The program

Part I: Examples
- Image deblurring
- Tomography
- Radiative transfer inversion
- Helioseismology

Part II: Mainly direct methods
- Fourier transforms
- Singular value decomposition
- Backus-Gilbert or Mollifier method

Part III: Mainly Iterative methods
- Noise and a priori knowledge
- Iteration algorithms
- Regularization by Tikhonov
- Nonlinear problems
- Gencode and Neural networks

FT-inversion: Convolution kernels

Convolution problems of the type (including generalizations to higher dimensions, \( L \) is the size of the domain)

\[
f(x) = \int_{L} K(x - x') g(x') \, dx'
\]

can be solved in principle by Fourier transform. We define

\[
\hat{f}(k) = \frac{1}{L} \int_{L} e^{-ikx} f(x) \, dx, \quad f(x) = \sum_{k} e^{ikx} \hat{f}(k) \, dx, \quad \frac{kL}{2\pi} \in \mathbb{Z}
\]

An FT of the convolution type integral equation yields a simple algebraic relation between the respective Fourier coefficients

\[
\hat{f}(k) = L \hat{K}(k) \hat{g}(k)
\]

We therefore have a problem for which an analytic inversion formula exists

\[
g(x) = \sum_{k} e^{ikx} \frac{\hat{f}(k)}{\hat{K}(k)}
\]

Other inversion problems can be brought into the form of a convolution problem by means of variable transforms. The solar limb equation is an example for a kernel of the division type

\[
f(x) = \int_{a}^{b} K\left(\frac{x}{x'}\right) g(x') \, dx'
\]

Use \( x = e^{y}, \ x' = e^{y'}, \ dx' = e^{y'} dy' \) to obtain

\[
f(x) = \int_{\ln a}^{\ln b} K\left(\frac{e^{y}-y'}{y-y'}\right) g(e^{y'}) e^{y'} dy'
\]

Many other inverse problems have analytic inversion formulas.
FT-inversion: X-ray transform

We have noted that the X-ray transform is also close to a convolution type problem → we can also approach it by FT

X-ray transform: \( f(x, e_\theta) = \int_{-R}^{R} g(x + s e_\theta) \, ds \)

Insert the inverse 3D Fourier transform of \( g \)
\[
g(x) = \sum_{k} e^{i k \cdot x} \hat{g}(k), \quad k = \frac{\pi l}{R}, \ l \in \mathbb{Z}^3
\]

into the X-ray transform for fixed \( e_\theta \)
\[
\int_{-R}^{R} g(x + s e_\theta) \, ds = \sum_{k} \hat{g}(k) \int_{-R}^{R} e^{i k \cdot (x + s e_\theta)} \, ds \simeq 2 R \sum_{k \perp e_\theta} e^{i k \cdot x} \hat{g}(k) \frac{2 R e^{i k \cdot x} \text{sinc}(R k \cdot e_\theta)}{2 R e^{i k \cdot x} \text{sinc}(R k \cdot e_\theta)}
\]

This is exactly of the form of a 2D FT in the image plane.

Hence \( \hat{f}(k, e_\theta) = 2 R \hat{g}(k) \) for \( k \) in the plane \( \perp e_\theta \).

\[ \text{Illustration of the Fourier reconstruction of the X-ray transform} \]

FT-inversion: The noise problem

The practical usefulness of analytic inversion formulas is limited if the data is contaminated with noise
\[
f(x) = \int_{-L}^{L} K(x - x') g(x') \, dx' + \epsilon(x)
\]
then
\[
\hat{f}(k) = L \hat{K}(k) \hat{g}(k) + \hat{\epsilon}(k)
\]
The noise is assumed of zero mean mean and correlation length \( L_\epsilon \).
Then its Fourier coefficients \( \hat{\epsilon}(k) \) are random complex numbers of zero mean and variance \( (L_\epsilon/L^2) \int_L e^2 \, dx \) independent of \( k \) as long as \( k < 2\pi/L_\epsilon \).

For the application of the analytic inversion formula the noise is a desaster:
\[
g(x) + g_\epsilon(x) = \sum_{k} e^{i k \cdot x} \hat{f}(k) + \sum_{k} e^{i k \cdot x} \hat{\epsilon}(k)
\]

\[ \text{Power spectra of data } (f + \epsilon) \text{ (solid) and kernel } K \text{ (dashed) before and after the inversion, i.e., division by } K \]

\[ \text{Illustration of the Fourier inversion of the X-ray transform} \]
**FT-inversion: Truncated spectrum**

We have to limit the spectrum to wave numbers $k_{\text{min}} < k_{\text{max}}$ where $k_{\text{max}}$ is given by the intersection of the noise level with the kernel spectral power when both, data and kernel spectra are normalized to the same value at $k = 0$

$$
\frac{\hat{K}(0)}{K(k_{\text{max}})} = \frac{\hat{f}^2(0) + \hat{\varepsilon}^2(0)}{\hat{\varepsilon}^2(0)} = 1 + \text{SNR}
$$

The signal-to-noise ratio is here defined as

$$
\text{SNR} = \frac{\hat{f}^2(0)}{\hat{\varepsilon}^2(0)} = \frac{(1/L^2) \left( \int_L f \, dx \right)^2}{(L_\varepsilon/L^2) \int_L \varepsilon^2 \, dx}
$$

As an example: image deblurring with Gaussian kernel in 1D

$$
K(x) \simeq \exp \left( -\frac{1}{2} \left( \frac{x}{L_K} \right)^2 \right), \quad \hat{K}(k) \simeq \exp \left( -\frac{1}{2} (kL_K)^2 \right)
$$

then

$$
\ln \left( \frac{\hat{K}(0)}{K(k_{\text{max}})} \right) \simeq \frac{1}{2} (k_{\text{max}}L_K)^2 = 2\pi^2 \left( \frac{L_K}{L} \right)^2 l_{\text{max}}^2
$$

$l_{\text{max}}$ is the maximum number of complex Fourier coefficients of the reconstruction.

$\rightarrow$ the number of independent image parameters of the reconstruction (factor 2 because the Fourier coefficients are complex)

$$
2l_{\text{max}} \simeq \frac{L}{L_K} \sqrt{\ln(1 + \text{SNR})}
$$

where $L/L_K$ are the number of independent image parameters which have been measured. With the knowledge of the kernel we can enhance the number of independent image parameters depending on SNR.
FT-inversion: Example with more noise

Example of a reconstruction. The Gaussian kernel has just a width corresponding to the distance of the two peaks in the original signal. Noise variance is 0.05², the spectrum is truncated at 2/3 $k_{\text{max}}$.

SVD-inversion: The basics

We need a generalization of FT for more general inverse problems

$$f(x) = \int K(x, x') g(x') \, dx' \iff f = Kg, \quad g \in \mathbb{R}^m, \ f \in \mathbb{R}^n$$

The basic idea is to construct a symmetric matrix from $K$ which has a complete orthogonal set of eigenvectors and real eigenvalues

$$\begin{pmatrix} 0 & K \\ K^T & 0 \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} = \lambda_i \begin{pmatrix} u_i \\ v_i \end{pmatrix} \quad \text{or} \quad K v_i = \lambda_i u_i, \quad K^T u_i = \lambda_i v_i$$

There are $m$ vectors of the $v_i$ which orthogonally span the modelspace $\mathbb{R}^m$ and $n$ vectors of the $u_i$ which orthogonally span the data space $\mathbb{R}^n$ because

$$K^T K v_i = \lambda_i^2 v_i, \quad K K^T u_i = \lambda_i^2 u_i$$

→ For every nonzero $\lambda_i$ there is also a negative one. There are at most $i_{\text{zero}} = \min(n,m)$ pairs of nonzero eigenvalues $\pm \lambda_i$

→ The action of $K$ is completely described by its singular value decomposition (to be read as a dyad)

$$K = \sum_{i=1}^{i_{\text{zero}}} u_i \lambda_i v_i, \quad v_i \in \mathbb{R}^m, \ u_i \in \mathbb{R}^n$$

where $v_i$ and $u_i$ are normalized to unity and all $\lambda_i$ chosen positive and ordered so that $\lambda_1 > \lambda_2 > \ldots \lambda_{i_{\text{zero}}} > 0$.

- The decomposition $K \rightarrow \left\{ v_i, u_i, \lambda_i \right\}$ $i = 1, i_{\text{zero}}$ can be found numerically for $i_{\text{zero}} < \text{about 1000}$. 
SVD-inversion: The nullspace

In most cases \( i_{\text{zero}} < \) both \( m \) and \( n \).

If \( i_{\text{zero}} < m \) not all model features are mapped into data space. There are nonequal models \( g \) which cannot be distinguished by the observation operation of \( K \). The space spanned by \( v_i \) with \( \lambda_i = 0 \) is the nullspace \( \mathcal{N}(K) \) of the observation operator.

If \( i_{\text{zero}} < n \) the observations do not cover the total data space. There are inconsistent vectors \( f \) which can impossibly be the result of an observation through \( K \). The space spanned by \( u_i \) with \( \lambda_i \neq 0 \) is the range \( \mathcal{R}(K) \) of observation operator.

As a generalized inverse to \( K \) we define

\[
K_{\text{SVD}}^{-1} = \sum_{i=1}^{i_{\text{zero}}} \lambda_i^{-1} v_i u_i
\]

SVD-inversion: The noise problem

If noise is added, the data \( f + \epsilon \) is almost certainly inconsistent. We assume the noise \( \epsilon \) to have zero mean and variance \( \sigma_\epsilon^2 \).

\[
K_{\text{SVD}}^{-1} \text{ ignores the part of the noise which falls out of the range } \mathcal{R}(K).
\]

The eigenvalues close to zero are problematic:

\[
g + g_\epsilon = K_{\text{SVD}}^{-1}(f + \epsilon) = \sum_{i=1}^{i_{\text{zero}}} \lambda_i^{-1} v_i \left[ \frac{(u_i \cdot f)}{\lambda_i} + \frac{(u_i \cdot \epsilon)}{\lambda_i} \right]
\]

because \( (u_i \cdot \epsilon) \) are random real numbers with zero mean and variance \( \sigma_\epsilon^2 \) (the \( u_i \) are normalized).

\[\rightarrow\] We have the same problem as in Fourier inversion. Depending on the noise level, we have to truncate the spectrum of \( K_{\text{SVD}}^{-1} \) below mode \( i_{\text{max}} \) to be determined from

\[
\frac{\lambda_1}{\lambda_{i_{\text{max}}}} \approx \sqrt{\left(\frac{u_i \cdot f}{\lambda_i}\right)^2 + \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2}}
\]

---

Mapping of \( K \) and its generalized inverse \( K_{\text{SVD}}^{-1} \) between data and model space. The “visible” part of model space is \( \mathcal{N}(K) \), the orthogonal complement of the null space \( \mathcal{N}(K) \).

Normalized SVD spectra of data \( (f + \epsilon) \) (solid) and kernel \( K \) (dashed) before and after the inversion, i.e., division by \( \lambda_i \).
SVD-inversion: 2D tomography, model and data

The grid is cylindrical with \( \varphi \) as azimuth angle and \( \rho \) as distance.

Data grid and image of original model with noise. \( r \) denotes the pixel number, \( \theta \) the view direction.

SVD-inversion: 2D tomography, the kernel

Tomography kernel matrix. Each subblock shows \( K \) for fixed \( \varphi \) and \( \theta \). Zero elements are blank.

Spectrum of kernel (crosses) and data (circles)
SVD-inversion: 2D tomography, the eigenfunctions

Eigenfunctions $\psi_i$ of some modes of the 2D tomography kernel

SVD-inversion: 2D tomography, reconstructions

Reconstructions for various truncation levels $\lambda_{\text{trunc}}/\lambda_1$
For most problems exact inverses $K^{-1}$ does not exist.

$\rightarrow$ The concept of matrix inverses needs to be generalized. Generalized inverses $K_{\text{gen}}^{-1}$ are defined through the four Moore-Penrose criteria for generalized inverses:

Instead of being a unit matrix, $K_{\text{gen}}^{-1}K$ and $KK_{\text{gen}}^{-1}$ are only required to be symmetric

\begin{align*}
(K_{\text{gen}}^{-1}K)^T &= K_{\text{gen}}^{-1}K \quad \text{(model resolution kernel)} \\
(KK_{\text{gen}}^{-1})^T &= KK_{\text{gen}}^{-1} \quad \text{(data resolution matrix)}
\end{align*}

and that they act as unit matrix at least in the “visible” model subspace $\mathcal{N}(K) \subset \mathbb{R}^m$ and the range $\mathcal{R}(K) \subset \mathbb{R}^n$, respectively,

\begin{align*}
KK_{\text{gen}}^{-1}K &= K \\
K_{\text{gen}}^{-1}KK_{\text{gen}}^{-1} &= K_{\text{gen}}^{-1}
\end{align*}

We find that $KK_{\text{SVD}}^{-1}$ satisfies these criteria however its truncated version

\[
K_{\text{SVD}}^{-1} = \sum_{i=1}^{i_{\text{trunc}}} v_i^T v_i \frac{1}{\lambda_i}
\]

with $i_{\text{trunc}} < i_{\text{zero}}$ satisfies only the first two Moore-Penrose criteria, because

\begin{align*}
KK_{\text{SVD}}^{-1}K &= \sum_{i=1}^{i_{\text{trunc}}} v_i v_i^T, \\
KK_{\text{SVD}}^{-1} &= \sum_{i=1}^{i_{\text{trunc}}} u_i u_i
\end{align*}

are projection operators onto only part of $\mathcal{N}(K) \subset \mathbb{R}^m$ and $\mathcal{R}(K)$, respectively.

---

**BG or mollifier inversion**: **Motivation**

Assume we have a continuous model and a discrete number of observations, i.e., $g \in \text{Hilbert space}$ and $f \in \mathbb{R}^n$. For each individual measurement $i = 1, \ldots, n$ we have

\[
f_i = \int K_i(x') g(x') \, dx'
\]

The problem is hopelessly underdetermined and a conventional inverse of $K(x)$ can never be achieved.

$\rightarrow$ We only want to obtain an estimate of $g(x)$ which should be a more or less localized average. Since the problem is linear this estimate must be a linear combination of the data. For each $x$ find coefficients $q(x)$ with

\[
\bar{g}(x) = \sum_{i=1}^{n} q_i(x) f_i = \int \sum_{i=1}^{n} q_i(x) K_i(x') g(x') \, dx'
\]

The resolution kernel (compare to Moore-Penrose definition) here is a suitable linear superposition of the individual forward kernels $K_i$.

The model resolution kernel $R(x, x')$ should satisfy

- **Localization** within width $w$

  \[
  R_w(x, x') \rightarrow 0 \quad \text{for} \quad |x - x'| > w
  \]

  \[
  R_w(x, x') \rightarrow \delta(x - x') \quad \text{as} \quad w \rightarrow 0
  \]

- **Normalization**

  \[
  \int R_w(x, x') \, dx' = 1
  \]
BG or mollifier inversion: SVD and noise

Assume we were able to construct a SVD of the kernel functions

\[ K(x) = \sum_{i=1}^{i_{\text{zero}}} u_i \lambda_i v_i(x), \quad p \leq n \]

then the \( v_i(x) \) span \( \mathcal{N}(K)^\perp \) completely and the resolution kernel \( R(x, x') \) has a representation in this basis. For simplicity we construct \( R(x, x') \) so that it is diagonal:

\[ R(x, x') = \sum_{i=1}^{i_{\text{zero}}} v_i(x) r_i v_i(x') \]

then the equivalent inverse is

\[ K_{\text{moll}}^{-1}(x) = \sum_{i=1}^{i_{\text{zero}}} v_i(x) \frac{r_i}{\lambda_i} u_i \quad \text{with} \quad K_{\text{moll}}^{-1}(x) K = R(x, x') \]

\[ \rightarrow \text{no truncation as in TSVD but gentle roll-off due to filter co-efficients } r_i. \]

If the observations \( f_i \) are contaminated with noise \( \epsilon_i \) then the estimate \( \bar{g} \) becomes affected as well:

\[ \bar{g}(x) + \bar{g}_e(x) = \sum_{i=1}^{i_{\text{zero}}} q_i(x) f_i + \sum_{i=1}^{i_{\text{zero}}} q_i(x) \epsilon_i \]

If the the noise has zero mean and variance \( \sigma^2 \) the error \( \bar{g}_e \) of the estimate has zero mean and variance \( \sigma^2 |q|^2 \). \( \rightarrow \) To confine the error due to data noise we need as additional requirement:

- Shortest possible \( q \)

\[ \sum_{i=1}^{i_{\text{zero}}} q_i^2(x) \rightarrow \text{minimum} \]

BG or mollifier inversion: Mollification

For each \( x \) try to find coefficients \( q_i(x) \) so that \( R_w(x, x') \) comes close to a desired mollifier function \( \delta_w(x, x') \) with width \( w \), i.e., solve (usually by SVD)

\[ \sum_{i=1}^{n} q_i(x) K_i(x') = \delta_w(x, x') + \text{res}(x') \]

where \( \text{res}(x') \in \mathcal{N}(K) \) is the part of the mollifier which falls into the nullspace of \( K \). Tune width \( w \) so that the error \( \propto |q| \) does not exceed given bounds.

Disadvantages:

- The above equation has to be solved for every \( x \) at which an estimate \( \bar{g} \) is required. Note, however, that the above equation is much easier to solve than the original problem because there is no noise involved.
- The computational overhead is large unless symmetries of the system reduce the number of resolution kernels \( R_w(x, x') \) to be calculated.

Advantages:

- For every \( x \) we not only obtain an estimate \( \bar{g} \) of the model but also a resolution kernel \( R_w(x, x') \) telling us which region \( \bar{g}(x) \) is representative of. We also obtain an individual error estimate \( \sigma_e |q| \) for each \( \bar{g} \).
- There is no need to discretize the model space
- The resolution kernels \( R_w(x, x') \) can be used again with different data if the kernels \( K_i(x') \) have not changed
Mollifiers $\delta$ and resolution kernels $R_w$ (left) and kernel coefficients $q_i$ (right) for the kernels $K_i$ in the top diagram. Resolution kernels are derived for $x$ (height) = 55 and different width $w$. Resolution kernels and mollifiers are practically identical.

Mollifiers $\delta$ and resolution kernels $R_w$ (left) and kernel coefficients $q_i$ (right) for the kernels $K_i$ in the top diagram. Resolution kernels are derived for $x$ (height) = 55 and different width $w$. Mollifiers are exactly box-shape.
BG or mollifier inversion
Noise comparison

The amount of noise in the estimate $\bar{g}(x)$ is $\sigma_\epsilon |q|$. Here, $\sigma_\epsilon$ is the standard deviation of the noise in the data, and $|q|$ is the length of the kernel coefficient vector given below.

![Length of q vs width of the resolution kernel for a Gaussian and box-shape mollifier](image)

BG or mollifier inversion: Backus-Gilbert approach

We do not specify the shape of the resolution kernel but only try to concentrate its width around a given $x$ by minimizing

$$
\int (x - x')^2 R(x, x') \, dx' = \int (x - x')^2 \left( \sum_{i=1}^{n} q_i(x) K_i(x') \right)^2 \, dx
$$

$$
= \sum_{i,j=1}^{n} q_i(x) q_j(x) \int (x - x')^2 K_i(x') K_j(x') \, dx \equiv (q(x) \cdot W(x) q(x))
$$

This expression has to be minimized along with $\sigma_\epsilon^2 |q(x)|^2$ (noise reduction) under the normalization constraint

$$
1 = \int R(x, x') \, dx' = \sum_{i=1}^{n} q_i(x) \int K_i(x') \, dx \equiv (q(x) \cdot p)
$$

Using Lagrangian multipliers $\alpha$ and $\beta$, the coefficient vector $q(x)$ is determined by

$$
(q \cdot W q) + \alpha \sigma_\epsilon^2 (q \cdot q) + \beta [(q \cdot p) - 1] \longrightarrow \text{minimum}
$$

for known matrix $W$ and vector $p$.

The result is

$$
q = \frac{1}{(p \cdot W + \alpha \sigma_\epsilon^2 1)^{-1} p} \left[ W + \alpha \sigma_\epsilon^2 1 \right]^{-1} p
$$

which has to be solved for every $x$. The parameter $\alpha$ serves to balance resolution vs noise and stabilize the inversion of the $n \times n$ matrix $W + \alpha \sigma_\epsilon^2 1$.

- The resulting $R(x, x')$ is well concentrated around $x$ but yet may not be well centered on $x$. Therefore, an additional constraint is sometimes used to obtain well centered resolution kernels.
BG or mollifier inversion: Tomography

In tomography the index $i$ stands for pixel number $r$ and view direction $\theta$. In 2D:

$$f_{r,\theta}(x) = \int K_{r,\theta}(x') g(x') dx' , \quad \text{where} \quad x \in \mathbb{R}^2 , \quad r \in \mathbb{R}$$

and $K_{r,\theta}(x')$ is the beam from pixel $r$ into direction $e_{\theta}$

$$K_{r,\theta}(x') = \begin{cases} 1 & \text{if } x' \text{ inside the beam } (r, \theta) \\ 0 & \text{else} \end{cases}$$

The mollifier method seeks $\overline{g}(x) = \sum_{r,\theta} q_{r,\theta}(x) f_{r,\theta}$ with $q_{r,\theta}$ so that

$$R(x, x') = \sum_{r,\theta} q_{r,\theta}(x) K_{r,\theta}(x') \rightarrow \delta(x - x')$$

hence, for each $x$ find coefficients $q_{r,\theta}(x)$ so that the resulting superposition of beams approaches a $\delta$ function at $x$.

In filtered backprojection tomography the special choice is

$$q_{r,\theta}(x) = w_{r-r_x} K_{r,\theta}(x) \quad \text{where} \quad r_x \text{ so that } K_{r,\theta}(x) \neq 0$$

This gives a symmetric resolution kernel $R(x, x')$ and

$$\bar{g}(x) = \sum_{\theta} K_{r,\theta}(x) \sum_r w_{r-r_x} f_{r,\theta}$$

Conclusions: What is the problem with inverse problems?

Kernel functions are “smooth” in the sense (Riemann-Lebesgue)

$$\int K(x, x') \left\{ \frac{\cos k x'}{\sin k x'} \right\} dx' \rightarrow 0 \quad \text{as} \quad k \rightarrow \infty$$

$\rightarrow f$ insensitive to the short wavelength structure in $g$

$\rightarrow$ solving for $g$ is an ill-posed problem (Hamadard):

- $g$ is either not unique (nullspaces)
- $g$ changes discontinuously with $f$ (small eigenvalues of $K$)

What is the solution to the problem with inverse problems? Replace the original problem by a series of solvable problems:

$$f(x) = \int K_\gamma(x, x') g_\gamma(x') dx' \quad \text{with} \quad \lim_{\gamma \rightarrow 0} K_\gamma(x, x') = K(x, x')$$

and set $g = \lim_{\gamma \rightarrow 0} g_\gamma$. Examples for the regularization parameter:

$$\gamma = 1/k_{\text{trunc}} \quad \text{in FT inversion}$$

$$= 1/i_{\text{trunc}} \quad \text{in SVD inversion}$$

$$= \text{width} w \quad \text{in mollification}$$

$$= \alpha \quad \text{in Backus-Gilbert inversion}$$

In practical cases, however, we have to stop at a finite $\gamma$ due to noise. The key problems are:

- to find the optimum value $\gamma^*$ of $\gamma$,
- to understand which features of $g_\gamma$, will survive if we could let $\gamma \rightarrow 0$
- which features $\lim_{\gamma \rightarrow 0} g_\gamma$ might have which $g_\gamma$ does not have.
- which contribution from $\mathcal{N}(K)$ has to be added to $g_\gamma$. 