The Mathematics of X-ray Tomography

Tatiana A. Bubba
Department of Mathematics and Statistics, University of Helsinki
tatiana.bubba@helsinki.fi

Summer School on Very Finnish Inverse Problems
Helsinki, June 3-7, 2019

fips Finnish Inverse Problems Society

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE

Finnish Centre of Excellence in Inverse Modelling and Imaging
2018-2025

TSVD as Spectral Filtering

We can regard the TSVD also as the result of a filtering operation, namely:

$$\boldsymbol{f}^{\mathsf{TSVD}} = \sum_{i=1}^{r} \frac{\boldsymbol{u}_i^T \boldsymbol{y}^\delta}{\sigma_i} \boldsymbol{v}_i = \sum_{i=1}^{\min(m,n)} \phi_i^{\mathsf{TSVD}} \frac{\boldsymbol{u}_i^T \boldsymbol{y}^\delta}{\sigma_i} \boldsymbol{v}_i$$

where r is the truncation parameter and

$$\phi_i^{\mathsf{TSVD}} = \begin{cases} 1 & i = 1, \ldots, r \\ 0 & \text{elsewhere} \end{cases}$$

are the **filter factors** associated with the method.

These are called **spectral filtering** methods because the SVD basis can be regarded as a spectral basis, since the vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are the eigenvectors of $\mathbf{K}^T\mathbf{K}$ and $\mathbf{K}\mathbf{K}^T$.

The Tikhonov Method

Let's now consider the following filter factors:

$$\phi_i^{\mathsf{TIKH}} = \begin{cases} \frac{\sigma_i^2}{\sigma_i^2+\alpha^2} & i = 1, \dots, \min(m,n) \\ 0 & \text{elsewhere} \end{cases}$$

which yield the reconstruction method:

$$\boldsymbol{f}^{\mathsf{TIKH}} = \sum_{i=1}^{\min(m,n)} \phi_i^{\mathsf{TIKH}} \frac{\boldsymbol{u}_i^T \boldsymbol{y}^\delta}{\sigma_i} \boldsymbol{v}_i = \sum_{i=1}^{\min(m,n)} \frac{\sigma_i \left(\boldsymbol{u}_i^T \boldsymbol{y}^\delta\right)}{\sigma_i^2 + \alpha^2} \boldsymbol{v}_i .$$

This choice of the filters result in a regularization technique called **Tikhonov method** and $\alpha > 0$ is the so-called **regularization parameter**.

The parameter α acts in the same way as the parameter r in the TSVD method: it controls which SVD components we want to damp or filter.

Tikhonov Regularization

Similarly to SVD being the solution of the least squares problem, also Tikhonov regularization can be understood as the solution of a minimization problem:

$$\boldsymbol{f}^{\mathsf{TIKH}} = \underset{\boldsymbol{f}}{\operatorname{argmin}} \left\{ \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \left\| \boldsymbol{f} \right\|_2^2 \right\}.$$

This problem is motivated by the fact that we clearly want $\left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2$ to be small, but we also wish to avoid that it becomes zero. Indeed, by taking the Moore-Pensore solution $\boldsymbol{f}^{\dagger}$ we would have

$$\|\boldsymbol{f}^{\dagger}\|_2^2 = \sum_{i=1}^{k} \frac{(\boldsymbol{u}_i^T \boldsymbol{y}^{\delta})^2}{\sigma_i^2}$$

which could become unrealistically large when the magnitude of the noise in some direction $\boldsymbol{u}_i$ greatly exceeds the magnitude of the singular value σ_i.

The above minimization problem ensures that both the norm of the residual $\mathbf{K}\boldsymbol{f}^{\mathsf{TIKH}} - \boldsymbol{y}^{\delta}$ and the norm of the solution $\boldsymbol{f}^{\mathsf{TIKH}}$ are somewhat small and α balances the trade-off between the two terms.

Normal Equation and Stacked Form for Tikhonov Regularization

The Tikhonov solution can be also formulated as a linear least squares problem:

$$\boldsymbol{f}^{\mathsf{TIKH}} = \operatorname*{argmin}_{\boldsymbol{f}} \left\| \begin{bmatrix} \mathbf{K} \\ \sqrt{\alpha}\,\mathbb{1} \end{bmatrix} - \begin{bmatrix} \boldsymbol{y}^{\delta} \\ 0 \end{bmatrix} \right\|_2^2 .$$

This is called **stacked form**. If we denote by $\widetilde{\mathbf{K}} = \begin{bmatrix} \mathbf{K} \\ \sqrt{\alpha}\,\mathbb{1} \end{bmatrix}$ and $\widetilde{\boldsymbol{y}^{\delta}} = \begin{bmatrix} \boldsymbol{y}^{\delta} \\ 0 \end{bmatrix}$ then the least square solution of the stacked form satisfies the normal equations:

$$\widetilde{\mathbf{K}}^T \widetilde{\mathbf{K}}\, \boldsymbol{f} = \widetilde{\mathbf{K}}^T\, \widetilde{\boldsymbol{y}^{\delta}}.$$

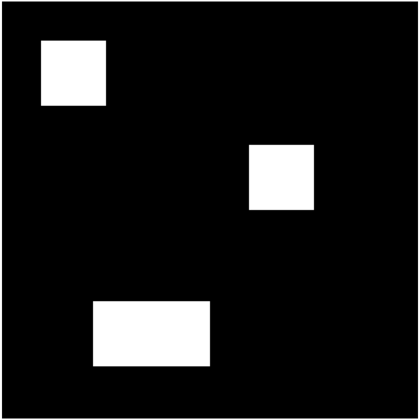
It is easy to check that

$$\widetilde{\mathbf{K}}^T \widetilde{\mathbf{K}} = \mathbf{K}^T\mathbf{K} + \alpha\,\mathbb{1} \qquad \text{and} \qquad \widetilde{\mathbf{K}}^T\, \widetilde{\boldsymbol{y}^{\delta}} = \mathbf{K}^T \boldsymbol{y}^{\delta}.$$

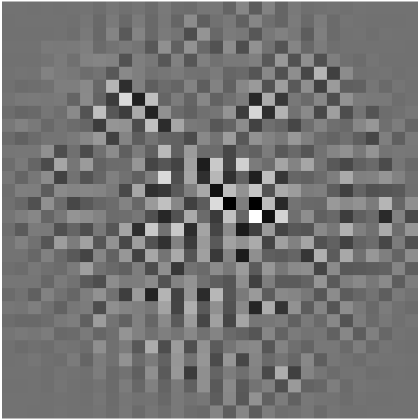
Hence we also have

$$\boldsymbol{f}^{\mathsf{TIKH}} = (\mathbf{K}^T\mathbf{K} + \alpha\,\mathbb{1})^{-1}\mathbf{K}^T \boldsymbol{y}^{\delta}.$$

Naive Reconstruction (Moore-Penrose Pseudoinverse)

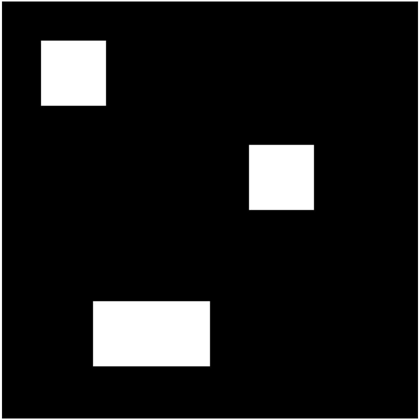


Original phantom

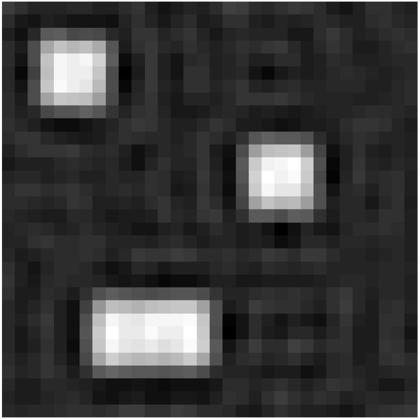


$f^\dagger$: RE $= 100\%$

Truncated SVD Regularization

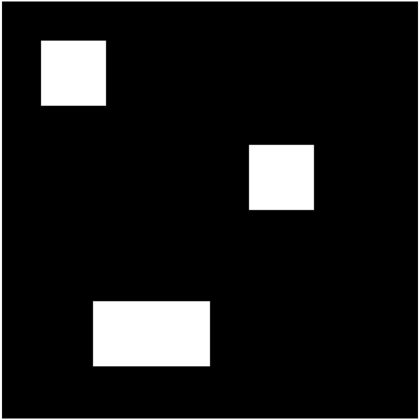


Original phantom

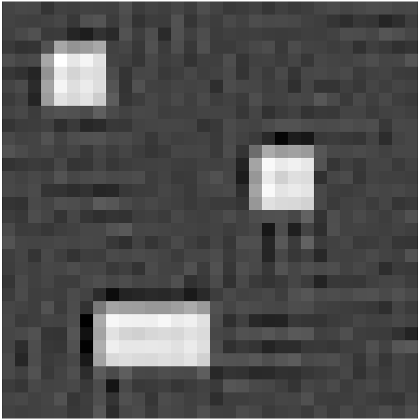


f^{TSVD}: RE = 35%

Tikhonov Regularization

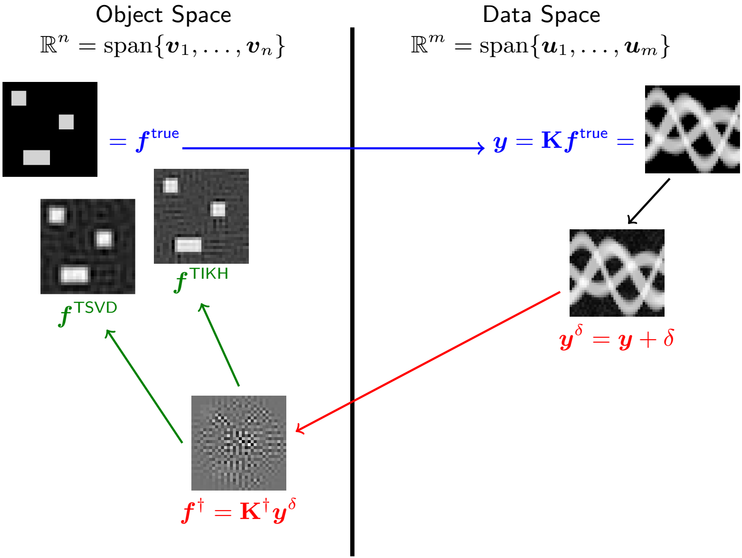


Original phantom



$\boldsymbol{f}^{\mathsf{TIKH}}$: RE $= 32\%$

Where Tikhonov Solution Stands in The Geometry of Ill-Conditioned Problems



About the Regularization Parameter

By looking at the minimization problem formulation of the Tikhonov solution

$$\boldsymbol{f}^{\mathsf{TIKH}} = \operatorname*{argmin}_{\boldsymbol{f}} \left\{ \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \left\| \boldsymbol{f} \right\|_2^2 \right\}$$

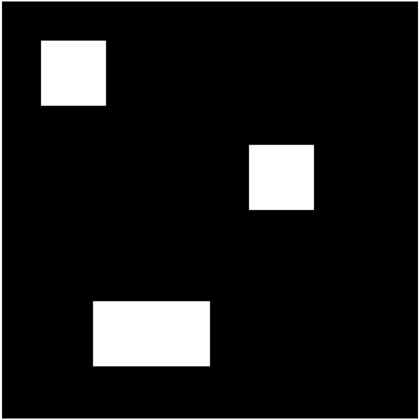
it is clear that:

- a **large** α results in strong regularity and possible over smoothing
- a **small** α small yields a good fitting, with the risk of over fitting.

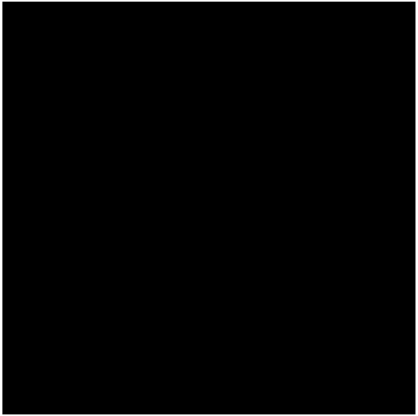
In general, choosing the regularization parameter for an ill-posed problem is not a trivial task and there are no rule of thumbs. Usually, it is a combination of good heuristics and prior knowledge of the noise in the observations.

Delving into this is out of the scope, but there are methods that can be found in the literature (Morozov's discrepancy principle, generalized cross validation, L-curve criterion), and more recent approaches tailored to specific problems.

Influence of the Choice of α in Tikhonov Regularization

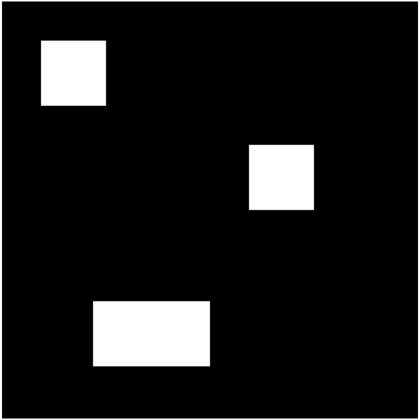


Original phantom

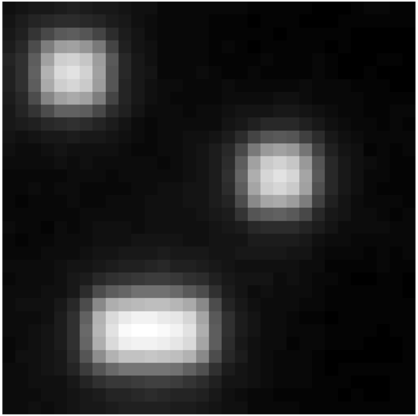


$\boldsymbol{f}^{\mathsf{TIKH}}$: $\alpha = 10^3$

Influence of the Choice of α in Tikhonov Regularization

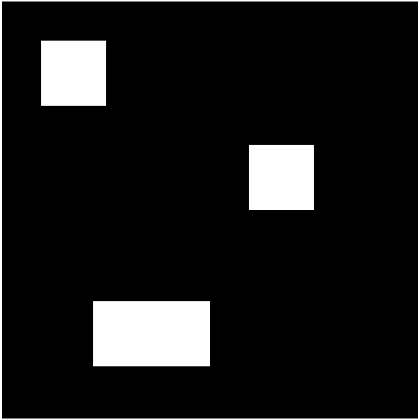


Original phantom

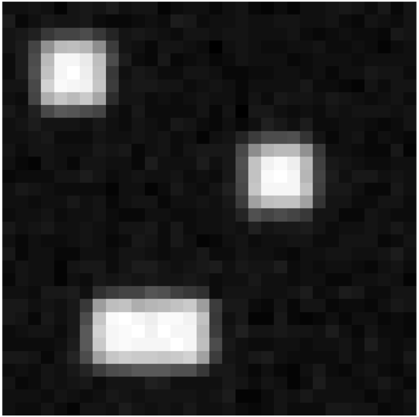


$\boldsymbol{f}^{\mathsf{TIKH}}$: $\alpha = 10^2$

Influence of the Choice of α in Tikhonov Regularization

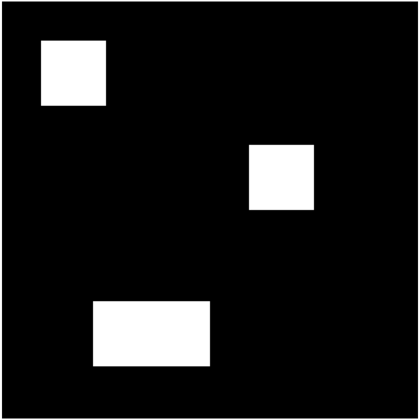


Original phantom

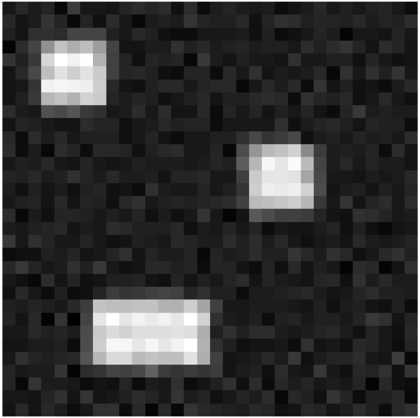


$\boldsymbol{f}^{\mathsf{TIKH}}$: $\alpha = 10$

Influence of the Choice of α in Tikhonov Regularization

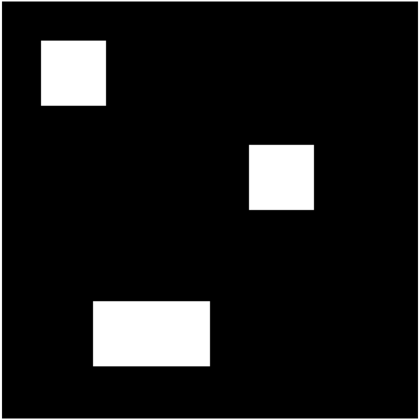


Original phantom

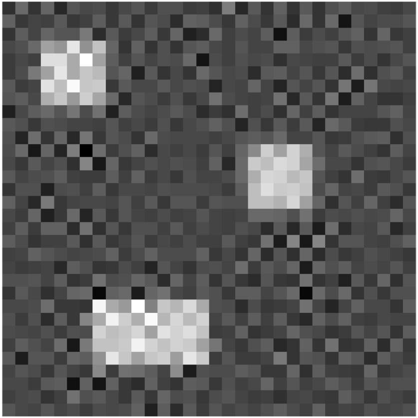


$\boldsymbol{f}^{\mathsf{TIKH}}$: $\alpha = 1$

Influence of the Choice of α in Tikhonov Regularization

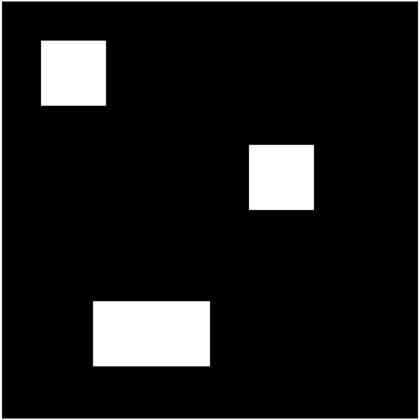


Original phantom

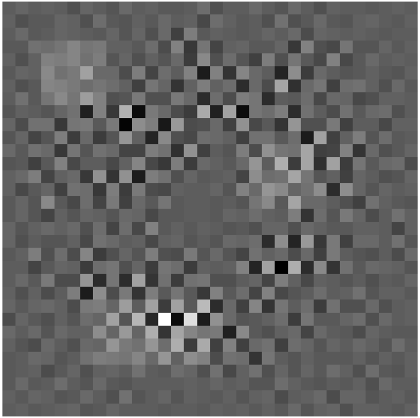


$\boldsymbol{f}^{\mathsf{TIKH}}$: $\alpha = 10^{-1}$

Influence of the Choice of α in Tikhonov Regularization

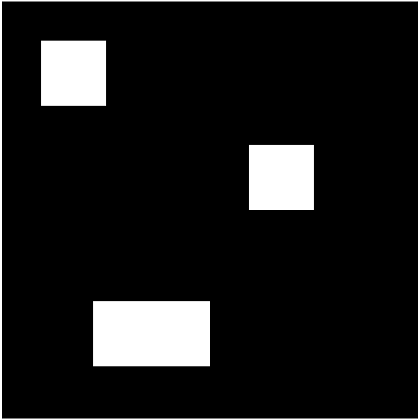


Original phantom

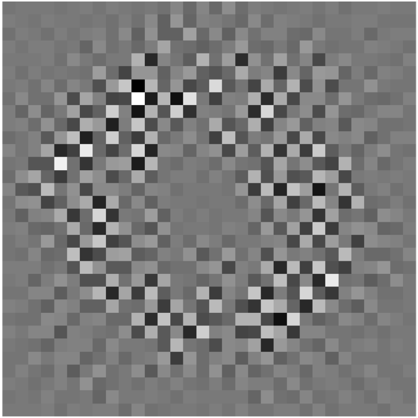


$\boldsymbol{f}^{\mathsf{TIKH}}$: $\alpha = 10^{-2}$

Influence of the Choice of α in Tikhonov Regularization



Original phantom



$\boldsymbol{f}^{\mathsf{TIKH}}$: $\alpha = 10^{-3}$

Generalized Tikhonov Regularization

Sometimes we have *a priori* information about the solution of the inverse problem. This can be incorporated in the minimization formulation of the Tikhonov method. For instance:

- $\boldsymbol{f}$ is close to a know $\boldsymbol{f}^*$

$$\boldsymbol{f}^{\text{GTIKH}} = \underset{\boldsymbol{f}}{\operatorname{argmin}} \left\{ \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \left\| \boldsymbol{f} - \boldsymbol{f}^* \right\|_2^2 \right\}$$

- $\boldsymbol{f}$ is known to be smooth

$$\boldsymbol{f}^{\text{GTIKH}} = \underset{\boldsymbol{f}}{\operatorname{argmin}} \left\{ \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \left\| L\boldsymbol{f} \right\|_2^2 \right\}$$

- $\boldsymbol{f}$ has similar smoothing properties as $\boldsymbol{f}^*$

$$\boldsymbol{f}^{\text{GTIKH}} = \underset{\boldsymbol{f}}{\operatorname{argmin}} \left\{ \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \left\| L(\boldsymbol{f} - \boldsymbol{f}^*) \right\|_2^2 \right\}$$

where L is a suitable operator.

Generalized Tikhonov Regularization

A common choice for generalized Tikhonov regularization is to take L as a discretized differential operator. For example, using forward differences:

$$L = \frac{1}{\Delta s}\begin{pmatrix} -1 & 1 & 0 & 0 & 0 & & \dots & 0 \\ 0 & -1 & 1 & 0 & 0 & & \dots & 0 \\ 0 & 0 & -1 & 1 & 0 & & \dots & 0 \\ \vdots & & & & \ddots & & & \vdots \\ \vdots & & & & & \ddots & & \vdots \\ 0 & & \dots & & 0 & -1 & 1 & 0 \\ 0 & & \dots & & 0 & 0 & -1 & 1 \\ 1 & & \dots & & 0 & 0 & 0 & -1 \end{pmatrix}$$

where Δs is the length of the discretization interval.

This choice promotes **smoothness** in the reconstruction.

Variational Regularization

In general, a minimization problem of the form:

$$\Gamma_\alpha(\boldsymbol{y}^\delta) = \operatorname*{argmin}_f \left\{ \frac{1}{2} \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^\delta \right\|_2^2 + \alpha\, \mathcal{R}(\boldsymbol{f}) \right\}$$

is called **variational formulation**:

- The data fidelity (or data fitting) term $\left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^\delta \right\|_2^2$ keeps the estimation of the solution close to the data under the forward physical system.
- The regularization parameter $\alpha > 0$ controls the trade-off between a good fit and the requirements from the regularization.
- $\mathcal{R}(\boldsymbol{f})$ incorporates **a priori** information or assumptions on the unknown $\boldsymbol{f}$. A **non** exhaustive list:
  - Tikhonov regularization: $\|\boldsymbol{f}\|_2^2$
  - Generalized Tikhonov regularization: $\|L\boldsymbol{f}\|_2^2$
  - Compress sensing or sparse regularization: $\|\boldsymbol{f}\|_0$ or $\|\boldsymbol{f}\|_1$ or $\|L\boldsymbol{f}\|_1$
  - Indicator functions of constraints sets: $\iota_{\mathbb{R}_+}(\boldsymbol{f})$
  - A combination of the above

ℓ^p Norms for $\mathbb{R}^n$

Let $\boldsymbol{f} \in \mathbb{R}^n$. The ℓ^p norms for $1 \leq p < \infty$ are defined by

$$\|\boldsymbol{f}\|_p = \left(\sum_{j=1}^{n} |\boldsymbol{f}_j|^p \right)^{1/p}.$$

Also important, but **not** a norm:

$$\|\boldsymbol{f}\|_0 \; = \; \lim_{p \to 0} \|\boldsymbol{f}\|_p^p \; = \; \big|\{j : f_j \neq 0\}\big|.$$

The ℓ^0 "norm" counts the number of non-zeros components in $\boldsymbol{f}$: this is used to measure sparsity.

Sparse Regularization

Finding the sparsest solution:

$$\underset{\boldsymbol{f}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta}\|_2^2 + \alpha \|L\boldsymbol{f}\|_0 \right\}$$

is known as Compressed Sensing (CS). However, the problem above is NP-hard, since it requires a combinatorial search of exponential size for considering all possible supports.

Under certain conditions on $L\boldsymbol{f}$ and $\mathbf{K}$, replacing ℓ^0 with ℓ^1 yields "similar" results. This relaxation leads to a convex problem:

$$\underset{\boldsymbol{f}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta}\|_2^2 + \alpha \|L\boldsymbol{f}\|_1 \right\}.$$

which is at the basis of optimization-based methods for CS.

About the Convex Relaxation

The formulation

$$\operatorname*{argmin}_{\boldsymbol{f}} \left\{ \frac{1}{2} \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \| L\boldsymbol{f} \|_1 \right\}.$$

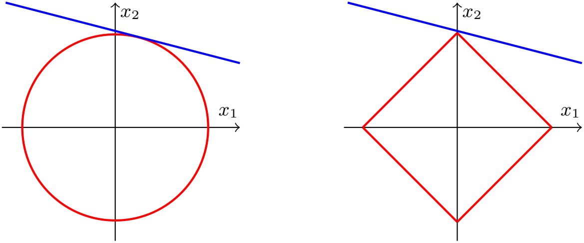
it is more easily solvable, but still **nonsmooth**. Also, it is **convex**, but not strictly convex. So why not using Tikhonov regularization?

About the Convex Relaxation

The formulation

$$\underset{\boldsymbol{f}}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \| L\boldsymbol{f} \|_1 \right\}.$$

it is more easily solvable, but still **nonsmooth**. Also, it is **convex**, but not strictly convex. So why not using Tikhonov regularization?



$|x_1|^2 + |x_2|^2 =$ const

$|x_1| + |x_2| =$ const

Total Variation Regularization

If we take $L = \nabla$ as the discrete differentiation matrix, the variational formulation

$$\boldsymbol{f}^{\mathsf{TV}} = \operatorname*{argmin}_{\boldsymbol{f}} \left\{ \frac{1}{2} \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \left\| \nabla \boldsymbol{f} \right\|_1 \right\}$$

is called **Total Variation**.

Total Variation (TV) regularization promotes sparsity in the derivative, in other words favouring piece-wise constantness.

TV, first introduced to face denoising problems (in 1992, the so-called ROF model) became a popular approach in many imaging processing tasks (including CT) due to its ability to **preserve or even favour reconstructions with sharp edges**.

Beyond Classical TV

- Total Generalized Variation (TGV^k)
  - defines a whole family of priors depending on the order of the derivative k
  - TGV^2 is suitable for piecewise smooth targets
- Total p-Variation (TpV)
  - $0 < p < 1$ refers to the norm
  - designs a nonsmooth and nonconvex problem
- Many many more: Higher Order TV, Directional TV, Anisotropic TV, ...

Wavelet-based Regularization

If we take $L = \mathbf{W}$ as the matrix associated to a certain wavelet transform, the variational formulation

$$\boldsymbol{f}^{\mathsf{WLET}} = \underset{\boldsymbol{f}}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \left\| \mathbf{W}\boldsymbol{f} \right\|_1 \right\}$$

promotes **sparsity on the wavelet coefficients**.

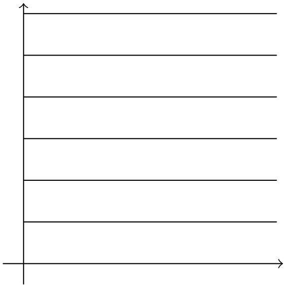
The idea behind wavelet-based regularization is that wavelet coefficients come with different magnitudes and the smallest ones are associated with noise. The ℓ_1-norm suppresses the small coefficients in favor of the largest ones, which are associated with edges and images dominant features.

Wavelets (widely used in image processing since 1990s) are a very common choice in CS approaches since they model images quite adequately.

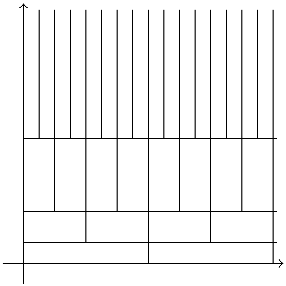
A Bit About Wavelets

Wavelets arose in 1980s to overcome some of the limitations of Fourier analysis.

Similarly to Fourier series, the idea is to "break" a signal into **building blocks**, but unlike Fourier series the building blocks are localized not only in the frequency domain but also in the space domain.



Time-frequency plane for the Fourier Transform.



Time-frequency plane for the wavelet transform.

A Bit About Wavelets

Different families of wavelets can be generated by considering different "parents":

- the scaling function $\phi \in L^2(\mathbb{R}^2)$, a low-pass filter, provides a rougher version of the signal itself;
- the (mother) wavelet $\psi \in L^2(\mathbb{R}^2)$, a high-pass filter, describes the details in the signal.

A wavelet system is generated by applying to both parents two operators:

- Isotropic dilation:

$$D_M \psi(x) = 2^{-\frac{j}{2}} \psi(2^j x),$$

  where $j \in \mathbb{Z}$ is the **scaling** parameter.

- Translation:

$$T_k \psi(x) = \psi(x - k)$$

  where $k \in \mathbb{Z}$ is the **location** parameter.

A Bit About Wavelets

The elements of a wavelet system are given by:

$$\psi_{jk}(x) = \{T_k D_j \psi(x) = 2^{-\frac{j}{2}}\psi(2^j x - k) \ : \ (j,k) \in \mathbb{Z} \times \mathbb{Z}\}$$

and similarly for the scaling function.

The wavelets coefficients are the result of the wavelet transform:

$$\mathcal{W} : f \ \longrightarrow \ \mathcal{W}f(j,k) = \langle f, \psi_{j,k} \rangle.$$

In practical applications these are computed using the language of filters, with convolutions and downsampling and upsampling operations.

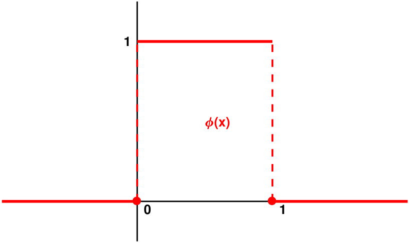
In particular in 2D, by considering tensor products, from the scaling and wavelet functions we get one scaling function but three wavelet functions (horizontal, vertical and diagonal):

$$\Phi(\boldsymbol{x}) = \phi(x_1)\phi(x_2),$$

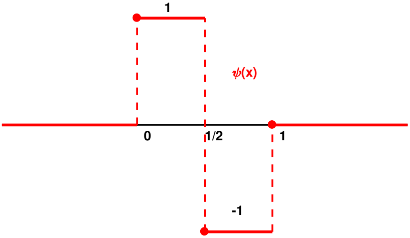
and

$$\Psi^1(\boldsymbol{x}) = \phi(x_1)\psi(x_2), \quad \Psi^2(\boldsymbol{x}) = \psi(x_1)\phi(x_2), \quad \Psi^3(\boldsymbol{x}) = \psi(x_1)\psi(x_2).$$

An Example: Haar Wavelets

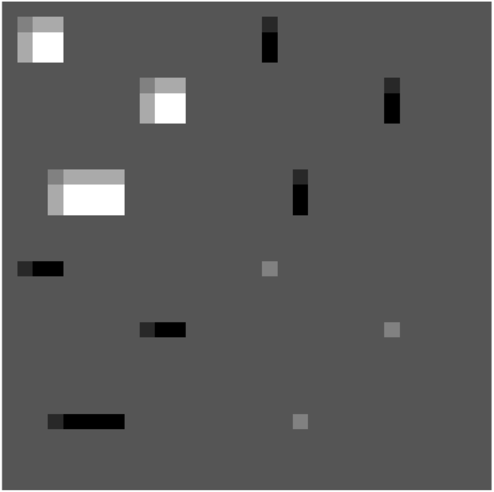


$$\phi(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$



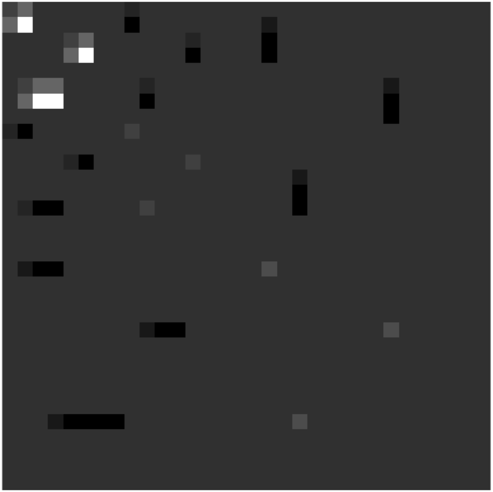
$$\psi(x) = \begin{cases} 1 & 0 \le x < \frac{1}{2} \\ 0 & \frac{1}{2} \le x < 1 \end{cases}$$

An Example: Haar Wavelet Transform of the Square Phantom



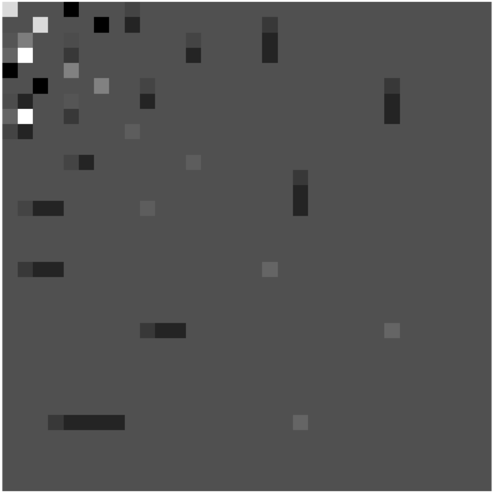
1-level Haar wavelet transform

An Example: Haar Wavelet Transform of the Square Phantom



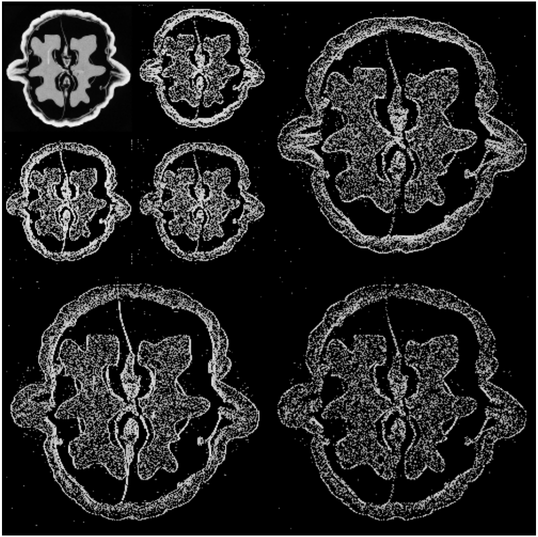
2-level Haar wavelet transform

An Example: Haar Wavelet Transform of the Square Phantom



3-level Haar wavelet transform

An Example: Haar Wavelet Transform of a Walnut



Constrained Regularization

In many cases, and CT is one of them, it is beneficial to include in the model a nonnegativity constraint:

$$\underset{\boldsymbol{f}}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \iota_{\mathbb{R}_+}(\boldsymbol{f}) \right\} \qquad \text{or} \qquad \underset{\boldsymbol{f} > 0}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 \right\},$$

where the inequality is meant component-wise.

The nonnegative constraint can also be coupled with other regularizers:

- Nonnegativity constrained Tikhonov regularization:

$$\boldsymbol{f}_+^{\mathsf{TIKH}} = \underset{\boldsymbol{f} > 0}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \left\| \boldsymbol{f} \right\|_2^2 \right\}$$

- Nonnegativity constrained sparse regularization:

$$\underset{\boldsymbol{f} > 0}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbf{K}\boldsymbol{f} - \boldsymbol{y}^{\delta} \right\|_2^2 + \alpha \left\| L\boldsymbol{f} \right\|_1 \right\}$$

How to Solve ℓ_1-type Problems?

- Approximating the absolute value function by

$$|t|_\beta = \sqrt{t^2 + \beta}.$$

Then the problem becomes smooth and we can use gradient-based minimization algorithms. This is often done for TV regularization (*smoothed TV*).

How to Solve ℓ_1-type Problems?

- Approximating the absolute value function by

$$|t|_\beta = \sqrt{t^2 + \beta}.$$

  Then the problem becomes smooth and we can use gradient-based minimization algorithms. This is often done for TV regularization (*smoothed TV*).

- Using algorithms for nonsmooth objective functions (primal-dual, forward-backward, Bregman iteration, ...). In general, these requires the computation of the proximal operator and depending wether there is or not an analytical closed form for it, the minimization problem can be rather challenging.

How to Solve ℓ_1-type Problems?

- Approximating the absolute value function by

$$|t|_\beta = \sqrt{t^2 + \beta}.$$

  Then the problem becomes smooth and we can use gradient-based minimization algorithms. This is often done for TV regularization (*smoothed TV*).

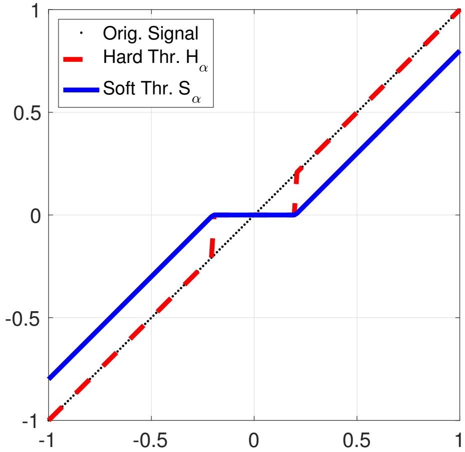
- Using algorithms for nonsmooth objective functions (primal-dual, forward-backward, Bregman iteration, ...). In general, these requires the computation of the proximal operator and depending wether there is or not an analytical closed form for it, the minimization problem can be rather challenging.

$\boldsymbol{f}^{\mathrm{TV}}$ and $\boldsymbol{f}^{\mathrm{WLET}}$ are special cases for which the proximal operator is easy and fast to compute, because is given by the **soft-thresholding** operator.

Hard- and Soft-Thresholding

$$S_\alpha(x) = \begin{cases} x + \frac{\alpha}{2} & \text{if } \; x \leq -\frac{\alpha}{2} \\ 0 & \text{if } \; |x| < \frac{\alpha}{2} \\ x - \frac{\alpha}{2} & \text{if } \; x \geq \frac{\alpha}{2} \end{cases}$$

$$H_\alpha(x) = \begin{cases} x & \text{if } \; x \leq -\frac{\alpha}{2} \\ 0 & \text{if } \; |x| < \frac{\alpha}{2} \\ x & \text{if } \; x \geq \frac{\alpha}{2} \end{cases}$$



Iterative Soft-Thresholding Algorithm (ISTA)

For instance, when $L = \mathbf{W}$ is the matrix associated with an orthogonal wavelet transform (*e.g.*, Haar wavelets), problems of the form:

$$\underset{\boldsymbol{f}\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}\left\|\mathbf{K}\boldsymbol{f} - \boldsymbol{y}^\delta\right\|_2^2 \; + \; \alpha \left\|\mathbf{W}\boldsymbol{f}\right\|_1 \right\}, \tag{1}$$

can be solved using an algorithm called Iterative Soft-Thresholding (ISTA) and the approximate solution is given by:

$$\boldsymbol{f}^{(i+1)} = \mathbf{W}^T S_\alpha \mathbf{W}\Big(\boldsymbol{f}^{(i)} + \mathbf{K}^T\big(\boldsymbol{y}^\delta - \mathbf{K}\boldsymbol{f}^{(i)}\big)\Big)$$

where S_α is the soft-thresholding operation.

Iterative Soft-Thresholding Algorithm (ISTA)

For instance, when $L = \mathbf{W}$ is the matrix associated with an orthogonal wavelet transform (*e.g.*, Haar wavelets), problems of the form:

$$\operatorname*{argmin}_{\boldsymbol{f}\in\mathbb{R}^n} \left\{ \frac{1}{2}\left\|\mathbf{K}\boldsymbol{f} - \boldsymbol{y}^\delta\right\|_2^2 \; + \; \alpha\left\|\mathbf{W}\boldsymbol{f}\right\|_1 \right\}, \tag{1}$$

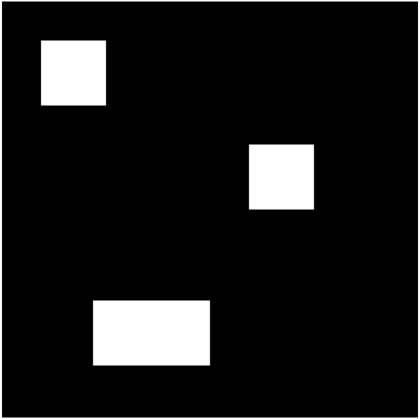
can be solved using an algorithm called Iterative Soft-Thresholding (ISTA) and the approximate solution is given by:

$$\boldsymbol{f}^{(i+1)} = \mathbf{W}^T S_\alpha \mathbf{W}\Big(\boldsymbol{f}^{(i)} + \mathbf{K}^T\big(\boldsymbol{y}^\delta - \mathbf{K}\boldsymbol{f}^{(i)}\big)\Big)$$

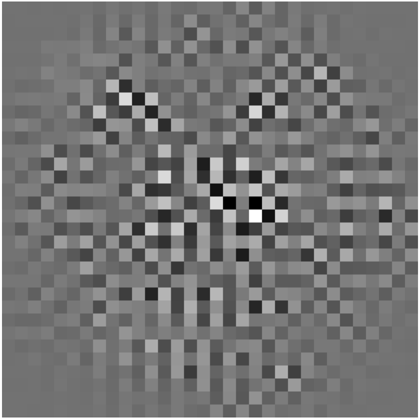
where S_α is the soft-thresholding operation.

There are many variants of ISTA to gain faster convergence (FISTA) or to extend it to non-orthogonal bases (or frames), or to include the non-negativity constraint (primal-dual fixed point, PDFP).

Naive Reconstruction (Moore-Penrose Pseudoinverse)

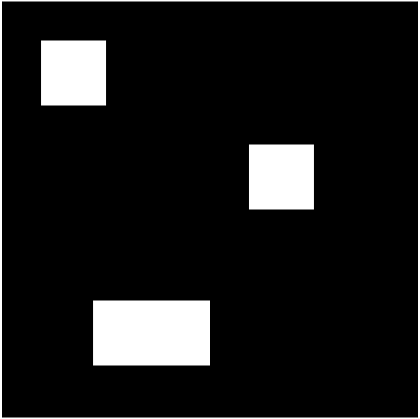


Original phantom

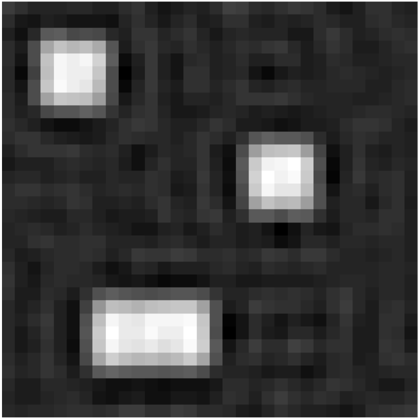


$\boldsymbol{f}^{\dagger}$: RE = 100%

Truncated SVD Regularization

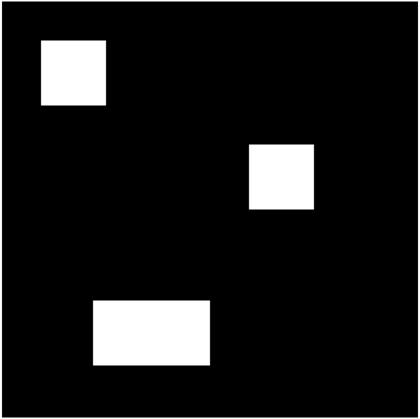


Original phantom

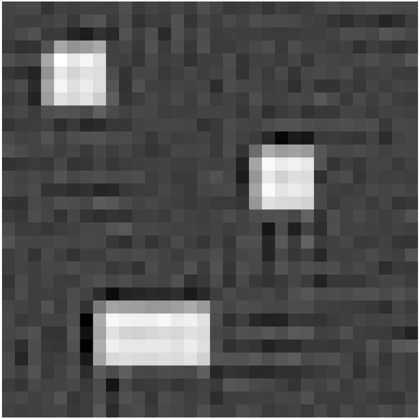


$\boldsymbol{f}^{\mathrm{TSVD}}$: RE $= 35\%$

Tikhonov Regularization

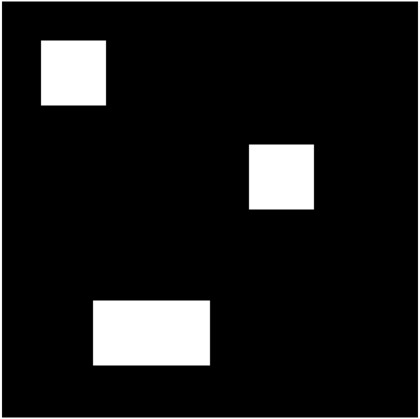


Original phantom

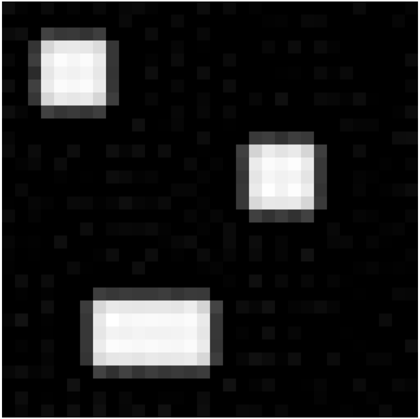


$\boldsymbol{f}^{\mathrm{TIKH}}$: RE = 32%

Nonnegativity Constrained Tikhonov Regularization

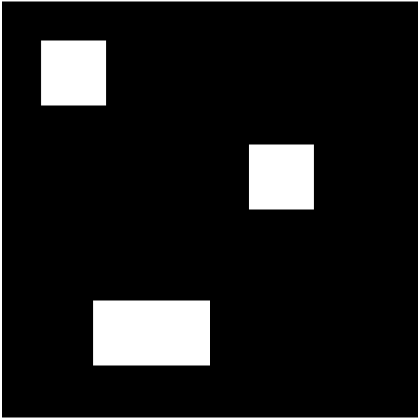


Original phantom

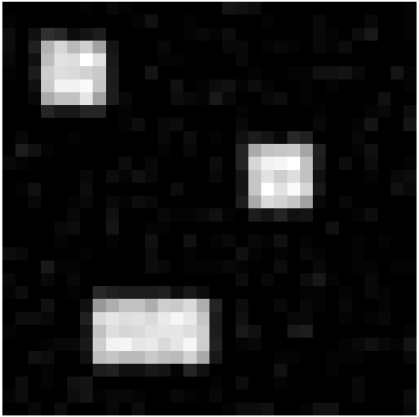


$\boldsymbol{f}_{+}^{\mathsf{TIKH}}$: RE = 13%

Nonnegativity Constrained Wavelet-based Regularization

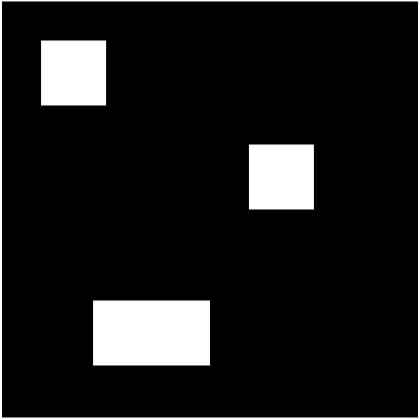


Original phantom

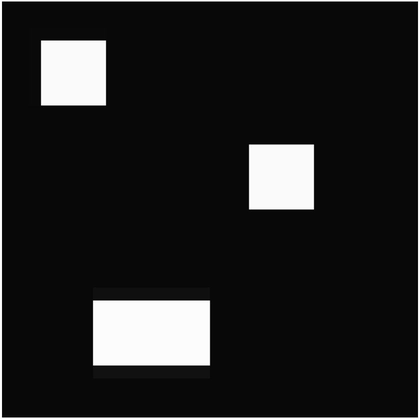


$\boldsymbol{f}_{+}^{\mathrm{WLET}}$: RE = 26%

Nonnegativity Constrained Total Variation Regularization



Original phantom



f_+^{TV}: RE $= 3\%$

Take-home message

- Uniqueness does not save us.
  Even with an injective forward map, failure of Hadamard's condition 3 means that we need regularization for solving the inverse problem.

- Non-uniqueness can be handled.
  Stable regularization strategy just needs enough *a priori* information for picking out a unique object among those with same data.

Take-home message

- Uniqueness does not save us.
  Even with an injective forward map, failure of Hadamard's condition 3 means that we need regularization for solving the inverse problem.

- Non-uniqueness can be handled.
  Stable regularization strategy just needs enough *a priori* information for picking out a unique object among those with same data.

Caveat. Regularization is not the only "cure" to ill-posedness: bayesian inversion and analytical strategies (designed *ad hoc* on the problem) are possible approaches as well.

Some References

- Inverse Problems:
  - Engl, Hanke & Neubauer, *Regularization of inverse problems*, 1996
  - Hansen, *Rank-deficient and discrete ill-posed problems*, 1998
  - Bertero & Boccacci, *Introduction to inverse problems in imaging*, 1998
  - Vogel, *Computational methods for inverse problems*, 2002
  - Hansen, *Discrete inverse problems*, 2010
  - Mueller & Siltanen, *Linear and Nonlinear Inverse Problems with Practical Applications*, 2012

- X-ray Tomography:
  - Deans, *The Radon Transform and Some of Its Applications*, 1983
  - Natterer, *The mathematics of computerized tomography*, 1986
  - Kak & Slaney, *Principles of computerized tomographic imaging*, 1988
  - Buzug: *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*, 2008
  - Natterer & Wübbeling, *Mathematical Methods in Image Reconstruction*, 2001
  - Epstein, *Introduction to the mathematics of medical imaging*, 2008

Some References

- Total variation and wavelets
  - Burger & Osher, *A guide to the TV zoo* (chapter 1 in Burger & Osher, *Level-Set and PDE-based Reconstruction Methods*, 2013)
  - Rudin, Osher & Fatemi, *Nonlinear total variation based noise removal algorithms*, 1992
  - Boggess & Narcowich, *A first course in wavelets with Fourier analysis*, 2009
  - Mallat, *A wavelet tour of signal processing*, 1999

- Optimization
  - Boyd & Vandenberghe, *Convex optimization*, 2004
  - Nocedal & Wright, *Numerical optimization*, 2006
  - Rockafellar, *Convex optimization*, 1996
  - Daubechies, Defrise & De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, 2004