A Robust Fully Correntropy–based Sparse Modeling Alternative to Dictionary Learning

Carlos A. Loza

Abstract Correntropy is a dependence measure that goes beyond Gaussian environments and optimizations based on Minimum Squared Error (MSE). Its ability to induce a metric that is fully modulated by a single parameter makes it an attractive tool for adaptive signal processing. We propose a sparse modeling framework based on the dictionary learning technique known as K–SVD where Correntropy replaces MSE in the sparse coding and dictionary update subroutines. The former yields a robust variant of Orthogonal Matching Pursuit while the latter exploits robust Singular Value Decompositions. The result is Correntropy–based dictionary learning. The data–driven nature of the approach combines two appealing features in unsupervised learning—robustness and sparseness—without adding hyperparameters to the framework. Robust recovery of bases in synthetic data and image denoising under impulsive noise confirm the advantages of the proposed techniques.

1 Introduction

Sparse modeling refers to the mechanisms involved in the learning of a linear generative model where the inputs are the result of sparse activations of selected vectors from an overcomplete basis. Its rationale comes from principles of parsimony where it is advantageous to represent a given phenomenon with as few variables as possible. Sparse modeling has been particularly appealing to two fields with (usually) different objectives and methodologies: statistics [2, 18, 4] and signal processing [13, 6, 5]. In neuroscience, Olshausen and Field [15] paved the way to what is currently known as dictionary learning—instead of using a fixed off–the–shelf basis, the authors proposed a fully data–driven learning scheme to estimate said basis, also known as dictionary.

Carlos A. Loza
Department of Mathematics. Universidad San Francisco de Quito. Quito, Ecuador, e-mail: cloza@usfq.edu.ec

In Image Processing and Computer Vision, sparse modeling is exploited for denoising [16, 7], inpainting [12], and demosaicking [11]. Most of these applications rely on K–SVD [1]—the well known dictionary learning technique that exploits a block coordinate descent approach to reach a local stationary point of a constrained optimization problem. Results are optimal under additive homogeneous Gaussian noise. Yet, if the underlying error deviates from normality, e.g. in the presence of outliers in the form of missing pixels or impulsive noise, the optimizers might introduce a bias.

Robust estimators are a principled scheme to deal with outliers in linear regimes [3]. One variant of said techniques is based on Correntropy [9]—the dependence measure that goes beyond Gaussian environments and their associated criterion for maximum likelihood estimation: Minimum Squared Error (MSE). Correntropy mimics induced metrics that are fully regulated via one main hyperparameter; if said scale parameter is chosen properly, the induced metric is robust against outliers. We harness such property to propose a novel dictionary learning approach where MSE criteria of K–SVD are replaced by robust metrics based on Correntropy. The result is Correntropy–based Dictionary Learning, or CDL.

Likewise K-SVD, CDL exploits fast sparse code estimators, such as Orthogonal Matching Pursuit (OMP), and iterative Singular Value Decompositions (SVD). Wang et. al proposed a Correntropy variant of OMP known as CMP [20] while Loza and Principe devised CK–SVD, a robust alternative to MSE–based SVD [10]. In the current work, we combine both approaches in a fully robust Correntropy–based sparse modeling framework for linear generative models. Synthetic data and image denoising under non–homogeneous impulsive noise confirm the robustness and sparseness of the solutions in addition to their superiority over K–SVD. The rest of the paper is organized as follows: Section 2 details the problem of robust dictionary learning and the proposed solutions. Section 3 summarizes the results, and, lastly, Section 4 concludes the paper and mentions potential further work.

2 Correntropy–based Sparse Modeling

Let $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N}$, $(\mathbf{y}_i \in \mathbb{R}^n)$ be a set of observations or measurements where each vector can be encoded as a sparse linear combination of predictors, also known as atoms, from an overcomplete basis, or dictionary $\mathbf{D} \in \mathbb{R}^{n \times K}$:

$$\mathbf{y} = \mathbf{D}\mathbf{x}_0 + \mathbf{n} \quad \text{s.t.} \quad ||\mathbf{x}_0||_0 = T_0 \tag{1}$$

where T_0 is the support of the ideal sparse decomposition $\mathbf{x}_0$, $||\cdot||_0$ represents the ℓ_0–pseudonorm, and $\mathbf{n}$ is the additive noise. The sparse coding problem aims to estimate $\mathbf{x}_0$ given $\mathbf{y}$ and a sparsity constraint. The sparse modeling problem generalizes to a full generative model where both sparse code and dictionary are unknown. Then for $\mathbf{Y}$, the constrained optimization becomes:

$$\min_{\mathbf{D},\mathbf{X}}\{||\mathbf{Y}-\mathbf{D}\mathbf{X}||_F^2\} \qquad \text{s. t.} \qquad \forall i,\ ||\mathbf{x}_i||_0 \leq T_0 \tag{2}$$

where $\mathbf{x}_i$ is the sparse code corresponding to the $\mathbf{y}_i$ entry and $||\cdot||_F$ stands for the Frobenius norm. The performance surface in (2) is non–convex; hence, typical greedy techniques are adopted instead. In this case, K–SVD [1] generalizes k–means by alternating between finding sparse codes, i.e. distributed representations of the inputs, and dictionary update in the form of SVD routines in a atom–by–atom scheme. Even though K–SVD admits any off–the–shelf sparse coding technique, Orthogonal Matching Pursuit (OMP) [19] is usually preferred due to its convergence properties, efficiency, and simple, intuitive implementation.

OMP and SVD–based routines are anchored on the underlying assumption of Gaussian errors. The former exploits Ordinary Least Squares (OLS) to sequentially update the active set of atoms, while the latter utilizes MSE as the cost function to update the dictionary elements. Both approaches are destined to introduce biases in the presence of outliers. We circumvent the Gaussianity assumption while incorporating robustness into the sparse modeling framework by exploiting Correntropy as the cost function in both K–SVD stages.

2.1 Correntropy–based OMP

OMP [19] aims to find a local solution to the sparse coding problem by iteratively selecting the most correlated atom in $\mathbf{D}$ to the current residual, i.e. for the j-th iteration:

$$\lambda_j = \underset{i\in\Omega}{\operatorname{argmax}}\,|\langle \mathbf{r}_{j-1}, \mathbf{d}_i\rangle| \tag{3}$$

where $\mathbf{r}_0 = \mathbf{y}$, $\Omega = \{1,2,\cdots,K\}$, $\mathbf{d}_i$ is the i-th column of $\mathbf{D}$, and $\langle\cdot,\cdot\rangle$ denotes the inner product operator. The resulting atom is then added to the active set via $\Lambda_j = \Lambda_{j-1}\cup\lambda_j$.

Lastly, the sparse code is updated as:

$$\mathbf{x}_j = \underset{\mathbf{x}\in\mathbb{R}^K,\operatorname{supp}(\mathbf{x})\subset\Lambda_j}{\operatorname{argmin}} ||\mathbf{y}-\mathbf{D}\mathbf{x}||_2 \tag{4}$$

which is solved via OLS. $\mathbf{r}_j$ is then updated as $\mathbf{r}_j = \mathbf{y}-\mathbf{D}\mathbf{x}_j$. Usually, OMP runs for a fixed number of iterations, L, or until the norm of the residue reaches a predefined threshold. The sparse code of (4) would be severely biased in the presence of outliers, i.e. each dimension in the input space would be equally weighted as a result of a non–robust estimator.

Correntropy [9] gauges the non–linear interactions between two random variable, X and Y, via a mapping to a reproducing kernel Hilbert space (RKHS):

$$V_\sigma(X,Y) = \mathbb{E}[g_\sigma(X-Y)] \tag{5}$$

where g_σ is the Gaussian kernel $g_\sigma(t) = \exp(-t^2/2\sigma^2)$ and σ, the kernel bandwidth, modulates the norm Correntropy will mimic (also known as CIM or Correntropy Induced Metric). The metric ranges from the ℓ_0–pseudonorm for small σ, ℓ_1–norm for increasing σ and ℓ_2–norm (defaults to MSE criterion) for large bandwidths. Hence, a proper choice of σ is able to incorporate robustness into the learning framework.

Correntropy Matching Pursuit (CMP) [20] replaces the MSE criterion of (4) by the robust CIM criterion, i.e.:

$$\mathbf{x}_j = \underset{\mathbf{x}\in\mathbb{R}^K,\text{supp}(\mathbf{x})\subset\Lambda_j}{\text{argmin}} \quad L_\sigma(\mathbf{y}-\mathbf{D}\mathbf{x}) \tag{6}$$

where $L_\sigma(\mathbf{e}) = \frac{1}{n}\sum_{i=1}^{n}\sigma^2(1-g_\sigma(\mathbf{e}[i]))$ is the simplified version of the CIM sample estimator. The non–convex nature of the CIM demands for alternative optimization techniques. As in [20], Half–Quadratic (HQ) optimization [14] yields a local minimum of the cost function via iterative minimizations of a convex enlarged parameter cost. The resulting adaptive σ hyperparameter, the weight vector that assesses the nature of the inputs (e.g. outliers vs. inliers), the sparse code for OMP iteration j, and the updated residue are estimated as:

$$\sigma_j^{(t+1)} = \left(\frac{1}{2n}\left|\left|\mathbf{y}-\mathbf{D}\mathbf{x}_j^{(t+1)}\right|\right|_2^2\right)^{\frac{1}{2}} \tag{7}$$

$$\mathbf{w}_j^{(t+1)}[i] = g_\sigma\left(\mathbf{y}[i]-\left(\mathbf{D}\mathbf{x}_j^{(t)}\right)[i]\right), \quad i=1,2,\ldots,n \tag{8}$$

$$\mathbf{x}_j^{(t+1)} = \underset{\mathbf{x}\in\mathbb{R}^K,\text{supp}(\mathbf{x})\subset\Lambda_j}{\text{argmin}} \left|\left|\sqrt{\mathbf{W}_j^{(t+1)}}(\mathbf{y}-\mathbf{D}\mathbf{x}_j)\right|\right|_2^2 \tag{9}$$

$$\mathbf{r}_j = \sqrt{\mathbf{W}_j}(\mathbf{y}-\mathbf{D}\mathbf{x}_j) \tag{10}$$

where t is the HQ iteration and $\mathbf{W}_j$ is the diagonal matrix version of $\mathbf{w}_j$. The theory behind HQ guarantees convergence of the sequences in question [14], i.e. $\lim_{t\to\infty}\mathbf{x}_j^{(t)} = \mathbf{x}_j$ and $\lim_{t\to\infty}\mathbf{w}_j^{(t)} = \mathbf{w}_j$. Eq. (9) is solved via classic OLS; hence the whole approach can also be framed as a weighted least squares problem. Hence, CMP weighs the inputs according to a Gaussian kernel—it emphasizes components from the underlying model family (linear in this case) and diminishes the influence of outliers. The result is a robust sparse code.

2.2 Correntropy–based Dictionary Update

K–SVD [1] is data–driven dictionary learning technique that exploits block coordinate descent to obtain a stationary point of (2). In practice, K–SVD alternates between sparse coding and dictionary element updates. The latter subroutine assumes both $\mathbf{X}$ and $K-1$ columns of $\mathbf{D}$ are fixed; then, the atom in question, $\mathbf{d}_k$, alongside

its support, i.e. $\mathbf{x}_T^k$ (k-th row of $\mathbf{X}$), are updated as:

$$\begin{aligned} ||\mathbf{Y}-\mathbf{D}\mathbf{X}||_F^2 &= \left|\left|\mathbf{Y}-\sum_{j=1}^{K}\mathbf{d}_j\mathbf{x}_T^j\right|\right|_F^2 \\ &= \left|\left|(\mathbf{Y}-\sum_{j\neq k}\mathbf{d}_j\mathbf{x}_T^j)-\mathbf{d}_k\mathbf{x}_T^k\right|\right|_F^2 \\ &= ||\mathbf{E}_k-\mathbf{d}_k\mathbf{x}_T^k||_F^2 \end{aligned} \tag{11}$$

where $\mathbf{E}_k$ is the error when the k-th atom is removed. The updated vector is estimated via SVD of $\mathbf{E}_k^R \in \mathbb{R}^{n\times m}$, which is a restricted version of the error matrix that only preserves the columns of $\mathbf{E}_k$ currently active for $\mathbf{d}_k$.

SVD is optimal only under the MSE criterion. Correntropy K–SVD or CK–SVD [10] replaces the SVD routines by robust alternatives that exploit the principle of Maximum Correntropy Criterion (MCC) [9]. Specifically, let $\mathbf{e}_i$ be the i–th column of $\mathbf{E}_k^R$ and $\mathbf{v}_i$ its low dimensional representation linearly mapped via the orthonormal projection matrix $\mathbf{U}$. The goal is to maximize a novel cost function $J(\mathbf{U})$ that mitigates the effect of outliers during said projection:

$$J(\mathbf{U}) = \sum_{i=1}^{m} g_\sigma(\mathbf{e}_i - \mathbf{U}\mathbf{v}_i) \tag{12}$$

As proposed by He et. al [8], HQ optimization is exploited to enlarge the parameter space and admit an iterative scheme that guarantees convergence to a local maximum. The adaptive σ hyperparameter, the weight vector that determines the influence of $\{\mathbf{e}_i\}_{i=1}^m$, and the projection matrix are equal to:

$$\left(\sigma_k^{(t)}\right)^2 = 1.06\times\min\left\{\sigma_E, \frac{R}{1.34}\right\}\times(m)^{-1/5} \tag{13}$$

$$\mathbf{p}_k^{(t+1)}[i] = -g\left(\sqrt{\mathbf{e}_i^T\mathbf{e}_i - \mathbf{e}_i^T(\mathbf{U}^{(t)})(\mathbf{U}^{(t)})^T\mathbf{e}_i}\right) \tag{14}$$

$$\mathbf{U}_k^{(t+1)} = \underset{\mathbf{U}}{\operatorname{argmax}}\,\mathrm{Tr}\left(\mathbf{U}^T\mathbf{E}_k^R\mathbf{P}_k^{(t+1)}(\mathbf{E}_k^R)^T\mathbf{U}\right) \tag{15}$$

where t is the HQ iteration and $\mathbf{P}_k^{(t)}$ is the diagonal matrix version of $\mathbf{p}_k^{(t)}$. Particularly, Eq. (13) uses Silverman's rule [17] to estimate the kernel bandwidth adaptively where σ_E is the standard deviation of the sequence $||\mathbf{e}_i - \mathbf{U}^{(t)}(\mathbf{U}^{(t)})^T\mathbf{e}_i||^2$ and R is its interquartile range. Eq. (15) is solved via classic SVD solvers where the updated atom is the eigenvector corresponding to the largest eigenvalue. In short, CK–SVD is a weighted PCA implementation that downplays the influence of outliers during the dictionary update stage of K–SVD.

2.3 Correntropy–based Dictionary Learning

As an algorithm based on block coordinate descent, K–SVD relies on effective sparse coders and SVD solvers to work iteratively to find a local solution. Yet, if any of the two subroutines yields biased estimates (due to outliers or non–Gaussian environments), it will directly affect the subsequent stage and lead to overall biased sparse codes and dictionary. Hence, it would be advantageous to incorporate robustness into both stages by leveraging the properties of Correntropy.

We propose a combined fully Correntropy–based sparse modeling framework. CDL or Correntropy–based Dictionary Learning alternates between robust sparse coding (CMP) and Correntropy–based Dictionary Update until convergence. On the one hand CMP downplays the influence of outliers (under a linear regime) in the observation vectors $\mathbf{Y}$, while on the other hand Correntropy–based Dictionary Update routines mitigate the effect of outliers (under MSE) in the estimated dictionary $\mathbf{D}$. Thus, CDL is able to deal with both types of outliers in a principled robust manner without any extra hyperparameters.

For completeness, we also propose a variant of K–SVD that uses CMP and MSE–based dictionary update: CMPDL, and reintroduce C–KSVD [10], a combination between OMP and Correntropy–based Dictionary Update. In this way, it is possible to assess which K–SVD stage is more sensitive to outliers and non–Gaussian scenarios.

3 Results

The first set of results focuses on robust sparse modeling with access to ground truth. The dictionary, $\mathbf{D} \in \mathbb{R}^{20\times 50}$, is generated by sampling a zero–mean uniform distribution with support $[-1,1]$. Each column is normalized to a unit ℓ_2–norm. Sparse codes are generated from a uniform random variable with support $[0,1]$ ($T_0 = 3$).

Then, 1500 20–dimensional samples are produced via linear combinations between sparse codes and dictionary. These samples are then affected by noise. We compare the performance of K–SVD, CMPDL, CK–SVD, and CDL by computing the inner product between atoms from estimated and generating dictionaries. Each dictionary learning technique alternates 40 times between sparse coding and dictionary update. The expected sparsity support is set to $L = 3$.

The first type of noise is additive Gaussian. Its SNR is varied from 0 to 20 dB. Table 1 details the performance of each algorithm as the average of 50 independent runs for each SNR case (upper rows of each cell). The table also summarizes the same metrics under additive Laplacian noise (lower rows of each cell). It is evident that the robust variants outperform K–SVD, with CDL being superior for most cases. The experiments with low SNR emphasize the fact that Correntropy–based variants are able to properly handle errors with long tail distributions.

Table 1: Average inner product between estimated and ground truth atoms under additive Gaussian (upper rows of cells) and Laplacian (lower rows of cells) noise. Best results are marked bold.

SNR	**Algorithm**			
(dB)	K–SVD	CMPDL	CK–SVD	CDL
0	0.67	0.71	0.68	**0.72**
	0.67	0.71	0.69	**0.75**
5	0.87	0.91	0.91	**0.94**
	0.86	0.91	0.92	**0.95**
10	0.98	0.98	**0.99**	**0.99**
	0.98	0.97	**0.99**	**0.99**
15	0.98	**0.99**	**0.99**	**0.99**
	0.98	0.98	**0.99**	**0.99**

The third type of noise is non–homogeneous in the form of missing entries in the observation vectors. In particular, a percentage of the components from each observation vector is set to zero. This rate is varied from 0 to 50%. Fig. 1 summarizes the average of 50 independent runs for each sparse modeling technique. The Correntropy–based variants consistently outperform K–SVD while CDL is superior in aggressive noise environments. The degradation of CDL under no missing pixels is worth investigating as further work.

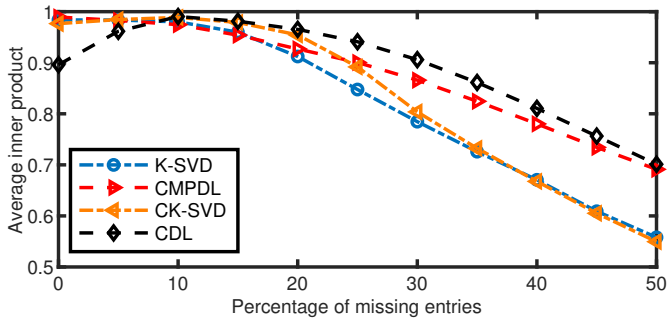


Fig. 1: Average inner product between estimated and ground truth atoms under missing entries type of noise.

The next set of results deals with image denoising. The approach proposed in [7] is exploited here. Essentially, the denoising mechanism invokes sparse modeling over local patches of the noisy image. Each patch is sparsely encoded with a constraint on the residue norm equal to 1.15σ, where σ is the standard deviation of homogeneous additive Gaussian noise. Then, local averaging over overlapping patches and global weighted averaging with the noisy image renders the estimated denoised example (Lagrange multiplier, λ, is set equal to 30 according to [7]).

For the current work, we choose $\sigma = 20$. Our framework is tested under a non–linear transformation in the form of impulsive noise, i.e. a percentage of affected pixels will be saturated to either 0 or 255 according to additive Gaussian noise with high power ($\sigma_{\text{imp}} = 100$ for this case). The rate of affected pixels is varied from 0 to 40%. In short, the original image goes first through an additive linear transformation with $\sigma = 20$ and then through a non–linear, inhomogeneous transform with $\sigma_{\text{imp}} = 100$. All possible overlapping vectorized 8×8 pixel patches constitute the observations of the sparse model. The initial dictionary $\mathbf{D} \in \mathbb{R}^{64\times 256}$ is chosen as the overcomplete Discrete Cosine Transform (DCT) basis. Lastly as suggested in [7], 10 alternating optimizations of the block coordinate descent routine are ran for each case.

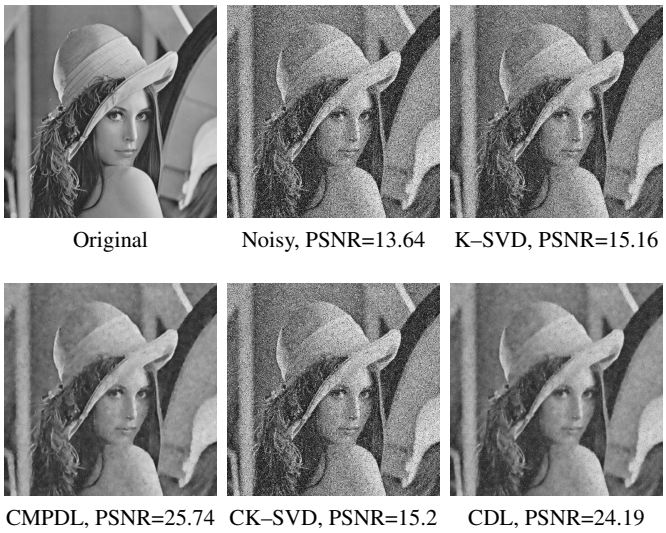
Table 2 details the average PSNR over 5 independent runs for each noise rate and five different well known gray–scale images of size 512×512. In general, CMPDL and CDL deliver the best performances while CK–SVD remains close to the K–SVD baseline. In particular, CMPDL and CDL are fairly consistent for a wide range of affected pixels. CMP seems to be the deciding factor here—it filters the inlier samples to the subsequent atom update stage, and, hence, reduces the estimation bias. On the other hand, OMP overrepresents the inputs and passes noisy examples to the K–SVD or CK–SVD dictionary update routines. This is confirmed in Table 3 where the average number of coefficients (per 8×8 block) in the sparse decompositions are compared. CMP–based variants clearly render a truly sparse representation, while other flavors overrepresent the input by encoding residual noise until the constraint on the residue norm is met. Therefore, Correntropy is more advantageous in the sparse coding subroutine than in the SVD solver. The details regarding the slight difference in PSNR between CMPDL and CDL are left as further work. Lastly, Fig. 2 illustrates the denoising results for a case of 40% rate of outlier pixels.

Table 2: Summary of denoising performance, PSNR (dB), under different impulsive noise (outliers) rate. Each cell reports four denoising techniques. Top left: K–SVD [7]. Top right: CMPDL. Bottom left: CK–SVD [10]. Bottom right: CDL. Best results are marked bold.

Outlier %	Barbara		Boats		House		Lena		Peppers		Average	
0	**30.80**	29.33	**30.33**	29.11	**33.17**	32.23	**32.38**	31.53	**30.77**	29.45	**31.49**	30.33
	29.84	28.73	29.85	28.39	32.43	31.50	31.85	30.85	30.25	28.81	30.84	29.66
10	20.95	**24.51**	20.76	**24.42**	20.98	**25.21**	20.99	**24.92**	20.86	**24.58**	20.91	**24.73**
	21.19	24.40	20.94	24.29	21.31	25.21	21.25	24.84	21.18	24.51	21.17	24.65
20	18.66	**23.76**	18.56	**24.06**	18.67	**24.73**	18.69	**24.44**	18.62	**23.98**	18.64	**24.19**
	18.55	23.57	18.37	23.78	18.49	24.62	18.48	24.25	18.71	23.81	18.52	24.01
30	16.66	**23.74**	16.54	**24.48**	16.61	**25.03**	16.65	**24.76**	16.66	**24.10**	16.62	**24.42**
	16.71	23.09	16.58	23.66	16.63	24.47	16.69	24.18	16.67	23.60	16.66	23.80
40	15.22	**24.28**	15.09	**24.88**	15.16	**25.99**	15.16	**25.76**	15.18	**24.58**	15.16	**25.09**
	15.24	22.16	15.12	23.30	15.21	24.30	15.19	24.20	15.22	22.88	15.20	23.37

Table 3: Grand average number of coefficients in sparse decompositions after block–based image denoising. Sparsest solutions are marked bold.

Outlier	**Algorithm**			
%	K–SVD	CMPDL	CK–SVD	CDL
0	0.85	**0.45**	1.17	0.46
10	3.72	**0.96**	3.96	**0.96**
20	7.98	**1.00**	8.11	**1.00**
30	11.12	**1.04**	13.47	1.05
40	13.07	**1.19**	17.99	1.32



Original Noisy, PSNR=13.64 K–SVD, PSNR=15.16

CMPDL, PSNR=25.74 CK–SVD, PSNR=15.2 CDL, PSNR=24.19

Fig. 2: Example of the denoising results for the image “Lena”. 40% of pixels are affected by impulsive noise.

4 Conclusion

We proposed a robust sparse modeling framework where Correntropy is exploited to reformulate the cost functions of both sparse coding and dictionary update stages of K–SVD. Experiments with synthetic data and image denoising confirm the robustness of the estimators and their potential in applications prone to outliers where sparsity is advantageous. In particular, Correntropy seemed to be more decisive when used in the sparse coding subroutine of K–SVD. Further work will involve in–depth analysis of the heuristics utilized to select the kernel widths and their connection to robust linear modeling [3]. In addition, the denoising mechanism proposed in [7] states empirical optimal hyperparameters for MSE–based cases. We believe different sets of hyperparameters might yield distinct stationary points for Correntropy–based optimizations that are worth investigating.

References

1. Aharon, M., Elad, M., Bruckstein, A., et al.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on signal processing **54**(11), 4311 (2006)
2. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Selected papers of hirotugu akaike, pp. 199–213. Springer (1998)
3. Andersen, R.: Modern methods for robust regression. 152. Sage (2008)
4. Barron, A., Rissanen, J., Yu, B.: The minimum description length principle in coding and modeling. IEEE Transactions on Information Theory **44**(6), 2743–2760 (1998)
5. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory **52**(2), 489–509 (2006)
6. Donoho, D.L., Johnstone, J.M.: Ideal spatial adaptation by wavelet shrinkage. biometrika **81**(3), 425–455 (1994)
7. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image processing **15**(12), 3736–3745 (2006)
8. He, R., Hu, B.G., Zheng, W.S., Kong, X.W.: Robust principal component analysis based on maximum correntropy criterion. IEEE Transactions on Image Processing **20**(6), 1485–1494 (2011)
9. Liu, W., Pokharel, P.P., Príncipe, J.C.: Correntropy: Properties and applications in non-gaussian signal processing. IEEE Transactions on Signal Processing **55**(11), 5286–5298 (2007)
10. Loza, C.A., Principe, J.C.: A robust maximum correntropy criterion for dictionary learning. In: Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on, pp. 1–6. IEEE (2016)
11. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: Computer Vision, 2009 IEEE 12th International Conference on, pp. 2272–2279. IEEE (2009)
12. Mairal, J., Elad, M., Sapiro, G.: Sparse representation for color image restoration. IEEE Transactions on image processing **17**(1), 53–69 (2008)
13. Mallat, S., Zhang, Z.: Matching pursuit with time-frequency dictionaries. Tech. rep., Courant Institute of Mathematical Sciences New York United States (1993)
14. Nikolova, M., Ng, M.K.: Analysis of half-quadratic minimization methods for signal and image recovery. SIAM Journal on Scientific computing **27**(3), 937–966 (2005)
15. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature **381**(6583), 607 (1996)
16. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena **60**(1-4), 259–268 (1992)
17. Silverman, B.W.: Density estimation for statistics and data analysis. Routledge (2018)
18. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) pp. 267–288 (1996)
19. Tropp, J.A., Gilbert, A.C.: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transactions on information theory **53**(12), 4655–4666 (2007)
20. Wang, Y., Tang, Y.Y., Li, L.: Correntropy matching pursuit with application to robust digit and face recognition. IEEE transactions on cybernetics **47**(6), 1354–1366 (2017)