is

$$u(\mathbf{x}) = (2\pi)^{-2} \iint v(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{k}. \tag{2.6.7b}$$

The penalty functional (2.6.5) is equivalent to

$$\mathcal{J}_2[u] = (2\pi)^{-2} \iint (W_0 + W_1 |\,\mathbf{k}\,|^2 + W_2 |\,\mathbf{k}\,|^4) |\,v\,|^2\, d\mathbf{k} + w\,|\,\mathbf{u} - \mathbf{d}\,|^2, \tag{2.6.8}$$

provided we assume that the domain $\mathcal{D}$ is the entire plane. Let the inverse transform of the reciprocal of the roughness weight in (2.6.8) be

$$C(\mathbf{x}) = (2\pi)^{-2} \iint (W_0 + W_1|\mathbf{k}|^2 + W_2|\mathbf{k}|^4)^{-1} e^{-i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{k}. \tag{2.6.9}$$

After some calculus, it may be seen that (2.6.8) becomes

$$\mathcal{J}_2[u] = \iiiint u(\mathbf{x}) W(\mathbf{x} - \mathbf{x}') u(\mathbf{x}')\, d\mathbf{x}\, d\mathbf{x}' + w\,|\,\mathbf{u} - \mathbf{d}\,|^2, \tag{2.6.10}$$

where

$$\iint W(\mathbf{x} - \mathbf{x}') C(\mathbf{x}' - \mathbf{x}'')\, d\mathbf{x}' = \delta(\mathbf{x} - \mathbf{x}''). \tag{2.6.11}$$

Thus there is close relationship between roughness penalties as in (2.6.5), and "nondiagonal sums" such as in (2.6.10). The latter penalty is in turn related to statistical estimation of a field having zero mean, and covariance

$$\overline{u(\mathbf{x})u(\mathbf{x}')} = C(\mathbf{x} - \mathbf{x}'). \tag{2.6.12}$$

Exercise 2.6.1
Verify all the calculus sketched above, and show that C as defined in (2.6.9) only depends upon $|\mathbf{x}|$. That is, the random field u is *isotropic*. □

Exercise 2.6.2
If $\mathcal{D}$ is bounded, what boundary conditions must $\hat{u}$ satisfy, in order to be an extremum of (2.6.5)? □

Exercise 2.6.3 (Wahba and Wendelberger, 1980)
Express $\hat{u}$ in terms of representers. What is the associated inner product? □

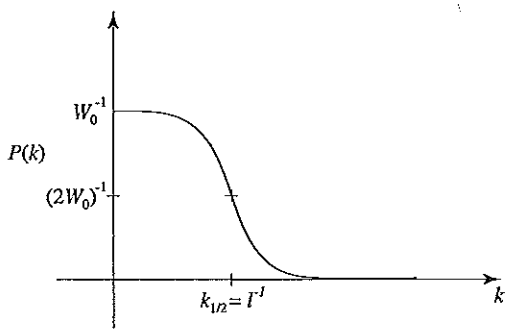

Figure 2.6.1 Power spectrum.

How should we choose W_0, W_1, and W_2? The inverse transform (2.6.9) yields, in particular, the hypothetical variance of $u(\mathbf{x})$:

$$\overline{u(\mathbf{x})^2} = C(0) = (2\pi)^{-2} \iint (W_0 + W_1 |\mathbf{k}|^2 + W_2 |\mathbf{k}|^4)^{-1} \, d\mathbf{k}, \tag{2.6.13}$$

hence W_0 may be chosen to set the variance once W_1/W_0 and W_2/W_0 have been chosen. For example, let us assume that

$$W_1/W_0 = 0, \quad W_2/W_0 = l^4 \tag{2.6.14}$$

for some length scale l. Then the hypothetical power spectrum of $u(\mathbf{x})$ is

$$P(k) = W_0^{-1}(1 + k^4 l^4)^{-1}, \tag{2.6.15}$$

where $k = |\mathbf{k}|$. Defining the half-power point $k_{\frac{1}{2}}$ by

$$P(k_{\frac{1}{2}})/P(0) = \frac{1}{2} \tag{2.6.16}$$

(see Fig. 2.6.1), we find that

$$k_{\frac{1}{2}} = l^{-1}. \tag{2.6.17}$$

The functional (2.6.5), with parameters obeying (2.6.14), penalizes scales shorter than l ($k \gg k_{\frac{1}{2}} = l^{-1}$) and fits the data more closely if $W_0 \ll w$.

Exercise 2.6.4
The "bell-shaped" covariance

$$C(\mathbf{x}) = \exp(-|\mathbf{x}|^2 l^{-2}) \tag{2.6.18}$$

is commonly used in optimal interpolation. Is there a corresponding smoothing norm, of the kind in (2.6.5)? □

In summary, there are at least two ways of implementing least-squares smoothing: with covariances or with smoothing norms. These can be precisely or imprecisely

matched, by choice of functional forms and parameters. It will be seen that the choice of implementation can be a matter of major convenience.

2.6.4 Embedding theorems

(The following two sections may be omitted from a first reading.) We began §2.6 with a discussion of quadratic penalty functionals used in the smoothing of data. It was seen that the smoothing field $\hat{u}(\mathbf{x})$ could have unacceptably singular behavior near the data points if the "smoothing norm" in the penalty functional were not chosen appropriately, that is, if the functional did not penalize derivatives of $u(\mathbf{x})$ of sufficiently high order. This was demonstrated by examining the solution of the Euler–Lagrange equation for $\hat{u}$, close to the data points. The examination was feasible since the functionals were quadratic in u and hence the Euler–Lagrange equations were linear, but we need not restrict ourselves in principle to quadratic functionals. There are powerful, theoretical guides that relate the mathematical smoothness of the estimate $\hat{u}$ to the differential order and algebraic power of the "smoothing norm" in the penalty functional. What follows is the crudest sketch of these so-called "embedding theorems" (Adams, 1975).

Let us suppose that the function $u = u(\mathbf{x})$ behaves algebraically near the point $\mathbf{x}_0$:

$$|u(\mathbf{x})| \sim Kr^{\alpha} \tag{2.6.19}$$

for small r, where $r = |\mathbf{x} - \mathbf{x}_0|$, K is a positive constant and α is a positive or negative constant. The point $\mathbf{x}$ is in n-dimensional space: $\mathbf{x} \in R^n$. We unrigorously infer that any m^{th}-order partial derivative of u is also algebraic near $\mathbf{x}_0$, with

$$\left|D^{(m)}u(\mathbf{x})\right| \sim K'r^{\alpha-m}, \tag{2.6.20}$$

where K' is another positive constant. Hence if we raise $D^{(m)}u$ to the power p and integrate over a bounded domain $\mathcal{D}$ that includes $\mathbf{x}_0$, then

$$\int \ldots \int_{\mathcal{D}} \left|D^{(m)}u(\mathbf{x})\right|^p d\mathbf{x} \sim K'' \int_0^R r^{(\alpha-m)p+n-1}\, dr, \tag{2.6.21}$$

where R is the radius of $\mathcal{D}$. The integral on the rhs of (2.6.21) is finite, provided

$$(\alpha - m)p + n - 1 > -1. \tag{2.6.22}$$

That is, if the integral on the lhs of (2.6.21) is finite, then

$$|u(\mathbf{x})| \sim Kr^{\alpha} < Kr^{m-n/p} \tag{2.6.23}$$

for small r. Provided $mp > n > (m-1)\,p$, a rigorous treatment (Adams, 1975, p. 98) would replace the conclusion (2.6.23) with the more conservative inequality

$$|\,u(\mathbf{x}) - u(\mathbf{x}_0)\,| < K'''\,|\mathbf{x} - \mathbf{x}_0\,|^{\lambda}, \tag{2.6.24}$$

where

$$0 < \lambda \leq m - n/p. \tag{2.6.25}$$

If we were to include a term like the lhs of (2.6.21) in our smoothing norm, and were to find the $\hat{u}$ that minimizes the penalty functional, then we could conclude that the lhs of (2.6.21) would be finite, and hence (2.6.24) must hold. The positivity of λ in (2.6.25) ensures that u is at least continuous at $\mathbf{x}_0$. If λ exceeds unity, then we can be sure that u is differentiable at $\mathbf{x}_0$, and so on: $\lambda > k$ implies $D^{(k)}u$ is continuous at $\mathbf{x}_0$.

For example, suppose $n = 2$ (we are in the plane: $\mathbf{x} = (x, y)$); suppose $m = 2$ (we include second derivatives) and suppose $p = 2$ (we have a quadratic smoothing norm as in (2.6.5)); then

$$mp = 4 > n = 2 \not> (m-1)p = 2, \tag{2.6.26}$$

and so we cannot even be sure that u is continuous. Nevertheless, we learned from the Euler–Lagrange equation (2.6.6) that $u \sim Kr^2 \ln r$, which is actually differentiable. Thus, the "embedding theorem" estimate of smoothness given in (2.6.25) is very conservative. The theorem would have us choose $p = 1.9$,

$$mp = 3.8 > n = 2 > (m-1)p = 1.9. \tag{2.6.27}$$

Such a fractional power would make the calculus of variations very awkward, but the penalty functional would be well defined and would have a minimum $\hat{u}$ with guaranteed continuity.

2.6.5 Combining hypotheses: harmonic means of covariances

We have been considering penalty functionals, schematically of the form

$$\mathcal{J}[u] = (Mu) \circ C_f^{-1} \circ (Mu) + \cdots, \tag{2.6.28}$$

where C_f is the hypothesized covariance of Mu, M being some linear differential operator or linear model operator in general. We might also hypothesize that B_u is the covariance of u, in which case we could form the penalty functional

$$\mathcal{J}[u] = (Mu) \circ C_f^{-1} \circ (Mu) + u \circ B_f^{-1} \circ u + \cdots. \tag{2.6.29}$$

What now is the effectively hypothesized covariance for u? Manipulations like integrations by parts yield

$$(Mu) \circ C_f^{-1} \circ (Mu) = u \circ MC_f^{-1}M \circ u \tag{2.6.30}$$

$$= u \circ C_u^{-1} \circ u, \tag{2.6.31}$$

where

$$C_u = M^{-1}C_f M^{-1}. \tag{2.6.32}$$

Think of C_u as the covariance of solutions of the model $Mu = f$, where f has covariance C_f. We can now identify the effectively hypothesized covariance:

$$\mathcal{J}[u] = u \circ C_u^{-1} \circ u + u \circ B_u^{-1} \circ u + \cdots \tag{2.6.33}$$

$$= u \circ A_u^{-1} \circ u, \tag{2.6.34}$$

where

$$A_u = \left(C_u^{-1} + B_u^{-1}\right)^{-1} \tag{2.6.35}$$

is the harmonic mean of the two covariances C_u and B_u.