On the Weighting for Convolutional Sparse Coding

Diego Carrera
STMicroelectronics, Italy
diego.carrera@st.com

Alessandro Foi
Tampere University, Finland
alessandro.foi@tuni.fi

Giacomo Boracchi
Politecnico di Milano, Italy
giacomo.boracchi@polimi.it

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

Introduction. We consider the recovery of a noise-free image $\mathbf{y}$ from a noisy image $\mathbf{z} \in \mathbb{R}^N$, where $\mathbf{z} = \mathbf{y} + \boldsymbol{\eta}$, and $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2)$, focusing on a recent denoising approach [1], [2] that assumes that $\mathbf{y}$ admits a sparse representation with respect to a convolutional dictionary expressed as a set of filters $\{\mathbf{d}_m\}$ [3], [4], [5]. Under this assumption, the denoised estimate of $\mathbf{y}$ is $\widehat{\mathbf{y}} = \sum_m \mathbf{d}_m * \widehat{\mathbf{x}}_m$, where $\widehat{\mathbf{x}}$ solves the convolutional sparse coding (CSC) problem

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2} \left\| \mathbf{z} - \sum_m \mathbf{d}_m * \mathbf{x}_m \right\|_2^2 + \lambda \sum_m \left\| \mathbf{w}_m \odot \mathbf{x}_m \right\|_1 , \quad (1)$$

$\odot$ denoting the element-wise product. A particularly effective heuristic criteria to set weights $\{\mathbf{w}_m\}$ in (1) has the form of correlation reciprocal [2]

$$\mathbf{w}_m[i] = \frac{1}{(\overline{\mathbf{d}}_m * \mathbf{y})[i]} = \frac{1}{(D_m^T \mathbf{y})[i]} , \quad (2)$$

where $\overline{\mathbf{d}}_m$ denotes the conjugate filter of $\mathbf{d}_m$, and D_m denotes the corresponding Toepliz matrix. In practice, smaller weights are being assigned where the image is more highly correlated with the filter $\mathbf{d}_m$. Directly implementing (2) is not possible as it involves $\mathbf{y}$, which thus needs to be replaced by a previous estimate. Here we investigate the rationale underpinning the weighting criteria in (2), and its connection with the WaveShrink algorithm [6], which is related to CSC, but more amenable to analysis due to its simpler form.

Oracle Thresholds for WaveShrink. The classical WaveShrink algorithm [6] solves the optimization problem

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2} \left\| z - D\mathbf{x} \right\|_2^2 + \left\| \mathbf{w} \odot \mathbf{x} \right\|_1 , \quad (3)$$

where $D \in \mathbb{R}^{N \times N}$ is an orthonormal dictionary, thus (3) is solved by applying the soft thresholding to each component of $D^T\mathbf{z}$:

$$\widehat{\mathbf{x}}[i] = \mathcal{S}_{\mathbf{w}[i]}((D^T\mathbf{z})[i]) , \quad (4)$$

where $\mathbf{w}[i]$ is the threshold used for the i-th component of $D^T\mathbf{z}$ and $\mathcal{S}_\omega(u) = \operatorname{sign}(u) \cdot \max(|u| - \omega, 0)$, for $\omega > 0$ and $u \in \mathbb{R}$. The estimated image is then $\widehat{\mathbf{y}} = D\widehat{\mathbf{x}}$.

In WaveShrink, the *oracle* weights $\mathbf{w}[i]$ can be defined as

$$\mathbf{w}[i] = \varphi_\sigma((D^T\mathbf{y})[i]), \quad (5)$$

where

$$\varphi_\sigma(x) = \arg\min_\omega \mathrm{MSE}_\sigma(\omega, x), \quad (6)$$

and the function $\mathrm{MSE}_\sigma(\omega, x) = \mathrm{E}\left\{ \left(\mathcal{S}_\omega((D^T\mathbf{z})[i]) - (D^T\mathbf{y})[i] \right)^2 \right\}$ denotes the mean square error between the estimated $\widehat{\mathbf{x}}[i]$ (4) and the corresponding noise-free coefficient $x = \mathbf{x}[i] = (D^T\mathbf{y})[i]$.

This rather simple denoising framework allows to derive the closed-form expression as in [6]

$$\begin{aligned} \mathrm{MSE}_\sigma(\omega, x) = {} & \sigma^2 + \omega^2 + \sigma^2(x - \omega)\phi_\sigma(-\omega - x) + \\ & + \sigma^2(-\omega - x)\phi_\sigma(\omega - x) + \\ & + (x^2 - \omega^2 - \sigma^2)\left[\Phi_\sigma(\omega - x) - \Phi_\sigma(-\omega - x)\right] , \end{aligned} \quad (7)$$

where ϕ_σ and Φ_σ respectively denote the probability density and the cumulative distribution function of $\mathcal{N}(0, \sigma^2)$. Fig. 1 illustrates MSE_1 and the oracle weights φ_1 obtained by numerical minimization of (7). Asymptotic analysis shows that $\varphi_\sigma(x) \sim 2\sigma^2\phi_\sigma(x)$ for large x, whereas $\varphi_\sigma(x) \sim \sigma^2 c/x$ for small x. Here c is the fixed point of the hyperbolic cotangent, i.e. $c = \operatorname{cotanh}(c)$, and approximately $c = 1.2$.

Oracle Waveshrink and Correlation Reciprocal Weights. We observe that the asymptotic rate for small x of the oracle weights matches that of the correlation reciprocal (2) in case of WaveShrink. Fig. 2 gives a further insight on this relation by comparing the output of soft thresholding (4) using either weights defined by the correlation reciprocal (2) (Fig. 2.b) or the oracle WaveShrink weights (5) (Fig. 2.c). We note that neither of the two corresponds to the standard soft thresholding (Fig. 2.a), since the threshold for $D^T\mathbf{z}[i]$ is selected based on the value $D^T\mathbf{y}[i]$, making them akin to a nonlinear counterpart of the classical Wiener filtering $D^T\mathbf{z}\frac{|x|^2}{|x|^2+\sigma^2}$, which is a linear minimum MSE estimator. Careful inspection of Figs. 1 and 2 also reveals that the two weighting schemes mainly differ at the plateau of the WaveShrink MSE surface, from which one can conclude that these differences are of minimal consequence and that the correlation reciprocal is essentially a close approximation of the oracle WaveShrink weights.

The same conclusions hold for other values of σ, as it can be shown that $\mathrm{MSE}_\sigma(\omega, x) = \mathrm{MSE}_1(\omega/\sigma, x/\sigma)\sigma^2$ and $\varphi_\sigma(x) = \varphi_1(x/\sigma)\sigma$, for any $\sigma > 0$ and $x, \omega \in \mathbb{R}$. Therefore, the surface and the curves in Fig. 1 can be obtained for $\sigma \neq 1$ through simple rescaling.

Weighting Scheme for CSC. CSC can be regarded as a generalization of the optimization problem solved by WaveShrink to the case of overcomplete and translation-invariant dictionaries. Therefore we define the weights for CSC by means of the function $\varphi_\sigma(x)$, even though here this does not guarantee the same optimality properties:

$$\mathbf{w}_m[i] = \varphi_\sigma((\overline{\mathbf{d}}_m * \mathbf{y})[i]). \quad (8)$$

The trend of φ_σ in Fig. 1 suggests that $\mathbf{w}_m[i]$ is small when $(\overline{\mathbf{d}}_m * \mathbf{y})[i]$ is large, and vice versa. In the following experiments we show that this weighting scheme performs very similarly to (2), in agreement with our previous analysis on orthonormal dictionaries.

Experiments. We consider natural image denoising through CSC with the weighting schemes in (2) and (8). As customary in sparsity-based denoising, the CSC is performed on a high-pass version of $\mathbf{z}$, preserving the complementary low-pass component. We compute a pilot estimate $\widehat{\mathbf{y}}_{\text{pilot}}$ by solving (1) with $\mathbf{w}_m[i] = 1$, then, we replace $\mathbf{y}$ with $\widehat{\mathbf{y}}_{\text{pilot}}$ in (2) and (8).

Fig. 3 shows the PSNR averaged over 50 noise realizations for the two considered weighting schemes and the pilot estimate (*Fixed Weights*). The performance of the two weighting schemes are very similar and both yield a substantial improvement w.r.t. the pilot estimate, consistently yielding an extra 0.5 dB for all values of σ.

Conclusions. Our study shows that the performance of convolutional sparse denoising can be substantially improved by suitable weighting schemes. We also show that, in the case of orthonormal dictionaries, the correlation reciprocal (the most effective weighting scheme in CSC), yields weights that are very similar to the oracle weights in WaveShrink. Moreover, when these oracle weights are employed in CSC, they provide very similar denoising performance to the correlation reciprocal.

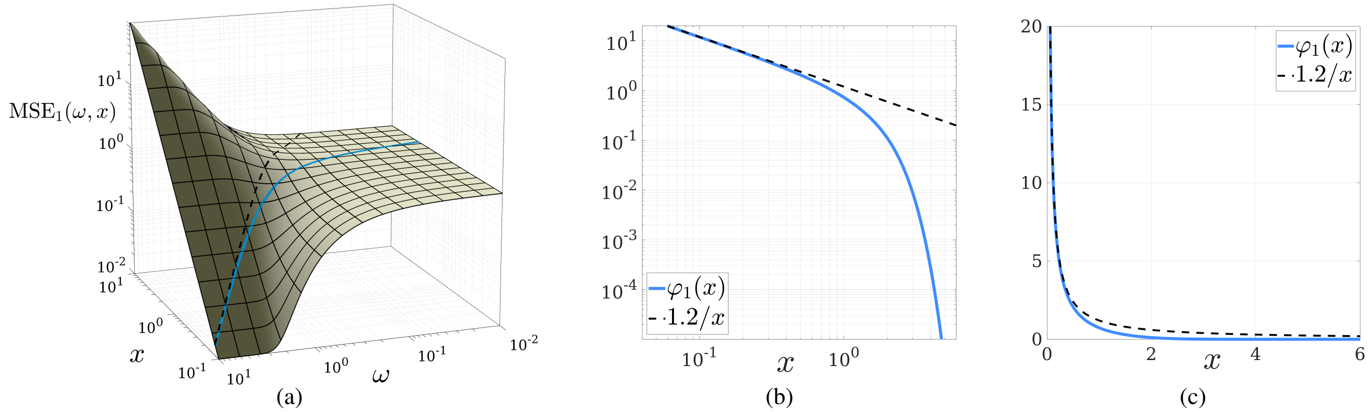

Fig. 1. (a) WaveShrink MSE_1 surface (7); the oracle minimizer φ_1 (6) is drawn in blue over the surface whereas the black dashed line corresponds to c/x, $c = 1.2$ being the fixed point of the hyperbolic cotangent; c/x is asymptotic to $\varphi_1(x)$ for small x, as shown also in the plots (b) and (c).

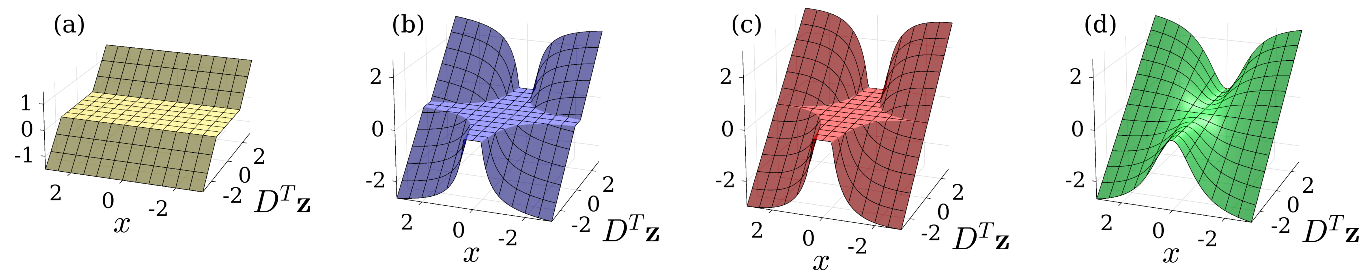

Fig. 2. Comparison of soft-thresholding output $\mathcal{S}_\omega(D^T\mathbf{z})$ with fixed threshold (weight) $\omega = 1.5\sigma$ (a), with weight $\omega = \frac{1}{x}$ defined by correlation reciprocal (2) (b), oracle WaveShrink weights $\omega = \varphi_\sigma(x)$ (5) (c), and the output of Wiener filtering $D^T\mathbf{z}\frac{|x|^2}{|x|^2+\sigma^2}$ (d). In all cases $\sigma = 1$.

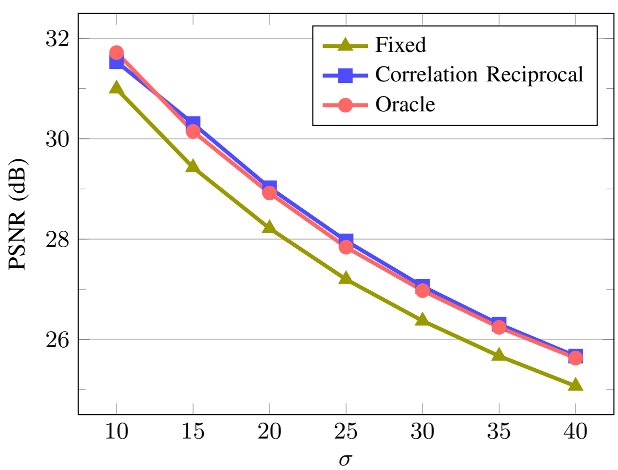

Fig. 3. PSNR averaged achieved by CSC leveraging three weighting schemes: fixed weights, correlation reciprocal (2), and oracle WaveShrink weights (5). The latter two were computed using the CSC estimate with fixed weights in place of the oracle in order to define the weights. The PSNR is averaged over five test images (Lena, Barbara, Man, Peppers, Cameraman) corrupted with Gaussian noise with standard deviation $\sigma \in \{10, 15, \ldots, 40\}$. The filters $\{\mathbf{d}_m\}$ are the synthesis filter of the Daubechies db3 wavelet with 4 decomposition levels. We follow the procedure suggested in [1] and compute the CSC on a high-pass version $\mathbf{z}_h$ of the noisy image $\mathbf{z}$.

References

[1] D. Carrera, G. Boracchi, A. Foi, and B. Wohlberg, "Sparse overcomplete denoising: aggregation versus global optimization," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1468–1472, 2017.

[2] B. Wohlberg, "Convolutional sparse coding with overlapping group norms," arXiv, Tech. Rep. 1708.09038, Aug. 2017.

[3] M. S. Lewicki and T. J. Sejnowski, "Coding time-varying signals using sparse, shift-invariant representations," in *Advances in Neural Information Processing Systems*, 1999, pp. 730–736.

[4] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks." in *CVPR*, 2010, p. 7.

[5] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301–315, 2016.

[6] A. G. Bruce and H.-Y. Gao, "Understanding WaveShrink: Variance and bias estimation," *Biometrika*, vol. 83, no. 4, pp. 727–745, 1996.