Inverse Problems in Space Physics

Bernd Inhester, May 2002

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

$$\hat{f}(k) = \frac{1}{L}\int_L \mathrm{e}^{-ikx} f(x)\, dx\ , \quad f(x) = \sum_k \mathrm{e}^{ikx}\hat{f}(k)\, dx\ , \quad \frac{kL}{2\pi} \in \mathbb{Z}$$

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 anlytic inversion formula exists

$$g(x) = \sum_k \mathrm{e}^{ikx} \frac{\hat{f}(k)}{\hat{K}(k)}$$

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

$$f(x) = \int_a^b K(\frac{x}{x'})\, g(x')\, dx'$$

Use $x = \mathrm{e}^y$, $x' = \mathrm{e}^{y'}$, $dx' = \mathrm{e}^{y'} dy'$ to obtain

$$f(x) = \int_{\ln a}^{\ln b} \underbrace{K(\mathrm{e}^{y-y'})}_{K'(y-y')}\ \underbrace{g(\mathrm{e}^{y'})\mathrm{e}^{y'}}_{g'(y')}\, dy'$$

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 $\longrightarrow$ we can also approach it by FT

$$\text{X-ray transform:} \quad f(\boldsymbol{x}, \boldsymbol{e}_\theta) = \int_{-R}^{R} g(\boldsymbol{x} + s\boldsymbol{e}_\theta)\, ds$$

Insert the inverse 3D Fourier transform of g

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{k}} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} \hat{g}(\boldsymbol{k}) \;, \quad \boldsymbol{k} = \frac{\pi}{R}\boldsymbol{l} \;,\; \boldsymbol{l} \in \mathbb{Z}^3$$

into the X-ray transform for fixed $\boldsymbol{e}_\theta$

$$\int_{-R}^{R} g(\boldsymbol{x} + s\boldsymbol{e}_\theta)\, ds = \sum_{\boldsymbol{k}} \hat{g}(\boldsymbol{k}) \underbrace{\int_{-R}^{R} \mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{x}+s\boldsymbol{e}_\theta)}\, ds}_{2R\,\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,\mathrm{sinc}(R\boldsymbol{k}\cdot\boldsymbol{e}_\theta)} \simeq 2R \sum_{\boldsymbol{k}\perp\boldsymbol{e}_\theta} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} \hat{g}(\boldsymbol{k})$$

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

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

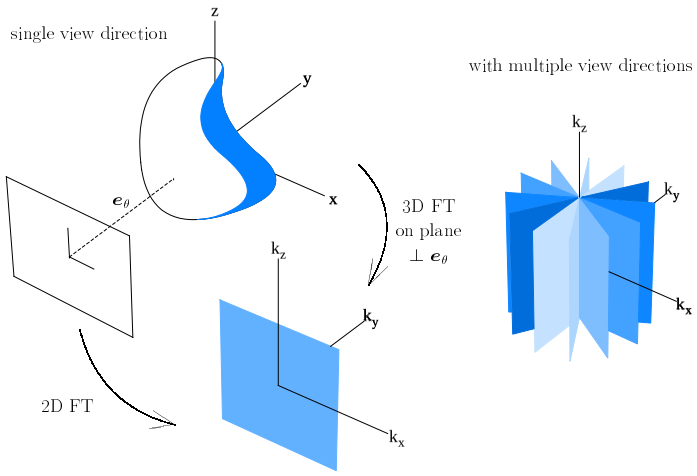


Illustration of the Fourier reconstruction of the X-ray transform

FT-inversion: The noise problem

The practical usefulnes of analytic inversion formulas is limited if the data is contaminated with noise

$$f(x) = \int_L K(x - x')\, g(x')\, dx' + \epsilon(x)$$

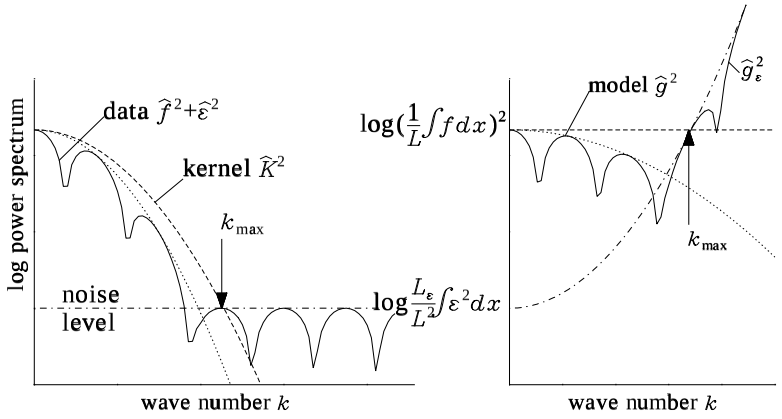
then

$$\hat{f}(k) = L\, \hat{K}(k)\, \hat{g}(k) + \hat{\epsilon}(k)$$

The noise is assumed of zero mean mean and correlation length L_ϵ. Then its Fourier coefficients $\hat{\epsilon}(k)$ are random complex numbers of zero mean and variance $(L_\epsilon/L^2) \int_L \epsilon^2 dx$ independent of k as long as k $< 2\pi/L_\epsilon$.

For the application of the analytic inversion formula the noise is a desaster:

$$g(x) + g_\epsilon(x) = \sum_k \mathrm{e}^{ikx} \frac{\hat{f}(k)}{\hat{K}(k)} + \sum_k \mathrm{e}^{ikx} \frac{\hat{\epsilon}(k)}{\hat{K}(k)}$$



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

FT-inversion: Truncated spectrum

We have to limit the spectrum to wave numbers $k_{\rm trunc} < k_{\rm max}$ where $k_{\rm max}$ 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

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

The signal-to-noise ratio is here defined as

$${\rm SNR} = \frac{\hat{f}^2(0)}{\hat{\epsilon}^2(0)} = \frac{(1/L^2)(\int_L f\,dx)^2}{(L_\epsilon/L^2)\int_L \epsilon^2\,dx}$$

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

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

then

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

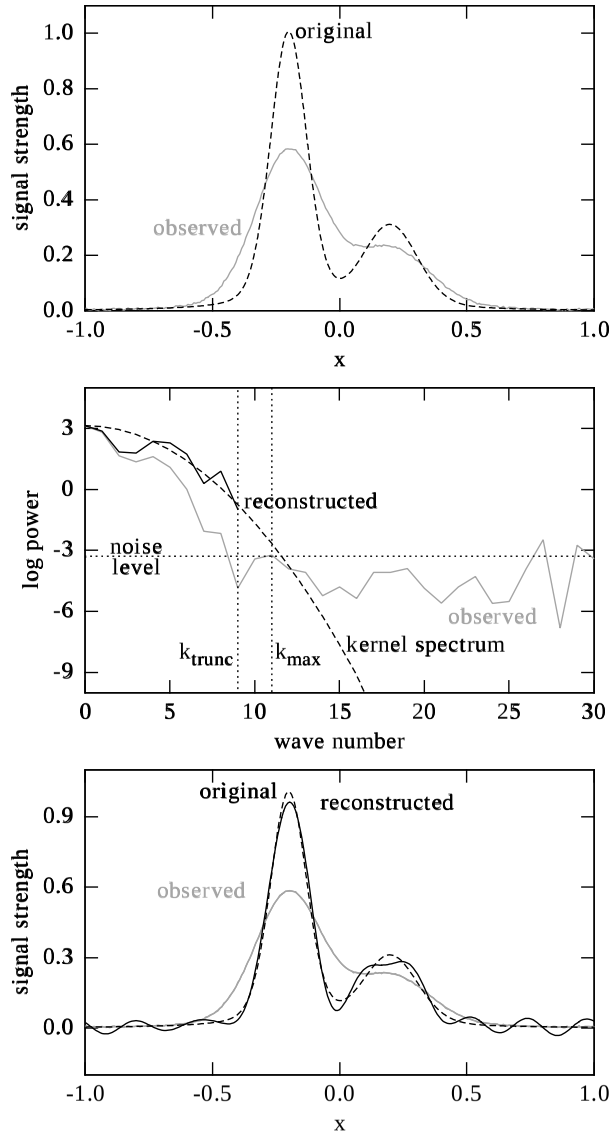
$l_{\rm max}$ is the maximum number of complex Fourier coefficients of the reconstruction.

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

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

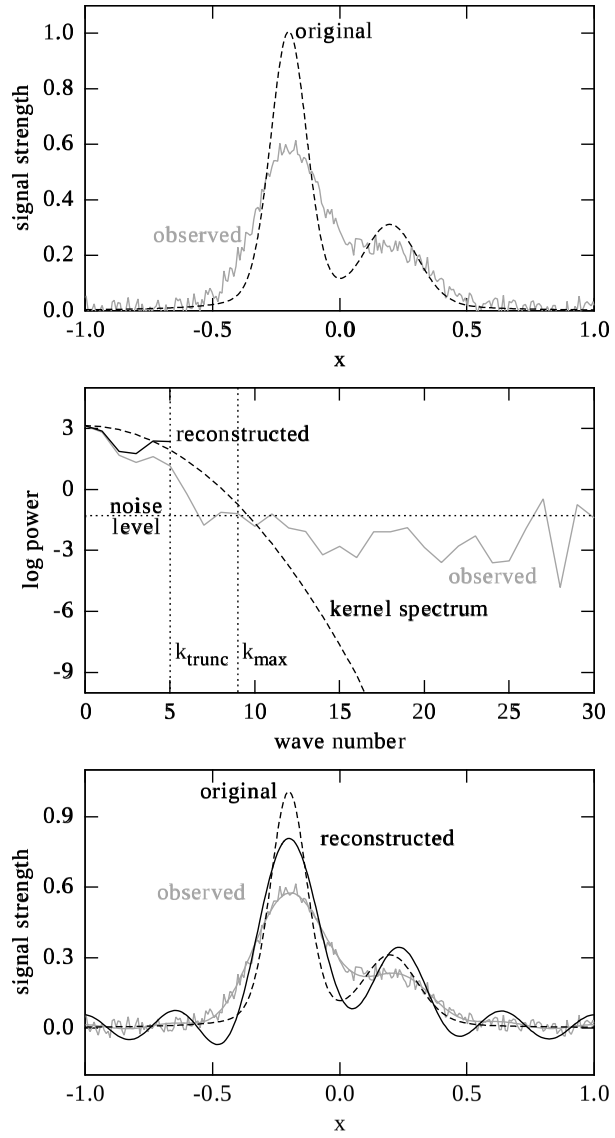
where L/L_K are the number of independent image parameters which have been measured. With the knowledge of the kernel we can enhance the number of independent image parameters depending on SNR.

FT-inversion: Example with little 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 0.005², the spectrum is truncated at 3/4 $k_{\rm max}$.

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 0.05^2, the spectrum is truncated at 2/3 $k_{\max}$.

SVD-inversion: The basics

We need a generalization of FT for more general inverse problems

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

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

$$\begin{pmatrix} 0 & \boxed{\boldsymbol{K}} \\ \boxed{\boldsymbol{K}^T} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix} = \lambda_i \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix} \quad \text{or} \quad \begin{matrix} \boldsymbol{K}\boldsymbol{v}_i = \lambda_i \boldsymbol{u}_i \\ \boldsymbol{K}^T\boldsymbol{u}_i = \lambda_i \boldsymbol{v}_i \end{matrix}$$

There are m vectors of the $\boldsymbol{v}_i$ which orthogonally span the modelspace $\mathbb{R}^m$ and n vectors of the $\boldsymbol{u}_i$ which orthogonally span the data space $\mathbb{R}^n$ because

$$\boldsymbol{K}^T\boldsymbol{K}\boldsymbol{v}_i = \lambda_i^2 \boldsymbol{v}_i\ , \quad \boldsymbol{K}\boldsymbol{K}^T\boldsymbol{u}_i = \lambda_i^2 \boldsymbol{u}_i$$

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

$\longrightarrow$ The action of $\boldsymbol{K}$ is completely described by its singular value decomposition (to be read as a dyad)

$$\boldsymbol{K} = \sum_{i=1}^{i_{\rm nzero}} \boldsymbol{u}_i \lambda_i \boldsymbol{v}_i\ , \quad \boldsymbol{v}_i \in \mathbb{R}^m\ ,\ \boldsymbol{u}_i \in \mathbb{R}^n$$

where $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$ are normalized to unity and all λ_i chosen positive and ordered so that $\lambda_1 > \lambda_2 > \ldots \lambda_{i_{\rm nzero}} > 0$.

• The decomposition $\boldsymbol{K} \rightarrow \{\boldsymbol{v}_i, \boldsymbol{u}_i, \lambda_i\}$ $i = 1, i_{\rm nzero}$ can be found numerically for $i_{\rm nzero} <$ about 1000.

SVD-inversion: The nullspace

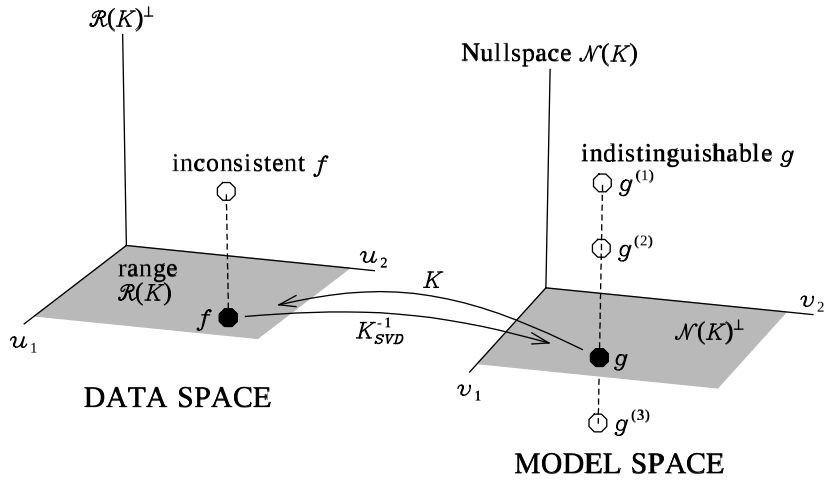
In most cases $i_{\mathrm{nzero}} <$ both m and n.

If $i_{\mathrm{nzero}} < m$ not all model features are mapped into data space. There are nonequal models $\boldsymbol{g}$ which cannot be distinguished by the observation operation of $\boldsymbol{K}$. The space spanned by $\boldsymbol{v}_i$ with λ_i = 0 is the nullspace $\mathcal{N}(\boldsymbol{K})$ of the observation operator.

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

As a generalized inverse to $\boldsymbol{K}$ we define

$$\boldsymbol{K}^{-1}_{\mathrm{SVD}} = \sum_{i=1}^{i_{\mathrm{nzero}}} \boldsymbol{v}_i \frac{1}{\lambda_i} \boldsymbol{u}_i$$



Mapping of $\boldsymbol{K}$ and its generalized inverse $\boldsymbol{K}^{-1}_{\mathrm{SVD}}$ between data and model space. The "visible" part of model space is $\mathcal{N}(\boldsymbol{K})^\perp$, the orthogonal complement of the null space $\mathcal{N}(\boldsymbol{K})$

SVD-inversion: The noise problem

If noise is added, the data $\boldsymbol{f} + \boldsymbol{\epsilon}$ is almost certainly inconsistent. We assume the noise $\boldsymbol{\epsilon}$ to have zero mean and variance σ^2_ϵ.

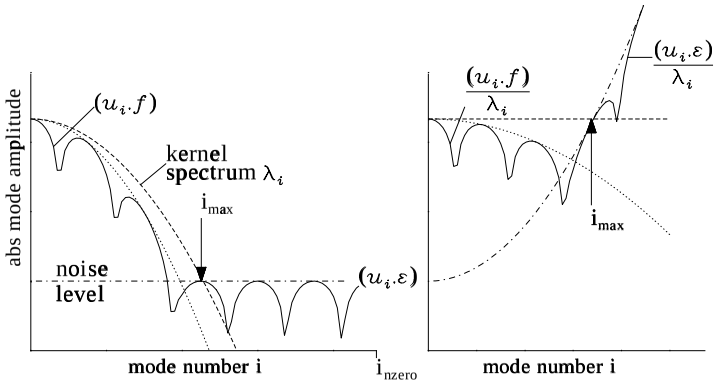
$\boldsymbol{K}^{-1}_{\mathrm{SVD}}$ ignores the part of the noise which falls out of the range $\mathcal{R}(\boldsymbol{K})$. Yet the eigenvalues close to zero are problematic:

$$\boldsymbol{g} + \boldsymbol{g}_\epsilon = \boldsymbol{K}^{-1}_{\mathrm{SVD}}(\boldsymbol{f} + \boldsymbol{\epsilon}) = \sum_{i=1}^{i_{\mathrm{nzero}}} \boldsymbol{v}_i \left[\frac{(\boldsymbol{u}_i \cdot \boldsymbol{f})}{\lambda_i} + \frac{(\boldsymbol{u}_i \cdot \boldsymbol{\epsilon})}{\lambda_i} \right]$$

because $(\boldsymbol{u}_i \cdot \boldsymbol{\epsilon})$ are random real numbers with zero mean and variance σ^2_ϵ (the $\boldsymbol{u}_i$ are normalized).

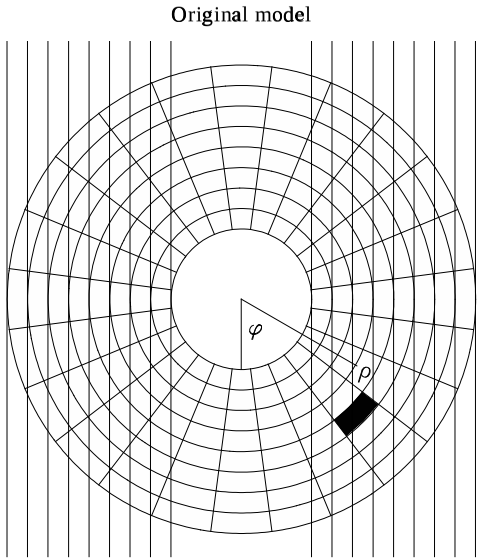
$\longrightarrow$ We have the same problem as in Fourier inversion. Depending on the noise level, we have to truncate the spectrum of $\boldsymbol{K}^{-1}_{\mathrm{SVD}}$ below mode i_{max} to be determined from

$$\frac{\lambda_1}{\lambda_{i_{\mathrm{max}}}} \simeq \frac{\sqrt{(\boldsymbol{u}_1 \cdot \boldsymbol{f})^2 + \sigma^2_\epsilon}}{\sigma_\epsilon}$$

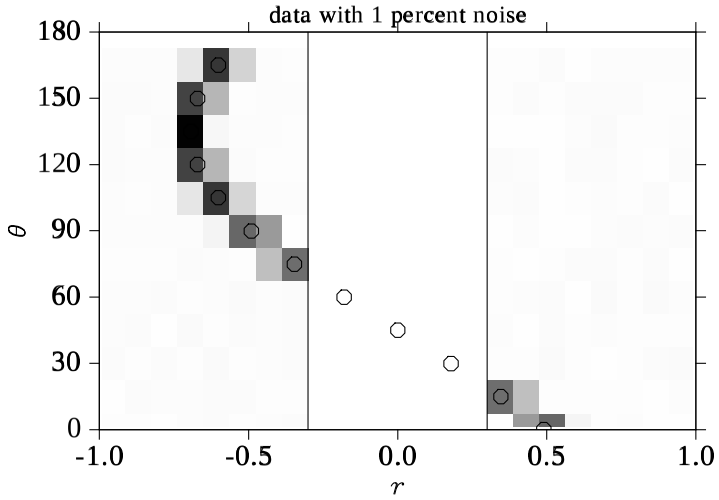


Normalized SVD spectra of data $(f+\epsilon)$ (solid) and kernel K (dashed) before and after the inversion, i.e., division by λ_i

SVD-inversion: 2D tomography, model and data

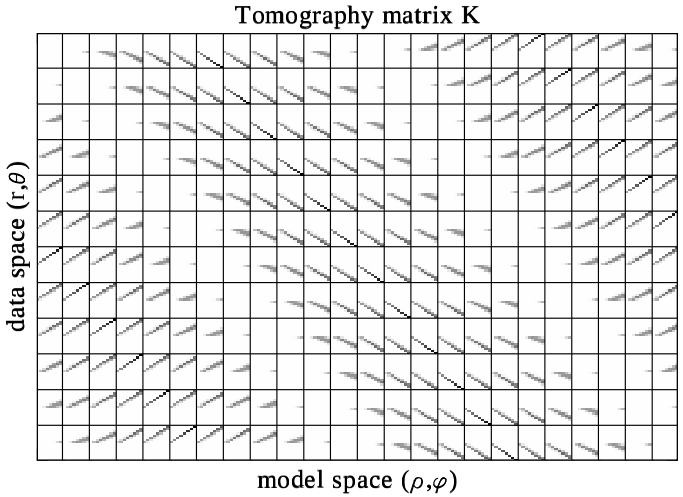


Tomography grid and model density. The grid is cylindrical with φ as azimuth angle and ρ as distance.

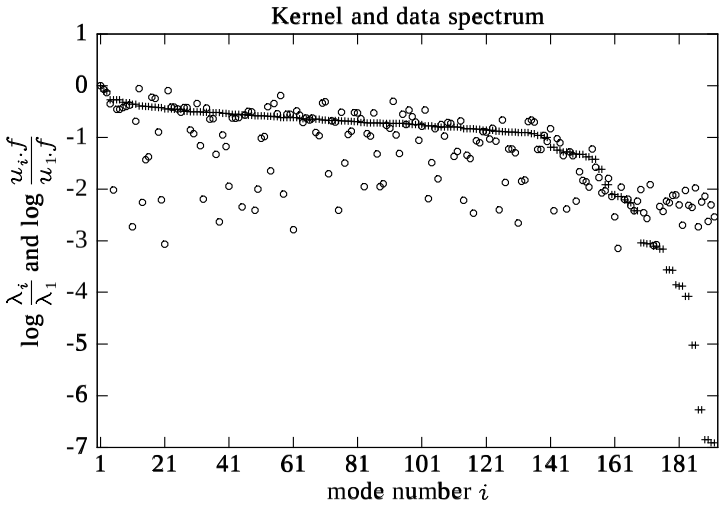


Data grid and image of original model with noise. r denotes the pixel number, θ the view direction.

SVD-inversion: 2D tomography, the kernel

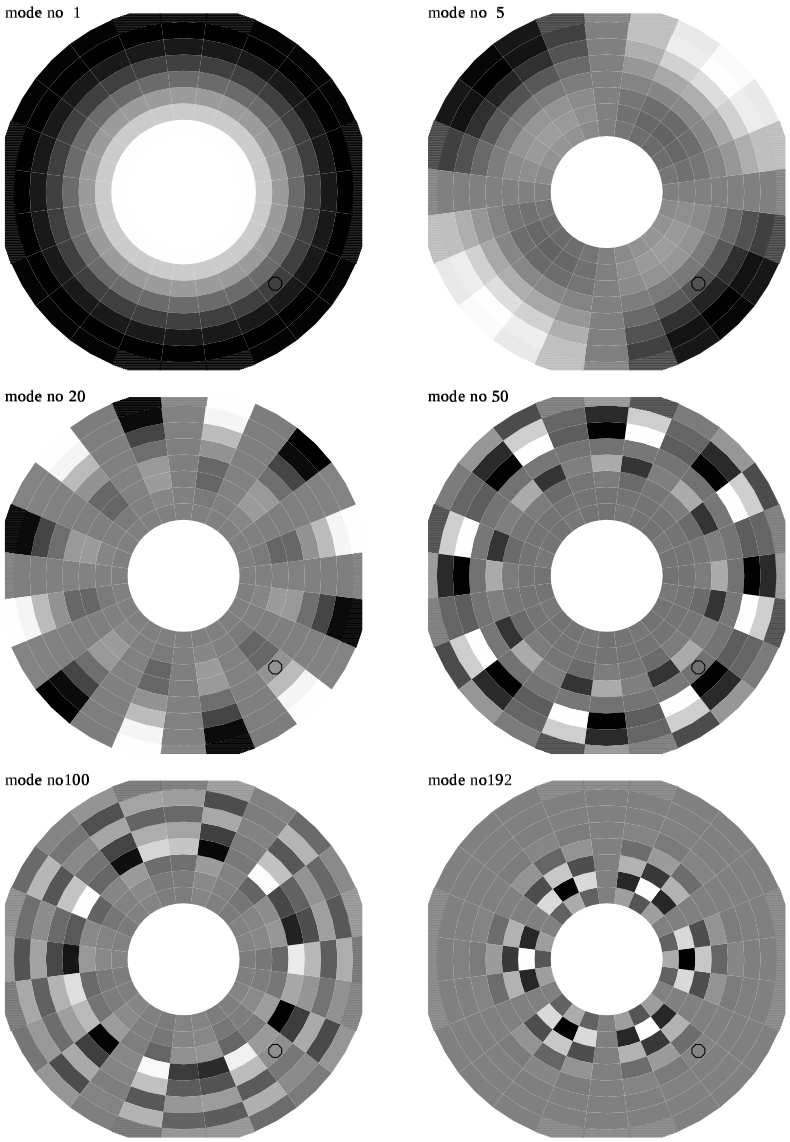


Tomography kernel matrix. Each subblock shows $\boldsymbol{K}$ for fixed φ and θ. Zero elements are blank.



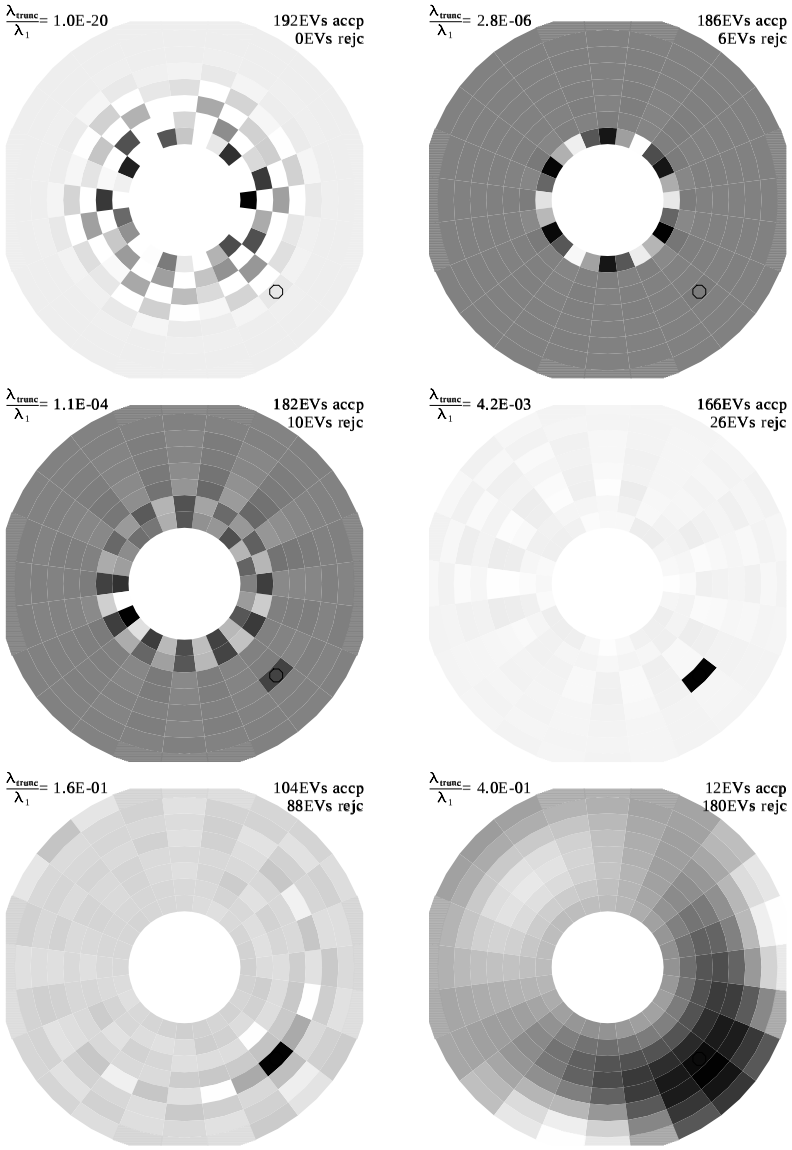
Spectrum of kernel (crosses) and data (circles)

SVD-inversion: 2D tomography, the eigenfunctions



Eigenfunctions $\boldsymbol{v}_i$ of some modes of the 2D tomography kernel

SVD-inversion: 2D tomography, reconstructions



Reconstructions for various truncation levels $\lambda_{\mathrm{trunc}}/\lambda_1$

SVD-inversion: Generalized inverses

For most problems exact inverses $\boldsymbol{K}^{-1}$ does not exist.
$\longrightarrow$ The concept of matrix inverses needs to be generalized. Generalized inverses $\boldsymbol{K}_{\rm gen}^{-1}$ are defined through the four Moore-Penrose criteria for generalized inverses :

Instead of being a unit matrix, $\boldsymbol{K}_{\rm gen}^{-1}\boldsymbol{K}$ and $\boldsymbol{K}\boldsymbol{K}_{\rm gen}^{-1}$ are only required to be symmetric

$$(\boldsymbol{K}_{\rm gen}^{-1}\boldsymbol{K})^T = \boldsymbol{K}_{\rm gen}^{-1}\boldsymbol{K} \qquad \text{(model resolution kernel)}$$
$$(\boldsymbol{K}\boldsymbol{K}_{\rm gen}^{-1})^T = \boldsymbol{K}\boldsymbol{K}_{\rm gen}^{-1} \qquad \text{(data resolution matrix)}$$

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

$$\boldsymbol{K}\boldsymbol{K}_{\rm gen}^{-1}\boldsymbol{K} = \boldsymbol{K}$$
$$\boldsymbol{K}_{\rm gen}^{-1}\boldsymbol{K}\boldsymbol{K}_{\rm gen}^{-1} = \boldsymbol{K}_{\rm gen}^{-1}$$

We find that $\boldsymbol{K}\boldsymbol{K}_{\rm SVD}^{-1}$ satisfies these criteria however its truncated version

$$\boldsymbol{K}_{\rm TSVD}^{-1} = \sum_{i=1}^{i_{\rm trunc}} \boldsymbol{v}_i \frac{1}{\lambda_i} \boldsymbol{u}_i$$

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

$$\boldsymbol{K}_{\rm TSVD}^{-1}\boldsymbol{K} = \sum_{i=1}^{i_{\rm trunc}} \boldsymbol{v}_i\boldsymbol{v}_i \ , \quad \boldsymbol{K}\boldsymbol{K}_{\rm TSVD}^{-1} = \sum_{i=1}^{i_{\rm trunc}} \boldsymbol{u}_i\boldsymbol{u}_i$$

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

BG or mollifier inversion: Motivation

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

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

The problem is hoplessly underdetermined and a conventional inverse of $\boldsymbol{K}(x)$ can never be achieved.

$\longrightarrow$ 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 $\boldsymbol{q}(x)$ with

$$\overline{g}(x) = \sum_{i=1}^{n} q_i(x)f_i = \int \underbrace{\sum_{i=1}^{n} q_i(x)K_i(x')}_{\text{model resolution kernel} \quad R(x,x')}\, g(x')\,dx'$$

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

$$R_w(x,x') \longrightarrow 0 \quad \text{for} \quad |x-x'| > w$$
$$R_w(x,x') \underset{w\to 0}{\longrightarrow} \delta(x-x')$$

• 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

$$\boldsymbol{K}(x) = \sum_{i=1}^{i_{\rm nzero}} \boldsymbol{u}_i \lambda_i v_i(x) \ , \quad p \leq n$$

then the $v_i(x)$ span $\mathcal{N}(\boldsymbol{K})^\perp$ 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:

$$R(x,x') = \sum_{i=1}^{i_{\rm nzero}} v_i(x) r_i v_i(x')$$

then the equivalent inverse is

$$\boldsymbol{K}^{-1}_{\rm moll}(x) = \sum_{i=1}^{i_{\rm nzero}} v_i(x) \frac{r_i}{\lambda_i} \boldsymbol{u}_i \quad \text{with} \quad \boldsymbol{K}^{-1}_{\rm moll}(x)\boldsymbol{K} = R(x,x')$$

$\longrightarrow$ no truncation as in TSVD but gentle roll-off due to filter coefficients r_i.

If the observations f_i are contaminated with noise ϵ_i then the estimate $\overline{g}$ becomes affected as well:

$$\overline{g}(x) + \overline{g}_\epsilon(x) = \sum_{i=1}^{i_{\rm nzero}} q_i(x)\, f_i + \sum_{i=1}^{i_{\rm nzero}} q_i(x)\, \epsilon_i$$

If the the noise has zero mean and variance σ_ϵ^2 the error $\overline{g}_\epsilon$ of the estimate has zero mean and variance $\sigma_\epsilon^2\ |\boldsymbol{q}|^2$. $\longrightarrow$ To confine the error due to data noise we need as additional requirement:

• Shortest possible $\boldsymbol{q}$

$$\sum_{i=1}^{i_{\rm nzero}} q_i^2(x) \rightarrow \quad \text{minimum}$$

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 $\overline{\delta}_w(x,x')$ with width w, i.e., solve (usually by SVD)

$$\sum_{i=1}^{n} q_i(x) K_i(x') = \overline{\delta}_w(x,x') + \mathrm{res}(x')$$

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

Disadvantages:

• The above equation has to be solved for every x at which an estimate $\overline{g}$ 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.

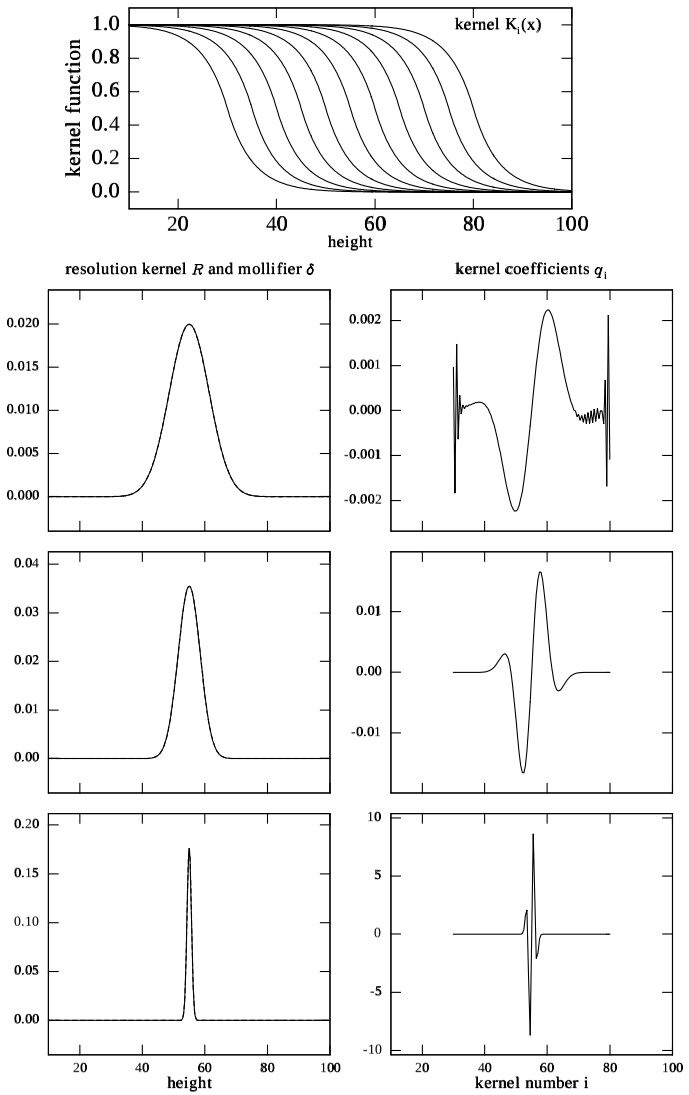
Advantages:

• For every x we not only obtain an estimate $\overline{g}$ of the model but also a resolution kernel $R_w(x,x')$ telling us which region $\overline{g}(x)$ is representative of. We also obtain an individual error estimate $\sigma_\epsilon|\boldsymbol{q}|$ for each $\overline{g}$.

• There is no need to discretize the model space

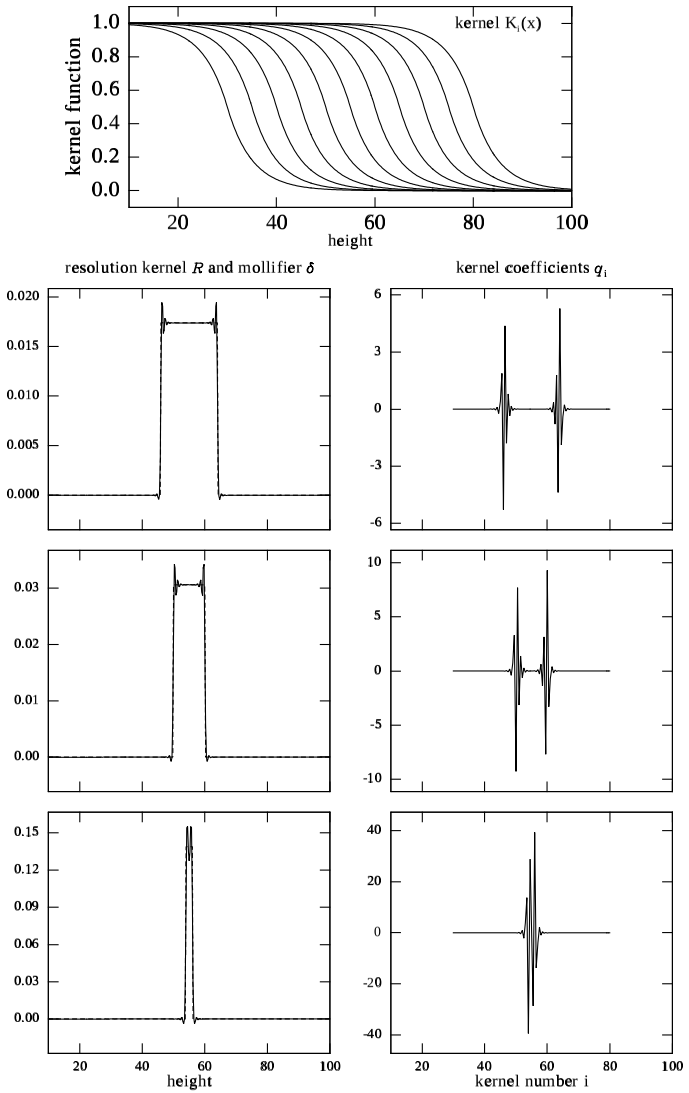
• The resolution kernels $R_w(x,x')$ can be used again with different data if the kernels $K_i(x')$ have not changed

BG or mollifier inversion: Gaussian mollifier



Mollifiers $\overline{\delta}$ 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.

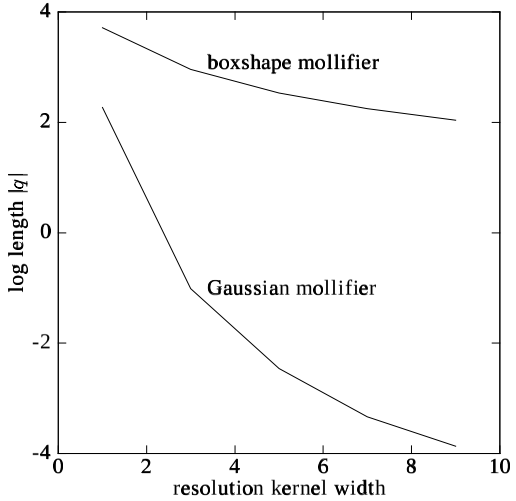
BG or mollifier inversion: Box-shape mollifier



Mollifiers $\overline{\delta}$ 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 $\overline{g}(x)$ is $\sigma_\epsilon|\boldsymbol{q}|$. Here, σ_ϵ is the standart deviation of the noise in the data, and $|\boldsymbol{q}|$ is the length of the kernel coefficient vector given below.



Length of $\boldsymbol{q}$ vs width w of the resolution kernel for a Gaussian and box-shape mollifier

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

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

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

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

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

Using Lagrangian multipliers α and β, the coefficient vector $\boldsymbol{q}(x)$ is determined by

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

for known matrix $\boldsymbol{W}$ and vector $\boldsymbol{p}$.

The result is $$\boldsymbol{q} = \frac{1}{(\boldsymbol{p}\cdot[\boldsymbol{W} + \alpha\sigma_\epsilon^2\mathbf{1}]^{-1}\boldsymbol{p})}[\boldsymbol{W} + \alpha\sigma_\epsilon^2\mathbf{1}]^{-1}\boldsymbol{p}$$

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

• 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 θ. In 2D:

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

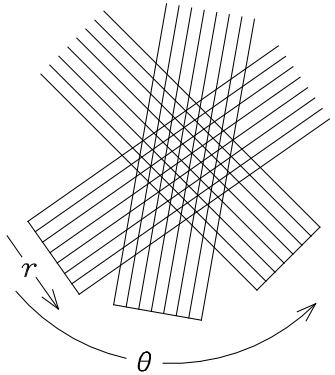
and $K_{(r,\theta)}(x')$ is the beam from pixel r into direction $\boldsymbol{e}_\theta$

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

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

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

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



In filtered backprojection tomography the special choice is

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

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

$$\overline{g}(x) = \sum_\theta \underbrace{K_{(r_x,\theta)}(x)}_{\text{backproj}} \underbrace{\sum_r w_{r-r_x} f_{(r,\theta)}}_{\text{filter}}$$

Conclusions: What is the problem with inverse problems ?

Kernel functions are "smooth" in the sense (Riemann-Lebesque)

$$\int K(x,x') \begin{Bmatrix} \cos kx' \\ \sin kx' \end{Bmatrix} dx' \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty$$

$\longrightarrow f$ insensitive to the short wavelength structure in g
$\longrightarrow$ 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 orginal problem by a series of solvable problems:

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

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

$$\begin{aligned} \gamma &= 1/k_{\text{trunc}} \quad \text{in FT inverstion} \\ &= 1/i_{\text{trunc}} \quad \text{in SVD inverstion} \\ &= \text{width } w \quad \text{in mollification} \\ &= \alpha \quad \text{in Backus-Gilbert inversion} \end{aligned}$$

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

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