The Maximum Entropy on the Mean Method for Image Deblurring

Rustum Choksi, Tim Hoheisel, Gabriel Rioux, and Christopher Scarvelis
Department of Mathematics and Statistics, McGill University, Montreal, QC H3G 1Y6, Canada
E-mail: \{rustum.choksi,tim.hoheisel\}@mcgill.ca, \{gabriel.rioux,christopher.scarvelis\}@mail.mcgill.ca

May 2020

Abstract. Image deblurring is a notoriously challenging ill-posed inverse problem. In recent years, a wide variety of approaches have been proposed based upon regularization at the level of the image or on techniques from machine learning. We propose an alternative approach, shifting the paradigm towards regularization at the level of the probability distribution on the space of images. Our method is based upon the idea of maximum entropy on the mean wherein we work at the level of the probability density function of the image whose expectation is our estimate of the ground truth. Using techniques from convex analysis and probability theory, we show that the approach is computationally feasible and amenable to very large blurs. Moreover, when images are embedded with symbology (a known pattern), we show how our method can be applied to approximate the unknown blur kernel with remarkable effects. While our method is stable with respect to small amounts of noise, it does not actively denoise. However, for moderate to large amounts of noise, it performs well by preconditioned denoising with a state of the art method.

Keywords: Image Deblurring, Maximum Entropy on the Mean, Kullback-Leibler Divergence, Convex Analysis, Optimization

1. Introduction

Ill-posed inverse problems permeate the fields of image processing and machine learning. Prototypical examples stem from non-blind (deconvolution) and blind deblurring of digital images. The vast majority of methods for image deblurring are based on some notion of regularization (e.g. gradient-based) at the image level. Motivated by our previous work \cite{22} for barcodes, we address general image deblurring at the level of the probability density function of the ground truth. Using Kullback-Leibler divergence as our regularizer, we present a novel method for both deconvolution and kernel (point
spread function) estimation via the expectation of the probability density with maximum entropy. This higher-level approach is known in information theory as maximum entropy on the mean and dates back to E.T. Jaynes in 1957 [10, 11]. Our approach is made computationally tractable as a result of two observations:

(i) Fenchel-Rockafellar duality transforms our infinite-dimensional primal problem into a finite-dimensional dual problem;

(ii) the sought expectation of the maximal probability distribution can be simply written in terms of known moment generating functions and the optimizer of the dual problem.

What is particularly remarkable about our higher-level method is that it effectively restores images that have been subjected to significantly greater levels of blurring than previously considered in the literature. While the method is stable with respect to small amounts of noise, it does not actively denoise; however, for moderate to large amounts of noise, it can readily be preconditioned by first applying expected patch log likelihood (EPLL) denoising [33].

We test and compare our method on a variety of examples (cf. Section 5). To start, we present an example of simple deconvolution (without additive noise) but for a range of blurs for which previous methods do not even converge (cf. Figure 1). Then we consider deconvolution with small to significant additive noise. We show that we can precondition with EPLL denoising to attain deconvolution results comparable with the state of the art (cf. Figure 3). We then address blind deblurring with the inclusion of a known shape (analogous to a finder pattern in a QR barcode [22]). In these cases, we can, preconditioning with EPLL denoising, blindly deblur with large blurs (cf. Figures 4, 5, 6). Given that our method relies on symbology, comparison with other methods is unfair (in our favour). However, we do provide comparison with the state of the art method of Pan et al. [19, 18] to demonstrate the power of our method in exploiting the symbology (finder pattern) to accurately recover the blur (point spread function).

Overall, we introduce a novel regularization methodology which is theoretically well-founded, numerically tractable, and amenable to substantial generalization. While we have directly motivated and applied our higher-level regularization approach to image deblurring, we anticipate that it will also prove useful in solving other ill-posed inverse problems in computer vision, pattern recognition, and machine learning.

Let us first mention current methods, the majority of which are based upon some notion of regularization at the level of the set of images. We then present a paradigm shift by optimizing at the level of the set of probability densities on the set of images.

1.1. Current Methods

The process of capturing one channel of a blurred image \( b \in \mathbb{R}^{n \times m} \) from a ground truth channel \( x \in \mathbb{R}^{n \times m} \) is modelled throughout by the relation \( b = c \ast x \), where \( \ast \) denotes the 2-dimensional convolution between the kernel \( c \in \mathbb{R}^{k \times k} (k < n, m) \) and the ground
truth; this model represents spatially invariant blurring. For images composed of more than one channel, blurring is assumed to act on a per-channel basis. We, therefore, derive a method to deblur one channel and apply it to each channel separately.

Current blind deblurring methods consist of solving

$$
\inf_{x \in \mathbb{R}^{n \times m}, c \in \mathbb{R}^{k \times k}} \left\{ R(x, c) + \frac{\alpha}{2} ||c * x - b||^2_2 \right\},
$$

where $R : \mathbb{R}^{n \times m} \times \mathbb{R}^{k \times k} \to \mathbb{R}$ serves as a regularizer which permits the imposition of certain constraints on the optimizers. This idea of regularization to solve ill-posed inverse problems dates back to Tikhonov [30]. Approaches that are not based on machine learning differ mostly in the choice of regularizer, examples include $L_0$-regularization, which penalizes the presence of non-zero pixels in the image or gradient [18]; weighted nuclear norm regularization, which ensures that the image or gradient matrices have low rank [21], and $L_0$-regularization of the dark channel, which promotes sparsity of a channel consisting of local minima in the intensity channel [19]. As it pertains to machine learning methods, other approaches have been employed including modelling the optimization problem as a deep neural network [28] and estimating the ground truth image from a blurred input without estimating the kernel using convolutional neural networks [16, 17, 29] or generative adversarial networks [14, 20].

The results achieved in these papers are comparable to the state of the art. However, to our knowledge, such methods have not been successfully applied to the large blurring regimes considered in this paper.

2. Preliminaries

We begin by recalling some standard definitions and establishing notation. We refer to [32] for convex analysis in infinite dimensions and [23] for the finite-dimensional setting. We follow [27] as a standard reference for real analysis.

Letting $X$ be a separated locally convex space, we denote by $X^*$ its topological dual. The duality pairing between $X$ and its dual will be written as $(\cdot, \cdot) : X \times X^* \to \mathbb{R}$ in order to distinguish it from the canonical inner product on $\mathbb{R}^d$, $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. For $f : X \to \mathbb{R} \equiv \mathbb{R} \cup \{-\infty, +\infty\}$, an extended real-valued function on $X$, the (Fenchel) conjugate of $f$ is $f^* : X^* \to \mathbb{R}$ defined by

$$
f^*(x^*) = \sup_{x \in X} \{(x, x^*) - f(x)\},
$$

using the convention $a - (-\infty) = +\infty$ and $a - (+\infty) = +\infty$ for every $a \in \mathbb{R}$. The subdifferential of $f$ at $\bar{x} \in X$ is the set

$$
\partial f(\bar{x}) = \{x^* \in X^* | (x - \bar{x}, x^*) \leq f(x) - f(\bar{x}) \forall x \in X\}.
$$

We define $\text{dom } f := \{x \in X | f(x) < +\infty\}$, the domain of $f$, and say that $f$ is proper if $\text{dom } f \neq \emptyset$ and $f(x) > -\infty$ for every $x \in X$. $f$ is said to be lower semicontinuous if $f^{-1}([-\infty, \alpha])$ is closed for every $\alpha \in \mathbb{R}$. 
A proper function $f$ is convex provided for every $x, y \in \text{dom } f$ and $\lambda \in (0, 1)$,
\[
f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y),
\]
if the above inequality is strict whenever $x \neq y$, $f$ is said to be strictly convex. If $f$ is proper and for every $x, y \in \text{dom } f$ and $\lambda \in (0, 1)$,
\[
f(\lambda x + (1 - \lambda)y) + \lambda(1 - \lambda)\frac{c}{2}||x - y||^2 \leq \lambda f(x) + (1 - \lambda)f(y),
\]
then $f$ is called $c$-strongly convex.

For any set $A \subseteq X$, the indicator function of $A$ is given by
\[
\delta_A : X \rightarrow \mathbb{R} \cup \{+\infty\}, \quad x \mapsto \begin{cases} 0, & x \in A, \\ +\infty, & \text{otherwise.} \end{cases}
\]

For any $\Omega \subseteq \mathbb{R}^d$, we denote by $\mathcal{P}(\Omega)$ the set of probability measures on $\Omega$. Let $\eta$ be a signed Borel measure on $\Omega$, we define the total variation of $\eta$ by
\[
|\eta|(\Omega) = \sup \left\{ \sum_{i \in \mathbb{N}} |\eta(\Omega_i)| : \bigcup_{i \in \mathbb{N}} \Omega_i = \Omega, \Omega_i \cap \Omega_j = \emptyset \ (i \neq j) \right\}.
\]

The set of all signed Borel measures with finite total variation on $\Omega$ will be denoted by $\mathcal{M}(\Omega)$. We say that a measure is $\sigma$-finite (on $\Omega$) if $\Omega = \bigcup_{i \in \mathbb{N}} \Omega_i$ with $|\mu(\Omega_i)| < +\infty$.

Let $\mu$ be a positive $\sigma$-finite Borel measure on $\Omega$ and $\rho$ be an arbitrary Borel measure on $\Omega$, we write $\rho \ll \mu$ to signify that $\rho$ is absolutely continuous with respect to $\mu$, i.e. if $A \subseteq \Omega$ is such that $\mu(A) = 0$, then $\rho(A) = 0$. If $\rho \ll \mu$ there exists a unique function $\frac{d\rho}{d\mu} \in L^1(\mu)$ for which
\[
\rho(A) = \int_A \frac{d\rho}{d\mu} \, d\mu, \quad \forall \ A \subseteq \Omega \text{ measurable.}
\]
The function $\frac{d\rho}{d\mu}$ is known as the Radon-Nikodym derivative (cf. [27, Thm. 6.10]).

The Kullback-Leibler divergence between $\rho, \mu \in \mathcal{M}(\Omega)$ is the functional
\[
\mathcal{K}(\rho, \mu) = \begin{cases} \int_\Omega \log \left( \frac{d\rho}{d\mu} \right) \, d\rho, & \rho, \mu \in \mathcal{P}(\Omega), \rho \ll \mu, \\ +\infty, & \text{otherwise.} \end{cases}
\]

For $\Omega \subseteq \mathbb{R}^d$, $\eta \in \mathcal{M}(\Omega)$ we denote, by a slight abuse of notation, $E_{\eta}[X]$ to be a vector whose $k$th component is $(E_{\eta}[X])_k = \int_\Omega x_k \, d\eta(x)$. Thus, $E_{\eta}[X]$ is a map from $\mathcal{M}(\Omega)$ to $\mathbb{R}^d$ whose restriction to $\mathcal{P}(\Omega)$ is known as the expectation of the random vector $X = [X_1, \ldots, X_d]$ associated with the input measure.

Finally, the smallest (resp. largest) singular value $\sigma_{\min}(C)$ (resp. $\sigma_{\max}(C)$) of the matrix $C \in \mathbb{R}^{m \times n}$ is the square root of the smallest (resp. largest) eigenvalue of $C^T C$. 
3. The MEM Method

3.1. Kullback-Leibler Regularized Deconvolution and the Maximum Entropy on the Mean Framework

**Notation:** We first establish some notation pertaining to deconvolution. The convolution operator \( c \ast \) will be denoted by the matrix \( C : \mathbb{R}^d \rightarrow \mathbb{R}^d \) acting on a vectorized image \( x \in \mathbb{R}^d \) for \( d = nm \) and resulting in a vectorized blurred image for which the \( k \)th coordinate in \( \mathbb{R}^d \) corresponds to the \( k \)th pixel of the image. We assume throughout that the matrix \( C \) is nonsingular, as is standard in image deblurring.

We recall that traditional deconvolution software functions by solving (1) with a fixed convolution kernel \( c \). Our approach differs from previous work by adopting the maximum entropy on the mean framework which posits that the state best describing a system is given by the mean of the probability distribution which maximizes some measure of entropy [10, 11]. As such, taking \( \Omega \subseteq \mathbb{R}^d \) to be compact, \( \mu \in \mathcal{P}(\Omega) \) to be a prior measure and \( b \in \mathbb{R}^d \) to be a blurred image, our approach is to determine the solution of

\[
\inf_{\rho \in \mathcal{P}(\Omega)} \left\{ K(\rho, \mu) + \frac{\alpha}{2} \| b - C\mathbb{E}_\rho[X] \|_2^2 \right\} = \inf_{\mathcal{P}(\Omega)} \{ f + g \circ A \},
\]

for

\[
f = K(\cdot, \mu), \quad g = \frac{\alpha}{2} \| b + (\cdot) \|_2^2, \quad A = -C\mathbb{E}(\cdot)[X].
\]

The following lemma establishes some basic properties of \( f \).

**Lemma 1.** The functional \( f : \mathcal{M}(\Omega) \rightarrow \mathbb{R} \) is proper, lower semicontinuous and strictly convex.

**Proof.** We begin with strict convexity of \( f \). Let \( x \in \Omega \) and \( t \in (0, 1) \) be arbitrary moreover let \( \rho_1 \neq \rho_2 \) be elements of \( \mathcal{P}(\Omega) \) and \( \rho_t = t \rho_1 + (1 - t) \rho_2 \). We have

\[
\log \left( \frac{d\rho_t(x)}{d\mu(x)} \right) \frac{d\rho_t}{d\mu}(x) = \log \left( \frac{t \frac{d\rho_1}{d\mu}(x) + (1 - t) \frac{d\rho_2}{d\mu}(x)}{t + (1 - t)} \right) \left( t \frac{d\rho_1}{d\mu}(x) + (1 - t) \frac{d\rho_2}{d\mu}(x) \right) \\
\leq t \log \left( \frac{d\rho_1}{d\mu}(x) \right) \frac{d\rho_1}{d\mu}(x) + (1 - t) \log \left( \frac{d\rho_2}{d\mu}(x) \right) \frac{d\rho_2}{d\mu}(x).
\]

The inequality is due to the log-sum inequality [5, Thm. 2.7.1], and since \( \rho_1 \neq \rho_2 \), \( \frac{d\rho_1}{d\mu} \) and \( \frac{d\rho_2}{d\mu} \) differ on a set \( E \subseteq \Omega \) such that \( \mu(E) > 0 \). The strict log-sum inequality therefore implies that the inequality is strict for every \( x \in E \). Since integration preserves strict inequalities,

\[
f(\rho_t) = \int_{\Omega \setminus E} \log \left( \frac{d\rho_t}{d\mu} \right) \frac{d\rho_t}{d\mu} d\mu + \int_E \log \left( \frac{d\rho_t}{d\mu} \right) \frac{d\rho_t}{d\mu} d\mu < t f(\rho_1) + (1 - t) f(\rho_2)
\]

so \( f \) is, indeed, strictly convex.
It is well known that \( f \) is lower semicontinuous and proper (cf. [6, Thm. 3.2.17]). Finally, for convex functions the notions of lower semicontinuity and weak lower semicontinuity are equivalent [32, Thm. 2.2.1].

Problem (3) is an infinite-dimensional optimization problem with no obvious solution and is thus intractable in its current form. However, existence and uniqueness of solutions thereof is established in the following remark.

**Remark 1.** First, the objective function in (3) is proper, strictly convex and lower semicontinuous since \( f \) is proper, strictly convex and lower semicontinuous whereas \( g \circ A \) is proper, continuous and convex.

Note that \( \mathcal{P}(\Omega) \) is weak* compact by the Riesz representation theorem [27, Thm. 6.14]. Since \( \Omega \) is compact, the notions of weak and weak* convergence of measures coincide, so [1, Thm. 3.2.1] establishes existence of solutions whereas strict convexity guarantees uniqueness.

Even with this theoretical guarantee, direct computation of solutions to (3) remains infeasible. In the sequel, a corresponding finite-dimensional dual problem will be established which will, along with a method to recover the expectation of solutions of (3) from solutions of this dual problem, permit an efficient and accurate estimation of the original image.

### 3.2. Dual Problem

In order to derive the (Fenchel-Rockafellar) dual problem to (3) we provide the reader with the Fenchel-Rockafellar duality theorem in a form expedient for our study, cf. e.g. [32, Cor. 2.8.5].

**Theorem 1** (Fenchel-Rockafellar Duality Theorem). Let \( X \) and \( Y \) be locally convex spaces and let \( X^* \) and \( Y^* \) denote their dual spaces. Moreover, let \( f : X \to \mathbb{R} \cup \{+\infty\} \) and \( g : Y \to \mathbb{R} \cup \{+\infty\} \) be convex, lower semicontinuous and proper functions and let \( A \) be a bounded linear operator from \( X \) to \( Y \). Assume that there exists \( \bar{y} \in A \operatorname{dom} f \cap \operatorname{dom} g \) such that \( g \) is continuous at \( \bar{y} \). Then

\[
\inf_{x \in X} \{ f(x) + g(-Ax) \} = \max_{y^* \in Y^*} \{ -f^*(A^*y^*) - g^*(y^*) \}
\]  

with \( A^* \) denoting the adjoint of \( A \). Moreover, \( \bar{x} \) is optimal in the primal problem if and only if there exists \( \bar{y}^* \in Y^* \) satisfying \( A^*\bar{y}^* \in \partial f(\bar{x}) \) and \( \bar{y}^* \in \partial g(-A\bar{x}) \).

In (5), the minimization problem is referred to as the primal problem, whereas the maximization problem is called the dual problem. Under certain conditions, a solution to the primal problem can be obtained from a solution to the dual problem.

**Remark 2** (Primal-Dual Recovery). In the context of Theorem 1, \( f^* \) and \( g^* \) are proper, lower semicontinuous and convex, also \( (f^*)^* = f \) and \( (g^*)^* = g \) [32, Thm. 2.3.3]. Suppose additionally that \( 0 \in \operatorname{int}(A^* \operatorname{dom} g^* - \operatorname{dom} f^*) \).
Let $\bar{y}^* \in \text{argmax}_Y \{-f^* \circ A^* - g^*\}$. By the first order optimality conditions, \[\text{Thm. 2.5.7}\]
\[
0 \in \partial (f^* \circ A^* + g) (\bar{y}^*) = A\partial f^*(A^*\bar{y}^*) + \partial g(\bar{y}^*),
\]
the second expression is due to \[\text{Thm. 2.168}\] (the conditions to apply this theorem are satisfied by assumption). Consequently, there exists $\bar{z} \in \partial g(A^*\bar{y}^*)$ for which $\bar{z} = -A\bar{x}$. Since $f$ and $g$ are proper, lower semicontinuous and convex we have \[\text{Thm. 2.4.2 (iii)}\]:
\[
A^*\bar{y}^* \in \partial f(\bar{x}), \quad \bar{y}^* \in \partial g(\bar{z}) = \partial g(-A\bar{x}).
\]
Thus Theorem 1 demonstrates that $\bar{x}$ is a solution of the primal problem, that is if $\bar{y}^*$ is a solution of the dual problem, $\partial f^*(A^*\bar{y}^*)$ contains a solution to the primal problem.

If, additionally, $f^*(A^*\bar{y}^*) < +\infty$ \[\text{Prop. 2.118}\] implies that, \[\text{cor}\]
\[
\bar{x} \in \partial f^*(A^*\bar{y}^*) = \text{argmax}_{x \in X} \{(x, A^*\bar{y}^*) - f(x)\}.
\]
We refer to \[\text{6}\] as the primal-dual recovery formula.

A particularly useful case of this theorem is when $A$ is an operator between an infinite-dimensional locally convex space $X$ and $\mathbb{R}^d$, as the dual problem will be a finite-dimensional maximization problem. Moreover, the primal-dual recovery is easy if $f^*$ is Gâteaux differentiable at $A^*\bar{y}^*$, in which case the subdifferential and the derivative coincide at this point \[\text{32 Cor. 2.4.10}\], so \[\text{6}\] reads $\bar{x} = \nabla f^*(A^*\bar{y}^*)$. Some remarks are in order to justify the use of this theorem.

**Remark 3.** It is clear that $\mathcal{P}(\Omega)$ is not a locally convex space, however it is a subset of the Banach space $\mathcal{M}(\Omega)$ normed by total variation $||\cdot||_{TV(\Omega)} = |\cdot|_\Omega$. The Riesz representation theorem \[\text{27 Thm. 2.14}\] identifies $(\mathcal{M}(\Omega), ||\cdot||_{TV(\Omega)})$ as the dual space of $(\mathcal{C}_0(\Omega), ||\cdot||_{\infty, \Omega})$ (continuous functions vanishing at infinity with the supremum norm) with duality pairing $(\cdot, \cdot) : \mathcal{C}_0(\Omega) \times \mathcal{M}(\Omega) \rightarrow \mathbb{R}$, $(\psi, \eta) \mapsto \int_\Omega \psi \eta$. However, in light of \[\text{2}\], $\text{dom} \mathcal{K}(\cdot, \mu) \subseteq \mathcal{P}(\Omega)$ so the inf in \[\text{3}\] can be taken over $\mathcal{M}(\Omega)$ or $\mathcal{P}(\Omega)$ interchangeably.

In the following we verify that $A$ is a bounded linear operator and compute its adjoint.

**Lemma 2.** The operator $A : \mathcal{M}(\Omega) \rightarrow \mathbb{R}^d$ in \[\text{4}\] is linear and bounded. Moreover, its adjoint is the mapping $z \in \mathbb{R}^d \mapsto \langle C^T z, \cdot \rangle$.

**Proof.** We begin by establishing boundedness of $E_{(\cdot)}[X] : \mathcal{M}(\Omega) \rightarrow \mathbb{R}^d$. Let $\rho \in \mathcal{M}(\Omega)$ be arbitrary, then
\[
||E_\rho[X]||_2^2 = \sum_{i=1}^d \left(\int_\Omega x_i \rho(x)\right)^2 \leq \sum_{i=1}^d \left(||x_i||_{\infty, \Omega} \rho(\Omega)\right)^2 \leq \kappa ||\rho||_{TV(\Omega)}^2.
\]
for $\kappa = \sum_{i=1}^{d} ||x_i||_{\infty,\Omega}^2$. The second inequality is due to the fact that $|\rho(\Omega)| \leq ||\rho||_{TV(\Omega)}$. Consequently, $A$ is bounded as

$$||A\rho||_2 = ||-C \mathbb{E}_\rho[X]||_2 \leq \sigma_{\text{max}}(C) ||\mathbb{E}_\rho[X]||_2 \leq \sigma_{\text{max}}(C) \sqrt{\kappa} ||\rho||_{TV}.$$  

(7)

Pertaining to linearity, let $\pi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ denote the projection onto the $i$-th coordinate (i.e. $\pi_i((x_1, \ldots, x_d)) = x_i$). Linearity of $\mathbb{E}[^{\cdot}|X]$ is clear since for $\rho, \eta \in \mathcal{M}(\Omega)$,

$$\mathbb{E}_\rho + \eta[X] = (\pi_1, \rho + \eta), \ldots, (\pi_d, \rho + \eta))$$

$$= \mathbb{E}_\rho[X] + \mathbb{E}_\eta[X],$$

so, $\langle C \mathbb{E}_\rho[X], z \rangle = \langle \mathbb{E}_\rho[X], C^T z \rangle = (\langle C^T z, \cdot \rangle, \rho)$, yielding $A^*(z) = (C^T z, \cdot)$.

We now compute the conjugates of $f$ and $g$, respectively and provide an explicit form for the dual problem of (3).

**Lemma 3.** The conjugate of $f$ in (4) is $f^* : \phi \in C_0(\Omega) \mapsto \log \left( \int_{\Omega} \exp(\phi) d\mu \right)$. In particular, $f^*$ is finite-valued. Moreover, for any $\phi \in C_0(\Omega)$, $\arg\max_{\mathcal{P}(\Omega)} \{(\phi, \cdot) - \mathcal{K}(\cdot, \mu)\} = \{\tilde{\rho}_\phi\}$, the unique probability measure on $\Omega$ for which

$$\frac{d\tilde{\rho}_\phi}{d\mu} = \frac{\exp \phi}{\int_{\Omega} \exp \phi d\mu}.$$  

(8)

**Proof.** We proceed by direct computation:

$$f^*(\phi) = \sup_{\rho \in \mathcal{M}(\Omega)} \{(\phi, \rho) - \mathcal{K}(\rho, \mu)\}$$

$$= \sup_{\rho \in \mathcal{P}(\Omega)} \{(\phi, \rho) - \mathcal{K}(\rho, \mu)\}$$

$$= \sup_{\rho \in \mathcal{P}(\Omega)} \left\{ \int_{\Omega} \log \left( \frac{\exp \phi}{\frac{d\rho}{d\mu}} \right) d\rho \right\},$$

where we have used the fact that dom $f \subseteq \mathcal{P}(\Omega)$ as noted in Remark 3. Note that $\exp \phi \in C_0(\Omega) \subseteq L^1(\rho)$ and since $t \mapsto \log t$ is concave, Jensen’s inequality [27, Thm. 3.3] yields

$$f^*(\phi) \leq \sup_{\rho \in \mathcal{P}(\Omega)} \left\{ \log \left( \int_{\Omega} \exp \phi \frac{d\rho}{d\mu} d\rho \right) \right\} = \log \left( \int_{\Omega} \exp \phi d\mu \right)$$  

(9)
Letting $\bar{\rho}_\phi$ be the measure with Radon-Nikodym derivative

$$\frac{d\bar{\rho}_\phi}{d\mu} = \frac{\exp \phi}{\int_\Omega \exp \phi \, d\mu},$$

one has that

$$(\phi, \bar{\rho}_\phi) - \mathcal{K}(\bar{\rho}_\phi, \mu) = (\phi, \bar{\rho}_\phi) - \int_\Omega \log \left( \frac{\exp \phi}{\int_\Omega \exp \phi \, d\mu} \right) \, d\bar{\rho}_\phi = \log \left( \int_\Omega \exp \phi \, d\mu \right),$$

so $\bar{\rho}_\phi \in \text{argmax}_{\mathcal{P}(\Omega)} \{(\phi, \cdot) - \mathcal{K}(\cdot, \mu)\}$ as $\bar{\rho}_\phi$ saturates the upper bound for $f^*(\phi)$ established in (9), thus $f^*(\phi) = \log \left( \int_\Omega \exp \phi \, d\mu \right)$. Moreover $\bar{\rho}_\phi$ is the unique maximizer since the objective is strictly concave.

With this expression in hand, we show that $\text{dom} f^* = C_0(\Omega)$. To this effect, let $\phi \in C_0(\Omega)$ be arbitrary and note that,

$$\exp (\phi(x)) \leq \exp \left( \max_\Omega \phi \right), \quad (x \in \Omega).$$

Thus,

$$f^*(\phi) = \log \left( \int_\Omega \exp \phi \, d\mu \right) \leq \log \left( \exp \left( \max_\Omega \phi \right) \right) = \max_\Omega \phi < +\infty,$$

since $C_0(\Omega) = C_b(\Omega)$. Since $\phi$ is arbitrary, $\text{dom} f^* = C_0(\Omega)$ and, coupled with the fact that $f^*$ is proper [32, Thm. 2.3.3], we obtain that $f^*$ is finite-valued.

**Lemma 4.** The conjugate of $g$ from (4) is $g^* : z \in \mathbb{R}^d \mapsto \frac{1}{2\alpha} ||z||_2^2 - \langle b, z \rangle$.

**Proof.** The assertion follows from the fact that $\frac{1}{2} ||\cdot||_2^2$ is self-conjugate [24, Ex. 11.11] and some standard properties of conjugacy [24, Eqn. 11(3)].

Combining these results we obtain the main duality theorem.

**Theorem 2.** The (Fenchel-Rockafellar) dual of (3) is given by

$$\max_{\lambda \in \mathbb{R}^d} \left\{ \langle b, \lambda \rangle - \frac{1}{2\alpha} ||\lambda||_2^2 - \log \left( \int_\Omega \exp \langle C^T \lambda, x \rangle \, d\mu \right) \right\}. \quad (10)$$

Given a maximizer $\bar{\lambda}$ of (10) one can recover a minimizer of (3) via

$$d\bar{\rho} = \frac{\exp \langle C^T \bar{\lambda}, \cdot \rangle}{\int_\Omega \exp \langle C^T \bar{\lambda}, \cdot \rangle \, d\mu} \, d\mu. \quad (11)$$

**Proof.** The dual problem can be obtained by applying the Fenchel-Rockafellar duality theorem (Theorem 1), with $f$ and $g$ defined in (4), to the primal problem

$$\inf_{\rho \in \mathcal{M}(\Omega)} \left\{ \mathcal{K}(\rho, \mu) + \frac{\alpha}{2} ||b - CE_\rho[X]||_2^2 \right\},$$

and substituting the expressions obtained in Lemmas 2, 3 and 4. All relevant conditions to apply this theorem have either been verified previously or are clearly satisfied.
Note that $0 \subseteq \text{dom} \ g^* = \mathbb{R}^d$ and $A^*0 = 0 \in C_0(\Omega)$, so

$$A^*(\text{dom} \ g^*) \setminus \text{dom} \ f^* \supseteq \{ \phi | \phi \in \text{dom} \ f^* \} = C_0(\Omega),$$

since $\text{dom} \ f^* = C_0(\Omega)$ by Lemma 3. Thus $0 \in \text{int} (A^* \text{dom} \ g^* - \text{dom} \ f^*) = C_0(\Omega)$, and Remark 2 is applicable. The primal-dual recovery formula (6) is given explicit form by Lemma 3 by evaluating $d\bar{\rho}(C^T\lambda)$. 

The utility of the dual problem is that it permits a staggering dimensionality reduction, passing from an infinite-dimensional problem to a finite-dimensional one. Moreover, the form of the dual problem makes precise the role of $\alpha$ in (3). Notably in [3, Cor. 4.9] the problem

$$\inf_{\rho \in \mathcal{P}(\Omega) \cap \text{dom} K(\cdot,\mu)} K(\rho, \mu) \quad \text{s.t.} \quad ||C E_{\rho}[X] - b||_2^2 \leq \frac{1}{2\alpha}$$

(12)

is paired in duality with (10). Thus the choice of $\alpha$ is directly related to the fidelity of $C E_{\rho}[X]$ to the blurred image. The following section is devoted to the choice of a prior and describing a method to directly compute $E_{\rho}[X]$ from a solution of (10).

### 3.3. Probabilistic Interpretation of Dual Problem

If no information is known about the original image, the prior $\mu$ is used to impose box constraints on the optimizer such that its expectation will be in the interval $[0, 1]^d$ and will only assign non-zero probability to measurable subsets of this interval. With this consideration in mind, the prior distribution should be the distribution of the random vector $X = [X_1, X_2, \ldots]$ with the $X_i$ denoting independent random variables with uniform distributions on the interval $[u_i, v_i]$. If the $k$th pixel of the original image is unknown, we let $[u_k, v_k] = [0 - \epsilon, 1 + \epsilon]$ for $\epsilon > 0$ small in order to provide a buffer for numerical errors.

However, if the $k$th pixel of the ground truth image was known to have a value of $\ell$, one can enforce this constraint by taking the random variable $X_k$ to be distributed uniformly on $[\ell - \epsilon, \ell + \epsilon]$. Constructing $\mu$ in this fashion guarantees that its support (and hence $\Omega$) is compact.

To deal with the integrals in (10) and (11) it is convenient to note that (cf. [25, Sec. 4.4])

$$\int_{\Omega} \exp\left(\langle C^T \lambda, x \rangle\right) d\mu = \mathbb{M}_X[C^T \lambda],$$

the moment-generating function of $X$ evaluated at $C^T \lambda$. Since the $X_i$ are independently distributed, $\mathbb{M}_X[t] = \prod_{i=1}^d \mathbb{M}_{X_i}[t]$ [25, Sec. 4.4], and since the $X_i$ are uniformly distributed on $[u_i, v_i]$ one has

$$\mathbb{M}_X[t] = \prod_{i=1}^d \frac{e^{t v_i} - e^{t u_i}}{t_i (v_i - u_i)},$$
and therefore the dual problem (10) with this choice of prior can be written as

$$\max_{\lambda \in \mathbb{R}^d} \left\{ \langle b, \lambda \rangle - \frac{1}{2\alpha} ||\lambda||_2^2 - \sum_{i=1}^{d} \log \left( \frac{e^{C_i^T \lambda v_i} - e^{C_i^T \lambda u_i}}{C_i^T \lambda (v_i - u_i)} \right) \right\}. \quad (13)$$

A solution of (13) can be determined using a number of standard numerical solvers. We opted for the implementation [4] of the L-BFGS algorithm due to its speed and efficiency.

Since only the expectation of the optimal probability measure for (3) is of interest, we compute the $i^{th}$ component of the expectation $\mathbb{E}_\rho[X]_i$ of the optimizer provided by the primal-dual recovery formula (11) via

$$\int_{\Omega} x_i e^{(C^T \lambda, x)} d\mu \int_{\Omega} e^{(C^T \lambda, x)} d\mu = \partial_i \log \left( \int_{\Omega} e^{(t, x)} d\mu \right) \bigg|_{t=C^T \lambda}.$$

Using the independence assumption on the prior, we obtain

$$\mathbb{E}_\rho[X] = \nabla_t \sum_{i=1}^{d} \log (M_{X_i}[t])$$

such that the best estimate of the ground truth image is given by

$$(\mathbb{E}_\rho[X])_i = \frac{v_i e^{C_i^T \lambda v_i} - u_i e^{C_i^T \lambda u_i}}{e^{C_i^T \lambda v_i} - e^{C_i^T \lambda u_i}} - \frac{1}{C_i^T \lambda}. \quad (14)$$

With (13) and (14) in hand, our entropic method for deconvolution can be implemented.

3.4. Exploiting Symbology for Blind Deblurring

In order to implement blind deblurring on images that incorporate a symbology, one must first estimate the convolution kernel responsible for blurring the image. This step can be performed by analyzing the blurred symbolic constraints. We propose a method that is based on the entropic regularization framework discussed in the previous sections.

In order to perform this kernel estimation step, we will use the same framework as (3) with $x$ taking the role of $c$. In the assumption that the kernel is of size $k \times k$, we take $\Omega = [0 - \epsilon, 1 + \epsilon]^{k^2}$ for $\epsilon > 0$ small (again to account for numerical error) and consider the problem

$$\inf_{\eta \in \mathcal{P}^2(\Omega)} \left\{ \frac{\gamma}{2} \left\| \mathbb{E}_\eta[X] \ast \tilde{x} - \tilde{b} \right\|_2^2 + \mathcal{K}(\eta, \nu) \right\}. \quad (15)$$

Here, $\gamma > 0$ is a parameter that enforces fidelity. $\tilde{x}$ and $\tilde{b}$ are the segments of the original and blurred image which are known to be fixed by the symbolic constraints. That is, $\tilde{x}$ consists solely of the embedded symbology and $\tilde{b}$ is the blurry symbology. By analogy with (3), the expectation of the optimizer of (15) is taken to be the estimated kernel. The role of $\nu \in \mathcal{P}(\Omega)$ is to enforce the fact that the kernel should be normalized and
non-negative (hence its components should be elements of $[0, 1]$). Hence we take its distribution to be the product of $k^2$ uniform distributions on $[0 - \epsilon, 1 + \epsilon]$. As in the non-blind deblurring step, the expectation of the optimizer of (15) can be determined by passing to the dual problem (which is of the same form as (13)), solving the dual problem numerically and using the primal-dual recovery formula (14). A summary of the blind deblurring algorithm is compiled in Algorithm 1. We would like to point out that the algorithm is not iterative, rather only one kernel estimate step and one deconvolution step are used.

This method can be further refined to compare only the pixels of the symbology which are not convolved with pixels of the image which are unknown. By choosing these specific pixels, one can greatly improve the quality of the kernel estimate, as every pixel that was blurred to form the signal is known; however, this refinement limits the size of convolution kernel which can be estimated.

Algorithm 1 Entropic Blind Deblurring

**Require:** Blurred image $b$, prior $\mu$, kernel width $k$, fidelity parameters $\gamma, \alpha$;

**Ensure:** Deblurred image $\bar{x}$

- $\nu \leftarrow$ density of $k^2$ uniformly distributed independent random variables
- $\lambda_\epsilon \leftarrow \text{argmax of analog of (13) for kernel estimate.}$
- $\bar{c} \leftarrow$ expectation of argmin of (15) computed via analog of (14) for kernel estimate evaluated at $\lambda_\epsilon$
- $\lambda_x \leftarrow \text{argmax of (13)}$
- $\bar{\bar{x}} \leftarrow$ expectation of argmin of (3) with kernel $\bar{c}$ computed via (14) evaluated at $\lambda_x$

**Return** $\bar{x}$

4. Stability Analysis for Deconvolution

In contrast to, say, total variation methods, our maximum entropy method does not actively denoise. However, its ability to perform well with a denoising preprocessing step highlights that is “stable” to small perturbations in the data. In this section, we show that our convex analysis framework readily allows us to prove the following explicit stability estimate.

**Theorem 3.** Let $b_1, b_2 \in \mathbb{R}^d$ be images obtained by convolving the ground truth images $x_1, x_2$ with the same kernel $c$. Let

$$
\rho_i = \arg\min_{\rho \in P(\Omega)} \left\{ \mathcal{K}(\rho, \mu) + \frac{\alpha}{2} ||C\mathbb{E}_\rho[X] - b_i|| \right\} \quad (i = 1, 2),
$$

then

$$
||\mathbb{E}_{\rho_1}[X] - \mathbb{E}_{\rho_2}[X]||_2 \leq \frac{2}{\sigma_{\min}\left(C\right)} ||b_1 - b_2||_2.
$$

Where $\sigma_{\min}(C)$ is the smallest singular value of $C$. 

The proof will follow from a sequence of lemmas. To this end we consider the
optimal value function for (3), which we denote \( v : \mathbb{R}^d \to \mathbb{R} \), as
\[
v(b) := \inf_{\rho \in P(\Omega)} \left\{ K(\rho, \mu) + \frac{\alpha}{2} \| C\mathbb{E}_\rho[X] - b \|^2 \right\} = \inf_{\rho \in P(\Omega)} \left\{ k(\rho, b) + h \circ L(\rho, b) \right\},
\]
where
\[
k : (\rho, b) \in \mathcal{M}(\Omega) \times \mathbb{R}^d \mapsto K(\rho, \mu), \quad h = \frac{\alpha}{2} \| \cdot \|^2, \quad L(\rho, b) = C\mathbb{E}_\rho[X] - b.
\]

The following results will allow us to conclude that \( \nabla v \) is (globally) \( \alpha \)-Lipschitz.

**Lemma 5.** The operator \( L \) in (17) is bounded and linear, its adjoint is the map
\( z \mapsto (\langle C^Tz, \cdot \rangle, -z) \in C_0(\Omega) \times \mathbb{R}^d \).

**Proof.** Linearity of this operator follows from the linearity of the expectation operator (cf. Lemma 2). As it pertains to boundedness, observe that
\[
\| C\mathbb{E}_\rho[X] - b \|_2 \leq \| C\mathbb{E}_\rho[X] \|_2 + \| b \|_2 \\
\leq \sigma_{\max}(C) \| \mathbb{E}_\rho[X] \|_2 + \| b \|_2 \\
\leq \kappa_1 (\| \rho \|_{TV} + \| b \|_2),
\]
where \( \kappa_1 = \max \{ 1, \sigma_{\max}(C)\sqrt{\kappa} \} \) with \( \kappa \) defined as in (7). The adjoint is obtained as in Lemma 2,
\[
\langle C\mathbb{E}_\rho[X] - b, z \rangle = \langle \langle C^Tz, \cdot \rangle, \rho \rangle + \langle b, -z \rangle.
\]

Next, we compute the conjugate of \( k + h \circ L \).

**Lemma 6.** The conjugate of \( k + h \circ L \) defined in (17) is the function
\[
(\phi, y) \in C_0(\Omega) \times \mathbb{R}^d \mapsto (K(\cdot, \mu))^*(\phi + \langle C^Ty, \cdot \rangle) + \frac{1}{2\alpha} \| y \|^2,
\]
where \( (K(\cdot, \mu))^* \) is the conjugate computed in Lemma 3.

**Proof.** Since dom \( h = \mathbb{R}^n \), \( h \) is continuous and \( k \) is proper, there exists \( x \in L \text{dom } k \cap \text{dom } h \) such that \( h \) is continuous at \( x \). The previous condition guarantees that, \( \{ 32 \text{ Thm. 2.8.3} \}
\[
(k + h \circ L)^*(\phi, y) = \min_{z \in \mathbb{R}^d} \{ k^*((\phi, y) - L^*(z)) + h^*(z) \}.
\]
The conjugate of \( k \) is given by
\[
k^*(\phi, y) = \sup_{\rho \in \mathcal{M}(\Omega)} \{ (\phi, \rho) + \langle y, b \rangle - K(\rho, \mu) \}.
\]
For \( y \neq 0 \), \( \sup_{\mathbb{R}^d} \langle y, \cdot \rangle = +\infty \). Thus,
\[
k^*(\phi, y) = \sup_{\rho \in \mathcal{M}(\Omega)} \{ (\phi, \rho) - K(\rho, \mu) \} + \delta_{\{0\}}(y) = (K(\cdot, \mu))^*(\phi) + \delta_{\{0\}}(y).
\]
The conjugate of $h$ was established in Lemma 4 and the adjoint of $L$ is given in Lemma 5. Substituting these expressions into (19) yields,

\[
(k + h \circ L)^*(\phi, y) = \min_{z \in \mathbb{R}^d} \left\{ (\mathcal{K}(\cdot, \mu))^*\left( (\phi - \langle C^T z, \cdot \rangle) + \delta_{\{0\}}(y + z) + \frac{\alpha}{2} \|z\|_2^2 \right) \right\} \\
= (\mathcal{K}(\cdot, \mu))^*(\phi + \langle C^T y, \cdot \rangle) + \frac{1}{2\alpha} \|y\|_2^2.
\]

\[
\begin{align*}
\text{The conjugate computed in the previous lemma can be used to establish that of the optimal value function.}
\end{align*}
\]

**Lemma 7.** The conjugate of $v$ in (16) is $v^* : y \in \mathbb{R}^d \mapsto (\mathcal{K}(\cdot, \mu))^*(\langle C^T y, \cdot \rangle) + \frac{1}{2\alpha} \|y\|_2^2$ which is $\frac{1}{\alpha}$-strongly convex.

**Proof.** We begin by computing the conjugate,

\[
v^*(y) = \sup_{b \in \mathbb{R}^d} \left\{ \langle y, b \rangle - \inf_{\rho \in \mathcal{M}(\Omega)} \{ k(\rho, b) + h \circ L(\rho, b) \} \right\} = \sup_{\rho \in \mathcal{M}(\Omega)} \left\{ (0, \rho) + \langle y, b \rangle - k(\rho, b) - h \circ L(\rho, b) \right\} = (k + h \circ L)^*(0, y).
\]

In light of (18), $v^*(y) = (\mathcal{K}(\cdot, \mu))^*(\langle C^T y, \cdot \rangle) + \frac{1}{2\alpha} \|y\|_2^2$ which is the sum of a convex function and a $\frac{1}{\alpha}$-strongly convex function and is thus $\frac{1}{\alpha}$-strongly convex. \qed

**Remark 4.** Theorem 2 establishes attainment for the problem defining $v$ in (16), so $\text{dom } v = \mathbb{R}^d$ and $v$ is proper. Moreover, [2, Prop. 2.152] and [2, Prop. 2.143] establish, respectively, continuity and convexity of $v$. Consequently, $(v^*)^* = v$ [32, Thm. 2.3.3] and since $v^*$ is $\frac{1}{\alpha}$-strongly convex, $v$ is Gâteaux differentiable with globally $\alpha$-Lipschitz derivative [32, Rmk. 3.5.3].

We now compute the derivative of $v$.

**Lemma 8.** The derivative of $v$ is the map $b \mapsto \alpha (b - C\mathbb{E}_\bar{\rho}[X])$, where $\bar{\rho}$ is the solution of the primal problem (3), which is given in (11).

**Proof.** Remark 4 guarantees that $v$ is differentiable on $\mathbb{R}^d$, so in particular $\nabla v(b)$ exists. Let $y = \nabla v(b)$, by [32, Thm. 2.4.2 (iii)], $y \in \mathbb{R}^d$ is the unique vector for which $\langle b, y \rangle = v(b) + v^*(y)$. Note that

\[
v(b) = \int_\Omega \log \left( \frac{\exp \langle C^T \tilde{\lambda}, \cdot \rangle}{\int_\Omega \exp \langle C^T \tilde{\lambda}, \cdot \rangle \, d\mu} \right) \, d\bar{\rho} + \frac{\alpha}{2} \|C\mathbb{E}_{\bar{\rho}}[X] - b\|_2^2
\]

\[
= \langle \tilde{\lambda}, C\mathbb{E}_{\bar{\rho}}[X] \rangle - \log \left( \int_\Omega \exp \langle C^T \tilde{\lambda}, \cdot \rangle \, d\mu \right) + \frac{\alpha}{2} \|C\mathbb{E}_{\bar{\rho}}[X] - b\|_2^2.
\]
and

\[ v^* (\lambda) = \log \left( \int_\Omega \exp \left( C^T \lambda, \cdot \right) d\mu \right) + \frac{1}{2\alpha} ||\lambda||_2^2. \]

Thus

\[ v(b) + v^* (\lambda) = \left\langle \lambda, C\mathbb{E}_\rho [X] \right\rangle + \frac{\alpha}{2} ||CE_\rho [X] - b||_2^2 + \frac{1}{2\alpha} ||\lambda||_2^2. \]

The first order optimality conditions for (10) imply that \( \lambda = \alpha (b - CE_\rho [X]) \), so

\[ v(b) + v^* (\lambda) = \left\langle \lambda, b - \frac{1}{\alpha} \lambda \right\rangle + \frac{1}{\alpha} ||\lambda||_2^2 = \left\langle b, \lambda \right\rangle. \]

Thus, \( \lambda \) is the unique vector satisfying \( v(b) + v^* (\cdot) = \left\langle b, \cdot \right\rangle \) and thus \( \nabla v(b) = \lambda = \alpha (b - CE_\rho [X]). \)

We now prove Theorem 3.

**Proof of Theorem 3.** By Lemma 7, \( v^* \) is \( \frac{1}{\alpha} \)-strongly convex, so \( \nabla v \) computed in Lemma 8 is globally \( \alpha \)-Lipschitz (cf. Remark 4), thus

\[ ||\nabla v (b_1) - \nabla v (b_2)||_2 \leq \frac{2}{\sigma_{\min}(C)} ||b_1 - b_2||_2 \]

and

\[ ||\nabla v (b_1) - \nabla v (b_2)||_2 = \alpha ||b_1 - b_2 + C(\mathbb{E}_{\rho_2} [X] - \mathbb{E}_{\rho_1} [X])||_2 \]

\[ \geq \alpha ||C (\mathbb{E}_{\rho_2} [X] - \mathbb{E}_{\rho_1} [X])||_2 - \alpha ||b_2 - b_1||_2 \]

\[ \geq \alpha \sigma_{\min}(C) ||\mathbb{E}_{\rho_1} [X] - \mathbb{E}_{\rho_2} [X]||_2 - \alpha ||b_1 - b_2||_2. \]

Consequently, \( ||\mathbb{E}_{\rho_1} [X] - \mathbb{E}_{\rho_2} [X]||_2 \leq \frac{2}{\sigma_{\min}(C)} ||b_1 - b_2||_2. \)

5. Numerical Results

We present results obtained using our method on certain simulated images. We begin with deconvolution, i.e. when the blurring kernel \( c \) is known. To start we present an example of a highly blurred image with no noise in Figure 1. A near-perfect estimate of the original image can be obtained provided the fidelity parameter \( \alpha \) is sufficiently large (in practice \( \alpha = 10^{15} \) yields good results). Figure 3 provides an example in which a blurry and noisy image has been deblurred using the non-blind deblurring method. We note that the method does not actively denoise blurred images, so a preprocessing step consisting of expected patch log likelihood (EPLL) denoising [33] is first performed. The resulting image is subsequently deblurred and finally TV denoising [26] is used to smooth the image. Note that for binary images such as text, TV denoising can be replaced by a thresholding step. Figure 2 demonstrates the effects of different magnitudes of noise on a uniformly coloured image.

Results for blind deblurring are compiled in Figure 4 and 6. In this case \( \gamma = 10^5 \) and \( \alpha = 10^6 \) provide good results in the noiseless case and \( \gamma = 10^3, \alpha = 10^4 \) is adequate for the noisy case, but these parameters require manual tuning to yield the best results.
Figure 1. Deconvolution: (a) is the original image. (b) is the 459 pixel wide convolution kernel. (c) is the blurred image. (d) is obtained from the blurred image using the non-blind deblurring method. (e) compares the fine details of the original image (left) and deblurred image (right). Despite the high degree of blurring, the reproduction matches the original almost exactly. Comparisons to Krishnan and Fergus’s fast deconvolution method [13] and the expected patch log likelihood deconvolution method [33] were performed, but neither method converged with such a large magnitude of blurring.

Figure 2. Examples of Noise: The leftmost image is a uniformly coloured image. The image is subsequently degraded by 1%, 2% and 5% additive Gaussian noise in the centre left, center right and rightmost images respectively.

5.1. The Effects of Noise

In the presence of additive noise, attempting to deblur images using methods that are not tailored for noise is generally ineffective. Indeed, the image acquisition model $b = c \ast x$ is replaced by $b = c \ast x + p$ where $p$ denotes the added noise. The noiseless model posits that the captured image should be relatively smooth due to the convolution, whereas the added noise sharpens segments of the image randomly, so the two models
Figure 3. Deconvolution with noise: (a) is the original scene. (b) is the 23 pixel wide convolution kernel. (c) is the blurred image which is further degraded with 1% Gaussian noise. (d) is the result obtained from the blurred image via the non-blind deblurring method. (e) is the result obtained using Krishnan and Fergus’s fast deconvolution method [13]. (f) compares the fine detail of our result (top) and the result of the fast deconvolution method (bottom).

are incompatible. However, Figures 3 and 4 show that our method yields good results in both deconvolution and blind deblurring when a denoising preprocessing step and a smoothing postprocessing step are utilized.

Remarkably, the blind deblurring method is more robust to the presence of additive noise in the blurred image. Indeed, accurate results were obtained with up to 5% Gaussian noise in the blind case whereas in the non-blind case, quality of the recovery diminished past 1% Gaussian noise. This is due to the fact that the preprocessing step fundamentally changes the blurring kernel of the image. We are therefore attempting to deconvolve the image with the wrong kernel, thus leading to aberrations. On the other hand, the estimated kernel for blind deblurring is likely to approximate the kernel modified by the preprocessing step, leading to better results. Moreover, a sparse (Poisson) prior was used in the kernel estimate for the results in Figure 4 so as to mitigate the effects of noise on the symbology.

Finally, we note that there is a tradeoff between the magnitude of blurring and the magnitude of noise. Indeed, large amounts of noise can be dealt with only if the blurring kernel is relatively small and for large blurring kernels, only small amounts of noise can be considered. This is due to the fact that for larger kernels, deviations in kernel estimation affect the convolved image to a greater extent than for small kernels.
Figure 4. Blind deblurring with and without noise: This figure compares the performance of our blind deblurring method with EPLL denoising preprocessing and TV denoising postprocessing to that of Pan et al.’s blind deblurring method [18] with an EPLL denoising preprocessing step when deblurring images with varying amounts of noise and different blurring kernels. The subfigures are arranged as follows: Left: Blurred and noisy image. Centre left: Original convolution kernel on top, our estimated kernel in the bottom left and Pan et al.’s estimated kernel in the bottom right. Centre right: The latent image obtained by our method. Right: The latent image obtained by Pan et al.’s method (a) is noiseless with a 33 pixel wide kernel. (b) has 1% Gaussian noise with a 27 pixel wide kernel. (c) has 2% Gaussian noise with a 17 pixel wide kernel and (d) has 5% Gaussian Noise with a 13 pixel wide kernel. Of note is that our method is consistently smoother, the differences are best viewed on a high resolution monitor.

6. The Role of the Prior

Our method is based upon the premise that a priori the probability density $\rho$ at each pixel is independent from the other pixels. Hence in our model, the only way to introduce correlations between pixels is via the prior $\mu$. Let us first recall the role of the prior
Figure 5. Blind text deblurring with and without noise: At the top is the original image. The subfigures are organised as follows: Top: Original convolution kernel on the right and estimated kernel on the left. Middle: Blurred and noisy image. Bottom: Deblurred image obtained using our method with an EPLL denoising preprocessing step and a thresholding postprocessing step. (a) is noiseless with a $57 \times 57$ pixel kernel. (b) has 1% Gaussian noise with a 45 pixel wide kernel.

Figure 6. Blind deblurring in color: (a) is the original image. (b) compiles the original $17 \times 17$ pixel convolution kernel on top, our estimated kernel on the bottom left and the kernel obtained by Pan et al.‘s method on the bottom right. (c) is the blurred image. (d) is the image obtained by performing blind deblurring on the previous blurred image using our method (without preprocessing or postprocessing). (e) is the image obtained via Pan et al.‘s blind deblurring method. (d) demonstrates that the fine details of the original image (top) are preserved in our deblurred image (bottom).
μ in the deconvolution (and ν in the kernel estimation). In deconvolution for general images, the prior μ was only used to impose box constraints; otherwise, it was unbiased (uniform). For deconvolution with symbology, e.g. the presence of a known finder pattern, this information was directly imposed on the prior. For kernel estimation, prior ν was used to enforce normalization and positivity of the kernel; but otherwise unbiased.

Our general method, on the other hand, facilitates the incorporation of far more prior information. Indeed, we seek a prior probability distribution μ over the space of latent images that possesses at least one of the following two properties:

(i) μ has a tractable moment-generating function (so that the dual problem can be solved via gradient-based methods such as L-BFGS),

(ii) It is possible to efficiently sample from μ (so that the dual problem can be solved via stochastic optimization methods).

As a simple example, we provide examples of the use of Bernoulli priors to model binary data and of Poisson priors to model sparsity. Figure 7 presents a comparison of deblurring a binary text image using different priors with the same choice of α = 2×10^{11}. In this case, sparsity has been used to promote the presence of a white background by inverting the colours of the channel during the deblurring process.

![Figure 7. Deconvolution with different priors:](image)

(a) is the original binary text image. (b) is the 207×207 pixel convolution kernel. (c) is obtained by blurring the text with the convolution kernel. (d) and (e) are the results of performing deconvolution on the previous blurred image using Bernoulli and Poisson priors respectively using α = 10^{10}. (f) and (g) were obtained by deconvolving with α = 10^6 with the two priors. (e) presents the fine detail of the deconvolution with α = 10^6 with the Bernoulli prior on the left and the Poisson prior on the right. Pixels which were black in the Bernoulli prior, but were gray in the Poisson prior have been made white manually in order to demonstrate the effect of a sparse prior.

More generally, we believe our method could be tailored to contemporary approaches for priors used in machine learning. In doing so, one could perform bling deblurring without the presence of any finder pattern. A natural candidate for such a prior μ is a generative adversarial network (GAN) (cf. [9]) trained on a set of instances from a class of natural images (such as face images). GANs have achieved
state-of-the-art performance in the generative modelling of natural images (cf. [12]) and it is possible by design to efficiently sample from the distribution implicitly defined by a GAN’s generator. Consequently, when equipped with a pre-trained GAN prior, our dual problem (12) would be tractable via stochastic compositional optimization methods such as the ASC-PG algorithm of Wang et al. in [31].

7. Discussion

It is surprising that inference schemes of the type considered in this paper have not yet been applied to image deblurring. Indeed, the principle of maximum entropy was first developed in a pair of papers published by E.T. Jaynes in 1957. Furthermore, the theory of Fenchel-Rockafellar duality is well established in the convex analysis literature and has found applications to solving maximum entropy estimation problems (cf. [7]).

Since our algorithm models blur as the convolution of the clean image with a single unknown blur kernel, it relies crucially on the spatial uniformity of the blur. This assumption may not hold in certain cases. For example, an image captured by a steady camera that contains a feature that moves during image capture will exhibit non-uniform motion blur. It may be of interest to explore extensions of this algorithm that divide the observed image into patches and estimate different blur kernels for each patch (cf. the motion flow approach proposed in [8]).

Finally, our method is flexible with respect to the choice of prior and as we briefly discussed in Section 6, this strongly alludes to future work on the incorporation of empirical priors obtained from techniques in machine learning.

Implementation Details

All figures were generated by implementing the methods in the Python programming language using the Jupyter notebook environment. Images were blurred synthetically using motion blur kernels taken from [15] as well as Gaussian blur kernels to simulate out of focus blur. The relevant convolutions are performed using fast Fourier transforms. Images that are not standard test bank images were generated using the GNU Image Manipulation Program (GIMP), moreover this software was used to add symbolic constraints to images that did not originally incorporate them. All testing was performed on a laptop with an Intel i5-4200U processor. The running time of this method depends on a number of factors such as the size of the image being deblurred, whether the image is monochrome or colour, the desired quality of the reproduction desired (controlled by the parameter $\alpha$) as well as the size of the kernel and whether or not it is given. If a very accurate result is required, these runtimes vary from a few seconds for a small monochrome text image blurred with a small sized kernel to upwards of an hour for a highly blurred colour image.
Acknowledgments

G.R. was partially supported by the NSERC CGS-M program, R.C. and T.H. were partially supported by the NSERC Discovery Grants program. R.C. would like to acknowledge and express his thanks to Pierre Maréchal, who suggested in 2014 the MEM approach for deblurring of barcodes (cf. [22]).

Bibliography


