Noname manuscript No.
(will be inserted by the editor)

Local stability and estimation of uncertainty for solutions to inverse problems

A. Yaremchuk[1], M. Yaremchuk[2], J. Schröter[3] and M. Losch[3]

[1] Andreyev Acoustics Institute, Moscow, Russia
[2] International Pacific Research Center, University of Hawaii, USA
[3] Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven, Germany

October 12, 2001

Abstract We present a method for studying local stability of a solution to an inverse problem and evaluate the uncertainty in determining true values of particular observables. The investigation is done under the assumption that only the Gaussian part of fluctuations about the local minimum of the cost (likelihood) function is essential. Our approach is based on the spectral analysis of the Hessian operator associated with the cost function at its extremal point and we put forward an effective iterative algorithm suitable for numerical implementation in case of a computationally large problem.

1 Introduction

The past decade was marked by a successful application of inverse techniques to the investigation of oceanic dynamics. However, the key point in formulating an inverse problem, namely specification of the cost (likelihood) function, remains more an art than a well-defined engineering procedure. Moreover, tuning the parameters and weights defining the cost function is very often done on the grounds of visual inspection of the output pictures and only deviations of the model-predicted fields from observations are used as a quantitative criterion. Also there is no tool which can help in understanding how much the output depends on the initial guess, neither do we have a reliable tool which can be used for estimating (in the framework of of the employed inverse problem) the accuracy of the obtained result.

Since in general actual geophysical inverse problems are computationally large and poorly conditioned, we also need a tool for the quantitative study of their stability and the efficiency of regularization. Theoretical principles for such an investigation are well known. Applications to geosciences have

been described, e.g., by Thacker (1989). This work is aimed at constructing a numerical scheme which can be used in practical applications.

Solutions to oceanographic inverse problems normally have the meaning of a maximum likelihood estimate. The logic of the maximum likelihood estimate implies that fluctuations of the solution about the most probable state must not be large, otherwise the estimate becomes useless. Also one should keep in mind that even when fluctuations about the most probable value are small and, in particular, there are no other local maxima of the probability distribution nearby, spontaneous transitions to far-distant locally-optimum solutions may occur and ruin the validity of the maximum likelihood estimate. We shall not consider these highly non-linear phenomena, but shall assume that the given solution makes sense. Therefore, it is natural to linearize the problem in the vicinity of the most probable point and treat the fluctuations as Gaussian noise (Thacker, 1989). The nature of linear problems and corresponding analytic formulas are simple, so our main concern is numerical implementation. Two particular examples will be given for illustrative purposes, one in a general form and the other will be related to the data assimilation problems arising in oceanography.

The outline of the paper is as follows. In section 2 we review basic principles for solving inverse problems via variational techniques emphasizing their probabilistic nature, discuss the maximum likelihood solution and analyze its uncertainty in the Gaussian approximation. In section 3 we describe an iterative algorithm suitable for numerical implementation. For convenience of reference we give a short summary in the appendix. In section 4 numerical examples are presented. Finally, general conclusions will be given in section 5.

2 Variational formulation

We consider regularized inverse problems of the following form: given a D-dimensional space X of possible solutions x, Riemannian metric $\langle dx|g|dx\rangle$ [we use standard angle-bracket notation (Dirac, 1981; Landau and Lifshitz, 1958a) to denote coupling between vectors and covectors] on it, which measures the quality of reconstruction (say, geodetic distance between the true solution and the reconstructed one or the amplitude of random noise contaminating the reconstructed solution), and the probability distribution

$$\mathcal{D}P(x) = Z^{-1}\, e^{-\mathcal{H}(x)}\, \mathcal{D}\mu(x), \qquad \int\limits_{x\in X} \mathcal{D}P(x) = 1, \tag{1}$$

where $\mathcal{H}$ is a scalar function, $\mathcal{D}\mu$ denotes the measure associated with metric g and Z is a normalization constant, find the most probable solution x_* such that

$$\mathcal{H}(x_*) = \min_{x\in X} \mathcal{H}(x). \tag{2}$$

In terms of statistical physics X has the meaning of the phase space, $\mathcal{D}\mu$ of canonical distribution, while probability $\mathcal{D}P$ is written in Gibbs' form (Landau and Lifshitz, 1958b). On the other hand, when formulated in the language of the optimal control theory, x is normally referred to as a control variable and $\mathcal{H}$ as a cost or likelihood function (e.g., Luong et al., 1998; Thacker, 1989).

When $\mathcal{H}$ has well-determined minima one can expect that the saddle-point approximation to (1) works well. In statistical physics it is known as the mean-field theory, in statistics as the maximum likelihood estimating (e.g., Nagelkerke, 1992), and within its framework we can consider X in the vicinity of the optimum point to be an affine space and g a constant matrix. Expanding $\mathcal{H}$ into a Taylor series,

$$\mathcal{H} = \mathcal{H}(x_*) + \frac{1}{2}\langle x - x_* | h | x - x_* \rangle + \cdots, \tag{3}$$

with h standing for a symmetric matrix of second derivatives of $\mathcal{H}$ with respect to control variables at stationary point x_*, and substituting (3) into (1), we find that the deviation $x - x_*$ of the control variable from the optimal one appears to be a Gaussian stochastic vector with zero mean and covariance matrix h^{-1},

$$\mathrm{Mean}_x \left[x - x_* \right] = 0, \quad \mathrm{Mean}_x \left[(x - x_*) \otimes (x - x_*) \right] = h^{-1}, \tag{4}$$

where $\mathrm{Mean}_x[\ldots]$ denotes mean value with respect to distribution (1).

Let Φ_α, $\alpha = 1, 2 \ldots$ be observables, i.e., some scalar functions of control variable x. Expanding Φ_α in powers of deviation $x - x_*$,

$$\Phi_\alpha = \Phi_\alpha(x_*) + \langle \phi_\alpha | x - x_* \rangle + \cdots, \quad \phi_\alpha = d\Phi(x_*),$$

we see that Φ_α are also Gaussian stochastic variables with the expected value equal to $\Phi_\alpha(x_*)$ and covariances

$$C_{\alpha\beta} \stackrel{\mathrm{def}}{=} \mathrm{Mean}_x \left\{ \left[\Phi_\alpha - \Phi_\alpha(x_*) \right] \left[\Phi_\beta - \Phi_\beta(x_*) \right] \right\} = \langle \phi_\alpha | \, h^{-1} | \phi_\beta \rangle. \tag{5}$$

Therefore, to leading order of the saddle-point approximation correlation functions of observables can be expressed as multi-linear combinations of matrix elements of the form (5). In particular, the mean squared deviation δ^2 of the true solution x from the maximum likelihood estimate x_* is given by

$$\delta^2 \stackrel{\mathrm{def}}{=} \mathrm{Mean}_x \left\{ \langle x - x_* | \, g \, | x - x_* \rangle \right\} = \mathrm{Tr} \left\{ h^{-1} g \right\}. \tag{6}$$

The value of δ also characterizes stability of the solution: if instead of a deep well centered at x_* the "landscape" of $\mathcal{H}$ looks like a valley, position of the deepest point becomes unstable and may be shifted by this distance along the bottom of the valley. At this point some clarification should be done. In our approximation the cost function is assumed quadratic and positive, and since all positive quadratic forms are equivalent to each other (by means of an appropriate linear transformation we always can turn h into a unity

matrix), all directions in the phase space also seem equivalent. However, one should take into account that we have a metric g for calculating the noise amplitude and, therefore, while reshaping the cost function care should be taken so that g is preserved.

We introduce the Hessian operator $H = g^{-1}h$ as the ratio of two quadratic forms and rewrite (5) and (6) as

$$C_{\alpha\beta} = \langle \phi_\alpha | H^{-1} g^{-1} | \phi_\beta \rangle, \quad \delta^2 = \mathrm{Tr}\left\{ H^{-1} \right\}. \tag{7}$$

Operator H is self-adjoint and positive with respect to Euclidean structure generated by quadratic form g and its spectral decomposition gives a full set of invariants for the pair g and h of quadratic forms. We denote with $\varepsilon_1, \ldots, \varepsilon_D$ its eigenvalues and corresponding eigenvectors with $\psi_1, \ldots, \psi_D$:

$$H\,\psi_k = \varepsilon_k\,\psi_k, \quad \langle \psi_i |\, g\, | \psi_k \rangle = \delta_{ik}, \quad i,k = 1,\ldots,D. \tag{8}$$

With this notation the covariance matrix of solution fluctuations takes the form

$$h^{-1} = \sum_k \frac{1}{\varepsilon_k}\, \psi_k \otimes \psi_k, \tag{9}$$

while (7) becomes

$$C_{\alpha\beta} = \sum_k \langle \phi_\alpha | \psi_k \rangle \frac{1}{\varepsilon_k} \langle \phi_\beta | \psi_k \rangle, \quad \delta^2 = \sum_k \frac{1}{\varepsilon_k}. \tag{10}$$

From (9) and (10) we see that along directions ψ_k corresponding to small eigenvalues $\varepsilon_k \to 0$ the profile of the cost function $\mathcal{H}$ is flat, position of x_* is unstable, and their contribution to the fluctuations amplitude δ is dominant. Also fluctuations of a particular observable, Φ_α, depend on whether its gradient is perpendicular to these eigenvectors or not. Even when the optimum point is unstable, certain quantities might be well observed if they are invariant with respect to shifts in unstable directions.

For a quantitative description we introduce the Källén-Lehmann spectral functions (see, e.g., Itzykson and Zuber, 1990) $F_H(\varepsilon)$ and $F_\alpha(\varepsilon)$ as follows:

$$dF_H(\varepsilon) \stackrel{\mathrm{def}}{=} \sum_k \delta(\varepsilon - \varepsilon_k)\, d\varepsilon, \quad F_H(0) = 0, \tag{11}$$

$$dF_\alpha(\varepsilon) \stackrel{\mathrm{def}}{=} \sum_k \delta(\varepsilon - \varepsilon_k) \left| \langle \phi_\alpha | \psi_k \rangle \right|^2 d\varepsilon, \quad F_\alpha(0) = 0, \quad \alpha = 1,2,\ldots \tag{12}$$

Both of them are monotonically increasing and exhibit jumps exactly at points coinciding with eigenvalues of the Hessian operator. Jumps of the first one at spectral points are equal to dimensions of corresponding invariant subspaces, jumps of the other are equal to squared amplitudes of

decomposition of ϕ_α into a superposition of eigenvectors. Each of these functions accumulates much more information than the corresponding entry into (7):

$$C_{\alpha\alpha} = \int\limits_0^{+\infty} \frac{1}{\varepsilon}\, dF_\alpha(\varepsilon), \qquad \delta^2 = \int\limits_0^{+\infty} \frac{1}{\varepsilon}\, dF_H(\varepsilon). \tag{13}$$

Also, convergence rate of many of iterative solvers in the vicinity of the optimal point x_* may be expressed through them. Thus, we consider the stability problem to be completely examined if we find a way for computing (7) and (11)–(12).

3 Computation

At present in case of a computationally large inverse problem a search for the optimum point x_* is done with the help of algorithms which perform a descend from the starting point to the nearest local minimum computing gradient of the cost function at each step. In contrast to the differential, which is completely determined by the cost function itself, the gradient also depends on the metric in the control space (Schwartz, 1967): $\nabla\mathcal{H} = g^{-1}d\mathcal{H}$. In view of (3) we have $\nabla\mathcal{H} = H|x - x_*\rangle$, therefore, $H|\psi\rangle$ for any vector ψ is available. We assume for the following that the product of the Hessian operator and a vector can always be computed. Also in practice g is either a diagonal matrix or differs from such a matrix by an operator of finite rank, so we also assume that from the computational point of view g may be treated as if it were diagonal.

The first formula in (7) suggests a simple way (Yaremchuk et al., 1998) for evaluation of the covariance matrix: solve equation $H\psi_\beta = g^{-1}\phi_\beta$ for ψ_β for all β and get $C_{\alpha\beta} = \langle\phi_\alpha|\psi_\beta\rangle$. The value of δ^2 may be obtained in a similar manner if we perform an additional averaging over an ensemble of random observables. Indeed, if ϕ is a Gaussian random vector with covariance matrix equal to g, then

$$\delta^2 = \mathrm{Tr}\left\{ \mathrm{Mean}_\phi \left[\phi \otimes \phi \right] \, H^{-1}g^{-1} \right\} = \mathrm{Mean}_\phi \left[\langle\phi|\, H^{-1}g^{-1}|\phi\rangle \right], \tag{14}$$

where $\mathrm{Mean}_\phi[\ldots]$ denotes averaging over ϕ. In practice we can only use a finite ensemble of independent realizations of a stochastic variable and, therefore, our estimate of the average value of the matrix element on the right-hand side of (14) will be approximate. The corresponding error may be expressed in terms of χ-distribution. In particular, employing an ensemble of five realizations, with probability of 90% we estimate the contribution from any eigenmode with accuracy not worse than five decibel. According to our experience it is enough to use three or even two realizations.

For computation of $H^{-1}g^{-1}|\phi\rangle$ any suitable iterative solver can be used. However, when (and in practice this is always the case) the Hessian operator is poorly conditioned, the result of such a computation does not make

much sense since it will crucially depend on routine's stopping criterion. It seems more meaningful to get an estimate of the spectral functions (11)–(12), examine their behavior, and choose the stopping criterion on these grounds in a favourable case or issue an "undetermined" verdict otherwise. For evaluation of (11)–(12) there are no library routines and we employ the method proposed by Yaremchuk and Schröter (1998). Given any function f of a complex variable regular at all points of the Hessian spectrum, we can apply it to the Hessian operator itself (e.g., Rudin, 1991) obtaining the following expression:

$$f(H) = \sum_k f(\varepsilon_k)\, |\psi_k\rangle\langle\psi_k| g. \tag{15}$$

Obviously this formula does not only work in the case of analytic functions, but distributions as well. Comparing it to (11)–(12) we see that

$$dF_H(\varepsilon) = \mathrm{Tr}\left\{\delta(\varepsilon - H)\right\} d\varepsilon = \mathrm{Mean}_\phi \left\{ \langle\phi|\, \delta(\varepsilon - H)\, g^{-1}|\phi\rangle \right\} d\varepsilon, \tag{16}$$

$$dF_\alpha(\varepsilon) = \langle\phi_\alpha|\, \delta(\varepsilon - H)\, g^{-1}|\phi_\alpha\rangle\, d\varepsilon. \tag{17}$$

In practice in case of a high-dimensional problem it is not possible to evaluate a function of an operator explicitly, because we do not know eigenvalues and eigenvectors beforehand. We only can evaluate a polynomial by successively computing $H|\psi\rangle$, $H^2|\psi\rangle, \ldots$ for any given vector $|\psi\rangle$. Thus, we may approximate distributions on the right-hand side of (16) and (17) by polynomials and evaluate them iteratively.

The most straightforward way to obtain a polynomial approximation to the delta distribution $\delta(\varepsilon - \varepsilon')$ is to use the orthogonal polynomial technique. Let $\left\{P_n(\varepsilon)\,|\, n = 0, 1, \ldots\right\}$ be a complete set of polynomials orthogonal with respect to $\rho(\varepsilon)\, d\varepsilon$, with $\rho(\varepsilon)$ being a positive weight function. Then

$$\delta(\varepsilon - H)\, g^{-1}|\phi\rangle = \rho(\varepsilon) \sum_{n=0}^{\infty} \frac{1}{h_n}\, P_n(\varepsilon)\, P_n(H)\, g^{-1}|\phi\rangle, \tag{18}$$

where $h_n = \int \left|P_n(\varepsilon)\right|^2 \rho(\varepsilon)\, d\varepsilon$. Here the main computational labor is required for evaluation of vectors $P_n(H)|\, g^{-1}\phi\rangle$, while subsequent summation is cheap. It makes sense to use the sequence $P_n(H)|\, g^{-1}\phi\rangle$ also for evaluation of $C_{\alpha\beta}$ and δ^2, say, by multiplying (18) by $\varepsilon^{-1} d\varepsilon \langle\phi|$ from the left and integrating (13) over the spectrum or, equivalently, by expanding H^{-1} in an infinite series of polynomials $P_n(H)$ and employing (7). However, it is more practical to compute $H^{-1/2}|\, g^{-1}\phi\rangle$ and use

$$C_{\alpha\beta} = \langle H^{-1/2} g^{-1}\phi_\alpha|\, g\, |H^{-1/2} g^{-1}\phi_\beta\rangle, \tag{19}$$

$$\delta^2 = \mathrm{Mean}_\phi \left[\langle H^{-1/2} g^{-1}\phi|\, g\, |H^{-1/2} g^{-1}\phi\rangle \right], \tag{20}$$

since $H^{-1/2}$ is less singular than H^{-1} and can be approximated more accurately.

In numerical applications the Hessian operator is always bounded, and without loss of generality we assume that its spectrum is contained in the subinterval $(0, 1)$ of the real axis. [Estimation of the maximum eigenvalue is relatively cheap and can be done, say, with the power method (Mathews, 1992)]. In the current investigation we use shifted Chebyshev polynomials of the second kind (Bateman, 1953), $\mathrm{U}_n(1-2\varepsilon)$, which correspond to

$$\rho(\varepsilon) = 4\sqrt{\varepsilon(1-\varepsilon)}, \quad h_n = \pi/2, \quad n = 0, 1, \ldots$$

and result in

$$H^{-1/2} g^{-1} |\phi\rangle = \frac{16}{\pi} \sum_{n=1}^{\infty} \frac{n}{4n^2 - 1} \, \mathrm{U}_{n-1}(1-2H) \, g^{-1} |\phi\rangle. \tag{21}$$

Vectors $|\mathrm{U}_n\rangle \stackrel{\mathrm{def}}{=} \mathrm{U}_n(1-2H)|\phi\rangle$ may be computed recursively:

$$\begin{gathered} |\mathrm{U}_0\rangle = g^{-1}|\phi\rangle, \qquad |\mathrm{U}_1\rangle = 2\,|\mathrm{U}_0\rangle - 4H\,|\mathrm{U}_0\rangle, \\ |\mathrm{U}_{n+1}\rangle = 2\,|\mathrm{U}_n\rangle - 4H\,|\mathrm{U}_n\rangle - |\mathrm{U}_{n-1}\rangle, \quad n = 1, 2, \ldots \end{gathered} \tag{22}$$

Spectral functions (11) and (12) may be represented in the form of a trigonometric series if we lift them to a unit circle via substitution $\varepsilon = \sin^2\theta/2$, $0 < \theta < \pi$:

$$F_H\left[\sin^2\!\big(\theta/2\big)\right] = \frac{1}{\pi} \sum_{n=0}^{\infty} \frac{\sin n\theta}{n} \, \mathrm{Mean}_{\phi}\left[\langle\phi|\mathrm{T}_n\rangle\right], \tag{23}$$

$$F_{\alpha}\left[\sin^2\!\big(\theta/2\big)\right] = \frac{1}{\pi} \sum_{n=0}^{\infty} \frac{\sin n\theta}{n} \, \langle\phi_{\alpha}|\mathrm{T}_n\rangle. \tag{24}$$

New vectors $|\mathrm{T}_n\rangle$ are formed from $|\mathrm{U}_n\rangle$ according to

$$|\mathrm{T}_n\rangle = \begin{cases} |\mathrm{U}_n\rangle, & n = 0, 1, \\ |\mathrm{U}_n\rangle - |\mathrm{U}_{n-2}\rangle, & n = 2, 3, \ldots \end{cases} \tag{25}$$

and are related to shifted Chebyshev polynomials of the first kind (Bateman, 1953) as follows:

$$|\mathrm{T}_n\rangle = \begin{cases} \mathrm{T}_0(1-2H)\, g^{-1}|\phi\rangle, & n = 0, \\ 2\mathrm{T}_n(1-2H)\, g^{-1}|\phi\rangle, & n = 1, 2, \ldots \end{cases} \tag{26}$$

Certainly in case of spectral functions F_{α} covectors ϕ_{α} should be substituted for ϕ in (22) and (26).

For numerical evaluation we have to truncate the infinite series (21) and (23)–(24). This procedure may be interpreted as multiplication of the expansion coefficients by the factors

$$w_n = \begin{cases} 1, & 0 \le n \le N-1, \\ 0, & N \le n, \end{cases} \tag{27}$$

or, more generally, as smoothing functions represented by the original series:

$$f(H) \mapsto \int_0^1 W(H,\varepsilon')\, f(\varepsilon')\, d\varepsilon', \quad W(\varepsilon,\varepsilon') = \sum_{n=0}^{\infty} \frac{w_n}{h_n} P_n(\varepsilon) P_n(\varepsilon')\, \rho(\varepsilon').$$

Here $f(H)$ stands for $\delta(\varepsilon - H)$ or $H^{-1/2}$ and the kernel $W(\varepsilon,\varepsilon')$ is a smoothed delta distribution determined by coefficients w_n. When lifted to the unit circle of Fourier frequencies θ, smoothing turns into convolution with the smoothing kernel. In signal processing smoothing operators are referred to as windows. It is well known that the Dirichlet window, given by (27), leads to the Gibbs effect and it is better to use a different one. For smoothing spectral functions (11) and (12) one should use a window that is represented by a strictly positive kernel $W(\varepsilon,\varepsilon')$ and maps monotonic functions into monotonic. Among such windows are the Cezàro kernel,

$$w_n = \begin{cases} 1 - n/N, & 0 \le n \le N-1, \\ 0, & N \le n, \end{cases} \tag{28}$$

and the Vallée-Poussin kernel (Hardy, 1949).

To construct an optimum window for computing $H^{-1/2}$ let us suppose that $C^{(\mathrm{est})}_{\alpha\beta}$ is an estimate of the covariance matrix. The relative error of an estimate may be defined as

$$\mathrm{error} \stackrel{\mathrm{def}}{=} \frac{\left| C_{\alpha\beta} - C^{(\mathrm{est})}_{\alpha\beta} \right|}{\sqrt{C_{\alpha\alpha}\, C_{\beta\beta}}}, \tag{29}$$

and our goal is to minimize this error choosing the best polynomial approximation $P(\varepsilon)$ to $\varepsilon^{-1/2}$. An optimal approximation to function ε^{-1} is well known (see, e.g., Axelsson and Barker, 1984, for the theory of the conjugate gradients method) and provides accuracy

$$\mathrm{error} \le \frac{1}{\cosh(N\theta_*)}, \qquad \cosh(\theta_*) \stackrel{\mathrm{def}}{=} \frac{1 + E_{\min}}{1 - E_{\min}}, \tag{30}$$

where $E_{\min} > 0$ bounds the spectrum of the Hessian operator from below. Accuracy in computing $\varepsilon^{-1/2}$ appears to be higher and the optimum approximating polynomial may be constructed as follows.

With the aid of Hölder's inequality it is easy to obtain a bound for the relative error in the following form:

$$\mathrm{error} \le \max_{E_{\min} < \varepsilon < 1} \left| 1 - \varepsilon P^2(\varepsilon) \right|. \tag{31}$$

The actual accuracy is normally much better than that given by (31), but the right-hand side of (31) is a guaranteed one and we shall search for the polynomial of degree $N-1$ which minimizes it.

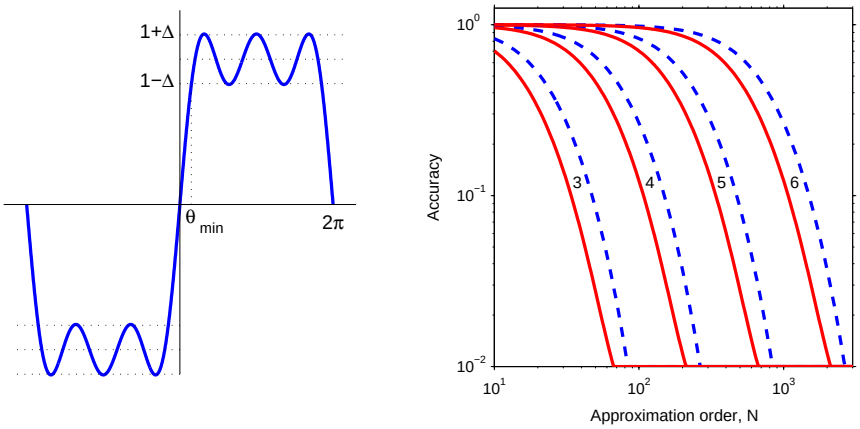

Fig. 1 The optimum approximation to Hilbert's transformer when $N = 3$ [$E_{\min} = \sin^2 \theta_{\min}/2$] (left) and relative accuracy as a function of N (right). Accuracy for the conjugate gradients inversion is shown by dashed curves, and for the $H^{-1/2}$ method by solid curves. The kth curve corresponds to $E_{\min} = 10^{-k}$.

Suppose that we have succeeded in finding an odd polynomial $F(s)$ of degree $2N-1$ such, that its maximum deviation Δ on the interval $\sqrt{E_{\min}} < |s| < 1$ from the function sign(s) is minimal. Then

$$P(\varepsilon) = \frac{1}{\sqrt{1+\Delta^2}} \frac{F\big(\sqrt{\varepsilon}\big)}{\sqrt{\varepsilon}}, \qquad \max_{E_{\min}<\varepsilon<1} \Big|1 - \varepsilon P^2(\varepsilon)\Big| \le \frac{2\Delta}{1+\Delta^2}.$$

Substituting $\sin(\theta/2)$ for $s = \sqrt{\varepsilon}$, we see that function $\mathcal{F}(\theta) = F\big[\sin(\theta/2)\big]$ may be interpreted as a finite-duration impulse response equiripple filter approximating Hilbert's transformer (e.g., Oppenheim and Schafer, 1989; Parks and Burrus, 1987, and Figure 1).

Given an approximate Hilbert transformer $\mathcal{F}(\theta)$ in the form of a Fourier series,

$$\mathcal{F}(\theta) = \sum_{n=0}^{\infty} c_n \sin\big[(n+1/2)\theta\big], \tag{32}$$

we can obtain $P(\varepsilon)$ as

$$P(\varepsilon) = \sum_{n=0}^{\infty} a_n \, \mathrm{U}_n(1-2\varepsilon), \quad a_n = \frac{c_n + c_{n+1}}{\sqrt{1+\Delta^2}}, \quad n = 0, 1, \ldots \tag{33}$$

If only a finite number of c_n are non-zero, $P(\varepsilon)$ is a polynomial, and series (21) should be replaced by

$$H^{-1/2} g^{-1} |\phi\rangle = \sum_{n=0}^{N-1} a_n \, |\mathrm{U}_n\rangle. \tag{34}$$

In signal processing the problem of digital filter design is well developed. We used the Parks–McClellan algorithm (IEEE, 1979) for computing the Fourier coefficients c_n. The accuracy (31) of the estimate depends on N and $E_{\min}$ and is better than that predicted by formula (30) for direct inversion of the Hessian operator via the conjugate gradients method (Figure 1).

However, it is impractical to attempt to estimate the minimal eigenvalue $E_{\min}$ beforehand. Instead it is natural to choose the necessary accuracy and the number of iterations we are ready to perform. These two numbers determine an approximation to Hilbert's transformer in accordance with Figure 1. The corresponding coefficients may be obtained via an iterative algorithm similar to Parks–McClellan's.

4 Numerical examples

To see the method in action we first demonstrate its performance in the case of a toy model which can be solved analytically and then apply it to estimating uncertainty in determining heat and mass fluxes across a hydrographic section in the North Atlantic Ocean obtained in the framework of a non-linear section inverse model (Nechaev and Yaremchuk, 1995).

Our toy problem is just a linear reconstruction of a 1D scalar field $u(x)$ on an interval $x \in (0,1)$ from direct observations $u_{\mathrm{data}}(x)$. We employ the likelihood function of the form

$$\mathcal{H} = \frac{\kappa^2}{2}\int_0^1 \big(\nabla u\big)^2 dx + \frac{m^2}{2}\int_0^1 \big(u - u_{\mathrm{data}}\big)^2 dx,$$

where κ is a regularization constant and m^{-1} is the amplitude of noise contaminating data. This choice results in the Hessian operator $H = m^2 - \kappa^2 \Delta$ with Δ standing for Laplacian with Neumann boundary conditions. The quality of reconstruction shall be determined by the L_2-norm.

For numerical implementation we specify a function $u(x)$ by its values u_s at the nodes $x_s = (s-1)\,\Delta x$, $\Delta x = 1/(D-1)$, $s = 1, \ldots, D$, and define the Laplacian with a finite-difference rule

$$\big(\Delta u\big)_s = \frac{1}{(\Delta x)^2}\left(u_{s+1} - 2u_s + u_{s-1}\right), \quad s = 1, \ldots, D,$$

assuming that $u_0 = u_2$ and $u_{D+1} = u_{D-1}$ (mirror reflection with respect to the boundaries). Quadratic form g is defined under the assumption that we interpolate with constants around each node; this leads to $g = \Delta x\,\mathrm{diag}\big(1/2, 1, \ldots, 1, 1/2\big)$ because boundary points only contribute to one half of the grid interval. Eigenvalues may be represented in the form

$$\varepsilon_k = \frac{m^2}{2E_{\min}}\left\{(1+E_{\min}) - (1-E_{\min})\,\cos\left[\pi\frac{k-1}{D-1}\right]\right\}, \quad k = 1, \ldots, D$$

where $E_{\min}^{-1} = 1 + 4\kappa^2 m^{-2}(\Delta x)^{-2}$ is the condition number of the system. For the following we choose noise amplitude to be unity ($m = 1$) and characterize the system by two parameters, D and $E_{\min}$.

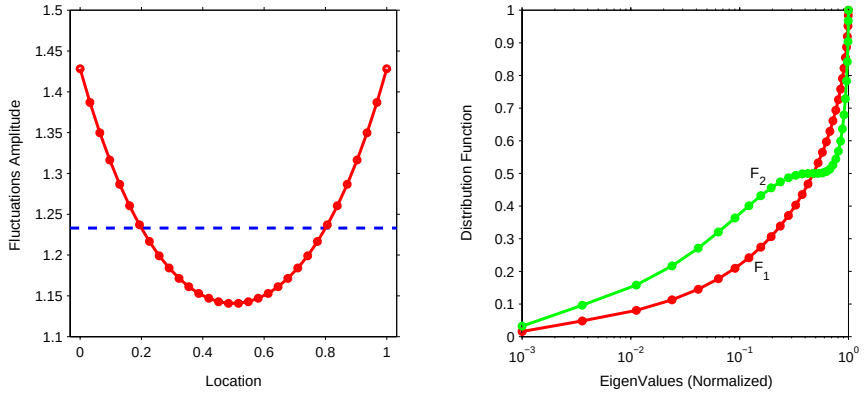

Fig. 2 Uncertainty in reconstruction of the observed field (left) and spectral distributions F_1 and F_2 for the first two observables u_1 and u_2, respectively (right). Deviation δ is shown by dashed line; parameters of the toy model are $D = 32$, $E_{\min} = 10^{-3}$.

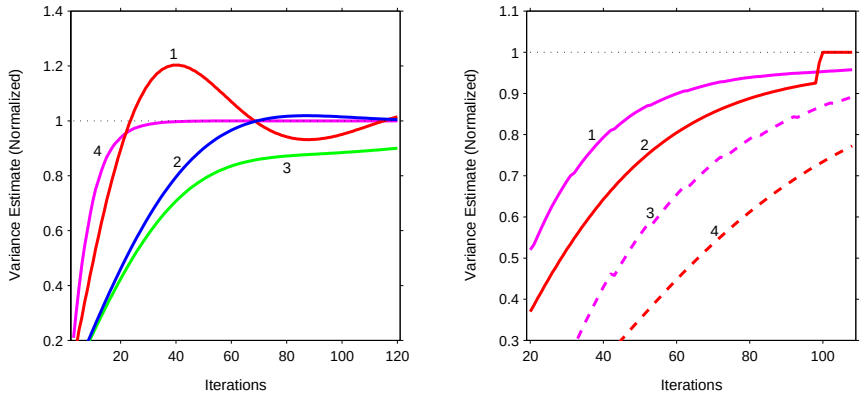

Fig. 3 Convergence of the $H^{-1/2}$ method with the Dirichlet (1), Hanning (2), Cezáro (3), and optimum (4) windows for the toy model at $D = 100$, $E_{\min} = 10^{-3}$ (left) and convergence of the $H^{-1/2}$ (1) and conjugate gradients (2) methods for the toy model at $D = 100$, $E_{\min} = 10^{-4}$ (right). Error bars for the $H^{-1/2}$ (3) and conjugate gradients (4) methods are given.

In Figure 2 we show the uncertainty in reconstructing field u at each grid point and spectral distributions F_1 and F_2 for the first two observables $\Phi_1[u] = u_1$ and $\Phi_2[u] = u_2$ which fluctuate most strongly. Note that while plotting spectral functions we made linear interpolation in between the spectral points instead of drawing true jumps — this renders the curves more readable and is indistinguishable from a step function when $D \to \infty$.

In Figure 3 we demonstrate the impact of different windowing functions on convergence rate of series (24) for the case of the most poorly deter-

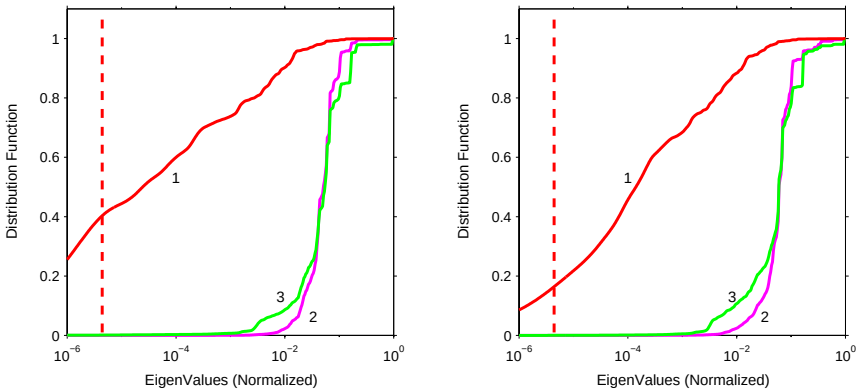

Fig. 4 Spectral distribution of the Hessian operator (1) and the observables that map independent variables into the integrated mass (2) and heat (3) transports when only smoothness as regularization is employed (left), and the same with an additional background regularization (right). The dashed line points at the spectral value determined by the uncertainty principle (due to truncation of infinite series representing delta-distribution) and shows the resolution in the spectral range. Note that in both cases distribution of the Hessian operator eigenvalues is not resolved completely.

mined observable Φ_1 and compare it with the convergence of the conjugate gradients method used for straightforward evaluation according to (7).

As a realistic example we present a non-linear analysis of hydrographic data. The model inverts temperature and salinity measurements from surface to bottom along the cruise track of a research vessel to obtain the flow field and thereby the mass and heat transport through the vertical plane beneath the surface track. Details of the model may be found in Nechaev and Yaremchuk (1995). The data set was produced artificially by integrating the 1/3° North Atlantic Model of the FLAME Group (Redler et al., 1998). The number of independent variables is of the order of 10^4 thus making a direct inversion of the Hessian matrix for calculating the uncertainties of integrated mass and heat transports impractical.

The quality of the inverse solution is evaluated by an Euclidean norm where the fluctuations of each independent physical variable are weighted by the inverse of an estimate of its horizontal variance. Since the control parameters are the independent physical variables normalized by the square root of their horizontal variance, the quadratic form g is represented by a unity matrix.

The model is regularized by imposing spatial smoothness on the modeled fields. In addition the deviation of the independent variables from a prior guess, the so-called background, can be penalized. From Figure 4 it becomes obvious that the gradients $d\Phi_\alpha$ of integrated transports are with a good accuracy orthogonal to the eigenvectors of the Hessian operator that correspond to small eigenvalues, thus making the estimation of the trans-

port uncertainties possible. The additional regularization with a background term shifts the "infrared" part of the Hessian spectrum not seen in the left frame of Figure 4 to the resolved part, but leaves the spectral distribution of the transport observables almost unchanged. This means that although we can not completely reconstruct the model state with the aid of the specified inverse problem and allowed CPU time, mass and heat transports seem robust with respect to the choice of regularization and can be estimated reasonably well. Employing the $H^{-1/2}$ method with the cutoff $E_{\min} = 10^{-3}$, we expect, on one hand, that more than 99% of elementary modes contributing to their variance are accurately resolved and, on the other hand, that we suppress numerical noise coming from the rest part of the spectrum.

5 Discussion

This paper deals with the problem of assigning confidence intervals to estimates of individual observables, determining amplitude of possible deviation of the true solution from the most probable one, and investigation of the solution stability. It should be stressed that we only consider numerical models of finite dimensions and do not investigate into their relations to the corresponding continuous prototypes. The outline of the peculiarities and shortcomings is as follows.

First, we confine ourselves to the Gaussian approximation and perform numerically a complete spectral analysis of the Hessian operator associated with the extremal point of the likelihood function. Evaluation of spectral distributions not only provides us with information about the impact of regularization on linear stability of the problem, but also shows what portion of the phase space becomes "visible" to iterative solvers after they perform a prescribed number of iterations. In case of high-dimensional poorly conditioned problems reliability of estimates comes to the fore. From this point of view our approach has an advantage over traditional linear systems solvers which employ stopping criteria based on checking the magnitude of the current relative change of the estimate. Theoretically, common solvers may stop even when a substantial contribution to the answer is still lacking or, on the contrary, may pass a solution and proceed further only amplifying numerical noise. In contrast, spectral analysis provides a criterion for the choice of a reasonable number of iterations: we only have to check that vital eigenvalues are resolved. However, it should be stressed that even if we are sure that 99.9% of eigenvalues are already resolved, there is no guarantee that the remaining 0.1% do not dominate in the true answer. But if we find that only 70% of the eigenvalues are resolved, we have a good reason to discard the current estimate and continue the iteration to improve the resolution.

Our approach is based on expansions of delta-functions and inverse square roots in a series of Chebyshev polynomials. On one hand, one should expect that expansions in a series of polynomials generated, say, by the

conjugate gradients method or any other method based on decomposition into Krylov's subspaces may converge faster than Chebyshev's. On the other hand, all these polynomials exhibit violent fluctuations in between the spectral points of the Hessian operator and can not be used for evaluation of spectral functions. In contrast, shifted Chebyshev polynomials of the first kind behave perfectly well over the whole spectral range and seem to be suitable for numerical computation. Also it is worth emphasizing that Chebyshev's expansions allow us to employ the entire power of 1D filter design and, given the number of iterations (or, equivalently, CPU time), to estimate the resolution beforehand.

Acknowledgments. We are grateful to D. Nechaev for his interest and help. AY thanks S. Becquey, H. Borth, G. Cortese, O. Eisen, L. Licari for numerous fruitful discussions and Alfred-Wegener-Institute for hospitality and support. This work was funded in part by the German CLIVAR Programme 03F0246B, AWI contribution No. 1726, and the Frontier System for Global Change, SOEST/IPRC contribution 4990/38.

Appendix

The computational algorithm of our study is aimed at estimating the spectral distribution of the Hessian operator associated with an objective function at its minimum and covariances of scalar observables. The user has to provide a subroutine that multiplies a vector by the Hessian, the number of calls to this operation that can be afforded, and the required accuracy.

The method is summarized as follows: spectral functions (11) and (12) store all necessary information about the Hessian spectrum and covariances of observables. We can compute and plot them iteratively together with estimates of uncertainty expressed by integrals (13). Formally the technique is based on (23)–(24) for spectral functions, and on (19)–(21) for fluctuation amplitude and covariances. All involved terms may be computed according to (22) and (25).

In practical computations infinite series (21) and (23)–(24) must be truncated. In order to avoid the Gibbs effect, which is introduced by simple truncation with the Dirichlet or boxcar window, regularizing filters should be applied. While an approximation to (23)–(24) may be obtained with any standard smoothing window, a filter for truncating (21) is constructed by minimizing the expected error in estimating the covariances. The resulting window is related to Hilbert's transform through (32)–(34). However, the standard algorithm for computing the Hilbert transform coefficients is not the most convenient one because it requires the user to supply the number of times he is prepared to multiply a vector by the Hessian *and* the desired resolution in the spectral range. Instead of resolution we propose to choose the required accuracy of approximation as the second parameter that defines the smoothing window.

In contrast to standard solvers our method offers an opportunity to check *a posteriori* the spectral functions via visual inspection in order to decide whether the choices were sufficient to resolve the part of the Hessian spectrum that is of interest to a given application.

References

Axelsson, O. and V. A. Barker (1984). *Finite Element Solution of Boundary Value Problems, Theory and Computation*. Academic Press.

Bateman, H, e. a. (1953). *Higher Transcendental Functions*, Volume 2. New York: McGraw-Hill.

Dirac, P. (1981). *The Principles of Quantum Mechanics* (Fourth ed.). Oxford: Clarendon Press.

Hardy, G. H. (1949). *Divergent Series*. Oxford: Clarendon Press.

IEEE (1979). *Algorithm 5.1 IEEE Press*. New York: John Wiley and Sons.

Itzykson, C. and J. Zuber (1990). *Quantum Field Theory*. New York: McGraw-Hill.

Landau, L. D. and E. A. Lifshitz (1958a). *Quantum Mechanics*. Oxford: Pergamon Press.

Landau, L. D. and E. A. Lifshitz (1958b). *Statistical Physics*. Oxford: Pergamon Press.

Luong, B., B. Jacques, and J. Verron (1998). A variational method for the resolution of a data assimilation problem in oceanography. *Inverse Problems 14*, 979–997.

Mathews, J. H. (1992). *Numerical Methods for Mathematics, Science and Engineering*. Englewood Cliffs NJ: Prentice Hall.

Nagelkerke, N. J. (1992). *Maximum Likelihood Estimation of Functional Relationships*. Berlin: Springer.

Nechaev, D. and M. Yaremchuk (1995). Application of the adjoint technique to processing of a standard section data set: World ocean circulation experiment section s4 along 67°S in the Pacific Ocean. *Journal of Geophysical Research 100*(C1), 865–879.

Oppenheim, A. V. and R. W. Schafer (1989). *Discrete-Time Signal Processing*. Englewood Cliffs NJ: Prentice-Hall.

Parks, T. W. and C. S. Burrus (1987). *Digital Filter Design*. New York: John Wiley and Sons.

Redler, R., K. Ketelsen, J. Deng, and C. Böning (1998). A high-resolution numerical model for the circulation of the Atlantic Ocean. In H. Lederer and F. Hertweck (Eds.), *Proceedings in the Fourth European SGI/CRAY MPP Workshop*, pp. 95–108. http://www.ifm.uni-kiel.de/to/FLAME.

Rudin, W. (1991). *Functional Analysis*. New York: McGraw-Hill.

Schwartz, L. (1967). *Analyse Mathématique*. Paris: Hermann.

Thacker, W. C. (1989). On the role of Hessian matrix in fitting models to data. *Journal of Geophysical Research 94*(C5), 6177–6196.

Yaremchuk, A. and J. Schröter (1998). Spectral analysis of symmetric operators: Application to the Laplace tidal model. *Journal of Computational Physics 147*, 1–21.

Yaremchuk, M., D. Nechaev, J. Schröter, and E. Fahrbach (1998). A dynamically consistent analysis of circulation and transports in the southwestern Weddell Sea. *Annales Geophysicae 16*, 1024–1038.