FAST ITERATIVELY REWEIGHTED LEAST SQUARES FOR L_P REGULARIZED IMAGE DECONVOLUTION AND RECONSTRUCTION

Xu Zhou[1], Rafael Molina[2], Fugen Zhou[1], and Aggelos K. Katsaggelos[3]

[1]Beihang University, Beijing, China
[2] Universidad de Granada, Granada, Spain
[3] Northwestern University, Evanston, USA
xuzhou@sa.buaa.edu.cn, rms@decsai.ugr.es, zhfugen@buaa.edu.cn, aggk@eecs.northwestern.edu

ABSTRACT

Iteratively reweighted least squares (IRLS) is one of the most effective methods to minimize the l_p regularized linear inverse problem. Unfortunately, the regularizer is nonsmooth and nonconvex when $0 < p < 1$. In spite of its properties and mainly due to its high computation cost, IRLS is not widely used in image deconvolution and reconstruction. In this paper, we first derive the IRLS method from the perspective of majorization minimization and then propose an Alternating Direction Method of Multipliers (ADMM) to solve the reweighted linear equations. Interestingly, the resulting algorithm has a shrinkage operator that pushes each component to zero in a multiplicative fashion. Experimental results on both image deconvolution and reconstruction demonstrate that the proposed method outperforms state-of-the-art algorithms in terms of speed and recovery quality.

Index Terms— Image restoration, image reconstruction, compressive sensing, nonconvex nonsmooth regularization, iteratively reweighted least squares

1. INTRODUCTION

In this work, we focus on the $l_2 - l_p$ minimization problem,

$$\min_x f(x) = \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|Rx\|_p \tag{1}$$

where $x \in \Re^n$, $y \in \Re^m$, λ is a nonnegative real number, $\|\cdot\|_p$ means l_p quasi-norm with $0 < p < 1$, A is a $m \times n$ ($n \geq m$) matrix and R is a $r \times n$ matrix. In particular, we assume that $Ker(A) \bigcap Ker(R) = 0$ so that the optimal solution of Eq. (1) exists (see [1]), and both $A^T A$ and $R^T R$ can be diagonalized by the same fast transform F(*e.g.*, DFT).

The $l_2 - l_p$ model has been widely applied in sparse signal recovery, such as image deconvolution [2, 3, 4, 5, 6, 7] and reconstruction [8, 9, 10]. Many published results [3, 4, 5, 6, 7, 11] suggest that $l_p(p < 1)$ regularization has better performance than l_1. This is because l_p not only enforces stronger sparsity than l_1 but it also better preserves edges. As a result, it renders a smooth image while significant image details are better recovered.

However, minimizing $l_2 - l_p$ is not a trivial task mainly because of its nonconvexity. One simple and effective method for minimizing $l_2 - l_p$ is the iteratively reweighted least squares method [3, 11, 12] ([11, 12] solve a constrained l_p minimization problem, in which the observation is noise free) and its variant version [13]. IRLS [3] and [13] with $\alpha = 2$ recursively solves the following reweighted linear equations,

$$(A^T A + \lambda R^T W^t R)x^{t+1} = A^T y \tag{2}$$

where W^t is a diagonal matrix with each component defined by $|Rx^t|$, *e.g.*, $W^t = diag(min(|Rx^t|, 0.01)^{p-2})$ [3]. One major drawback of IRLS is that solving Eq. (2) is computationally expensive because no fast transform can diagonalize $A^T A + \lambda R^T W^t R$. Nevertheless, as we will show soon, Eq. (2) can be efficiently solved by making use of ADMM, under the assumption that both $A^T A$ and $R^T R$ can be diagonalized by the same fast transform.

Related Work. For the $l_2 - l_1$ problem, the most efficient optimization method is ADMM, also known as split Bregman [14] or augmented Lagrangian. ADMM achieves state-of-the-art speed by splitting the original problem into simpler subproblems, which can be easily solved by computationally inexpensive operators (*e.g.*, DFT and shrinkage operator). *Iterative Shrinkage-Thresholding* (IST) based algorithms, such as [15] and [16], are also very useful for this problem. For the $l_2 - l_p$ problem, most recent approaches use smooth approximation [1, 5, 6, 17] and then search the stationary point of approximation function by trust region methods, quasi-Newton iteration or gradient projection. Krishnan and Fergus [4] propose a fast algorithm with variable splitting [8], in

which the nonconvex subproblem is solved by a lookup-table (LUT) method. Instead of formulating a nonconvex objective function, Chartrand suggests using a generalized shrinkage operator for image reconstruction [18, 19].

In this work, we first prove the convergence of the IRLS method in [3] from the perspective of majorization minimization (MM) [20]. By formulating the solution of reweighted linear equations as the minimizer of a particular quadratic function, we propose an ADMM to accelerate IRLS. Each ADMM iteration includes a shrinkage operator that moves every component of the input vector toward zero in a multiplicative way. Experiments show that the proposed method decreases the objective of Eq. (1) efficiently and converges to the limit points at linear rate.

2. FAST IRLS

Since $f(x)$ is not differentiable, we use the following smooth function $J_\varepsilon(x)$ to approximate it,

$$J_\varepsilon(x) = \frac{1}{2}\|Ax - y\|_2^2 + \lambda \sum_i \varphi(R_i x) \qquad (3)$$

where R_i means the ith row vector of R and for $v \in \Re$, $\varphi(v)$ is given by

$$\varphi(v) = \begin{cases} \frac{p}{2}\varepsilon^{p-2}v^2 + \frac{2-p}{2}\varepsilon^p, & \text{if } |v| < \varepsilon \\ |v|^p, & \text{if } |v| \geq \varepsilon \end{cases} \qquad (4)$$

In essence, φ can be viewed as the *Huber* function [1] since they are identical after a linear transform. We prefer this kind of function because λ in J_ε is not scaled by p and f is upper bounded by J_ε. The major merit of the *Huber* approximation is the perfect approximation for $\forall |v| \geq \varepsilon$, but the shortcoming is that it is not second order differentiable at ε. Other smooth approximations can be found in [5, 6, 17].

It is conceivable that a local minimizer of J_ε is also a local minimizer of f when $\varepsilon \to 0$ (based on the lower bound of nonzero entries of $|Rx|$ and continuity of f). Instead of searching for a global minimizer of Eq. (3), we look for a stationary point of Eq. (3). By setting $\nabla J_\varepsilon(x) = 0$, we have

$$\nabla J_\varepsilon(x) = A^T Ax - A^T y + \lambda R^T D_\varepsilon Rx = 0 \qquad (5)$$

where the $D_\varepsilon = diag(min(p\varepsilon^{p-2}, p|Rx|^{p-2}))$. In what follows, we first derive IRLS [3] from the perspective of MM and subsequently show that IRLS [3] converges to a stationary point of J_ε. Finally, we present the fast IRLS algorithm.

Making use of the conjugate concave function principles [21], see also [13] for a definition for the penalty function $\varphi(v) = (|v|^\alpha + \varepsilon)^{\frac{p}{\alpha}-1}(\alpha \geq 1)$, we introduce an auxiliary function

$$G_\varepsilon(x, w) = \frac{1}{2}\|Ax - y\|_2^2 + \frac{\lambda}{2}\sum_i w_i (R_i x)^2 - \frac{p^{1-q}}{q}w_i^q \quad (6)$$

where $0 < w_i \leq p\varepsilon^{p-2}$ and $\frac{2}{p} + \frac{1}{q} = 1$. Notice that, for any $v \geq 0$,

$$\varphi(v) = \min_{0 < w \leq p\varepsilon^{p-2}} \frac{1}{2}wv^2 - \frac{p^{1-q}}{2q}w^q \qquad (7)$$

Furthermore, for a given v, the solution of

$$w_v = \arg\min_{0 < w \leq p\varepsilon^{p-2}} \frac{1}{2}wv^2 - \frac{p^{1-q}}{2q}w^q \qquad (8)$$

is given by

$$w_v = \min(p\varepsilon^{p-2}, p|v|^{p-2}) \qquad (9)$$

For a given $x \in \Re^n$, let $w_x \in \Re^r$ and define its *ith* component $w_x(i)$ as

$$w_x(i) = min(p\varepsilon^{p-2}, p|R_i x|^{p-2}) \qquad (10)$$

Consequently, we have

$$\begin{aligned} J_\varepsilon(x) &= G_\varepsilon(x, w_x) & (11) \\ J_\varepsilon(x') &\leq G_\varepsilon(x', w_x) \quad \forall x' \neq x & (12) \end{aligned}$$

Hence, using the MM framework [20], given an iteration point x^t, if we can find a x^{t+1} satisfying

$$G_\varepsilon(x^{t+1}, w_{x^t}) \leq G_\varepsilon(x^t, w_{x^t}). \qquad (13)$$

It then follows that

$$\begin{aligned} J_\varepsilon(x^{t+1}) &= G_\varepsilon(x^{t+1}, w_{x^{t+1}}) \leq G_\varepsilon(x^{t+1}, w_{x^t}) \\ &\leq G_\varepsilon(x^t, w_{x^t}) = J_\varepsilon(x^t) & (14) \end{aligned}$$

Consequently, the sequence $J_\varepsilon(x^t)$, for $t = 1, 2, ...$, is nonincreasing as long as Eq. (13) holds. In addition, since G_ε is continuous in both x and w, any accumulation point x^* of the MM sequence x^t is a stationary point of $J_\varepsilon(x)$ (see [2]), and $J_\varepsilon(x^t)$ decreases monotonically to $J_\varepsilon(x^*)$.

Let's now proceed to find x^{t+1} satisfying Eq. (13). For a given w_{x^t}, a typical choice for x^{t+1} is the minimizer of $G_\varepsilon(x, w_{x^t})$. Thus, x^{t+1} has the form of

$$x^{t+1} = \arg\min_x \frac{1}{2}x^T A^T Ax - (A^T y)^T x + \frac{1}{2}x^T R^T W^t Rx \qquad (15)$$

where $W^t = \lambda diag(w_{x^t})$. Obviously, the above quadratic function is strictly convex and its minimizer is given by

$$x^{t+1} = (A^T A + R^T W^t R)^{-1} A^T y \qquad (16)$$

Eq. (16) has an IRLS form like [3] but with a p-scaled weight. This IRLS contains two steps only, updating the weights and solving Eq. (16). Better (closer to the desirable) weights probably yield better results. To show this, suppose W^t is formed by x^*, a stationary point of J_ε; it then follows that $x^{t+1} = x^*$, indicating that x^* is also the fixed point of Eq. (16). However, note that since the coefficient matrix of Eq. (16) can not be diagonalized by DFT, solving Eq. (16) may require tens or even hundreds of CG iterations (see [2, 3, 7]), which is very computationally expensive.

Algorithm 1 Fast IRLS Algorithm

Require: $y, A, R, \lambda, p, T, K, \varepsilon_{min}, \varepsilon_{max}$.
1: precompute $\Lambda_A, \Lambda_R, \beta = p\lambda\varepsilon_{min}^{p-2}, Y = FA^T y$
2: $x = y$ (or $x = Y$), $d = 0, \beta_a = p\lambda\varepsilon_{max}^{p-2}$
3: **for** $t = 1$ to T **do**
4:     $W = diag(min(\beta, p\lambda|Rx^t|^{p-2}))$
5:     **for** $k = 1$ to K **do**
6:        update v using Eq.(19)
7:        update d using Eq.(20)
8:        update x using Eq.(18)
9:     **end for**
10: **end for**
11: **return** x

By making use of ADMM, we can find the solution to Eq. (15) by a faster method than solving the linear system Eq. (16). To this end, we first introduce an auxiliary variable $v = Rx$. Accordingly, the unconstrained problem Eq. (15) becomes a constrained problem of the form

$$x^{t+1} = \arg\min_x \frac{1}{2}x^T A^T Ax - (A^T y)^T x + \frac{1}{2}v^T W^t v$$
$$\text{subject to } v = Rx \qquad (17)$$

Since Eq. (17) is convex, it can be efficiently solved by AD-MM (see page 15 in [22]). The resulting IRLS algorithm is shown in Alg.1, in which Λ_A and Λ_R are the eigenvalues of $A^T A$ and $R^T R$ respectively, and the updates for x, v and the dual variable d are given by

$$x = F^{-1}(\Lambda_A + \beta_a\Lambda_R)^{-1}(Y + \beta_a FR^T(v - d)) \qquad (18)$$

$$v = \beta_a(W^t + \beta_a I)^{-1}(Rx + d) \qquad (19)$$

$$d = d + Rx - v \qquad (20)$$

where β_a in Eq. (18) is the penalty parameter which enforces the constraint $Rx = v$ and controls the convergence rate of ADMM. Notice that, the v update Eq. (19) can be considered as a shrinkage operator since it moves all components of the input vector toward zero in a multiplicative style.

Since we are looking for a fixed point of Eq.(16), we adopt the *warm start* strategy like in [7] to further accelerate IRLS. In particular, it is conceivable that Alg.1 with $K = 1$ is faster than Alg.1 with $K \geq 2$ when the same number of x updates is performed, based on the intuition that better weights yield better result. In fact, when $K \geq 2$, the weights are fixed in the $k - loop$, while Alg.1 with $K = 1$ updates the weights immediately after updating x. In addition, if $p = 1$, Alg.1 with $K = 1$ is similar to the well known ADMM-l_1 (e.g., [14]) with a noticeable difference of the shrinkage operator (see Eq. (19)). We observe that our Alg.1 with $K = 1$ is faster than ADMM-l_1 for 20 iterations which is a typical iteration number in image deconvolution.

Compared with the IRLS method in [3], the proposed fast IRLS Alg.1 has two major advantages. First, it makes use of a
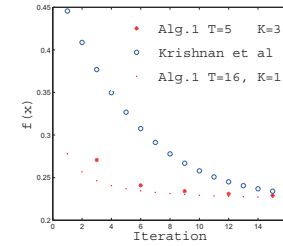

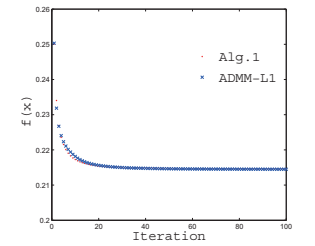(a) Cameraman 256×256 (b) Degraded Image

(c) [4], p=0.8, ISNR=8.78   (d) Alg.1, p=0.8, ISNR=8.83

(e) Evolution of $f(x)$, p=0.8   (f) Evolution of $f(x)$, p=1

Fig. 1. Deconvolution results and evolutions of $f(x)$

fast transform F and ADMM so that it reaches state-of-the-art speed, whereas IRLS [3] cannot utilize F. Second, it allows ε_{min} to be zero without causing numerical problems, whereas IRLS [3] cannot do this. It should be noted that, in order to have $J_\varepsilon(x^{t+1}) \leq J_\varepsilon(x^t)$, we only need an approximate solution of Eq.(15), so that Eq.(13) holds. Luckily, ADMM provides such a good solution at a very low computation cost.

3. EXPERIMENTS

In this section, we examine the performance of the proposed algorithm on image deconvolution and reconstruction problems. All experiments are carried out using MATLAB 7.11 on a 1.8 GHz laptop with 4 GB of memory. For simplicity, we use only two first-order derivative filters ($dx = [1 - 1]$ and $dy = [1 - 1]^T$) to enforce sparsity, while second-order derivative filters [3, 7] and tight frames (like wavelet, $R^T R = I$) can be used to improve the results. Let $Rx = [dx \otimes x; dy \otimes x]$ ($\otimes$

Table 1. Image deconvolution with $l_{0.8}$ regularization

Method	Iterations	Time	$f(x)$	ISNR
[4] with LUT	16	0.62s	0.2318	8.78dB
Alg.1 K=1	16	0.43s	0.2266	8.59dB
Alg.1 K=3	15	0.23s	0.2291	8.83dB

Table 2. Image deconvolution with l_1 regularization

| Method	Iteration=10		Iteration=20		Total
	$f(x)$	ISNR	$f(x)$	ISNR	Time
[14]	0.2181	8.61dB	0.2156	8.63dB	1.44s
Alg.1	0.2173	8.69dB	0.2155	8.68dB	1.46s

means 2-D convolution) and assume periodic boundary conditions, $R^T R$ can be diagonalized by the 2-D DFT. All input images are scaled to the range $[0, 1]$. The MATLAB code is available at website `https://www.researchgate.net/publication/262534911_FIRLS_ICIP2014`.

3.1. Image deconvolution

The *cameraman* image is adopted for image deconvolution. Fig. 1(b) shows the observation image blurred by a 9×9 uniform kernel and added Gaussian noise such that $BSNR = 40dB$. For $l_{0.8}$ deblurring, we choose $\lambda = 0.00002$ for Alg.1 and the algorithm [4]. $\varepsilon_{min} = 2/255$ is adopted for Alg.1 to avoid enforcing too strong sparsity. $\varepsilon_{max} = 20/255$ is used for Alg.1. Table.1 shows the ISNRs of three algorithms as well as the time and objective value. It is clear that Alg.1 with $K = 3$ has the best performance in terms of time and ISNR, yet Alg.1 with $K = 1$ has the smallest objective value but yields the poorest image. Two deblurred results of the three are shown in Fig.1(c-d). As we can see in Fig.1(e), Alg.1 with $K = 1$ has the best performance in minimizing $f(x)$. To further show this, we compare it with [14] in l_1 minimization. We set $\lambda = 0.00003$ and $\beta_a = \lambda(20/255)^{-1}$ for the two, and $\varepsilon_{min} = 0$ is used for the strongest sparsity promotion. The evolutions of $f(x)$ are shown in Fig.1(f) and more details can be found in Table.2. We should mention that the objective function of [14] is lower than ours after 50 iterations.

3.2. Image reconstruction

We choose 256×256 Shepp-Logan phantom for image reconstruction. The mask with 7 radial lines and the corresponding back projection are shown in Fig.2(b) and Fig.2(c), respectively. It is reported in [9] that $p = 0.5$ has the best performance for no less than 10 radial lines. We use the same value for Alg.1 with $K = 1$ and set $\lambda = 10^{-14}$, since the noise level is zero. Again, $\varepsilon_{min} = 0$ is used for the best approximation. ε_{max} should be carefully chosen because the penalty parameter β_a depends on it. We find that $\varepsilon_{max} = 4/255$ works very well. As we show in Fig.2(d), the ISNR reaches

(a) Shepp-Logan   (b) 7 radial lines   (c) Back Projection

(d) Evolution of ISNR   (e) Evolution of $\ln f(x)$
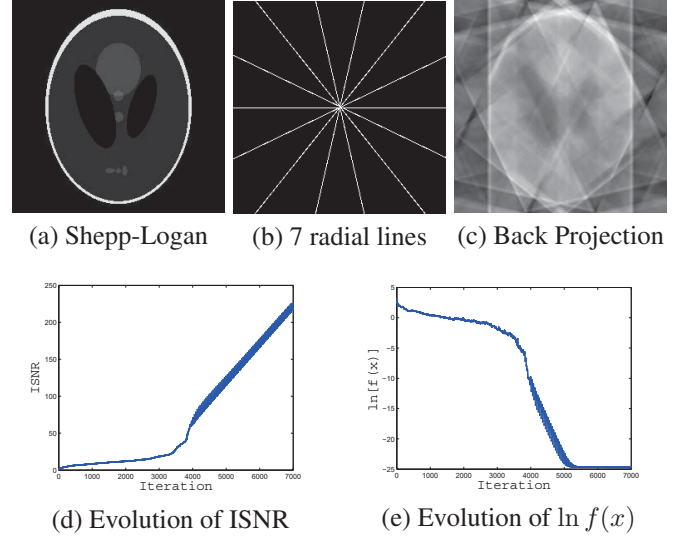
Fig. 2. Image reconstruction with $l_{0.5}$ regularization. Reconstruction is perfect, so not presented

226.3dB after 7000 iterations and 150s. The worst pixel error is 8.57×10^{-11}, whereas [18] takes more than 8000 iterations to obtain perfect reconstruction of error 6.58×10^{-10} from 9 radial lines. A perfect reconstruction from 6 radial lines can be found in [19], but we note that the regularization term used in [19] is not l_p. The evolution of $\ln f(x)$ is shown in Fig.2(e). It is interesting that $f(x)$ does not drop monotonically but still reaches the limit after 5500 iterations.

4. CONCLUSIONS

In this paper, a fast IRLS algorithm for L_p regularized image deconvolution and reconstruction is presented. We show that a weighted linear equation whose coefficient matrix cannot be diagonalized by any fast transform, can be solved by ADMM efficiently. From the perspective of MM, we prove that IRLS [3] converges to a stationary point of J_ε. Experiments show that the proposed algorithm reaches the state-of-the-art speed and yields competitive results. We believe the proposed idea, using ADMM to solve a weighted linear system, can be applied in trust region method [1, 17] as well as other nonsmooth nonconvex minimization problems, such as [23].

5. REFERENCES

[1] M. Hintermüller and T. Wu, "Nonconvex TV^q-models in image restoration: analysis and a trust-region regularization based superlinearly convergent solver," *Siam Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1385–1415, 2013.

[2] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Total

variation-based image deconvolution: A majorization-minimization approach," in *ICASSP*. IEEE, 2006.

[3] A. Levin, R. Fergus, F. Durand, and W.T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph*, vol. 26, no. 3, 2007.

[4] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-laplacian priors," in *NIPS*, 2009.

[5] M. Nikolova, M.K. Ng, and C.P. Tam, "Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction," *IEEE Trans. Image Process*, vol. 19, no. 12, pp. 3073–3088, 2010.

[6] X. Chen, M.K. Ng, and C. Zhang, "Non-lipschitz l_p-regularization and box constrained model for image restoration," *Image Processing, IEEE Transactions on*, vol. 21, no. 12, pp. 4709–4721, 2012.

[7] X. Zhou, F. Zhou, X. Bai, and B. Xue, "A boundary condition based deconvolution framework for image deblurring," *Journal of Computational and Applied Mathematics*, vol. 261, no. 0, pp. 14 – 29, 2014.

[8] Y. L. Wang, J. F. Yang, W. T. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *Siam Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[9] Rick Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *Signal Processing Letters, IEEE*, vol. 14, no. 10, pp. 707–710, 2007.

[10] E. Candés, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *J. Fourier Anal. Appl*, vol. 14, pp. 877–905, 2008.

[11] Rick Chartrand and Wotao Yin, "Iteratively reweighted algorithms for compressive sensing," in *ICASSP*. IEEE, 2008.

[12] I. Daubechies, R. DeVore, M. Fornasier, and C.S. Gunturk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure. Appl. Math*, vol. 63, pp. 1–38, 2010.

[13] Zhaosong Lu, "Iterative reweighted minimization methods for l_p regularized unconstrained nonlinear programming," *arXiv:1210.0066*, 2012.

[14] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *Siam Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[15] J.M. Bioucas-Dias and M.A.T. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *Image Processing, IEEE Transactions on*, vol. 16, pp. 2992–3004, 2007.

[16] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *Siam Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, March 2009.

[17] X. Chen, L. Niu, and Y. Yuan, "Optimality conditions and a smoothing trust region newton method for non-lipschitz optimization," *SIAM Journal on Optimization*, vol. 23, no. 3, pp. 1528–1552, 2013.

[18] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: Mri reconstruction from very few data," in *International Symposium on Biomedical Imaging*. IEEE, 2009.

[19] R. Chartrand and B. Wohlberg, "Generalized shrinkage and penalty functions," in *Global Conference on Signal and Information Processing*. IEEE, 2013.

[20] K. Lange, *Optimization, Springer Texts in Statistics*, Springer-Verlag, 2004.

[21] S. Babacan, R. Molina, M. Do, and A. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *ECCV*, 2012.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122, 2011.

[23] X. Zhou, F. Zhou, and X. Bai, "Blind deconvolution using a nondimensional gaussianity measure," in *ICIP*. IEEE, 2013.