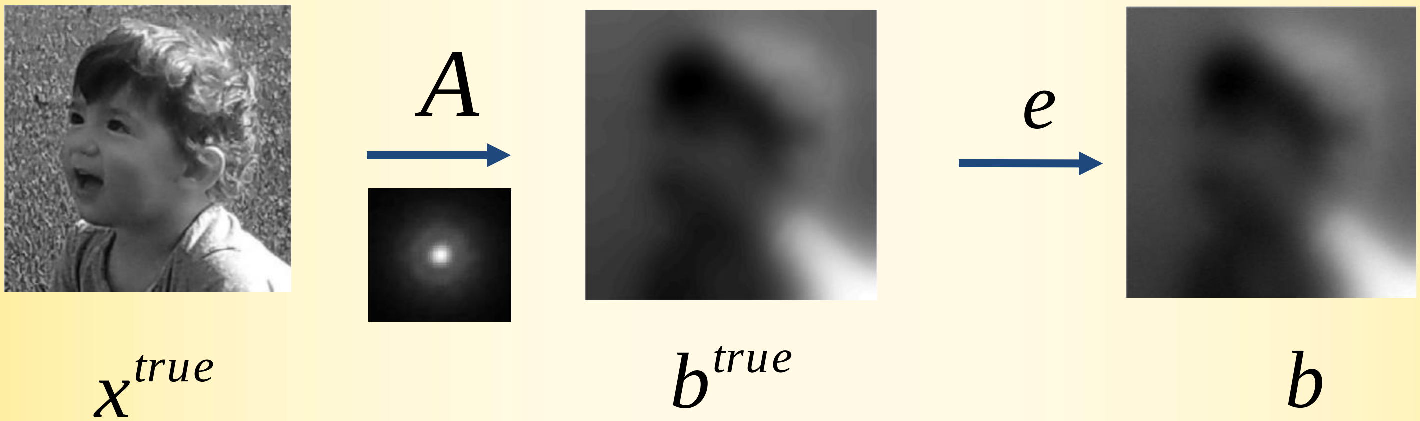
A Learning Approach for Computing Regularization Parameters for General-Form Tikhonov

Julianne Chung[1] and Malena I. Espanol[2]

[1] Virginia Tech, Blacksburg, VA
[2] The University of Akron, Akron, OH

SIAM Annual Meeting
Pittsburgh, PA
July 10-14th, 2017

Image Deblurring



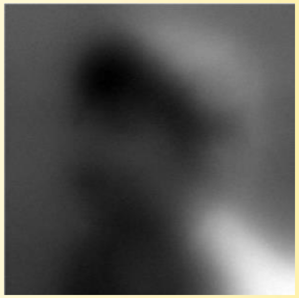
$$Ax = b^{true} + e = b$$

Image deblurring problem: Try to reconstruct the true image from a blurred and noisy one.

The Naïve Solution

x^{true}?

A^{-1}



b

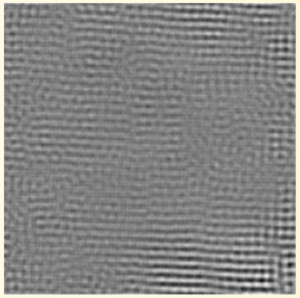
...but $x = A^{-1}b =$ 

Image deblurring is an ill-posed inverse problem: small perturbations in the data may result in large errors in the solution.

Tikhonov Regularization

Tikhonov in standard form :

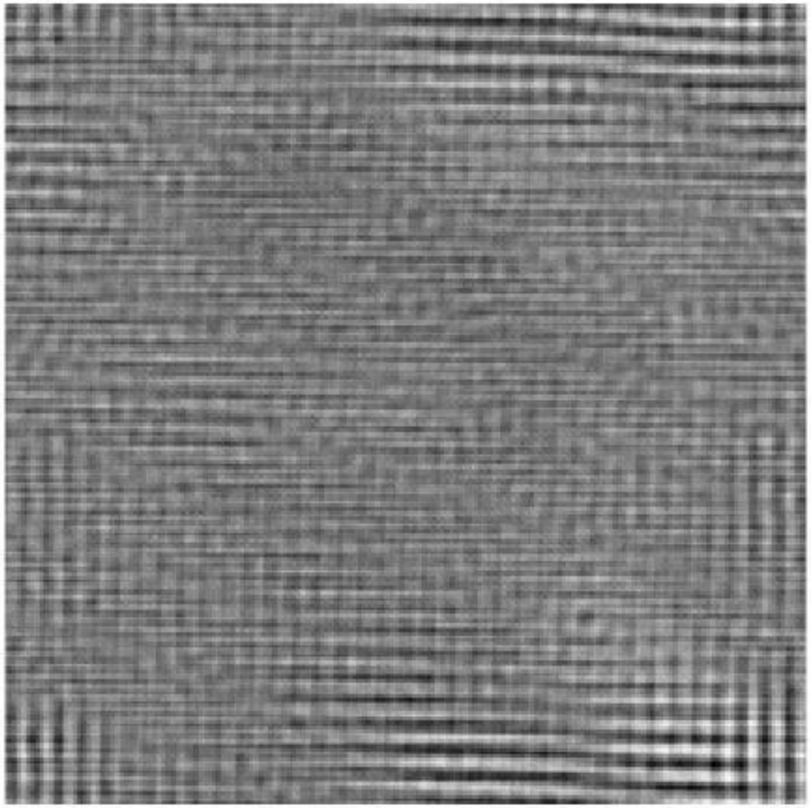
$$x_{Tik} = \underset{x}{\arg\min} \| Ax - b \|_2^2 + \lambda^2 \|x\|_2^2$$

Tikhonov in general form :

$$x_{Tik} = \underset{x}{\arg\min} \| Ax - b \|_2^2 + \lambda^2 \|Lx\|_2^2$$

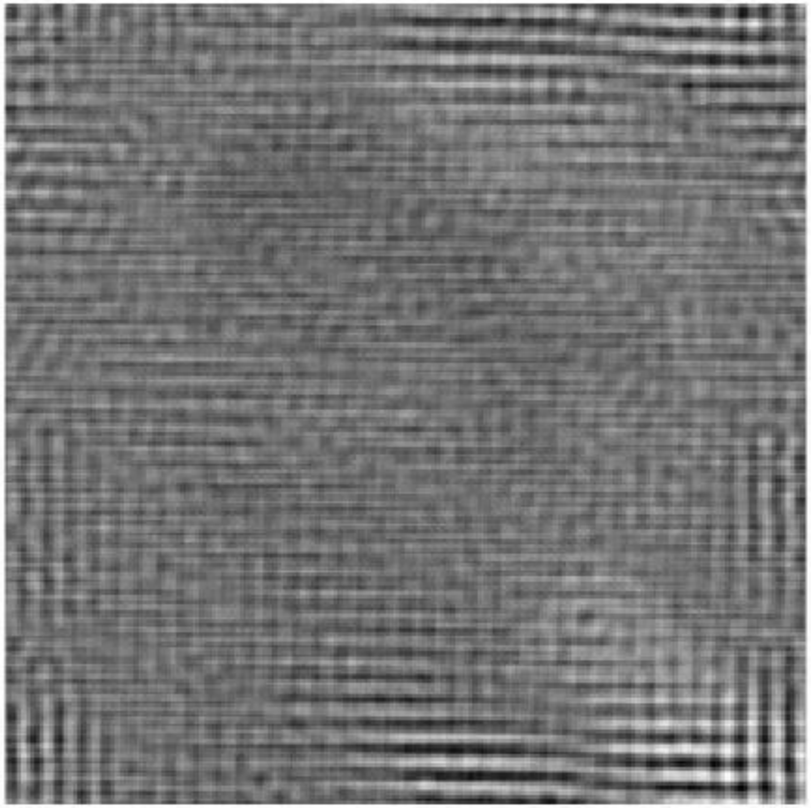
We regularize by adding information we know about the solution.

How important is λ?



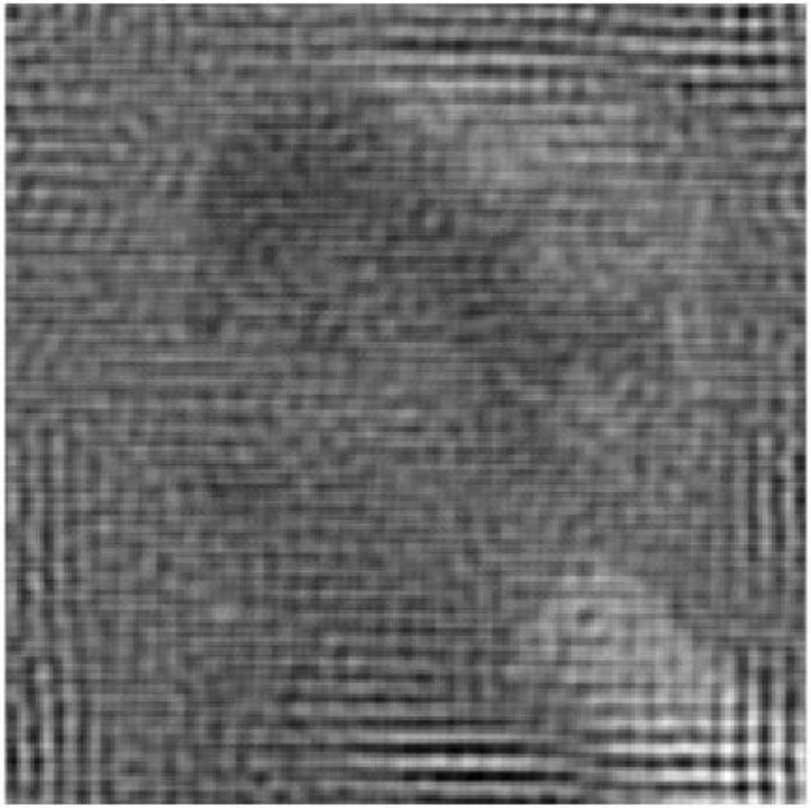
$$\lambda = 0.0010$$

How important is λ?



$$\lambda = 0.0017$$

How important is λ?



$$\lambda = 0.0028$$

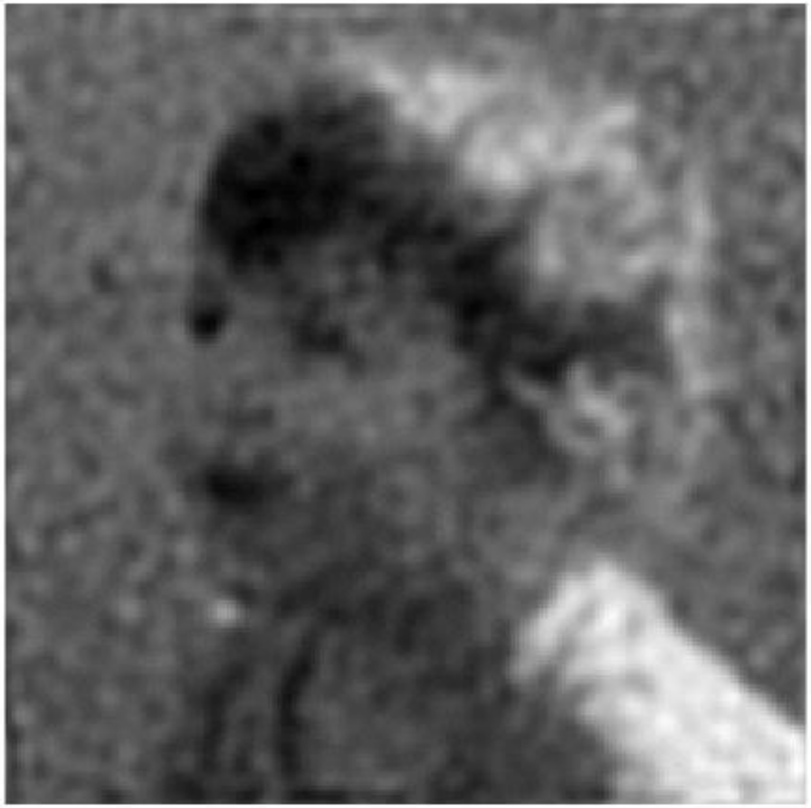
How important is λ?

$$\lambda = 0.0046$$

How important is λ?

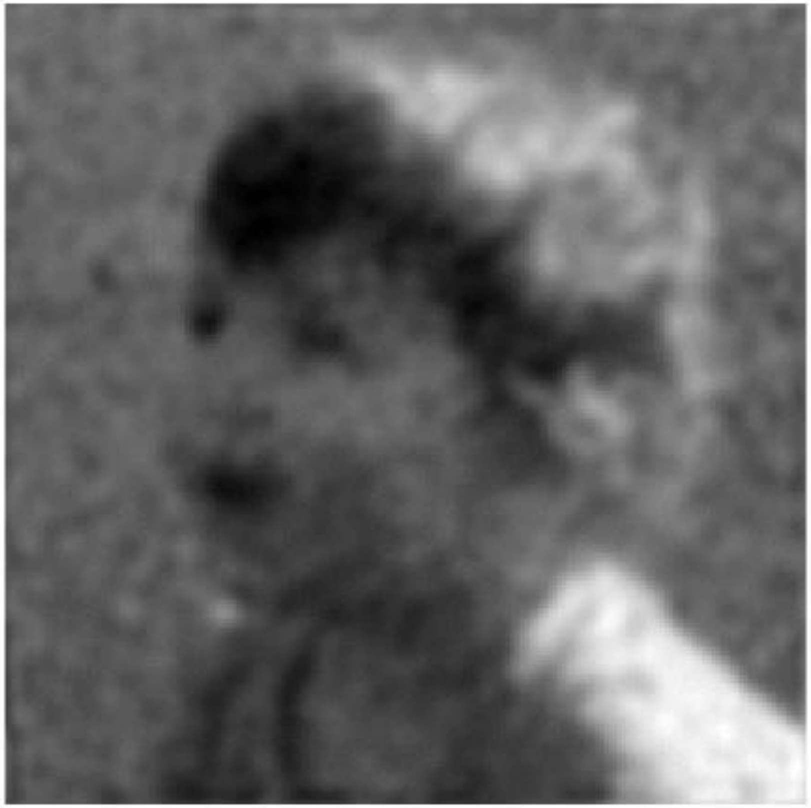
$$\lambda = 0.0077$$

How important is λ?



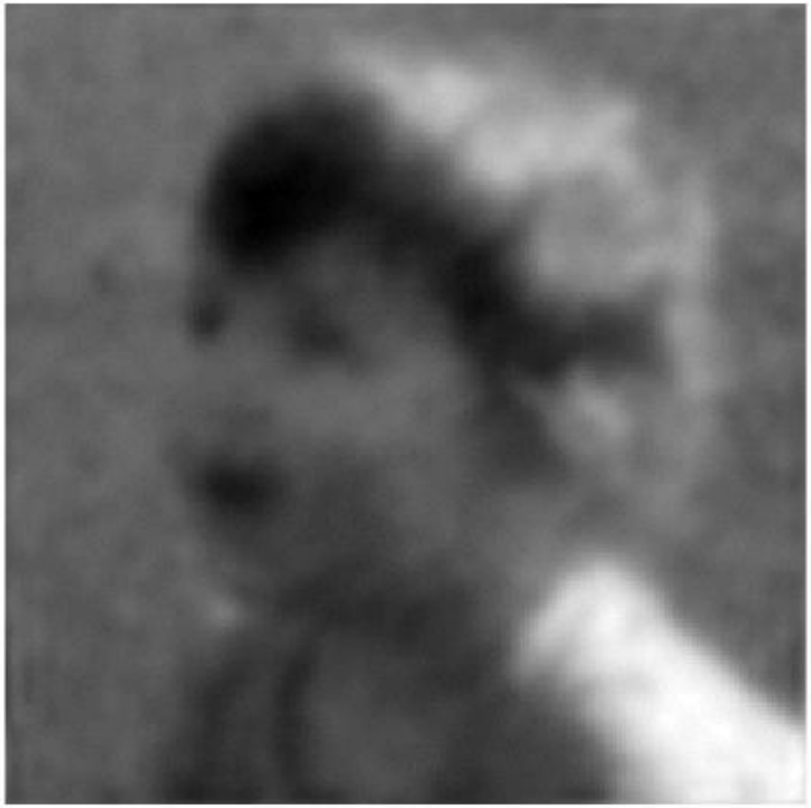
$$\lambda = 0.0129$$

How important is λ?



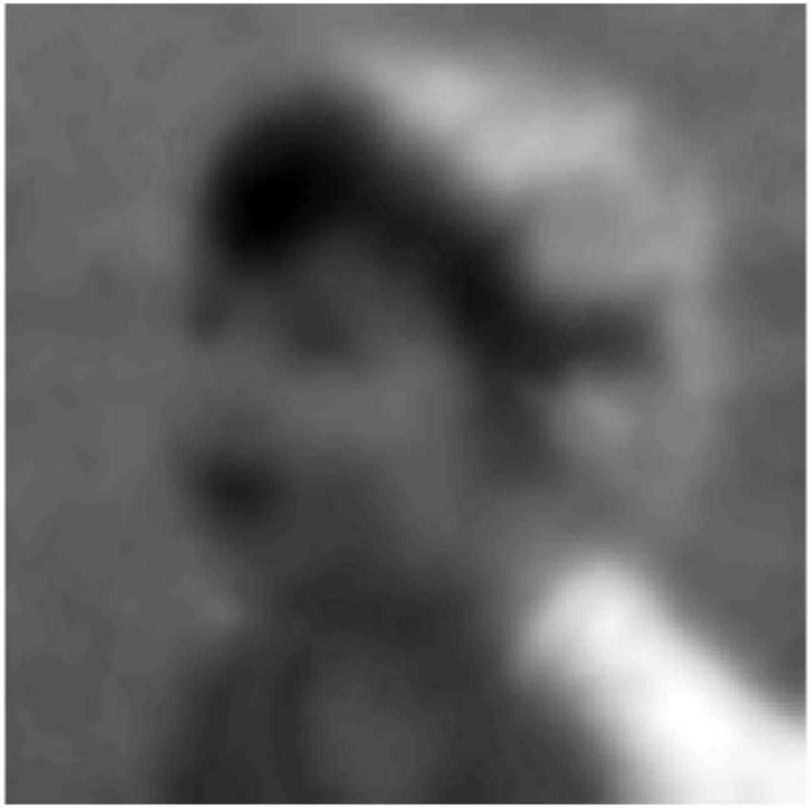
$$\lambda = 0.0215$$

How important is λ?



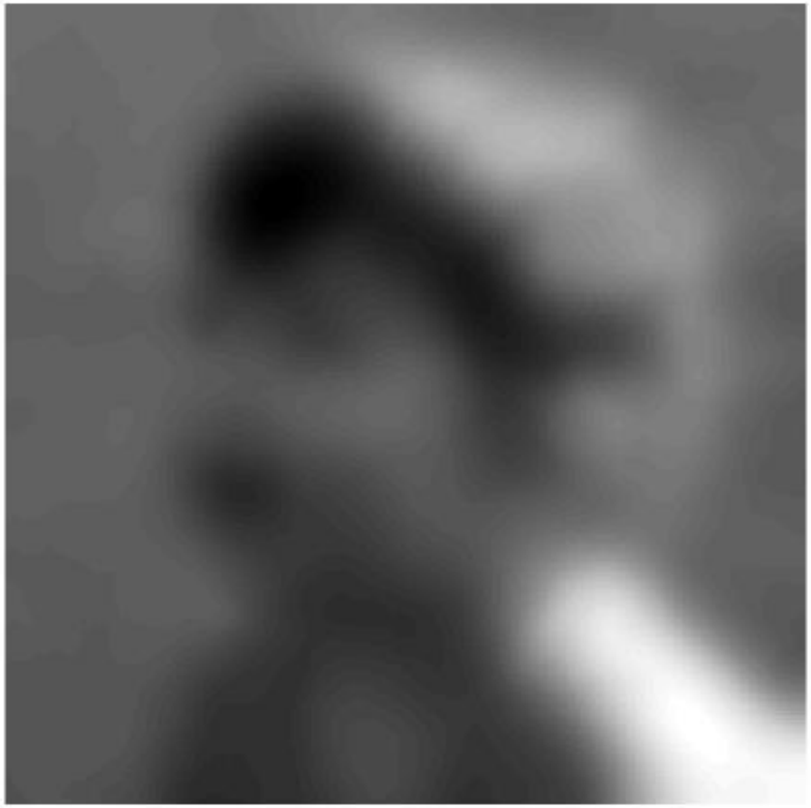
$$\lambda = 0.0359$$

How important is λ?



$$\lambda = 0.0359$$

How important is λ?

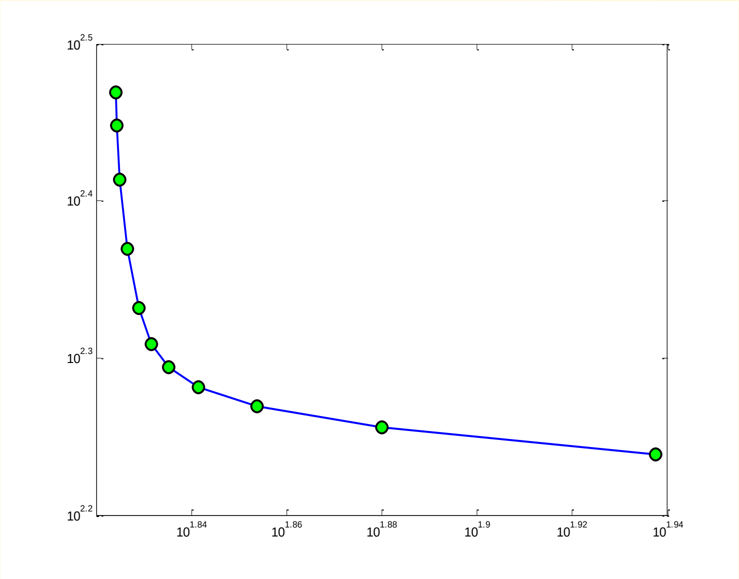


$$\lambda = 1$$

How to choose λ

L-curve:

$$\log \|Lx\|_2$$



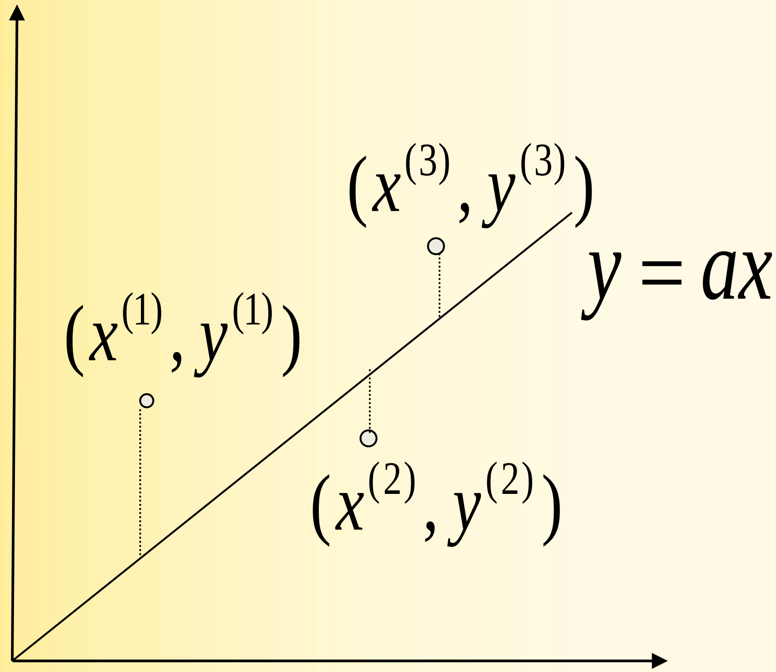
$$\log \|Ax - b\|_2$$

Other methods: Discrepancy Principle

Generalized Cross Validation (GCV)

Unbiased Predictive Risk Estimator (UPRE)

Learning from data



Given data points $\left(x^{(k)}, y^{(k)}\right)$,

find the parameter a such that

$$\min_a \sum_{k=1}^{3} [y_a^{(k)} - y^{(k)}]^2$$

subject to $y_a^{(k)} = ax^{(k)}$.

Learning λ from data

Given training images

$$\left(x_{\text{true}}^{(k)}, b^{(k)}\right) \quad k = 1,2,\ldots,K,$$

find λ such that

$$\min_{\lambda} \frac{1}{K} \sum_{k=1}^{K} \left\| x_{\lambda}^{(k)} - x_{\text{true}}^{(k)} \right\|_p^p$$

subject to

$$x_{\lambda}^{(k)} = \arg\min_{x} \left\| Ax - b^{(k)} \right\|_2^2 + \lambda^2 \left\| Lx \right\|_2^2 \quad \text{for} \quad k = 1,\ldots,K$$

Designing optimal spectral filters for inverse problems by Chung, Chung, and O'Leary (2011)
Learning regularization parameters for general-form Tikhonov by Chung and Espanol (2017)

A bilevel optimization problem

$$\min_{\lambda} \frac{1}{K} \sum_{k=1}^{K} \left\| x_{\lambda}^{(k)} - x_{\text{true}}^{(k)} \right\|_p^p$$

subject to

$$x_{\lambda}^{(k)} = \arg\min_{x} \left\| Ax - b^{(k)} \right\|_2^2 + \lambda^2 \left\| Lx \right\|_2^2 \quad \text{for} \quad k = 1, \ldots, K$$

A Bilevel Optimization Approach for Parameter Learning in Variational Models by Kunisch and Pock (2013)

Bilevel Approaches for Learning of Variational Imaging Models by Calastroni et al. (2015)

Bilevel Parameter Learning for Higher-order Total Variarion Regularization Models by De Los Santos et al. (2015)

A simpler minimization problem

Notice that the solution

$$x_\lambda^{(k)} = \arg\min_x \| Ax - b^{(k)} \|_2^2 + \lambda^2 \|Lx\|_2^2 \quad \text{for} \quad k = 1, \ldots, K$$

has a explicit solution

$$x_\lambda^{(k)} = (A^T A + \lambda^2 L^T L)^{-1} A^T b^{(k)}$$

Therefore, we only need to solve

$$\min_\lambda \frac{1}{K} \sum_{k=1}^{K} \left\| (A^T A + \lambda^2 L^T L)^{-1} A^T b^{(k)} - x_{\text{true}}^{(k)} \right\|_p^p$$

Generalized SVD

We have the following joint decomposition of A and L:

$$A = PCZ^{-1} \qquad L = \overline{P}SZ^{-1}$$

where

- $C = diag(c_1, ..., c_n) \in R^{nxn}$ and $S = diag(s_1, ..., s_n) \in R^{pxn}$

  $1 \geq c_1 \geq ... \geq c_{\min\{n,p\}} \geq 0$ and $0 \leq s_1 \leq ... \leq s_{\min\{n,p\}} \leq 1$

  $\mathrm{C^T C + S^T S = I}$
- P and $\overline{P}$ orthogonal, Z nonsingular

Finding the optimal λ

Using the $GSVD$ of A and L, we have that

$$x_\lambda^{(k)} = \sum_{i=1}^{n} \phi_i \frac{p_i^T b^{(k)}}{c_i} z_i = Z\Gamma^{(k)} \phi(\lambda),$$

where $\phi_i = \dfrac{c_i^2}{c_i^2 + \lambda_i^2 s_i^2}$ are called the filter factors,

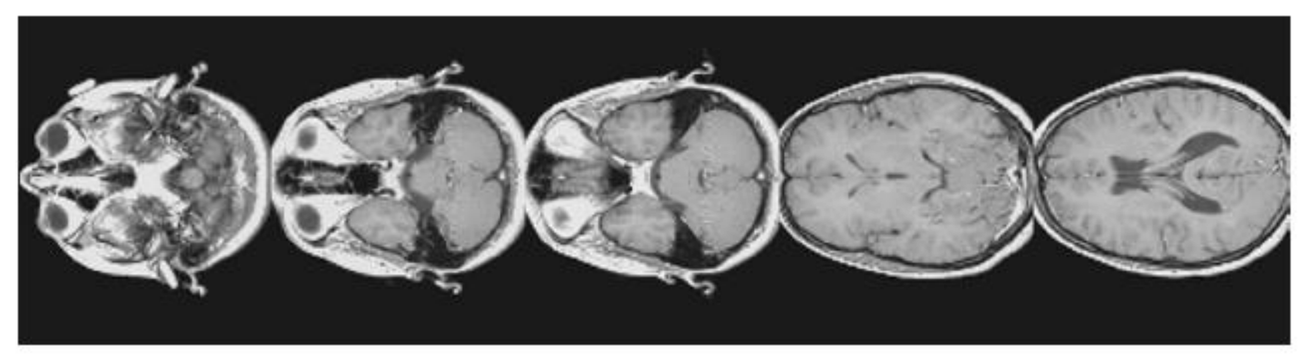
$$\Gamma^{(k)} = diag\left(\frac{p_1^T b^{(k)}}{c_1}, \ldots, \frac{p_n^T b^{(k)}}{c_n} \right) \text{ and } \phi(\lambda) = (\phi_1, \ldots, \phi_n).$$

Then

$$\min_\lambda \frac{1}{K} \sum_{k=1}^{K} \left\| Z\Gamma^{(k)} \phi(\lambda) - x_{\text{true}}^{(k)} \right\|_p^p$$

can easily be solved by standard root finding methods.

Numerical results: 1D Deconvolution



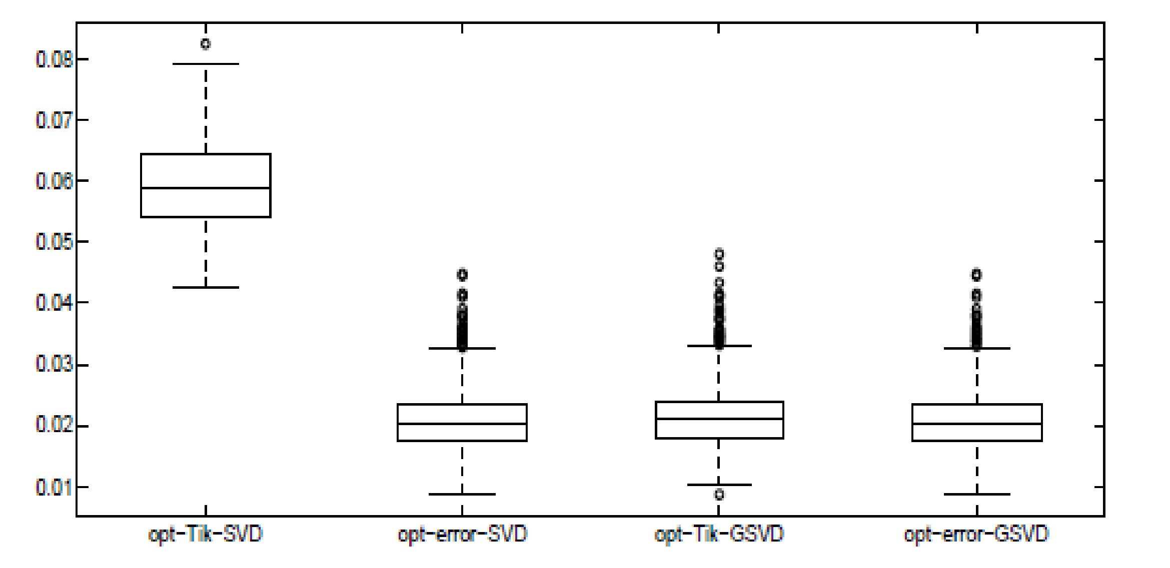
Training data : 200 columns of each MRI = 1000 signals of size 256 each.

We apply a Gaussian blur + random noise : 0.2 - 0.25.

$$L = \begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ & \ddots & \ddots & & \\ & & -1 & 1 & \\ & & & & -1 \end{pmatrix}$$

Numerical results: Box and whisker plots

Validation data : 1000 signals. $\mathrm{eERR} = \|x_\lambda - x_{\mathrm{true}}\|_2 / \|x_{\mathrm{true}}\|_2$



opt – Tik – SVD: $A = U\Sigma V^T$

$$x_\lambda^{(k)} = \sum_{i=1}^{n} \phi_i \frac{u_i^T b^{(k)}}{\sigma_i} v_i, \quad \text{where } \phi_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2},$$

opt – error – SVD:

$$x^{(k)} = \sum_{i=1}^{n} \phi_i \frac{u_i^T b^{(k)}}{\sigma_i} v_i$$

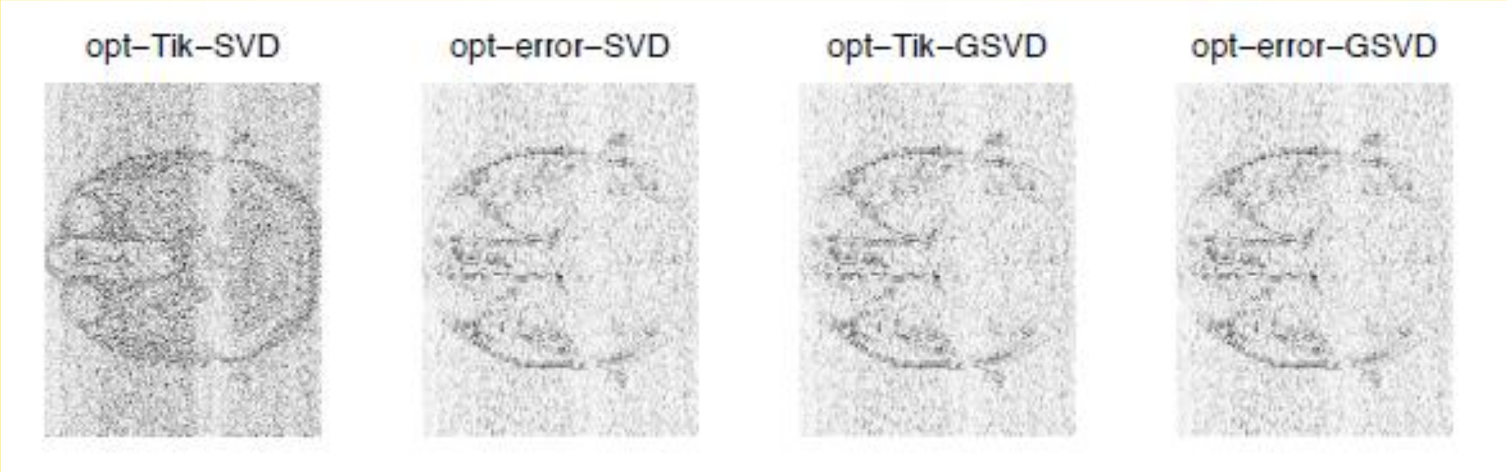
opt – Tik – GSVD:

$$x_\lambda^{(k)} = \sum_{i=1}^{n} \phi_i \frac{p_i^T b^{(k)}}{c_i} z_i, \quad \text{where } \phi_i = \frac{c_i^2}{c_i^2 + s_i^2 \lambda^2}$$

opt – error – GSVD:

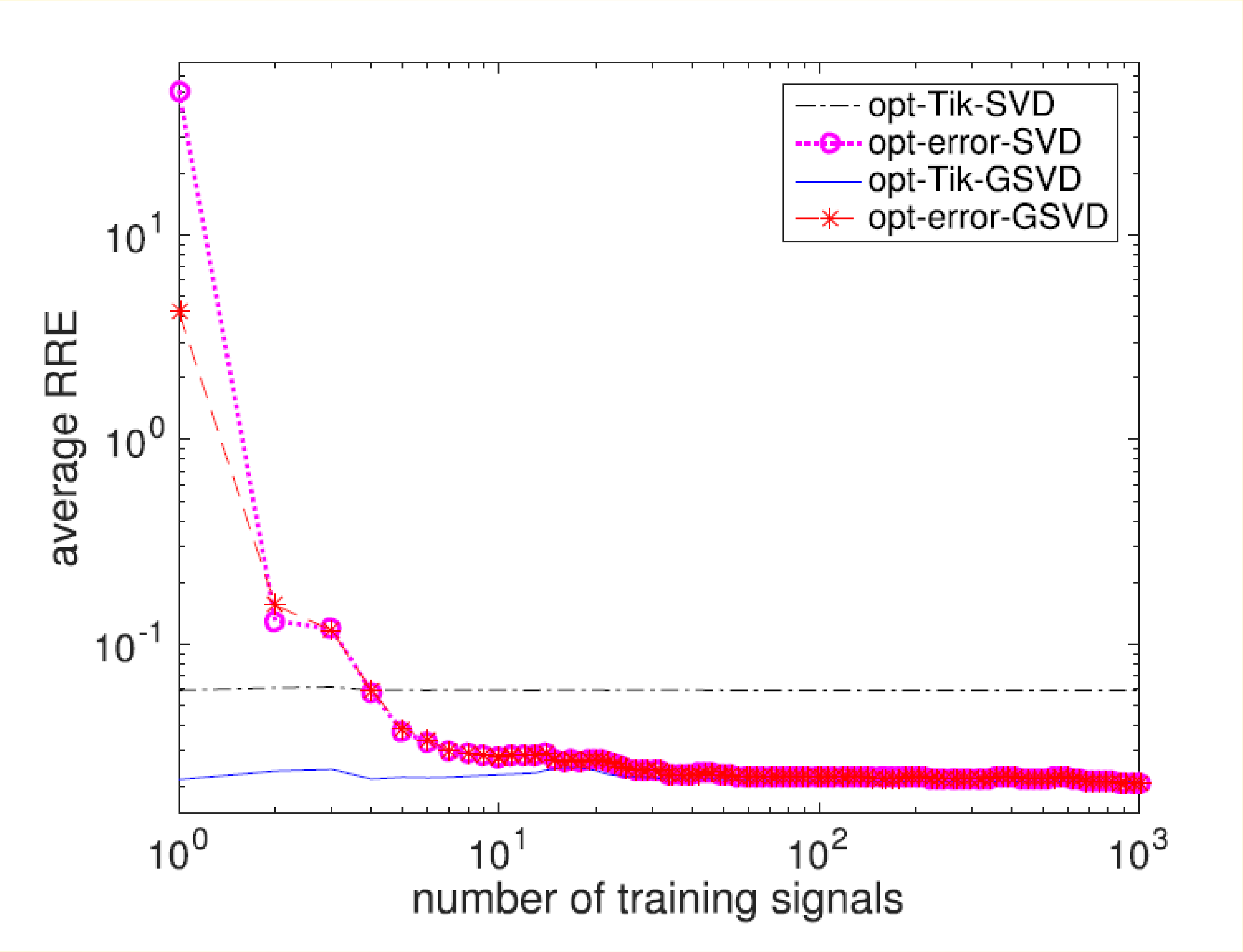
$$x^{(k)} = \sum_{i=1}^{n} \phi_i \frac{p_i^T b^{(k)}}{c_i} z_i$$

Numerical results



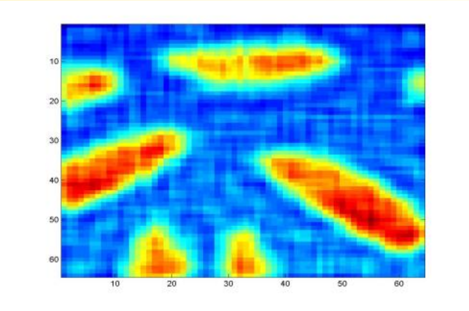
	Training Set		Validation Set	
	average RRE	STD	average RRE	STD
opt-Tik-SVD	5.912e-02	6.713e-03	5.929e-02	6.839e-03
opt-error-SVD	2.145e-02	5.931e-03	2.092e-02	5.182e-03
opt-Tik-GSVD	2.202e-02	5.986e-03	2.143e-02	5.277e-03
opt-error-GSVD	2.144e-02	5.927e-03	2.091e-02	5.180e-03
MSE-Tik-GSVD	2.142e-02	5.185e-03	2.091e-02	4.855e-03

Numerical results: Pareto Curve

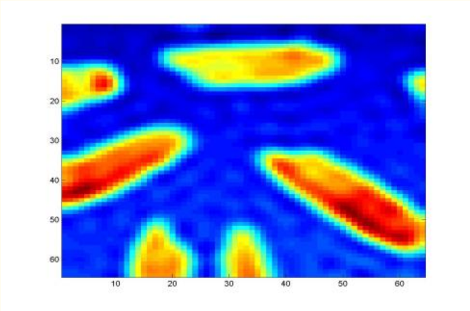


Multi-parameter Tikhonov

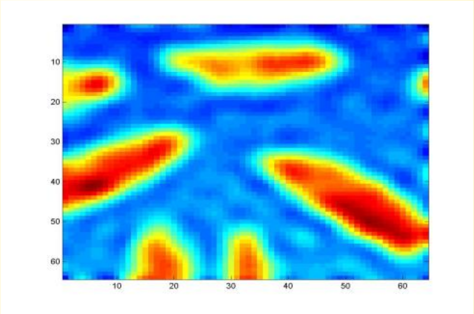
The selection of the regularization matrix L is very important and difficult



L = Identity



L = Derivative Op.



L = Laplacian Op.

…but what if we do not have to choose.

$$x_\lambda = \underset{x}{\arg\min} \| Ax - b \|_2^2 + \sum_i \lambda_i^2 \|L_i x\|_2^2$$

Multi-parameter Tikhonov

$$\min_{x} \| Ax - b \|_2^2 + \sum_{j=1}^{J} \lambda_j^2 \left\| L_j x \right\|_2^2$$

Previous work include:

M. Belge, M. E. Kilmer, and E. L. Miller. C. 2002
Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu. 2003
S. Lu and S. V. Pereverzev. 2011
F. S. Viloche Bazan, L. S. Borges, and J. B. Francisco. 2012
Z. Wang. 2012
S. Gazzola and P. Novati. 2013

Learning the <u>vector</u> λ from data

Given training images

$$\left(x_{\text{true}}^{(k)}, b^{(k)}\right) \quad k = 1,2,\ldots,K,$$

find λ such that

$$\min_{\lambda} \frac{1}{K} \sum_{k=1}^{K} \left\| x_{\lambda}^{(k)} - x_{\text{true}}^{(k)} \right\|_2^2$$

subject to

$$x_{\lambda}^{(k)} = \arg\min_{x} \left\| Ax - b^{(k)} \right\|_2^2 + \sum_i \lambda_i^2 \left\| L_i x \right\|_2^2 \quad \text{for} \quad k = 1,\ldots,K$$

The SD case

Assume A and L_j are simultaneously diagonalizable

$$A = Q^* C Q \quad \text{and} \quad L_j = Q^* S_j Q$$

Multi-parameter Tikhonov solution can be writen

$$x_\lambda = \sum_{i=1}^{n} \phi_i \frac{q_i^T b}{c_i} q_i$$

where ϕ_i are the filter factors defined by

$$\phi_i = \frac{|c_i|^2}{|c_i|^2 + \sum_{j=1}^{J} \lambda_j^2 |s_{i,j}|^2}$$

- Structured matrices common in image processing[1]
  - Periodic boundary conditions - Q^* represents 2D DFT matrix
  - Double symmetry of PSFs and reflexive boundary conditions - Q^* represents 2D DCT matrix

Finding the optimal λ

$$x_\lambda^{(k)} = \sum_{i=1}^{n} \phi_i \frac{q_i^T b^{(k)}}{c_i} q_i = Q\Gamma^{(k)}\phi(\lambda)$$

where

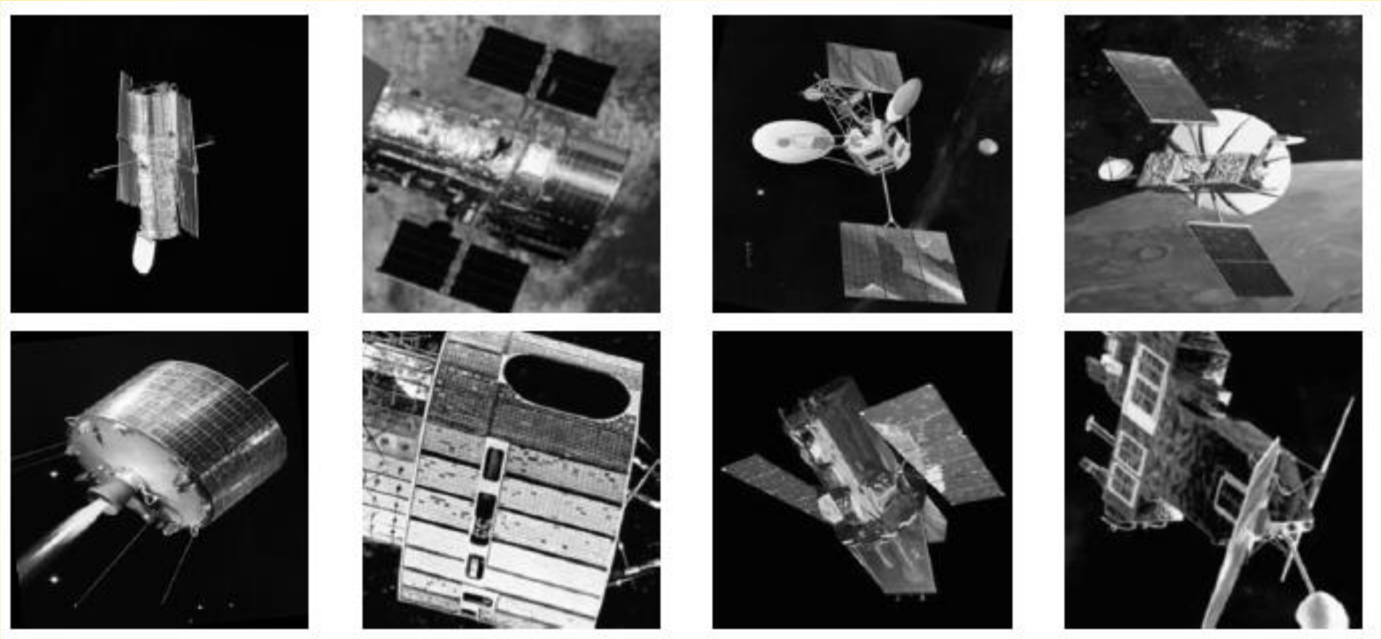
$$\Gamma^{(k)} = diag\left(\frac{q_1^T b^{(k)}}{c_1}, \ldots, \frac{q_n^T b^{(k)}}{c_n}\right) \text{ and } \phi(\lambda) = (\phi_1, \ldots, \phi_n)$$

$$\min_\lambda \frac{1}{K} \sum_{k=1}^{K} \left\| Q\Gamma^{(k)}\phi(\lambda) - x_{\text{true}}^{(k)} \right\|_p^p$$

It is easy to compute the Jacobian and Hessian explicitely and solve using Gauss - Newton Method.

Numerical results

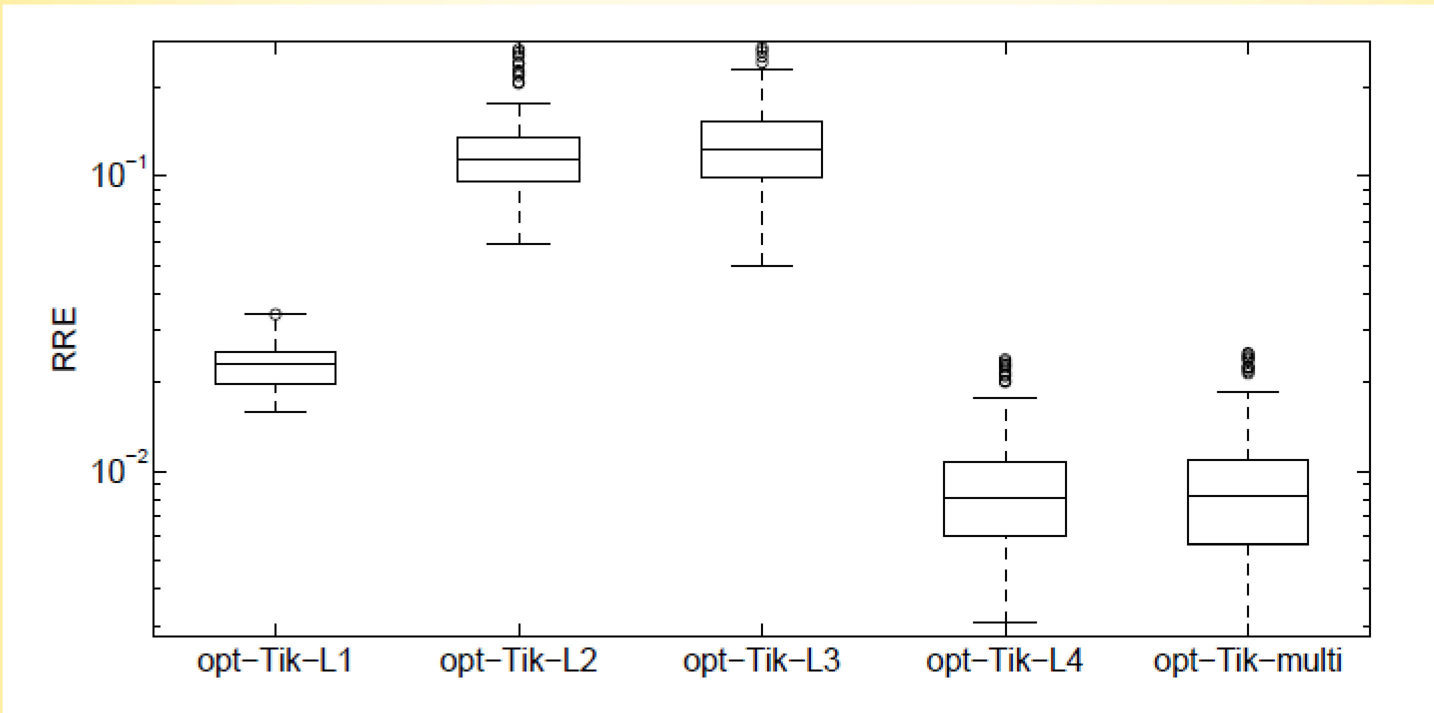
Sample of training and validation images:



Training and validation data: 80 images = 8 images * 10 rigid rotations.

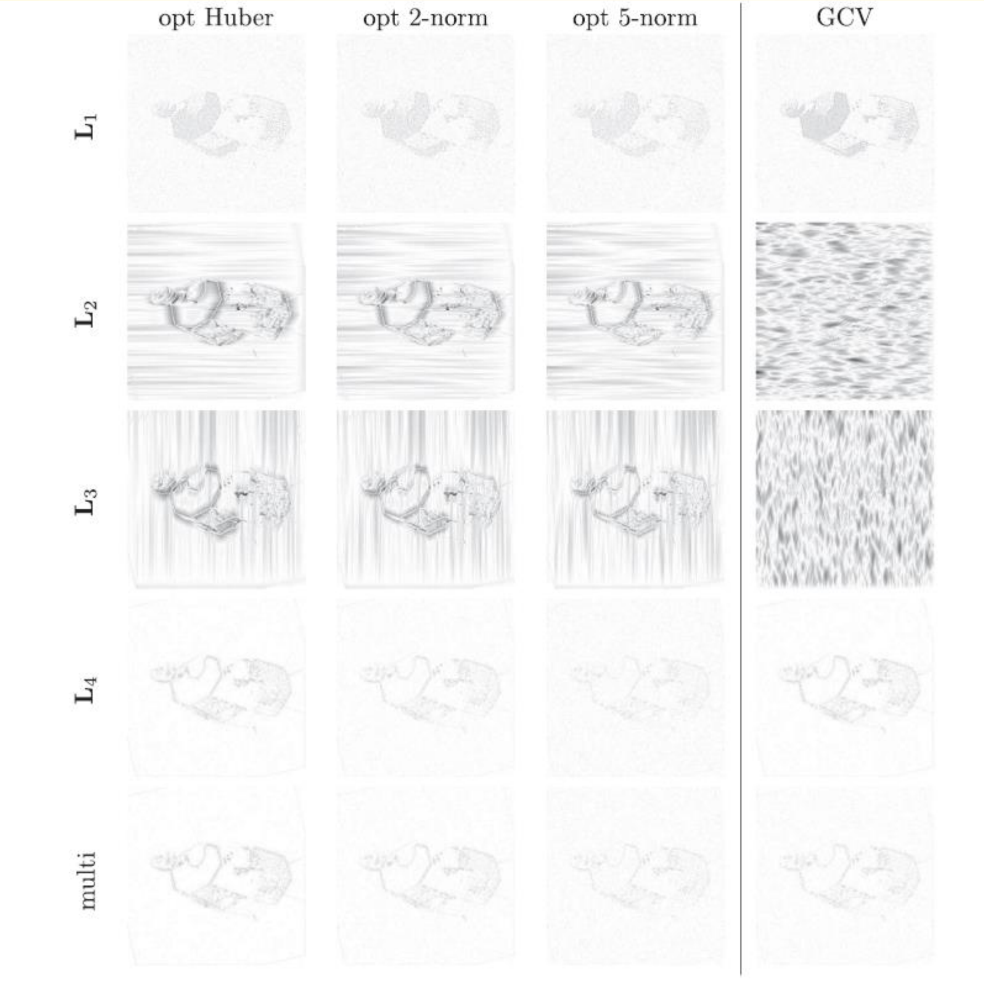
We apply a Gaussian blur + random noise: 0.1 - 0.15.

Numerical results: Box and whisker plots



$\mathbf{L}_1 = \mathbf{I}$, $\begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ $\begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

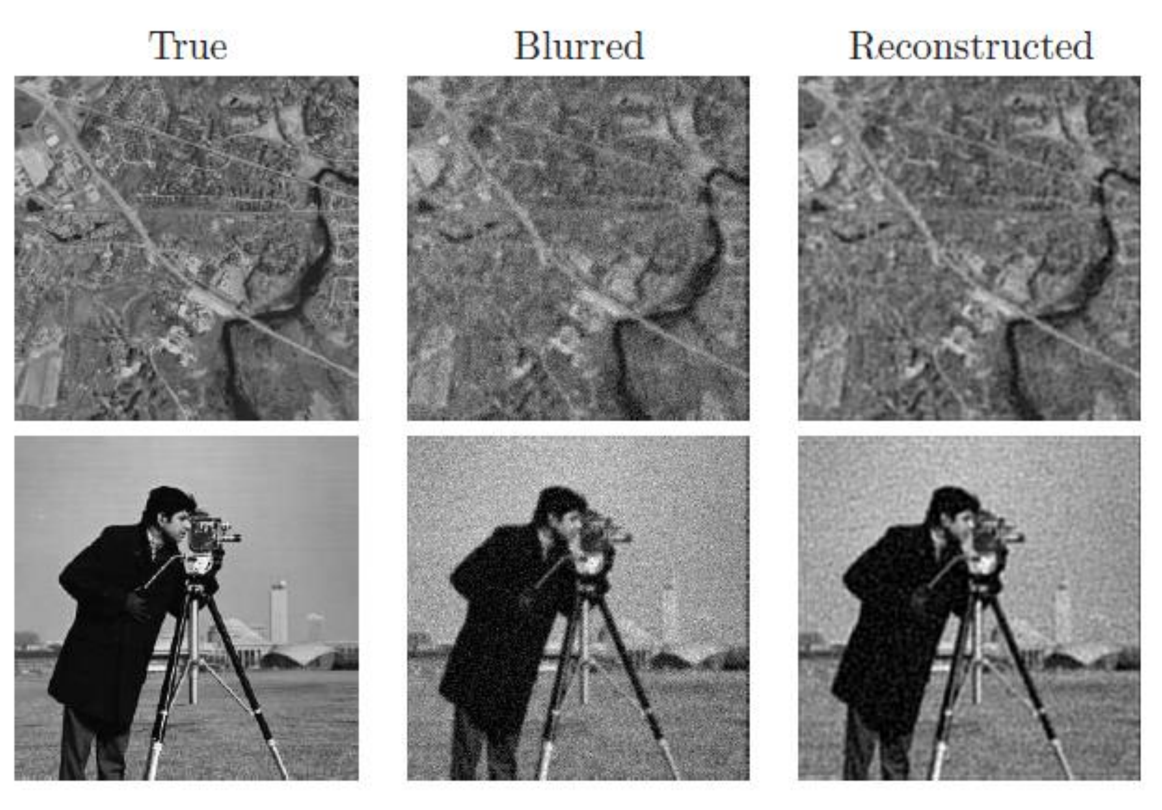
Numerical results



Numerical results

	Huber	2-norm	5-norm
opt-L_1	1.672e-01 (4.164e-02)	2.324e-02 (4.375e-03)	1.232e-04 (9.215e-05)
opt-L_2	3.718e-01 (1.411e-01)	1.257e-01 (4.904e-02)	1.118e-02 (9.886e-03)
opt-L_3	3.816e-01 (1.422e-01)	1.303e-01 (4.669e-02)	1.145e-02 (9.391e-03)
opt-L_4	9.368e-02 (4.586e-02)	9.766e-03 (5.457e-03)	3.383e-05 (3.436e-05)
opt-multi	9.393e-02 (4.806e-02)	9.829e-03 (5.993e-03)	3.469e-05 (3.817e-05)
GCV-multi	9.411e-02 (4.739e-02)	8.870e-03 (5.172e-03)	5.076e-05 (4.711e-05)

How it performs in different images



	Ave. Difference	Ave. SSIM Difference	Rel. Reconstruction Errors
Cell Image	71.6595	2.348e-01	1.13e-02
Fabric Image	83.6700	0.482e-01	2.50e-02
Aerial Map	90.3935	0.432e-01	1.75e-02
Camera Man	114.4114	1.379e-01	1.55e-02

Conclusions

- Computing regularization parameters for Tikhonov regularization is in general a difficult task
- We developed a learning approach to compute regularization parameters for one-parameter and multi-parameter Tikhonov from training data
- Optimal parameters can be computed off-line
- The optimal parameters work well for images outside the family of training data