Data Driven Inverse Problems
Lecture 1 : Introduction to Inverse Problems

Simon R. Arridge[1]

[1] Department of Computer Science, University College London, UK

Data Driven Inverse Problems
Autumn School
Sep $20^{\mathrm{th}} - 22^{\mathrm{nd}}$ 2023

Outline

Outline

- Inverse Problems arise in physics and engineering whenever we wish, indirectly, to recover parameters of a model from observations mapped by a *Forward Operator*

$$g^{\mathrm{obs}} = A(f) + \delta g \quad \leftrightarrow \quad f^{\dagger} = A^{-1}(g^{\mathrm{obs}})$$

- We think of A as a mapping between spaces X and Y. In the continuous setting these are typically Hilbert or Banach spaces. In the discrete setting they are typically Euclidean spaces.
- A may be linear or nonlinear, and A^{-1} is rarely available except in simple cases. Thus in (classical) inverse problems, the estimate $f^{\dagger}$ has to be obtained from an algorithm, acting as a proxy for A^{-1}

- Inverse Problems are traditionally discussed in terms of Hadamard's criterion for a problem to be *well-posed* (1902).
    1. A solution should *exist*
    2. A solution should be *unique*
    3. The error in the solution should depend stably on noise in the data
- If one or more of these criteria fail, the problem is *ill-posed* : these are precisely the problems we study!
- A solution may not exist if the data is outside the range of A, and it may be non-unique if A has a non-trivial kernel (Null-Space)
- In case of non-existance, a solution is usually chosen whose image in the range is closest to the data in a suitable distance.
- Similarly in the case of non-uniqueness, a solution is chosen with zero projection in the Null-space of A.
- The most serious problem in practice is instabilty w.r.t. noise.

As a motivational example, consider the problem

$$\frac{\partial f(x,t)}{\partial t} = \kappa \Delta f(x,t) \quad x \in \mathbb{R}, t \in \mathbb{R}_+ \tag{1}$$

with initial condition $f(x,0) = f_0(x)$. This corresponds to a model for the temperature of an infinite metal bar, with heat conductivity κ and initial temperature f_0 at time $t = 0$, evaluated forwards in time.
Equation (1) is formally solved by Fourier transforming with respect to x :

$$\frac{\partial \hat{f}(k,t)}{\partial t} = -\kappa k^2 \hat{f}(k,t) \quad \Leftrightarrow \quad \hat{f}(k,t) = \hat{f}_0(k)\mathrm{e}^{-\kappa k^2 t}. \tag{2}$$

Making use of the convolution theorem (see eq. (10)) we obtain the solution

$$f(x,t) = f_0(x) * \mathrm{e}^{-\frac{x^2}{4\kappa t}} = \int_{\mathbb{R}} f_0(x')\mathrm{e}^{-\frac{(x-x')^2}{4\kappa t}} \mathrm{d}x'. \tag{3}$$

- Equation (3) tells us that the temperature distribution $f_T(x)$ at some time $T > 0$ is found from the initial distribution $f_0(x)$ by convolution with a Gaussian of variance $\sigma^2 = 2\kappa T$. This defines the *forward problem*

$$f_T(x) = \mathcal{A}\, f_0(x). \tag{4}$$

as a linear integral operator.

- Equation (3) tells us that the temperature distribution $f_T(x)$ at some time $T > 0$ is found from the initial distribution $f_0(x)$ by convolution with a Gaussian of variance $\sigma^2 = 2\kappa T$. This defines the *forward problem*

$$f_T(x) = \mathcal{A}\, f_0(x). \tag{4}$$

  as a linear integral operator.
- More generally we call the Gaussian $G_\sigma = \exp[-x^2/4\kappa T]$ the *Green's function* for the PDE given by eq. (1).

- Equation (3) tells us that the temperature distribution $f_T(x)$ at some time $T > 0$ is found from the initial distribution $f_0(x)$ by convolution with a Gaussian of variance $\sigma^2 = 2\kappa T$. This defines the *forward problem*

$$f_T(x) = \mathcal{A}\, f_0(x). \tag{4}$$

  as a linear integral operator.
- More generally we call the Gaussian $G_\sigma = \exp[-x^2/4\kappa T]$ the *Green's function* for the PDE given by eq. (1).
- We want to solve the inverse problem, i.e. we would like an operator

$$f_0(x) = \mathcal{A}^{-1}\, f_T(x). \tag{5}$$

- Equation (3) tells us that the temperature distribution $f_T(x)$ at some time $T > 0$ is found from the initial distribution $f_0(x)$ by convolution with a Gaussian of variance $\sigma^2 = 2\kappa T$. This defines the *forward problem*

$$f_T(x) = \mathcal{A}\, f_0(x). \tag{4}$$

  as a linear integral operator.
- More generally we call the Gaussian $G_\sigma = \exp[-x^2/4\kappa T]$ the *Green's function* for the PDE given by eq. (1).
- We want to solve the inverse problem, i.e. we would like an operator

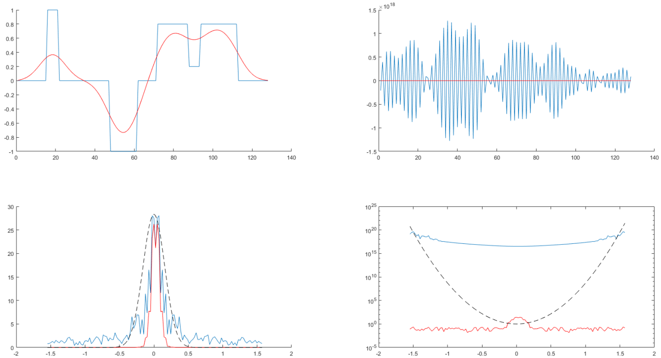
$$f_0(x) = \mathcal{A}^{-1}\, f_T(x). \tag{5}$$

- Proceeding in the same way we would have

$$\hat{f}_0(k) = \hat{f}_T(k)\mathrm{e}^{\kappa k^2 T} \Leftrightarrow f_0(x) = \mathcal{F}^{-1}\left[\frac{\hat{f}_T(k)}{\mathrm{e}^{-\kappa k^2 T}}\right] \tag{6}$$

  from which we see a number of difficulties

1. No direct integral transform expression for $\mathcal{A}^{-1}$ exists, because the function $\mathrm{e}^{\kappa k^2 T}$ in eq. (2) is unbounded and has no inverse FT.
2. Solving eq. (5) using division in Fourier space will lead to amplification of any measurement noise; thus let $f_T^\delta(x) = f_T(x) + e(x)$ where $e(x) \sim N(0, \varsigma)$ is (white noise) drawn from an isotropic Normal distribution, then:
$$\mathcal{F}^{-1}\left[\frac{\hat{f}_0(k) + \hat{e}(k)}{\mathrm{e}^{-\kappa k^2 T}}\right] = f_0(x) + \eta_x(x)$$
where $\eta_x(x)$ is drawn from a distribution whose variance increases exponentially with the square of spatial frequency k
3. proceeding instead to solve eq. (5) using the *backward* heat equation
$$\frac{\partial f(x,t)}{\partial t} = -\kappa \Delta f(x,t) \quad t < T \tag{7}$$
with "final" condition $f(x,t) = f_T(x)$ is unstable and rapidly leads to noise dominating the signal.



a) initial state f_0 (blue) and f_T (red); b) final state f_T with 1% noise, and backward propagation to recovered inital state $f_0^\dagger$; c) Fourier transform of a) with Green's function $\hat{G}_\sigma(k)$ (dashed line); d) Fourier transform of b) with reciprocal Green's function $1/\hat{G}_\sigma(k)$ (dashed line). Note the log scale in d) arising due to the exponential growth in the high frequencies.

Introduction

Singular Value Decomposition

- Linear inverse problems can be characterized by the spectrum of the forward operator A which is defined in terms of its singular values
- $Af = \sum_j u_j w_j \langle v_j, f \rangle$ with $\{u_j\}$ orthonormal w.r.t inner product in Y, $\{v_j\}$ orthonormal w.r.t inner product in X and

$$Av_j = w_j u_j \qquad A^* u_j = w_j v_j \qquad \text{where } A^* \text{ is the adjoint of } A$$

- It follows that the *Moore-Penrose* inverse of A can be formed : $A^\dagger g = \sum_j v_j \frac{\langle u_j, g \rangle}{w_j}$.
- **Low rank problems** are those for which the $w_j = 1$ up to the rank of A and zero afterwards. Examples : inpainting and super resolution in imaging
- **Mildly Ill-posed problems** are those for which w_j decays polynomially. Examples : X-Ray CT.
- **Strongly ill-posed problems** are those for which w_j decays exponentially. Examples : Image deconvolution

Let's look at a few examples in more depth

Outline

- As we have seen, the *forward* problem for the heat equations is solved by convolution with a Gaussian. Therefore the *inverse problem* for the heat equations is a *deconvolution* problem.
- More generally, deconvolution problems arise whenever a signal is degraded by a filtering process that acts as a weighted average.
- Let us write an integral transform operator $\mathcal{A} : X \to X$ as

$$g(x) = \int_\Omega K(x, x')f(x')\mathrm{d}x' \tag{8}$$

- A strict definition of convolution requires that the kernel in eq. (8) is *stationary*, i.e. $K(x, y) \to K(x - y) \equiv K(\cdot)$. Then $K(x)$ is a filter kernel, usually of local support, that acts equally over every point in the signal f.
- Physically this arises due to a measurement operation such as a camera with an aperture [1].
- When Ω is the infinite interval, or finite with periodic boundary conditions, then stationary convolution obeys the convolution theorem eq. (10).

[1] In imaging systems the convolution filter is known as the *Point Spread Function* (PSF)

Reminder :

Definition (Convolution)

The convolution of two functions $f(x), h(x), x \in \mathbb{R}^d$ in d dimensions is defined for functions whose absolute integral is bounded i.e.
$f(x) \in L_1(\mathbb{R}^d) \Rightarrow \int_{R^d} |f(x)| \mathrm{d}^d x < \infty$, produces another function $g(x) \in L_1(\mathbb{R}^d)$, given by

$$g(x) = (f * h)(x) := \int_{\mathbb{R}^d} f(x) h(x - x') \mathrm{d}^d x' \ . \tag{9}$$

Convolution is commutative. Let $\hat{f}(\mathbf{k}), \hat{h}(\mathbf{k}), \hat{g}(\mathbf{k}), \mathbf{k} \in \mathbb{R}^d$ be the d-dimensional Fourier transforms of f, h, g and $\mathbf{k}$ be the *wave-number* in d dimensions. The *Convolution Theorm* states

$$\hat{g}(\mathbf{k}) = \hat{f}(\mathbf{k}) \hat{g}(\mathbf{k}) \tag{10}$$

- The convolution operator defined in eq. (10) has the property that its (right) singular functions are the harmonic functions, i.e.

$$K(x) * \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}} = \hat{K}(\mathbf{p})\, \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}} \tag{11}$$

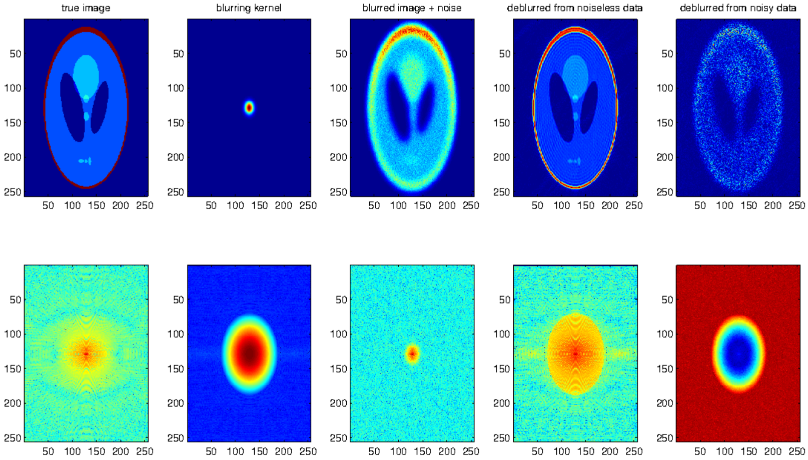
  for any given point **p** in **k**-space.
- Although formally the (stationary) deconvolution problem is solved when $K(x)$ is known by division in the frequency domain, i.e.

$$\text{Fourier Deconvolution} \qquad f(x) = \mathcal{F}^{-1}_{\mathbf{k}\to x}\left[\frac{\hat{g}(\mathbf{k})}{\hat{K}(\mathbf{k})}\right] \tag{12}$$

  we note
  1. Fourier Deconvolution will fail if $\hat{K}(\mathbf{k})$ is zero at any point. This is equivalent to saying that the Null-space of eq. (9) is spanned by the harmonic functions with frequencies in the zero set of $\hat{K}(\mathbf{k})$
  2. Even if there are no zeros of $\hat{K}(\mathbf{k})$ eq. (12) becomes unstable wherever the values become small.

Image Deconvolution

Example : Fourier Domain Deconvolution



Outline

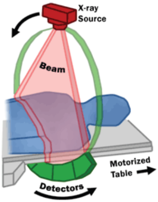
- X-rays were first observed and documented in 1895 by Wilhelm Conrad Roentgen, a German scientist who found them quite by accident when experimenting with vacuum tubes.
- 1^{st} Nobel Prize in Physics to Roentgen "in recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him"

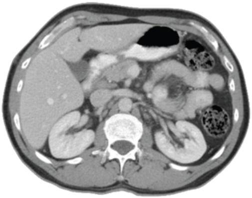
Tomography
The Radon Transform

If we consider a slice $\mathcal{M}$ of a 3D object as a 2D image, then we can consider the **X-Ray attenuation coefficient** as a 2D function

$$f(x,y) \equiv \mu_{\mathrm{a}}\big|_{(x,y)\in\mathcal{M}}$$

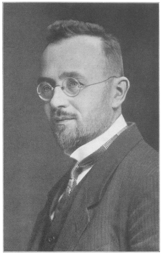
A single X-Ray projection image is formed by integrating along rays in the plane $\mathcal{M}$ resulting in one line of data.
Obviously, the line of data is not enough to recover the 2D function $f(x,y)$.
In the 1960s the pioneering work of Hounsfield and McCormack considered using projections at *multiple view angles* combined with *image reconstruction* to build an image of the interior of objects. They received the Nobel Prize in 1979



The mathematics behind X-Ray tomography was based on the work of Johannes Radon in 1917.

The mathematics behind X-Ray tomography was based on the work of Johannes Radon in 1917.
The forward projection operator of the Radon transform in 2D is an integral operator mapping a function $f(x,y)$ to a function $g(s,\theta)$, by integrating along the parallel lines $\mathbf{r}\cdot\hat{\mathbf{n}} = s$



$$g = \mathcal{R}_{\rm 2D} f$$

Note : $\hat{\mathbf{n}} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$ is the unit vector perpendicular to the direction of the ray.

$$g(s,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\underbrace{\delta(s-(y\cos\theta - x\sin\theta))}_{K(x,y,\theta,s)}\mathrm{d}x\mathrm{d}y \quad = \int_{\mathbf{r}\cdot\hat{\mathbf{n}}=s} f(x,y)\mathrm{d}\ell \tag{13}$$

Tomography
The Radon Transform

The mathematics behind X-Ray tomography was based on the work of Johannes Radon in 1917.
The forward projection operator of the Radon transform in 2D is an integral operator mapping a function $f(x, y)$ to a function $g(s, \theta)$, by integrating along the parallel lines $\mathbf{r} \cdot \hat{\mathbf{n}} = s$

$$g = \mathcal{R}_{\mathrm{2D}} f$$

Note : $\hat{\mathbf{n}} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$ is the unit vector perpendicular to the direction of the ray.

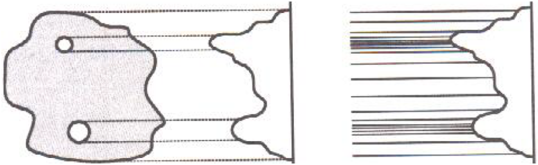
$$g(s,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\underbrace{\delta(s - (y\cos\theta - x\sin\theta))}_{K(x,y,\theta,s)}\mathrm{d}x\mathrm{d}y \quad = \int_{\mathbf{r}\cdot\hat{\mathbf{n}}=s} f(x,y)\mathrm{d}\ell \tag{13}$$

Note that the operator has a kernel which is a function of four variables, two of which are integrated out.

To get the adjoint we apply the rule of integrating with the kernel over the other two variables

The adjoint maps a function in (s, θ) to one in (x, y); it is known as *Back-projection* $h = \mathcal{R}^*_{\mathrm{2D}} b$

$$h(x, y) = \int_{-\infty}^{\infty} \int_0^{\pi} b(s, \theta) \delta(s - (y \cos\theta - x \sin\theta)) \mathrm{d}\theta \mathrm{d}s \tag{14}$$



Left : one projection of a 2D project. Right : the backprojection of a single projection gives an image which is translationally uniform in the direction of the projection on the left. Code `radonex1proj`

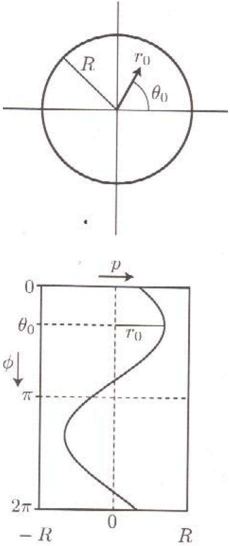
To understand the Radon Transform we should look at the data space.
Consider the forward projection of a single point
$f^\delta = \delta(\boldsymbol{r} - \boldsymbol{r}_0) \equiv \delta(x - x_0, y - y_0)$
The Radon Transform of this object is

$$
\begin{aligned}
g(s,\theta) &= \mathcal{R}_{\mathrm{2D}} f^\delta \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x - x_0, y - y_0)\delta(s - (y\cos\theta - x\sin\theta))\mathrm{d}x\mathrm{d}y \\
&= \delta(s - (y_0\cos\theta - x_0\sin\theta)) = \delta(s - r_0\cos(\theta_0 - \theta))
\end{aligned}
$$

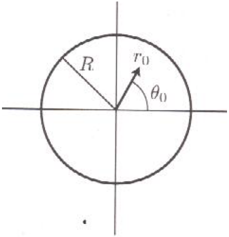
where r_0, θ_0 are the polar coordinates of $\boldsymbol{r}_0$

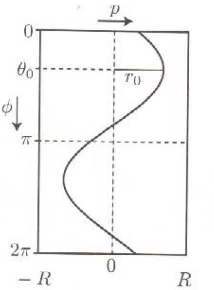
- The above shows that the projection of a δ-function in 2D is a δ-function in the 1D projection at each angle.

- The above shows that the projection of a δ-function in 2D is a δ-function in the 1D projection at each angle.
- If we plot the position of this δ-function over angle we see its position varies over the space variable s in a sinusoidal manner. The amplitude of the sine function is the radial position r_0 and the phase is the polar angle θ_0. This gives rise to the term **Sinogram Space** for the data space of the Radon Transform. Code `ShowSinogram`



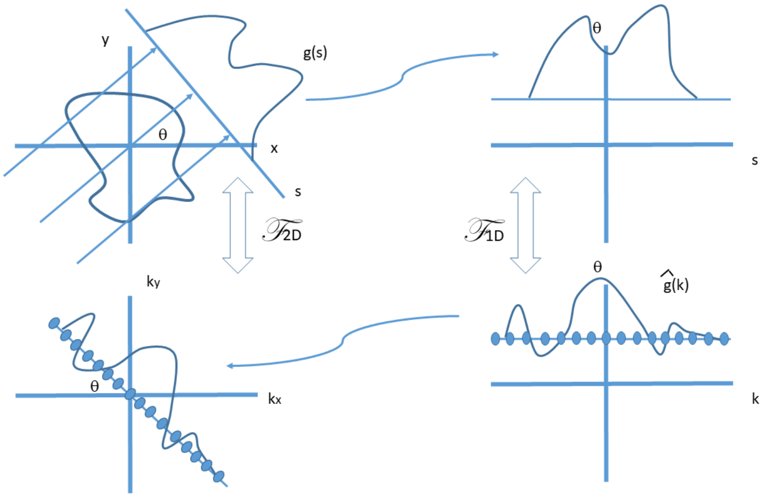
- The above shows that the projection of a δ-function in 2D is a δ-function in the 1D projection at each angle.
- If we plot the position of this δ-function over angle we see its position varies over the space variable *s* in a sinusoidal manner. The amplitude of the sine function is the radial position r_0 and the phase is the polar angle θ_0. This gives rise to the term **Sinogram Space** for the data space of the Radon Transform. Code `ShowSinogram`
- Since the Radon Transform is linear, the sinogram of a more general object is the linear superposition of the sine functions of its individual pixels.





There is a remarkable relationship between the Fourier Transform and the Radon Transform which is expressed via the **Central Slice Theorem** (CLT) This states that The 1D Fourier Transform $\mathcal{F}_{s\to k}$ of the Radon Transform of a 2D function $f(x,y)$ is the same as the 2D Fourier Transform $\mathcal{F}_{x\to k_x, y\to k_y}$ of $f(x,y)$ sampled along the line $k_x = -k\sin\theta, k_y = k\cos\theta$.

$$
\begin{aligned}
\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}ks} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(s-(y\cos\theta - x\sin\theta))\mathrm{d}x\mathrm{d}y\mathrm{d}s &= \\
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}ks} f(x,y)\delta(s-(y\cos\theta - x\sin\theta))\mathrm{d}s\mathrm{d}x\mathrm{d}y &= \\
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}k(y\cos\theta - x\sin\theta)} f(x,y)\mathrm{d}x\mathrm{d}y &= \\
\hat{F}(-k\sin\theta, k\cos\theta)
\end{aligned}
$$


y
g(s)
θ
x
s
θ
s
$\mathcal{F}_{2D}$
$\mathcal{F}_{1D}$
θ
$\hat{g}(k)$
ky
θ
kx
k

The Central Slice Theorem is the key step to finding an exact reconstruction formula for the Radon Transform.

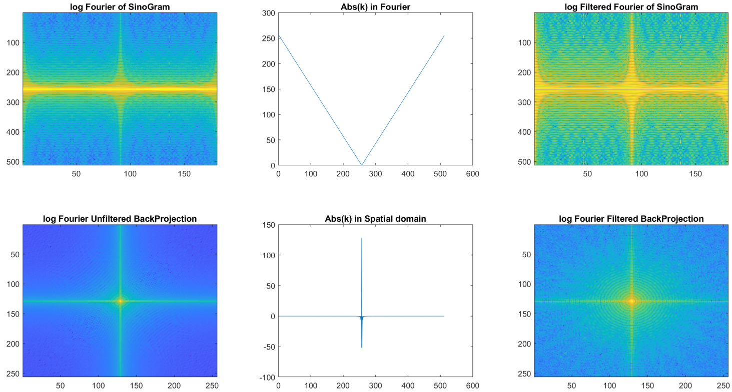
$$f(x,y) = \frac{1}{(2\pi)^2}\int_0^{\pi}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}k\boldsymbol{r}\cdot\hat{\mathbf{n}}}\mathcal{F}_{1D}\left[\mathcal{R}f(x,y)\right]|k|\mathrm{d}k\mathrm{d}\theta \qquad (15)$$

This comes from Fourier Inversion formula with $\mathbf{k} = (-k\sin\theta, k\cos\theta)$:

$$\begin{aligned} f(x,y) &= \frac{1}{(2\pi)^2}\int_{\infty}^{\infty}\int_{-\infty}^{\infty}\hat{F}(\mathbf{k})\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\boldsymbol{r}}\mathrm{d}\mathbf{k} \\ &= \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{\infty}\hat{F}(k\hat{\mathbf{n}})\mathrm{e}^{\mathrm{i}k\boldsymbol{r}\cdot\hat{\mathbf{n}}}k\mathrm{d}k\mathrm{d}\theta \\ &= \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{\infty}\mathcal{F}_{1D}\left[\mathcal{R}f(x,y)\right]\mathrm{e}^{\mathrm{i}k\boldsymbol{r}\cdot\hat{\mathbf{n}}}k\mathrm{d}k\mathrm{d}\theta \\ &= \frac{1}{(2\pi)^2}\int_0^{\pi}\int_{-\infty}^{\infty}\mathcal{F}_{1D}\left[\mathcal{R}f(x,y)\right]\mathrm{e}^{\mathrm{i}k\boldsymbol{r}\cdot\hat{\mathbf{n}}}|k|\mathrm{d}k\mathrm{d}\theta \end{aligned}$$

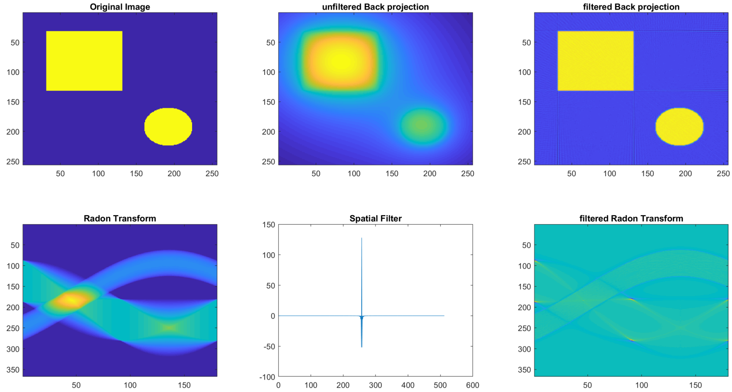
where the last equality follow from the symmetry property $\hat{g}(-k,-\hat{\mathbf{n}}) = \hat{g}(k,\hat{\mathbf{n}})$. Code `ExplicitFilterBP`

Tomography
Filtered BackProjection



Multiplication in Fourier Domain has an equivalent convolution in the spatial domain.

$$h(s) = \mathcal{F}^{-1}_{k \to s} |k|$$

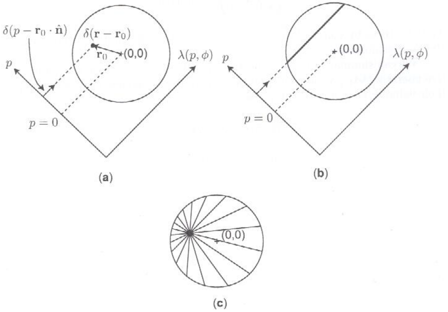


The filter with $|k|$ is the difference between filtered and unfiltered back-projection.
Unfiltered backprojection can be written as

$$\mathcal{R}^*_{\mathrm{2D}}\mathcal{R}_{\mathrm{2D}} f = \int_{\infty}^{\infty} \frac{f(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\mathrm{d}\boldsymbol{r}'$$

This is a convolution operation with a stationary kernel $\frac{1}{|r|}$.
This is remarkable since neither $\mathcal{R}_{\mathrm{2D}}$ nor $\mathcal{R}^*_{\mathrm{2D}}$ is shift-invariant.

a) A single 2D δ-function and its projection at angle ϕ.
b) backprojection of the 1D projection in a).
c) Sum of many back-projections. This sum limits to a $1/r$ function as the number of projection angles approaches infinity.
Code `ShowUFBconv`



The fact that $\mathcal{R}_{2\mathrm{D}}^{*}\mathcal{R}_{2\mathrm{D}}$ is a convolution means that it has harmonic functions as its eigenfunctions.
This in turn means that the SVD of the Radon Transform has harmonic functions as its singular functions in the image space.
What are the corresponding data space (i.e. sinogram space) functions ?
They are also sinusoidal in the spatial variable :

$$\mathcal{R}_{2\mathrm{D}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\boldsymbol{r}} \mapsto \sqrt{|k|}\mathrm{e}^{iks}\delta(\mathbf{k}\cdot\hat{\mathbf{n}}_{\perp}) = \frac{1}{\sqrt{|k|}}\mathrm{e}^{iks}\delta(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}_{\perp})$$

They are non-zero only if the angle of projection corresponds exactly to the direction of the plane-wave $\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\boldsymbol{r}}$; in any other direction the integral crosses the positive and negative components of the plane-wave in equal proportions, leading to a value of zero. Code `ShowRadonSVD`.

Outline

Many forward problems are based on the measurement of a field that obeys a PDE with appropriate boundary conditions.
These include steady-state such as EIT, time-harmonic such as ISP, and time-domain problems such as UST.
Writing a general PDE as

$$\mathcal{L}(f)U(x) = q(x)\,, \qquad x \in \Omega \tag{16}$$
$$\mathcal{B}U(x) = v\,, \qquad x \in \partial\Omega \tag{17}$$

the measurements result from sampling the field U at specified detectors that may vary in space and/or time, i.e.

$$y = \mathcal{M}U \tag{18}$$

The *forward* problem is to solve the mapping from parameters f to measurements, i.e.

$$y = \mathcal{A}(f) = \mathcal{M}\mathcal{L}^{-1}(f)\, q. \tag{19}$$

The corresponding inverse problem is non-linear and usually strongly ill-posed.
A typical strategy is to linearise the problem around an initial state f_0 and use an iterative descent scheme for minimising a (regularised) objective function such as

$$\mathcal{E} = \frac{1}{2}\|g - \mathcal{A}(f)\|_2^2 + \alpha\Psi(f) \tag{20}$$

Writing the Fréchet derivative of $\mathcal{A}$ as

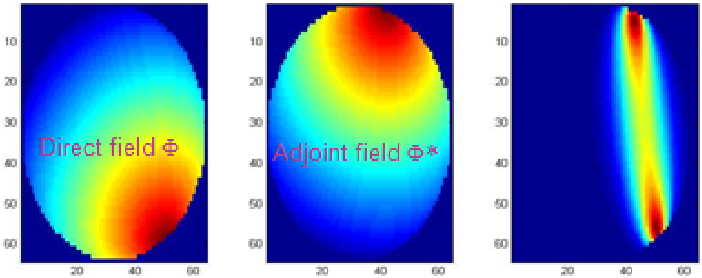
$$\mathcal{A}'(f) = \mathcal{M}\mathcal{L}^{-1}(f)\frac{\partial \mathcal{L}}{\partial f}\mathcal{L}^{-1}(f)\, q \tag{21}$$

we can write the gradient of the data-fit functional in eq.(20) as

$$
\begin{aligned}
\nabla \frac{1}{2}\|g - \mathcal{A}(f)\|_2^2 &= \mathcal{A}^{'*}(f)[g - \mathcal{A}(f)] && (22)\\
&= \left\langle g - \mathcal{A}(f), \mathcal{A}'(f) \right\rangle && (23)\\
&= \left\langle g - \mathcal{A}(f), \mathcal{M}\mathcal{L}^{-1}(f)\frac{\partial \mathcal{L}}{\partial f}\mathcal{L}^{-1}(f)\, q \right\rangle && (24)\\
&= \left\langle \underbrace{\mathcal{L}^{-1*}(f)\mathcal{M}^*\left[g - \mathcal{A}(f)\right]}_{U^*}, \frac{\partial \mathcal{L}}{\partial f} U \right\rangle && (25)\\
&= \left\langle U^*, \frac{\partial \mathcal{L}}{\partial f} U \right\rangle && (26)
\end{aligned}
$$

Here U^* is an *adjoint field* representing the backpropagation of the measurement error from the detectors into the domain, solving

$$
\mathcal{L}^*(f)U^* = \mathcal{M}^*\left[g - \mathcal{A}(f)\right] \qquad (27)
$$



Direct field Φ
Adjoint field Φ*
10
20
30
40
50
60
20
40
60

Let's now consider the time-dependent case :

$$\frac{\partial U(x,t)}{\partial t} - \mathcal{L}(f)U(x,t) = q(x)\ , \qquad x \in \Omega, t \in \mathbb{R}_+$$

One way to solve this problem is through a *Forward Euler* scheme :

$$U_T = \underbrace{(\mathsf{I} + \delta\mathcal{L}) \circ (\mathsf{I} + \delta\mathcal{L}) \circ \cdots \circ\ (\mathsf{I} + \delta\mathcal{L})}_{\text{N times}}\, q$$

Our Fréchet derivative (c.f.eq.29) becomes

$$\mathcal{A}'(f) = \sum_{j=1}^{N} \mathcal{M}\, \underbrace{(\mathsf{I} + \delta\mathcal{L}) \circ (\mathsf{I} + \delta\mathcal{L}) \circ \cdots \circ\ (\mathsf{I} + \delta\mathcal{L})}_{\text{N-1-j times}} \circ \frac{\partial \delta\mathcal{L}}{\partial f} \circ \underbrace{(\mathsf{I} + \delta\mathcal{L}) \circ (\mathsf{I} + \delta\mathcal{L}) \circ \cdots \circ\ (\mathsf{I} + \delta\mathcal{L})}_{\text{j times}}\, q \qquad (28)$$

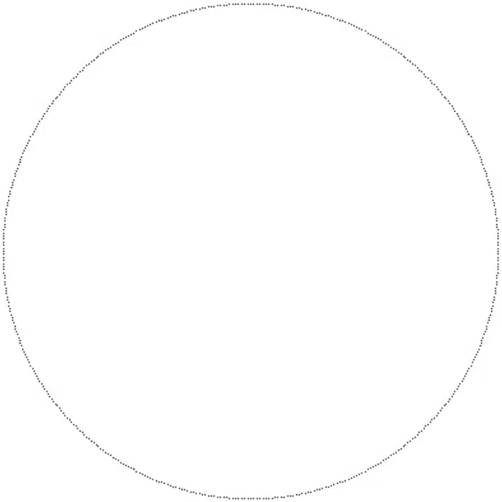
The gradient of data-fit term (cf eq.26) becomes a convolution in time

$$\nabla \frac{1}{2}\|g - \mathcal{A}(f)\|_2 = \sum_{j=1}^{N} \left\langle U^{*(N-1-j)}, \frac{\partial \mathcal{L}}{\partial f} U^{(j)} \right\rangle \tag{29}$$

where now $U^*(x,t)$ is the *time-reversed adjoint field* representing the backpropagation of the measurement error from the detectors into the domain, solving

$$\frac{\partial U^*(x,t)}{\partial t} + \mathcal{L}^*(f)U^*(x,t) = \mathcal{M}^*\left[g - \mathcal{A}(f)\right] \tag{30}$$

Non-Linear Inverse Problems
Wave propagation: Example



Outline

Summary

- The basic concepts of Inverse Problems
- Model problems : Heat equation, Image Deconvolution, X-Ray tomography
- Ideas of non-linear inverse problems