How to choose the regularization parameter in the total variation (TV) functional?

Heuristics: Rullgård **2008**

Balancing ℓ^1 and TV: Clason, Jin & Kunisch **2010**

Local variance: Dong, Hintermüller & Rincon-Camacho **2011**

Discrepancy principle: Wen & Chan **2012**

S-curve method: Kolehmainen, Lassas, Niinimäki & S **2012**

Dantzig estimation: Frick, Marnitz & Munk **2012**

Quasi-optimality principle and Hanke-Raus rules: Kindermann, Mutimbu & Resmerita **2013**

KKT system: Chen, Loli Piccolomini & Zama **2014**

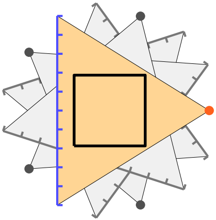
Discrepancy principle: Toma, Sixou & Peyrin **2015**

Cross validation, Stein's unbiased risk estimates, L-curve method, . . .

No single choice rule works perfectly for all applications. Therefore, it is good to have a collection of rules.

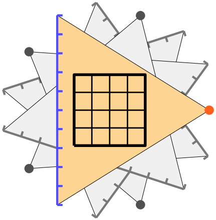
The continuous tomographic model needs to be approximated using a discrete model

Continuous model:



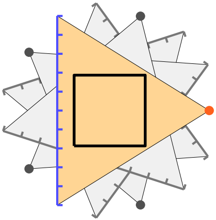
In this schematic setup we have 5 projection directions and a 10-pixel detector. Therefore, the number of data points is **50**.

Discrete model:



The resolution of the discrete model can be freely chosen according to computational resources

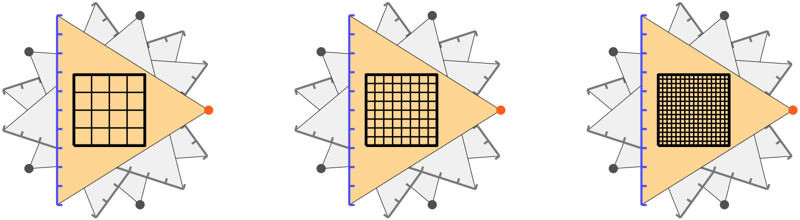
Continuous model:



In this schematic setup we have 5 projection directions and a 10-pixel detector. Therefore, the number of data points is **50**.

The number of degrees of freedom in the three discrete models below are **16**, **64** and **256**, respectively.

Discrete models:



We define the total variation (TV) norm consistently for continuous and discrete cases

Continuous anisotropic TV norm for attenuation coefficient $f : \Omega \to \mathbb{R}$:

$$\int_\Omega \left(\left| \frac{\partial f}{\partial x_1} \right| + \left| \frac{\partial f}{\partial x_2} \right| \right) dx.$$

Discrete anisotropic TV norm for an image matrix of size $n \times n$:

$$\frac{1}{n} \sum |f_\kappa - f_{\kappa'}| \,,$$

where the sum is over horizontally and vertically neighboring pixel values f_κ and $f_{\kappa'}$.

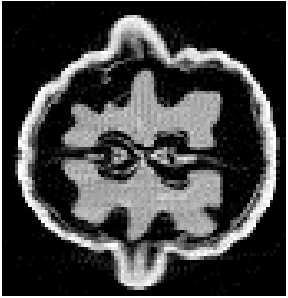
The above is based on this approximate two-dimensional computation:

$$\int_\Omega \left| \frac{f(x_1 + \frac{1}{n}, x_2) - f(x_1, x_2)}{1/n} \right| dx \approx (1/n)^2 \sum \left| \frac{f_\kappa - f_{\kappa'}}{1/n} \right| ,$$

where the sum is over horizontally neighboring pixel values f_κ and $f_{\kappa'}$.

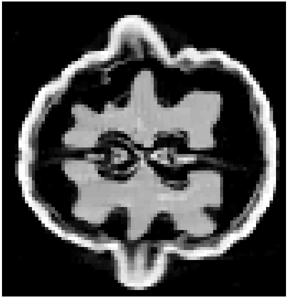
Low-noise TV reconstructions of a walnut using several regularization parameters

$\alpha = 0.001$



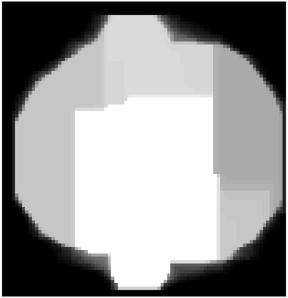
Too small α

$\alpha = 1$



Just right α

$\alpha = 1000$

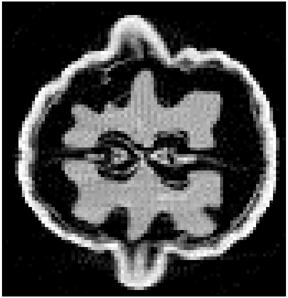


Too large α

Computations by Kati Niinimäki using a primal-dual interior point method.

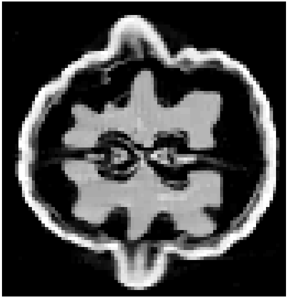
Low-noise TV reconstructions of a walnut using several regularization parameters

$\alpha = 0.001$



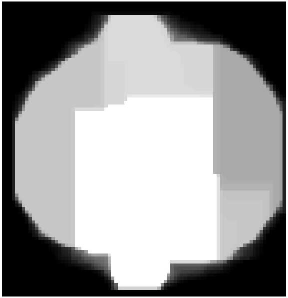
Too small α

$\alpha = 1$



Just right α

$\alpha = 1000$

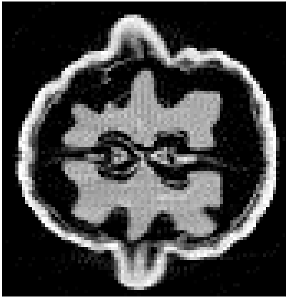


Too large α

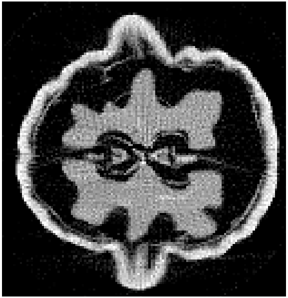
What happens when we compare reconstructions at different resolutions?

Low-noise TV reconstructions of a walnut at many resolutions using $\alpha = 0.001$

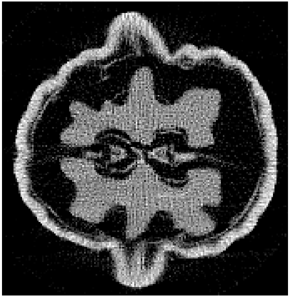
128×128



192×192

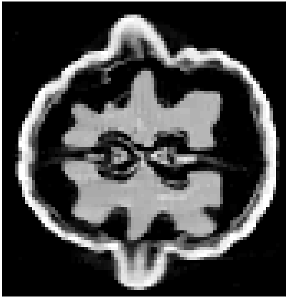


256×256

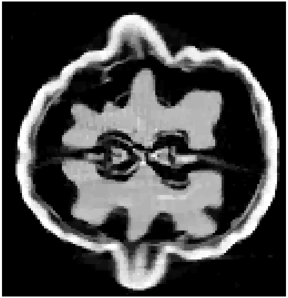


Low-noise TV reconstructions of a walnut at many resolutions using $\alpha = 1$

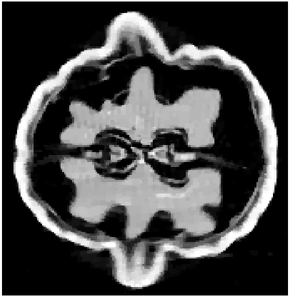
128×128



192×192

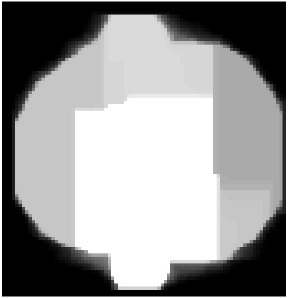


256×256

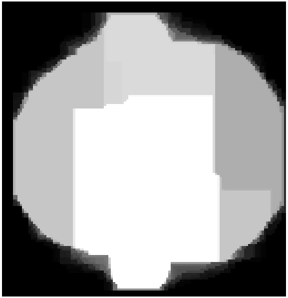


Low-noise TV reconstructions of a walnut at many resolutions using $\alpha = 1000$

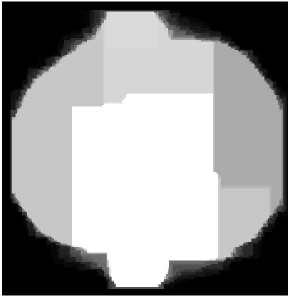
128×128



192×192



256×256



TV norms of low-noise reconstructions with various resolutions and parameters α

α	Resolution 128×128	192×192	256×256
10^{-4}	1.51	2.29	3.64
10^{-3}	1.51	2.29	3.46
10^{-2}	1.50	2.23	2.97
10^{-1}	1.43	1.85	1.93
10^{0}	1.08	1.11	1.11
10^{1}	0.78	0.78	0.77
10^{2}	0.48	0.48	0.48
10^{3}	0.12	0.12	0.12
10^{4}	0.04	0.04	0.04
10^{5}	0	0	0
10^{6}	0	0	0

TV norms of low-noise reconstructions with various resolutions and parameters α

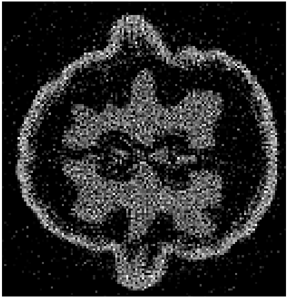
α	Resolution 128×128	192×192	256×256
10^{-4}	1.51	2.29	3.64
10^{-3}	1.51	2.29	3.46
10^{-2}	1.50	2.23	2.97
10^{-1}	1.43	1.85	1.93
10^{0}	1.08	1.11	1.11
10^{1}	0.78	0.78	0.77
10^{2}	0.48	0.48	0.48
10^{3}	0.12	0.12	0.12
10^{4}	0.04	0.04	0.04
10^{5}	0	0	0
10^{6}	0	0	0

What happens when we add noise to the data?

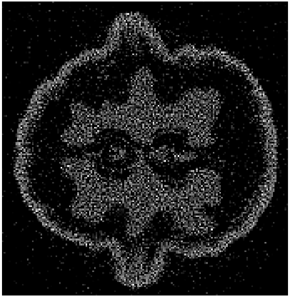
5% noise TV reconstructions of a walnut at many resolutions using $\alpha = 0.001$

128×128

192×192

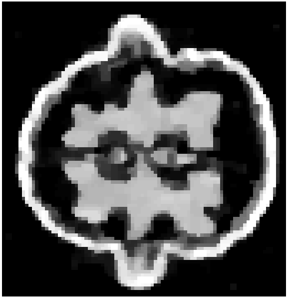


256×256

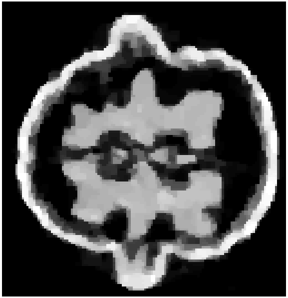


5% noise TV reconstructions of a walnut at many resolutions using $\alpha = 10$

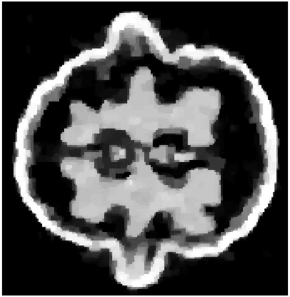
128×128



192×192

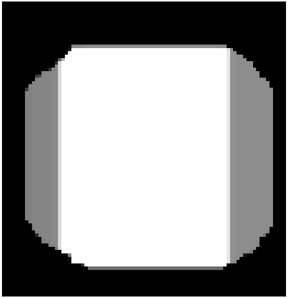


256×256

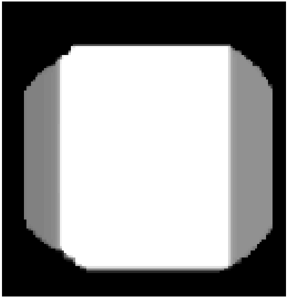


5% noise TV reconstructions of a walnut at many resolutions using $\alpha = 10000$

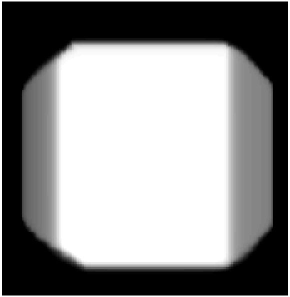
128×128



192×192



256×256



TV norms of reconstructions using various noise levels, resolutions and parameters α

	Low noise			5% noise		
α	128^2	192^2	256^2	128^2	192^2	256^2
10^{-4}	1.51	2.29	3.64	2.42	5.05	8.71
10^{-3}	1.51	2.29	3.46	2.43	5.05	8.59
10^{-2}	1.50	2.23	2.97	2.42	5.01	8.59
10^{-1}	1.43	1.85	1.93	2.37	4.83	8.16
10^{0}	1.08	1.11	1.11	1.99	3.50	5.12
10^{1}	0.78	0.78	0.77	0.86	0.86	0.88
10^{2}	0.48	0.48	0.48	0.48	0.48	0.48
10^{3}	0.12	0.12	0.12	0.12	0.12	0.12
10^{4}	0.04	0.04	0.04	0.04	0.04	0.04
10^{5}	0	0	0	0	0	0
10^{6}	0	0	0	0	0	0

[Niinimäki, Lassas, Hämäläinen, Kallonen, Kolehmainen, Niemi & S, SIAM Journal on Imaging Sciences 2016]

How about theory?

There are some related results in the literature

1992 **Vainikko**: *On the discretization and regularization of ill-posed problems with noncompact operators*

We use geometric arguments similar to those here:

1995 **Chambolle**: *Image Segmentation by Variational Methods: Mumford and Shah Functional and the Discrete Approximations.*

These works consider TV functionals and Γ-convergence when discretization is refined, but without a measurement operator:

2009 **Chambolle, Giacomini & Lussardi**
2012 **Gennip & Bertozzi**
2013 **Bellettini, Chambolle & Goldman**
2013 **Trillos & Slepčev**

This paper achieves a result analogous with ours, using wavelet frames in the finite-dimensional functionals:

2012 **Cai, Dong, Osher & Shen**

Assumptions on the linear forward map $\mathcal{A}$

Assume either **(A)** or **(B)** about the linear operator $\mathcal{A}$:

(A) $\mathcal{A} : L^2(D) \to L^2(\Omega)$ is compact and $\mathcal{A} : L^1(D) \to \mathcal{D}'(\Omega)$ is continuous with some open and bounded set $\Omega \subset \mathbb{R}^2$.
This covers the case of classical Radon transform with image domain D and sinogram domain Ω. We denote the set of distributions by $D'(\Omega)$.

(B) $\mathcal{A} : L^1(D) \to \mathbb{R}^M$ is bounded.
This covers the practically important discrete pencil beam model of tomographic measurements.

Definition of discrete and continuous regularization functionals

Let D be the square $[0,1]^2 \subset \mathbb{R}^2$. Use anisotropic $BV(D)$ norm

$$\|u\|_{BV} = \|u\|_{L^1} + V(u) = \|u\|_{L^1} + \int_D \Big(\Big|\frac{\partial u(x)}{\partial x_1}\Big| + \Big|\frac{\partial u(x)}{\partial x_2}\Big| \Big) dx.$$

Define $S_\infty : BV(D) \to \mathbb{R}$ and $S_j : BV(D) \to \mathbb{R} \cup \{\infty\}$ by

$$S_\infty(u) = \|\mathcal{A}u - m\|^2_{L^2(\Omega)} + \alpha_1 \|u\|_{L^1(D)} + \alpha V(u)$$

with positive regularization parameters $\alpha_1 > 0$ and $\alpha > 0$, and

$$S_j(u) = \begin{cases} S_\infty(u), & \text{for } u \in \text{Range}(T_j), \\ \infty, & \text{for } u \notin \text{Range}(T_j). \end{cases}$$

Linear operator T_j projects to functions that are piecewise constant on a regular $2^j \times 2^j$ square pixel grid.

Our main theorem ensures the convergence of regularized solutions as resolution grows

- There exists a minimizer $\widetilde{u}_j \in \arg\min(S_j)$ for all $j = 1, 2, 3, \ldots$
- There exists a minimizer $\widetilde{u}_\infty \in \arg\min(S_\infty)$.
- Any sequence $\widetilde{u}_j \in \arg\min(S_j)$ of minimizers has a subsequence $\widetilde{u}_{j(\ell)}$ that converges weakly in $BV(D)$ to some limit $w \in BV(D)$. Furthermore, $\lim_{\ell\to\infty} V(\widetilde{u}_{j(\ell)}) = V(w)$.
- The limit w is a minimizer: $w \in \arg\min(S_\infty)$.

[Niinimäki, Lassas, Hämäläinen, Kallonen, Kolehmainen, Niemi & S, SIAM Journal on Imaging Sciences 2016]

What can we say about the proposed method?

Benefits of our multiresolution TV parameter choice method:

- simple definition,
- easy implementation, and
- no need of *a priori* information about the noise amplitude.

Also, it seems to perform well for real tomographic data.

Downside: several reconstructions need to be computed. Also, it is still unclear why the method works so nicely: if there is convergence for any α in theory, what is the instability we are observing?

The method can be tried out with 3D tomography (it works!) and with other inverse problems and regularizers.

The S-curve method determines a regularization parameter value giving the right sparsity level