

Multichannel Image Blind Deconvolution for Noisy Measurements

João Rabello Alvim, Kenji Nose-Filho, Renato Rocha Lopes

Abstract—This paper analyzes Multichannel Blind Deconvolution (MBD) techniques over noiseless and noisy data. We employ both REgularization by Denoising (RED) and the multichannel blind criterion as components of an approach to regularize the ill-posed inverse problem of image deblurring. Tests with different optimization techniques, such as fixed step, global search and the Barzilai-Browein step direction were performed in synthetic data and the results were compared with other techniques in the literature. Low Root Mean Squared Error (RMSE) kernels and image estimates were achieved in the deblurring of noisy images. This was achieved by combining fixed-step RED and the global search method in an alternating optimization algorithm.

Keywords—Processing, Multichannel Blind Deconvolution, Inverse Problems, Non-Linear Optimization, SIMO systems

I. INTRODUCTION

Blind deconvolution is a thriving research topic inside the scope of inverse problems [1], [2], [3]. In signal processing, the necessity of solving such problems arises when systems require the estimate of an unknown entity from its interactions with the environment and within the processing pipeline itself.

Linear interactions between signals and systems are described through the operation of convolution. In seismic analysis, as exposed in [1], the reflectivity function of the subsurface, which is related to its geological properties, is usually estimated through a deconvolution algorithm. In fields like astronomical and microscopical imaging [4], [2], often some deconvolution technique is applied over the acquired images in order to reconstruct a picture closer to reality. In satellite and cosmic imaging, in general, the time dependent configuration of particles in the atmosphere allow the use of MBD techniques for deblurring purposes when the pictures are taken with the correct sample rate [5].

Although recent works in MBD proposed new regularization techniques, such as [6] [7], there is a lack of models in literature that incorporate prior information about natural images. Recent work uses convolutional neural networks to encode such information through real data, as can be seen in [3]. Nevertheless a regularization model that mathematically induces solutions in the natural images domain can be a useful tool for deconvolution purposes. As the denoiser technology further advances, Romano et al. describe in [8] how to take advantage of such systems to solve inverse problems with the RED algorithm. In this context, the present paper combines modern optimization algorithms for MBD to the RED regularization.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, grant number 88887.640925/2021-00 and the National Council for Scientific and Technological Development (CNPq), grant number 310824/2021-4.

II. NOTATION

Lower case letters such as y_i , k_i and x represents images, that are mathematical equivalent to matrices or 2D continuous signals depending on the context. Since the proposed technique is used in a multichannel framework, the sub-indices i and j denotes the i -th or j -th samples of the multichannel array. At the same time, super-indices (like x^l) indicate the l -th term of a sequence.

The set of multiple image samples or blurring kernels are represented by lower case letters and a "vector hat" indication, to highlight the multidimensional array-like structure. The symbols \vec{k} and \vec{y} represent respectively the array denoted by all the blurring kernels and the array of all captured image samples and $y_1, \dots, y_m \in \vec{y}$.

Upper case letters such as Y_i , K_i , \vec{Y} and \vec{K} are the Fourier transforms (FT's) of their lower case counterparts. The FT of an array represents the transformation applied to each entry of the array. The symbol " A^* " represents the complex conjugate of the matrix A . The discrete convolution between two signals x and y is represented by the expression $x*y$, and the element-wise product between two matrices (X and Y) is expressed as $X \cdot Y$.

III. MATHEMATICAL MODEL

The multichannel model can be represented as the degradation of an image signal x of a set composed by $m \geq 2$ blurring kernels $\{k_i\}_{i=1}^m$ resulting in the set of m samples $\{y_i\}_{i=0}^m \mid y_i = x * k_i$ in a noiseless scenario.

Using known properties of the convolution it can be seen that for a given acquired image $y_i = x * k_i \mid k_i \in \vec{k}$:

$$k_j * y_i = k_j * (k_i * x) = k_i * (k_j * x) = k_i * y_j \quad (1)$$

In [9] it was shown that, based on (1), the estimation of the blurring kernels of each channel can be achieved by minimizing the cost function:

$$\mathcal{L}(\vec{y}, \vec{k}) = \sum_{i \neq j} \|k_i * y_j - k_j * y_i\|_2^2 \quad (2)$$

So the optimization task to recover the original signals becomes:

$$\begin{aligned} \min_{\vec{k}} \quad & \mathcal{L}(\vec{k}, \vec{y}) \\ \text{s.t.} \quad & \|k_i\|_2 = 1 \end{aligned} \quad (3)$$

Problem's (3) solution is the set of kernels that minimizes the cross-channel error in (2). The norm constraint $\|k_i\|_2 = 1$ assures that the found kernel is proportional to the actual

solution up to a scalar constant (scaling ambiguity), as stated in [9].

Most of the complexity of computing \mathcal{L} arises from the discrete convolution, operation equivalent to a matrix multiplication with complexity $c > O(n^2)$, as stated in [10]. One approach to reduce the overall complexity is to use the convolution theorem, which, by transporting the signals to the frequency domain, transforms the convolution operation in a point-wise product, as discussed in [11] for the z -transform. In this paper, we'll use a similar approach using the Fast Fourier Transform (FFT) of the images and blurring kernels.

A. Noiseless model

To solve (3) it is necessary to compute the gradient of the cost function in (2), which is done as follows:

$$\begin{aligned}\nabla_{k_i} \mathcal{L} &= \sum_{i \neq j} \frac{\partial}{\partial k_i} \|k_i * y_j - k_j * y_i\|_2^2 \\ &= \sum_{i \neq j} 2 \frac{\partial}{\partial k_i} (y_j * k_i)^T (k_i * y_j - k_j * y_i) \\ &= 2 \sum_{i \neq j} \mathcal{F}^{-1} \{Y_j^* \cdot (K_i \cdot Y_j - K_j \cdot Y_i)\}\end{aligned}\quad (4)$$

Using $\nabla_{k_i} \mathcal{L}$ defined in (4), problem (3) can be solved by some gradient descent algorithm.

B. Noisy data model and RED regularization

In order to further improve the deblurring capabilities of the proposed system, the cost function should incorporate the image reconstruction error. We also propose the RED regularization term to be included in the problem model. By introducing both regularization terms in (2), we obtained the following cost function with penalty terms μ and λ :

$$\begin{aligned}\mathcal{L}(x, \vec{k}, \vec{y}) &= \sum_{i \neq j} \|y_j * k_i - y_i * k_j\|_2^2 \\ &+ \frac{\mu}{2} \sum_{i=1}^m \|x * k_i - y_i\|_2^2 + \lambda x^T (x - f(x))\end{aligned}\quad (5)$$

The last term in (5) is the RED regularization, which favors solutions x that are orthogonal to the denoised residual $x - f(x)$, where $f(\cdot)$ is a denoiser. Romano et al. states in [8] that if f satisfies some conditions, then $\nabla_x x^T (x - f(x)) \approx x - f(x)$. By this approximation the gradient of the cost function in the x domain becomes:

$$\begin{aligned}\nabla_x \mathcal{L} &= \frac{\mu}{2} \sum_{i=1}^m \frac{\partial}{\partial x} \|x * k_i - y_i\|_2^2 + \lambda (x - f(x)) \\ &= \mu \sum_{i=1}^m \mathcal{F}^{-1} \{K_i^* \cdot (X \cdot K_i - Y_i)\} + \lambda (x - f(x))\end{aligned}\quad (6)$$

The steepest direction given for the cost function in (5) in the kernel domain is given by:

$$\begin{aligned}\nabla_{k_i} \mathcal{L} &= \mathcal{F}^{-1} \{Y_j^* \cdot (K_i \cdot Y_j - K_j \cdot Y_i)\} \\ &+ \mu \sum_{i=1}^m \mathcal{F}^{-1} \{X^* (X \cdot K_i - Y_i)\}\end{aligned}\quad (7)$$

IV. METHODOLOGY

In this section, two main aspects of the methodology were discussed into two subsections: the algorithms used to achieve the deconvolution and its implementation details.

A. Algorithms

1) *Fixed Step Size*: The fixed step size approach consists in setting a constant value to the gradient vector in the descent process. The iteration of two successive values of the sequence $\{x\}_{k=0}^n$ is given by the expression: $x^{i+1} = x^i - \rho \nabla \mathcal{L}$ with ρ being a constant given by: $0.5 / \|\nabla \mathcal{L}\|_2$. Although this is not a very usual minimization strategy, since the ℓ_2 -norm of the gradient vector modulates the distance traveled in each step; this heuristic (dividing the gradient vector by its norm) was developed during the experiments to avoid trivial solutions and numerical divergences to infinity.

2) *Barzilai-Browein (BB)*: The gradient descent direction proposed by Barzilai and Browein in [12] uses two distinct step sizes to approximate the Newton's method direction. Since the computation of the Hessian matrix and its inverse can be very costly, the algorithm consists in approximating the second derivative by the difference between the gradient computed at the current and last iteration of the descent. The steps σ' and σ'' (which can be used interchangeably) are calculated by the following equations:

$$\sigma' = \frac{\langle \text{vec}(\Delta x), \text{vec}(\Delta g) \rangle}{\langle \text{vec}(\Delta g), \text{vec}(\Delta g) \rangle}\quad (8)$$

$$\sigma'' = \frac{\langle \text{vec}(\Delta x), \text{vec}(\Delta x) \rangle}{\langle \text{vec}(\Delta x), \text{vec}(\Delta g) \rangle}\quad (9)$$

Where the operator " $\langle \cdot, \cdot \rangle$ " represents the scalar product between the vectorized matrices Δx and Δg which are respectively given by:

$$\Delta x = x^l - x^{l-1}\quad (10)$$

$$\Delta g = \nabla f(x^l) - \nabla f(x^{l-1})\quad (11)$$

3) *Global Search*: This line search method consists in partitioning the search direction (steepest descent) into $\delta = 20$ intervals between 0 and $\|\nabla \mathcal{L}\|_2$. These values were defined by fine-tuning the parameters during the execution step, where a complexity \times performance trade-off was the main tradeoff, since the cost function is evaluated δ times.

4) *Fixed point RED for noisy data*: To derive a fixed-point strategy to minimize the cost function in (5), we follow [8] and first assume that the denoised images in two subsequent iterations are approximately the same. Thus, the RED term in (5) can be replaced by $(x^l)^T (x^l - f(x^{l-1}))$. Using the gradient in (6), we calculate the fixed-point solution to $\nabla_x \mathcal{L} = 0$ as:

$$\begin{aligned}\nabla_x \mathcal{L}(x^l, x^{l-1}) &= \frac{\mu}{2} \sum_{i=1}^m \mathcal{F}^{-1} \{K_i^* \cdot (X^l \cdot K_i - Y_i)\} \\ &+ \lambda (x^l - f(x^{l-1})) = 0\end{aligned}\quad (12)$$

By transposing the convolution and differential operators to the frequency domain and isolating x^l it is possible to find



Fig. 1: Picture used for benchmark testing.

an iterative formula that converges to the fixed point where $\nabla_x \mathcal{L}(x^*) = 0 \implies \{x^l\}_{l=0}^{\infty} \rightarrow x^*$, see [8]. Once rearranged, the described iterative formula becomes:

$$x^l = \mathcal{F}^{-1} \left\{ \frac{\lambda \mathcal{F}\{f(x^{l-1})\} + \mu \sum_{i=1}^m K_i^* Y_i}{\lambda + \mu \sum_{i=1}^m K_i^* K_i} \right\} \quad (13)$$

V. RESULTS

The results presented in this paper are gathered into two subsections: Noiseless and noisy synthetic data. Both sections evaluate and discuss the interesting emerging solutions for each case and isolates the different aspects of each technique as well as simulation details about the experiments.

A. Simulation

The synthetic data was generated by using the starfish picture in Figure 1 convolved with different pairs of blurring kernels. Additive White Gaussian Noise (AWGN) was synthesized to fit a 30 dB SNR level.

The simulations were divided into two main categories: deblurring with noiseless synthetic data and deblurring with noisy synthetic data. For the noiseless tests, only the multichannel criterion - Equation (2) - was used to model the system. As shown in [9], the convergence rates of different line search methods were compared for 10^5 gradient descent iterations. This rate can be evaluated by the inspection of Root Mean Squared Error (RMSE) between real and estimated image as well as real and estimated kernels.

To deblur noisy inputs, an alternating optimization approach was taken: within 100 external "macro-iterations", two minimization processes are successively performed. A fixed point RED minimization in the image reconstruction domain, with 500 "micro-iterations", followed by a classical fixed step steepest descent algorithm in the kernel reconstruction domain with the same number of micro-iterations. The regularization constants in Equation (5) were set to $\lambda = 0.5$ and $\mu = 1.0$

B. Noiseless Data

Figure 2 shows the RMSE between the original and estimated kernels at each iteration. In the noiseless case, it is interesting to see an emerging pattern: the least computational complex search algorithms (fixed step) achieves competitive performances when compared to the most computationally expensive algorithm (the global search approach). However the latter method produces a more stable behavior in the descent.

Figure 3 shows the estimates of the kernel and image reconstruction for the different line search approaches for 100,000 gradient descent iterations. One can see that both fixed

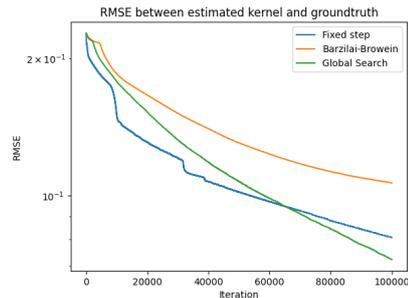


Fig. 2: RMSE for each method for 100'000 iterations.

step and Global search techniques achieve nice reconstructed image while the Barzilai-Browein method achieved a slightly larger RMSE than the other algorithms by the same number of iterations.

The image reconstructions shown in Figure 3, for the noiseless tests, were obtained by the division of the Fourier Transforms of the estimated kernels by their respective blurred picture. Mathematically this procedure can be written as $\hat{x}_i = \mathcal{F}^{-1}\{Y_i/\hat{K}_i\}$. For the noisy simulations, \hat{x} was obtained by directly optimizing the cost function.

C. Noisy Data

Figures 5 and 4 illustrate the progression of the gradient descent sequences for both the cost function and RMSE for two different sets of 4×4 blurring kernels: a simplified set, characterized for the impulse and anti-impulse, and a randomly generated set. Both cost functions decay in a very steep angle at the beginning of the process for all kernel scenarios. By evaluating the graphs in Figures 4, one can notice that kernel domain RMSE is monotonically decreasing, behavior which reveals stability in the algorithm. However the RMSE of the image reconstruction shows an average descending behavior, but no monotonicity is achieved. Figure 5 reveals that the three interest costs are monotonically decreasing for the simplified kernels.

By evaluating the outcomes shown in Figures 6a and 6b, one can see that, for the simplified kernel case, both the kernel and image reconstructions are close enough to their references. For the random kernels another outcome is observed: the image reconstruction converges to a close representation of the reference image, moreover the kernel estimates seem to converge to local minima of the cost function, compromising convergence of the kernel reconstructions.

VI. CONCLUSIONS

When observing the proposed algorithm for the noiseless synthetic images, it is interesting to notice how gradient descent performs well for the deblurring purpose. In some cases the different line search methods can improve radically the behavior of the descent. Usually a trade-off between computational cost and numerical stability have to be made, but, as seen in the results, there is no great differences between approaches. In this sense we experienced better deconvolution results with the least complex method.

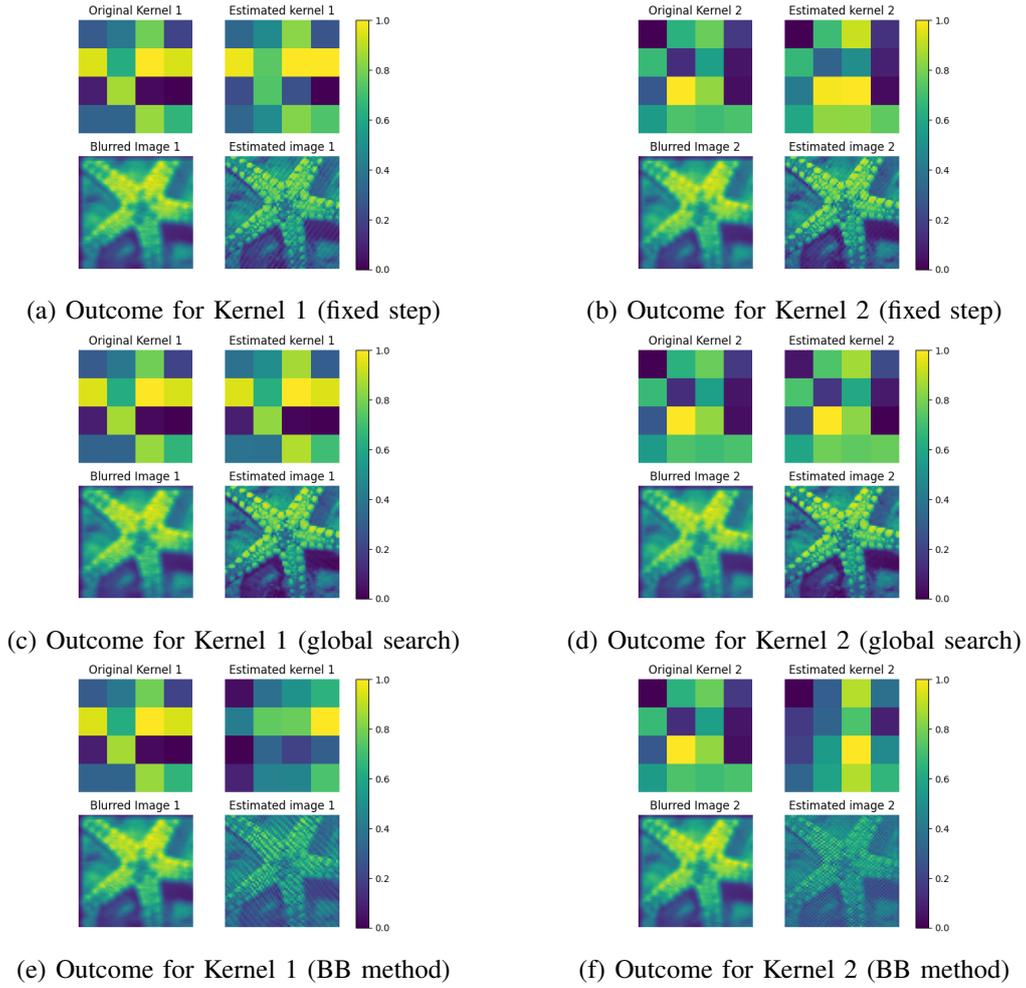


Fig. 3: Outcome for the different methods for 100,000 iterations and two blurring kernels.

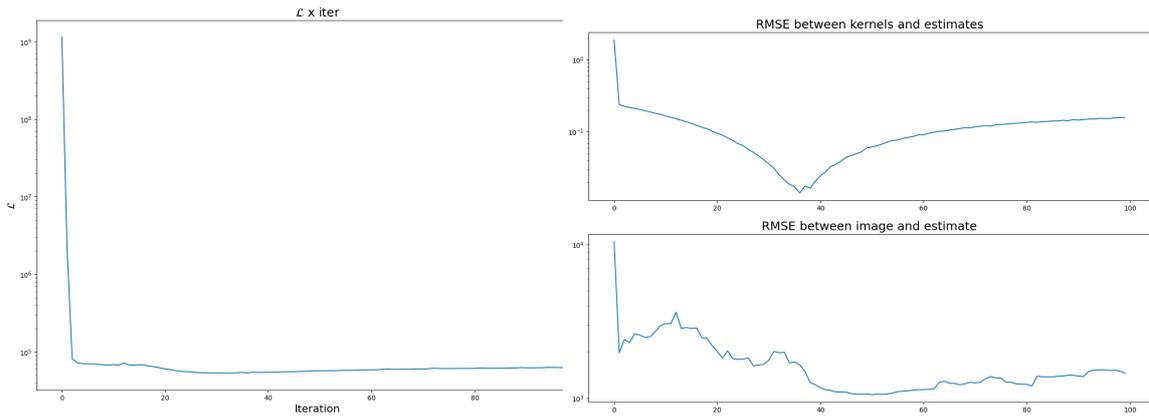


Fig. 4: Cost value (left) and RMSE of kernel and image estimates for the simplified set of kernels (right) per iteration.

Nevertheless, noiseless synthetic data represents an idealized scenario. Most real images carry some level of additive noise and deconvolution models must deal with such adversities. The proposed algorithm considers noise in its formulation and uses denoiser technologies to assist in the regularization of the problem. As the use of multiple sensors becomes more and more accessible and the technologies behind denoisers are further developed, efforts in unifying these and other

technologies should be done by the scientific community. This paper tries to contribute in this sense by showing how different techniques (such as MBD and RED) can complement one another.

In order to further advance the results shown in this paper, work is being developed by evaluating other RED implementations besides the fixed point approach, such as steepest descent and the Alternating Direction Method of Multipliers (ADMM)

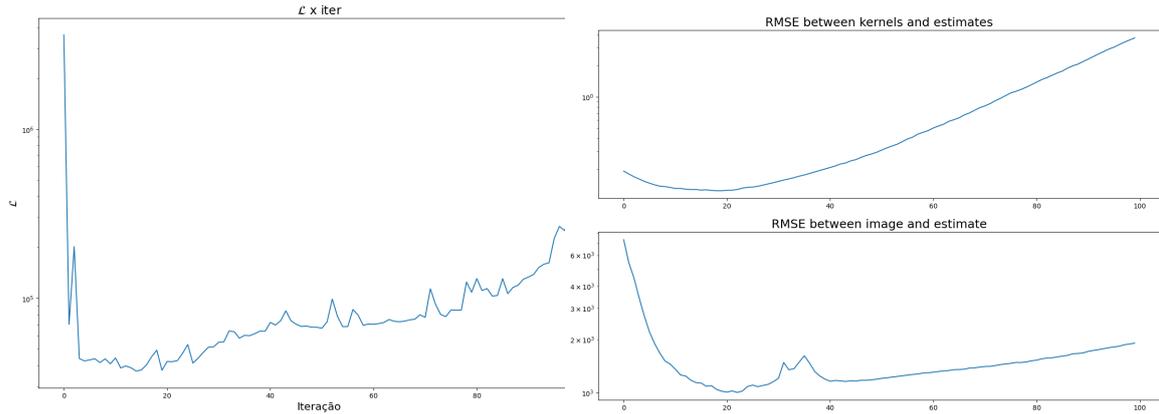
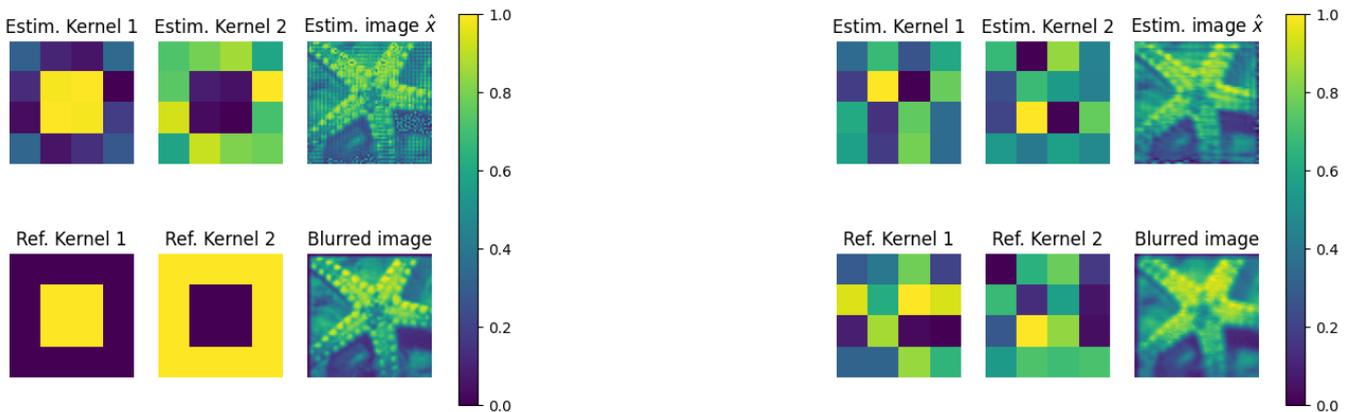


Fig. 5: Cost value (left) and RMSE of kernel and image estimates for the randomized set of kernels (right) per iteration.



(a) Obtained reconstructions from the image and the simplified kernels.

(b) Obtained reconstructions from the image and the random kernels.

Fig. 6: Outcomes from noisy deconvolution.

methods. Testing the deblurring capabilities of the proposed approach in real data, evaluating different initialization techniques - taking into consideration application specific kernel and priors - are also future perspectives for the presented work.

REFERENCES

- [1] K. Nose-Filho, A. K. Takahata, R. Lopes, and J. M. T. Romano, "Improving Sparse Multichannel Blind Deconvolution with Correlated Seismic Data: Foundations and Further Results," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 41–50, Mar. 2018. [Online]. Available: <http://ieeexplore.ieee.org/document/8310697/>
- [2] A. Ahmed, "Blind Deconvolution Using Modulated Inputs," *IEEE Trans. Signal Process.*, vol. 68, pp. 374–387, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/8931024/>
- [3] G. Chen, Z. Gao, Q. Wang, and Q. Luo, "Blind de-convolution of images degraded by atmospheric turbulence," *Applied Soft Computing*, vol. 89, p. 106131, Apr. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1568494620300715>
- [4] A. Yang, X. Jiang, and D. Day-Uei Li, "Multi-frame blind deconvolution of atmospheric turbulence degraded images with mixed noise models," *Electron. Lett.*, vol. 54, no. 4, pp. 206–208, Feb. 2018. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1049/el.2017.4277>
- [5] D. J. Lee, "Supersampling multiframe blind deconvolution resolution enhancement of adaptive optics compensated imagery of low earth orbit satellites," *Opt. Eng.*, vol. 41, no. 9, p. 2238, Sep. 2002. [Online]. Available: <http://opticalengineering.spiedigitallibrary.org/article.aspx?doi=10.1117/1.1497615>
- [6] X.-X. Wei, L. Zhang, and H. Huang, "High-quality blind defocus deblurring of multispectral images with optics and gradient prior," *Opt. Express*, vol. 28, no. 7, p. 10683, Mar. 2020. [Online]. Available: <https://opg.optica.org/abstract.cfm?URI=oe-28-7-10683>
- [7] L. Shi and Y. Chi, "Manifold Gradient Descent Solves Multi-Channel Sparse Blind Deconvolution Provably and Efficiently," *arXiv:1911.11167 [cs, eess, math, stat]*, Apr. 2021, arXiv: 1911.11167. [Online]. Available: <http://arxiv.org/abs/1911.11167>
- [8] Y. Romano, M. Elad, and P. Milanfar, "The Little Engine That Could: Regularization by Denoising (RED)," *SIAM J. Imaging Sci.*, vol. 10, no. 4, pp. 1804–1844, Jan. 2017. [Online]. Available: <https://epubs.siam.org/doi/10.1137/16M1102884>
- [9] Lang Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998. [Online]. Available: <http://ieeexplore.ieee.org/document/720247/>
- [10] V. V. Williams, Y. Xu, Z. Xu, and R. Zhou, "New bounds for matrix multiplication: from alpha to omega," 2023.
- [11] K. Nose-Filho, A. K. Takahata, R. Lopes, and J. M. Romano, "A fast algorithm for sparse multichannel blind deconvolution," *Geophysics*, vol. 81, no. 1, pp. V7–V16, 2016.
- [12] J. Barzilai and J. M. Borwein, "Two-Point Step Size Gradient Methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 1988. [Online]. Available: <https://academic.oup.com/imajna/article-lookup/doi/10.1093/imanum/8.1.141>