Regularization in Tomography

Dealing with Ambiguity and Noise

Per Christian Hansen
Technical University of Denmark

HDtomo
HIGH-DEFINITION TOMOGRAPHY

DTU Compute
Department of Applied Mathematics and Computer Science

About Me ...

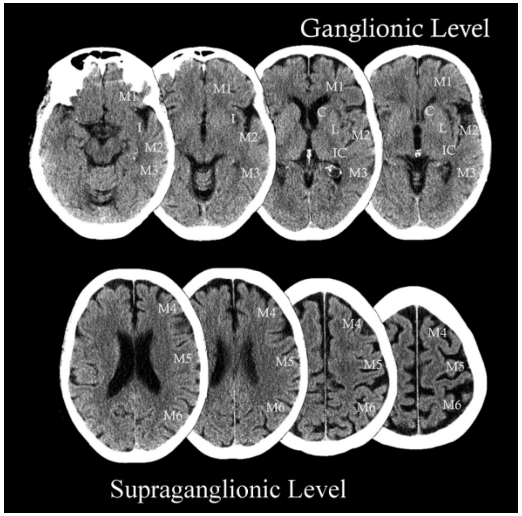
- Professor of Scientific Computing at DTU
- Interests: inverse problems, tomography, regularization algorithms, matrix computations, image deblurring, signal processing, Matlab software, ...
- Head of the project High-Definition Tomography, funded by an ERC Advanced Research Grant.
- Author of several Matlab software packages.
- Author of four books.

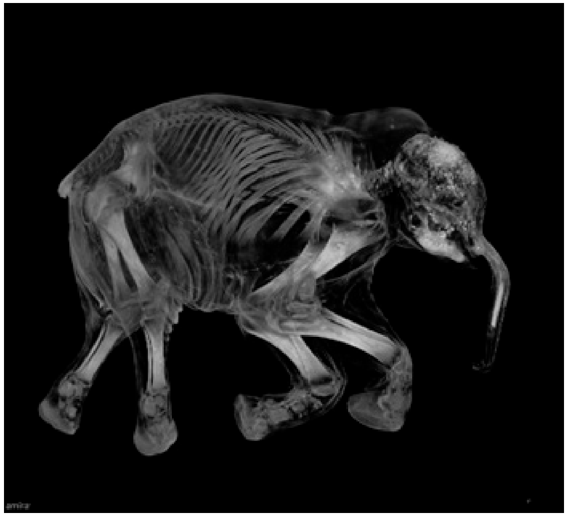
HDtomo
HIGH - DEFINITION TOMOGRAPHY

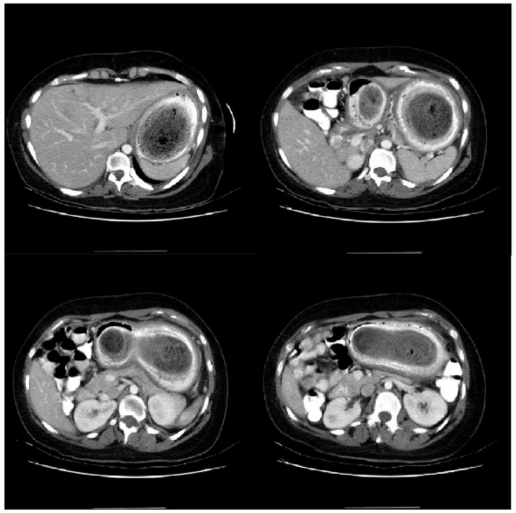
Tomographic Reconstructions are Amazing!

Tomographic reconstructions are routinely computed each day.

Our reconstruction algorithms are so reliable that we sometimes forget we are actually dealing with *inverse problems* with inherent stability problems.







These algorithms are successful because they automatically incorporate *regularization* techniques that, in most cases, handle very well the stability issues.

This talk is intended for scientists who need a "brush up" on the underlying mathematics of some common tomographic reconstruction algorithms.

Outline of Talk

We take a basic look at the *inverse problem* of “plain vanilla” absorption CT reconstruction and the associated stability problems:

- solutions are very sensitive to data errors,
- solutions may fail to be unique.

We demonstrate how *regularization* is used to avoid these problems:

- We make the reconstruction process stable by
- incorporate regularization in reconstruction algorithm.

Webster

Reg·u·lar·ize – to make regular by conformance to law, rules, or custom.

Reg·u·lar – constituted, conducted, scheduled, or done in conformity with established or prescribed usages, rules, or discipline.

In tomography: we make the problem, or the solution, more regular in order to prevent it from being dominated by noise and other artefacts.

We look at the principles of different regularization techniques and show that they have different *impact* in the computed reconstructions.

The Origin of Tomography

Johan Radon, *Über die Bestimmung von Funktionen durch ihre Integralwerte Längs gewisser Mannings-faltigkeiten*, Berichte Sächsische Akadamie der Wissenschaften, Leipzig, Math.-Phys. Kl., 69, pp. 262-277, 1917.

Main result: An object can be perfectly reconstructed from a full set of projections.

NOBELFÖRSAMLINGEN KAROLINSKA INSTITUTET

THE NOBEL ASSEMBLY AT THE KAROLINSKA INSTITUTE

11 October 1979

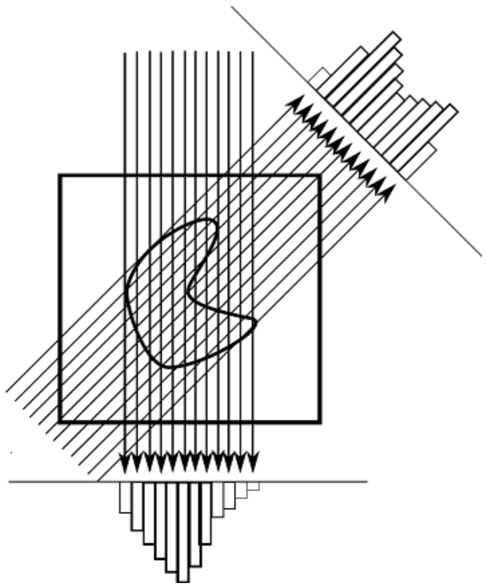
The Nobel Assembly of Karolinska Institutet has decided today to award the Nobel Prize in Physiology or Medicine for 1979 jointly to

Allan M Cormack and Godfrey Newbold Hounsfield

for the "development of computer assisted tomography".

The Radon Transform

The principle in parallel-beam tomography: send parallel rays through the object at different angles, measure the damping.

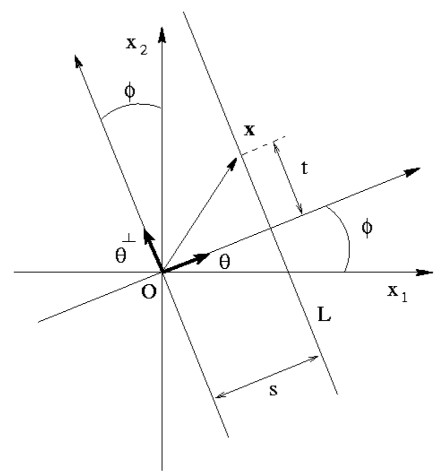


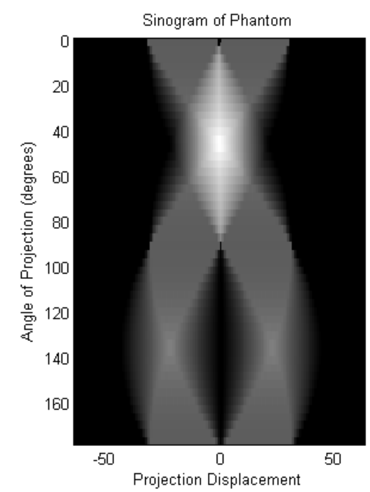
$$f(\mathbf{x}) = \text{2D object / image}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\hat{f}(\phi, s) = \text{sinogram / Radon transform}$$

Line integral along line defined by ϕ and s:

$$\hat{f}(\phi, s) = \int_{-\infty}^{\infty} f\left(s \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} + \tau \begin{bmatrix} -\sin\phi \\ \cos\phi \end{bmatrix} \right) d\tau$$





The Inverse Radon Transform

Let R denote the Radon transform, such that

$$\hat{f} = R\, f \quad \Leftrightarrow \quad f = R^{-1}\,\hat{f}$$

How to conveniently write the inverse Radon transform:

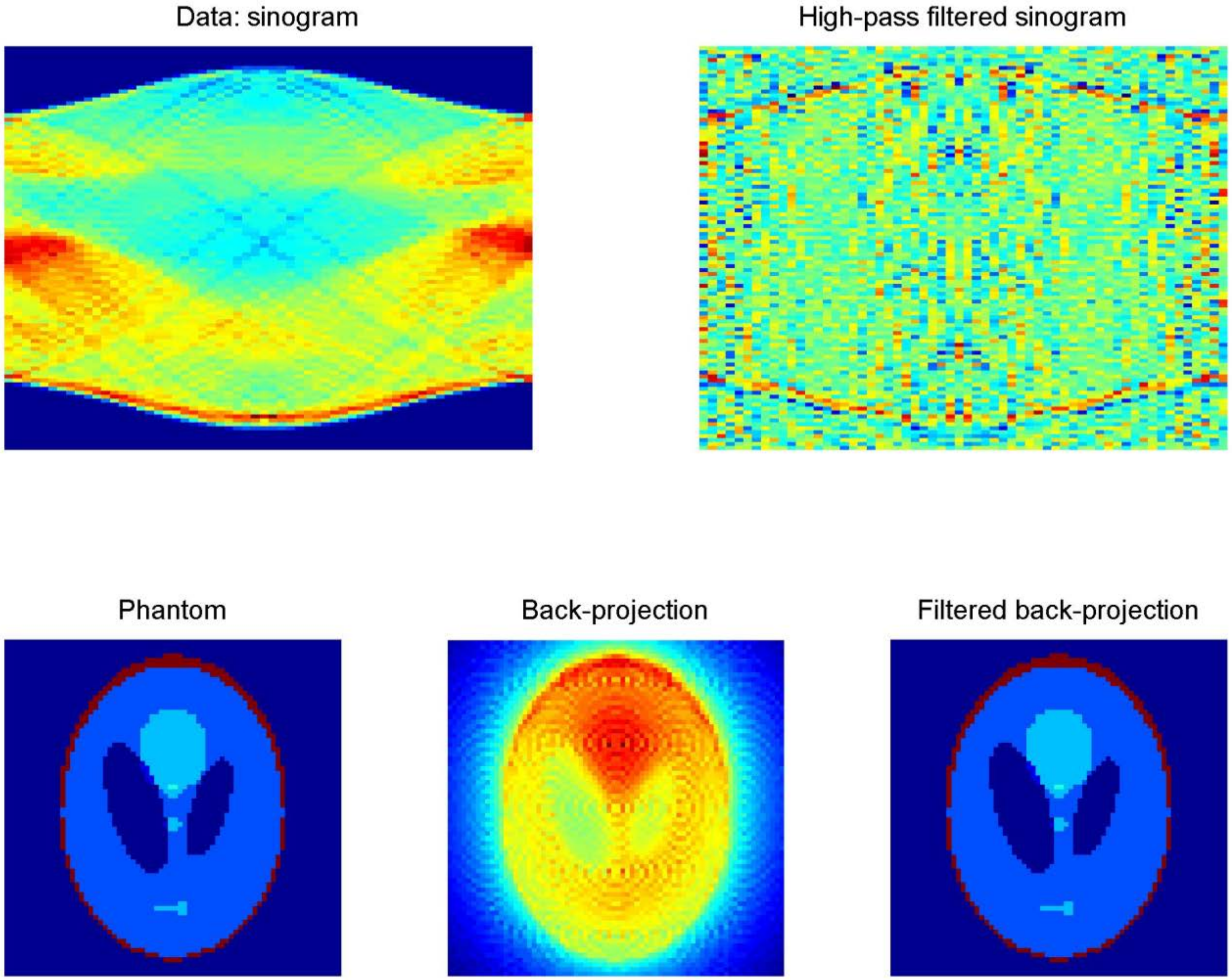
$$\begin{aligned} R^{-1} &= c\,(-\Delta)^{1/2} R^*, \qquad c = \text{constant} \\ R^* &= \text{backprojection (dual transform)} \\ \Delta &= \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2 = \text{Laplacian} \\ (-\Delta)^{1/2} &= \text{high-pass filter} \qquad \mathcal{F}\left((-\Delta)^{1/2}\xi\right)(\omega) = |\omega|\mathcal{F}(\xi)(\omega) \end{aligned}$$

Not precisely how we compute it.

The operators $(-\Delta)^{1/2}$ and R^* commute – this leads to the *filtered back projection* (FBP) algorithm:

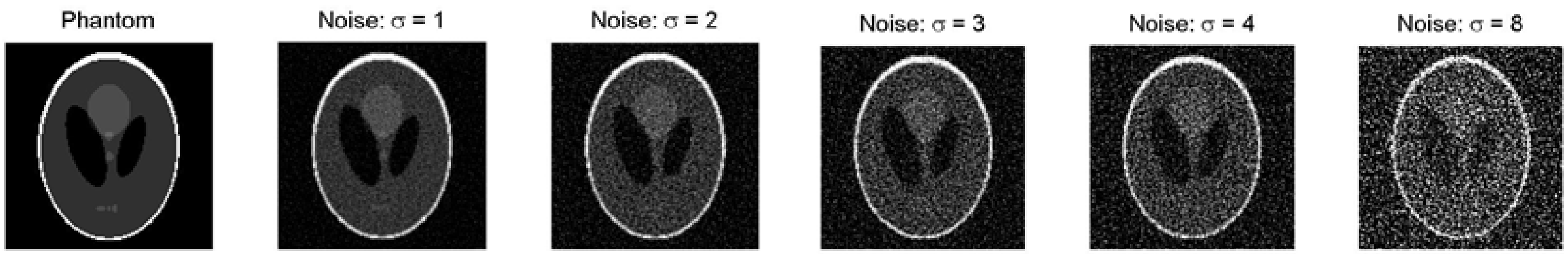
$$R^{-1} = c\,R^*(-\Delta)^{1/2} \quad \rightarrow \quad f = R^{-1}\hat{f} = c\,R^*(-\Delta)^{1/2}\hat{f}.$$

Matlab Check …



“Naïve” FBP is Very Sensitive to Noise

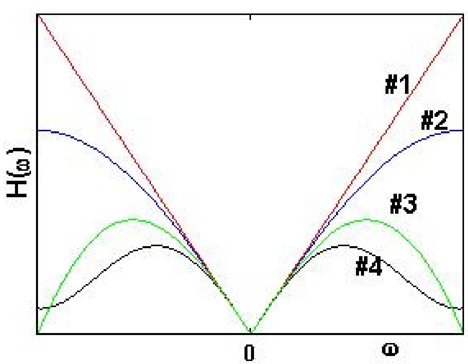
The high-pass filter $|\omega|$ in ”naive” FBP amplifies high-frequency noise in data.



The solution is to insert an *additional filter* than dampens higher frequencies:

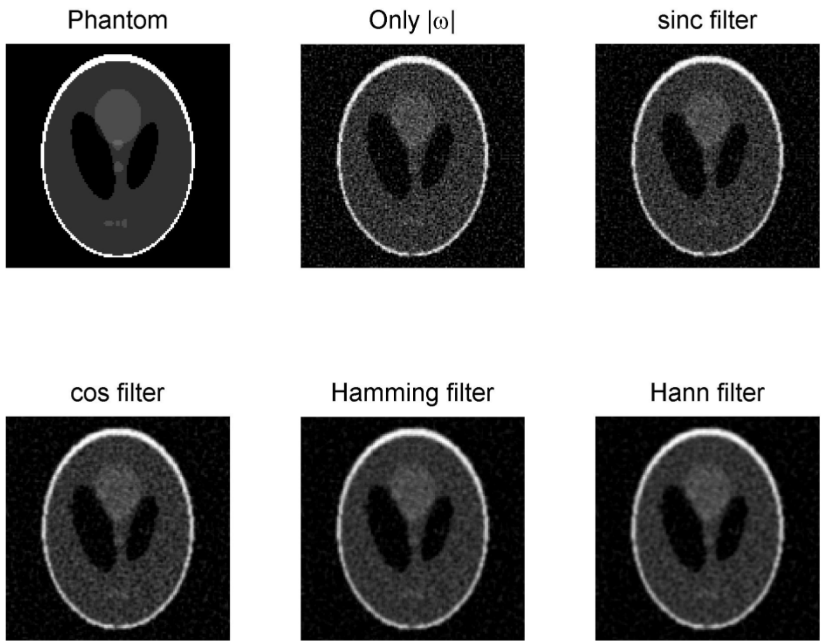
$$|\omega| \;\rightarrow\; \psi(\omega)\cdot|\omega|$$

This is regularization!

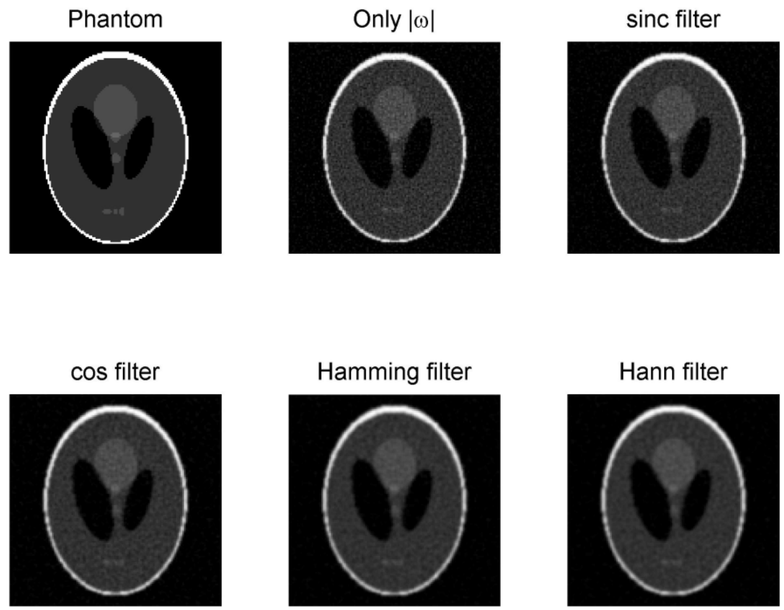


1. Only $|\omega|$
2. sinc filter (”Shepp-Logan”)
3. cos filter
4. Hamming filter

FBP + Low-Pass Filter Suppresses Noise



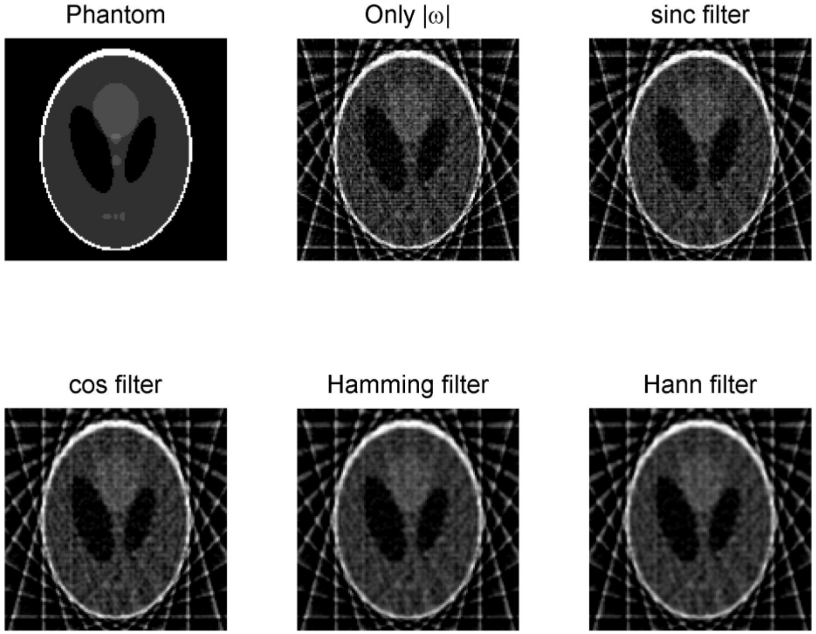
180 projections



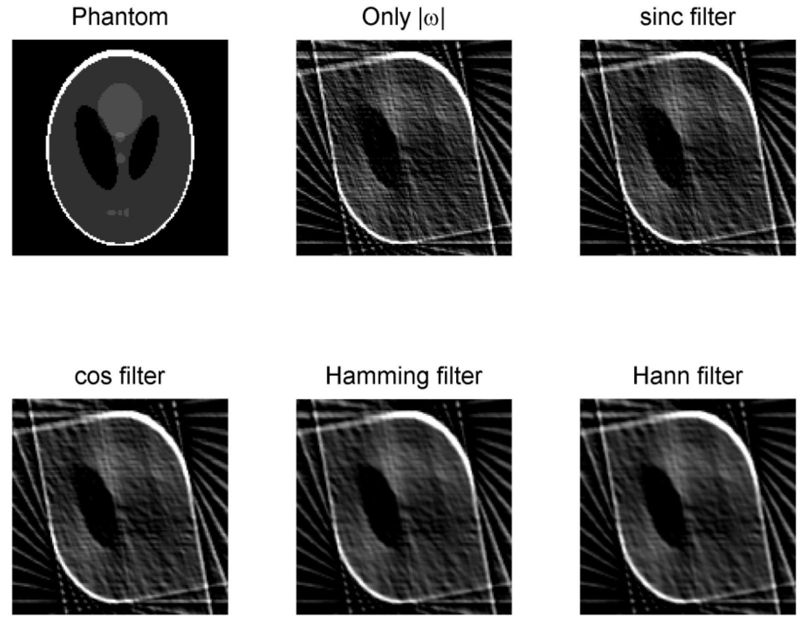
1000 projections

More data is better!
But we loose some details due to the filter (low-pass = smoothing).

FBP with Few Projections



Projection angles
15:15:180



Projection angles
10:10:100

Less data creates trouble!

Now the problem is *under-determined* and artifacts appear.

Setting Up an Algebraic Model

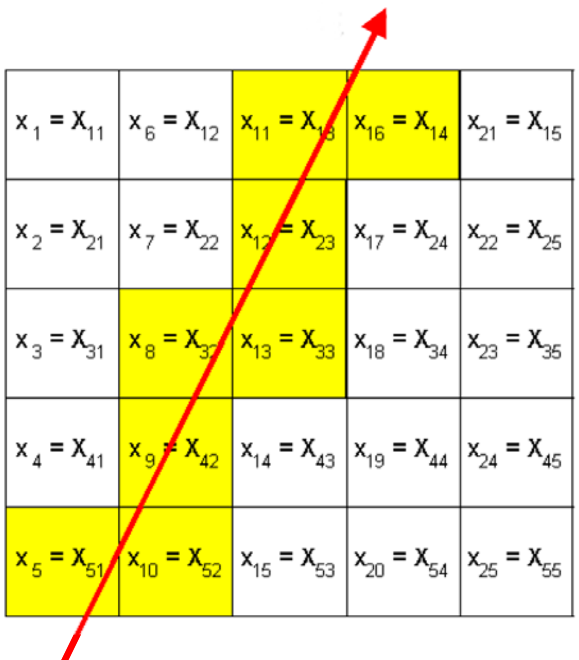
To understand these issues better, let us switch to an algebraic formulation!

Damping of *i*-th X-ray through domain:

$$b_i = \int_{\mathrm{ray}_i} \chi(\mathbf{s})\, d\ell, \qquad \chi(\mathbf{s}) = \text{attenuation coef.}$$

Assume $\chi(\mathbf{s})$ is a constant x_j in pixel *j*, leading to:

$$b_i = \sum_j a_{ij}\, x_j, \qquad a_{ij} = \text{length of ray } i \text{ in pixel } j.$$



<u>This leads to a large linear system:</u>

$$A\, x = b \qquad b = \bar{b} + e$$

"Geometry"

Image

Projections

Noise

$$\bar{b} = A\, \bar{x}$$

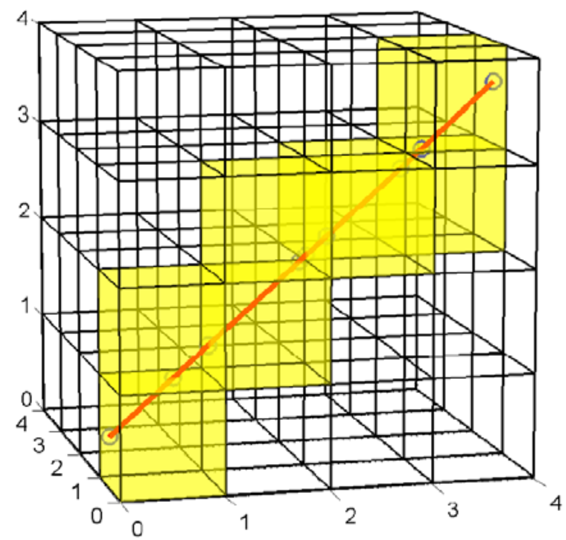
$$\bar{x} = \text{exact image}$$

More About the Coefficient Matrix, 3D Case

$$\boxed{b_i = \textstyle\sum_j a_{ij}\, x_j , \qquad a_{ij} = \text{length of ray } i \text{ in voxel } j.}$$

To compute the matrix element a_{ij} we simply need to know the intersection of ray i with voxel j. Let ray i be given by the line

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + t \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} , \qquad t \in \mathbb{R}.$$



The intersection with the plane $x = p$ is given by

$$\begin{pmatrix} y_j \\ z_j \end{pmatrix} = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} + \tfrac{p - x_0}{\alpha} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} , \qquad p = 0, 1, 2, \ldots$$

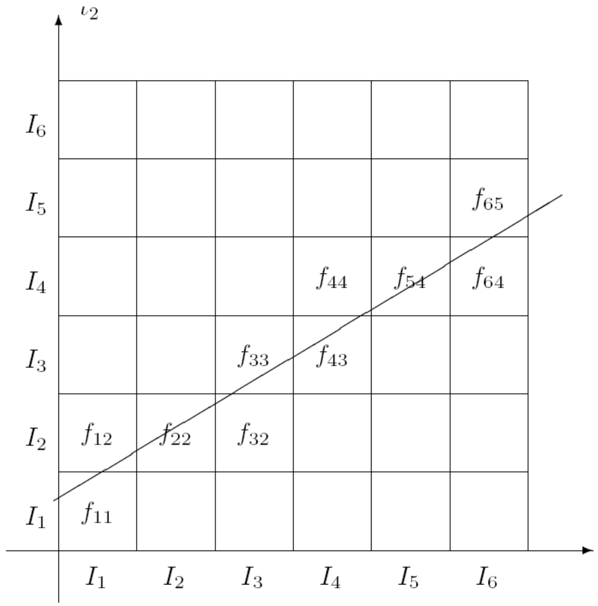
with similar equations for the planes $y = y_j$ and $z = z_j$.

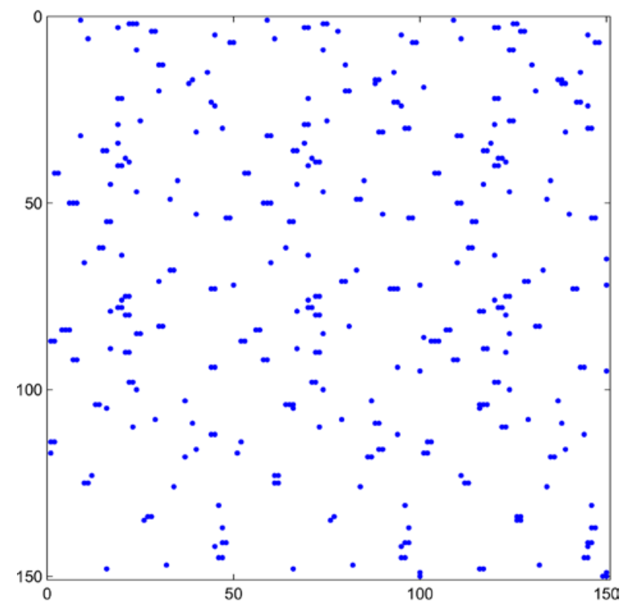
From these intersetions it is easy to compute the ray length in voxel j.

Siddon (1985) presented a fast method for these computations.

The Coefficient Matrix is Very Sparse

Each ray intersects only a few cells, hence A is very sparse.



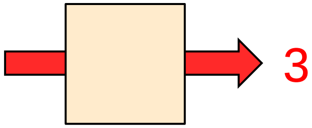


Many rows are structurally orthogonal, i.e., the zero/nonzero structure is such that their inner product is zero (they are orthogonal).

This sparsity plays a role in the convergence and the success of some of the iterative methods.

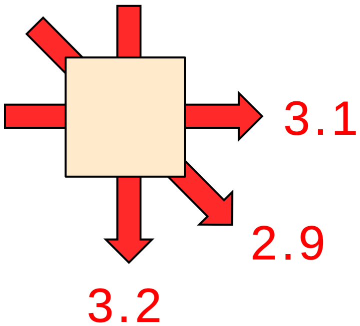
The Simplest Case: A Single Pixel

No noise:



$$1 \cdot x^* = 3$$

Now with noise in the measurements – least squares solution:

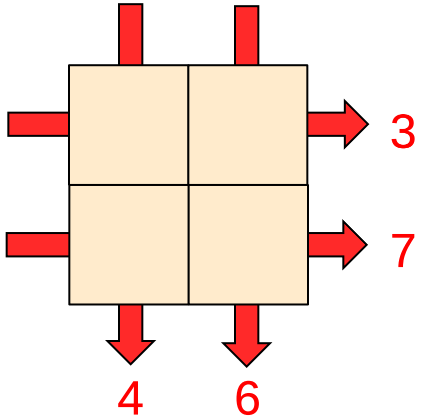


$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} x = \begin{pmatrix} 3.1 \\ 2.9 \\ 3.2 \end{pmatrix} \qquad x_{\mathrm{LSQ}} = A^{\dagger} b = 3.067$$

We know from statistics that cov(x_{LSQ}) is proportional to m^{-1}, where m is the number of data. So more data is better.

Let us immediately continue with a 2 × 2 image ...

Analogy: the “Sudoku” Problem – 数独



$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \\ 4 \\ 6 \end{pmatrix}$$

This matrix in rank deficient and there are infinitely many solutions (c = constant):

$$\begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} = \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline \end{array} + c \times \begin{array}{|c|c|} \hline -1 & 1 \\ \hline 1 & -1 \\ \hline \end{array}$$

Prior: solution is integer and non-negative

0	3
4	3

1	2
3	4

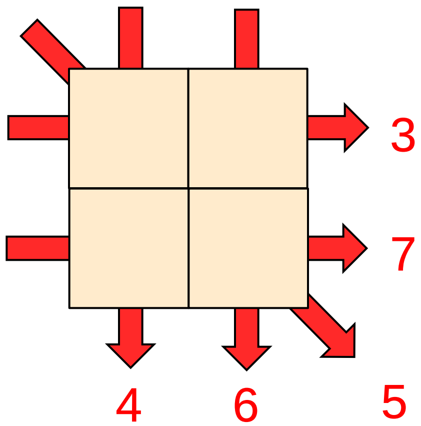
2	1
2	5

3	0
1	6

More Rays is Better

With *enough rays*, the problem has a unique solution.

Here, one more ray is enough to ensure a full-rank matrix:



$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \\ 4 \\ 6 \\ 5 \end{pmatrix}$$

The solution is now unique but it is still sensitive to the noise in the data.

The “difficulties” associated with the discretized tomography problem are closely linked with properties of the matrix *A*:

- The sensitivity of the solution *x* to data errors is characterized by cond(*A*), the condition number of *A*, defined as cond(A) = $\| A \| \cdot \| A^{-1} \|$.
- Uniqueness of the solution *x* is characterized by rank(*A*), the rank of the matrix *A* (the number of linearly independent rows or columns).

Characterization of Noise Sensitivity

Assume that *A* has full rank, and consider the two problems:

$$A\,\bar{x} = \bar{b} \text{ (no noise)} \qquad\qquad A\,x \approx b = \bar{b} + e$$

Perturbation theory gives an upper bound for the solution errors:

$$\frac{\|\bar{x} - x_{\mathrm{LSQ}}\|_2}{\|\bar{x}\|_2} \;\square\; \mathrm{cond}(A) \cdot \frac{\|e\|_2}{\|\bar{b}\|_2}$$

If cond(A) is too large for our liking, then we must modify the way we compute our solution – such that the modified solution is less sensitive.

The is at the heart of regularization!

SVD Analysis – How to Reduce Sensitivity

Recall the two relations:

$$x_{\mathrm{LSQ}} = A^\dagger b = (A^T A)^{-1} A^T b, \qquad f = R^{-1} \hat{f}.$$

We introduce the singular value decomposition (SVD):

$$A = U\,\Sigma\,V^T = \sum_i u_i\,\sigma_i\,v_i^T, \qquad U, V \text{ orthogonal}, \quad \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots).$$

Algebraic reconstruction

$$\begin{aligned} x_{\mathrm{LSQ}} &= A^T (U\,\Sigma^{-2} U^T)\,b \\ &= (V\,\Sigma^{-2} V^T)\,A^T b \end{aligned}$$

Tikhonov: filter the singular values

$$\Sigma^{-2} \;\to\; \Phi^2\,\Sigma^{-2}$$

$$\Phi = \Sigma^2\,(\Sigma^2 + \lambda^2 I)^{-1}$$

$$\mathrm{cond}(A) = \sigma_1/\sigma_n \;\to\; \sigma_1/\lambda$$

Inverse Radon transform

$$\begin{aligned} f &= R^* (-\Delta)^{1/2} \hat{f} \\ &= (-\Delta)^{1/2} R^* \hat{f} \end{aligned}$$

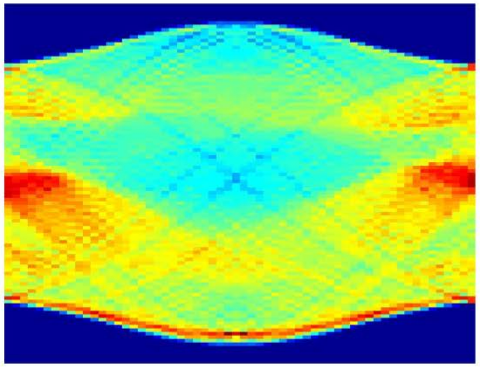
FBP: add a filter here in the frequency domain

$$|\omega| \;\to\; \Psi(\omega)\cdot|\omega|$$

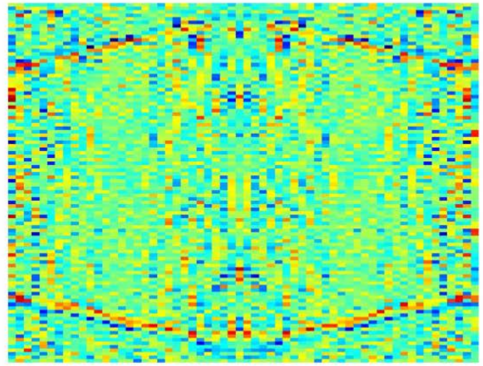
In both methods, we loose the details associated with high frequencies.

Matlab Check ...

Data: sinogram

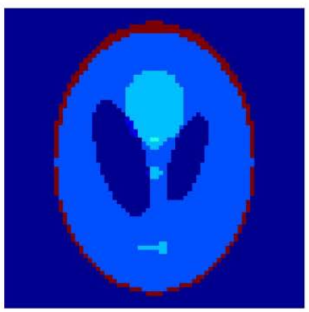


High-pass filtered sinogram

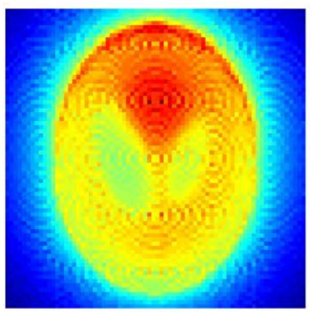


```
N = 3*24;
theta = 3:3:180;

[A,b,x] = paralleltomo(N,theta,[],N);
[U,S,V] = svd(full(A));

lt = length(theta);
Si = reshape(b,length(b)/lt,lt);

bf =  U*pinv(S)'*pinv(S)*U'*b;
SF = reshape(bf,length(b)/lt,lt);

Xbp = reshape(A'*b,N,N);
Xrec = reshape(A'*bf,N,N);
```

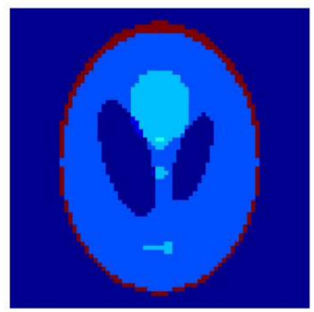
Phantom



Back-projection



Filtered back-projection



Dealing with a Nonunique Solution

The system $A\,x \approx b$ fails to have a unique solution when rank(A) < n, where n is the number of unknowns (the number of columns in A).

- This can happen when the distribution of rays is badly chosen (we saw an example of this in the 2 x 2 example).
- The more common situation is when we less data than unknowns (i.e., too few rays penetrating the object); this happens, e.g.,
  - if we need to reduce the X-ray dose,
  - or if we have limited time to perform the measurements.

Infinitely many solutions of the general forms:

$$x = x_0 + x_\perp, \quad x_\perp \in \mathcal{N}(A)$$

Minimum-norm solution:

$$x_{\rm MN} = A^\dagger b = A^T (A\,A^T)^{-1} b$$

$$x_{\rm MN} \in \mathcal{R}(A^T) \;\Rightarrow\; x_{\rm MN} \perp \mathcal{N}(A)$$

Radon – the limited-angle case:

$$f = f_0 + f_\perp, \quad f_\perp \in \mathcal{N}(R_{\rm la})$$

$$f = R_{\rm la}^* (-\Delta)^{1/2} \hat{f}$$

$$f \in \mathcal{R}(R_{\rm la}^*) \;\Rightarrow\; f_\perp = 0$$

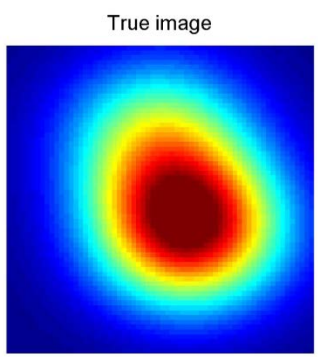
Appreciation of Minimum-Norm Solutions

The minimum-norm solution deals – in a way – in a very logical way with $\mathcal{N}(A)$, the null space of A: don't try to reconstructruct this component.

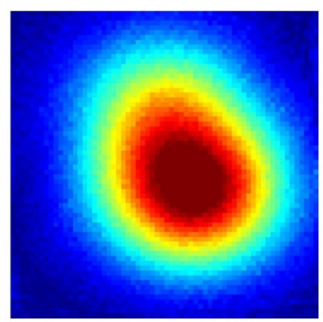
The same is true for the filtered solutions: the filter effectively dampens the highly-sensitive components corresponding to small singular values σ_i.

So: if the subspace $\mathcal{R}(A^T)$ captures the main features of the object to be reconstructed, then this is a good approach.

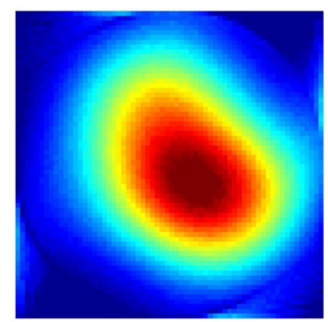
Example: underdetermined, limited-angle problem, angles 5,10,15,…,120.



A smooth test image



Tikhonov



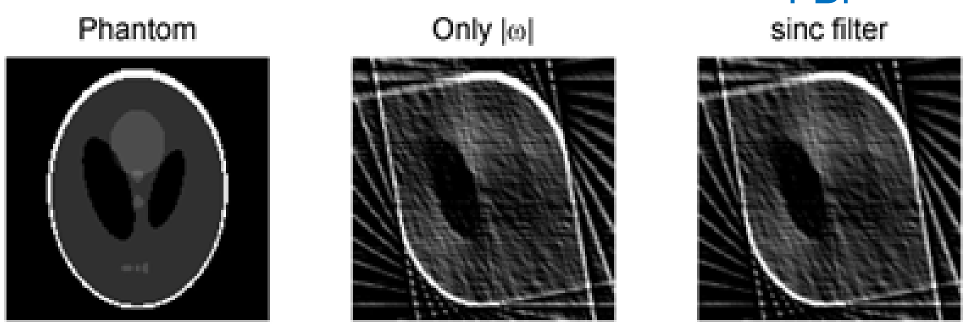
Another filtered solution: Cimmino

Critique of Minimum-Norm Solutions

The minimum-norm solution – while mathematically "nice" – is not guaranteed to provide a good reconstruction; it "misses" information in the $\mathcal{N}(A)$ or $\mathcal{N}(R_{\mathrm{la}}^*)$.

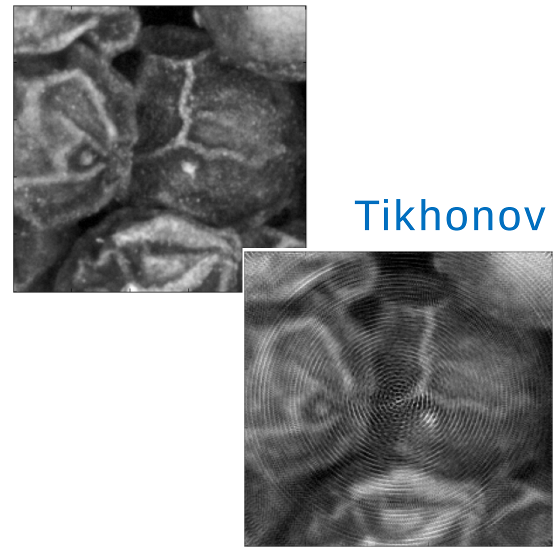
Examples: underdetermined, limited-angle problems.

FBP FBP



True

Tikhonov



Notice that for the limited-angle problem, FBP misses certain geometric structures in the image, associated with the missing projection angles.

These structures are precisely those in the null space of R_{la}, see, e.g.:

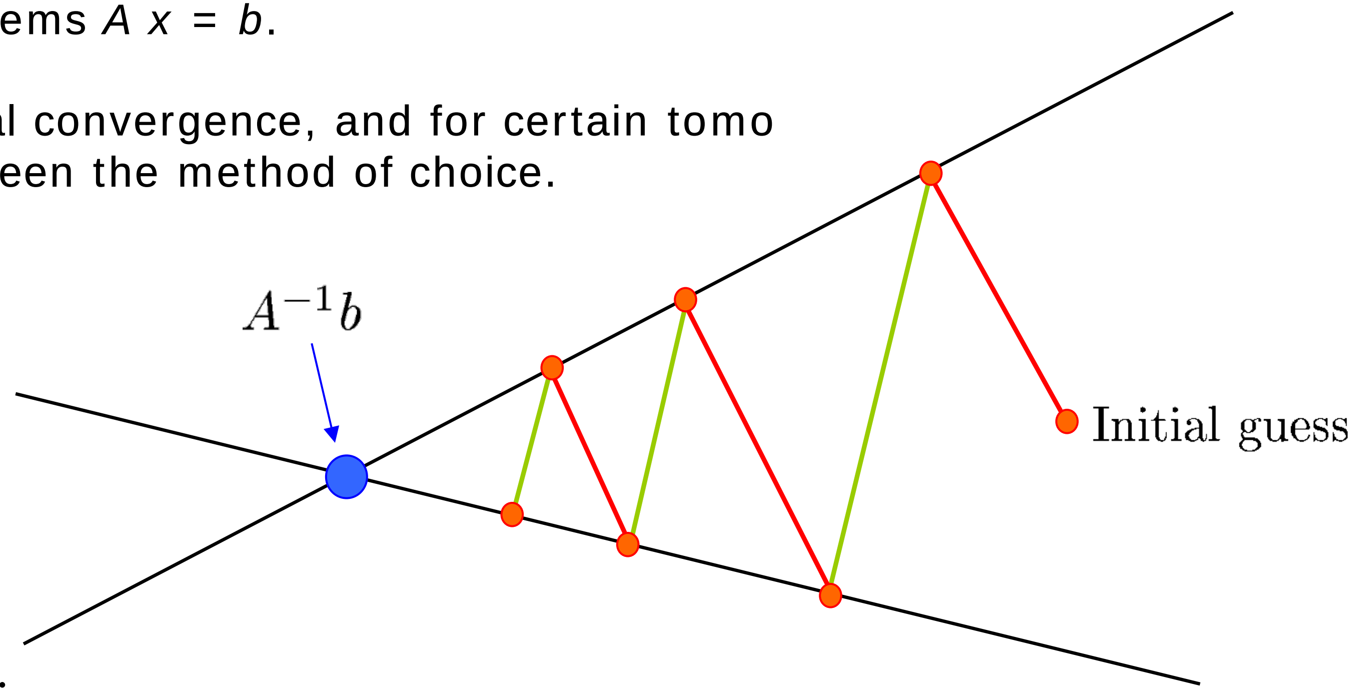
- Jürgen Frikel, *Sparse regularization in limited angle tomography*, Appl. Comput. Harmon. Anal., 34 (2013), 117–141.

We can find other ways to deal with effectively underdetermined problems!

Approach 1: ART

The Algebraic Reconstruction Method (*ART*) – also known as *Kaczmarz*'s method – was originally developed to solve full-rank square problems $A\,x = b$.

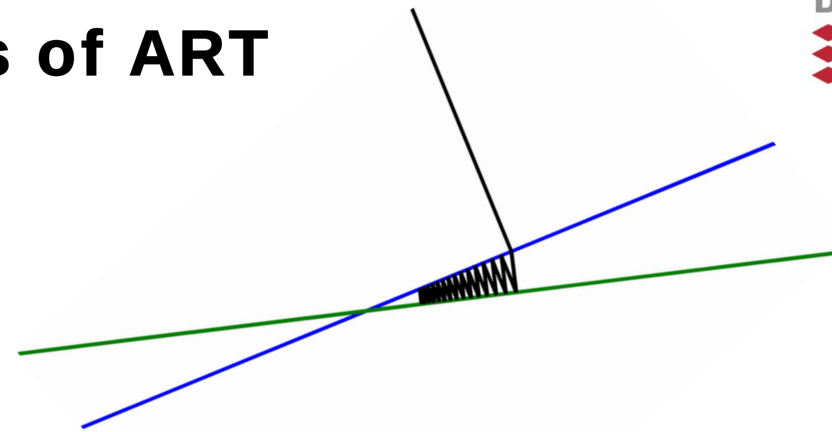
ART has fast initial convergence, and for certain tomo problems is has been the method of choice.



$$
\begin{aligned}
&\text{for } k = 1, 2, 3, \ldots \\
&\qquad i = k \bmod (\#\ \text{rows}) \\
&\qquad x^{k+1} = x^k + \omega \, \frac{b_i - a_i^T x^k}{\|a_i\|_2^2} \, a_i \qquad a_i^T = i\text{th row of } A \\
&\text{end}
\end{aligned}
$$

Regularizing Properties of ART

After some iterations the method slows down – and at this time we are often “close enough” to the desired solution.



During the first iterations, the iterates x^k capture “important” information in b, associated with the exact data $\bar{b} = A\,\bar{x}$.

- In this phase, the iterates x^k approach the exact solution $\bar{x}$.

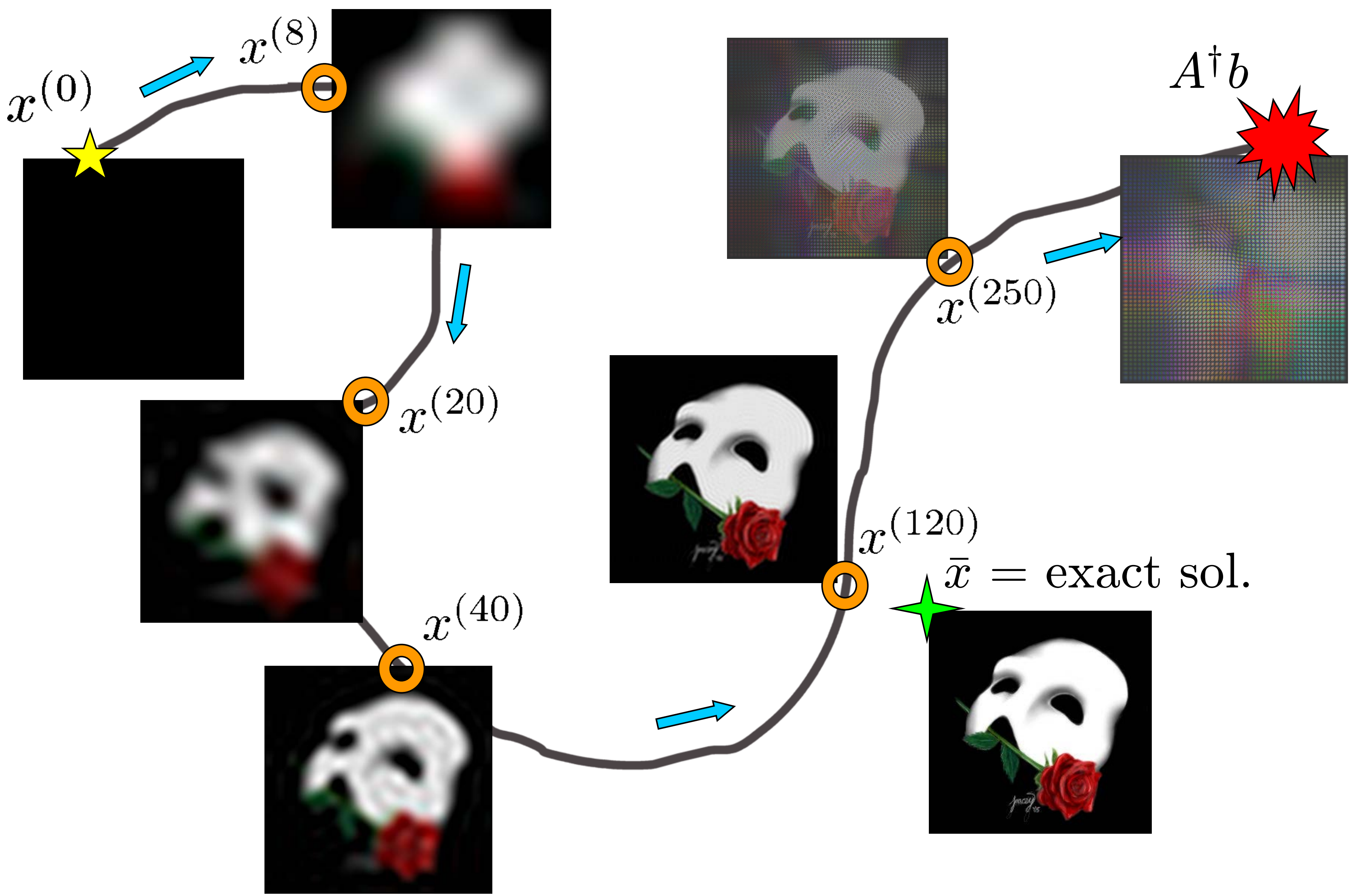
At later stages, the iterates starts to capture undesired noise components.

- Now the iterates x^k diverge from the exact solution and they approach the undesiredleast squares solution x_{LSQ}.

This behavior is called *semi-convergence*, a term coined by Natterer (1986):

”... even if [the iterative method] provides a satisfactory solution after a certain number of iterations, it deteriorates of the iteration goes on.”

Illustration of Semi-Convergence

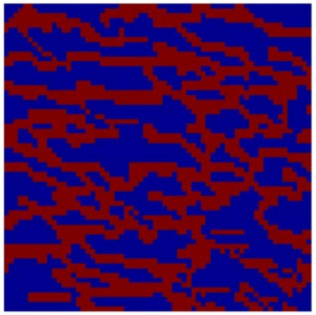


Appreciation of ART

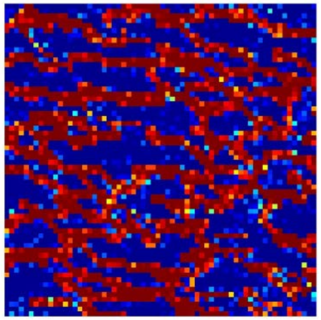
Experience shows that ART can give great solutions, with surprisingly many details. It does **not** produced filtered solutions.

Example with a binary test image:

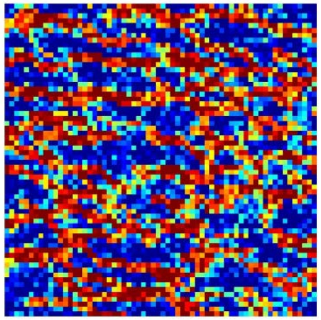
True image



ART



Tikhonov



In this example we have utilized that ART can incorporate inequality or box constraints – here we required pixel values between 0 and 1.

Many users of ART don't notice the semi-convergence – basically because the method dramatically "slows down" at this stage.

Appreciation of ART – Contd.

Why can ART give so good solutions with high-frequency components?

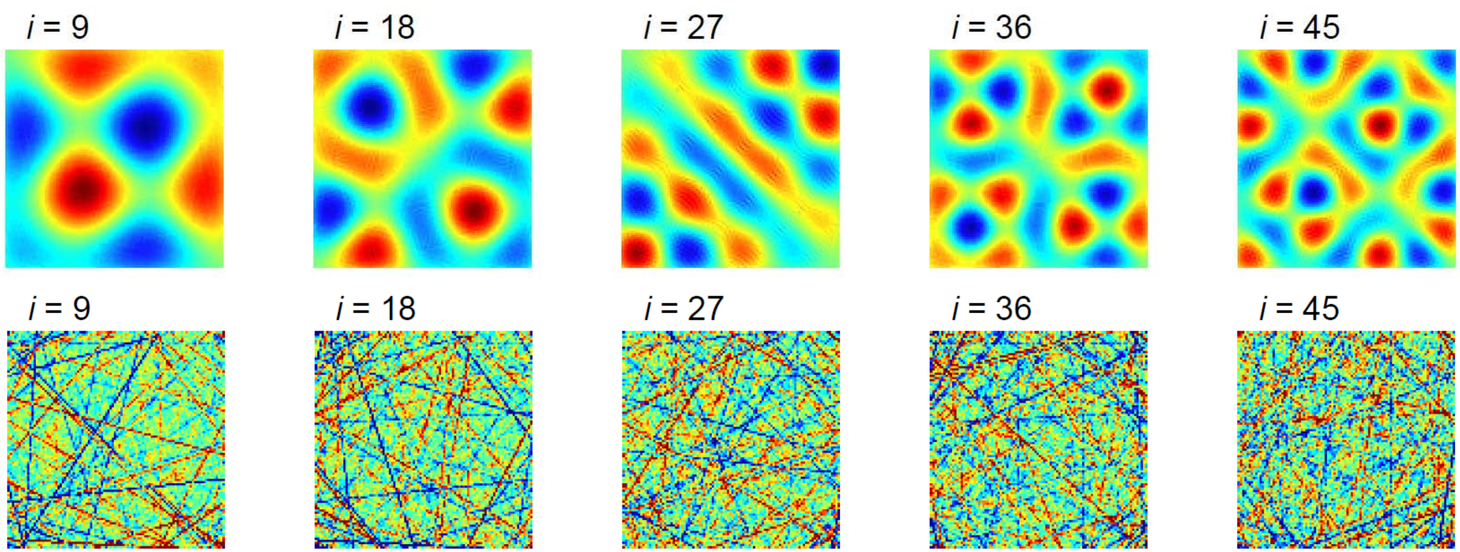
- It does not correspond to spectral filtering.
- It includes components in the null space, which may be desirable.
- A full theoretical understanding of its superiority is still missing …

Towards some insight.

A certain variant, *Symmetric ART*, can actually be expressed in a certain ortho-normal basis – and this basis includes the needed high-frequency components!

SVD basis

Basis for *sym. ART*

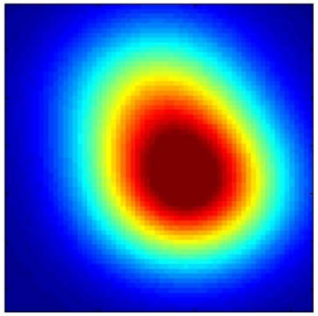


Critique of ART

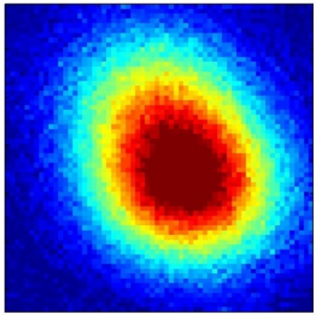
The reconstruction and regularization properties of ART are solely associated with the semi-convergence, and not suited for all types of problems.

Example with a smooth test image:

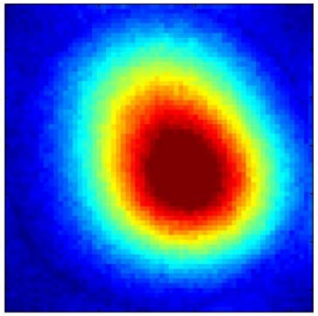
True (smooth)



ART



Tikhonov



There is also a need for more general regularization methods ... next slide.

Approach 2: Variational Regularization

In these methods, the regularization is explicit in the formulation of the problem to be solved:

$$\min_x \{\text{misfit}(A,b,x) + \lambda\cdot\text{penalty}(L,x)\} \qquad \text{subject to} \qquad x \in \mathcal{C}$$

Different noise:
Gaussian: $\|b - A\,x\|_2^2$
Laplace: $\|b - A\,x\|_1$
Poisson: $\|\text{diag}(\log(Ax)b - A\,x\|_1$
Etc.

Norm/energy: $\|x\|_2^2$
Flatness : $\|L_1\,x\|_2^2$
Roughness : $\|L_2\,x\|_2^2$
Piecewise smooth : $\|L_1\,x\|_1$
Etc.

Give smooth solutions

Let's look at this case, known as Total Variation (TV)

Nonnegativity: $x \geq 0$
Box constr.: $\ell \square x \square u$
Etc.

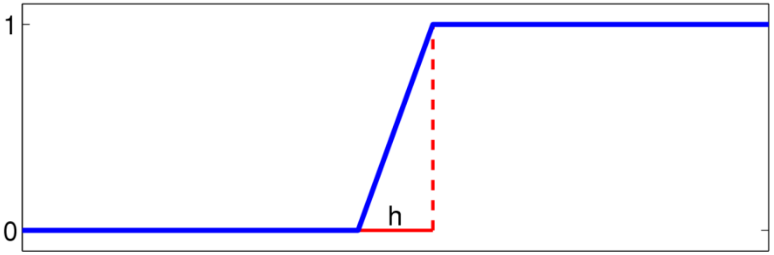
Total Variation Allows Steep Gradients

1-D continuous formulation:

$$TV(g) = \left\| g' \right\|_1 = \int_\Omega |g'(t)| \, dt$$

Example (2-norm penalizes steep gradients, TV doesn't):

$$\begin{aligned} TV(g) &= 1 \\ \|g'\|_2^2 &= \int_\Omega g'(t)^2 dt = 1/h \end{aligned}$$



2-D and 3-D continuous TV formulations:

$$TV(g) = \left\| \, \|\nabla g\|_2 \, \right\|_1 = \int_\Omega \|\nabla g(\mathbf{t})\|_2 \, d\mathbf{t}$$

TV Produces a Sparse Gradient Magnitude

Underlying assumption or prior knowledge: the image consists (approx.) of regions with constant intensity.

Hence the gradient magnitude (2-norm of gradient in each pixel) is sparse.

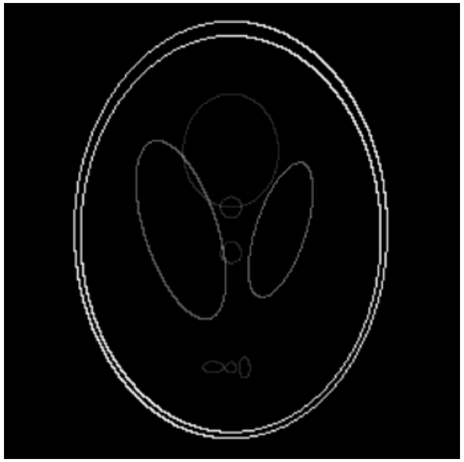
TV = 1-norm of the gradient magnitude,

= sum of 2-norm of gradients.

Experience shows that the TV prior is often so "strong" that it can compensate for a reduced amount – or quality – of data.

image

gradient magnitude



3 % non-zeros

Conclusions – What to Take Home

We demonstrated that tomographic reconstruction problems (inverse problems) have stability problems:

- The solution is always sensitive to noise.
- The solution may not be unique.

We looked at different common reconstruction algorithms and explained how the incorporate regularization:

- Filtered back projection – via a low-pass filter
- Tikhonov – via filtering of SVD components
- ART (Kaczmarc) – by stopping the iterations (semi-convergence)
- Variational methods – the regularization is explicit.

We saw that all these algorithms have their advantages and disadvantages, and introduce different artifacts in the solutions.