Unbiased Risk Estimation for Sparse Analysis Regularization

Charles Deledalle[1], Samuel Vaiter[1], Gabriel Peyré[1], Jalal Fadili[3] and Charles Dossal[2]
[1]CEREMADE, Université Paris–Dauphine — [2]GREY'C, ENSICAEN — [3]IMB, Université Bordeaux I

Problem statement

Consider the convex but non-smooth **Analysis Sparsity Regularization** problem

$$x^\star(y,\lambda) \in \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|y - \Phi x\|^2 + \lambda \|D^* x\|_1 \qquad (\mathcal{P}_\lambda(y))$$

which aims at inverting

$$y = \Phi x_0 + w$$

by promoting sparsity and with

- $x_0 \in \mathbb{R}^N$ the unknown image of interest,
- $y \in \mathbb{R}^Q$ the low-dimensional noisy observation of x_0,
- $\Phi \in \mathbb{R}^{Q \times N}$ a linear operator that models the acquisition process,
- $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_Q)$ the noise component,
- $D \in \mathbb{R}^{N \times P}$ an analysis dictionary, and
- $\lambda > 0$ a regularization parameter.

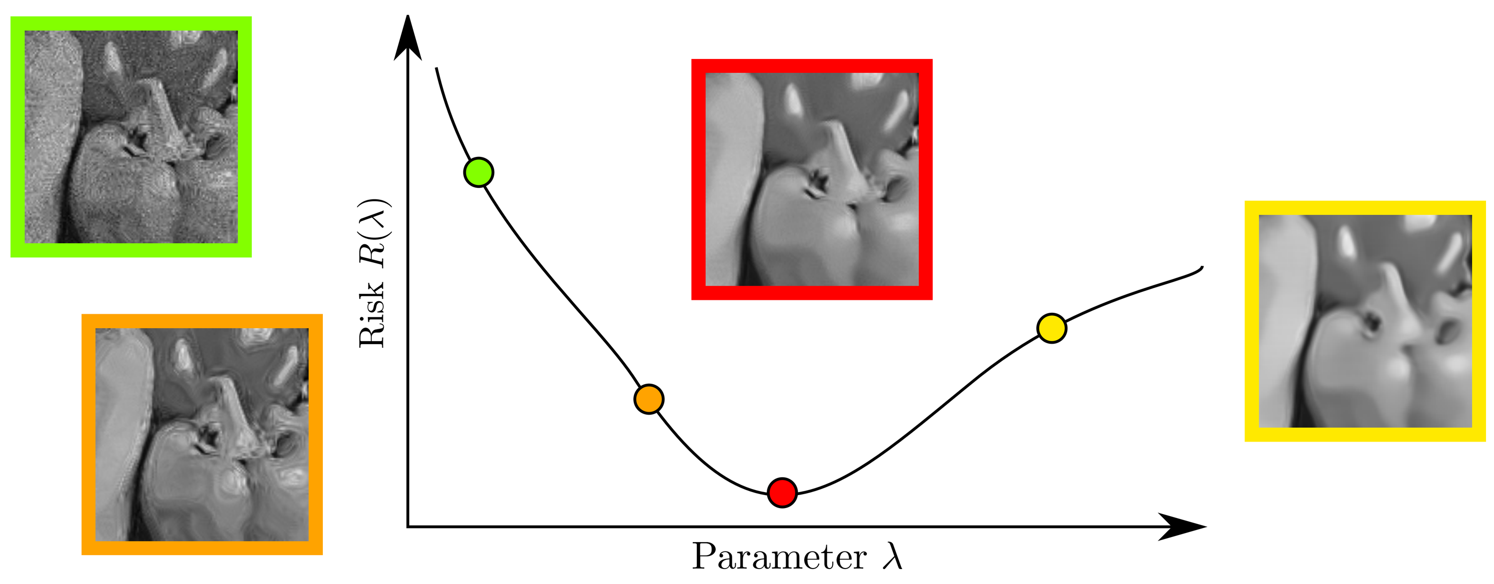
How to choose the value of the parameter λ?

Risk-based selection of λ

- Risk associated to λ: measure of the expected quality of $x^\star(y,\lambda)$ wrt x_0,

$$R(\lambda) = \mathbb{E}_w \|x^\star(y,\lambda) - x_0\|^2 .$$

- The optimal (theoretical) λ minimizes the risk.



The risk is unknown since it depends on x_0.
Can we estimate the risk solely from $x^\star(y,\lambda)$?

Risk estimation

- Assume $y \mapsto \Phi x^\star(y,\lambda)$ is weakly differentiable (*a fortiori* uniquely defined).

Prediction risk estimation via SURE

- The Stein Unbiased Risk Estimator (SURE):

$$\mathrm{SURE}(y,\lambda) = \|y - \Phi x^\star(y,\lambda)\|^2 - \sigma^2 Q + 2\sigma^2 \underbrace{\operatorname{tr}\left(\frac{\partial \Phi x^\star(y,\lambda)}{\partial y}\right)}_{\text{Estimator of the DOF}}$$

is an unbiased estimator of the prediction risk [Stein, 1981]:

$$\mathbb{E}_w(\mathrm{SURE}(y,\lambda)) = \mathbb{E}_w(\|\Phi x_0 - \Phi x^\star(y,\lambda)\|^2) .$$

Projection risk estimation via GSURE

- Let $\Pi = \Phi^*(\Phi\Phi^*)^+\Phi$ be the orthogonal projector on $\ker(\Phi)^\perp = \mathrm{Im}(\Phi^*)$,
- Denote $x_{\mathrm{ML}}(y) = \Phi^*(\Phi\Phi^*)^+ y$,
- The Generalized Stein Unbiased Risk Estimator (GSURE):

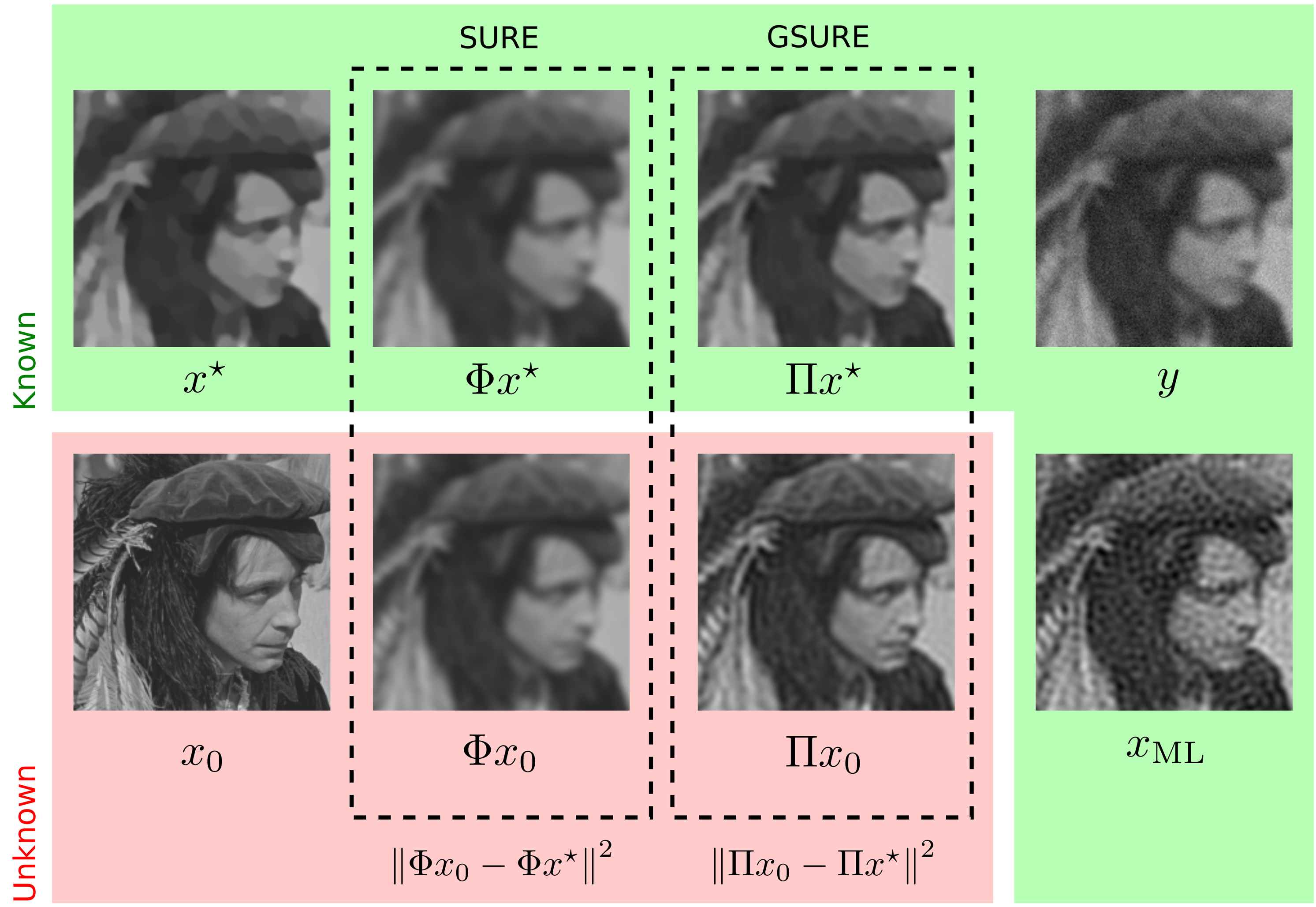
$$\mathrm{GSURE}(y,\lambda) = \|x_{\mathrm{ML}}(y) - \Pi x^\star(y,\lambda)\|^2 - \sigma^2 \operatorname{tr}((\Phi\Phi^*)^+) + 2\sigma^2 \operatorname{tr}\left((\Phi\Phi^*)^+ \frac{\partial \Phi x^\star(y,\lambda)}{\partial y}\right)$$

is an unbiased estimator of the projection risk [Vaiter et al., 2012]

$$\mathbb{E}_w(\mathrm{GSURE}(y,\lambda)) = \mathbb{E}_w(\|\Pi x_0 - \Pi x^\star(y,\lambda)\|^2)$$

(see also [Eldar, 2009, Pesquet et al., 2009, Vonesch et al., 2008] for similar results).

Illustration of risk estimation



(here, $x^\star$ denotes $x^\star(y,\lambda)$ for an arbitrary value of λ)

How to estimate the quantity $\operatorname{tr}\left((\Phi\Phi^*)^+ \frac{\partial x^\star(y,\lambda)}{\partial y}\right)$?

Main notations and assumptions

- Let $I = \mathrm{supp}(D^* x^\star(y,\lambda))$ be the support of $D^* x^\star(y,\lambda)$,
- Let $J = I^c$ be the co-support of $D^* x^\star(y,\lambda)$,
- Let D_I be the submatrix of D whose columns are indexed by I,
- Let $s_I = \mathrm{sign}(D^* x^\star(y,\lambda))_I$ be the subvector of $D^* x^\star(y,\lambda)$ whose entries are indexed by I,
- Let $\mathcal{G}_J = \operatorname{Ker} D_J^*$ be the "cospace" associated to $x^\star(y,\lambda)$,
- To study the local behaviour of $x^\star(y,\lambda)$, we impose Φ to be "invertible" on $\mathcal{G}_J$:

$$\mathcal{G}_J \cap \operatorname{Ker} \Phi = \{0\},$$

- It allows us to define the matrix

$$A^{[J]} = U(U^* \Phi^* \Phi U)^{-1} U^*,$$

where U is a matrix whose columns form a basis of $\mathcal{G}_J$,
- In this case, we obtain an implicit equation:

$$x^\star(y,\lambda) \text{ solution of } \mathcal{P}_\lambda(y) \Leftrightarrow x^\star(y,\lambda) = \hat{x}(y,\lambda) \triangleq A^{[J]}\Phi^* y - \lambda A^{[J]} D_I s_I .$$

Is this relation true in a neighbourhood of (y,λ)?

Theorem (Local Parameterization)

- *Even if the solutions $x^\star(y,\lambda)$ of $\mathcal{P}_\lambda(y)$ might be not unique, $\Phi x^\star(y,\lambda)$ is uniquely defined.*
- *If $(y,\lambda) \notin \mathcal{H}$, for $(\bar{y},\bar{\lambda})$ close to (y,λ), $\hat{x}(\bar{y},\bar{\lambda})$ is a solution of $\mathcal{P}(\bar{y},\bar{\lambda})$ where*

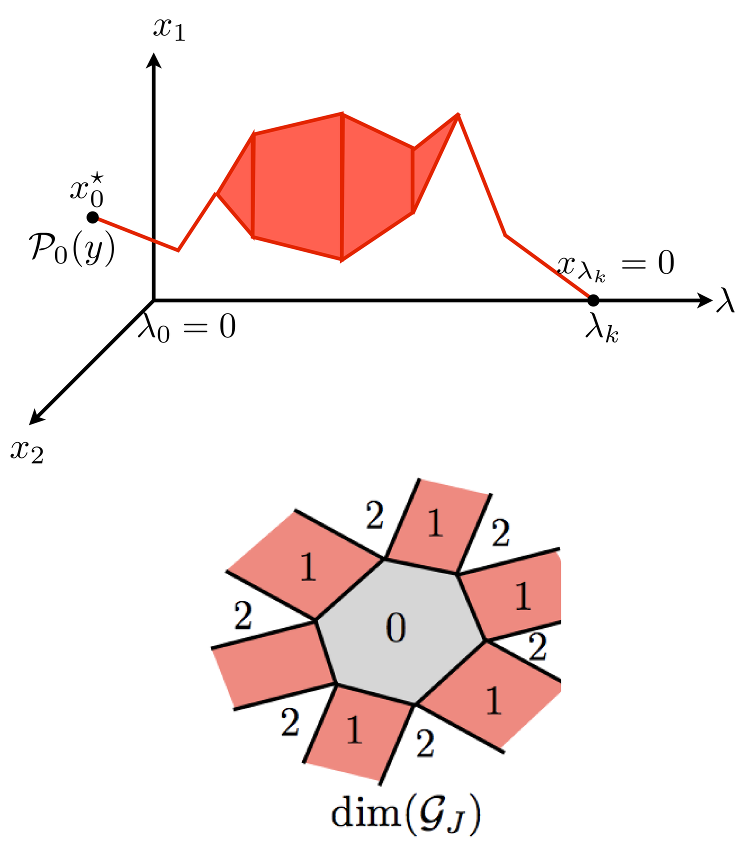
$$\hat{x}(\bar{y},\bar{\lambda}) = A^{[J]}\Phi^* \bar{y} - \bar{\lambda} A^{[J]} D_I s_I .$$

- *Hence, it allows us writing*

$$\frac{\partial \Phi x^\star(y,\lambda)}{\partial y} = \Phi A^{[J]} \Phi^* ,$$

- *Moreover, the DOF can be estimated by*

$$\operatorname{tr}\left(\frac{\partial \Phi x^\star(y,\lambda)}{\partial y}\right) = \dim(\mathcal{G}_J) .$$



Can we compute this quantity efficiently?

Computation of GSURE

- One has for $Z \sim \mathcal{N}(0, \mathrm{Id}_P)$,

$$\operatorname{tr}\left((\Phi\Phi^*)^+ \frac{\partial \Phi x^\star(y,\lambda)}{\partial y}\right) = \mathbb{E}_Z(\langle \nu(Z),\, \Phi^*(\Phi\Phi^*)^+ Z \rangle)$$

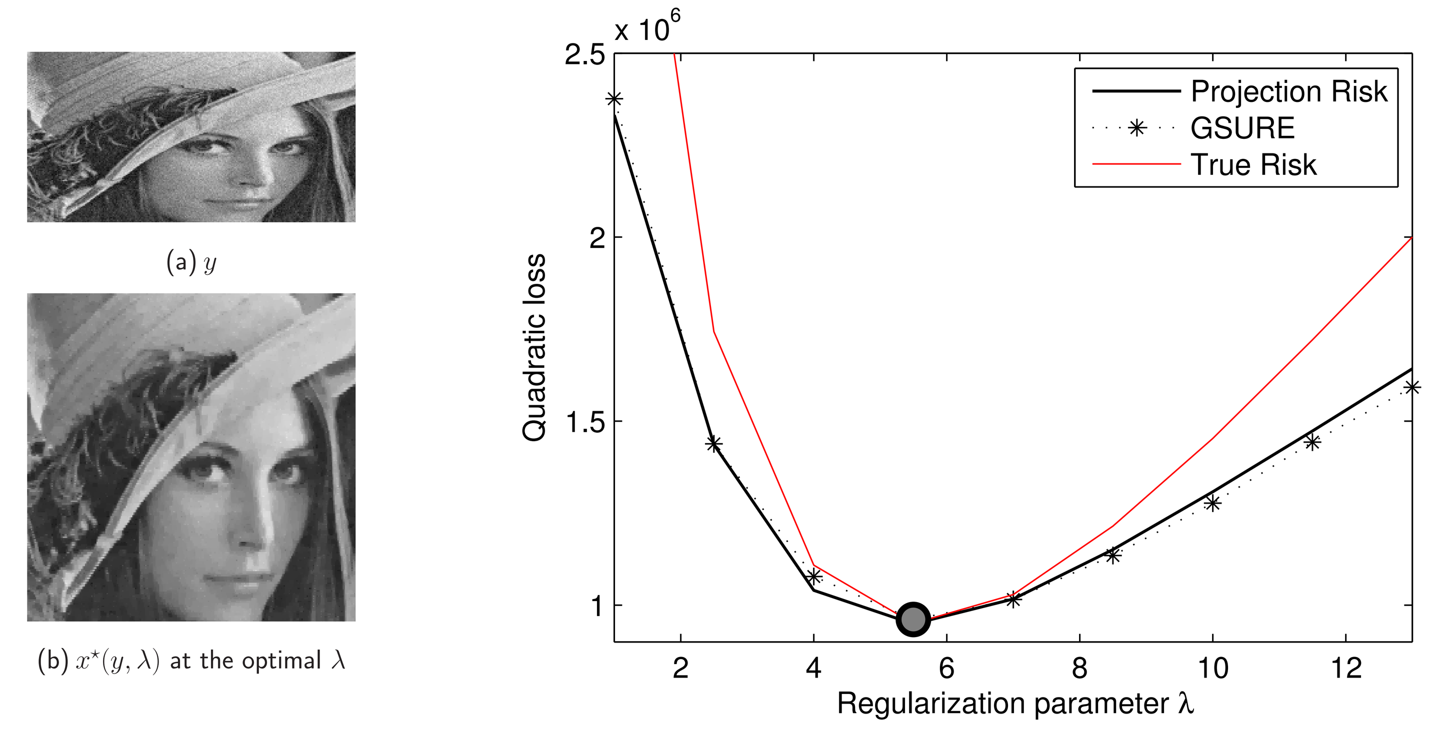
where, for any $z \in \mathbb{R}^P$, $\nu = \nu(z)$ solves the following linear system

$$\begin{pmatrix} \Phi^*\Phi & D_J \\ D_J^* & 0 \end{pmatrix} \begin{pmatrix} \nu \\ \tilde{\nu} \end{pmatrix} = \begin{pmatrix} \Phi^* z \\ 0 \end{pmatrix} .$$

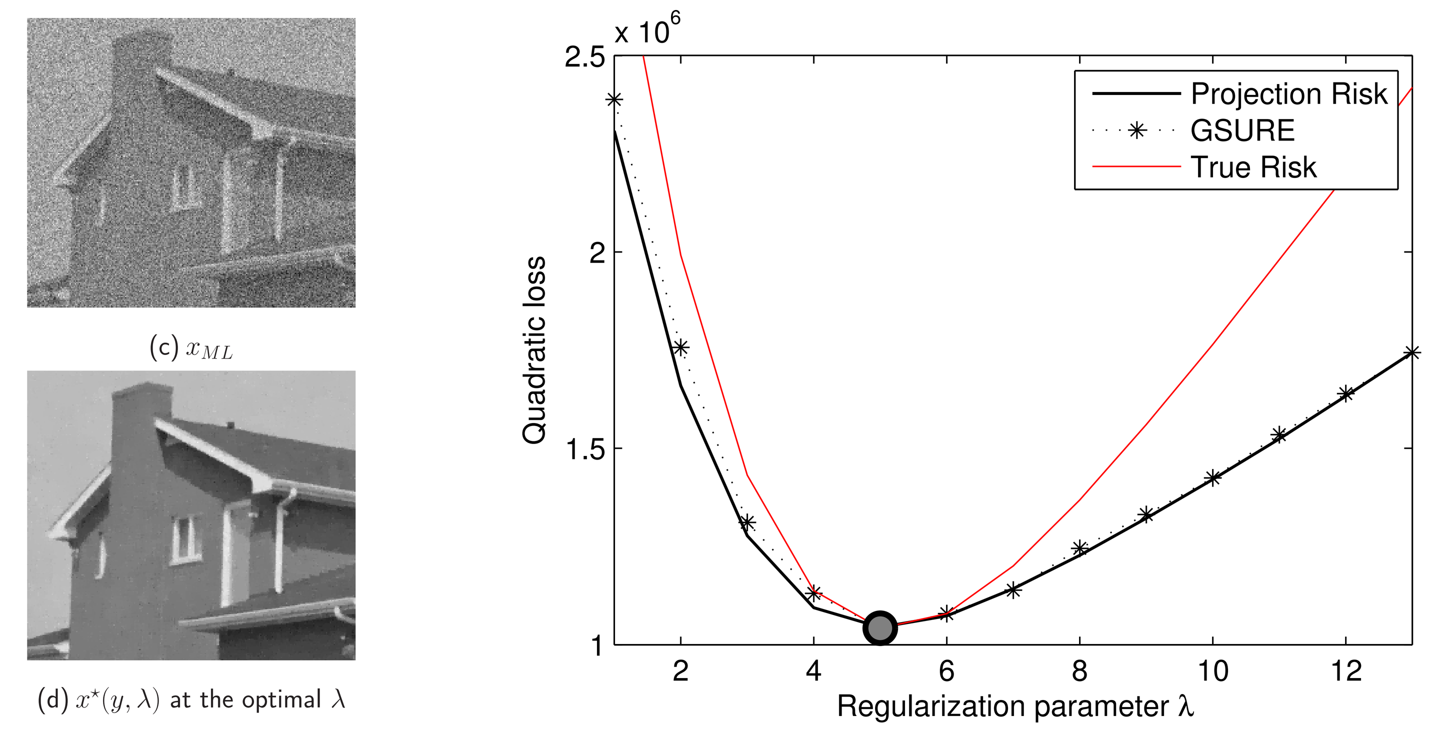
- In practice, with law of large number, the empirical mean is replaced for the expectation.
- The computation of $\nu(z)$ is achieved by solving the linear system with a conjugate gradient solver.

Numerical example

Super-resolution using (anisotropic) Total-Variation



Compressed-sensing using multi-scale wavelet thresholding



Perspectives: How to efficiently minimizes $\mathrm{GSURE}(y,\lambda)$ wrt λ?

References

Eldar, Y. C. (2009).
Generalized SURE for exponential families: Applications to regularization.
IEEE Transactions on Signal Processing, 57(2):471–481.

Pesquet, J.-C., Benazza-Benyahia, A., and Chaux, C. (2009).
A SURE approach for digital signal/image deconvolution problems.
IEEE Transactions on Signal Processing, 57(12):4616–4632.

Stein, C. (1981).
Estimation of the mean of a multivariate normal distribution.
The Annals of Statistics, 9(6):1135–1151.

Vaiter, S., Deledalle, C., Peyré, G., Dossal, C., and Fadili, J. (2012).
Local behavior of sparse analysis regularization: Applications to risk estimation.
Arxiv preprint arXiv:1204.3212.

Vonesch, C., Ramani, S., and Unser, M. (2008).
Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint.
In *ICIP*, pages 665–668. IEEE.