Iterative Regularization Methods for Inverse Problems: Lecture 1

Thorsten Hohage

Institut für Numerische und Angewandte Mathematik
Georg-August Universität Göttingen

Madrid, April 4, 2011

Outline

Keller's definition of an inverse problem

> *"We call two problems* inverses *of one another if the formulation of each involves all or part of the solution of the other. Often, for historical reasons, one of the two problems has been studied extensively for some time, while the other is newer and not so well understood. In such cases, the former problem is called the direct problem, while the latter is called the inverse problem."*

J.B. Keller. Inverse Problems. *Am. Math. Mon., 83:107-118*, 1976

some examples ...

What are the questions to which the answer is:

1. Chicken Suzuki
2. Washington, Irving
3. Nine, W

from J.B. Keller. Inverse Problems. *Am. Math. Mon., 83:107-118*, 1976

some examples ...

What are the questions to which the answer is:

1 answer: Chicken Suzuki

1 question: What is the name of the only surviving kamikaze pilot?

2 answer: Washington, Irving

2 question: What is the capital of the United States, Max?

3 answer: Nine, W

3 question: Do you spell your name with a V, Herr Wagner?

from J.B. Keller. Inverse Problems. *Am. Math. Mon., 83:107-118*, 1976

some examples ...

What are the questions to which the answer is:

1 answer: Chicken Suzuki

1 question: What is the name of the only surviving kamikaze pilot?

2 answer: Washington, Irving

2 question: What is the capital of the United States, Max?

3 answer: Nine, W

3 question: Do you spell your name with a V, Herr Wagner?

from J.B. Keller. Inverse Problems. *Am. Math. Mon., 83:107-118*, 1976

some examples ...

What are the questions to which the answer is:

1 answer: Chicken Suzuki

1 question: What is the name of the only surviving kamikaze pilot?

2 answer: Washington, Irving

2 question: What is the capital of the United States, Max?

3 answer: Nine, W

3 question: Do you spell your name with a V, Herr Wagner?

from J.B. Keller. Inverse Problems. *Am. Math. Mon., 83:107-118*, 1976

some examples ...

What are the questions to which the answer is:

1 answer: Chicken Suzuki

1 question: What is the name of the only surviving kamikaze pilot?

2 answer: Washington, Irving

2 question: What is the capital of the United States, Max?

3 answer: Nine, W

3 question: Do you spell your name with a V, Herr Wagner?

from J.B. Keller. Inverse Problems. *Am. Math. Mon., 83:107-118*, 1976

some examples ...

What are the questions to which the answer is:

1 answer: Chicken Suzuki

1 question: What is the name of the only surviving kamikaze pilot?

2 answer: Washington, Irving

2 question: What is the capital of the United States, Max?

3 answer: Nine, W

3 question: Do you spell your name with a V, Herr Wagner?

from J.B. Keller. Inverse Problems. *Am. Math. Mon., 83:107-118*, 1976

some examples ...

What are the questions to which the answer is:

1 answer: Chicken Suzuki

1 question: What is the name of the only surviving kamikaze pilot?

2 answer: Washington, Irving

2 question: What is the capital of the United States, Max?

3 answer: Nine, W

3 question: Do you spell your name with a V, Herr Wagner?

from J.B. Keller. Inverse Problems. *Am. Math. Mon., 83:107-118*, 1976

Hadamard's definition of well-posedness

Definition

A problem is called *well-posed* if

1. there exists a solution to the problem (existence),
2. there is at most one solution to the problem (uniqueness),
3. the solution depends continuously on the data (stability).

Otherwise the problem is called ill-posed.

ill-posedness of inverse problems

In many cases one of two problems, which are inverse to each other, is ill-posed. In this case we call it the inverse problem and the other one the direct of forward problem. All inverse problems we will consider in the following are ill-posed.

discussion of the first two Hadamard criteria

- Existence of a solution to the inverse problem is clear if the data space is defined as set of solutions to the direct problem.
- A solution may fail to exist if the data are perturbed by noise. This problem will be addressed below.
- Uniqueness of a solution to an inverse problem is often not easy to show. Obviously, it is an important issue.
- If uniqueness is not guaranteed by the given data, then either additional data have to be observed or the set of admissible solutions has to be restricted using a-priori information on the solution. In other words, a remedy against non-uniqueness can be a reformulation of the problem.

discussion of the third Hadamard criterion

- Among the three Hadamard criteria, a failure to meet the third one is most delicate to deal with.
- In this case inevitable measurement and round-off errors can be amplified by an arbitrarily large factor and make a computed solution completely useless.
- Until the beginning of the last century it was generally believed that for natural problems the solution will always depend continuously on the data. ’natura non facit salti.’
- Only in the second half of the last century it was realized that a huge number of problems aring in science and technology are ill-posed in any reasonable mathematical setting.

ill-posedness in terms of operator equations

Suppose the inverse problem can be formulated as an operator equation

$$F(x) = y$$

where x denotes the unknown solution and y the given data.
Then the inverse problem is well-posed in the sense of Hadamard if

1. F is surjective (existence)
2. F in injective (uniqueness)
3. F^{-1} is continuous (stability)

Typically, the third condition is violated for inverse problems!

Outline

example 1: numerical differentiation

problem: Given a noisy signal $g : \mathbb{R} \to \mathbb{R}$ estimate its derivative.

A classical example where this appears is:
estimation of the density of a random variable

- Let $X_1, \ldots, X_n$ be independent copies of a random variable X with values in $[0, 1]$ and unknown density f.
- We can estimate the distribution function

$$g(x) := \int_0^x f(t)\,\mathrm{d}t = \mathbf{P}(X \leq x)$$

of f from our data by

$$g_n(x) := \frac{1}{n}\#\{i : X_i <= x\}\,.$$

operator formulation

We define the forward problem to be the evaluation of the integral

$$(T_{\mathrm{D}} f)(x) := \int_0^x f(t)\,\mathrm{d}t \qquad \text{for } x \in [0,1]$$

for a given $f \in C([0,1])$. The inverse problem consists in solving the equation

$$T_{\mathrm{D}} f = g$$

for a given $g \in C([0,1])$ satisfying $g(0) = 0$, or equivalently computing $f = g'$.

ill-posedness of numerical differentiation

- Obviously, the equation $T_{\mathrm{D}} f = g$ has a solution f in $C([0,1])$ if and only if $g \in C^1([0,1])$.
- The inverse problem is well-posed if we choose the norm $\|g\|_\infty + \|g'\|_\infty$ in the image space.
- However, the error in our data is only bounded with respect to the supremum norm. ´

ill-posedness of numerical differentiation

Let us assume that we are given noisy data $g^\delta \in C([0,1])$ satisfying

$$\|g^\delta - g\|_\infty \leq \delta$$

with noise level $0 < \delta < 1$. The functions

$$g_n^\delta(x) := g(x) + \delta \sin \frac{nx}{\delta}, \qquad x \in [0,1],$$

$n = 2, 3, 4, \ldots$ satisfy this error bound, but for the derivates

$$(g_n^\delta)'(x) = g'(x) + n \cos \frac{nx}{\delta}, \qquad x \in [0,1]$$

we find that

$$\|(g_n^\delta)' - g'\|_\infty = n.$$

Hence, the error in the solutions tends blows up without bound as $n \to \infty$ although the error in the data is bounded by δ. This shows the ill-posedness of the problem.

example 2: backwards heat equation

forward problem: Given $f \in L^2([0,1])$ find $g(x) = u(x,T)$ $(T > 0)$ where $u : [0,1] \times [0,T] \to \mathbb{R}$ satisfies

$$\begin{aligned} \frac{\partial}{\partial t} u(x,t) &= \Delta u(x,t), && x \in (0,1), t \in (0,T), \\ u(0,t) = u(1,t) &= 0, && t \in (0,T], \\ u(x,0) &= f(x), && x \in [0,1]. \end{aligned}$$

Physical interpretation: f may describe a temperature profile at time $t = 0$. On the boundaries of the interval $[0,1]$ the temperature is kept at 0. The task is to find the temperature at time $t = T$.

backwards heat equation: inverse problem

The inverse problem consists in finding the initial temperature given the temperature at time $t = T$.

inverse problem: Given $g \in L^2([0, 1])$, find initial values $f \in L^2([0, 1])$ such that the corresponding solution u to the direct problem satisfies $u(\cdot, T) = g$.

separation of variables

Let $f_n := \sqrt{2}\int_0^1 \sin(\pi n x) f(x)\,\mathrm{d}x$ denote the Fourier coefficients of f with respect to the complete orthonormal system $\{\sqrt{2}\sin(\pi n \cdot) : n = 1, 2, \dots\}$ of $L^2([0,1])$. A separation of variables leads to the formal solution

$$u(x,t) = \sqrt{2}\sum_{n=1}^{\infty} f_n e^{-\pi^2 n^2 t}\sin(n\pi x).$$

Introducing the operator $T_{\mathrm{BH}} : L^2([0,1]) \to L^2([0,1])$ by

$$(T_{\mathrm{BH}} f)(x) := \int 2\sum_{n=1}^{\infty}\left(e^{-\pi^2 n^2 T}\sin(n\pi x)\sin(n\pi y)\right) f(y)\,\mathrm{d}y,$$

we may formulate the inverse problem as an integral equation of the first kind

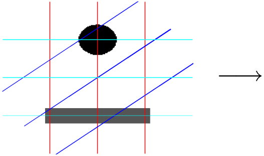
$$T_{\mathrm{BH}} f = g.$$

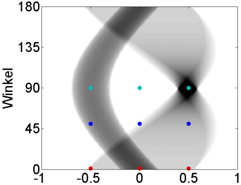
severe ill-posedness of the inverse problem

- The direct solution operator T_{BH} damps out high frequency components with an exponentially descreasing factor $e^{-\pi^2 n^2 T}$.
- Therefore, in the inverse problem a data error in the nth Fourier component of g is amplified by the factor $e^{\pi^2 n^2 T}$! This shows that the inverse problem is severely ill-posed.
- The inverse problem does not have a solution for arbitrary $g \in L^2([0,1])$.

example 3: computerized tomography

forward problem: Given the absorpion coefficient for x-rays in a slice of the body, compute all shadow images by integration!





inverse problem: Given all shadow images, find the absorption coefficient!

Radon transform

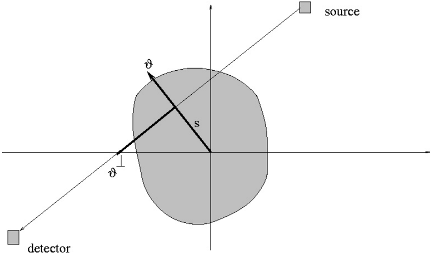
- Mathematically the forward problem of computarized tomography is described by the following Radon transform

$$R : L^2(B) \to L^2(S^1 \times [-1,1])$$

with $B := \{x \in \mathbb{R}^2 : |x| \leq 1$ and $S^1 := \partial B$.

- For a direction $\vartheta \in S^1$ define an orthogonal direction by $\vartheta^\perp := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \vartheta$, extend $f \in L^2(B)$ by 0 to $\mathbb{R}^2$, and define

$$(Rf)(\vartheta, s) := \int_{-\infty}^{\infty} f(s\vartheta + t\vartheta^\perp)\,\mathrm{d}t.$$

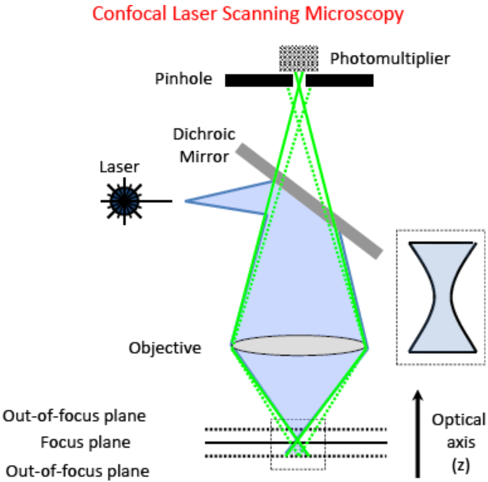


Johann Radon, 1887–1956

example 4: decovolution problems in imaging

confocal fluorescence microscopy

- Focused light excites fluorescent markers in the object.
- 3d imaging of livinig cells



the Hubble space telescope

- In early 1990 the Hubble Space Telescope was launched into the low-earth orbit outside of the disturbing atmosphere in order to provide images with a unprecedented spatial resolution.
- Unfortunately, soon after launch a manufacturing error in the main mirror was detected, causing severe spherical aberrations in the images.
- Therefore, before the space shuttle Endeavour visited the telescope in 1993 to fix the error, astronomers employed inverse problem techniques to improve the blurred images.

deblurring problem

- The true image f and the blurred image g are related by a first kind integral equation

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(x,y;x',y')f(x',y')\,\mathrm{d}x'\,\mathrm{d}y' = g(x,y)$$

  where k is the blurring function.
- $k(\cdot;x_0,y_0)$ describes the blurred image of a point source at (x_0,y_0).
- It is usually assumed that k is spatially invariant, i.e.

$$k(x,y;x',y') = h(x-x',y-y'), \qquad x,x',y,y' \in \mathbb{R}.$$

  h is called *point spread function*.

operator equation

- For a spatially invariant psf the direct problem is described by the convolution operator

$$(T_{DB}f)(x,y) := \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-x', y-y')f(x',y')\,\mathrm{d}x'\,\mathrm{d}y'.$$

- The inverse problem is then described by the operator equation

$$T_{DB}f = g\,.$$

- In principle the solution can be computed by the Fourier convolution theorem:

$$f = \frac{1}{2\pi}\mathcal{F}^{-1}(1/\hat{h})\mathcal{F}g.$$

- The multiplication by $1/\hat{h}$ is unstable since $\hat{h} := \mathcal{F}h$ vanishes asymptotically for large arguments. Therefore the inverse problem is ill-posed.

example 5: inverse scattering problems

acoustic waves:

$c =$ speed of sound

$U(t, x) =$ velocity potential:

$-\frac{\partial U}{\partial t} =$ pressure

$\mathrm{grad}\, U =$ density $\cdot$ velocity

$\omega =$ frequency

$k = \omega / c =$ wave number

- wave equation: $\frac{\partial^2 U}{\partial t^2} = c^2 \Delta U$
- time-harmonic waves: $U(x, t) = \Re\left(u(x) e^{-i\omega t}\right)$
- Helmholtz equation: $\Delta u + k^2 u = 0$
- boundary conditions:
  - sound-soft obstacles Ω: pressure vanishes at the boundary, i.e. $u = 0$ on $\partial\Omega$
  - sound-hard obstacles Ω: normal component of velocity vanishes at the boundary, i.e $\frac{\partial u}{\partial \nu} = 0$ on $\partial\Omega$.

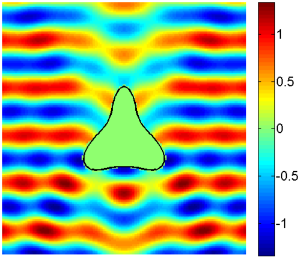
forward problem

$\Omega \subset \mathbb{R}^m$ compact obstacle, $\mathbb{R}^m \setminus \Omega$ connected
$u_i(x) = e^{-ikx\cdot d}$ incident plane wave with direction $|d| = 1$
$k \approx (\mathrm{diam}\,\Omega)^{-1}$ wave number

Given the domain Ω and the incident field u_i find scattered field u_s such that the total field $u := u_i + u_s$ satisfies

1. the Helmholtz equation $\Delta u + k^2 u = 0$ in $\mathbb{R}^m \setminus \Omega$
2. the Sommerfeld radiation condtion $r^{\frac{m-1}{2}} \left(\frac{\partial u_s}{\partial r} - iku_s\right) \overset{r\to\infty}{\longrightarrow} 0$
3. either the boundary condition $u = 0$ or $\frac{\partial u}{\partial \nu} = 0$ on $\partial\Omega$

total field for Neumann boundary condition
and incident wave from below



inverse obstacle scattering problem

- The scattered field u_s has the asymptotic behavior

$$u_s(x) = \frac{e^{ik|x|}}{|x|^{\frac{m-1}{2}}} \left(u_\infty \left(\frac{x}{|x|} \right) + O\left(\frac{1}{|x|} \right) \right), \qquad |x| \to \infty .$$

  The amplitude factor u_∞, which is an analytic function of the direction $\frac{x}{|x|}$ is called the far field pattern or scattering amplitude of u_s.
- Given far field patterns u_∞ corresponding to one or many incident waves u_i, find the shape of the scatterer Ω!

example 6: Electrical Impedance Tomography (EIT)

Let $\Omega \subset \mathbb{R}^m$ describe an electrically conducting medium.

u	voltage	$\sigma > 0$	conductivity
$E = -\operatorname{grad} u$	electric field	j	electric current
H	magnetic field	$I = \sigma \frac{\partial u}{\partial \nu}$	boundary current

- derivation of partial differntial equation in Ω:
  Ohm's law: $j = \sigma E = -\sigma \operatorname{grad} u$
  Maxwell eq.: $\operatorname{rot} H = j + \partial_t(\epsilon E)$
  Apply div and use identity $\operatorname{div}\operatorname{rot} = 0$ and assumption $\partial_t E = 0$:
  $$0 = \operatorname{div} j = -\operatorname{div} \sigma \operatorname{grad} u \qquad \text{in } \Omega.$$
- conservation of charge (Gauß law): $\int_{\partial\Omega} I \,\mathrm{d}s = 0$
- Since the voltage is only determined up to a constant, we can normalize u by $\int_{\partial\Omega} u \,\mathrm{d}s = 0$.

Electrical Impedance Tomography (EIT)

Direct problem: Given σ, determine the voltage $u|_{\partial\Omega}$ on the boundary for all current distributions I satisfying the compatibility condition $\int_{\partial\Omega} I \,\mathrm{d}s = 0$ by solving the elliptic boundary value problem

$$-\operatorname{div} \sigma \operatorname{grad} u = 0 \qquad in \Omega,$$
$$\sigma \frac{\partial u}{\partial \nu} = I \qquad on \partial\Omega$$

with the normalization condition $\int_{\partial\Omega} u \,\mathrm{d}s = 0$. In other words, determine the Neumann-to-Dirichlet map $\Lambda_\sigma : H_\diamond^{-1/2}(\partial\Omega) \to H_\diamond^{1/2}(\partial\Omega)$ defined by $\Lambda_\sigma I := u|_{\partial\Omega}$.

Inverse problem: Given measurements of the voltage distributation $u|_{\partial\Omega}$ for all current distributations I, i.e. given the Neumann-to-Dirichlet map Λ_σ, reconstruct the conductivity σ.

Outline

singular value decomposition (SVD)

- $\mathcal{X}, \mathcal{Y}$ Hilbert spaces
- $T \in L(\mathcal{X}, \mathcal{Y})$ compact, $\dim T(\mathcal{X}) = \infty$
- P orthogonal projection onto $N(T) := \{f \in \mathcal{X} : Tf = 0\}$.

Theorem

There exist singular values $\sigma_0 \geq \sigma_1 \geq \cdots > 0$ and orthonormal systems $\{u_0, u_1, \ldots\} \subset \mathcal{X}$ and $\{v_0, v_1, \ldots\} \subset \mathcal{Y}$ such that

$$f = \sum_{n=0}^{\infty} \langle f, u_n \rangle u_n + Pf,$$

$$Tf = \sum_{n=0}^{\infty} \sigma_n \langle f, u_n \rangle v_n$$

for all $f \in \mathcal{X}$. The σ_n are uniquely determined by T and satisfy $\lim_{n\to\infty} \sigma_n = 0$. $\{(\sigma_n, u_n, v_n)\}$ is called a singular system of T.

alternative formulation of the SVD

- Assume for simplicity that $N(T) = \{0\}$ and $\overline{T(\mathcal{X})} = \mathcal{Y}$.
- Then $\{u_n : n \in \mathbb{N}\} \subset \mathcal{X}$ and $\{v_n : n \in \mathbb{N}\} \subset \mathcal{Y}$ are Hilbert bases.
- Denote by $U : \mathcal{X} \to l^2(\mathbb{N}_0)$, $Uf := \sum_{n=0}^{\infty} \langle f, u_n \rangle \, u_n$ and $V : \mathcal{Y} \to l^2(\mathbb{N})$, $Vg := \sum_{n=0}^{\infty} \langle g, v_n \rangle \, v_n$ the corresponding unitary operators, and $\Sigma : l^2(\mathbb{N}) \to l^2(\mathbb{N})$, $(\Sigma a)_n := \sigma_n a_n$.
- Then the following diagram commutes:

$$
\begin{array}{ccc}
\mathcal{X} & \xrightarrow{T} & \mathcal{Y} \\
U \big\downarrow & & V \big\downarrow \\
l^2(\mathbb{N}) & \xrightarrow{\Sigma} & l^2(\mathbb{N})
\end{array}
$$

- In other words $T = V^* \Sigma U$. (This generalizes the matrix SVD.)

sequence model

- By the previous formulation of the SVD, every operator equation $Tf = g$ with a compact operator in Hilbert spaces can be reduced to a multiplication operator Σ on the sequence space $l^2(\mathbb{N})$.
- In other words, the operator equation $Tf = g$ decouples via the SVD into a countable number of scalar linear equations

$$\sigma_n f_n = g_n, \qquad n \in \mathbb{N}$$

  for the Fourier coefficients $f_n := \langle f, u_n \rangle$ and $g_n := \langle g, v_n \rangle$.
- Recall, however, that $\lim_{n\to\infty} \sigma_n = 0$. Therefore, $\left(\frac{g_n}{\sigma_n}\right)_{n\in\mathbb{N}}$ does not necessarily belong to $l^2(\mathbb{N})$ for $(g_n)_{n\in\mathbb{N}} \in l^2(\mathbb{N})$. This yields:

Picard criterion

Theorem (Picard)

Under the assumptions above the equation $Tf = g$ *is solvable if and only if the Picard criterion*

$$\sum_{n=0}^{\infty} \frac{1}{\sigma_n^2} \, |\langle g, v_n \rangle|^2 < \infty$$

is satisfied. Then the solution is given by

$$f = \sum_{n=0}^{\infty} \frac{1}{\sigma_n} \langle g, v_n \rangle \, u_n \, .$$

sequence model with noise

- Now assume that the data $g = Tf^\dagger$ are not given exactly, but only with some measurement errors err,

$$g^\delta = g + \mathrm{err}$$

with a known bound $\|\mathrm{err}\| \leq \delta$.

- Solving the equations $\sigma_n f_n^\delta = g_n^\delta$ yields

$$f_n^\delta = \frac{g_n^\delta}{\sigma_n} = f_n^\dagger + \frac{\mathrm{err}_n}{\sigma_n}.$$

- Since $\lim_{n\to\infty} \sigma_n = 0$, this shows that we have

infinite noise amplification!

truncated SVD (spectral cut-off)

- One remedy is to choose some cut-off parameter α and define

$$f_n^\delta := \begin{cases} \frac{g_n^\delta}{\sigma_n}, & \text{if } \sigma_n^2 > \alpha \\ 0 & \text{else.} \end{cases}$$

- This ensures that noise amplification is bounded by $\frac{1}{\sqrt{\alpha}}$.
- Disadvantage: The computation of the truncated SVD requires explicite knowledge of an SVD of T.
  This is only known in a few exceptional cases, and the numerical computation of an SVD is often prohibitively expensive.

Tikhonov regularization

- f_α^δ = miminum of the strictly convex, quadratic functional

$$J_\alpha(f) := \|Tf - g^\delta\|_{\mathcal{Y}}^2 + \alpha\|f - f_0\|_{\mathcal{X}}^2$$

  $\alpha > 0$ regularization parameter, $f_0 \in \mathcal{X}$ initial guess .

- First order optimality conditions: (Necessary and sufficient due to strict convexity) For all $h \in \mathcal{X}$ we have

$$0 = J'_\alpha[f_\alpha^\delta]h = 2\left\langle Tf_\alpha^\delta - g^\delta, Th\right\rangle_{\mathcal{Y}} + 2\alpha\left\langle f_\alpha^\delta - f_0, h\right\rangle_{\mathcal{X}} .$$

- explicit formula:

$$f_\alpha^\delta = (T^*T + \alpha I)^{-1}\left(T^*g^\delta + \alpha f_0\right)$$

iterated Tikhonov regularization

- Once we have computed the Tikhonov solution f_α^δ we may find a better approximation by applying Tikhonov regularization again using f_α^δ as initial guess f_0.
- This leads to iterated Tikhonov regularization:

$$
\begin{aligned}
f_{\alpha,0}^\delta &:= 0 \\
f_{\alpha,m}^\delta &:= (T^*T + \alpha I)^{-1}(T^*g^\delta + \alpha f_{\alpha,m-1}^\delta), \qquad m \geq 1
\end{aligned}
$$

- Note that only one operator $T^*T + \alpha I$ has to be inverted to compute $f_{\alpha,n}^\delta$ for any $n \in \mathbb{N}$. If we use, e.g., the *LU* factorization to apply $(T^*T + \alpha I)^{-1}$, the computation of $f_{\alpha,n}^\delta$ for $n \geq 2$ is not much more expensive than the computation of $f_{\alpha,1}^\delta$.

Landweber iteration

- Introduce the output-least-squares functional

$$J(f) := \| Tf - g^\delta \|_{\mathcal{Y}}^2$$

- The direction of steepest descent is given at f is given by $-\mathrm{grad} J(f) = -T^*(Tf - g^\delta)$.
- This leads to the following iteration formula for minimizing J known as Landweber iteration:

$$f_{n+1}^\delta := f_n^\delta - \mu T^*(Tf - g^\delta)$$

$\mu > 0$ is a step-length parameter.

- If $f_0 = 0$, an induction argument shows that

$$f_n^\delta = \sum_{j=0}^{n-1} (I - \mu T^* T)^j \mu T^* g^\delta \, .$$

functional calculus

- Let $A : \mathcal{X} \to \mathcal{X}$ be compact and self-adjoint, and let $\{u_n : n \in \mathbb{N}\} \subset \mathcal{X}$ be a Hilbert basis of eigenvectors with

$$Au_n = \lambda_n u_n \,.$$

- For a bounded function $\varphi : \sigma(A) \to \mathbb{R}$ on the spectum $\sigma(A) := \{\lambda_n : n \in \mathbb{N}\} \cap \{0\}$ define $\varphi(A) \in L(\mathcal{X})$ by

$$\varphi(A)f := \sum_{n=0}^{\infty} \varphi(\lambda_n) \langle f, u_n \rangle \, u_n$$

- Properties:
  - $p(A) = \sum_{j=0}^{m} p_j A^j$ if $p(t) = \sum_{j=0}^{m} p_j t^j$
  - $(\alpha\varphi + \beta\psi)(A) = \alpha\varphi(A) + \beta\psi(A)$ for $\alpha, \beta \in \mathbb{R}$
  - $(\varphi \cdot \psi)(A) = \varphi(A)\psi(A)$
  - $\|\varphi(A)\| \leq \sup_{t\in\sigma(A)} |\varphi(t)|$

spectral regularization methods

- Consider a family of functions $q_\alpha : \sigma(T^* T) \to \mathbb{R}$, $\alpha > 0$ such that
  - $\lim_{\alpha \to 0} q_\alpha(t) = \frac{1}{t}$ for all $t > 0$.
  - $\sup_{t \in \sigma(A)} |q_\alpha(t)| \le \frac{C_e}{\alpha}$ for all $\alpha > 0$.
- Define a regularized solution to $Tf = g^\delta$ by

$$f_\alpha^\delta := q_\alpha(T^* T) T^* g^\delta$$

- Examples:

regularization method	$q_\alpha(t)$
spectral cut-off	$\frac{1}{t}\chi_{[\alpha,\infty)}(t)$
Tikhonov regularization	$\frac{1}{\alpha+t}$
iterated Tikhonov	$\frac{(t+\alpha)^n-\alpha^n}{t(t+\alpha)^n}$
Landweber iteration $n = 1/\alpha$	$\sum_{j=0}^{n-1}(1-\mu t)^j$

estimation of the approximation error

- We study the error of the reconstruction $f_\alpha := q_\alpha(T^*T)T^*g$ for exact data $g = Tf^\dagger$.
- This so-called approximation error is given by

$$f^\dagger - f_\alpha = (I - q_\alpha(T^*T)T^*T)f^\dagger = r_\alpha(T^*T)f^\dagger$$

  where $r_\alpha(t) := 1 - tq_\alpha(t)$.
- Assumption: $|r_\alpha(t)| \leq C_r$ for all $\alpha, t \geq 0$
- Examples:

regularization method	$q_\alpha(t)$	$r_\alpha(t)$
spectral cut-off	$\frac{1}{t}\chi_{[\alpha,\infty)}(t)$	$\chi_{[0,\alpha)}(t)$
Tikhonov regularization	$\frac{1}{\alpha+t}$	$\frac{\alpha}{\alpha+t}$
iterated Tikhonov	$\frac{(t+\alpha)^n-\alpha^n}{t(t+\alpha)^n}$	$\left(\frac{\alpha}{\alpha+t}\right)^m$
Landweber iteration $n = 1/\alpha$	$\sum_{j=0}^{n-1}(1-\mu t)^j$	$(1-\mu t)^n$

estimation of the propagated data noise error

- The total error for noisy data can be estimated as follows:

$$\|f^\dagger - f_\alpha^\delta\| \leq \|f^\dagger - f_\alpha\| + \|f_\alpha - f_\alpha^\delta\|$$

- For the propagated data noise error we obtain the estimate

$$\begin{aligned}
\|f_\alpha - f_\alpha^\delta\|^2 &= \|q_\alpha(T^*T)T^*(g - g^\delta)\|^2 = \|T^*q_\alpha(TT^*)\mathrm{err}\|^2 \\
&= \langle TT^*q_\alpha(TT^*), q_\alpha(TT^*)\mathrm{err}\rangle \\
&\leq \|TT^*q_\alpha(TT^*)\| \, \|q_\alpha(TT^*)\| \|err\|^2 \\
&\leq \left(\sup_{t\geq 0} tq_\alpha(t)\right) \left(\sup_{t\geq 0} q_\alpha(t)\right) \delta^2 \\
&\leq \frac{(1 + C_r)C_e}{\alpha}\delta^2
\end{aligned}$$

Outline

regularization methods: notation

- We consider a family of *continuous* (not necessarily linear) operators $R_\alpha : \mathcal{Y} \to \mathcal{X}$ defined for α in some index set A which approximate the *unbounded* operator T^{-1}
- Let $\overline{\alpha} : (0, \infty) \times \mathcal{Y} \to A$ be a parameter choice rule. For a given noisy data $g^\delta \in \mathcal{Y}$ and noise level $\delta > 0$ such that $\|g^\delta - g\| \leq \delta$ the exact solution is approximated by

$$T^{-1} g \approx R_{\overline{\alpha}(\delta, g^\delta)} g^\delta.$$

Examples:

- Tikhonov regularization: $R_\alpha = (\alpha I + T^* T)^{-1} T^*$
- spectral cutoff: $R_\alpha g := \sum_{\{n : \sigma_n \geq \alpha\}} \frac{1}{\sigma_n} \langle g, g_n \rangle \varphi_n$

regularization methods: definition

Definition

- The pair $(R, \overline{\alpha})$ is called a regularization method for the problem $T\varphi = g$ if the *worst case error* tends to 0 with the noise level, i.e.

$$\sup \left\{ \left\| R_{\overline{\alpha}(\delta, g^\delta)} g^\delta - T^{-1} g \right\| : g^\delta \in \mathcal{Y}, \|g^\delta - g\| \le \delta \right\} \to 0 \qquad \text{as } \delta \to 0$$

  for all $g \in R(T)$.
- $\overline{\alpha}$ is called an a-priori parameter choice rule if $\overline{\alpha}(\delta, g^\delta)$ depends only on δ. Otherwise $\overline{\alpha}$ is called an a-posteriori parameter choice rule.

error decomposition

Let $\alpha \in A = \mathbb{R}$ and assume that R_α are linear operators with $R_\alpha g \to T^{-1} g$ as $\alpha \to 0$ for all $g \in R(T)$. Then the total error can be decomposed by the triangle inequality

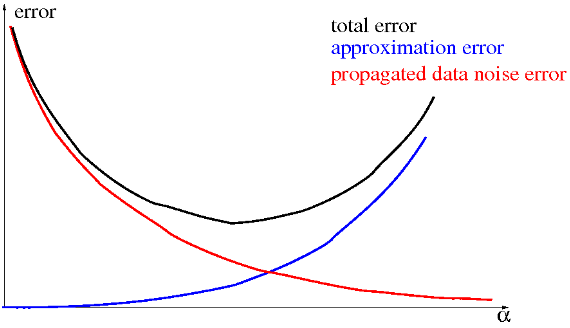
$$\|R_\alpha g^\delta - T^{-1} g\| \leq \|R_\alpha g^\delta - R_\alpha g\| + \|R_\alpha g - T^{-1} g\|$$

into

- a propagated data noise error $\|R_\alpha g^\delta - R_\alpha g\| \leq \delta \|R_\alpha\|$, which explodes as $\alpha \to 0$ and
- an approximation error $\|R_\alpha g - T^{-1} g\|$, which tends to 0 as $\alpha \to 0$

We have a trade-off between accuracy (small α) and stability (large α).

balancing the error components



Morozov's discrepancy principle

For a fixed parameter $\tau \geq 1$ choose the largest $\alpha > 0$ for which $\| Tf_\alpha^\delta - g^\delta \| \leq \delta$. Here $f_\alpha^\delta := R_\alpha g^\delta$ denotes the reconstruction for the regularization parameter α.

$$\alpha(\delta, g^\delta) := \sup\{\alpha > 0 : \| Tf_\alpha^\delta - g^\delta \| \leq \tau\delta\}$$

Do not try to fit the noise!

For iterative methods such as Landweber iteration, the discrepancy principle consists in stopping the iteration at the first index N for which $\| T\varphi_N^\delta - g^\delta \| \leq \tau\delta$.

numerical realization of the discrepancy principle

- For Tikhonov regularization and most other regularization methods the function $\alpha \mapsto \|Tf_\alpha^\delta - g^\delta\|$ is monotonely increasing. Therefore, $\overline{\alpha}(\delta, g^\delta)$ can be found by a simple bisection algorithm.
- Faster convergence can be achieved by Newton's method applied to $\beta := 1/\alpha$ as unknown. The function $f(\beta) := \|T\varphi_{1/\beta}^\delta - g^\delta\|^2$ turns out to be convex and monotonely decreasing. Therefore, Newton's method is globally convergent.

error-free parameter choice rules

Theorem (Bakushinskiĭ)

Let $T : \mathcal{X} \to \mathcal{Y}$ be one-to-one with dense range. Assume there exists a regularization method $(R_\alpha, \overline{\alpha})$ for $T\varphi = g$ with a parameter choice rule $\overline{\alpha}(\delta, g^\delta)$, which depends only on g^δ, but not on δ. Then T^{-1} is continuous.

We consider regularization methods $(R_\alpha, \overline{\alpha})$ which satisfy the following assumption:

$$R_\alpha : \mathcal{Y} \to \mathcal{X},\ \ \alpha \in A \subset (0,\infty) \text{ are linear operators and}$$
$$\lim_{\delta\to 0} \sup\left\{\overline{\alpha}(\delta, g^\delta) : g^\delta \in \mathcal{Y}, \|g^\delta - g\| \le \delta\right\} = 0.$$

Then R_α converges pointwise to T^{-1}:

$$\lim_{\alpha\to 0} R_\alpha g = T^{-1} g \qquad \text{for all } g \in R(T).$$

Theorem

Assume that T^{-1} is unbounded and the assumptions above hold true. Then

- *the operators R_α cannot be uniformly bounded with respect to α and*
- *the operators $R_\alpha T$ cannot be norm convergent to I as $\alpha \to 0$.*

arbitrarily slow convergence

Theorem

Assume that there exist a regularization method $(R_\alpha, \overline{\alpha})$ *for* $T\varphi = g$ *and a continuous function* $\varphi : [0, \infty) \to [0, \infty)$ *with* $\varphi(0) = 0$ *such that*

$$\sup \left\{ \left\| R_{\overline{\alpha}(\delta, g^\delta)} g^\delta - T^{-1} g \right\| : g^\delta \in \mathcal{Y}, \|g^\delta - g\| \leq \delta \right\} \leq \varphi(\delta)$$

for all $g \in R(T)$ *with* $\|g\| \leq 1$ *and all* $\delta > 0$*. Then* T^{-1} *is continuous.*