A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors

Alexis Benichoux, Emmanuel Vincent, Rémi Gribonval

A FUNDAMENTAL PITFALL IN BLIND DECONVOLUTION WITH SPARSE AND SHIFT-INVARIANT PRIORS

Alexis Benichoux[1]*, Emmanuel Vincent*[2]*, Rémi Gribonval*[3]

[1]Université Rennes 1, IRISA - UMR6074, Campus de Beaulieu, 35042 Rennes, France
[2]INRIA Nancy - Grand Est, 615 rue du Jardin Botanique, 54600 Villers-ls-Nancy, France
[3]INRIA, Centre de Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes, France

ABSTRACT

We consider the problem of blind sparse deconvolution, which is common in both image and signal processing. To counter-balance the ill-posedness of the problem, many approaches are based on the minimization of a cost function. A well-known issue is a tendency to converge to an undesirable trivial solution. Besides domain specific explanations (such as the nature of the spectrum of the blurring filter in image processing) a widespread intuition behind this phenomenon is related to scaling issues and the nonconvexity of the optimized cost function. We prove that a fundamental issue lies in fact in the intrinsic properties of the cost function itself: for a large family of shift-invariant cost functions promoting the sparsity of either the filter or the source, the only global minima are trivial. We complete the analysis with an empirical method to verify the existence of more useful local minima.

***Index Terms*—** blind deconvolution, sparsity, MAP failure, deblurring, dereverberation

1. INTRODUCTION

The goal of blind deconvolution is to recover an unknown source signal $s \in \ell^2(\mathbb{Z}^d)$ from a filtered observation $x = a * s \in \ell^2(\mathbb{Z}^d)$ when the filter $a \in \ell^2(\mathbb{Z}^d)$ is unknown.

The ill-posed nature of the blind problem implies the introduction of prior knowledge. In particular, for each solution (a, s), and $\lambda \in \mathbb{R}^*$, $(\lambda a, \frac{1}{\lambda} s)$ is also a solution : this is known as the scaling ambiguity. In many physical problems, energy conservation assumptions avoid this ambiguity. For instance, early approaches based on Minimum Entropy Deconvolution [1] used scale-invariant cost functions under the assumption of statistical whiteness of the source.

In many practical image processing or audio processing scenarios, the statistical whiteness of s is not a reasonable assumption, and other types of prior knowledge are required, as well as ways to exploit them. Among the range of existing approaches [2–5], many approaches aim to minimize a cost function involving a quadratic data fidelity term and additional priors derived from the ℓ_p norm over the source signal and/or the filter. These approaches are often referred to as maximum a posteriori (MAP) in connection with Bayesian modeling and estimation. In the image processing literature, several priors on the source s are widely used based on image statistics or gradient domain sparsity [6–9]. In audio signal processing, sparse priors have been considered over the time-domain or the short time Fourier transform (STFT) representation of the source s [10–12]. A sparse prior on the filter a was also introduced in [13].

The form of these cost functions is reminiscent of those arising in matrix factorization problems such as sparse principal component analysis (PCA) [14], sparse non-negative matrix factorization (NMF) [15], dictionary learning [16], or independent component analysis [17]. In such matrix factorization problems, empirical as well as theoretical results have shown the validity of the approaches based on the minimization of these cost functions. In contrast, for blind deconvolution, many works show both theoretically and practically that "MAP" approaches often fail: they output the blurred observation x and a trivial Dirac filter [3].

A domain specific explanation of this failure phenomenon in image processing [3] blames the nature of the spectrum of the blurring filter, and was used to guide the design of alternative approaches such as the marginalization of the distribution over the filter [6, 18], the addition of an edge detection step [19], time varying priors [7], or re-weighted priors [20]. Another widespread intuition behind this phenomenon is related to scaling issues and the nonconvexity of the optimized cost function. Non-convexity was heuristically first dealt with using alternate optimization of a and s [21]. Recent algorithms which have been proposed for the simultaneous estimation of a and s [22] using proximal methods are only known to converge to a stationary point of the cost function.

In this paper, we provide two novel explanations of this failure phenomenon. First, we show that a large family of cost functions naturally arising in the context of blind deconvolution are in fact fundamentally flawed. The cost function itself is to blame, not the algorithm to minimize it: under mild conditions, all its global minima are trivial. Second, we also provide an empirical local study of the cost function arising from typical sparsity inducing audio priors. Inspired by the char-

acterization of ℓ^1 minima used in dictionary learning [16], we observe that the desired solution is a local minimum of the cost function only when both the filter and the sources are sufficiently sparse. Besides providing a new interpretation to a number of experimental observations, these results can help the design of improved cost functions by providing some guarantees on their minima independently of the algorithm chosen to minimize them.

This paper is organized as follows. The cost functions considered in the paper are described in Section 2. We display in Section 3 our main result on the global minima. Section 4 is dedicated to the local study in an audio processing example. We conclude in Section 5. The proofs of the results are given in the appendix.

2. REGULARIZATION WITH PRIORS

The observation $x \in \ell^2(\mathbb{Z}^d)$ is modeled as a convolutive mixture of the source $s \in \ell^2(\mathbb{Z}^d)$ with the filter $a \in \ell^2(\mathbb{Z}^d)$ plus some noise n, that is for all $t \in \mathbb{Z}^d$:

$$x(t) = (a*s)(t) + n(t) := \sum_{\tau \in \mathbb{Z}^d} a(\tau) s(t-\tau) + n(t). \quad (1)$$

To circumvent its natural ill-posedness, a widespread approach is to use priors on a and s (making the problem rather myopic than blind), which typically leads to regularized optimization problems of the type

$$\min_{a,s} \lambda \|x - a * s\|_2^2 + p(a,s) \quad (2)$$

where the penalty function $p(a,s)$ captures the prior.

The design and exploitation of signal priors is a wide research field and it has proven to be successful for underdetermined inverse problems in general. In particular, it is well known that sparsity-inducing priors such as the ℓ^p norm $\|s\|_p$ and $\|a\|_p$ with $p < 2$ can provide computationally efficient solutions that are accurate provided that s and a are indeed sparse or at least "compressible".

Because of the intrinsic scaling ambiguity of the blind deconvolution problem, some naive priors $p(a,s)$ should be avoided. In particular, it was shown that it is a bad idea to only enforce a source prior while using a uniform prior on the filter (which means no regularization on a) [3]. Denoting $\|\cdot\|$ a regularization norm penalty on s, this would lead to the optimization problem

$$\min_{a,s} \lambda \|x - a * s\|_2^2 + C\|s\|. \quad (3)$$

Such function has been pointed out to be dramatically sensitive to the scaling ambiguity.

Lemma 1 *[3, Claim 1] Let $a_0, s_0 \in \ell^2(\mathbb{Z}^d)$. The global minima of*

$$\mathcal{L} : (a,s) \mapsto \lambda \|a_0 * s_0 - a * s\|_2^2 + C\|s\|. \quad (4)$$

are never reached. There exists a^k, s^k such that

$$\lim_{k\to\infty} s^k = 0, \quad \textit{and} \lim_{k\to\infty} \mathcal{L}(a^k, s^k) = 0.$$

As a consequence (3) has no solution. Due to this remark, we only consider approaches that depends upon a prior on a.

From now on we consider a regularization $\|\cdot\|$ on s which is a translation invariant seminorm.

Definition 1 *A translation invariant seminorm on $\ell^2(\mathbb{Z}^d)$ is a function $\|\cdot\| : \ell^2(\mathbb{Z}^d) \to \mathbb{R}$ which satisfies $\forall u, v \in E$*

(i). $\|u+v\| \le \|u\| + \|v\|$

(ii). $\forall \lambda \in \mathbb{R}, \quad \|\lambda u\| = |\lambda| \|u\|$

(iii). $\forall k \in \mathbb{Z}, \quad \|u(\cdot - k)\| = \|u(\cdot)\|$

The only difference with a norm is that there can be nonzero vectors u such that $\|u\| = 0$.

Such penalty appear in many practical scenarios. Typical image applications [6, 18, 20] introduce the gradient sparsity inducing seminorm $\|s\| = \|\nabla s\|_p$ with $p \in [0.5, 0.8]$. Typical audio applications [11, 12] use an STFT matrix Φ and regularize in the time-frequency plane with the sparsity inducing norm $\|s\| = \|\Phi s\|_p$.

In case of a sparse prior on the filter, the deconvolution problem is often stated as

$$\min_{a,s} \lambda \|x - a * s\|_2^2 + \|a\|_1 + C\|s\|. \quad \text{(P1)}$$

Alternatively one can add a scaling constraint [11, 20] on a, resulting in a different problem

$$\min_{a,s} \lambda \|x - a * s\|_2^2 + C\|s\| \quad \text{s.t. } \|a\|_1 = 1. \quad \text{(P2)}$$

Note that in image processing, the estimation of the gradient instead of the image itself is often subject to a regularization framework. Our study apply to the gradient domain regularized estimation problems [6–9], which are a variants of (P1).

$$\min_{a,s} \lambda \|\nabla x - a * \nabla s\|_2^2 + \|a\|_1 + C\|\nabla s\|. \quad (5)$$

The formulations (P1)-(P2) are quite similar to common matrix factorization approaches arising in dictionary learning [16], sparse PCA [14], non-negative matrix factorization [15], etc, where the goal is to factor a matrix $\mathbf{X}$ as $\mathbf{X} = \mathbf{AS}$ while promoting certain properties of the factors $\mathbf{A}$ and $\mathbf{S}$. However, in contrast to matrix factorization approaches which often exhibit good practical performance, we will show that the cost functions appearing in (P1)-(P2) have fundamentally problematic properties. Although they are not equivalent, both problems (P1) and (P2) fail to characterize a non trivial solution, for any value of the parameters C or λ.

3. PITFALLS OF GLOBAL MINIMA

3.1. Main result

Given a mixture x, we show here that the global minima of (P2) and (P1) are trivial reconstructions, in the sense that the estimated filter is equal to a Dirac pulse. Let δ_0 be the Dirac pulse such that $\forall y \in \ell^2(\mathbb{Z}^d)$, $\delta_0 * y = y$.

Proposition 1 *Let $\|\cdot\|$ be a translation invariant seminorm. For all $a, s \in \ell^2(\mathbb{Z}^d)$, $0 < p \leq 1$, and $C > 0$, there exist $\mu_-, \mu^+ \in \mathbb{R}_+^*$ such that $\forall \mu \in [\mu_-, \mu^+]$*

$$\|\mu\delta_0\|_p + C\|\frac{1}{\mu}a * s\| \leq \|a\|_p + C\|s\|. \tag{6}$$

Remarks :

- We can extend the proposition to an even more general case, we may consider a family of linear transformations $(\mathcal{T}_t)_{t\in E}$ such that

$$\forall t \in \mathbb{Z} \quad x(t) = \sum_{\tau\in E} a(\tau)\mathcal{T}_t(s)(\tau), \tag{7}$$

  and a norm $\|\cdot\|$ invariant under these transformations. For example, the case of the circular convolution $x = a \otimes s$ of finite length signals in $\mathbb{R}^T$ corresponds to

$$\forall t \in E \quad \mathcal{T}_t(s)(\tau) = s\,(t - \tau \mod T) \tag{8}$$

  with $E = \{1, \ldots, T\}$.

- If $\|\cdot\|$ is a semi-quasinorm, i.e. satisfies instead of (i)

$$\|u + v\|^q \leq \|u\|^q + \|v\|^q$$

  for $q \geq 0$, the same result can be obtained under the condition $p \leq q$. This allows to treat the case $\|\cdot\| = \|\cdot\|_q$ with $0 < p \leq q \leq 1$.

- There is no uniqueness result, but if $p < 1$ or if $\|\cdot\|$ is strictly convex, equality in (6) implies $a = \delta_0$ up to a pure delay (the proof is provided in the Appendix).

We now derive a direct corollary suitable for the noisy case, without exact reconstruction of x, which corresponds to the practical situations described by (P1).

Corollary 1 *Let $x \in \ell^2(\mathbb{Z}^d)$, $0 < p \leq 1$, $C > 0$, $\lambda > 0$. There exists $\mu \geq 0, \hat{a}, \hat{s} \in \ell^2(\mathbb{Z}^d)$ such that $(\mu\delta_0, \frac{1}{\mu}\hat{a} * \hat{s})$ is a global minimum of*

$$\mathcal{L} : (a, s) \mapsto \lambda\|x - a * s\|_2^2 + \|a\|_p + C\|s\|. \tag{9}$$

Finally, we show that problem (P2) has a trivial global minimum.

Corollary 2 *Let $x \in \ell^2(\mathbb{Z}^d)$, $C > 0$, $\lambda > 0$. There exists $\hat{a}$, $\hat{s} \in \ell^2(\mathbb{Z}^d)$ such that $(\delta_0, \hat{a} * \hat{s})$ is a global minimum of*

$$\mathcal{L} : (a, s) \mapsto \lambda\|x - a * s\|_2^2 + C\|s\| \quad \text{s.t. } \|a\|_1 = 1. \tag{10}$$

In the case when $p < 1$ or $\|\cdot\|$ is strictly convex, all global minima of (9) and (10) are trivial.

4. LOCAL MINIMA

Globally solving of (P1) without knowing that the global minimum is trivial is a priori computationally challenging, since the optimized cost function is non convex. Optimization problems of a similar nature appear in the context of matrix factorization, and alternate estimation algorithms have been designed to address them. Such algorithms are never guaranteed to converge to the global minimum but at best to a stationary point of (P1). In the case of blind deconvolution, since the global minimum is trivial, convergence to a local minimum can in fact be a blessing: provided that the seeked solution (a, s) is indeed close to a local minimum, one can envision to exploit side information to initialize the algorithm in a good basin of attraction and converge to a useful solution. We describe now on a particular case how to experimentally check if the original signal is a local minimum.

4.1. Local analysis of (P1) in the ℓ^1 case

There is no local minimum in general, unless the constant C is wisely chosen. We can easily derive the following result from Proposition 1.

Corollary 3 *If $(\hat{a}, \hat{s})$ is a local minimum of*

$$(a, s) \mapsto \lambda\|x - a * s\|_2^2 + \|a\|_1 + C\|s\|, \tag{11}$$

then $C = \frac{\|\hat{a}\|_1}{\|\hat{s}\|}$.

Therefore we assume in the following that $C = \frac{\|\hat{a}\|_1}{\|\hat{s}\|}$.

Then, in the particular case of an ℓ^1 penalty $\|\cdot\| = \|\cdot\|_1$ on the source s, we derive a characterization of local minima. The computation is detailed in [16] in a general setting not specific to deconvolution. We do not reproduce here the methodology. It is nevertheless a verification that can be reproduced in different applications before the design of an algorithm. In a nutshell, there exists two matrices $\mathbf{A}, \mathbf{B}$ and a vector $\mathbf{c}$ that can be computed from $\hat{a}$ and $\hat{s}$, which lead to the necessary condition

$$\sup_{h\in\ker\mathbf{A}} \frac{|<\mathbf{c}, h>|}{\|\mathbf{B}h\|_1} \leq 1, \tag{12}$$

where a strict inequality is a sufficient condition. The quantity on the left-hand side of (12) can be computed using convex optimization.

4.2. Experimental analysis

We wish to test condition (12) in a typical audio situation. The cost function uses the $(m_f \times n_f) = (32 \times 16)$ STFT matrix Φ and the ℓ^1 norm $\|s\| = \|\Phi s\|_1$ and an ℓ^1 norm on the filters.

For $T = 256$, we generate a pair of uniform random signals $a \in \mathbb{R}^T, s \in \mathbb{R}^T$ for each pair of sparsity ratios $\rho_a = \frac{\|a\|_0}{T}, \rho_s = \frac{\|\Phi\|_0}{m_f n_f}$. Choosing $C = 1$ we rescale them

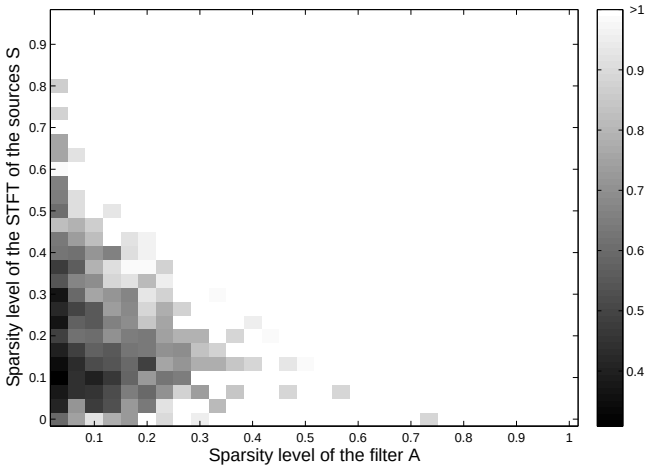

Fig. 1. Estimation of (12) : white areas are not local minima

to satisfy $\|\hat{a}\|_1/\|\hat{s}\| = 1$. We obtain an estimate of condition (12) with a convex optimization toolbox and provide Fig. 1 the resulting array. This indicates that the inequality is violated for non sparse filter and sources (high values of ρ_a, ρ_s), whereas for $\rho_a + \rho_s \leq 0.4$, the original solution is often a local minimum of (P1), even though it cannot be its global minimum.

5. CONCLUSION

We explored some of the theoretical limitations to the blind sparse deconvolution problem, for several typical approaches. The consequences of this pitfall are omnipresent in both image and audio processing frameworks. Our study gives a new interpretation to many well-known experimental failures and a justification to the choice of scaled constrained models in the past. An inspired example is the ℓ^1/ℓ^2 scaled sparsity regularizer [20] in the gradient domain

$$\min_{a,s} \lambda\|\nabla x - a * \nabla s\|_2^2 + C\frac{\|\nabla s\|_1}{\|\nabla s\|_2} \text{ s.t. } \|a\|_1 = 1, \quad \text{(P3)}$$

which to our knowledge does not admit any trivial reconstruction as a solution. This regularizer may however not apply in certain contexts, especially in audio, and our results can help the design of improved cost functions in these contexts by providing some guarantees on their minima independently of the algorithm chosen to minimize them. Besides, the local study proves that such approaches are still relevant under sparsity hypotheses. Further work is needed to extend our results to the multichannel multisource case, in order to address blind source separation problems.

Appendix

Proof of Lemma 1

Simply observe that $\lim_{n\to\infty} \mathcal{L}(na, \frac{1}{n}s) = 0$. □

Proof of Proposition 1

First, we minimize $g : \mu \in \mathbb{R}_+^* \mapsto \|\mu\delta_0\|_p + C\|\frac{1}{\mu}x\|$ and obtain with $\hat{\mu} = \sqrt{\frac{C}{\|\delta_0\|_p}\|x\|}$ an optimum $g(\hat{\mu}) = 2\sqrt{C\|\delta_0\|_p\|x\|} = 2\sqrt{C\|x\|}$.

On the other hand, as a consequence of the invariance of $\|\cdot\|$ we obtain, for $0 < p \leq 1$,

$$\begin{aligned} \|a * s\|^p &\leq \left(\sum_{\tau\in\mathbb{Z}^d} |a(\tau)| \cdot \|s(\cdot - \tau)\|\right)^p && (13)\\ &= (\|a\|_1\|s\|)^p && (14)\\ &\leq \|a\|_p^p\,\|s\|^p && (15)\\ \|x\| &\leq \|a\|_p\|s\| && (16) \end{aligned}$$

When $p < 1$ the inequality in (15) is strict unless $a = \delta_0$ up to a pure delay. The strict convexity of $\|\cdot\|$ also restricts the equality cases in (13) if a is not a Dirac. The last inequality gives an upper bound to the minimum of g :

$$\begin{aligned} g(\hat{\mu}) &\leq 2\sqrt{C\|a\|_p\,\|s\|} && (17)\\ &\leq \|a\|_p + C\|s\|. && (18) \end{aligned}$$

The last line uses the inequality $\forall u, v \in \mathbb{R}, 2uv \leq u^2 + v^2$.

In addition, a wide range of μ satisfies the conclusion of Proposition 1, namely $\mu \in [\frac{\|a\|_p + C\|s\| - \sqrt{\Delta}}{2}, \frac{\|a\|_p + C\|s\| + \sqrt{\Delta}}{2}]$, where $\Delta = \|a\|_p + C\|s\|^2 - 4\|x\|$. Surprisingly, the trivial mixture without scaling factor (δ_0, x) is lower than the original for large values of C. Formally, (6) is satisfied for $\mu = 1$ if $C \geq \frac{2|\|x\|-1|-\|a\|_p}{\|s\|}$. □

Proof of Corollary 1

First, $\mathcal{L}$ is coercive so $\operatorname{argmin}\mathcal{L} \neq \emptyset$. Let $\hat{a}, \hat{s}$ be a minimum of $\mathcal{L}$. Using Proposition 1 there exists μ such that $\mathcal{L}(\mu\delta_0, \frac{1}{\mu}\hat{a} * \hat{s}) \leq \mathcal{L}(\hat{a}, \hat{s})$, and $(\mu\delta_0, \frac{1}{\mu}\hat{a} * \hat{s}) \in \operatorname{argmin}\mathcal{L}$. □

Proof of Corollary 2

Suppose $(\hat{a}, \hat{s})$ is a solution of (10), and simply recall (16), $C\|\hat{a} * \hat{s}\| \leq C\|\hat{a}\|_1\|s\| = C\|\hat{s}\|$. Then for (a, s) such that $\|a\|_1 = 1$,

$$\|x - \delta * (\hat{a} * \hat{s})\|_2^2 + C\|\hat{a} * \hat{s}\| \leq \|x - a * s\|_2^2 + C\|s\|$$

and $(\delta_0, \hat{a} * \hat{s})$ is also a solution of (10). □

This work was supported in part by the European Research Council, PLEASE project ERC-StG-2011-277906.

6. REFERENCES

[1] D. L. Donoho, "On minimum entropy deconvolution," *Applied Time Series AnalysisII, Academic Press*, pp. 565–609, 1981.

[2] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, 1996.

[3] A. Levin, Y. Weiss, F. Durand, and W. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1964–1971.

[4] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *Speech and Audio Processing, IEEE Transactions on*, vol. 5, no. 5, pp. 425–437, 1997.

[5] P. A. Naylor and N. G. Gaubitch, *Speech Dereverberation*. Springer, 2010.

[6] R. Fergus, B. Singh, A. Hertzmann, S. Roweis, and W. Freeman, "Removing camera shake from a single photograph," in *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3. ACM, 2006, pp. 787–794.

[7] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3. ACM, 2008, p. 73.

[8] S. Cho and S. Lee, "Fast motion deblurring," in *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5. ACM, 2009, p. 145.

[9] H. Shen, L. Du, L. Zhang, and W. Gong, "A blind restoration method for remote sensing images," *Geoscience and Remote Sensing Letters, IEEE*, vol. 9, no. 6, pp. 1137–1141, 2012.

[10] M. Wu and D. Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 3, pp. 774–784, 2006.

[11] H. Kameoka, T. Nakatani, and T. Yoshioka, "Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, pp. 45–48.

[12] K. Kumar, R. Singh, B. Raj, and R. Stern, "Gammatone sub-band magnitude-domain dereverberation for ASR," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 4604–4607.

[13] Y. Lin, J. Chen, Y. Kim, and D. Lee, "Blind channel identification for speech dereverberation using ℓ_1-norm sparse learning," in *Advances in Neural Information Processing Systems 20*, 2007.

[14] R. Jenatton, G. Obozinski, and F. Bach, "Structured sparse principal component analysis," *arXiv preprint arXiv:0909.1440*, 2009.

[15] F. Theis, K. Stadlthanner, and T. Tanaka, "First results on uniqueness of sparse non-negative matrix factorization," in *Proceedings of the 13th European Signal Processing Conference (EUSIPCO?05)*, 2005.

[16] R. Gribonval and K. Schnass, "Dictionary identification - sparse matrix-factorization via ℓ^1-minimisation," *Information Theory, IEEE Transactions on*, vol. 56, no. 7, pp. 3523–3539, 2010.

[17] M. Babaie-Zadeh, C. Jutten, and A. Mansour, "Sparse ica via cluster-wise pca," *Neurocomputing*, vol. 69, no. 13, pp. 1458–1466, 2006.

[18] S. Babacan, R. Molina, M. Do, and A. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," *European Conference on Computer Vision (ECCV) Firenze, Italy*, 2012.

[19] J. Money and S. Kang, "Total variation minimizing blind deconvolution with shock filter reference," *Image and Vision Computing*, vol. 26, no. 2, pp. 302–314, 2008.

[20] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 233–240.

[21] T. Chan and C. Wong, "Convergence of the alternating minimization algorithm for blind deconvolution," *Linear Algebra and its Applications*, vol. 316, no. 1, pp. 259–285, 2000.

[22] J. Bolte, P. Combettes, and J. Pesquet, "Alternating proximal algorithm for blind image recovery," in *Proceedings of the IEEE International Conference on Image Processing. Hong-Kong*, 2010.