JOHANNES KEPLER UNIVERSITY LINZ | JKU

Optimality Criteria and Minimization Schemes for Parameter Choice

Frank Bauer

Fuzzy Logic Laboratorium Linz

19.02.2009

Introduction

In many areas of modern technology one needs the solution of inverse problems, examples:

- Deconvolution of noisy images
- Geodesy (e.g. satellite based gravitation and magnetic field measurements)
- Medical Imaging (e.g. EIT, Computerized Tomography, PET, Parallel MRI)
- Non-destructive testing (e.g. inner heat sources or conductivity from surface measurements)
- Learning Theory

All of these problems have one thing in common: small errors in the measurements can result in really big deviations in the reconstructions.

Structure

The presentation is now structured as follows:

- Some definitions
- Description of different models for solutions and noise components
- Optimality criteria
- Parameter choice in general
- Quasi-Optimality
- Some applications

Inverse Problems

In this talk we consider

$$F(x) = y$$

where

- x and y are in a separable Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$
- F is either
  - badly conditioned linear operator or
  - compact linear operator or
  - its Fréchet derivative $F'[\overline{x}]$ around x is compact
- We can just measure

$$y^\delta = y + \delta\xi$$

where ξ is some normalized noise component.

Singular Value Decomposition

In a number of cases it will be much easier to consider the purely linear case, then we denote F by A.

For mostly formal reasons (this will not be needed in most methods) we will require the singular value decomposition.

$$y^\delta = Ax + \delta\xi = \sum_{k=1}^{\infty} \left(\sigma_k \langle x, u_k \rangle + \delta \langle \xi, v_k \rangle\right) v_k$$

Regularization Methods

There exists a big number of regularization methods for various situations

- Spectral-Cut-Off Regularization (TSVD)
- Tikhonov Regularization
- Landweber Regularization
- Iterated Gauß-Newton Regularization
- ...

We will write x_n^δ for the n^{th} regularized solution and x_n^0 for the noise-free n^{th} regularized solution. I.e. we write:

$$x_n^\delta = A_n^{-1} y^\delta$$

$$x_n^0 = A_n^{-1} y$$

Assume that for all x

$$\|x_n^0 - x\| \searrow 0 \qquad \text{for } n \to \infty$$

Parameter Choice

A solution procedure for an inverse problem now consists out of

- A regularization method returning $\{x_n^\delta\}_{n\in\mathbb{N}}$
- A parameter choice procedure, selecting one particular solution $x_{n_*}^\delta$.

A parameter choice method can be, e.g.:

- A priori, i.e. $n_* = n_*(x, \delta)$
- Noise dependent (mostly on some information of the noise), i.e. $n_* = n_*(\delta\xi, y^\delta)$;
  - most classical methods fulfill $n_* = n_*(\delta, y^\delta)$
  - some other methods need an independent realization $\delta\widetilde{\xi}$ of the noise, i.e. $n_* = n_*(\delta\widetilde{\xi}, y^\delta)$
- Data driven, i.e. $n_* = n_*(y^\delta)$

Stochastic Noise

Stochastic noise:

- ξ Gaussian random variable
- $\mathbb{E} \langle \xi, v \rangle = 0 \qquad \forall v \in \mathcal{Y}$
- $\mathbb{E} \langle \xi, v_1 \rangle \langle \xi, v_2 \rangle = \langle v_1, K v_2 \rangle \qquad \forall v_1, v_2 \in \mathcal{Y}$
  where K is a positive definite self-adjoint matrix.

An assumption in this case is that $K = (AA^*)^\nu$. It follows that K diagonalizes trivially in the same singular system as A does and that the singular values of K are monotone.

Deterministic Noise

Classically one uses deterministic noise

$$\|\xi\| \leq 1$$

which is an appropriate model for discretization errors but *not* for measurement errors.

Restricted Deterministic Noise

The problem with this definition of noise is, that one has no information at all, where the energy of the noise lies, in contrary to the stochastic case.
This can be changed by an additional condition (for some $\beta > 0$) in the case of Tikhonov regularization with $\alpha_n = q^n$:

$$D_{up}(t) \geq \sum_{\{k:|\sigma_k^2 \geq t\}} \frac{\langle \xi, v_k \rangle^2}{\sigma_k^2} \geq D_{low}(t) \quad \forall 0 \leq t \leq \beta$$
$$E_{up}(t) \geq \sum_{\{k:|\sigma_k^2 \leq t\}} \sigma_k^2 \langle \xi, v_k \rangle^2 \geq E_{low}(t) \quad \forall 0 \leq t \leq \beta$$

$$\frac{E_{low}(qt)}{4q^2 E_{up}(t)} \geq E_1 > 1 \qquad \frac{4E_{up}(qt)}{q^2 E_{low}(t)} < E_0 < \infty$$
$$\frac{D_{low}(qt)}{4D_{up}(t)} \geq D_1 > 1 \qquad \frac{4D_{up}(qt)}{D_{low}(t)} < D_0 < \infty$$

Prior on x

Bayes case:

- x Gaussian random variable
- $\mathbb{E}\langle x, u\rangle = 0 \qquad \forall v \in \mathcal{X}$
- $\mathbb{E}\langle x, u_1\rangle \langle x, v_2\rangle = \langle u_1, Lu_2\rangle \qquad \forall u_1, u_2 \in \mathcal{X}$
  where L is a positive definite self-adjoint matrix.

An assumption in this case is that $L = (A^*A)^{\mu+1/2}$. It follows that L diagonalizes trivially in the same singular system as A does and that the singular values of L are monotone.

Deterministic x

Classically one uses deterministic solutions and source conditions, i.e.

$$\|(A^*A)^{-\mu}x\| \leq R$$

for some $R \in \mathbb{R}$ fixed.

Restricted deterministic x

As for the noise for some methods this is too unspecific. For Tikhonov-regularization in the non-saturated case we will additionally want to have fulfilled:
There exist constants $\gamma > 0,\ 1 > \eta > 0,\ \ C_{\gamma,\eta} > 0$ and $D_{\gamma,\eta} > 0$ such that for all $0 \leq t \leq \gamma$

$$D^2_{\gamma,\nu} t^{2\nu} \geq \sum_{\{k:|\sigma_k^2 \leq t\}} \langle x, u_k \rangle^2 \geq C^2_{\gamma,\eta} t^{2\eta}$$

B., F. and Kindermann, S., *The Quasi-Optimality Criterion for Classical Inverse Problems*, Inverse Problems Vol. 24, 035002 (20 pp) (2008)

Notation

We will now write (formally!):

- $x \in \mathcal{S} \, (\subset \mathcal{X})$
- $\xi \in \mathcal{N}$

Later on we will specify in what regime the specific parameter choice methods work in.
Another problem is, that depending on the regime, we will later on need quantities in the form

- Deterministic: $\sup_{x \in \mathcal{S}} f(x)$
- Stochastic: $\mathbb{E}_{x \in \mathcal{S}} f(x)$

We will now write for both situations

- $\mathbb{S}_{x \in \mathcal{S}} f(x)$

Optimality

There are several different concepts of optimality

- Convergence
- Rates
- Oracle inequalities

Convergence and rates

Convergence: It holds for all $x \in \mathcal{S}$ and $\xi \in \mathcal{N}$

$$\lim_{\delta \to 0} \|x - x_{n_*}^{\delta}\|^2 = 0$$

respectively in a more strict regime:

$$\lim_{\delta \to 0} \mathbb{S}_{x \in \mathcal{S}} \mathbb{S}_{\xi \in \mathcal{N}} \|x - x_{n_*}^{\delta}\|^2 = 0$$

Rates: There exists a function $f_{\mathcal{S},\mathcal{N}}(\delta) \searrow 0$ for $\delta \to 0$ such that

$$\mathbb{S}_{x \in \mathcal{S}} \mathbb{S}_{\xi \in \mathcal{N}} \|x - x_{n_*}^{\delta}\|^2 = O(f_{\mathcal{S},\mathcal{N}}(\delta))$$

Obviously, rate results are stronger than convergence results.

Rates discussed

Convergence rates have some specific problems

- the constant in front
- for what δ starts the rate to kick in
- is the rate optimal, or could it be improved for most solutions?

However, one of the biggest problems is the following one:

- for every finite dimensional realization of an inverse problem, we have rate $O(\delta)$, despite
  - the error is almost always bigger than the smallest eigenvalue.
  - for regularization methods like TSVD we *do not* see the difference in the result between the finite and infinite dimensional setting.

Oracle Inequalities

Oracle inequalities exist in a number of different forms, one of them is, e.g.:

$$\mathbb{S}_{x\in\mathcal{S}}\mathbb{S}_{\xi\in\mathcal{N}}\|x - x^\delta_{n_*}\|^2 \leq C_{\mathcal{S},\mathcal{N}} \min_n \mathbb{S}_{x\in\mathcal{S}}\mathbb{S}_{\xi\in\mathcal{N}}\|x - x^\delta_n\|^2$$

Another possibility is

$$\|x - x^\delta_{n_*}\| \leq C_{\mathcal{S},\mathcal{N}} \min_n \|x - x^0_n\| + \|x^0_n - x^\delta_n\|$$

Normally, the latter one just holds in very specialized regimes for $\mathcal{S}$ and $\mathcal{N}$ respectively just with "high probability".

Oracle Inequalities (2)

- Obviously, oracle inequalities imply rate results, as long the regularization method is converging.
- The concept works in a finite and infinite dimensional setting.
- One can show, that oracle inequalities are actually stronger than rate results. (Super-efficiency ...)
  See e.g.
  B., F., Hohage, T. and Munk, A., *Iteratively Regularized Gauss-Newton Method for Nonlinear Inverse Problems with Random Noise*, SIAM Journal on Numerical Analysis (accepted 2009)
  and references therein.

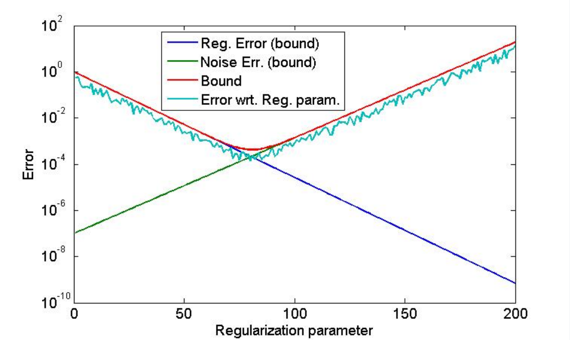
Parameter Choice

Now we will turn our attention to parameter choice, namely:

- Lepskij- Balancing Principle
- Quasi-Optimality

Both of them are particular in the sense, that they work in the space $\mathcal{X}$ only.

Lepskij-Balancing (Prerequisites)



Assume that we have

- The error $\|x_n^\delta - x\|$ which is bounded by the sum of
- a monoton. increasing function $\phi(n)$ and
- a monoton. decreasing function $\psi(n)(= \|x_n^0 - x\|)$.

ϕ and ψ are not changing faster than exponentially.

Lepskij-Balancing (Definition)

Assume that

- $\phi(n)$ is known (e.g. $\phi(n) = \delta\|A_n^{-1}\|$ in the deterministic noise case)
- A parameter N is known where $n_* < N$, e.g. $N = \psi^{-1}(\delta)$ in the deterministic case.

and define:

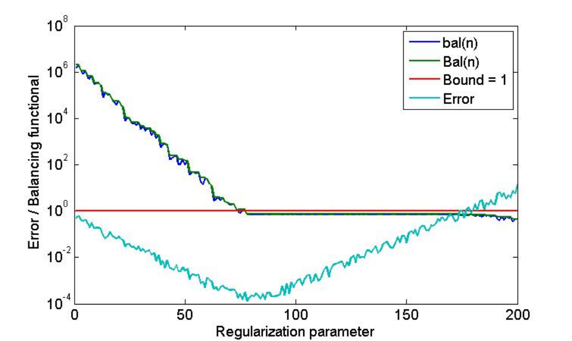
$$bal(n) = \max_{n<m\leq N} \frac{\|x_n^\delta - x_m^\delta\|}{4\phi(m)}$$

$$Bal(n) = \max_{n\leq m<N} bal(n)$$

$$n_* = \underset{n}{\operatorname{argmax}}\{Bal(n) > 1\}$$

Let furthermore $n_\#$ such that $n_\# = \operatorname{argmax}_n\{\phi(n) < \psi(n)\}$.

Lepskij-Balancing (Prerequisites)



Observations:

- Curve has L-shape
- It passes the bound 1 slightly before the minimum of $\|x_n^\delta - x\|$

Lepskij Balancing (Result)

It holds (deterministic noise)

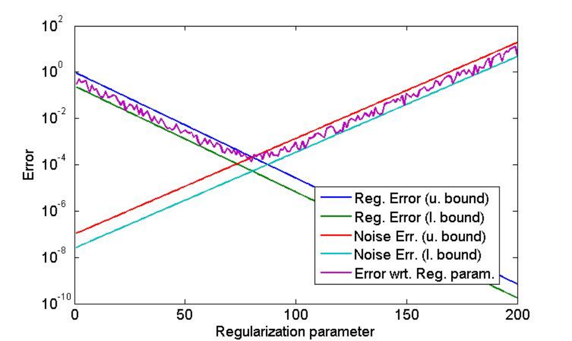
$$\|x^\delta_{n_*} - x\| \leq c \min_n \|x^0_n - x\| + \phi(n)$$

respectively (stochastic noise)

$$\mathbb{S}_{\mathcal{N}} \|x^\delta_{n_*} - x\| \leq c \log(\delta^{-1}) \min_n \|x^0_n - x\| + \phi(n)$$

both for linear and non-linear inverse problems.
This is a kind of oracle inequality bounding the error.

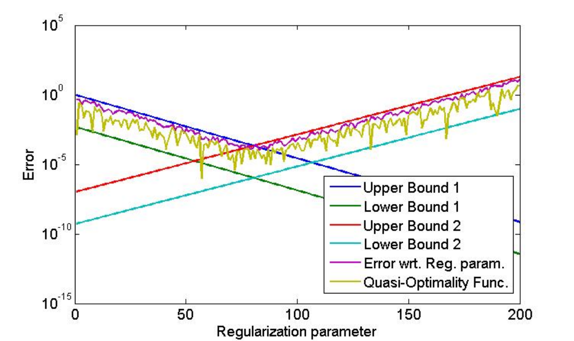
Quasi-Optimality (Prerequisites)



Assume that

- The regularization error and the noise error is upper and lower bounded
- Both of them are changing exponentially

Quasi-Optimality (Definition)



The quasi-optimality principle is defined by

$$n_* = \operatorname*{argmin}_n \|x_n^\delta - x_{n+1}^\delta\|$$

Quasi-Optimality (Idea of Proof)

All proofs work along the following lines

- In average, the result holds
- The probability that the minimum is attained somewhere far away from this average point is getting exponentially smaller
- Using the Hölder inequality we can show an oracle inequality.

This concept works for a number of minimization schemes, the minimal prerequisites they need to fulfill are rather low.

Quasi-Optimality (Result)

We can show in a Bayesian setting with stochastic noise:

$$\mathbb{E}_{x\in\mathcal{S}}\mathbb{E}_{\xi\in\mathcal{N}}\|x_{n_*}^\delta - x\|^2 \leq c \min_n \mathbb{E}_{x\in\mathcal{S}}\mathbb{E}_{\xi\in\mathcal{N}}\|x_n^\delta - x\|^2$$

Similar results can be shown in the other, partly or fully non-stochastic, cases.

Bayesian case:

B., F. and Reiss, M., *Regularisation independent of the noise level: an analysis of quasi-optimality*, Inverse Problems Vol. 24, 055009 (16pp) (2008)

Conclusion

We have shown

- Lepskij Balancing principle
  - Robust and working in more or less all know cases.
  - Needs rather strict knowledge about the noise behavior.
- Quasi-optimality principle
  - Needs no information about the noise ξ or the solution x.
  - Works under special conditions, which seem to be fulfilled in many practical cases
  - Needs great care choosing the set of regularization parameters.

Both of them yield oracle inequalities and therefore are (potentially) better than comparable classical parameter choice methods.

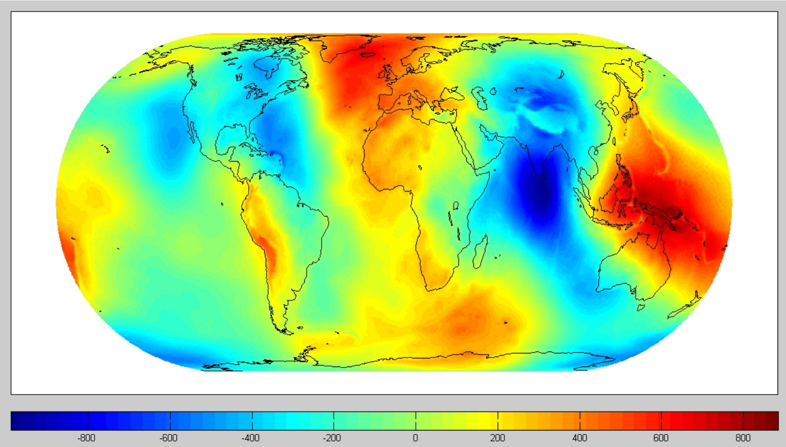
Gravitation

The satellite GOCE measures the second derivative of the gravitational field in a height of about 300km. One wants to reconstruct the gravitational field on the Earth

- For simulation we added correlated noise to the data
- Spectral cut-off along the degree of the spherical harmonics.
- The number of data for this kind of satellite missions is normally no problem. Therefore we assume that we can split the data into two data sets and that the noise is not correlated.
  Out of these two data sets one can estimate the correlated noise.
- The regularization parameter is determined via Lepskij-Balancing and Quasi-optimality.

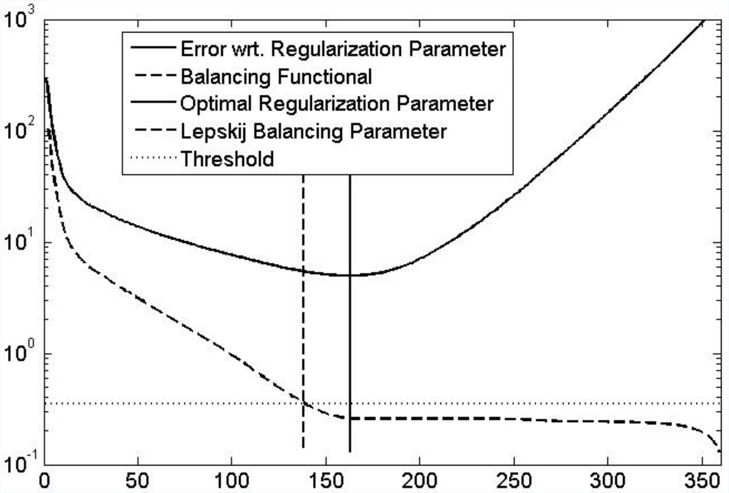
Original

The EGM96-gravitation model:



Lepskij balancing

Plot of the error and the parameter



Quasi-Optimality

Plot of the error and the parameter