

# Blind Deconvolution via Lower-Bounded Logarithmic Image Priors

Daniele Perrone, Remo Diethelm, and Paolo Favaro

Department of Computer Science and Applied Mathematics  
University of Bern  
Neubrückestrasse 10  
3012 Bern, Switzerland

{perrone,favaro}@iam.unibe.ch, remo.diethelm@outlook.com

**Abstract.** In this work we devise two novel algorithms for blind deconvolution based on a family of logarithmic image priors. In contrast to recent approaches, we consider a minimalistic formulation of the blind deconvolution problem where there are only two energy terms: a least-squares term for the data fidelity and an image prior based on a lower-bounded logarithm of the norm of the image gradients. We show that this energy formulation is sufficient to achieve the state of the art in blind deconvolution with a good margin over previous methods. Much of the performance is due to the chosen prior. On the one hand, this prior is very effective in favoring sparsity of the image gradients. On the other hand, this prior is non convex. Therefore, solutions that can deal effectively with local minima of the energy become necessary. We devise two iterative minimization algorithms that at each iteration solve convex problems: one obtained via the primal-dual approach and one via majorization-minimization. While the former is computationally efficient, the latter achieves state-of-the-art performance on a public dataset.

**Keywords:** blind deconvolution, majorization-minimization, primal-dual, image prior

## 1 Introduction

In the past decade, several high-performing blind deconvolution schemes using Bayesian principles have been proposed [1, 5, 6, 8, 9, 11, 17, 21–23]. The first step in the Bayesian framework is to devise a statistical distribution for both the gradients of the sharp image and the measurement noise or the model error. This joint distribution is used to pose a *maximum a posteriori* (MAP) problem, which yields point estimates for both the sharp image and the blur kernel. Also, one can marginalize the joint distribution with respect to one of the unknown random variables (typically, the sharp image) and then solve maximum a posteriori of the marginalized distribution. However, marginalization is typically computationally intractable. Thus, a variational Bayes upper bound is used together with several approximations such as independence of the random variables and explicit

simplified models of their distributions [21]. Whether one chooses one approach or another, the final algorithm is always an alternating minimization scheme. One iteratively improves the estimates of sharp image, blur kernel and some additional auxiliary variables [9]. The main differences among these schemes lie in how the coefficients weighing each term in the energy being minimized are updated at each iteration. Currently, the general wisdom is probably that the variational Bayes approach yields a better performance than the more classical MAP approach on the joint distribution. This belief is also reinforced by arguments showing how classical MAP approaches, such as total variation blind deconvolution [4], have fundamental shortcomings that would prevent them from achieving the desired solution [10, 21]. In contrast to those findings, in this paper we introduce two novel alternating minimization algorithms that can be cast as classical MAP approaches and that yield state of the art performance. Albeit only experimentally, one can then conclude that there is no inherent advantage in using a MAP or a variational Bayes approach. Moreover, other critical limitations of MAP [10] are overcome by using an alternating minimization with a delayed scaling [14].

As in the vast majority of blind deconvolution algorithms, our approach uses an image prior that strongly encourages sparsity in the image gradients of the reconstructed sharp image. We propose to use the logarithm of the norm of the gradients. While this prior was already introduced in [1] in a variational Bayes framework, here we use it in a MAP approach. Furthermore, to avoid the trivial solution (a constant), we introduce a lower-bound in the norm. This bound is essential to the correct functioning of the prior and needs to be carefully balanced with the data-fidelity term to yield a sparse gradient solution. The other challenges that we address are the non convexity of the image prior even when blur is given and the limited computational efficiency of the alternating minimization scheme. We do so by using two techniques: *majorization-minimization* [7] and the *primal-dual* method [3]. In the first case a tight upper bound of the image prior is obtained and iteratively updated. This algorithm achieves high accuracy. In the second case the Legendre-Fenchel transform and the proximal operator are used to produce iterations that mostly work on independent 1D updates, and that can therefore be executed in parallel. This algorithm is instead highly computationally-efficient.

## 2 Blind Deconvolution

Consider the following model for a blurry image  $f$

$$f = k * u + n \quad (1)$$

where  $k$  is the camera blur (or point spread function),  $u$  is the sharp image and  $n$  is the sensor noise. In this model, blur does not change across the image. This assumption does not hold in real scenes with depth variation and/or with general camera motions. Given both the blurry image  $f$  and the blur  $k$ , the estimation of the sharp image  $u$  is a *deblurring* problem. When instead only the blurry image

$f$  is given, the problem of estimating both the sharp image  $u$  and the blur  $k$  is called *blind deconvolution*.

A widely used framework for solving deblurring or denoising (when  $k$  is the Dirac delta) is to look for a solution to the following minimization problem

$$u = \arg \min_u \|u\|_{TV} + \lambda \|k * u - f\|_2^2 \quad (2)$$

where the first term corresponds to the ubiquitous Total Variation (TV) of  $u$  [16],  $\lambda > 0$  is a regularization constant and the second term corresponds to the data fitting error. This problem is convex and therefore the global minimum can be achieved very efficiently. Often, however, one does not know the blur  $k$  and is therefore faced with the more challenging blind deconvolution problem, which is non convex. Currently, several approaches can successfully obtain high-quality results [1, 5, 6, 8, 9, 11, 17, 21–23]. A formulation of blind deconvolution inspired by eq. (2) is

$$\begin{aligned} u, k = \arg \min_{u, k} \|u\|_{TV} + \lambda \|k * u - f\|_2^2 \\ \text{s.t. } \quad k \succcurlyeq 0, \quad \|k\|_1 = 1, \end{aligned} \quad (3)$$

which has already been proposed in the past [4, 24]. This formulation, however, suffers from several limitations. Firstly, the global minima of this problem are the no-blur solutions, where  $u = f$  and  $k = \delta$ , up to translation [14, 10, 21]. Secondly, this is a non convex problem in both  $u$  and  $k$ . Thus, the solution obtained via an iterative method depends on the initialization of the unknowns. Despite the limitations outlined above, early alternating minimization implementations [4] for eq. (3) converged to desirable solutions. While many of current algorithms are derived via variational Bayes arguments or based on edge enhancements and noise suppression, they eventually result in alternating minimization schemes each resembling that used to solve eq. (3) [9]. The key differences are the introduction of additional auxiliary variables and the dynamic update of regularization parameters. As demonstrated experimentally, any such variation leads to a quite different performance. Moreover, most recent approaches choose to solve this problem by working on gradients of the images, or, more in general, filtered images rather than directly on the intensities of the images. Then, once the blur  $k$  has been estimated, a final deblurring step is performed to obtain the sharp image  $u$ .

In this paper we show that two minimalistic optimization schemes working directly on the image intensities are sufficient to achieve state of the art performance. This allows us to conclude that any additional modification to the alternating minimization scheme, including working on filtered images, is not essential to converge to the desired solution. Before doing so, however, we briefly review relevant work and clarify differences when compared to our solutions.

### 3 Prior Work

As mentioned in the Introduction, most approaches in blind deconvolution can be described as *maximum a posteriori* (MAP) approaches. The MAP approach

relies on an explicit definition of the joint probability

$$p(u, k|f) \propto p(f|u, k)p(u)p(k), \quad (4)$$

where  $p(f|u, k)$  is a generative model of the noise,  $p(u)$  is a prior of the sharp image and  $p(k)$  is a prior of the blur. Commonly used sharp image priors approximate the heavy-tail distribution of natural image gradients [18] via sparsity-inducing norm of the gradients of  $u$ . The  $\ell_2$  norm of the gradients (isotropic total variation), or the  $\ell_1$  norm of the derivatives (anisotropic total variation) are classical choices [4]. In contrast to other sparsity-inducing norms, total variation (TV)[16] has the desirable property of being convex. However, it also introduces a loss of contrast in the recovered sharp image [14, 19]. Other methods use heuristics to encourage sharp gradients [5, 22, 17], or some reweighing strategy of the norm of the gradients [8, 9]. The latter methods aim at approximating the  $l_0$  “norm” of the gradients, as proposed also by Xu et al. [23]. In this paper we also encourage sparsity in the gradients. However, we use the logarithm of TV, which yields a simple energy term while providing a good approximation to the number of nonzero gradients. Indeed, this prior has already demonstrated promising results in blind deconvolution [1, 21] and denoising [13].

Despite the widespread use of the above MAP formulation, finding the mode of the posterior probability of  $u$  and  $k$  has received many criticisms. Levin et al. [10] and Perrone and Favaro [14] have shown that a large class of commonly used image priors favors the blurry image instead of the sharp one. Because of such limitation, Levin et al. [11] suggested to marginalize over all possible sharp images  $u$  and thus to solve the reduced problem

$$\max_k p(k|f) = \min_k -\log p(k|f) = \min_k -\log \int_u p(u, f|k)p(k)du. \quad (5)$$

In general, the integral in problem (5) is intractable. Therefore, typically one looks for an approximate solution. A common approach is to minimize an upper bound of  $-\log p(k|f)$  using a variational Bayes strategy [1, 6, 11, 21]. So far, this class of methods has shown better performance compared to methods that directly solve the MAP problem (4).

Despite their apparent superior performance, Wipf and Zhang [21] have shown that methods that solve (5) using a variational Bayes strategy are equivalent to a MAP strategy as in (4). They experimentally show that with an  $\ell_p$  norm with  $p \ll 1$ , MAP approaches are able to favor the right sharp solution. They also argue that a variational Bayes approach should be preferred because it is more robust when minimizing a highly non-convex function. The conclusions given by Wipf and Zhang [21] suggest that minimizing a cost functional as in (4) is not limited per se, as long as one finds a minimization strategy that carefully avoids its local minima. In this paper, we propose two non-variational Bayes strategies to minimize a functional based on a logarithmic non-convex prior and show that they can achieve state-of-the-art results.

## 4 A Logarithmic Image Prior

In this section we introduce our image prior. From a Bayesian perspective, natural images can be described as having a sparse collection of gradients [18]. Hence, one could employ sparsity-inducing priors of the image gradients. However, another point of view is that blurring results in the average of shifted and scaled replicas of the same image gradients. The likelihood that such replicas combine to cancel each other is statistically insignificant. Vice versa, this averaging is more likely to multiply the number of gradients by the number of nonzero elements in the blur. Thus, a different perspective is that, in the context of deblurring, the role of an image prior is to favor solutions that have as few gradients as possible regardless of their magnitude. Both points of view lead to the same principle, *i.e.*, one should choose as prior

$$\text{Number of non zero elements of } (|\nabla u|) \doteq \|\nabla u\|_0 \quad (6)$$

where  $\|\cdot\|_0$  denotes the  $l_0$  “norm” (the Hamming distance to zero) and  $\nabla u$  is the 2-D gradient of  $u$ . Unfortunately, optimization with this prior is very challenging and, typically, smoother alternatives such as  $\ell_p$  norms  $\|\nabla u\|_p^p$ , with  $0 < p < 1$ , are used. In this work we also consider a prior with a similar behavior and simple form.

Let us consider the discrete setting. In the 2D discrete case, we have images with  $N \times M$  pixels. The  $(i, j)$ -th entry of the blurry image  $u$  will be denoted by  $u_{i,j}$ . The first order (discrete) derivatives of  $u$  will be denoted by  $\nabla u \doteq [u_{i+1,j} - u_{i,j} \quad u_{i,j+1} - u_{i,j}]^T$ . As image prior we propose using the following logarithmic prior<sup>1</sup>

$$\log \|\nabla u\|_{2,\epsilon}^p \doteq \sum_{i=1}^N \sum_{j=1}^M \log \|\nabla u_{i,j}\|_{2,\epsilon}^p = \frac{p}{2} \sum_{i=1,j=1}^{N,M} \log \|\nabla u_{i,j}\|_{2,\epsilon}^2 \quad (7)$$

with  $p > 0$  and where

$$\|\nabla u_{i,j}\|_{2,\epsilon}^2 \doteq (u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 + \epsilon^2 \quad (8)$$

for  $\epsilon > 0$  so that the argument of the logarithm is never 0. The parameter  $\epsilon$  leads to a lower bound for this prior equal to  $MNp \log \epsilon$ . We can then formulate our blind deconvolution problem as

$$\begin{aligned} u, k &= \arg \min_{u,k} \lambda \|k * u - f\|_2^2 + \log \|\nabla u\|_{2,\epsilon}^p \\ \text{s.t.} \quad & k \succcurlyeq 0, \quad \|k\|_1 = 1. \end{aligned} \quad (9)$$

Notice how the role of  $\epsilon$  is fundamental. If  $\epsilon = 0$  then the optimal solution will always be  $u = 0$  for any  $\lambda$ . To understand how  $\epsilon$ ,  $\lambda$  and  $p$  relate, consider the

<sup>1</sup> Although we choose an  $\ell_2$  norm, any  $\ell_q$  norm could be used. However, we have found experimentally that for a wide set of values in  $q$  this makes little difference in the final performance.

following limit

$$\lim_{\epsilon \rightarrow 0} \frac{p}{2} + \frac{1}{\log(1/\epsilon^2)} \log \|\nabla u\|_{2,\epsilon}^p = \frac{p}{2} \|\nabla u\|_0 \quad (10)$$

which shows how the log prior approximates the desired  $l_0$  “norm”. Now, assume that  $0 < \epsilon \leq 1$  and we substitute  $\lambda$  in problem (9) with  $-\lambda p \log \epsilon$ . Then, in the limit for  $\epsilon \rightarrow 0$  we are solving

$$\begin{aligned} u, k &= \arg \min_{u,k} \lambda \|k * u - f\|_2^2 + \|\nabla u\|_0 \\ \text{s.t.} \quad &k \succeq 0, \quad \|k\|_1 = 1. \end{aligned} \quad (11)$$

Finally, to avoid the degenerate constant solution we can compare two cases: one when  $u = \text{constant}$  and one when  $u = f$  and  $k = \delta$ . The idea is to make sure that the cost function favors the no-blur solution over the constant one. We can therefore plug the two cases in the cost of problem (9) and obtain the following inequality

$$\log \|\nabla f\|_{2,\epsilon}^p < \lambda \|\bar{f} - f\|_2^2 + \frac{p}{2} MN \log \epsilon^2 \quad (12)$$

or, alternatively,

$$\log \left\| \frac{1}{\epsilon} \nabla f \right\|_{2,1}^p < \lambda \|\bar{f} - f\|_2^2. \quad (13)$$

Then, we use Jensen’s inequality and the fact that the logarithm is a concave function to obtain an upper bound of the left hand side of eq. (13)

$$\frac{p}{2} \sum_{i=1,j=1}^{N,M} \log \left[ \left\| \frac{1}{\epsilon} \nabla f_{i,j} \right\|_{2,1}^2 \right] \leq \frac{pMN}{2} \log \left[ \frac{1}{MN} \sum_{i=1,j=1}^{N,M} \left\| \frac{1}{\epsilon} \nabla f_{i,j} \right\|_{2,1}^2 \right]. \quad (14)$$

Then, if we choose  $\epsilon$  such that

$$\epsilon > \sqrt{\frac{\frac{1}{MN} \sum_{i=1,j=1}^{N,M} \|\nabla f_{i,j}\|_2^2}{e^{\frac{2\lambda}{pMN} \|f - \bar{f}\|_2^2} - 1}}} \quad (15)$$

where  $\bar{f}$  is the average value of  $f$ , the degenerate constant solution will be avoided. Also, notice that  $\frac{2\lambda}{pMN} \|f - \bar{f}\|_2^2 > 0$  and  $\frac{1}{MN} \sum_{i=1,j=1}^{N,M} \|\nabla f_{i,j}\|_2^2 > 0$  unless  $f$  is constant (in this case  $u$  constant is a plausible solution and it should not be avoided). This means that an  $\epsilon$  that satisfies eq. (15) always exists.

Finally, to solve problem (9) we use the alternating minimization scheme

<pre> initialize     <math>k^1 = k_1</math> iterate <math>t \in [1, \dots, T]</math>     <math>u^{t+1} = \arg \min_u \lambda \ k^t * u - g\ _2^2 + \log \ \nabla u\ _{2,\epsilon}^p</math>     <math>k^{t+1} = \arg \min_k \ k * u^{t+1} - f\ _2^2</math>     s.t. <math>k \succcurlyeq 0, \quad \ k\ _1 = 1.</math> </pre>	(16)
---	------

While the iteration in the blur  $k$  entails solving a convex problem, and we solve it as in [4], the minimization in the update of the sharp image  $u$  is non convex and requires more attention. To this purpose we introduce two solvers: one based on a primal-dual approach and another on majorization-minimization.

#### 4.1 A Primal-Dual Solver

Recall the deblurring problem (given the blur  $k^t$ ) in Algorithm (16); here we rewrite it as

$$u = \arg \min_u \sum_{i=1, j=1}^{N, M} ((k^t * u)_{i,j} - f_{i,j})^2 + \frac{1}{\mu} \log \|\nabla u_{i,j}\|_{2,\epsilon}^2 \quad (17)$$

where  $\mu = 2\lambda/p$ . By using the primal-dual approach of Chambolle and Pock [3] we obtain the following minimax problem

$$u = \arg \min_u \max_{z_1, z_2} \langle k^t * u, z_1 \rangle - F_1^*(z_1) + \langle \nabla u, z_2 \rangle - F_2^*(z_2) \quad (18)$$

where  $F_1^*$  and  $F_2^*$  are conjugate functions of  $F_1$  and  $F_2$  respectively, and we have defined

$$F_1(x) \doteq \sum_{i=1, j=1}^{N, M} (x_{i,j} - f_{i,j})^2, \quad F_2(\xi) \doteq \frac{1}{\mu} \sum_{i=1, j=1}^{N, M} \log \|\xi_{i,j}\|_{2,\epsilon}^2. \quad (19)$$

The conjugate functions can be computed via the Legendre-Fenchel (LF) transform [15] and are convex by construction. Thus problem (18) is a convex approximation in all variables  $z_1$ ,  $z_2$  and  $u$  of the original problem (17). Notice that the convex approximation provided by the primal-dual formulation may not lead to the minima of the original non convex cost.

Our general primal-dual algorithm to solve problem (18) is

<pre> <math>z_1^{n+1} = \text{prox}_{\sigma F_1^*}(z_1^n + \sigma k^t * \bar{u}^n)</math> <math>z_2^{n+1} = \text{prox}_{\sigma F_2^*}(z_2^n + \sigma \nabla \bar{u}^n)</math> <math>u^{n+1} = u^n - \tau (k^t * z_1^{n+1} + \nabla \cdot z_2^{n+1})</math> <math>\bar{u}^{n+1} = u^{n+1} + \theta(u^{n+1} - u^n)</math> </pre>	(20)
---	------

where  $k_-^t$  denotes the mirrored blur kernel  $k^t$  (along both axes),  $n$  is the iteration index,  $\theta \in (0, 1]$  and  $\tau\sigma\|K\|^2 < 1$ , with  $\tau, \sigma > 0$ , where  $K$  is the matrix operator implementing both the blur  $k$  and the finite difference operator  $\nabla$ . Two of the 4 iterations in the above algorithm are defined based on the proximity operator. The proximity operator  $\text{prox}_{\sigma F_1^*}$  is computed via

$$\begin{aligned}\text{prox}_{\sigma F_1^*}(z) &= z - \sigma \text{prox}_{F_1/\sigma}(z/\sigma) \\ &= z - \sigma \arg \min_x \frac{1}{2} \left\| \frac{z}{\sigma} - x \right\|_2^2 + \sigma F_1(x) \\ &= \frac{1}{\sigma + 1} (z - \sigma f).\end{aligned}\quad (21)$$

The proximity operator  $\text{prox}_{\sigma F_2^*}$  is instead computed via

$$\text{prox}_{\sigma F_2^*}(z) = z - \sigma \arg \min_x \frac{1}{2} \left\| \frac{z}{\sigma} - x \right\|_2^2 + \sigma F_2(x). \quad (22)$$

Since the minimization problem is separable, let us consider the solution obtained for only one element  $x_{i,j}$  and  $z_{i,j}$  of the variables  $x$  and  $z$  respectively. With an abuse of notation, instead of the element-wise cumbersome notation  $x_{i,j}$  and  $z_{i,j}$  we simply refer to  $x$  and  $z$  in the next equations. We use the representation  $x \doteq \rho w$ , where  $\rho \geq 0$  and  $\|w\|_2 = \|z\|_2/\sigma$ . Then, let  $\xi = z/\sigma$  and we have

$$\arg \min_x \frac{1}{2} \|\xi - x\|_2^2 + \sigma F_2(x) = \arg \min_{\rho, w} \frac{\rho^2}{2} \left\| \frac{\xi}{\rho} - w \right\|_2^2 + \frac{\sigma}{\mu} \log \left( \rho^2 \frac{\|z\|_2^2}{\sigma^2} + \epsilon^2 \right). \quad (23)$$

Notice that the logarithmic term now depends only on  $\rho$ . Hence, we can first solve the minimization problem with respect to  $w$ . By simplifying the least squares term we obtain

$$\begin{aligned}\arg \min_{w, \|w\|=\|z\|/\sigma} \frac{\rho^2}{2} \left\| \frac{\xi}{\rho} - w \right\|_2^2 &= \arg \min_{w, \|w\|=\|z\|/\sigma} \|\xi\|_2^2 / \rho^2 + \|w\|_2^2 - 2\langle \xi/\rho, w \rangle \\ &= \arg \min_{w, \|w\|=\|z\|/\sigma} \|\xi\|_2^2 / \rho^2 + \|z\|_2^2 / \sigma^2 - 2\langle \xi/\rho, w \rangle \\ &= \arg \max_{w, \|w\|=\|z\|/\sigma} \langle \xi, w \rangle\end{aligned}\quad (24)$$

which immediately yields  $w = \frac{\|z\|_2}{\sigma\|\xi\|_2} \xi = z/\sigma$ . By substituting the expression of  $w$  back into eq. (23) we finally have

$$\begin{aligned}\arg \min_x \frac{1}{2} \|\xi - x\|_2^2 + \sigma F_2(x) &= \xi \cdot \arg \min_{\rho} \frac{1}{2} \|\xi - \rho z/\sigma\|_2^2 + \frac{\sigma}{\mu} \log \left( \rho^2 \frac{\|z\|_2^2}{\sigma^2} + \epsilon^2 \right) \\ &= \xi \cdot \arg \min_{\rho} \frac{\mu}{2\sigma} (1 - \rho)^2 \|\xi\|_2^2 + \log \left( \rho^2 \|\xi\|_2^2 + \epsilon^2 \right).\end{aligned}\quad (25)$$

**Table 1.** The primal-dual algorithm.

<p><b>initialize</b></p> $h^1 = h_1$ <p><b>iterate</b> <math>t \in [1, \dots, T]</math></p> <p style="padding-left: 20px;"><b>iterate</b> <math>n \in [1, \dots, N_0]</math></p> $z_1^{n+1} = \frac{1}{\sigma + 1} (z_1^n + \sigma(k^t * \bar{u}^n - f))$ $z_2^{n+1} = \left(1 - \mathcal{H}\left(\frac{z_2^n + \sigma \nabla \bar{u}^n}{\sigma}, \mu, \epsilon, \sigma\right)\right) (z_2^n + \sigma \nabla \bar{u}^n)$ $\tilde{u}^{n+1} = \tilde{u}^n - \tau (k_-^t * z_1^{n+1} + \nabla \cdot z_2^{n+1})$ $\bar{u}^{n+1} = \tilde{u}^{n+1} + \theta(\tilde{u}^{n+1} - \tilde{u}^n)$ <p style="padding-left: 20px;"><b>end iterate</b> <math>n</math></p> $u^{t+1} = \tilde{u}^{N_0+1}$ $h^{t+1} = \arg \min_k \ k * u^{t+1} - f\ _2^2$ <p style="padding-left: 40px;">s.t. <math>k \succcurlyeq 0, \quad \ k\ _1 = 1</math></p> <p><b>end iterate</b> <math>t</math></p>
---

We can define  $\mathcal{H}$  as the solution of the 1D problem

$$\mathcal{H}(\xi, \mu, \epsilon, \sigma) = \arg \min_{\rho} \frac{\mu}{2\sigma} (\rho - 1)^2 \|\xi\|_2^2 + \log(\rho^2 \|\xi\|_2^2 + \epsilon^2) \quad (26)$$

and build it into a lookup table.<sup>2</sup> The proximity operator  $\text{prox}_{\sigma F_2^*}$  can then be written as

$$\text{prox}_{\sigma F_2^*}(z) = \left(1 - \mathcal{H}\left(\frac{z}{\sigma}, \mu, \epsilon, \sigma\right)\right) z. \quad (27)$$

The final algorithm is summarized in Table 1. Notice how several operations are parallelizable, thus leading to a very efficient implementation.

## 4.2 A Majorization-Minimization Approach

As a more accurate alternative to the primal-dual algorithm, one could use a majorization-minimization (MM) approach [7], in a similar manner as proposed by Candes *et al.* [2]. In the MM approach one defines an upper bound functional  $\psi(u|u^t)$  given the current estimate  $u^t$  at time  $t$ . This upper bound must satisfy the following properties

$$\psi(u|u^t) \geq \sum_{i=1}^N \sum_{j=1}^M \log \|\nabla u_{i,j}\|_{2,\epsilon}^p, \quad \psi(u^t|u^t) = \sum_{i=1}^N \sum_{j=1}^M \log \|\nabla u_{i,j}^t\|_{2,\epsilon}^p. \quad (28)$$

<sup>2</sup> Notice that the 1D problem leads to a third order polynomial equation for which closed-form solutions are known.

**Table 2.** The majorization-minimization algorithm.

<b>initialize</b> $h^1 = h_1$ <b>iterate</b> $t \in [1, \dots, T]$ $u^{t+1} = \arg \min_u \sum_{i=1, j=1}^{N, M} \lambda ((k^t * u)_{i,j} - f_{i,j})^2 + \frac{\ \nabla u_{i,j}\ _2^p}{\ \nabla u_{i,j}^t\ _2^p}$ $h^{t+1} = \arg \min_k \ k * u^{t+1} - f\ _2^2$ <p style="text-align: center;">s.t. <math>k \succcurlyeq 0, \quad \ k\ _1 = 1</math></p> <b>end iterate</b> t
--

Then, one can apply the following iterative scheme and provably reach a local minimum of the original function

$$u^{t+1} = \arg \min_u \sum_{i=1, j=1}^{N, M} \lambda ((k * u)_{i,j} - f_{i,j})^2 + \psi(u|u^t). \quad (29)$$

As upper bound we consider using the Taylor expansion of the logarithm around the  $t$ -th estimate of  $\|\nabla u\|_{2,\epsilon}^p$  up to the first term

$$\psi(u|u^t) = \sum_{i=1, j=1}^{N, M} \log \|\nabla u_{i,j}^t\|_{2,\epsilon}^p + \frac{\|\nabla u_{i,j}\|_{2,\epsilon}^p - \|\nabla u_{i,j}^t\|_{2,\epsilon}^p}{\|\nabla u_{i,j}^t\|_{2,\epsilon}^p}. \quad (30)$$

The properties (28) hold because of the concavity of the logarithm function. Finally, by plugging  $\psi$  in eq. (29) we obtain the following update

$$\begin{aligned} u^{t+1} &= \arg \min_u \sum_{i=1, j=1}^{N, M} \lambda ((k * u)_{i,j} - f_{i,j})^2 \\ &\quad + \log \|\nabla u_{i,j}^t\|_{2,\epsilon}^p + \frac{\|\nabla u_{i,j}\|_{2,\epsilon}^p - \|\nabla u_{i,j}^t\|_{2,\epsilon}^p}{\|\nabla u_{i,j}^t\|_{2,\epsilon}^p} \\ &= \arg \min_u \sum_{i=1, j=1}^{N, M} \lambda ((k * u)_{i,j} - f_{i,j})^2 + \frac{\|\nabla u_{i,j}\|_{2,\epsilon}^p}{\|\nabla u_{i,j}^t\|_{2,\epsilon}^p}. \end{aligned} \quad (31)$$

so that the majorization-minimization algorithm can be summarized in Table 2. Notice the similarity with reweighed least squares algorithms when  $p = 2$ .

## 5 Experiments

We evaluated the proposed algorithms on the dataset from Levin et al. [10]. The dataset is made of 4 images of size  $255 \times 255$  pixels blurred with 8 different

blurs, and it is provided with ground truth sharp images and blurs. Therefore it is possible to use metrics that take into account the intrinsic difficulty of each blur, such as the SSD ratio proposed in [10]. This ratio can be computed by

$$r = \frac{\sum_{i=1, j=1}^{N, M} (u_{i,j}^{k^e} - u_{i,j}^g)^2}{\sum_{i=1, j=1}^{N, M} (u_{i,j}^{k^g} - u_{i,j}^g)^2} \quad (32)$$

where  $u^g$  is the ground truth sharp image,  $u^{k^g}$  is the image obtained by solving a non-blind deconvolution problem with the ground truth blur, and  $u^{k^e}$  is the image obtained by solving a non-blind deconvolution problem with the estimated blur. For each method the same parameters are used for all the 32 blurry images of the dataset. For all the tests we used the non-blind deconvolution algorithm from Levin et. al. [12], where for each method we carefully selected the regularization parameter in order to have the best SSD ratio.

In Fig. 1 we show the cumulative histogram of the SSD ratios for several methods in the literatures and for our proposed algorithms (Log-TV MM and Log-TV PD). The MM algorithm achieves error ratio equal to 1 for more than 50% of the images, clearly outperforming the methods from Wipf and Zhang [21] and Babacan et. al. [1], and, for most error ratios, the method of Sun *et al.* [20]. Our primal-dual method is on par with high performing variational Bayesian algorithms such as the one from Levin et. al. [11]. In Fig. 2 we also show the cumulative histogram of the SSD errors, while in Fig. 3 we show some of the sharp images obtained on this dataset<sup>3</sup>.

For our methods we used the same regularization parameter  $\lambda = 30000$ ,  $\epsilon = 0.001$ ,  $p = 2$  and 3500 iterations for each pyramid level. For the primal-dual algorithm we set  $N_0 = 1$ ,  $\tau = 0.005$  and  $\sigma = \frac{1}{32\tau}$ . The parameter values have been found experimentally. We used a pyramid scheme where the input image and the blur are down sampled at each level by  $\sqrt{2}$ , and the parameter  $\lambda$  is divided by the number 2.1. The number of levels of the pyramid is computed such that at the top level the blur kernel has a support of 3 pixels. For the other methods we used the estimates provided by the authors, or we ran their algorithm using the tuning that gives the best results. The primal-dual method has the desirable feature of being parallelizable and therefore faster, but at the cost of being too coarse (due to the convex approximation of the logarithmic prior), thus unable to achieve the same accuracy of the MM algorithm.

## 6 Conclusions

In this paper we presented solutions to blind deconvolution based on a logarithmic image prior. The chosen prior is as effective as  $\ell_p$  norms with  $p < 1$  on the image gradients, while at the same time leading to simple optimization schemes despite its non convexity. To solve blind deconvolution with this image prior we propose a computationally efficient scheme via a primal-dual approach

<sup>3</sup> A list of all the experiments is available at [www.cvg.unibe.ch/dperrone/logtv/](http://www.cvg.unibe.ch/dperrone/logtv/)

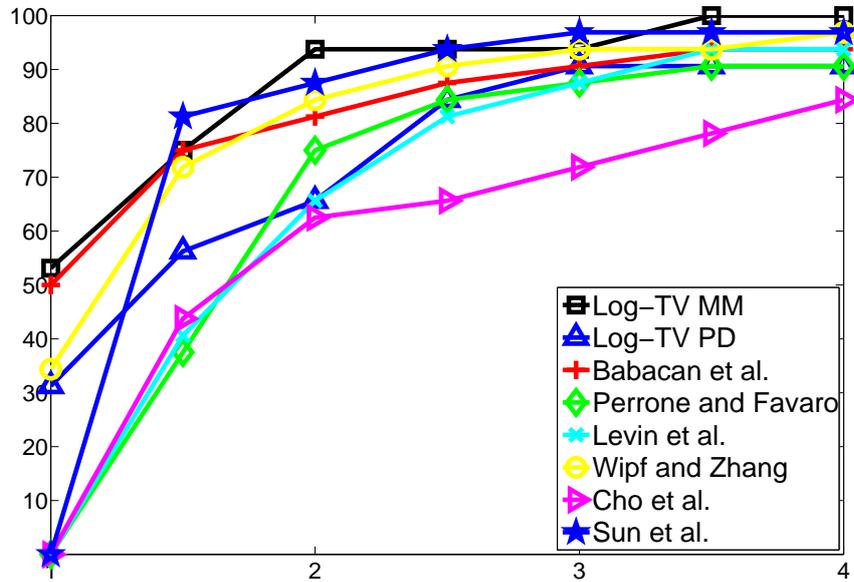


Fig. 1. Cumulative histogram of SSD ratio results on the dataset [10].

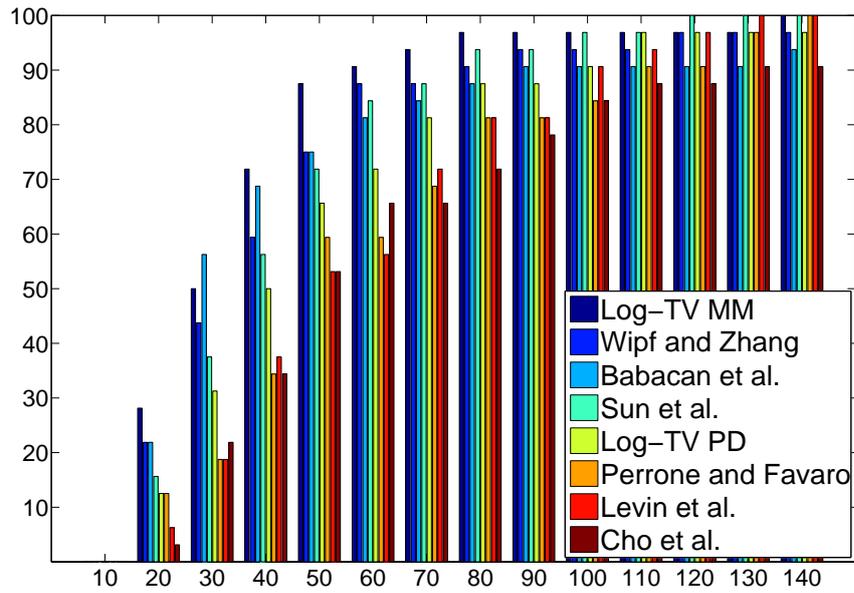
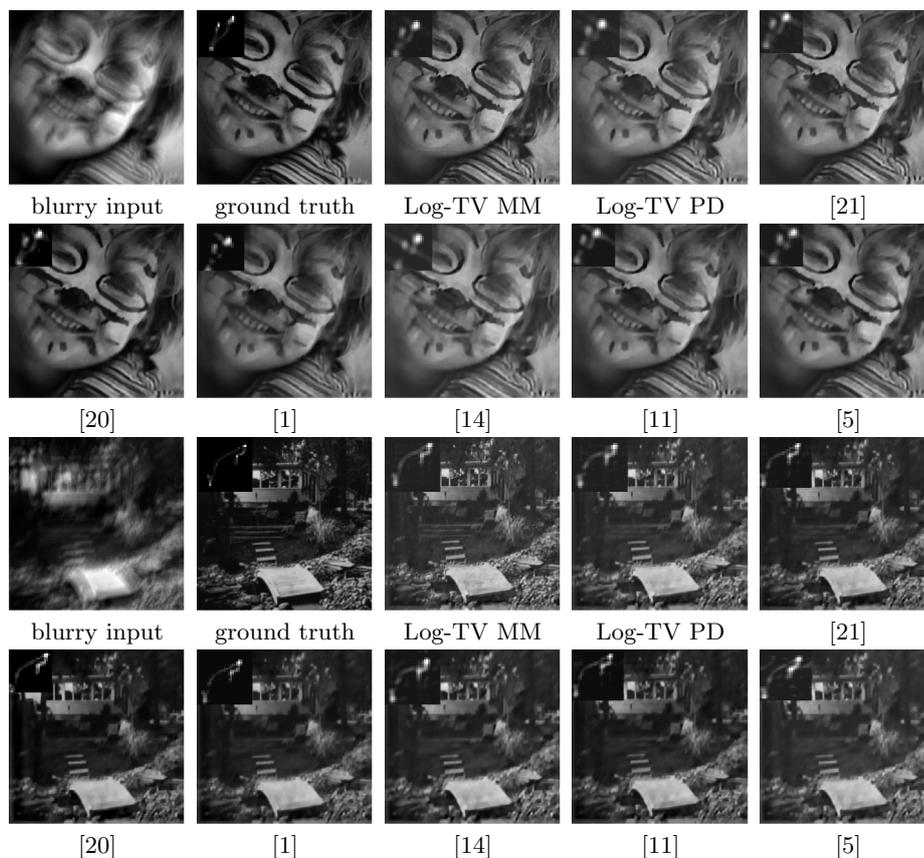


Fig. 2. Cumulative histogram of SSD results per image of the database [10].



**Fig. 3.** Examples of deblurred images from Levin et al. [10] dataset.

and a high-accuracy scheme via the majorization-minimization approach. Both approaches perform well and converge very robustly.

## References

1. Babacan, S., Molina, R., Do, M., Katsaggelos, A.: Bayesian blind deconvolution with general sparse image priors. In: *Computer Vision – ECCV 2012, Lecture Notes in Computer Science*, vol. 7577, pp. 341–355. Springer Berlin Heidelberg (2012)
2. Candes, E.J., Wakin, M.B., Boyd, S.: Enhancing sparsity by reweighted  $l_1$  minimization. *Journal of Fourier Analysis and Applications* 14(5), 877–905 (2008)
3. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision* 40(1), 120–145 (2011)
4. Chan, T., Wong, C.K.: Total variation blind deconvolution. *Image Processing, IEEE Transactions on* 7(3), 370–375 (Mar 1998)

5. Cho, S., Lee, S.: Fast motion deblurring. *ACM Trans. Graph.* 28(5), 145:1–145:8 (Dec 2009)
6. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. *ACM Trans. Graph.* 25(3), 787–794 (Jul 2006)
7. Hunter, D., Lange, K.: A tutorial on mm algorithms. *The American Statistician* 58, 30–37 (2004)
8. Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. pp. 233–240 (June 2011)
9. Krishnan, D., Bruna, J., Fergus, R.: Blind deconvolution with re-weighted sparsity promotion. *CoRR abs/1311.4029* (2013)
10. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Understanding blind deconvolution algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(12), 2354–2367 (2011)
11. Levin, A., Weiss, Y., Durand, F., Freeman, W.: Efficient marginal likelihood optimization in blind deconvolution. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. pp. 2657–2664 (June 2011)
12. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.* 26(3) (Jul 2007)
13. Ochs, P., Chen, Y., Brox, T., Pock, T.: ipiano: Inertial proximal algorithm for nonconvex optimization. *SIAM J. Imaging Sciences* 7(2), 1388–1419 (2014)
14. Perrone, D., Favaro, P.: Total variation blind deconvolution: The devil is in the details. In: *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. pp. 2909–2916 (June 2014)
15. Rockafellar, R.: *Convex Analysis*. Princeton University Press (1970)
16. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D* 60(1-4), 259–268 (Nov 1992)
17. Shan, Q., Jia, J., Agarwala, A.: High-quality motion deblurring from a single image. *ACM Trans. Graph.* 27(3), 73:1–73:10 (Aug 2008)
18. Srivastava, A., Lee, A., Simoncelli, E.P., Zhu, S.C.: On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision* 18, 17–33 (January 2003)
19. Strong, D., Chan, T.: Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems* 19(6), S165 (2003)
20. Sun, L., Cho, S., Wang, J., Hays, J.: Edge-based blur kernel estimation using patch priors. In: *Computational Photography (ICCP), 2013 IEEE International Conference on*. pp. 1–8 (April 2013)
21. Wipf, D., Zhang, H.: Analysis of bayesian blind deconvolution. In: *Energy Minimization Methods in Computer Vision and Pattern Recognition*. pp. 40–53. Springer (2013)
22. Xu, L., Jia, J.: Two-phase kernel estimation for robust motion deblurring. In: *Proceedings of the 11th European Conference on Computer Vision: Part I*. pp. 157–170. *ECCV’10*, Springer-Verlag, Berlin, Heidelberg (2010)
23. Xu, L., Zheng, S., Jia, J.: Unnatural l0 sparse representation for natural image deblurring. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. pp. 1107–1114 (June 2013)
24. You, Y.L., Kaveh, M.: A regularization approach to joint blur identification and image restoration. *Image Processing, IEEE Transactions on* 5(3), 416–428 (1996)