Sparse overcomplete denoising: aggregation versus global optimization

IEEE Signal Processing Letters, Volume 24, Issue 10, 2017

Diego Carrera, Giacomo Boracchi
Politecnico di Milano, Italy
{diego.carrera, giacomo.boracchi}@polimi.it

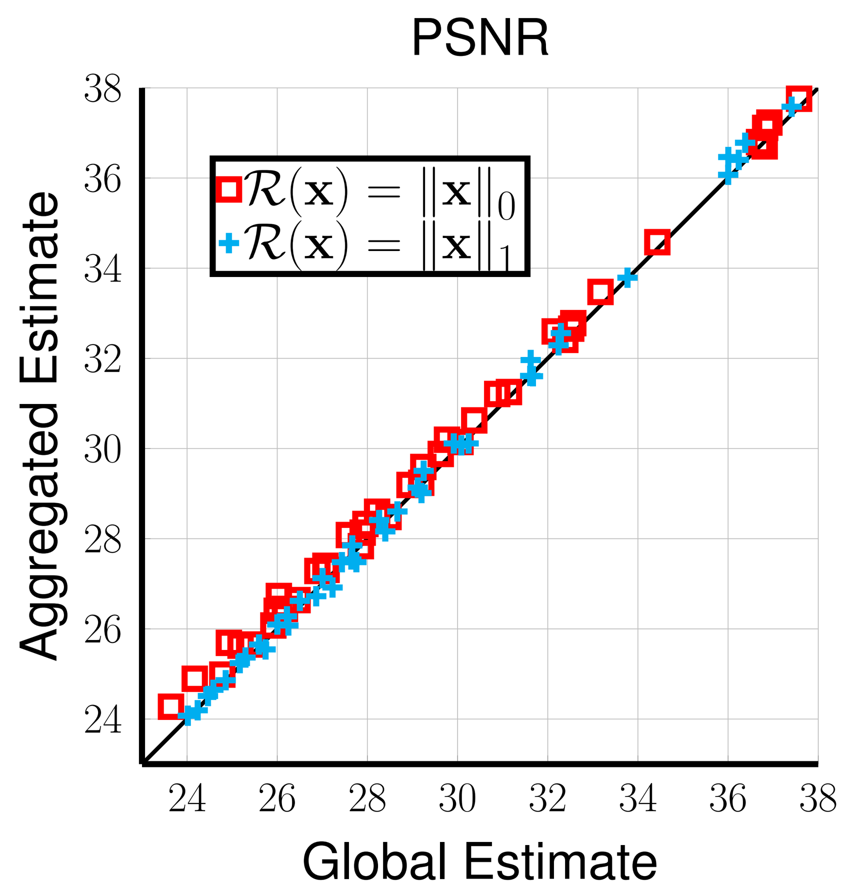
Alessandro Foi
Tampere University of Technology, Finland
alessandro.foi@tut.fi

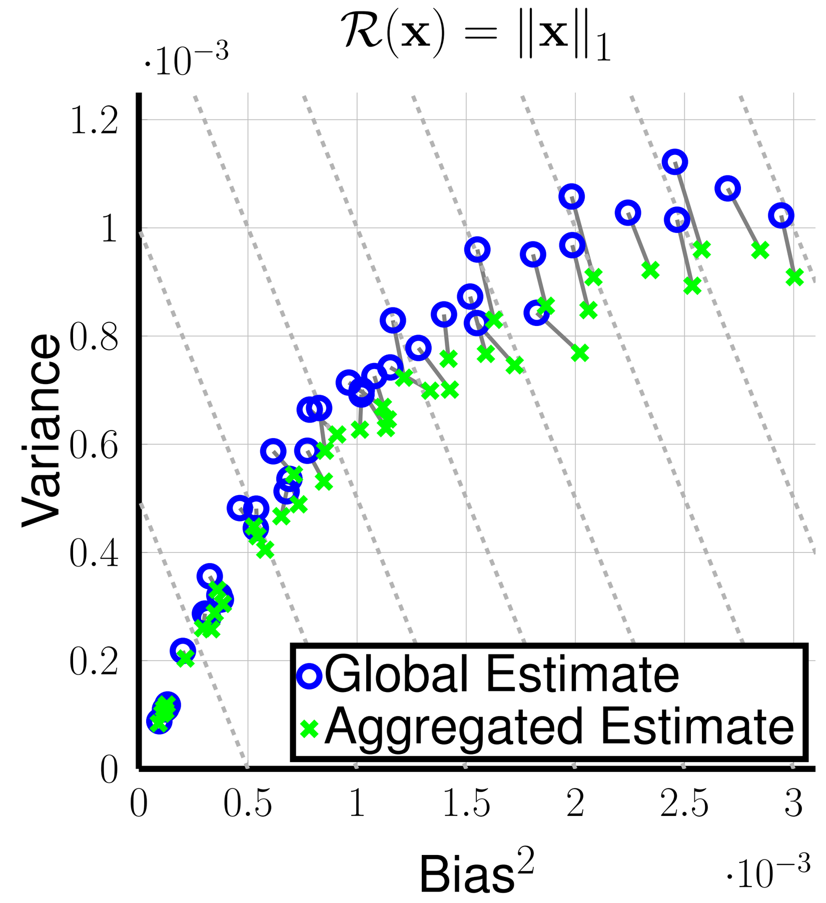
Brendt Wohlberg
Los Alamos National Laboratory, USA
brendt@lanl.gov

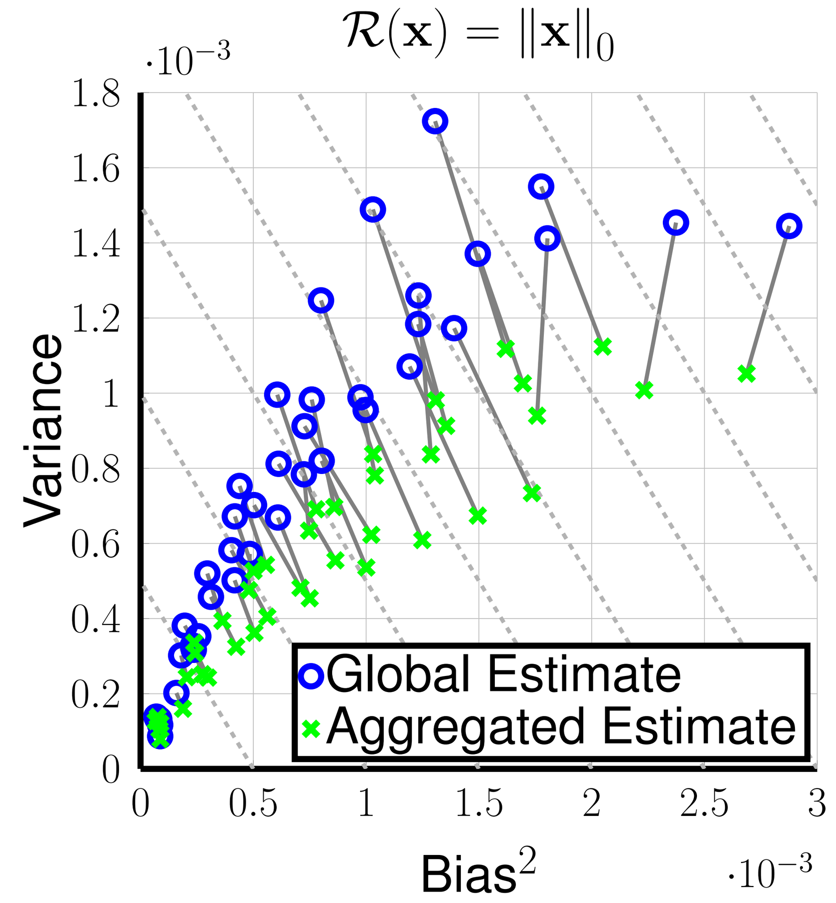
Highlights

- We consider the problem of **translation-invariant denoising** by sparse coding w.r.t. an overcomplete dictionary.
  We compare two approaches:
  - Directly solving a **global optimization** as done in *convolutional sparse coding*.
  - Solving multiple partial optimization problems and **aggregate the partial estimates** as in *cycle spinning*.
- We analyze both approaches by decomposing their mean squared error into the **bias** and **variance** components.
- We show that on natural images **global optimization features a lower bias and larger variance** than aggregation of partial estimates.
- **Global optimization is superior only when images admit a very sparse representation, while for natural images the two perform comparably**.

Results on Natural Images







Results under Extreme Sparsity

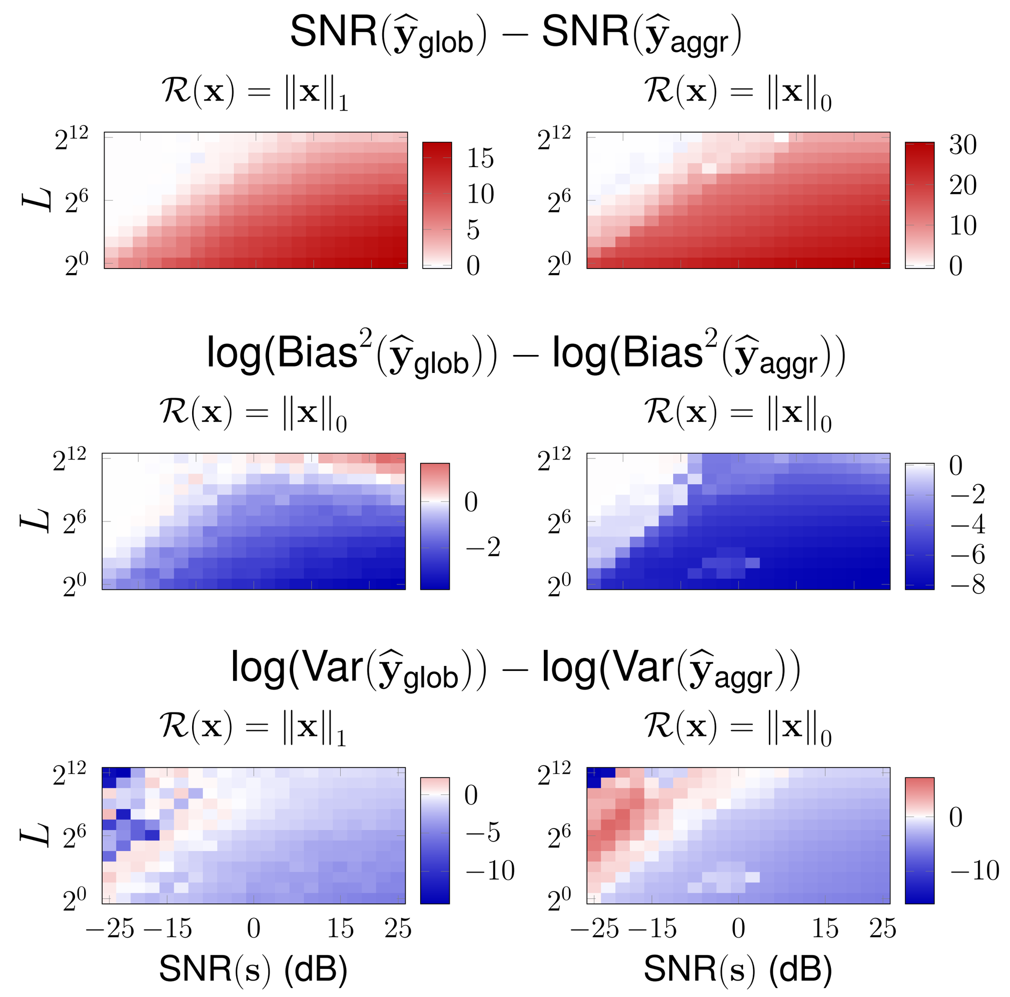


Image Denoising

We consider images corrupted by white Gaussian noise:

$$\mathbf{s} = \mathbf{y} + \boldsymbol{\eta}, \qquad \boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2). \tag{1}$$

Given an **orthonormal basis** $D_1 \in \mathbb{R}^{N \times N}$ and all its **translates** $D_i \in \mathbb{R}^{N \times N}$, the denoising estimate $\widehat{\mathbf{y}}$ is computed as:

$$\widehat{\mathbf{y}} = D\widehat{\mathbf{x}}, \qquad D = (D_1 \cdots D_N) \in \mathbb{R}^{N \times N^2}, \tag{2}$$

where $\widehat{\mathbf{x}} \in \mathbb{R}^{N^2}$ is assumed to be sparse. We focus on sparse coding problem, ignoring issues related to dictionary learning.

Aggregation of Partial Estimates:

For each translate D_i, we solve:

$$\widehat{\mathbf{x}}_i = \arg\min_{\mathbf{u} \in \mathbb{R}^N} \frac{1}{2} \|D_i \mathbf{u} - \mathbf{s}\|_2^2 + \lambda \mathcal{R}(\mathbf{u}), \quad i \in \{1, \dots, N\}, \tag{3}$$

where $\mathcal{R}(\cdot) = \|\cdot\|_1$ or $\mathcal{R}(\cdot) = \|\cdot\|_0$. The final estimate $\widehat{\mathbf{y}}_{\text{aggr}}$ is obtained **aggregating** the N estimates $D_i\widehat{\mathbf{x}}_i$:

$$\widehat{\mathbf{y}}_{\text{aggr}} = \frac{1}{N} \sum_{i=1}^{N} D_i \widehat{\mathbf{x}}_i = D \frac{(\widehat{\mathbf{x}}_0^T \cdots \widehat{\mathbf{x}}_N^T)^T}{N} = D\widehat{\mathbf{x}}_{\text{aggr}}. \tag{4}$$

Global optimization:

We **jointly** consider **all the translates**:

$$\widehat{\mathbf{x}}_{\text{glob}} = \arg\min_{\mathbf{x} \in \mathbb{R}^{N^2}} \frac{1}{2} \|D\mathbf{x} - \mathbf{s}\|_2^2 + \lambda \mathcal{R}(\mathbf{x}). \tag{5}$$

This problem can be formulated in a **convolutional** form (Zeiler et al., *IEEE CVPR* 2010):

$$\widehat{\mathbf{x}}_{\text{glob}} = \arg\min_{\mathbf{x} \in \mathbb{R}^{N^2}} \frac{1}{2} \left\| \sum_{m=1}^{M} \mathbf{d}_m * \mathbf{x}_{[m]} - \mathbf{s} \right\|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \tag{6}$$

and the final estimate is

$$\widehat{\mathbf{y}}_{\text{glob}} = D\widehat{\mathbf{x}}_{\text{glob}} = \sum_{m=1}^{M} \mathbf{d}_m * \widehat{\mathbf{x}}_{[m]}. \tag{7}$$

Regularization and solutions:

When $\mathcal{R}(\mathbf{u}) = \|\mathbf{u}\|_1$ the solution of (3) is given by the **soft-thresholding** operator $\mathcal{S}_\lambda \colon \mathbb{R}^N \to \mathbb{R}^N$

$$\widehat{\mathbf{x}}_i = \mathcal{S}_\lambda(D_i^T \mathbf{s}), \qquad [\mathcal{S}_\lambda(\mathbf{u})]_j = \mathrm{sign}(u_j) \cdot \max(|u_j| - \lambda, 0), \quad j \in \{1, \dots, N\}. \tag{8}$$

Problem (5-6) can be solved using an efficient implementation in the Fourier domain of the ADMM algorithm (Wohlberg, *IEEE TIP* 2016).
When $\mathcal{R}(\mathbf{u}) = \|u\|_0$ the solution of (3) is given by the **hard-thresholding** operator $\mathcal{H}_\lambda \colon \mathbb{R}^N \to \mathbb{R}^N$

$$\widehat{\mathbf{x}}_i = \mathcal{H}_\lambda(D_i^T \mathbf{s}), \qquad [\mathcal{H}_\lambda(\mathbf{u})]_j = u_j \cdot 1_{\{|u_j| > \lambda\}}, \qquad j \in \{1, \dots, N\}. \tag{9}$$

Problem (5) can be solved using Iterative Thresholding Algorithm (Kowalski, *IEEE ICIP* 2014). However, (5) is not convex and this algorithm converges only to a **local minimum**.

Experiments and Discussions

Settings:

- D_1 is the Daubechies db3 wavelet dictionary with 4 decomposition levels.
- λ is tuned each time to achieve the lowest mean squared error.
- **Natural images**: Lena, Barbara, Man, Peppers, Cameraman corrupted by noise with $\sigma \in \{5, 10, \dots, 40\}$. Results are averaged over 50 realizations of noise.
- **Very Sparse Synthetic images**: 128×128 noise-free image $\mathbf{y} = D\mathbf{x}$, where $\mathbf{x}$ has L nonzero components at random positions. σ is such that $\mathrm{SNR}(\mathbf{s}) = \tau$. We set $L \in \{2^0, 2^1, \dots, 2^{12}\}$, $\tau \in \{-25, -22.5, \dots, 25\}$, 50 realizations of $\mathbf{y}$ for each pair (L, τ), and 50 realizations of $\mathbf{s}$ for each $\mathbf{y}$.

Results and Concluding Remarks:

- Estimates from **global optimization** are characterized by a **lower bias** and **larger variance** than the aggregation of partial estimates.
- We speculate that the **larger variance** of global estimates is due to the **high redundancy** of D, since shifted atoms are highly correlated.
- The two approaches achieve **similar performance** when $\mathcal{R} = \|\cdot\|_1$.
- When $\mathcal{R} = \|\cdot\|_0$ the variance of global estimates is even larger. This is due to the non-convexity of the global optimization problem, since we can compute only **local minima** that are subject to the particular noise realization.
- It may be unreasonable to approach the **natural image denoising** by the computationally demanding convolutional sparse coding, as this does not outperform cycle spinning.