The Analysis of an Ill-Posed Problem Using Multi-Scale Resolution and Second-Order Adjoint Techniques

I. M. Navon

Department of Mathematics and

School of Computational Science and Information Technology

Florida State University, Tallahassee, Florida

Outline of talk

- Ill-posed problems
- Singular Value decomposition as a tool for regularization
- Outline of regularization methods
- Application: regularization of ill-posed problem of parameter estimation.
- Numerical and implementation aspects
- Other applications of SVD
  - adaptive observations
  - model reduction
- Summary and conclusions

Ill-posed and inverse problems

The concept goes back to Hadamard(1923). He defined a problems as ill-posed if the solution is not unique or if it is not continuous function of the data.

This means an arbitrary small perturbation of the data can cause an arbitrarily large perturbation of the solution.

Inverse problems arise quite naturally in many areas of science and engineering.

What is an inverse problems?

Determining internal structure of a physical system from the systems measured behavior or determining unknown input that gives rise to a measured output signal.

Examples include

- acoustics(Santosa *et al.* 1984)
- computerized tomography
- inverse geo-seismology
- mathematical biology
- optics and image restoration
- remote sensing

Example:

Astronomical image deblurring

Input is the night sky

Blurring system is the telescope and the atmosphere.

”Output” is the record blurred image.

Goal: reconstruct the input, *i.e.* the unblurred image given a mathematical description of blurring effects of telescope and atmosphere.

Example

Computerized tomography

Input is an X-ray source

System is the object being scanned (often the human brain)

Output is the measured damping of X-rays.

Goal is to reconstruct the “system” *i.e.* scanned object from information about locations of X-ray sources and measurements of their damping.

Tools and concepts of regularization

Important theoretical as well as computational tools in connection with rank-deficient and discrete ill-posed pbs.

Refs Hanson(1971),Varah(1973), Stewart(1993),Per Christian Hansen(1998).

Let $A \in R^{m \times n}$ be a rectangular or square matrix and assume without loss of generality($m \geq n$).

The SVD of A is a decomposition of the form:

$$A = U\Sigma V^T = \sum_{i=1}^{n} u_i \sigma v_i^T, \tag{1}$$

where

$$U = (u_1, u_2, \cdots, u_n) \in R^{m\times n},$$
$$V = (v_1, v_2, \cdots, v_n)$$

are matrices with orthonormal columns

$$U^T U = V^T V = I_n$$

and where the diagonal matrix

$$\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_n)$$

has nonnegative diagonal elements in nonincreasing order s.t.

$$\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n \geq 0$$

the numbers σ_i are singular values of A, while the vectors u_i and v_i are left and right singular vectors of A, respectively.

The SVD is defined for any m and n.

If $m < n$, simply apply (1) to A^T and interchange U and V.

The SVD of A provides sets of orthonormal basis vectors, s.t. that the matrix becomes diagonal when transformed into these two bases.

The singular values are always well conditioned w.r.t. perturbations.

If A is perturbed by a matrix $\mathbf{F}$, then $||\mathbf{F}||_2$ is an upper bound for absolute perturbation of each singular value.

Singular values σ_i can be defined as stationary values of $\frac{||Ax||_2}{||x||_2}$.

We have

$$A^T A = V \Sigma^2 V^T$$

and

$$A A^T = U \Sigma^2 U^T$$

SVD of A is strongly linked to eigenvalue decomposition of symmetric semi definite matrices AA^T and AA^T.

SVD is unique for a given matrix up to a sign change in pair(u_i, v_i) except for singular vectors associated with multiple singular values.

Characteristic features of SVD

Singular values σ_i decay gradually to zero with no particular gap in the spectrum.

When the dimension of A increase–this increases number of small singular values.

Left and right singular vectors u_i and v_i tend to have more sign changes in their elements as the index i increases, *i.e.* as σ_i decreases.

To see how SVD provides insight into ill-conditioning of A, consider following relations:

$$\left.\begin{array}{ll} Av_i = \sigma_i u_i, & \|Av_i\|_2 = \sigma_i \\ A^T u_i = \sigma_i v_i, & \|A^T u_i\|_2 = \sigma_i \end{array}\right\} i = 1, 2, \cdots, n.$$

We see that small singular value σ_i compared to

$$\sigma_1 = \|A\|_2$$

means that there exists a certain linear combination of columns of A characterized by elements of the right singular vector v_i, *s.t.* $\|Av_i\|_2 = \sigma_i$ is small. The same holds for u_i and the rows of A.

This means that a situation with one or more small σ_i implies A is nearly rank of deficient and vectors u_i and v_i associated with the small σ_i are the numerical null vectors of A^T and A respectively.

Thus the matrix in a discrete ill-posed problem is always ill conditioned and its numerical null-space is spanned by vectors with many sign changes.

The classical algorithm for computing SVD of a dense matrix is due to Golub, Kahan and Reinsch(1965,1970) .

Subroutines for computing the SVD of a dense matrix are available in most math software libraries today such as LAPACK Library.

The computational cost of SVD algorithm for dense matrices is $14mn^2 + 8n^3$ to compute full SVD.

There are rank-revealing methods when A has a numerical $\varepsilon-$rank r_ε and there is a well-determined gap between singular values σ_{r_ε} and $\sigma_{r_\varepsilon+1}$ where $\sigma_{r_\varepsilon} > \varepsilon \geq \sigma_{r_\varepsilon+1}$.

Such methods are the rank revealing QR and LU factorization.

The QR factorization of an $n \times m$ matrix is given by

$$A = QR$$

where $Q \in R^{m \times m}$ is orthogonal.

$R \in R^{m \times n}$ is upper triangular,$m \geq n$.

If A has full column rank, the first n columns of Q form an orthonormal basis for range(A).

Methods for rank deficient problems

Consider given matrix A as a noisy representation of a mathematically rank deficient matrix and replace A by a matrix close to A and mathematically rank deficient rank$-k$ matrix A_k defined as

$$A_k = \sum_{i=1}^{k} u_i \sigma_i v_i^T$$

where $\sigma_{k+1}, \cdots, \sigma_n$ small singular values are set to zero. which means we project ill-conditioned matrix A onto set of rank r_ε matrices.

Direct regularization

Developed by Tichonov(1963) and Phillips(1963).

Include or incorporate “a priori” assumptions about size and smoothness of the desired solution in the form of smoothing function

$$\min\{\|Ax-b\|_2^2+\lambda^2\|L(x-x^*)\|_2^2\}$$

$\|Ax\|_2$ residual

$\lambda-$ regularization parameter

$\Omega_x=\|L(x-x^*)\|_2$ discrete smoothing norm

The smoothing norm is in standard form if L is the identity matrix.

The key idea of Tichonov is to incorporate a priori assumptions about size and smoothness of the desired solution. Transformed into

$$\min\{\|\bar{A}\bar{x} - \bar{b}\|_2^2 + \lambda^2\|\bar{x} - \bar{x}^*\|_2^2\}$$

Typical Tichonov regularization leads to minimization

$$\min\{\|Ax - b\|_2^2 + \lambda^2\|Lx\|_2^2\}$$

where the regularization parameter λ controls the weight given to minimization of regularization term relative to the minimization of the residual norm.

Applications of Regularization Methods

Estimation of inflow boundary conditions is of significance in many heat transfer engineering problem. The estimation problem for inverse heat conduction is important when direct measurements on the considered boundary are either unreliable or impossible. Such examples are

- Ablation problems where satellites reenter the atmosphere,
- Control of welding processes, ovens or quenching baths

One searches in the inverse heat conduction problem for an unknown boundary condition $q(t) \in \mathcal{V}$, $\mathcal{V}$ being the parameter space, by minimizing an objective function

$$J(q) = ||\Phi(q) - \hat{Z}||^2_{\mathcal{Z}} \tag{2}$$

where the mapping to the space of observations $\Phi : q \in \mathcal{V} -> z \in \mathcal{Z}$ is known, q denotes a function to be determined, $\hat{Z}$ is a noisy observation and $\mathcal{Z}$ and $\mathcal{V}$ are Hilbert spaces.

Ill-posedness of the Inverse Heat Conduction Problem (IHCP) is characterized by the fact that the minimum solution of (2) does not depend continuously on the given observation $\hat{Z}$, i.e. for any tolerance $\epsilon > 0$ the solution set $S_\epsilon = \{q \in \mathcal{V} | J(q) < \epsilon^2\}$ is unbounded.

To stabilize the solution two approaches are possible:

- Modify the objective function by adding a penalization term
- Modify sought function space by discarding its ill-posed parts

This consists of two stages:

- Find the solution set S_ϵ
- Select a convenient solution

If one linearizes the mapping Φ (i.e.using linear mapping $A : \mathcal{V} -> \mathcal{Z}$) the objective functional is replaced by

$$J(q) = ||A\ q - \hat{Z}||^2_{\mathcal{Z}} \tag{3}$$

Let $\mathcal{N}(A)$ be the null space of A. If x^* is solution of $A\ x = \hat{Z}$ then

$$x^* + \mathcal{N}(A) \tag{4}$$

is the solution set of $A\ x = \hat{Z}$. To solve $A\ x = \hat{Z}$ we decompose $\mathcal{V}$ into

$$\mathcal{V} = \mathcal{N}(A) \oplus \mathcal{N}(A)^\perp \tag{5}$$

(it can be shown that $range(\mathcal{N}(A)^\perp) = null(A^T)$) and we restrict the solution to be only on $\mathcal{N}(A)^\perp$.

Directional sensitivity

The quantity $\frac{||Ax||}{||x||}$ is called directional sensitivity of mapping A in direction x and provides characteristics about the solution set S_ϵ.

For a given direction x, if sensitivity is small, the solution set S_ϵ contains a large interval in that direction.

Otherwise an error in the noisy data could be amplified by a factor $\frac{||x||}{||Ax||}$ in the solution.

For each subspace $\mathcal{U} \subset \mathcal{V}$ we introduce the following boundary functions

$$B_{inf}(\mathcal{U}) = \inf_{x \in \mathcal{U} \setminus \{0\}} \frac{||Ax||}{||x||}$$

$$B_{sup}(\mathcal{U}) = \sup_{x \in \mathcal{U} \setminus \{0\}} \frac{||Ax||}{||x||}$$

For numerical reasons $\mathcal{V}$ is decomposed into

$$\mathcal{V} = \mathcal{V}^+ \oplus \mathcal{V}^-$$

such that $\mathcal{V}^-$ is a numerically null space and the restriction of A on $\mathcal{V}^+$ is numerically invertible i.e. $B_{sup}(\mathcal{V}^-) < \epsilon$ and $B_{inf}(\mathcal{V}^+) > \epsilon$ for some $\epsilon > 0$.

Decomposition of space $\mathcal{V}$ with SVD

Let A^* be the adjoint of A. A^*A is a symmetric compact non-negative linear operator on $\mathcal{V}$ and one obtains an orthonormal set v_k of eigenfunctions of A^*A

$$A^*Av_k = \lambda_k^2 v_k \quad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq 0 \tag{6}$$

λ_k are called right singular values of A and v_k are called singular functions.

For each $\epsilon > 0$ we can decompose $\mathcal{V}$ into

$$\mathcal{V} = \mathcal{V}_\epsilon^+ \oplus \mathcal{V}_\epsilon^-$$

where $\mathcal{V}_\epsilon^+$ is spanned by the set of singular functions $\{v_1, \cdots, v_l\}$ and $\mathcal{V}_\epsilon^-$ is spanned by the set $\{v_{l+1}, v_{l+2}, \cdots\}$ with $\lambda_l \geq \epsilon \geq \lambda_{l+1}$.

For a numerically stable inverse solution of A we need to discard the ill-posed parts $\mathcal{V}_\epsilon^-$ and limit the solution only on $\mathcal{V}_\epsilon^+$.

This corresponds to the so-called truncated singular value decomposition method TSVD.

When SVD is difficult to implement one can attempt to find two suboptimal subspaces $\mathcal{U}_\epsilon^-, \mathcal{U}_\epsilon^-$ for each given $\epsilon > 0$ such that $B_{sup}(\mathcal{U}_\epsilon^-) < \epsilon$ and $B_{inf}(\mathcal{U}_\epsilon^+) > \epsilon$.

A wavelet basis is useful for such a decomposition (See Wavelets: Theory and Applications, Erlebacher,Hussaini and Jameson Eds 1996). The SVD is replaced by a multi-scale(wavelet) decomposition.

Using Mallat's multiresolution analysis (1989) wavelet theory yields a series of embedded subsets of $\mathcal{V}$

$$\mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2 \cdots \subset \mathcal{V}_j \subset \mathcal{V}$$

and $\mathcal{V}_j = \mathcal{V}_0 \oplus \mathcal{W}_0 \oplus \mathcal{W}_1 \oplus \cdots \oplus \mathcal{W}_{j-1}$
where the notation $\oplus$ (i.e. $\mathcal{V}_0 = \mathcal{V}_1 \oplus \mathcal{W}_1$)
indicates that the vectors in $\mathcal{V}_1$ are orthogonal to the vectors in $\mathcal{W}_1$ and the space $\mathcal{V}_0$ is simply decomposed into these two component subspaces. Here $\mathcal{V}_j$ is a subspace spanned over the wavelet mother function of scale j.

$$\mathcal{W}_j = span\{H_j^i(x) = 2^{\frac{j}{2}} H(2^j x - i) | i \in \mathbf{Z}\}$$

with $\mathbf{Z}$ the set of integers and H the Haar basis functions.

Using above decomposition,for every scale j the space $\mathcal{V}$ can be decomposed in

$$\mathcal{V} = \mathcal{V}_j \oplus \mathcal{V}_j^{\perp}$$

where $\mathcal{V}_j^{\perp} = \mathcal{W}_j \oplus \mathcal{W}_{j+1} \oplus \cdots$ is the subspace of high scale wavelets.

Work of Liu et al. (1995) demonstrated for Haar basis that $B_{sup}(V_j^{\perp}) \leq C2^{-\frac{j}{2}}$ and for every ϵ one can find a j such that $B_{sup}(V_j^{\perp}) \leq \epsilon$.

Thus wavelet transformation provides an ordering of subspaces $\mathcal{U}_j^{\perp}$ according to the scale j and the minimal eigenvalues.

The discrete wavelet transform is used for approximation of control function

$$f(y) = a_0\phi_0(y) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} a_k^j W(2^j y - k)$$

written in finite-dimensional form as $f_i = W a_k^j, 1 \leq i \leq N$, j is the scale, k the translation and N the number of controls.

Application to the 2-D parabolized Navier-Stokes equations

Consider the problem of estimating unknown inflow parameters

$$f_\infty(y) = (\rho(y), U(y), V(y), T(y))$$

on the inflow boundary from measurements in a flow field section $f^{obs}(x_m, y_m)$.

A cost functional consisting of the discrepancy of measured and calculated flow parameters is minimized.

We consider laminar flow which is supersonic along x coordinate

$$\frac{\partial(\rho U)}{\partial x}+\frac{\partial(\rho V)}{\partial y}=0$$

$$U\frac{\partial U}{\partial x}+V\frac{\partial U}{\partial y}+\frac{1}{\rho}\frac{\partial P}{\partial x}=\frac{1}{Re\rho}\frac{\partial^2 U}{\partial x^2}$$

$$U\frac{\partial V}{\partial x}+V\frac{\partial V}{\partial y}+\frac{1}{\rho}\frac{\partial P}{\partial y}=\frac{4}{3Re\rho}\frac{\partial^2 V}{\partial y^2}$$

$$U\frac{\partial e}{\partial x}+V\frac{\partial e}{\partial y}+(k-1)e\left(\frac{\partial U}{\partial y}+\frac{\partial V}{\partial y}\right)=\left(\frac{K}{\rho RePr}\frac{\partial^2 e}{\partial y^2}+\frac{4}{3Re\rho}\left(\frac{\partial U}{\partial y}\right)^2\right)$$

where $e=C_V T=\frac{RT}{k-1}$, $\quad P=(K-1)\rho e=\rho RT$, and
$0<x<1, \quad 0<y<1.$

The entrance Boundary Conditions (for $x = 0$) are as follows:

$$e(0, y) = e_\infty(y)$$
$$\rho(0, y) = \rho_\infty(y)$$
$$U(0, y) = U_\infty(y)$$
$$V(0, y) = V_\infty(y)$$

The outflow condition $\frac{\partial f}{\partial y} = 0$ is used at $y = 0$ and $y = 1$.

We search for inflow parameters

$$f_\infty(y) = (\rho(y), U(y), V(y), e(y))$$

using outflow data by minimizing the cost functional

$$J(f_\infty(y)) = \int_{x=0}^{x_{max}} \int_{y=0}^{1} \left(f^{obs}(x,y) - f(x,y) \right)^2 \delta(x-x_m)\delta(y-y_m)dxdy$$

Since we deal with a nonlinear problem we will use the Hessian of the cost functional with respect to the control variables instead of the A^*A operator.

One can define a Lagrangian

$$\mathcal{L}(f_\infty, f, \Psi) = J(f_\infty(y)) + \int\int_\Omega \left(\frac{\partial(\rho U)}{\partial x} + \frac{\partial(\rho V)}{\partial y} \right) \Psi_\rho(x,y)dxdy$$

$$+ \int\int_\Omega \left(U\frac{\partial U}{\partial x} + V\frac{\partial U}{\partial y} + \frac{1}{\rho}\frac{\partial P}{\partial x} - \frac{1}{Re\rho}\frac{\partial^2 U}{\partial x^2} \right) \Psi_U(x,y)dxdy$$

$$+ \int\int_\Omega \left(U\frac{\partial V}{\partial x} + V\frac{\partial V}{\partial y} + \frac{1}{\rho}\frac{\partial P}{\partial y} - \frac{4}{3Re\rho}\frac{\partial^2 V}{\partial y^2} \right) \Psi_V(x,y)dxdy$$

$$+ \int\int_\Omega \left(U\frac{\partial e}{\partial x} + V\frac{\partial e}{\partial y} + (k-1)e\left(\frac{\partial U}{\partial y} + \frac{\partial V}{\partial y} \right) \right.$$

$$\left. - \left(\frac{K}{\rho RePr}\frac{\partial^2 e}{\partial y^2} - \frac{4}{3Re\rho}\left(\frac{\partial U}{\partial y} \right)^2 \right) \right) \Psi_e(x,y)dxdy$$

Here the Lagrange multipliers (adjoint parameters) are

$$\Psi_\rho(x,y), \Psi_U(x,y), \Psi_V(x,y), \Psi_e(x,y) \in \mathbf{H}^{1,2}(\Omega)$$

where $\mathbf{H}^{1,2}(\Omega)$ is the Hilbert space of functions having derivatives of first order in x and derivatives and derivatives of second order in y .

Fischer Information Matrix

An approach using Fischer information matrix (approximating the Hessian of the cost function in vicinity of the solution) can be used for estimation of problem ill-posedness.

For the cost functional

$$J(u) = \sum_i (f_i^{obs} - f_i(u))^2$$

u being the vector of control parameters, the Hessian of the cost functional with respect to the control parameters has the form

$$H_{jk} = \frac{\partial^2 J}{\partial u_{jk}^2} = \sum_i \frac{\partial f_i}{\partial u_j} \frac{\partial f_i}{\partial u_k} - 2 \sum_i (f_i^{obs} - f_i) \frac{\partial^2 f_i}{\partial u_k \partial u_k}$$

Let $S_{ij} = \frac{\partial f_i}{\partial u_j}$ denote the rectangular sensitivity matrix and let us define $M_{ij} = \frac{\partial f_k}{\partial u_i}\frac{\partial f_k}{\partial u_j}$ as the information matrix.

Let σ be the data error dispersion then Fisher matrix is defined by

$$\tilde{M}_{ij} = \sum_k \frac{1}{\sigma_k^2}\frac{\partial f_k}{\partial u_i}\frac{\partial f_k}{\partial u_j}$$

The inverse matrix $D = \tilde{M}^{-1}$ is a dispersion matrix of the control parameters u_j.

The magnitude of the Fischer information matrix minimal eigenvalue can be comparable to the data error, i.e. $\lambda_{min} \approx \sigma^2$, i.e. it can be used to estimate the problem's ill-posedness.

Algorithmic form of sequential multiscale algorithm

Using adjoint methods can provide a cheap way for the calculation of the Hessian action and consequently the bounds of its eigenvalues spectrum.

We may then examine subspaces from the perspective of minimal eigenvalue of the Hessian and *a priori* known data error.

The algorithm then consists of the following stages:

1) Calculate minimal eigenvalue of Hessian of cost functional. If $\lambda_{min} \geq \sigma^2$ (with σ being the data error dispersion) the problem is considered well-posed. Otherwise if $\lambda_{min} < \sigma^2$ we perform the following operations.

2) Control space is transformed into wavelet space

3) Coefficients of the smallest scale are discarded (i.e. we effectively decrease control space dimension by a factor of 2).

4) Recalculate Hessian minimal value for next control space. If $\lambda_{min} \geq \sigma^2$ then problem is considered well-posed in this subspace. If not then the following scale is discarded and the iteration is repeated.

First and Second Order Adjoint

The solution of the adjoint problem is used for gradient of cost calculation, while the action of the Hessian is obtained by using second order adjoint.

Let x denote the marching coordinate, f the vector of flow parameters, C the observation operator, J the cost function measuring the discrepancy between model calculation and the observation and let U be the vector of control parameters (in this case the inflow boundary parameters).

The forward problem

$$\frac{df}{dx} = F(f)$$

$$f(0) = U$$

$$J(u) = \frac{1}{2}\int\int (Cf - f^{obs})^2 dxdy$$

First order adjoint (Ψ are adjoint variables)

$$\frac{d\Psi}{dx} + \frac{\partial F^T}{\partial f}\Psi = C^T(Cf - f^{obs})$$

$$\Psi(1) = 0$$

$$\nabla J(U) = -\Psi(0)$$

Tangent problem

$$\frac{\partial \hat{f}}{\partial x} = \frac{\partial F}{\partial f}\hat{f}$$

$$\hat{f}(0) = u$$

Second order adjoint problem (tangent to first order adjoint)

$$\frac{\partial \hat{\Psi}}{\partial x} + \left[\frac{\partial^2 F}{\partial f^2}\hat{f}\right]\Psi + \left[\frac{\partial F^T}{\partial f}\right]\hat{\Phi} = C^T C\hat{f}$$

$$\hat{\Phi}(1) = 0$$

$$\nabla \hat{J}(U) = H(U)u = -\hat{\Phi}(0)$$

Thus the Hessian action on the vector U may be obtained by solving sequentially the above four initial boundary value problems:

1. Forward problem (X is increasing)
2. First order adjoint problem (x is decreasing)
3. Tangent linear problem (x is increasing)
4. Second order adjoint problem (x is decreasing)

The typical cost of finding the action of a Hessian of the cost on a vector is $\approx 4N$. One can use a less accurate method i.e. the Hessian free finite-difference expression

$$Hdu = (grad(u + hdu) - grad(u))/h$$

which is about as costly as the exact approach.

The data error is modeled by a normally distributed random values δf with dispersion $\sigma = f^{obs} + \delta f$.

The flow field was computed using a finite-difference method marching along x with first order of accuracy in x and second order in y. The pressure gradient for supersonic flow is computed from energy and density.

Same grid was used for adjoint problem solution but integration is in the reverse direction beginning at $x = x_{max}$. The grid has 50 points along y and 100 points along x.

A limited-memory quasi-Newton L-BFGS large-scale unconstrained optimization was used to perform the minimization. Iterations of the type

$$x_{m+1} = Hx_m$$

and

$$\lambda = \frac{max(x_{m+1})}{max(x_m)}$$

are used for obtaining maximum eigenvalues λ_{max}.

The minimal eigenvalue is calculated by a method of shifted iteration

$$M * E - H$$

where M is the eigenvalues majorant and E is the unit matrix.

The minimum eigenvalue can also be calculated using the Rayleigh quotient algorithm

$$R(x) = \frac{(Hx,x)}{(x,x)}, \quad \min R(x) = \lambda_{\min}$$
$$grad(R(x)) = c(Hx - R(x)x)$$

Iterations were performed using the steepest–descent method of the form

$$x^{k+1} = x^k - \alpha(Hx - R(x)x)$$

When minimal eigenvalues are relatively large (10^{-3}) both algorithms obtain similar results. For smaller eigenvalues Rayleigh quotient turned out to be more accurate.

Results

Tests were carried out for a typical error of $\sigma = 0.01$. Iterations are stopped at $\delta < \sigma_0$ where $\delta = (J/N)^{1/2}$ is the normalized cost function and σ_0 is the data error.

A Mach number of 4 was used with Reynolds numbers varying in the range $10^2 - 10^7$.Low Reynolds numbers(high viscosity) correlate with small eigenvalues and poor quality of parameters obtained.(See Fig 2)

For the wavelet transformation a Daubechies-20 transformation was used. Calculations were performed for estimating the variation of λ_{min} eigenvalue with change of scale j.

The discrete wavelet transformation was used for control functions approximation. $\lambda_{\min}$ is increasing as Reynolds number increases.

The wavelet transformation allows us to use only $\ln(N)$ calculation of eigenvalues instead of N–when the entire Hessian spectrum is used.

We compared wavelet regularization at different scales with the Tichonov regularization.

Standard Tichonov regularization transform the cost function into

$$J(f_\infty(y)) = \int_x \int_y (f^{obs}(y) - f(x_{\max}, y))^2 \, dy \, dx + \alpha \iint f_\infty(y, x))^2 \, dx \, dy$$

α being the regularization parameter.

This variant usually results in overly smoothed inflow parameters.

One can consider addition of a second–order Tichonov regularization term

$$\alpha(f_{i+1} - 2f_i + f_{i-1})^2$$

to the cost functional $J(f)$.

This provides a much better quality of regularization and the result is similar in quality to the result obtained using the wavelet transformation.

The search for the magnitude of a suitable Tichonov regularization parameter α is not transparent and does not provide for a counterpart of the automatic procedure of comparing minimal eigenvalue and data error used in the wavelet regularization.

Summary and Conclusions

- Numerical tests demonstrate a monotonous decrease of the Hessian minimal eigenvalues as the scale of the wavelet transformation decreases.
- An algorithm of regularization based on multi-scale resolution (wavelet analysis) and Hessian minimal eigenvalue calculation. was tested
- It is based on using the solution of second order adjoint of the problem.
- Numerical tests using a 2-D parabolized Navier–Stokes equation model show applicability of this algorithm to the solution of inverse convection problem–allowing optimal estimation of inflow parameters from measurements in a down stream flow field section.

- Use of adjoint method is efficient when a large number of control parameters has to be estimated.
- A comparison with a zeroth order Tichonov regularization method reveals that the latter yields unacceptable results(oversmoothing).
- The approach outlined here can be readily extended to other problems where ill-posedness is present and when adjoint parameter estimation is carried out.

Regularization in optimal control

Many typical optimal control problems have the form:
Given

1. *State variable* ϕ
2. *Control or design parameters* g
3. *Cost or objective function* $J(\phi, g)$
4. *Constraint equations* $F(\phi, g) = 0$

find controls g **and states** ϕ **such that** $J(\phi, g)$ **is minimized subject to** $F(\phi, g) = 0$

Many optimization problems set in this manner are ill-posed in the sense that solutions with bounded controls do not exist.

This happens when the objective cost functional to be minimized does not explicitly depend on the controls or design parameters. One must then limit size of the controls. There are two ways to do this

- Penalize the objective functional with some norm of the control, i.e. instead of minimizing a functional $\Psi(\phi)$ one would minimize

$$J(\phi, g) = \Psi(\phi) + \sigma ||g||^2$$

- Place constraints on the size of the admissible controls to be within a bounded set, i.e. for a suitable norm $||g|| \leq K$

Regularization of the functional in optimal control

Functionals to be minimized are often not “well-behaved”. They may not be convex -or depend weakly on some design variables.

Discretized functionals may exhibit spurious numerical local minima which will slow down the convergence to desired minima or prevent it altogether.

Regularization of functional by addition of penalty terms can improve optimizer performance at least in the early stages of minimization.

Regularization reduces the number of optimization iterations since one works –at least in the beginning– with “easier” functional.

Regularization changes position of minima. An idea is to reduce the size of the penalty or eliminate it once one is “near” the minima.

This may result in inaccuracies and requires choice of penalty parameters and knowledge when to lower the values of penalty parameters.

In Navier-Stokes equations there are still open questions about such approaches.

References

[1] J. Liu, B. Gurrier, C. Denard, A sensitivity decomposition for regularized solution of inverse heart conduction problems by wavelets, Inverse Problems 11 (1995) 1177-1187.

[2] Z. Wang, I.M. Navon, F.X. LeDimet, X. Zou, The second order adjoint analysis: Theory and applications, Meteorol. Atmos. Phys. 50 (1992) 3-20.

[3] V.Yu. Terebigh, The restoration of images using minimal a priori information, Uspehi Fizicheskih Nauk, 165 (2) (1995) 143-176 (in Russian).

4] O.M. Alifanov, E.A. Artyukhin, S.V. Rumyantsev, Extreme Methods for Solving Ill-Posed Problems with Applications to Inverse Heat Transfer Problems, Begell House Publishers, 1996.

[5] I. Daubechies, Orthonormal basis of compactly supported wavelets, Commun. Pure Appl. Math. 41 (7) (1988) 909-996.

[6] F.X. Le Dimet and O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations, Tellus 38A ,2(1986) 97-110.

[7] Liu, D.C. Jorge Nocedal, On the limited memory BFGS method for large scale minimization, Math. Prog. 45 (1989) 503-528.

[8] A.K. Alekseev, On estimation of spatial distribution of entrance boundary parameters from flow field measurements for supersonic flow, Math. Model. 11 (12) (1999) 33-44.

[9] A.V. Tikhonov, V. Arsenin, Solution of Ill-posed Problems. Winston & Sons, Washington, DC, 1977, p. 224.

[10] Z. Wang, I.M. Navon, X. Zou, F.X. LeDimet, A truncated-Newton optimization algorithm in meteorology applications with analytic Hessian/vector products, Comput. Optimization Applications 4 (1995) 241-262.

[11] J. Liu, A sensitivity analysis for least-squares ill-posed problems using Haar basis, SIAM J. Numer. Anal. 31 (5) (1994) 1486-1496.

[12] G. W. Stewart , On the early history of the singular value decomposition. SIAM Review, 35 (1993), 551-566.

[13] R.J. Hanson , A numerical method for solving Fredholm integral equations of the first kind using singular values. SIAM J. Numer. Anal. 8 (1971), 616-622.

[14] J.M. Varah : On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems SIAM J. Numer. Anal.,10 (1973), 257-267.

[15] Per Christian Hansen : Rank Deficient and Discrete Ill-posed problems. SIAM , Philadelphia, 247pp (1998)

[16] G.H. Golub and W Kahan : Calculating the singular values and pseudo-inverse of a matrix. SIAM J. Numer Anal. Ser. B, 2 (1965) pp 205-224.

[16b] G.H. Golub and C. Reinsh, Singular value decomposition and least squares solutions. Numer. Math. ,14 , (1970),403-420.

[17] D. L. Phillips , A technique for the numerical solution of certain integral equations of the first kind. J. Assoc. Comput. Mach., 9,(1962) ,84-97.

[18] G. Erlebacher, M. Y. Hussaini and L. Jameson (eds) 1996: Wavelets: Theory and Applications. Oxford University Press.

[19] S. Mallat, (1989). ``A Theory for Multiresolution Signal Decomposition: the Wavelet Representation,'' IEEE Transactions on Pattern Analysis and Machine Intelligence, 11:674--693.

[20] Hadamard J., 1923, Lectures on the Cauchy Problem in Linear Partial Differential Equations. Yale University Press, New Haven

[21] Fadil Santosa, William W. Symes, Yih-Hsing Pao and Charles Holland (Eds), 1984: Inverse Problems of Acoustic & Elastic Waves. ISBN: 0898710502 - Society for Industrial and Applied Mathematics .

[22] Tichonov, A. N., 1963: Regularization of ill-posed problems. Dokd. Akad. Nauk, USSR, 153, 49-52.

[23] Francois X. Le Dimet, I. M. Navon and Dacian N. Daescu : Second Order Information in Data Assimilation Monthly Weather Review , 130, No. 3 ,629-648 (2002)