The Maximum Entropy on the Mean Method for Image Deblurring: Applying Fenchel-Rockafellar Duality in Finite and Infinite Dimensions

Gabriel Erik Rioux

Master of Science

Department of Mathematics and Statistics

McGill University

Montreal, Quebec

July, 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

DEDICATION

To my parents, for their endless support.

ACKNOWLEDGEMENTS

The author would like to express his gratitude to Rustum Choksi and Tim Hoheisel for their invaluable mentorship throughout both his Master of Science degree as well as his undergraduate research. Their patience and keen insight were integral in ensuring the success of the various projects undertaken throughout this degree. The author would additionally like to thank Christopher Scarvelis for his contributions to both papers presented in this thesis, Pierre Maréchal for proposing the use of the maximum entropy on the mean framework for image deblurring, and Maxime Laborde for his collaboration on an unrelated project. Finally, the author acknowledges the support of the NSERC CGS-M program throughout this degree.

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Image deblurring is an inverse problem which has seen a surge of activity in recent years due to the advent of machine learning-based approaches. As such, traditional methods consisting of optimizing a fidelity term coupled with a regularizer over the set of all possible images have fallen in popularity. In the following, a novel approach to this problem is proposed, based upon the maximum entropy on the mean method. It consists of optimizing at the level of the set of probability distributions on the set of all images and employing an entropic regularization. The theory is first described in the context of barcode deblurring and, subsequently, for the deblurring of general images. The problem afforded by the principle of maximum entropy on the mean is intractable (it is finite-dimensional, but prohibitively large in the former case and infinite-dimensional in the latter). Nevertheless, a judicious application of the Fenchel-Rockafellar duality theorem affords a finite-dimensional dual problem which can be solved using standard optimization software, as well as a formula to recover a solution of the original problem from that of its dual counterpart. Numerical experiments are provided to demonstrate the strength of this method.

ABRÉGÉ

Récemment, le problème inverse de débrouillage d'image a fait l'objet d'une résurgence due à la popularité de l'apprentissage automatique. Ainsi, les méthodes classiques, basées sur la régularisation d'un problème d'optimisation sur l'ensemble de toutes les images, sont de moins en moins étudiées. Dans la présente thèse, il sera question d'une nouvelle approche basée sur la méthode du maximum d'entropie sur la moyenne. Cette dernière consiste à optimiser à l'échelle des lois de probabilité sur l'ensemble des images et d'employer une régularisation entropique. D'abord, cette méthode sera décrite dans le contexte de débrouillage de code-barres et, ensuite, pour le débrouillage d'images génériques. Dans le premier cas, le problème à résoudre a une dimensionalité immense tandis que dans la seconde, la dimensionalité est infinie. Néanmoins, une application judicieuse du théorème de dualité de Fenchel-Rockafellar accorde un problème dual de dimension finie qui peut être résolu numériquement avec des logiciels génériques, ainsi qu'une formule permettant de passer d'une solution du problème dual à une solution du problème d'origine. Des résultats numériques sont également fournis afin de démontrer la puissance de cette méthode.

TABLE OF CONTENTS

CHAPTER 1
Introduction

The problem of image deblurring is perhaps the most ubiquitous inverse problem encountered in day to day life. Indeed, with the advent of smartphones, billions of people have the ability to take digital photos at a moments notice. If, in haste, one captures an image before the camera is stable the resulting image will likely be corrupted by motion blur. Similarly, if the subject of the image is in movement, motion blur may manifest itself in the captured photo. These two examples are among the most common causes of blur in an image. Another important example is that of out of focus blur, which occurs when the focus of the camera is incorrectly set.

In addition to salvaging personal photographs, research in deblurring is motivated by important applications such as medical imaging [1, 2, 9, 50, 59, 60], astronomical imaging [13, 18, 33, 43] (a particularly exciting application was the reconstruction of images from the M87 Event Horizon Telescope, yielding the "first picture" of a black hole [14]), and fast and accurate reading of barcodes [15, 17, 29, 30, 41, 47, 53] among myriad others.

The deblurring literature is split into two broad categories, the first consists of methods that are based on machine learning and the second, those that are not. The method described in this manuscript does not employ the techniques of machine learning, hence the theory discussed in the sequel will be limited to more classical regularization-based techniques. The reader interested in machine learning may find the following references interesting [12, 36, 44, 45, 48, 54, 61]. The following section describes the methodology employed by most regularization-based deblurring software.

1.1 Classical Framework for Image Deblurring

1.1.1 Image Acquisition

In order to deblur a given image, one must first understand the mechanism by which blurred images are created. Throughout, it is assumed that blurring occurs on a per channel, thus an RGB image (consisting of one red channel, one green channel, and one blue channel) would have all three of its channels blurred independently. As such, let $x \in \mathbb{R}^{n\times m}$ denote one channel of the ground truth 2-dimensional image and let $c \in \mathbb{R}^{k\times k}$ $k < n, m$ (k odd) denote the convolution kernel. Then the blurred image $b \in \mathbb{R}^{n\times m}$ is obtained via the relation

$$c * x = b$$

with $*$ denoting the 2-dimensional linear convolution operator; Figure 1–1 presents an example of the simulated image acquisition process and an example calculation is provided subsequently in Example 1. This model represents spatially invariant blurring as opposed to spatially variant blurring, wherein different segments of the image are blurred with different kernels (see [8, 11, 34, 58] for examples of methods tailored to spatially variant blurring). The assumption of spatial invariance is made throughout this manuscript.

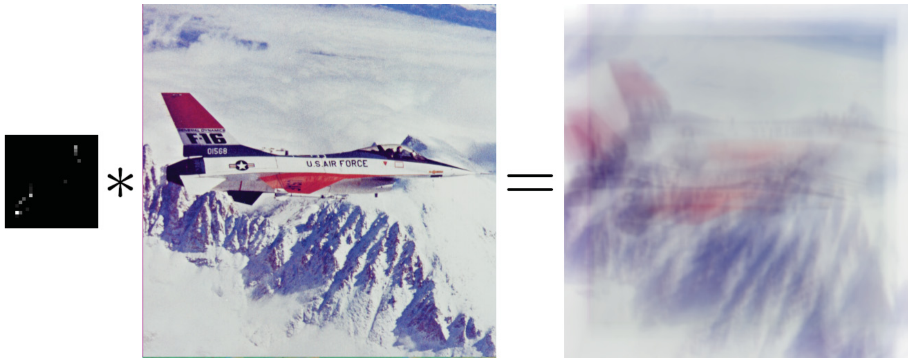


Figure 1–1: The blurring kernel c on the left, convolved with the original image x in the middle yields the blurred image b on the right.

Example 1 (2D Convolution)**.** *Consider a channel* $A \in \mathbb{R}^{n \times m}$ *as well as its vectorization* $a \in \mathbb{R}^{nm}$ *obtained by taking the rows of* A *and stacking them together into a vector, i.e.*

$$a = \left((A^T)_1, (A^T)_2, \dots, (A^T)_n\right),$$

with A_i *denoting the* i*-th column of* A*.*

Moreover, let $c \in \mathbb{R}^{2\ell-1 \times 2\ell-1}$ *denote the convolution kernel. Then the convolution between* c *and* A *can be written componentwise as*

$$[c * A]_{i,j} = \sum_{k=1}^{n} \sum_{l=1}^{m} c_{\ell+i-k,\ell+j-l} a_{k,l},$$

using the convention that $a_{i,j} = 0$ *if* $i, j < 1$*,* $i > n$*, or* $j > m$ *and* $c_{i,j} = 0$ *if* $i, j < 1$ *or* $i, j > 2\ell - 1$*.*

One can generate a matrix $C \in \mathbb{R}^{nm \times nm}$ acting on the vectorization a of the image A as follows. Let

$$C_i = \begin{bmatrix} c_{i,\ell} & c_{i,\ell-1} & \dots & c_{i,1} & 0 & \dots \\ c_{i,\ell+1} & c_{i,\ell} & \dots & c_{i,2} & c_{i,1} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ c_{i,2\ell-1} & c_{i,2\ell-2} & \dots & \dots & \dots & \dots \\ 0 & c_{i,2\ell-1} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix},$$

be $n \times n$ matrix blocks for $i \in \{1, 2, \dots, 2\ell - 1\}$. Then the convolution matrix C is the following m block by m block matrix,

$$C = \begin{bmatrix} C_\ell & C_{\ell-1} & \dots & C_1 & 0 & \dots \\ C_{\ell+1} & C_\ell & \dots & C_2 & C_1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ C_{2\ell-1} & C_{2\ell-2} & \dots & \dots & \dots & \dots \\ 0 & C_{2\ell-1} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

In practice, convolutions can be efficiently handled using fast Fourier transform based convolution routines (cf. [16]), thus nullifying the need to store the convolution matrix or to directly perform these sums.

Note that the image acquisition model heretofore described is noiseless. In order to explicitly account for additive noise $p \in \mathbb{R}^{n \times m}$, the model $b = c * x + p$ can be used. The most common types of noise are white noise and salt-and-pepper noise [38]. White noise is assumed to be drawn from a Gaussian with unknown variance and mean zero; it is typically caused by issues with the image or by the analogue to digital conversion of the signal. The latter corrupts the image by turning a percentage of the pixels black or white. It can be caused by damaged sensors in the camera.

Throughout, the theory is developed in the assumption that the image is captured without noise and that the blurring is spatially invariant. Now that the image acquisition process is well understood, a framework for regularization-based deblurring methods can be described.

1.1.2 Standard Deblurring Methodology

First, consider the problem of non-blind deblurring (deconvolution), which consists of taking a given blurred image b and recovering the latent image x when the convolution kernel c, or an approximation thereof, is known.

1.1.2.1 Deconvolution

In this setting, a naive approach to deconvolution consists of inverting the convolution operator. As aforementioned (see Example 1) $c* : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}$ can be represented as a matrix $C \in \mathbb{R}^{nm\times nm}$ acting on a vectorized version of the image $x \in \mathbb{R}^{nm}$ (this notation will be used interchangeably). The blurring matrix C is typically non-singular [38], however the same reference indicates that the condition number of C (the square root of the ratio of the largest and smallest eigenvalues of $C^T C$) is very large, thus inversion is very sensitive to round-off error [24].

The standard approach to solving ill-conditioned problems consists of regularizing it by imposing additional structure on solutions. Indeed, the majority of deconvolution software solve some variant of the following problem

$$\min_{x \in \mathbb{R}^{n \times m}} \left\{ R(x) + \frac{\alpha}{2} \, ||c * x - b||_2^2 \right\}. \tag{1.1}$$

In (1.1), $R : \mathbb{R}^{n \times m} \to \mathbb{R}$ is a regularizer which is used to enforce certain constraints on the optimizer. The second term promotes fidelity of $c * x$ to b without necessarily imposing equality, the parameter α is used to balance the importance of this term against that of the regularizer. The case where the kernel is unknown can now be described.

1.1.2.2 Blind Deblurring

For generic blurred images, the convolution kernel c is, a priori, unknown and thus the problem in (1.1) must be modified. The most common approach [7] is to solve

$$\min_{\substack{x \in \mathbb{R}^{n \times m} \\ c \in \mathbb{R}^{k \times k}}} \left\{ R(x, c) + \frac{\alpha}{2} \, ||c * x - b||_2^2 \right\}, \tag{1.2}$$

where the optimization is performed over the image and kernel simultaneously. Often the regularizer is of the form $R(x, c) = R_x(x) + R_c(c)$ such that solving (1.2) is tantamount to the iterative resolution of the following problems [10]:

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

Of note is that one must first initialize c (or x if one starts with the kernel estimation step) with some matrix in order to implement this approach, then the image x (resp. the kernel c) used in the second problem is taken to be the solution of the first problem. The routine then continues, using the kernel c (resp. the image x) obtained from the second problem to solve the first. This iterative method is repeated until some stopping condition is met. Since the size of the initial blurring kernel is unknown, a coarse-to-fine approximation is often adopted. It consists of starting with a small kernel, say 11×11 pixels, and increasing the kernel size after a certain number of iterations [10]. Some common choices of regularizers will now be discussed.

1.1.2.3 Regularization for Deblurring

The majority of literature pertaining to regularization-based deblurring consists of proposing different types of regularizers and analysing their effects on solutions of (1.2). In particular, the image regularizer (R_x) should penalize the presence of structures in the deblurred image which do not occur in natural images whereas the kernel regularizer (R_c) should enforce characteristics common to convolution kernels. For example, in text images the intensity of the pixels is essentially bimodal with peaks at 0 corresponding to the text and 1 corresponding to the background. For kernels, a natural assumption is that they have must non-negative components summing to 1. We list some common regularizers which can be used for either the image or the kernel estimation (thus, we denote the regularizer by R and the kernel or image by z) and briefly describe their effects.

Example 2 (Examples of Regularizers)**.** *Throughout, the matrix $D : \mathbb{R}^d \to \mathbb{R}^{2d}$ will denote the discrete gradient matrix for vectorized matrices. That is, $Dz = [D_x z, D_y z]$ with $D_x, D_y : \mathbb{R}^d \to \mathbb{R}^d$ denoting the finite difference approximations of the derivatives of the image in the horizontal and vertical directions respectively. For example, for an image $A \in \mathbb{R}^{n\times m}$ with vectorization $a \in \mathbb{R}^{nm}$, discretizing the derivatives using first-order forward differences yields*

$$D_x a = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} * A, \qquad D_y a = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} * A.$$

Figure 1–2 demonstrates the action of these derivative operators on an image.

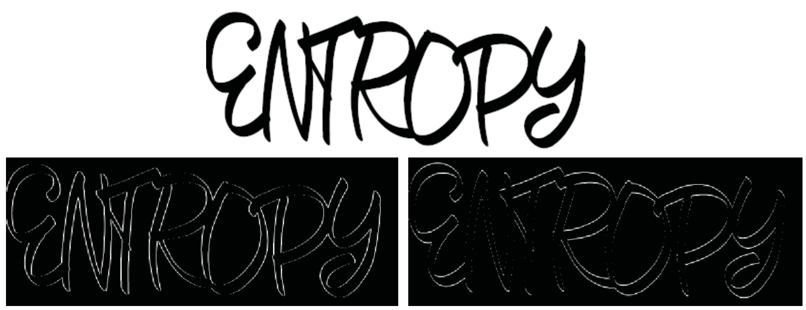


Figure 1–2: A visual depiction of the derivative operations on 2D images. The bottom left and bottom right images are the horizontal and vertical derivatives of the top image respectively. Note that the absolute value of the derivatives is taken, as its magnitude is arbitrary (indeed, first-order backward differences could have been used).

With this notation in hand, we present some examples of regularizers.

1. ***Tikhonov Regularization:*** *Perhaps the most well-known regularizer for ill-posed problems is the ℓ_2 penalty, dating back to the work of Tikhonov [49]. Here, $R(z) = ||\Gamma z||_2^2$ for some matrix Γ. For $\Gamma = D$, the result is to enforce*

smoothness of solutions, effectively minimizing large jumps in intensity between neighbouring pixels. This assumption is natural for most types of images, whence justifying its use in many methods [11, 19, 24, 31, 35]. For $\Gamma = \alpha I$ *with* $\alpha \in \mathbb{R}$ *and* I *the identity matrix, solutions with small coefficients will be favoured. However, since the norm squares the contribution of each coefficient, coefficients* z_i *such that* $|z_i| > 1$ *are over-penalized and those with* $0 < |z_i| < 1$ *are under-penalized. Such behaviour is not necessarily desirable, especially when* $\Gamma = D$*, as smooth textures will have essentially* 0 *gradient, whereas edges will have large gradient*[1] *. This behaviour can be improved by employing the* ℓ^1 *penalty.*

2. ℓ_1 ***Penalty:*** *The* ℓ_1 *penalty is given by* $R(z) = ||z||_1 = \sum_{i=1}^{d} |z_i|$ *[24]. It is not dissimilar to Tikhonov regularization, however, it is more permissive of large coefficients and less permissive of small but non-zero coefficients [58]. It is often used in conjunction with other regularizers [23, 25, 46]. Again, this penalty can be composed with some matrix* A *in order to regularize some quantity other than the intensity of the image. The case* $A = D$ *bears the name of (anisotropic) total variation regularization.*

3. ***Total Variation:*** *In (anisotropic) total variation regularization,* $R(z) = ||Dz||_1$*. As aforementioned, the gradient of the image naturally contains*

[1] This also depends on what numeric representation is being used to store the intensity values. For example, one can encode the intensity of a channel as a number in $[0, 1]$ or as an integer between 0 and 255 (8-bit encoding). The maximal gradient in the first case is 1 whereas in the second it is 255.

some large components at the edges [24] and thus should not be overly penalized. The total variation penalty, therefore, yields a sharper restoration than the Tikhonov regularizer. It has seen much popularity in the field of image deblurring [20, 28, 37, 51, 52, 55], due in part to the popularity of the Rudin-Osher-Fatemi method for image denoising [42].

4. ℓ_0 ***Penalty:*** *The* ℓ_0 *penalty is given by* $R(z) = \mathrm{card}(I)$, *with* $I = \{i = 1, \ldots, d : z_i \neq 0\}$ *i.e. the number of non-zero components in the signal [39, 40]. This regularizer clearly promotes sparse solutions and treats all non-zero components equally. As with other regularizers, one can penalize the gradient of the image using the* ℓ^0 *penalty.*

The approach for image deblurring proposed in this manuscript is based upon the principle of maximum entropy on the mean as described in the following section.

1.2 The Principle of Maximum Entropy on the Mean

The principle of maximum entropy has its roots in the study of statistical mechanics. Put simply, statistical mechanics consists of studying the properties of large bodies of particles, so large in fact that determining the equations of motion of each particle (as in classical mechanics) is infeasible [27]. Since the number of particles is so large, statistical laws which have no analogue in smaller systems arise due to the larger number of degrees of freedom. Thus, one studies such systems by means of probability distributions on the set of all possible states that the system may take; quantities of interest, such as energy, position and magnetization of the system are described via their expected value.

In his landmark papers [21, 22] Jaynes extolled the virtues of a reinterpretation of statistical mechanics based on the principle that thermodynamic entropy and information-theoretic entropy represent the same quantity. From this viewpoint, given partial information about a system, the least biased probability distribution describing this system subject to the given constraints is that which maximizes some measure of entropy.

Throughout it will be assumed that such prior information can be encoded via a prior probability distribution μ. In such a case, the Kullback-Leibler relative entropy [26] is the canonical choice of entropy, as it permits a measurement of the discrepancy between two measures.

Later, Dacunha-Castelle and Gamboa provided a mathematical formulation of the principle of maximum entropy on the mean. It consists of seeking the maximum entropy distribution among all distributions satisfying a moment constraint. Thus, the principle of maximum entropy on the mean comprises solving a convex program (in finite or infinite dimensions) subject to finitely many affine constraints. These constraints can, in turn, be emulated via the addition of a convex penalty which penalizes deviations from the constraint set. For such problems, approaches based on Fenchel-Rockafellar duality are often auspicious. Indeed, the Fenchel-Rockafellar duality theorem establishes a dual problem to

$$\inf_X \{f + g \circ A\},$$

where (X, τ), (Y, τ') are separated locally convex spaces with (topological) duals X^* and Y^* respectively, A is a continuous linear operator from X to Y and

$f : X \to \bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ and $g : Y \to \bar{\mathbb{R}}$ are proper, convex functions. In the context of the maximum entropy on the mean method, A encodes the affine constraints, g acts as a fidelity term for these constraints and f acts as a regularizer (in this case the relative entropy).

The dual problem afforded by this approach is a maximization problem over Y^* which, in some cases, is more readily solved than the original problem. In particular, if $Y = \mathbb{R}^d$ the dual problem is finite-dimensional, this case bears the name of partially finite convex programming [5, 6]. Under certain assumptions, a formula can be established that relates solutions of the dual problem to those of the primal problem. Explicitly, the dual problem reads

$$\max_{Y^*} \left\{ -f^* \circ A^* - g^* \right\},$$

with f^* and g^* denoting the convex conjugates of f and g, and A^* denoting the adjoint of A. Of note is that the conjugate of the relative entropy has a well-known formula which we will derive in the sequel. Indeed, for continuous densities, the conjugate of the entropy is the Gibbs measure, which is ubiquitous in statistical physics [27, Eq. 28.8] and probability theory [62]. In the discrete case, it is given by the softmax function which has seen applications in information theory [21], statistical mechanics [27, Eq. 28.3], deep learning [63], and economics [64].

With these preliminary notions in hand, the maximum entropy on the mean method can be applied to the problem of image deblurring. Previously, the application of the maximum entropy on the mean method for solving linear inverse problems has been considered [3, 4, 32]; mostly in the context of deblurring astronomical images. These approaches differ from the one presented in this manuscript in the following ways:

- First, they consider the constraint on the mean to be a hard constraint, that is, the optimization is performed over distributions whose mean, once convolved with the (known) kernel is equal to the blurred image, so no fidelity term is needed. This approach necessitates the verification of a qualification condition which depends on the chosen prior. By penalizing deviations from this constraint (as in our proposed approached, also called a soft constraint) rather than strictly enforcing it, an anologous qualification condition is trivially satisfied for any choice of prior. Moreover, the dual problem obtained in our setting has a strongly convex objective function, which is ideal for optimization and permits the derivation of a stability estimate with respect to the blurred input image and the corresponding deblurred image. Whether or not such an estimate holds when a hard constraint is used is unclear (in particular only strict convexity of the objective is guaranteed in this case [3], the same reference provides a short sensitivity analysis in the limit of small variations).
- Next, using a soft constraint, it is possible to perform blind deblurring, that is, both a kernel estimation step as well as a deconvolution step can be written using a unified framework. An approach using a hard constraint is not amenable to blind deblurring, as in the deconvolution step one will have imperfect knowledge of the kernel. In particular, even if the estimated convolution matrix is non-singular it is not guaranteed that any image (i.e.

any vector in $[0,1]^{nm}$) will be identical to the blurred image once convolved. The same issues arise when performing the kernel estimation step. With a soft constraint, one can enforce that the mean of the solution is an image via the regularizer and use a fidelity parameter to lessen the impact of such imperfections in the estimated kernel and image.

- Finally, in [3, 4], noise is explicitly accounted for. That is, the image vector x is augmented with a noise vector n and the image acquisition process is written as

$$b = \begin{bmatrix} C & I \end{bmatrix} \begin{bmatrix} x \\ \\ n \end{bmatrix} = Cx + n.$$

  where I denotes the identity matrix. Thus in the maximum entropy on the mean method one optimizes over the product measure $\rho \otimes \eta$, with $\mathbb{E}_\rho[\boldsymbol{X}]$ approximating the image and $\mathbb{E}_\eta[\boldsymbol{X}]$ approximating the noise. In the proposed method, noise is handled implicitly by the regularizer. The latter approach to denoising is favored in contemporary deblurring methods, as one simply promotes the qualities desired in the image (e.g. smoothness) using the regularizer. In the explicit model, it is unclear how the corrections from n interact with the estimate of the image x in the optimization procedure.

These ideas are further developed in the following chapters; a brief summary of their contents is provided in the following section.

1.3 Summary of Chapters

In Chapter 2, the maximum entropy on the mean method for deblurring binary images is studied. In this context, the set of all $n \times m$ images has cardinality

2^{nm}, since any given pixel can only take one of two intensity values. Thus, the problem of interest is 2^{nm}-dimensional and the probability distributions over the set of images are discrete. The Fenchel-Rockafellar duality theorem affords a tremendous dimensionality reduction, resulting in a dual problem that is nm-dimensional, as well as a formula for recovering the mean of the solution of the original problem from the solution of the dual one. Numerical experiments are equally performed which demonstrate the strength of this method in both blind and non-blind deblurring.

In Chapter 3, a generalization of the techniques presented in chapter 2 are presented. Indeed, general images whose channels take intensity values in $[0, 1]$ are considered. Thus, the set of all possible images is infinite-dimensional (discounting discretization of colours due to colour spaces) and the probability distributions on this set are continuous. The deblurring problem is, therefore, infinite-dimensional, whereas the dual problem afforded by the duality theorem is, again, nm-dimensional. A recovery formula is also established which permits the evaluation of the mean of the maximum entropy distribution after the numerical resolution of the dual problem. Examples are equally provided that serve as a proof of concept of the potential of this method.

CHAPTER 2
Paper 1: Blind Deblurring of Barcodes via Kullback-Leibler Divergence

Authors: Gabriel Rioux[1], Christopher Scarvelis[1], Rustum Choksi[1], Tim Hoheisel[1], and Pierre Maréchal[2]

Author Ordering: Rioux and Scarvelis equally share the role of first author. Choksi, Hoheisel and Maréchal are then listed alphabetically.

[1]Department of Mathematics and Statistics, McGill University, Montreal, QC H3G 1Y6, Canada

[2]Institut de Mathématiques de Toulouse, Université Paul Sabatier, Toulouse 31 062, France

Abstract

Barcode encoding schemes impose symbolic constraints which fix certain segments of the image. We present, implement, and assess a method for blind deblurring and denoising based entirely on Kullback-Leibler divergence. The method is designed to incorporate and exploit the full strength of barcode symbologies. Via both standard barcode reading software and smartphone apps, we demonstrate the remarkable ability of our method to blindly recover simulated images of highly

blurred and noisy barcodes. As proof of concept, we present one application on a real-life out of focus camera image.

2.1 Introduction

Deblurring in image processing addresses a notoriously difficult ill-posed problem. In this article we present a novel algorithm for deblurring and denoising of barcodes. The strength of our method lies in its effective incorporation (at all stages) of the precise symbology of barcodes. In principle, our method could apply to any class of images possessing some a priori set structure. We present and test the method for barcodes for the following reasons: (i) Barcodes remain ubiquitous objects for the encoding of information, and are the simplest class of images which follow a fixed symbology. (ii) For large amounts of blurring and noise, there is a less ambiguous test of the success of the algorithm than the *eye norm* – their readability by standard commercial software and smartphone *apps*.

One dimensional (1D) UPC barcodes remain popular for coding merchandise while QR (*Quick Response*) barcodes, a type of matrix 2D barcode [38, 4], are increasingly popular because of the ubiquity of smartphone cameras. While barcode readers and smartphone apps are well-developed, the issue of deblurring and denoising barcodes remains of considerable interest with the presence of motion blur from hand movement and noise intrinsic to the camera sensor. The interplay between deblurring and barcode symbology is important for the successful use of mobile smartphones [26, 17, 49, 50, 13, 53]. Methods for deblurring and denoising of barcode signals are well-developed; for example, many techniques have been presented in academic articles (see, for example, [12, 9, 16, 18, 51], [21], [25], [24], [27],

[29], [30], [33], [34], [35], [40], [41], [47], [50], [52]) while implemented algorithms are hidden in commercial software (for example, open source readers like *Zbar* and apps like Apple's *QR Reader*).

The majority of general state-of-the-art blind deblurring methods approach the problem in two steps. The first step is to estimate the blurring kernel and the second is to use non-blind deblurring methods to estimate the original image using the approximative kernel (cf. [11], Chapter 1 of [42] and the references therein, [28]). Our approach follows this structure, however we present novel kernel estimation and deblurring methods that are based on an approach known as the *Method of Maximum Entropy on the Mean* (MMEM) through the Kullback-Leibler divergence. To this end, we do not attempt to find the cleaned image directly but rather we find its probability density function over all binary arrays. We then take, as our best guess of the cleaned image, its (thresholded) expectation. While this particular use of entropy and the Kullback-Leibler divergence has a well-established record of success in many areas of information theory (cf. [15, 6]), we believe this is the first implementation for deblurring of barcodes. In fact, while the Kullback-Leibler divergence appears in the highly-cited deblurring paper of Fergus et al. [19], to our knowledge this particular approach is also new within the wider context of image deblurring. As can be seen in Fig. 2–7-2–10, our method is quite remarkable in its ability to blindly deblur and denoise data. In each case, the only information used to reconstruct the barcode from the simulated blurred and noisy signal is the QR symbology (cf. Fig. 2–1). Software (Zbar and smartphones) were all unable to read the initial signal; however, all can read our processed versions. To our knowledge, we are unaware of any other simple method which can produce such dramatic results.

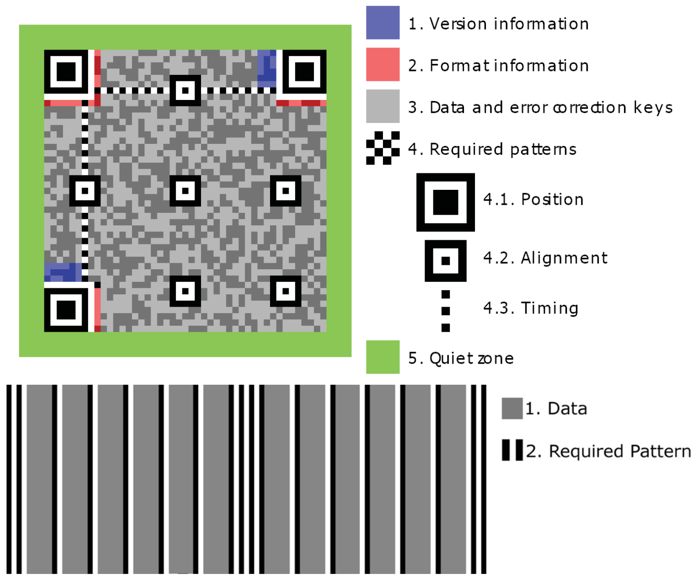


Figure 2–1: A depiction of the symbolic constraints in UPC-A and QR codes (Source (top image): Wikipedia [4] (image by Bobmath, CC BY-SA 3.0).

The principle of maximum entropy was introduced by E.T. Jaynes in 1957 [22, 23]. This principle states that among all probability distributions that are compatible with given moments, the least biased is the one that maximizes the entropy. If prior knowledge on the unknown distribution is available, then the Kullback-Leibler relative entropy is the method of choice. A particular occurrence of it is named the MMEM which was introduced by Dacunha-Castelle and Gamboa [15], and implemented later on in various applications (see e.g. [6, 37]). Applying the MMEM entails solving a convex program in possibly infinite dimensions under finitely many affine equality constraints. This type of problem is efficiently approached by means of Fenchel-Rockafellar duality [43, 7, 8, 36]. In our application we consider a finite dimensional problem; albeit one of "very high" dimension.

We briefly outline our entropic barcode method; the details of the algorithm are presented in Section 2.2.

2.1.1 Outline of our Kullback-Leibler Approach

We model a barcode by a vector $\boldsymbol{x} \in \{0,1\}^N$ of N independent Bernoulli random variables with x_i denoting the i-th bar for a UPC barcode or the i-th module for a QR code. For UPC barcodes $N = 95$ while for QR barcodes N ranges from 441 to 31329. We model the blurring of the barcode $\boldsymbol{x}$ via discrete linear convolution of the form $\boldsymbol{b} = C\boldsymbol{x} \equiv \boldsymbol{c} * U\boldsymbol{x}$, where $C \in \mathbb{R}^{Nm \times N}$, m is an *upscaling factor* as explained at the start of Section 2.2 and $U \in \mathbb{R}^{Nm \times N}$ is the matrix that upscales $\boldsymbol{x}$. C is therefore responsible for upscaling and blurring $\boldsymbol{x}$ with *point spread function* (PSF) $\boldsymbol{c} \in \mathbb{R}^{Nm}$, also known as the blur kernel, and $\boldsymbol{b} \in \mathbb{R}^{Nm}$ the observed blurry signal. Let us for the moment assume that C is known (this is the case for non-blind deblurring). The number of possible barcodes is 2^N and we let $\boldsymbol{p}$ be a *probability mass function* (PMF) defined over the space of barcodes $\{0,1\}^N$. Hence, $\boldsymbol{p} \in \mathbb{R}^{2^N}$, where the k-th component of $\boldsymbol{p}$, denoted p_k (this notation will be used throughout) represents the probability assigned to the k-th barcode in $\{0,1\}^N$. Now, given a prior distribution μ over the space of barcodes, we minimize the function

$$p \mapsto \sum_{i=1}^{2^N} p_i \log\left(\frac{p_i}{\mu_i}\right) \; + \; \gamma \, \| \, C\mathbb{E}_{\boldsymbol{p}}[\boldsymbol{x}] - \boldsymbol{b} \, \|^2 \tag{2.1}$$

over all PMFs $\boldsymbol{p}$. With the solution $\bar{\boldsymbol{p}}$ in hand, our cleaned (processed) barcode is then the thresholded expectation of $\bar{\boldsymbol{p}}$, $\mathbb{E}_{\bar{\boldsymbol{p}}}[\boldsymbol{x}]$. Ideally $\bar{\boldsymbol{p}}$ should be a 1-hot vector such that it gives full weight to a single barcode. We consider a uniform prior, a prior based entirely on the symbology (i.e. one which assigns probability 0 to any $\boldsymbol{x}$ which does not respect the symbology), and for UPC-A barcodes an empirically generated prior based upon a database of 10^6 barcodes.

Even with C known, problem (2.1) with our range of N, is numerically intractable. To this end, we employ the following strategy. First, we exploit Fenchel-Rockafellar duality with a significantly simplified dual problem which has Nm degrees of freedom as opposed to 2^N for the primal problem (2.1). While this presents a fundamental reduction in complexity, it is still too costly to compute $\bar{\boldsymbol{p}}$ via the solution to the dual problem. On the other hand, we do not need to find $\bar{\boldsymbol{p}}$ but rather its expectation, and to this end we present a probabilistic version of the dual which allows for the quick and efficient computation of $\bar{\boldsymbol{x}} = \mathbb{E}_{\bar{\boldsymbol{p}}}[\boldsymbol{x}]$.

The above outlines the method when C is known. For blind deblurring, i.e. when C is unknown, we perform an iterative process which couples the above with an entropy based optimization (cf. (2.5) in the following section) to estimate $\boldsymbol{c}$ from the observed signal $\boldsymbol{b}$ and $\bar{\boldsymbol{x}}$, where $\bar{\boldsymbol{x}}$ is the outcome of the previous entropic image estimation. The iteration begins with an initial estimation of $\boldsymbol{c}$ based upon $\boldsymbol{b}$ and $\bar{\boldsymbol{x}} = \mathbb{E}_{\mu}[\boldsymbol{x}]$.

2.2 The Entropic Blind Deblurring Method

Throughout, the process of capturing an image will be modeled via $\boldsymbol{b} = \boldsymbol{c} * U\boldsymbol{x} = C\boldsymbol{x}$ where $\boldsymbol{x} \in \{0,1\}^N$ is the original barcode, $C \in \mathbb{R}^{Nm \times N}$ is a matrix that upsamples and blurs the image via discrete linear convolution by the PSF $\boldsymbol{c}$ and $\boldsymbol{b} \in \mathbb{R}^{Nm}$ is the acquired image. We model the unknown true barcode $\boldsymbol{x}$

as a vector $\boldsymbol{X} = (X_1, ..., X_N)$ of N independent Bernoulli random variables and recall $N \in \mathbb{N}$ is the total number of the barcode modules. We let $m \in \mathbb{N}$ be an upscaling factor, as the pixels of a camera will seldom align in a one-to-one manner with the bars of the barcode. For example, if $m = 3$, one module of a QR code will correspond to a block of 3×3 pixels rather than just one pixel. Moreover, upscaling is necessary in our model to consider realistic quantities of blurring as demonstrated in Fig. 2–4.

We represent the probability mass function as a vector $\boldsymbol{p} \in \Delta_{2^N}$, where the i-th component of $\boldsymbol{p}$ corresponds to the probability $p(x^i)$ of the i-th binary sequence in $\{0, 1\}^N$ under some arbitrary ordering of the set. We use the symbol Δ_n to denote the unit simplex in $\mathbb{R}^n$ defined as

$$\Delta_n = \left\{ \boldsymbol{u} \in \mathbb{R}^n \,:\, \sum_{i=1}^{n} u_i = 1,\, u_i \geq 0 \ (i = 1, \ldots, n) \right\}.$$

The unit simplex Δ_n is the space of probability distributions over a finite sample space of cardinality n.

The Kullback-Leibler relative entropy quantifies the divergence between two probability distributions and is defined in [31, Eqn. 2.4] as

$$\mathscr{K}(\boldsymbol{p}; \mu) = \begin{cases} \sum_{i \in I} p_i \log\left(\frac{p_i}{\mu_i}\right) & \text{for } \boldsymbol{p} \in \Delta_{2^n} \\ +\infty & \text{otherwise} \end{cases},$$

where $I = \{j \,:\, \mu_j > 0\}$. Working in the convention that $0 \log 0 = 0$ and that $p_k = 0$ for some k if and only if $\mu_k = 0$, summing over I is equivalent to summing from $i = 1, \ldots, 2^N$, as if $\mu_j = 0$ for some j, the j-th summand is 0. These constraints ensure that the entropy term is well-defined. Moreover, μ denotes a prior probability distribution which encodes certain characteristics which a valid barcode should exhibit.

The constraint on the mean can be rephrased by noting that $\mathbb{E}_{\boldsymbol{p}}(\boldsymbol{x}) = A\boldsymbol{p}$ where $A \in \{0,1\}^{N\times 2^N}$ is a matrix formed by ordering the set of all binary sequences of length N and letting the i-th column of A be the i-th element of this ordering. Thus, A computes the expectation value. This constraint will be enforced by means of the penalty function

$$\boldsymbol{p} \mapsto \gamma \left|\left|\boldsymbol{c} * UA\boldsymbol{p} - \boldsymbol{b}\right|\right|^2 . \tag{2.2}$$

Here, $\gamma > 0$ is a scalar which can be varied in order to penalize deviations from the mean to a variable extent. We make precise that the standard Euclidean norm will be used throughout. This form of penalization is flexible enough to permit the presence of additive noise in the image acquisition process without needing to explicitly account for it. Indeed, with noise, the observed barcode $\boldsymbol{b}$ will not generally be equal to $\boldsymbol{c} * UA\boldsymbol{p}$ for any $\boldsymbol{p}$, hence a hard constraint on the mean enforcing that $\boldsymbol{c} * UA\boldsymbol{p}$ must equal $\boldsymbol{b}$ is inadequate.

In the case of blind deblurring, we seek to determine the PMF $\bar{\boldsymbol{p}}$ and the convolution kernel $\bar{\boldsymbol{c}}$ which solve

$$\inf_{\boldsymbol{p},\boldsymbol{c}} \left\{ \mathscr{K}(\boldsymbol{p};\mu) + \mathscr{K}(\boldsymbol{c};\nu) + \gamma \left|\left|\boldsymbol{c} * UA\boldsymbol{p} - \boldsymbol{b}\right|\right|^2 \right\} , \tag{2.3}$$

as $\bar{\boldsymbol{p}}$ will allow us to estimate the original barcode and the PSF responsible for the blurring is unknown. In this equation, μ and ν are distinct prior probability distributions and the particular characteristics that μ and ν encode will be discussed in the following sections, as they play a fundamentally different role. In this framework, the Kullback-Leibler divergence also guarantees that $\bar{\boldsymbol{p}}$, $\bar{\boldsymbol{c}}$ are elements of the 2^N-simplex by its very definition. The utility of this property will be made clear later. Our approach to tackling problem (2.3) is by iteratively coupling the following subproblems.

1. Image estimation based on $\boldsymbol{c}$ (non-blind deblurring): Determine $\bar{\boldsymbol{p}}$ as a solution of

$$\inf_{\boldsymbol{p}} \left\{ \mathscr{K}(\boldsymbol{p}; \mu) + \tfrac{\alpha}{2} \left|\left| \boldsymbol{c} * UA\boldsymbol{p} - \boldsymbol{b} \right|\right|^2 \right\}. \tag{2.4}$$

   Here, based on the PSF $\boldsymbol{c}$ we determine an approximation of the image trough $A\bar{\boldsymbol{p}}$.

2. Kernel estimation based on $\boldsymbol{p}$: Determine $\bar{\boldsymbol{c}}$ as a solution of

$$\inf_{\boldsymbol{c}} \left\{ \mathscr{K}(\boldsymbol{c}; \nu) + \tfrac{\beta}{2} \left|\left| \boldsymbol{c} * UA\boldsymbol{p} - \boldsymbol{b} \right|\right|^2 \right\}. \tag{2.5}$$

   Here, based on the image $A\boldsymbol{p}$ we approximate the PSF $\bar{\boldsymbol{c}}$.

Alternating between image and kernel estimation is common in state of the art deblurring methods (see e.g. [10], [39]). In the following section we first discuss how to solve the problems (2.4) and (2.5), respectively, and then we discuss the coupling mechanism which constitutes the basis for our algorithm.

2.2.1 The Image Estimation

Throughout this section, (2.4) will be referred to as the primal problem. Recalling that both the convolution and expectation operators can be written in matrix form, we define $M = CA$ with $M \in \mathbb{R}^{Nm \times 2^N}$ for the sake of convenience.

We note, moreover that solving this problem is not a straightforward endeavour, as it is a 2^N-dimensional minimization problem. Even in the simpler case of UPC-A encoding, a barcode is composed of 95 bars, hence $\boldsymbol{p} \in \mathbb{R}^{2^{95}}$. In such a high dimensional minimization problem, attempting to compute a solution directly is infeasible and thus an alternative method must be determined to solve (2.4).

2.2.1.1 A Convex Analytic Approach to Solving the Primal Problem

We employ Fenchel-Rockafellar duality for a first simplification of the problem (2.4). To this end, we present a brief exposition of this duality scheme following [44, Example 11.41]: For $\phi : \mathbb{R}^\ell \to \mathbb{R} \cup \{+\infty\}$ its domain is dom $\phi := \{\boldsymbol{x} \in \mathbb{R}^\ell \mid \phi(\boldsymbol{x}) < +\infty\}$. Its conjugate $\phi^* : \mathbb{R}^\ell \to \mathbb{R} \cup \{\pm\infty\}$ is given by $\phi^*(\boldsymbol{y}) = \sup_{\boldsymbol{x}}\{\boldsymbol{y}^T\boldsymbol{x} - \phi(\boldsymbol{x})\}$ and the subdifferential of ϕ at $\bar{\boldsymbol{x}} \in$ dom ϕ is $\partial\phi(\bar{\boldsymbol{x}}) := \{\boldsymbol{v} \mid g(\boldsymbol{x}) \geq g(\bar{\boldsymbol{x}}) + \boldsymbol{v}^T(\boldsymbol{x} - \bar{\boldsymbol{x}})\ (\boldsymbol{x} \in \text{dom}\ \phi)\}$.

Given two lower semicontinuous convex functions with nonempty domain $k : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$, a matrix $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$ this duality scheme makes a connection between the optimization problem

$$\min_{\boldsymbol{x}} k(\boldsymbol{x}) + h(\boldsymbol{b} - A\boldsymbol{x}), \tag{2.6}$$

called the *primal* problem, with its associated *dual* problem

$$\max_{\boldsymbol{y}} \boldsymbol{b}^T\boldsymbol{y} - k^*(A^T\boldsymbol{y}) - \psi^*(\boldsymbol{y}). \tag{2.7}$$

Fenchel-Rockafellar duality now states that, under the qualification condition

$$\boldsymbol{b} \in \text{int } \left(A\text{dom k} + \text{dom h}\right),$$

the optimal value of the primal and dual problem coincide and that, given a solution $\bar{\boldsymbol{y}}$ of the dual problem, a solution of the primal can be recovered from the relation $\bar{\boldsymbol{x}} \in \partial k^*(A^T\bar{\boldsymbol{y}})$. We will from now on refer to (2.4) as the primal problem. To apply the Fenchel-Rockafellar scheme, we need to compute the conjugates of the functions in play. The conjugate of $\mathscr{K}(\cdot;\mu)$ can be computed by considering the log exp function $\log\exp : \boldsymbol{y} \mapsto \log\left(\sum_{i=1}^n \exp(y_i)\right)$ and noting that

$$\log\exp^*(\boldsymbol{q}) = \begin{cases} \sum_{i=1}^n q_i \log(q_i) & \text{for } \boldsymbol{q} \in \Delta_n \\ +\infty & \text{otherwise} \end{cases},$$

as discussed in [44, Ex. 11.12]. Observe that we can express the Kullback-Leibler entropy as

$$\mathscr{K}(\boldsymbol{p};\mu) = \sum_{i=1}^{2^N} p_i \log(p_i) - \langle \boldsymbol{p}, \log\mu\rangle.$$

As log exp is finite-valued and convex (hence lower semicontinuous and proper), the Fenchel-Moreau theorem (see e.g. [44, Theorem 11.1]) yields $\log\exp = (\log\exp^*)^*$. Therefore, also using [44, Eq. 11(3)], the conjugate of $\mathscr{K}(\cdot;\mu)$ is given by

$$\mathscr{K}^*(\boldsymbol{q};\mu) = \log\left(\sum_{i=1}^{2^N} \mu_i \exp\left(q_i\right)\right). \tag{2.8}$$

The same reference [44, Eq. 11(3)] together with [44, Ex. 11.11] also gives

$$\left(\frac{\alpha}{2}\,||\cdot||^2\right)^* = \frac{1}{2\alpha}\,||\cdot||^2 \quad (\alpha > 0).$$

We now obtain the dual problem by setting

$k = \mathscr{K}(\,\cdot\,, \mu),\ h = \frac{\alpha}{2}\,||\cdot||^2$. Hence

$$\sup_{\boldsymbol{\lambda}} \left\{ \langle \boldsymbol{b}, \boldsymbol{\lambda} \rangle - \left(\tfrac{\alpha}{2}\,||\cdot||^2 \right)^* (\boldsymbol{\lambda}) - \mathscr{K}^*(M^T \boldsymbol{\lambda}; \mu) \right\},$$

is the resulting dual problem with $\boldsymbol{\lambda}$ denoting the dual variable, $\boldsymbol{b}$ the acquired image, $M = CA$ and μ the prior. Substituting the conjugates computed previously into this expression, this problem can be written explicitly as

$$\sup_{\boldsymbol{\lambda}} \left\{ \langle \boldsymbol{b}, \boldsymbol{\lambda} \rangle - \tfrac{1}{2\alpha}\,||\boldsymbol{\lambda}||^2 - \log \left(\sum_{i=1}^{2^N} \mu_i \exp\left(M_i^T \boldsymbol{\lambda} \right) \right) \right\}. \tag{2.9}$$

We note that, on $\mathbb{R}^m$, the domain of $\frac{\alpha}{2}\,||\cdot||^2$ is the entire space, so

$$\boldsymbol{b} \in \operatorname{int}\left(M \operatorname{dom}(\mathscr{K}) + \operatorname{dom}\left(\frac{\alpha}{2}\,||\cdot||^2 \right) \right)$$

is trivially satisfied. This condition ensures that the optimal value of (2.4) is attained for at least one $\bar{\boldsymbol{p}} \in \Delta_{2^N}$, this is a property of the duality scheme that has been used. Note moreover that $\bar{\boldsymbol{p}}$ is guaranteed to be an element of the unit simplex as otherwise the Kullback-Leibler divergence takes on a value of infinity. Similarly, since

$$\mathbf{0} \in \operatorname{int}\left(M^T \operatorname{dom}\left(\frac{\alpha}{2}\,||\cdot||^2 \right)^* - \operatorname{dom}\left(\mathscr{K}^* \right) \right),$$

the optimal value of (2.9) is also attained for at least one $\bar{\boldsymbol{\lambda}}$. Together, these conditions ensure that these problems share the same finite optimal value. Moreover, given a solution $\bar{\boldsymbol{\lambda}}$ of (2.9) one can perform primal-dual recovery via

$$\bar{\boldsymbol{p}} = \nabla \mathscr{K}^*(M^T \bar{\boldsymbol{\lambda}}; \mu), \tag{2.10}$$

which is another property of this duality scheme. The previous equation is formulated in terms of the gradient, as $\mathscr{K}^*$ is differentiable at every point of its domain such that its subgradient at a given point is a singleton, namely its gradient at that point by [45, Thm. 25.1].

One of the advantages of this dual formulation is that solving the primal problem, a minimization problem in $\boldsymbol{p} \in \mathbb{R}^{2^N}$, is now analogous to solving the dual problem, a maximization problem in $\boldsymbol{\lambda} \in \mathbb{R}^{Nm}$ and recovering a solution to the primal problem via (2.10). This foray into Fenchel-Rockafellar duality has therefore yielded a tremendous dimensionality reduction. Despite this amelioration, solving the dual problem is still intractable as the conjugate of the entropy contains an immense sum over 2^N elements and the matrix M has dimensions $Nm \times 2^N$ hence it cannot be stored in memory for computations for large N.

2.2.1.2 Exploiting the Probabilistic Structure of the Dual Problem

Recall that, by definition, $M = CA$ where A is a $N \times 2^N$ matrix whose columns consist of the binary sequences of length of N. In particular, A_i^T is the i-th binary sequence in some arbitrarily chosen ordering of $\{0,1\}^N$. The sum in (2.9) can therefore be rewritten as

$$\sum_{i=1}^{2^N} \mu(A_i^T) \exp(\langle A_i^T, C^T \boldsymbol{\lambda} \rangle). \tag{2.11}$$

This expression equals $\mathbb{E}_\mu\left[\exp\langle C^T\boldsymbol{\lambda}, \boldsymbol{X}\rangle\right]$ with $\mathbb{E}_\mu[h(\boldsymbol{X})]$ denoting the expected value of the random variable $h(\boldsymbol{X})$, where $\boldsymbol{X}$ has probability distribution μ [46, Def. 1 p.141]. This expectation is simply the moment generating function (MGF) $M_{\boldsymbol{X}}$ of $\boldsymbol{X}$ evaluated at $C^T\boldsymbol{\lambda}$. By assumption, $\boldsymbol{X} = (X_1, \ldots, X_N)$ where the X_i are independent Bernoulli random variables. Therefore, using [46, Thm. 5 p.155], the MGF in (2.11) can be written as

$$\prod_{i=1}^{N} M_{X_i}(C_i^T\boldsymbol{\lambda}). \tag{2.12}$$

The MGF of a Bernoulli random variable is made explicit in [46, Sect. 5.2.2 p.180]. Hence (2.12) is equivalent to

$$\prod_{i=1}^{N}(1-\rho_i+\rho_i\exp(C_i^T\boldsymbol{\lambda})),$$

where ρ_i is the probability that the i-th bar in $\boldsymbol{x}$ is white. Replacing the sum in (2.8) with this product yields the following expression for the conjugate of the Kullback-Leibler divergence:

$$\mathscr{K}^*(M^T\boldsymbol{\lambda};\mu) = \sum_{i=1}^{N}\log\left(1-\rho_i+\rho_i\exp\left(C_i^T\boldsymbol{\lambda}\right)\right). \tag{2.13}$$

This expression is easily evaluated given some $\boldsymbol{\lambda}$. Using this form for $\mathscr{K}^*$ renders the dual problem (2.9) tractable via standard numerical optimization algorithms. However, we recall that $\bar{\boldsymbol{p}} \in \Delta_{2^N}$, hence determining an expression for $\bar{\boldsymbol{p}}$ is infeasible regardless of the fact that we can solve the dual problem. We opt therefore to recover the original image directly from $\bar{\boldsymbol{\lambda}}$.

2.2.1.3 Determining the Original Image From the Argmax of the Dual Problem

In the following we seek to compute the expectation of (2.10) which serves as the estimate of the original image. Performing this calculation naively leads to (2.14) which includes the large matrix M. Thus we use the probabilistic argument of the previous section to derive an analogous expression (2.15) which can be computed explicitly.

Given an optimal solution $\bar{\boldsymbol{p}}$ of the primal problem (2.4), we can recover an estimate of the original image $\boldsymbol{x}$ via $\bar{\boldsymbol{x}} = A\bar{\boldsymbol{p}}$. We refer to $\bar{\boldsymbol{x}}$ as an estimate of $\boldsymbol{x}$, as the penalty function (2.2) does not guarantee that $C\bar{\boldsymbol{x}} = \boldsymbol{b}$. Using our expression from (2.10), we can write

$$\bar{\boldsymbol{x}} = A\nabla\mathscr{K}^*(M^T\bar{\boldsymbol{\lambda}};\mu).$$

We first write $\nabla\mathscr{K}^*(M^T\bar{\boldsymbol{\lambda}};\mu)$ componentwise yielding

$$\left[\nabla\mathscr{K}^*(M^T\bar{\boldsymbol{\lambda}};\mu)\right]_k = \begin{cases} \frac{\mu_k \exp(M_k^T\bar{\boldsymbol{\lambda}})}{\sum_{i=1}^{2^N}\mu_i\exp(M_i^T\bar{\boldsymbol{\lambda}})} & \text{for } k\in I, \\ 0 & \text{otherwise} \end{cases}$$

such that $\bar{\boldsymbol{x}}$ can be written componentwise by multiplying the previous expression by A, i.e.

$$\bar{x}_k = \frac{\sum_{i=1}^{2^N} a_{ki}\mu_i\exp\left(M_i^T\bar{\boldsymbol{\lambda}}\right)}{\sum_{i=1}^{2^N}\mu_i\exp(M_i^T\bar{\boldsymbol{\lambda}})}. \tag{2.14}$$

Here a_{ij} is the value in the i-th row of the j-th column of A. We now consider

$$\nabla\log\left(\sum_{i=1}^{2^N}\mu_i\exp\left\langle A_i^T,\cdot\right\rangle\right),$$

the k-th component of which is simply

$$\frac{\sum_{i=1}^{2^N} a_{ki}\mu_i \exp\langle A_i^T, \cdot \rangle}{\sum_{i=1}^{2^N} \mu_i \exp\langle A_i^T, \cdot \rangle}.$$

Evaluating this expression at the point $C^T\bar{\boldsymbol{\lambda}}$ demonstrates that it is equivalent to (2.14) (since $M = CA$) with the advantage that we can simplify it using the same probabilistic argument that was derived previously. Thus, $\boldsymbol{x}$ can be estimated whilst bypassing the matrix A via:

$$\bar{\boldsymbol{x}} = \nabla \sum_{i=1}^{N} \log(1 - \rho_i + \rho_i \exp(C_i^T \bar{\boldsymbol{\lambda}})). \tag{2.15}$$

Consequently, once the argmax of the dual problem has been determined we can estimate the original image directly, without determining A or the PMF $\bar{\boldsymbol{p}}$.

Each step of the image estimation has now been made computationally tractable.

2.2.1.4 A Summary of the Steps for Image Estimation

The previous developments can be summarized in the following procedure for deblurring an image for which the convolution kernel is known or approximated.

First, a prior μ must be formed. This prior will assign a probability of being white to each bar, so encoding the symbolic constraints of the barcode of interest into the prior will ensure that the solution to (2.4) is at least correct on these bars. Other types of priors and a more detailed discussion of the construction of this symbolic prior is found in the results section.

Next, the dual problem (2.9) with the expression for the conjugate of the Kullback-Leibler divergence given in (2.13) is solved. This step can be performed efficiently by standard optimization software. Our choice of algorithm is discussed in the results section.

Finally, an estimate of the initial image is determined via (2.15). The resulting image will not be identical to $\boldsymbol{x}$ due to rounding errors and the choice of tolerance in the optimization algorithm.

We choose to subsequently perform a thresholding step to guarantee that all of the segments of the barcode are either 0 or 1. This step ensures that the barcode will be readable if it was accurately deblurred and thus the information encoded in the original image can be extracted if our method has succeeded.

2.2.2 The Kernel Estimation

We now focus on solving (2.5), keeping in mind that it shares a similar paradigm to (2.4). Again, since the convolution is linear and discrete, $\boldsymbol{c} * UA\boldsymbol{p}$ can be written as $X\boldsymbol{c}$. We enforce that $\boldsymbol{c} \in \mathbb{R}^{Nm}$, as the convolution kernel should not be larger than the size of the image. Thus, $X \in \mathbb{R}^{95m \times 95m}$ such that (2.5) can be solved directly as a constrained minimization problem, since it is not as high-dimensional a problem as (2.4). However mimicking the previous foray into Fenchel-Rockafellar duality will yield a simpler unconstrained analogue to this primal problem.

2.2.2.1 Advantages of the Dual Formulation

The dual problem to (2.5) is nearly identical to (2.9), hence we simply state the dual problem using the same duality scheme:

$$\sup_{\boldsymbol{\lambda}} \left\{ \langle \boldsymbol{b}, \boldsymbol{\lambda} \rangle - \tfrac{1}{2\beta} \left|\left|\boldsymbol{\lambda}\right|\right|^2 - \mathscr{K}^*(X^T \boldsymbol{\lambda}; \nu) \right\}. \tag{2.16}$$

The same argument used to show that (2.4) and (2.9) share the same optimal value and that this solution is attained in both problems implies that (2.5) and (2.16) satisfy the same property. Consequently, the argmin $\bar{\boldsymbol{c}}$ of (2.5) is given by

$$\bar{\boldsymbol{c}} = \nabla \left(\log \left(\sum_{i=1}^{95m} \nu_i \exp(X_i^T(\cdot)) \right) \right) (\bar{\boldsymbol{\lambda}}), \tag{2.17}$$

in the same vein as (2.10). Here, $\bar{\boldsymbol{\lambda}}$ denotes the argmax of the unconstrained dual problem.

Regularizing problem (2.5) via a Kullback-Leibler divergence term guarantees that the optimal kernel estimate $\bar{\boldsymbol{c}}$ is nonnegative and that its elements sum to 1 as explained. This property is characteristic of any normalized blur kernel which is precisely the type of PSF that occurs in image acquisition. Moreover, ν is used to limit the size of the considered kernel by setting all but a square of the desired width centred at the middle of its matrix representation to 0 and setting uniform values summing to 1 in this square. Hence, adopting a coarse-to-fine approach as in [21], [48] is analogous to simply increasing the size of the kernel being considered at each step which can be accomplished by varying ν.

2.2.3 The Algorithm

We summarize the development of the prior probability distributions and outline an algorithm that implements our blind deblurring method.

Barcode symbologies impose constraints which typically fix certain segments of an image. We outline a method to generate a prior which captures these constraints. Recall that we have modeled a barcode by a vector of N independent Bernoulli random variables. The distribution of a Bernoulli random variable is completely determined by a single parameter ρ. As a barcode is a vector of independent Bernoulli random variables, its distribution is determined by N parameters ρ_i as in equation (2.13) above, where each ρ_i represents the probability that the i-th bar in a barcode is white. This suggests that a natural prior distribution μ has the following probability mass function

$$\mu(\boldsymbol{x}) = \prod_{i=1}^{N} \rho_i^{x_i}(1-\rho_i)^{1-x_i}$$

where x_i is the i-th bar of $\boldsymbol{x}$. We let $\rho_i = 0$ if the i-th bar is fixed as black by the barcode symbology, $\rho_i = 1$ if the i-th bar is fixed as white, and $\rho_i = \frac{1}{2}$ if the i-th bar is not fixed by the symbology. (This choice reflects our lack of prior knowledge of the state of the i-th bar when it is not fixed by a symbology.)

The algorithm is summarized with references to the relevant equations in Alg. 1. The algorithm features two loops. The outer loop iterates through a set of fixed widths for our kernel estimate. The inner loop repeatedly solves problems (2.4) and (2.5).

We begin by setting $i = 1$ and hence initially assume that the size of the convolution kernel is $2i + 1 = 3$. We take our initial best guess of the true barcode to be the image which is black or white in regions which are fixed as such by the relevant barcode symbology, and gray in all other regions. (See the lower image in

Fig. 2–1 for an example of such an initial guess for UPC barcodes.) We substitute this image for $A\bar{\boldsymbol{p}}$ in the kernel estimation step (2.5), solve its dual problem (2.16), and then compute (2.17) to obtain our initial estimate $\bar{\boldsymbol{c}}$ of the true convolution kernel. We then use this estimate $\bar{\boldsymbol{c}}$ to solve the image estimation problem (2.4) via the methods outlined in the previous sections and obtain a first estimate $\bar{\boldsymbol{x}}$ of the true barcode.

This estimated barcode is subsequently read by a software barcode scanner. If the barcode is readable, the algorithm terminates successfully. If the algorithm does not terminate after the first iteration of the inner loop, we continue to iterate through the alternating kernel estimation and image estimation steps, each time substituting our image estimate for $A\bar{\boldsymbol{p}}$ in the kernel estimation step, and our subsequent kernel estimate for $\boldsymbol{c}$ in the image estimation step. This yields progressive improvements in the estimates of the convolution kernel and the image.

If the barcode is still not readable after a fixed number of iterations (5 in our implementation), we infer that the initial width $2i + 1$ of the estimated convolution kernel was too small. We therefore increment i and iterate through the inner loop once again. We iterate through the outer loop until the width of the convolution kernel reaches that of the image. If the barcode is still not readable at this point, the algorithm terminates unsuccessfully.

We have therefore set up a framework that permits blind deblurring for both QR and UPC-A barcodes that effectively utilizes prior knowledge of their respective structure.

Algorithm 1 Entropic Blind Deblurring

Require: Blurred image $\boldsymbol{b}$, prior μ

  for $i = 1$ to (width of $\boldsymbol{b}$)/2, **do**

    $\bar{\boldsymbol{x}} \leftarrow \mu$

    width $\leftarrow (2i + 1)$

    for $j = 1$ to 5 **do**

      $\bar{\boldsymbol{c}} \leftarrow$ (2.17) at argmax of (2.16) with size width

      $C \leftarrow \bar{\boldsymbol{c}}$ written as a convolution matrix

      $\bar{\boldsymbol{x}} \leftarrow$ (2.15) at argmax of (2.9) using expression (2.13)

      threshold $\bar{\boldsymbol{x}}$

      if $\bar{\boldsymbol{x}}$ is readable **then**

        return $\bar{\boldsymbol{x}}$

      end if

    end for

  end for

  return $\bar{\boldsymbol{x}}$

Ensure: approximated image $\bar{\boldsymbol{x}}$

2.3 Results

In the following, we will discuss some of the results obtained while testing our method. We will refer to graphs in the online supplementary material for more details of our experiments. First, we explain the methodology used to generate all of the relevant quantities used for testing.

2.3.1 Implementation Details

We began by generating barcodes in both the UPC-A and QR symbologies. In the case of UPC-A barcodes, we considered 200000 valid barcodes from the Open Product Data database [2]. In the case of QR codes, we chose various phrases to encode and used an online QR code generator to get the relevant images. Each phrase was encoded in all four levels of error tolerance supported by the symbology, namely low (7% tolerance), medium (15% tolerance), quartile (25% tolerance) and high (30% tolerance) as explained in [4].

In order to blur images synthetically, normalized PSFs were generated. For Gaussian convolution kernels, this process is straightforward in both the 1D and 2D cases, as for a PSF of width k one need only sample a 1D or 2D Gaussian function with mean 0 at k points on an interval centred at 0. For box blurs in the 1D case, we simply initialize all k points of the PSF to the value $\frac{1}{k}$. For linear motion blurs in the 2D case, the motion blur is simply a line through the centre of the kernel at a prescribed angle. Examples of these kernels are compiled in Fig. 2–2.

Figure 2–2: A graphical depiction of the the types of kernel used with width 5. The left kernel generates Gaussian blur and the right one generates linear motion blur at an angle of $\frac{\pi}{4}$. Note that they are normalized such that the intensity values sum to 1.

Moreover, general motion blur kernels such as those in [32] were tested on QR codes, yielding adequate results on barcodes of a reasonable size relative to the kernel as demonstrated in Fig. 2–3 (other examples are presented in Fig. Sup. 1 of the supplementary material). In what follows we concentrate on Gaussian and linear motion blur as their testing can be readily automated.

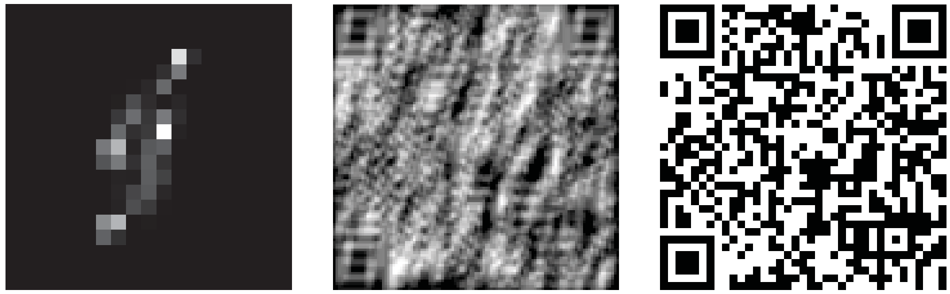


Figure 2–3: The leftmost image is a general motion blur kernel. The center image is the corresponding blurred QR code. The rightmost image is the result obtained by applying our method which is readable. The kernel was normalized such that it's intensity values summed to 1 prior to blurring and the QR code was upscaled by a factor of 3.

The barcodes are upscaled prior to convolution in order to access a greater range of blurring magnitudes. Indeed, even blur kernels of width 3 (the smallest tested) produce dramatic quantities of blur when one bar is the size of one pixel as demonstrated in Fig. 2–4. Thus, upscaling allows us to consider more realistic levels of blurring.

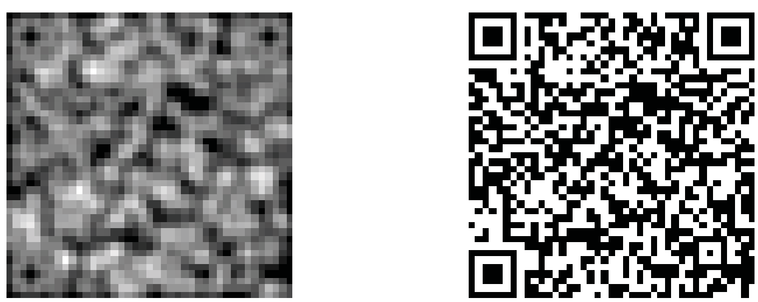


Figure 2–4: This figure presents an image blurred with the smallest 2D Gaussian kernel on the left hand side performed on a QR code which has not been upscaled. Note that this magnitude of blur is rather large, hence the need for upscaling the image prior to convolution. The right hand side is the result we obtain upon applying our blind deblurring algorithm with a symbolic prior. The right hand QR code can be read by any conventional QR code reader.

In preliminary testing, the PSF was turned into a convolution matrix C by examining the result of the discrete convolution as demonstrated in [20, Fig. 4.7] and inferring a matrix that performs the same operation. A similar method was used to determine the matrix X in (2.16). We worked with the 0 boundary conditions in the formation of these matrices, as it is simpler to construct the convolution matrix in this case. In reality the barcode is encased in a white quiet zone so one can simply invert the colours of the captured signal such that the 0 boundary condition in the inverted image is equivalent to a 1 boundary condition in the original. Hence, in a real image, the 0 boundary condition would be sufficient, as the barcode would not be convolved with data outside this quiet zone for blur kernels of a reasonable size.

Thereafter, the convolutions were performed by means of the fftconvolve method from the scipy python library [1]. The advantage to this approach is that it is both faster and less memory intensive than forming the convolution matrix and storing it in memory. We make precise that the Hermitian adjoint of the discrete convolution operator is obtained by performing a discrete convolution with one of the arrays reversed about its axes as discussed in [14, Sect. 5.1.1]. Hence, passing from the matrix methodology to this one is akin to replacing the transposed matrices in the dual problems and the recovery of the solutions to the primal problems by the adjoint of the corresponding convolution.

We employ a downscaling step once we have estimated our image by averaging together the blocks of pixels that correspond to one pixel once upscaled. We subsequently round the pixel intensities to the nearest integer as discussed previously. The utility of this step is highlighted in Fig. 2–5. The critical task of decoding the QR estimate is delegated to the Zbar Python implementation provided by [5] which permits automation for checking readability of the iterates during testing. Hence, if a barcode is readable post thresholding, we terminate the algorithm and return the data which has been decoded. We use both Zbar and various smartphone applications in order to compare the performance of our algorithm to state of the art QR code scanners.

Figure 2–5: This figure demonstrates the utility of the thresholding step. The left hand side is the deblurred image prior to the downscaling and rounding. The original image was subjected to linear motion blur with large kernel size at an angle of $-\frac{\pi}{4}$. Note that some degree of distortion along a diagonal axis remains prior to thresholding and downscaling. The right side displays the barcode post-thresholding; it is readable.

We equally examine how our method performs in the presence of noise. We consider both additive Gaussian and salt and pepper noise throughout, as they are the most common in practice. To generate Gaussian noise, we generate a matrix of the same size as the image to which each pixel is associated a random sample from a normal distribution with prescribed variance and simply add both matrices to add noise to the image. Similarly, to generate salt and pepper noise of a given percentage, we generate at each pixel a random real number between 0 and 1.

If this number is higher than our prescribed percentage, we add nothing to the image. If it is lower, we randomly choose between 0 or 1; if 0 is selected, the pixel is made black in the image, if 1 is selected, it is made white. Visuals are provided below to better illustrate the magnitude and types of noise.

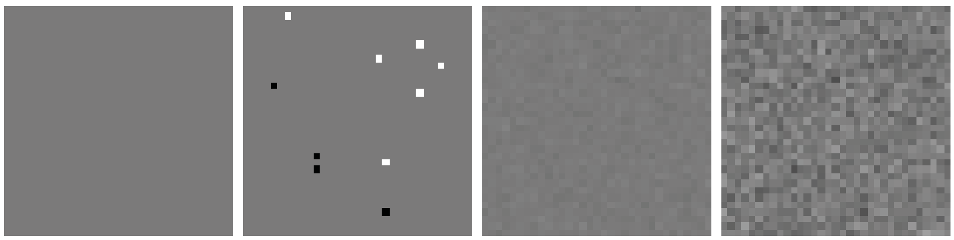


Figure 2–6: A demonstration of types and magnitudes of noise tested. The leftmost image represents the noiseless case, the centre-left image depicts 1% salt and pepper noise, the centre-right image is 0.01 variance Gaussian noise and the rightmost image is 0.05 variance Gaussian noise.

All that remains before our algorithm can be tested is to make explicit how the various priors are generated in the different symbologies. Without considering the intricacies of the various encoding schemes, it is possible to form a uniform prior in which every bar is given a probability of 0.5 of being white. It is obvious that such a prior will not perform as well as a symbolic one which encodes all of the fixed modules within a symbology and assigns a uniform probability of 0.5 to the bars that have not been fixed. Note that QR codes have varying size, hence a prior must be generated for the various sizes. In the UPC-A case, a third prior was equally constructed in which our library of more than 200000 UPC-A barcodes was analyzed and each bar was given a probability reflecting the percentage of barcodes of the library having that bar white, this prior is referred to as empirical. Testing for UPC-A barcodes was performed first and it was deemed that the symbolic and empirical priors yield similar performance. As no tangible performance improvements were expected, we did not construct an empirical prior for QR codes.

With this framework in place to generate blurred and noisy barcodes, we are ready to test the performance of our method.

2.3.2 Non-Blind Deblurring

As mentioned previously, non-blind deblurring can be done by performing the image estimate step with the exact kernel $\boldsymbol{c}$ known. We wish to determine the performance of our method for this step and examine the effects of prior choice. Moreover, we wish to quantify the flexibility of our method with respect to the presence of noise in the acquired image.

2.3.2.1 Non-Blind Deblurring for UPC-A barcodes

In order to gauge the performance of this method for non-blind deblurring of 1D barcodes, we begin by observing its noiseless performance.

We do so by choosing 5 random barcodes, upscaling them by a factor of 5 and blurring them with progressively larger blur kernels until the barcode was no longer successfully readable when using the method. We repeat this process for every prior as well as with both Gaussian and box blurs. The results of the five barcodes are then averaged in order to provide a general idea of the non-blind, noiseless performance. We set $\alpha = 1000000$ in order to give great importance to the error term thus incentivizing the proximity of $C\mathbb{E}_{\boldsymbol{p}}[\boldsymbol{x}]$ to $\boldsymbol{b}$. The results of this test are shown in Table 2–1.

Prior	**Type of Blur**	**Cut-off Width**
Uniform	Gaussian	173.0
	Box	129.8
Symbolic	Gaussian	259.8
	Box	210.6
Empirical	Gaussian	259.4
	Box	259.4

Table 2–1: This table compares the performance of the various priors in the presence of both types of blur. The cut off width is the width of the kernel at which the method first fails.

We note that the empirical prior outperforms the symbolic prior in the case of box blur specifically and that they both outperform the uniform prior by a significant margin. Clearly, the structure encoded in the non-uniform priors account for their superior performance. Moreover, despite the assumption that the empirical prior should encode some form of correlation between the various bars, it performs essentially identically to the symbolic prior. Therefore, the intrinsic symbology of the barcode appears to take precedence over any additional structure gained by a statistical learning approach. Finally, the blur widths at which these priors first fail are so large that they would not occur in real life applications, this is a testament to the strength of our method in the case of noiseless image acquisition for UPC-A barcodes.

As for the performance of this method as it pertains to a noisy image acquisition process, we determine a cutoff variance for Gaussian noise for various blur widths before which we can read all of the blurred and noisy barcodes generated.

Again, we pick 5 random barcodes and begin by blurring with blur width 3 Gaussian noise. Next, we iteratively increase the variance of the additive Gaussian noise that is added to the image until we first fail to successfully deblur the barcode. We then increase the blur width by 2 and repeat this procedure until we reach a width such that even the lowest variance noise (0.005) cannot be read. At this point, we repeat the entire process with a box blur.

We note that salt and pepper noise was not tested for the UPC-A symbology, as in our one-dimensional formulation, this type of noise is equivalent to changing the color of the entire bar. In practical applications this noise would only effect a segment of a bar which our model is not designed to account for.

In these tests, $\alpha = 1000$ in order to account for the fact that $\boldsymbol{b}$ will not be in the range of C. These tests are performed with the three different priors and the results are compiled in Fig. Sup. 2 of the supplementary material. We note again that the symbolic and empirical priors outperform the uniform prior by a significant margin for larger blur widths.

We note again that the symbolic and empirical priors outperform the uniform prior by a significant margin for larger blur widths.

2.3.2.2 Non-Blind Deblurring for QR Codes

We test our method for QR codes by picking five of the encoded messages and determining the blur width at which our method first fails to recover the information contained in the QR code. The barcodes are upscaled by a factor of 3 in order to consider a greater range of blurring kernels. We proceed similarly to the UPC-A testing, however, rather than considering different types of priors, we compare the different levels of error tolerance and use only a symbolic prior. We equally compare our method to the ZBar algorithm by attempting to read the blurred barcode prior to deblurring it. If it fails to read, we deblur the barcode using our method and verify if the image is now readable. Throughout these tests, we set $\alpha = 10000000$ in order to enforce the constraint on the mean. The averaged results of this testing are shown in Table 2–2.

Error Tolerance	**Type of Blur**	**Cut-off Width (ZBar)**	**Cut-off Width (Ours)**
Low	Gaussian	5.0	29.4
	Motion	6.6	30.6
Medium	Gaussian	5.4	32.2
	Motion	7.0	37.8
Quartile	Gaussian	5.4	33.8
	Motion	7.0	50.2
High	Gaussian	5.8	35.4
	Motion	7.0	65.8

Table 2–2: This table compares the performance of the various error tolerances in QR codes in the presence of different types of blur. The cut off width is the width of the kernel for which the method first fails.

We note in particular that the algorithm performs noticeably better in the presence of motion blur as compared to Gaussian blur. Letting l denote the width of the blur, motion blur kernels yield convolutions such that the value at one point is determined by the values of l points, whereas in the Gaussian case, l^2 points are considered. Hence, Gaussian blurs produce more dramatic blurring for the same size of kernel, so this observation is reasonable. The various error tolerances perform as expected, with the low tolerance performing the worst and the high tolerance performing the best. We note, moreover, that with every error tolerance, we are able to successfully recover images that are blurred far beyond what can be considered normal for real life applications.

In order to gauge the effects of additive noise on this method, we again approach the problem in a similar fashion to the 1-dimensional problem. The effects of salt and pepper noise can now be studied, as QR codes include a degree of error correction. Hence, even if the noise does not permit us to reconstruct the same QR code as the original, the deblurred image may still be read. We compile our results in Fig. Sup. 3 of the supplementary material, here α was set at 750 to promote flexibility.

We note that, as expected, our method is more robust to Gaussian noise as it is less dramatic, changing the value of almost every module by a small amount as opposed to salt and pepper noise which makes a certain number of modules black or white. In comparing the various error tolerances, it is clear that the low level performs much worse than the others which perform almost identically. Regardless, the amounts of noise considered are far larger than those that would occur in realistic applications hence, this method performs adequately for reasonable noise levels.

Finally, the method employed by ZBar performs exceptionally well for very small quantities of blur. However our method greatly outperforms it for every blur width other than the smallest one.

2.3.3 Blind Deblurring

We now test the performance of our method as it pertains to the problem of blind deblurring.

2.3.3.1 Blind Deblurring for UPC-A Barcodes

We wish to quantify the performance of our blind deblurring method in the presence of additive noise at various blur widths. We employ an identical series of tests to those used in the non-blind case. The only difference is that we utilize our non-blind method to deblur the resulting images. The noiseless results are shown in Table 2–3.

Prior	**Type of Blur**	**Cut off Width**
Uniform	Gaussian	19.4
	Box	19.0
Symbolic	Gaussian	76.6
	Box	75.0
Empirical	Gaussian	74.2
	Box	89.0

Table 2–3: This table compiles the blur widths at which our method first fails to recover the information contained in the UPCA barcode when using the blind deblurring method.

Comparing these results to those obtained with the non-blind method reveals, unsurprisingly, that the blind method is not as robust to the size of the blur width. This is to be expected, as the blur kernel is being estimated rather than provided. However, we note that these results are certainly adequate for real life conditions.

We equally examine the case in which noise is present in the image acquisition process by performing tests identical to those used in the non-blind section. The results are compiled in Fig. Sup. 4 of the supplementary material.

We note that the symbolic and empirical priors perform nearly identically. Notably, they greatly outperform their uniform counterpart. Moreover, this blind method actually appears to outperform the non-blind method. This slight improvement may be due to the fact that the non-blind method enforces the use of the original convolution kernel of which $\boldsymbol{b}$ is often not in the range of. The blind deblurring method offers greater flexibility in terms of the convolution kernel and hence with the thresholding step we employ, it is conceivable that performance is improved.

2.3.3.2 Blind Deblurring for QR Barcodes

In the noiseless case, we employ the same tests as we did in non-blind deblurring. The results are shown in Table 2–4.

As expected, the results of this method are not as good as those obtained with its non-blind analogue. They are however acceptable for real world image acquisition.

We equally perform the same tests as those considered in the non-blind case and compile the results in Fig. Sup. 5 of the supplementary material.

The various levels of error tolerance perform as expected and we note that the flexibility with respect to both the Gaussian and salt and pepper noise is

Error Tolerance	**Type of Blur**	**Cut-off Width (ZBar)**	**Cut-off Width (Ours)**
Low	Gaussian	5.0	9.0
	Motion	6.6	25.0
Medium	Gaussian	5.0	10.2
	Motion	7.0	27.0
Quartile	Gaussian	5.4	10.2
	Motion	7.0	33.8
High	Gaussian	5.8	10.6
	Motion	7.0	31.8

Table 2–4: This table compares the performance of the various error tolerances in QR codes in the presence of different types of blur. The cut off width is the width of the kernel for which this blind method first fails.

considerable, as they greatly surpass what could be considered reasonable at low widths of blur.

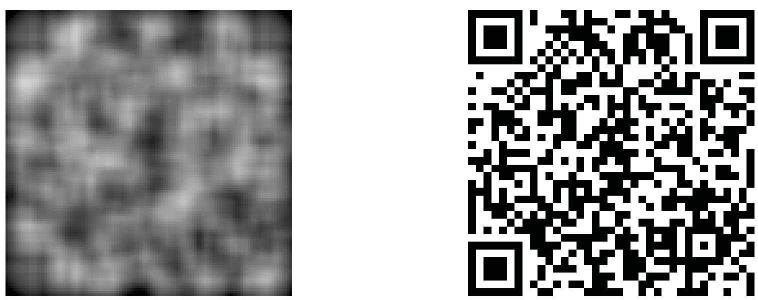


Figure 2–7: This figure demonstrates the strength of our method even in the case where very large Gaussian blur is present. The right hand side is the result we obtain upon applying our blind deblurring algorithm with a symbolic prior. The right hand QR code can be read by any conventional QR code reader.

2.3.4 Possible Improvements

Given that our aim throughout was to explain our method and demonstrate the power of symbology in barcode reconstruction, we did not explore some avenues for improving it. We list some improvements for those who wish to

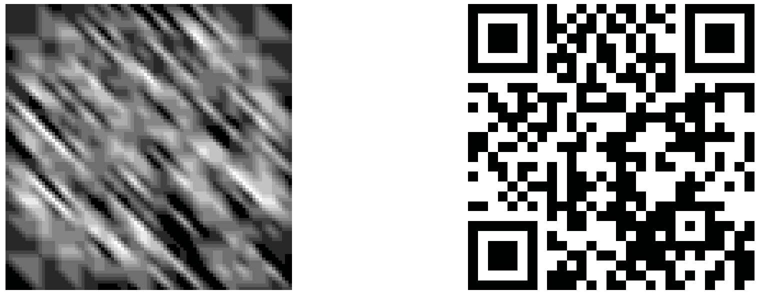


Figure 2–8: This figure presents a blurred image on the left hand side. The blurring is a linear motion blur of kernel size 9 with angle $-\frac{\pi}{4}$ which has been performed on a 29×29 pixel QR code upscaled to 87×87 pixels. The right hand side is the result we obtain upon applying our blind deblurring algorithm with a symbolic prior. The right hand QR code can be read by any conventional QR code reader.

Figure 2–9: On the left, a QR code with an upscaling factor of 3, subject to width 11 motion blur at an angle of $\frac{\pi}{4}$ with 1% salt and pepper noise is presented. The right hand side is the result we obtain upon applying our blind deblurring algorithm with a symbolic prior. The right hand QR code can be read by any conventional QR code reader.

implement them. First, we did not attempt to optimize solving either of the dual problems of interest, opting rather to implement the stock l-bfgs algorithm from the scipy package [54]. A further analysis of these two problems could yield a tailor-made approach to solving these problems that outperforms our current approach. This modification could significantly improve the run time of the

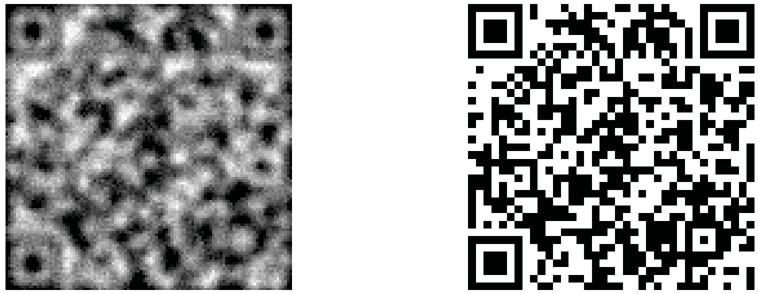


Figure 2–10: On the left, a QR code with an upscaling factor of 3, subject to width 7 Gaussian blur with 0.05 variance Gaussian noise is presented. The right hand side is the result we obtain upon applying our blind deblurring algorithm with a symbolic prior. The right hand QR code can be read by any conventional QR code reader.

algorithm, but would not likely improve its accuracy in terms of reproducing the original barcode.

Next, our preliminary tests were performed in the Python programming language using the Jupyter Notebook application as well as in the Matlab computing environment. All final testing was performed in Python, which is known to be slower than C/C++ especially as speed was not the main consideration when the code was written, as explained in [3]. One could conceivably decrease runtimes in a significant fashion by rewriting the code in a different language or simply by optimizing the code already written.

Moreover, the parameters α and β used during testing were determined empirically during testing. A more detailed analysis of these parameters may prove fruitful in enhancing the performance of this method depending on the context in which it is used.

Furthermore, recall the algorithm terminates either when the approximate barcode has been read or when the iterations terminate. Thus, no supplementary processing is performed on the intermediate approximations of the image before attempting to read them. In the case of UPC-A barcodes, we are simply verifying each segment against a dictionary of digits, hence if even a single bar fails the entire barcode will fail to read and the algorithm will continue to run. It would be possible to implement a method that would attempt to correct errors in the barcode prior to the reading step which could improve performance. Moreover, determining optimal values for the number of iterations to perform and verifying if one could increase the size of the blurring kernel quicker could significantly reduce runtime.

2.4 Conclusion and Real World Applications

Throughout this article, our focus here has been on developing and testing a novel entropy-based method to solve a difficult ill-posed problem: blind and non-blind barcode deblurring of barcodes. The strength of our method lies in its effective exploitation of the symbology innate to barcodes. Using various barcode reading software packages, our results were tested on simulated images with moderate amounts of noise and large amounts of blurring. A natural question is to what extend our method can be used for real life camera images, i.e. industrial applications. Such applications are in no way immediate from our current set up. Note that our method depends heavily on the symbology and it is assumed that the scaling is uniform throughout the image; thus any implementation would require a significant amount of preprocessing to obtain data to which our algorithm can be directly applied. This dependence on symbology suggests that our method is ill-suited for situations where the blur is not uniform. While the general details of such preprocessing are beyond the scope of this article, we include, as proof of concept for the applicability of this method to real life situations, one example in Fig. 2–11. Here, we present a picture of a barcode with significant out of focus blur. Zbar and our smart phones are unable to read this picture. Fig. 2–12 is the readable barcode obtained by applying our method after isolating the barcode from the image. As explained in Section 2.3.1, the boundary conditions were accounted for by inverting the colours of the signal before the method was applied.

Figure 2–11: Out of focus image of a QR code.

Acknowledgment

Gabriel Rioux and Christopher Scarvelis would like to acknowledge funding they received from the McGill SURA and NSERC USRA respectively which permitted their contribution to this research. Rustum Choksi and Tim Hoheisel were partially supported by NSERC (Canada) Discovery Grants. The authors

Figure 2–12: Result of applying our method to a processed version of Fig. 2–11.

would like to thank Jerome Gilles for helpful comments on a previous draft as well as the anonymous reviewers for their useful feedback.

REFERENCES

[1] Fftconvolve. URL `https://docs.scipy.org/doc/scipy-0.16.1/reference/generated/scipy.signal.fftconvolve.htm`.

[2] Open product data website. URL `https://product-open-data.com/`.

[3] Python speed. URL `https://wiki.python.org/moin/PythonSpeed`.

[4] Qr code. URL `https://en.wikipedia.org/wiki/QR_code`.

[5] Zbar bar code reader. URL `http://zbar.sourceforge.net/`.

[6] J. F. Bercher, G. L. Besnerais, and G. Demoment. The Maximum Entropy on the Mean Method, Noise and Sensitivity. In Skilling, John and Sibisi, Sibusiso, editor, *Maximum Entropy and Bayesian Methods*, pages 223–232. Springer Netherlands, 1996.

[7] J. Borwein and A. Lewis. Partially Finite Convex Programming, Part I: Quasi Relative Interiors and Duality Theory. *Mathematical Programming*, 57:15–48, 1992.

[8] J. Borwein and A. Lewis. Partially finite convex programming, part II: Explicit lattice models. *Mathematical Programming*, 57:49–83, 1992.

[9] L. Chen. A Directed Graphical Model for Linear Barcode Scanning from Blurred Images. *ACCV 2013*, pages 524–535, 2013.

[10] H. Cho, S. Lee, and J. Wang. Text Image Deblurring Using Text-Specific Properties. *ECCV 2012*, pages 524–537, 2012.

[11] S. Cho and S. Lee. Fast Motion Deblurring. *ACM Trans. Graph.*, 28(5): 145:1–145:8, 2009.

[12] R. Choksi, Y. van Gennip, and A. Oberman. Anisotropic Total Variation Regularized L^1-Approximation and Denoising/Deblurring of 2D Bar Codes. *Inverse Problems and Imaging*, 5(3):591–617, 2011.

[13] C.-H. Chu, D.-N. Yang, and M.-S. Chen. Image stabilization for 2d barcode in handheld devices. *ACM Multimedia 2007*, pages 697–706, 2007.

[14] J. F. Clærbout. *Earth Soundings Analysis: Processing versus Inversion.* Stanford University Press, 1993.

[15] D. Dacunha-Castelle and F. Gamboa. Maximum d'entropie et problème des moments. *Annales de l'IHP Probabilités et statistiques*, 26(4):567–596, 1990.

[16] M. Dodangeh, I. N. Figueiredo, and G. Gonçalves. Spatially adaptive total variation deblurring with split Bregman technique. *IET Image Processing*, 12 (6), 2018.

[17] O. Eisaku, H. Hiroshi, and A. Lim. Barcode readers using the camera device in mobile phones. *CW 2004*, pages 260–265, 2004.

[18] S. Esedoglu. Blind Deconvolution of Bar Code Signals. *Inverse Problems*, 20: 121–135, 2004.

[19] R. Fergus, B. Singh, A. Hertzmann, S. Roweis, and W. Freeman. Removing Camera Shake from a Single Photograph. *SIGGRAPH 2006*, pages 787–794, 2006.

[20] P. Hansen, J. Nagy, and D. O'Leary. *Deblurring Images.* Society for Industrial and Applied Mathematics, 2006.

[21] M. Iwen, F. Santosa, and R. Ward. A Symbol-Based Algorithm for Decoding Bar Codes. *SIIMS*, 6(1):56–77, 2013.

[22] E. T. Jaynes. Information Theory and Statistical Mechanics. *Phys. Rev.*, 106: 620–630, May 1957.

[23] E. T. Jaynes. Information Theory and Statistical Mechanics. II. *Phys. Rev.*, 108:171–190, Oct 1957.

[24] E. Joseph and T. Pavlidis. Bar Code Waveform Recognition Using Peak Locations. *IEEE Trans. Pattern Anal. Machine Intell.*, 16(6):630–640, 1994.

[25] E. Joseph and T. Pavlidis. Deblurring of Bilevel Waveforms. *IEEE Trans. Image Process.*, 16(6):630–640, 1994.

[26] J. Kim and H. Lee. Pervasive 2D Barcodes for Camera Phone Applications. *Opt. Express*, 15(22):14817–14837, 2007.

[27] J. Kim and H. Lee. Joint Nonuniform Illumination Estimation and Deblurring for Bar Code Signals. *Opt. Express*, 15(22):14817–14837, 2007.

[28] J. Kotera, F. ˇ Sroubek, and P. Milanfar. Blind Deconvolution Using Alternating Maximum a Posteriori Estimation with Heavy-Tailed Priors. *CAIP 2013*, 2013.

[29] S. Kresic-Juric. Edge Detection in Bar Code Signals Corrupted by Integrated Time-Varying Speckle. *Pattern Recognition Letters*, 38(12):2483–2493, 2005.

[30] S. Kresic-Juric, D. Madej, and F. Santosa. Applications of Hidden Markov Models in Bar Code Decoding. *Pattern Recognition Letters,*, 27(14):1665–1672, 2006.

[31] S. Kullback and R. Leibler. On Information and Sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951.

[32] A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding and Evaluating Deconvolution Algorithms. *CVPR 2009*, pages 1964–1971, 2009.

[33] S. Li, J. Shang, Z. Duan, and J. Huang. Fast Detection Method of Quick Response Code Based on Run-Length Coding. *IET Image Processing*, 12(4): 546–551, 2018.

[34] N. Liu, X. Zheng, H. Sun, and X. Tan. Two-Dimensional Bar Code Out-of-Focus Deblurring via the Increment Constrained Least Squares filter. *Pattern Recognition Letters*, 34(124–130), 2013.

[35] N. Liu, Y. Du, and Y. Xu. QR Codes Blind Deconvolution Algorithm Based on Binary Characteristic and L0 Norm Minimization. *Pattern Recognition Letters*, 111(117–123), 2018.

[36] P. Maréchal. A Note on Entropy Optimization. In M. Lassonde, editor, *Approximation, Optimization and Mathematical Economics*, pages 205–211. Physica-Verlag HD, 2001.

[37] P. Maréchal and A. Lannes. Unification of some Deterministic and Probabilistic Methods for the Solution of Linear Inverse Problems via the Principle of

Maximum Entropy on the Mean. *Inverse Problems*, 78(1), 1997.

[38] R. Palmer. *The Bar Code Book: A Comprehensive Guide to Reading, Printing, Specifying, Evaluating, and Using Bar Code and Other Machine-Readable Symbols*. Trafford Publishing, 2007.

[39] J. Pan, Z. Hu, Z. Su, and M. Yang. Deblurring Text Images via L0-Regularized Intensity and Gradient Prior. *CVPR 2014*, pages 2901–2908, 2014.

[40] D. Parikh and G. Jancke. Localization and Segmentation of a 2D High Capacity Color Barcode. *WACV 2008*, pages 1–6, 2008.

[41] H. Pu, M. Fan, J. Yang, and J. Lian. Quick Response Barcode Deblurring via Doubly Convolutional Neural Network. *Multimedia Tools and Applications*, 78:897–912, 2018.

[42] A. Rajagopalan and R. Chellappa. *Motion Deblurring: Algorithms and Systems*. Cambridge University Press, 2014.

[43] R. Rockafellar. *Conjugate Duality and Optimization*. SIAM, 1974.

[44] R. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, 2009.

[45] R. T. Rockafellar. *Convex analysis*. Princeton University Press, 1970.

[46] V. Rohatgi and A. M. E. Saleh. *An Introduction to Probability and Statistics*. Wiley, 2001.

[47] G. Sörös, S. Semmler, L. Humair, and O. Hilliges. Fast Blur Removal for Wearable QR Code Scanners. *ISWC 2015*, pages 117–124, 2015.

[48] L. Sun, S. Cho, J. Wang, and J. Hays. Edge-based Blur Kernel Estimation using Patch Priors. *ICCP 2013*, pages 1–8, 2013.

[49] J. Thielemann, H. Schumann-Olsen, H. Schulerud, and T. Kirkhus. Handheld PC with Camera Used for Reading Information Dense Barcodes. *CVPR 2014*, 2014.

[50] W. Turin and R. Boie. Bar Code Recovery via the EM Algorithm. *IEEE Trans. Image Process.*, 46(2):354–363, 1998.

[51] Y. van Gennip, P. Athavale, J. Gilles, and R. Choksi. A Regularization Approach to Blind Deblurring and Denosing of QR Barcodes. *IEEE Trans. Image Processing*, 24(9):2864–2873, 2006.

[52] M. Wang, J. Liu, and E. Fang. Accelerating Stochastic Composition Optimization. *Journal of Machine Learning Research*, 18:1–23, 2017.

[53] H. Yang, C. Alex, and X. Jiang. Barization of Low-Quality Barcode Images Captured by Mobile Phones using Local Window of Adaptive Location and Size. *IEEE Trans. Image Process.*, 21(1):418–425, 2012.

[54] C. Zhu, R. H. Byrd, and J. Nocedal. L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for Large Scale Bound Constrained Optimization. *ACM Trans. Math. Softw*, 23(2):550–560, 1997.

Interlude

With this method for barcodes and, more generally, binary images in hand, it is natural to consider an extension for generic images. This transition appears rather daunting at first, as it represents a passage from a finite, albeit dimensionally immense, primal problem to an infinite-dimensional primal problem. However, at least heuristically, one assumes that every finite-dimensional formula should have an infinite-dimensional analogue and that the two should not be overly dissimilar. This is indeed the case, although the analysis is more delicate in the latter case and, as such, the extension is non-trivial. Moreover, a stability result is derived with respect to pertubations in the input (blurred) image in this setting.

The following chapter is therefore devoted to extending the previously described method. Throughout, the assumption that some known symbology has been embedded into the image is made. The author concedes that, unlike in the case of barcodes, this assumption is rather strong and limits the applicability of the method. However, the author is hopeful that different choices of prior may permit blind deblurring in the absence of symbology.

CHAPTER 3
Paper 2: The Maximum Entropy on the Mean Method for Image Deblurring

Submitted for publication in the journal Inverse Problems on May 4 2020

Authors: Rustum Choksi, Tim Hoheisel, Gabriel Rioux, and Christopher Scarvelis

Author Ordering: The authors are listed alphabetically.

Department of Mathematics and Statistics, McGill University, Montreal, QC H3G 1Y6, Canada

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.

3.1 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 [28] 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 [15, 16]. 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 [40].

We test and compare our method on a variety of examples (cf. Section 3.5). To start, 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 3–1). We then address blind deblurring with the inclusion of a known shape (analogous to a finder pattern in a QR barcode [28]). In these cases, we can, preconditioning with EPLL denoising, blindly deblur with large blurs (cf. Figures 3–2, 3–3, 3–4). 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. [24, 25] 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.

3.1.1 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 $b = c * x$, where $*$ 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

$$\inf_{\substack{x\in\mathbb{R}^{n\times m}\\ c\in\mathbb{R}^{k\times k}}} \left\{ R(x,c) + \frac{\alpha}{2}\,||c * x - b||_2^2 \right\}, \tag{3.1}$$

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 and $\alpha > 0$ is a fidelity parameter. This idea of regularization to solve ill-posed inverse problems dates back to Tikhonov [36]. 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 [25]; weighted nuclear norm regularization, which ensures that the image or gradient matrices have low rank [27], and L_0-regularization of the dark channel, which promotes sparsity of a channel consisting of local minima in the intensity channel [24]. As it pertains to machine learning methods, other approaches have been employed including modelling the optimization problem as a deep neural network [34] and estimating the ground truth image from a blurred input without estimating the kernel using convolutional neural networks [22, 23, 35] or generative adversarial networks [18, 26].

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.

3.2 Preliminaries

We begin by recalling some standard definitions and establishing notation. We refer to [39] for convex analysis in infinite dimensions and [30] for the finite-dimensional setting. We follow [33] as a standard reference for real analysis.

Letting (X, τ) 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 $f : X \to \bar{\mathbb{R}} \equiv \mathbb{R} \cup \{-\infty, +\infty\}$, an extended real-valued function on X, the (Fenchel) conjugate of f is $f^* : X^* \to \bar{\mathbb{R}}$ 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 $\operatorname{dom} f := \{x \in X | f(x) < +\infty\}$, the domain of f, and say that f is proper if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty$ for every $x \in X$. f is said to be lower semicontinuous if $f^{-1}([-\infty, \alpha])$ is τ-closed for every $\alpha \in \mathbb{R}$.

A proper function f is convex provided for every $x, y \in \operatorname{dom} f$ 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 $x, y \in \operatorname{dom} f$ and $\lambda \in (0,1)$,

$$f(\lambda x + (1-\lambda)y) + \lambda(1-\lambda)\frac{c}{2}\left|\left|x-y\right|\right|^2 \leq \lambda f(x) + (1-\lambda)f(y),$$

then f is called c-strongly convex.

For any set $A \subseteq X$, the indicator function of A is given by

$$\delta_A : X \to \mathbb{R} \cup \{+\infty\}, \qquad x \mapsto \begin{cases} 0, & x \in A, \\ +\infty, & \text{otherwise.} \end{cases}$$

For any $\Omega \subseteq \mathbb{R}^d$, we denote by $\mathcal{P}(\Omega)$ the set of probability measures on Ω. Let η be a signed Borel measure on Ω, we define the total variation of η by

$$|\eta|(\Omega) = \sup\left\{\sum_{i\in\mathbb{N}} |\eta(\Omega_i)| \,|\, \cup_{i\in\mathbb{N}}\Omega_i = \Omega, \Omega_i \cap \Omega_j = \emptyset \; (i \neq j)\right\}.$$

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

Let μ be a positive σ-finite Borel measure on Ω and ρ be an arbitrary Borel measure on Ω, we write $\rho \ll \mu$ to signify that ρ is absolutely continuous with respect to μ, 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 $\frac{\mathrm{d}\rho}{\mathrm{d}\mu} \in L^1(\mu)$ for which

$$\rho(A) = \int_A \frac{\mathrm{d}\rho}{\mathrm{d}\mu}\,\mathrm{d}\mu, \qquad \forall\, A \subseteq \Omega \text{ measurable.}$$

The function $\frac{\mathrm{d}\rho}{\mathrm{d}\mu}$ is known as the Radon-Nikodym derivative (cf. [33, Thm. 6.10]).

The Kullback-Leibler divergence between $\rho, \mu \in \mathcal{M}(\Omega)$ is the functional

$$\mathcal{K}(\rho, \mu) = \begin{cases} \int_\Omega \log\left(\frac{\mathrm{d}\rho}{\mathrm{d}\mu}\right)\mathrm{d}\rho, & \rho, \mu \in \mathcal{P}(\Omega), \rho \ll \mu, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.2}$$

For $\Omega \subseteq \mathbb{R}^d$, $\eta \in \mathcal{M}(\Omega)$ we denote, by a slight abuse of notation, $\mathbb{E}_\eta[\boldsymbol{X}]$ to be a vector whose k^{th} component is $(\mathbb{E}_\eta[\boldsymbol{X}])_k = \int_\Omega x_k \mathrm{d}\eta(x)$. Thus, $\mathbb{E}_{(\cdot)}[\boldsymbol{X}]$ 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 $\boldsymbol{X} = [X_1, \cdots, X_d]$ 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 $C^T C$.

3.3 The MEM Method

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

Notation: We first establish some notation pertaining to deconvolution. The convolution operator $c*$ will be denoted by the matrix $C : \mathbb{R}^d \to \mathbb{R}^d$ 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 (3.1) 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 [15, 16]. 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

$$\inf_{\rho \in \mathcal{P}(\Omega)} \left\{ \mathcal{K}(\rho, \mu) + \frac{\alpha}{2} \left|\left| b - C\mathbb{E}_\rho[\boldsymbol{X}] \right|\right|_2^2 \right\} = \inf_{\mathcal{P}(\Omega)} \left\{ f + g \circ A \right\}, \tag{3.3}$$

for

$$f = \mathcal{K}(\cdot, \mu), \quad g = \frac{\alpha}{2} \left|\left| b + (\cdot) \right|\right|_2^2, \quad A = -C\mathbb{E}_{(\cdot)}[\boldsymbol{X}]. \tag{3.4}$$

The following lemma establishes some basic properties of f.

Lemma 1. *The functional $f : \mathcal{M}(\Omega) \to \bar{\mathbb{R}}$ is proper, weak* 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

$$\begin{aligned}\log\left(\frac{\frac{\mathrm{d}\rho_t}{\mathrm{d}\mu}(x)}{t + (1-t)}\right)\frac{\mathrm{d}\rho_t}{\mathrm{d}\mu}(x) &= \log\left(\frac{t\frac{\mathrm{d}\rho_1}{\mathrm{d}\mu}(x) + (1-t)\frac{\mathrm{d}\rho_2}{\mathrm{d}\mu}(x)}{t + (1-t)}\right)\left(t\frac{\mathrm{d}\rho_1}{\mathrm{d}\mu}(x) + (1-t)\frac{\mathrm{d}\rho_2}{\mathrm{d}\mu}(x)\right) \\ &\leq t\log\left(\frac{\mathrm{d}\rho_1}{\mathrm{d}\mu}(x)\right)\frac{\mathrm{d}\rho_1}{\mathrm{d}\mu}(x) + (1-t)\log\left(\frac{\mathrm{d}\rho_2}{\mathrm{d}\mu}(x)\right)\frac{\mathrm{d}\rho_2}{\mathrm{d}\mu}(x).\end{aligned}$$

The inequality is due to the log-sum inequality [9, Thm. 2.7.1], and since $\rho_1 \neq \rho_2$, $\frac{\mathrm{d}\rho_1}{\mathrm{d}\mu}$ and $\frac{\mathrm{d}\rho_2}{\mathrm{d}\mu}$ 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,

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

so f is, indeed, strictly convex.

It is well known that the restriction of f to $\mathcal{P}(\Omega)$ is weak* lower semicontinuous and proper (cf. [10, Thm. 3.2.17]). Since $f \equiv +\infty$ on $\mathcal{M}(\Omega)\backslash\mathcal{P}(\Omega)$, f preserves these properties. $\square$

Problem (3.3) 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* (3.3) *is proper, strictly convex and weak* lower semicontinuous since f is proper, strictly convex and weak* lower semicontinuous whereas $g \circ A$ is proper, weak* continuous and convex.*

Now, recall that the Riesz representation theorem [12, Cor. 7.18] identifies $\mathcal{M}(\Omega)$ as being isomorphic to the dual space of $(C(\Omega), ||\cdot||_\infty)$. Hence, by the Banach-Alaoglu theorem, [12, Thm. 5.18] the unit ball of $\mathcal{M}(\Omega)$ in the norm-induced topology[1] *($\mathbb{B}^*$) is weak*-compact.*

[1] The norm here is given by the total variation, we make precise that the weak* topology will be the only topology used in the sequel.

Since $\operatorname{dom} f \subseteq \mathcal{P}(\Omega) \subseteq \mathbb{B}^*$*, standard theory for the existence of minimizers of* τ*-lower semicontinuous functionals on* τ*-compact sets [1, Cor. 3.2.3] imply that* (3.3) *has a solution and strict convexity of* f *guarantees that it is unique.*

Even with this theoretical guarantee, direct computation of solutions to (3.3) 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 (3.3) from solutions of this dual problem, permit an efficient and accurate estimation of the original image.

3.3.2 Dual Problem

In order to derive the (Fenchel-Rockafellar) dual problem to (3.3) we provide the reader with the Fenchel-Rockafellar duality theorem in a form expedient for our study, cf. e.g. [39, Cor. 2.8.5].

Theorem 1 (Fenchel-Rockafellar Duality Theorem)**.** *Let* (X, τ) *and* (Y, τ') *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 (in their respective topologies) and proper functions and let* A *be a continuous 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*

$$\inf_{x \in X} \{f(x) + g(-Ax)\} = \max_{y^* \in Y^*} \{-f^*(A^* y^*) - g^*(y^*)\} \tag{3.5}$$

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 (3.5), 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 $(g^*)^* = g$ [39, Thm. 2.3.3]. Suppose additionally that $0 \in \operatorname{int}(A^* \operatorname{dom} g^* - \operatorname{dom} f^*)$.*

Let $\bar{y}^* \in \operatorname{argmax}_{Y^*} \{-f^* \circ A^* - g^*\}$. By the first order optimality conditions, [39, Thm. 2.5.7]

$$0 \in \partial\left(f^* \circ A^* + g\right)(\bar{y}^*) = A\partial f^*(A^*\bar{y}^*) + \partial g(\bar{y}^*),$$

the second expression is due to [4, Thm. 2.168] (the conditions to apply this theorem are satisfied by assumption). Consequently, there exists $\bar{z} \in \partial g^*(\bar{y}^*)$ and $\bar{x} \in \partial f^*(A^*\bar{y}^*)$ for which $\bar{z} = -A\bar{x}$. Since f and g are proper, lower semicontinuous and convex we have [39, Thm. 2.4.2 (iii)]:

$$A^*\bar{y}^* \in \partial f(\bar{x}), \qquad \bar{y}^* \in \partial g(\bar{z}) = \partial g(-A\bar{x}).$$

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, $f^*(A^*\bar{y}^*) < +\infty$ [4, Prop. 2.118] implies that,

$$\bar{x} \in \partial f^*(A^*\bar{y}^*) = \operatorname{argmax}_{x \in X} \left\{(x, A^*\bar{y}^*) - f(x)\right\}. \tag{3.6}$$

We refer to (3.6) *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 [39, Cor. 2.4.10], so (3.6) 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 $\mathcal{P}(\Omega)$ endowed with any topology is not a locally convex space, however it is a subset of $\mathcal{M}(\Omega)$. Previously, $\mathcal{M}(\Omega)$ was identified with the dual of $(C(\Omega), ||\cdot||_\infty)$, thus the dual of $\mathcal{M}(\Omega)$ endowed with its weak* topology $(\mathcal{M}(\Omega), w^*)^*$ can be identified with $C(\Omega)$ [8, Thm. 1.3] with duality pairing $(\phi, \rho) \in C(\Omega) \times \mathcal{M}(\Omega) \mapsto \int_\Omega \phi d\rho$. Since* $\operatorname{dom} f \subseteq \mathcal{P}(\Omega)$, *so the* inf *in* (3.2) *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 $A : \mathcal{M}(\Omega) \to \mathbb{R}^d$ in* (3.4) *is linear and weak* continuous. Moreover, its adjoint is the mapping $z \in \mathbb{R}^d \mapsto \left\langle C^T z, \cdot \right\rangle \in C(\Omega)$.*

Proof. We begin by demonstrating weak* continuity of $\mathbb{E}_{(\cdot)}[\boldsymbol{X}] : \mathcal{M}(\Omega) \to \mathbb{R}^d$. Letting $\pi_i : \mathbb{R}^d \to \mathbb{R}$ denote the projection of a vector onto its $i-$th coodinate, we have

$$\mathbb{E}_\rho[\boldsymbol{X}] = ((\pi_1, \rho), \ldots, (\pi_n, \rho)) \tag{3.7}$$

Thus, A is the composition of a weak* continuous operator from $\mathcal{M}(\Omega)$ to $\mathbb{R}^d$ and a continuous operator from $\mathbb{R}^d$ to $\mathbb{R}^d$ and hence is weak* continuous.

Eq. (3.7) equally establishes linearity of A, since the duality pairing is a bilinear form.

The adjoint can be determined by noting that

$$\langle \mathbb{E}_\rho[\boldsymbol{X}], z \rangle = \sum_{i=1}^{d} \int_\Omega x_i \mathrm{d}\rho(x) z_i = \int_\Omega \sum_{i=1}^{d} x_i z_i \mathrm{d}\rho(x) = (\langle z, \cdot \rangle, \rho),$$

so,

$$\langle C\mathbb{E}_\rho[\boldsymbol{X}], z \rangle = \left\langle \mathbb{E}_\rho[\boldsymbol{X}], C^T z \right\rangle = (\left\langle C^T z, \cdot \right\rangle, \rho),$$

yielding $A^*(z) = \left\langle C^T z, \cdot \right\rangle$. $\square$

We now compute the conjugates of f and g, respectively and provide an explicit form for the dual problem of (3.3).

Lemma 3. *The conjugate of f in (3.4) is $f^* : \phi \in C(\Omega) \mapsto \log\left(\int_\Omega \exp(\phi)\, d\mu\right)$. In particular, f^* is finite-valued. Moreover, for any $\phi \in C(\Omega)$,*

$$\operatorname{argmax}_{\mathcal{P}(\Omega)} \{(\phi, \cdot) - \mathcal{K}(\cdot, \mu)\} = \{\bar{\rho}_\phi\},$$

the unique probability measure on Ω for which

$$\frac{d\bar{\rho}_\phi}{d\mu} = \frac{\exp \phi}{\int_\Omega \exp \phi \, d\mu}. \tag{3.8}$$

Proof. We proceed by direct computation:

$$\begin{aligned} f^*(\phi) &= \sup_{\rho \in \mathcal{M}(\Omega)} \{(\phi, \rho) - \mathcal{K}(\rho, \mu)\} \\ &= \sup_{\rho \in \mathcal{P}(\Omega)} \{(\phi, \rho) - \mathcal{K}(\rho, \mu)\} \\ &= \sup_{\rho \in \mathcal{P}(\Omega)} \left\{ \int_\Omega \log \left(\frac{\exp \phi}{\frac{\mathrm{d}\rho}{\mathrm{d}\mu}} \right) \mathrm{d}\rho \right\}, \end{aligned}$$

where we have used the fact that $\operatorname{dom} f \subseteq \mathcal{P}(\Omega)$ as noted in Remark 3. Note that $\exp \phi \in C(\Omega) \subseteq L^1(\rho)$ and since $t \mapsto \log t$ is concave, Jensen's inequality [33, Thm. 3.3] yields

$$f^*(\phi) \leq \sup_{\rho \in \mathcal{P}(\Omega)} \left\{ \log \left(\int_\Omega \frac{\exp \phi}{\frac{\mathrm{d}\rho}{\mathrm{d}\mu}} \mathrm{d}\rho \right) \right\} = \log \left(\int_\Omega \exp \phi \, \mathrm{d}\mu \right) \tag{3.9}$$

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

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

one has that

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

so $\bar{\rho}_\phi \in \operatorname{argmax}_{\mathcal{P}(\Omega)} \{(\phi, \cdot) - \mathcal{K}(\cdot, \mu)\}$ as $\bar{\rho}_\phi$ saturates the upper bound for $f^*(\phi)$ established in (3.9), thus $f^*(\phi) = \log \left(\int_\Omega \exp \phi \, \mathrm{d}\mu \right)$. Moreover $\bar{\rho}_\phi$ is the unique maximizer since the objective is strictly concave.

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

$$\exp(\phi(x)) \leq \exp\left(\max_{\Omega} \phi\right), \qquad (x \in \Omega).$$

Thus,

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

since $C(\Omega) = C_b(\Omega)$ by compactness of Ω. Since ϕ is arbitrary, dom $f^* = C(\Omega)$ and, coupled with the fact that f^* is proper [39, Thm. 2.3.3], we obtain that f^* is finite-valued. □

Lemma 4. *The conjugate of g from* (3.4) *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 [29, Ex. 11.11] and some standard properties of conjugacy [29, Eqn. 11(3)]. □

Combining these results we obtain the main duality theorem.

Theorem 2. *The (Fenchel-Rockafellar) dual of* (3.3) *is given by*

$$\max_{\lambda \in \mathbb{R}^d} \left\{ \langle b, \lambda \rangle - \frac{1}{2\alpha} ||\lambda||_2^2 - \log\left(\int_{\Omega} \exp\left\langle C^T\lambda, x\right\rangle d\mu(x)\right)\right\}. \tag{3.10}$$

Given a maximizer $\bar{\lambda}$ *of* (3.10) *one can recover a minimizer of* (3.3) *via*

$$d\bar{\rho} = \frac{\exp\left\langle C^T\bar{\lambda}, \cdot\right\rangle}{\int_{\Omega} \exp\left\langle C^T\bar{\lambda}, \cdot\right\rangle d\mu} d\mu. \tag{3.11}$$

Proof. The dual problem can be obtained by applying the Fenchel-Rockafellar duality theorem (Theorem 1), with f and g defined in (3.4), to the primal problem

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

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 $\mathbf{0} \subseteq \operatorname{dom} g^* = \mathbb{R}^d$ and $A^*\mathbf{0} = 0 \in C(\Omega)$, so

$$A^*(\operatorname{dom} g^*) - \operatorname{dom} f^* \supseteq -\operatorname{dom} f^* = \{\phi | -\phi \in \operatorname{dom} f^*\} = C(\Omega),$$

since $\operatorname{dom} f^* = C(\Omega)$ by Lemma 3. Thus $0 \in \operatorname{int}(A^* \operatorname{dom} g^* - \operatorname{dom} f^*) = C(\Omega)$, and Remark 2 is applicable. The primal-dual recovery formula (3.6) is given explicit form by Lemma 3 by evaluating $\mathrm{d}\bar{\rho}_{\langle C^T\bar{\lambda}, \cdot \rangle}$.

□

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 α in (3.3). Notably in [5, Cor. 4.9] the problem

$$\inf_{\rho \in \mathcal{P}(\Omega) \cap \operatorname{dom} \mathcal{K}(\cdot, \mu)} \mathcal{K}(\rho, \mu) \quad \text{s.t.} \|C\mathbb{E}_\rho[\boldsymbol{X}] - b\|_2^2 \leq \frac{1}{2\alpha} \tag{3.12}$$

is paired in duality with (3.10). Thus the choice of α is directly related to the fidelity of $C\mathbb{E}_\rho[\boldsymbol{X}]$ to the blurred image. The following section is devoted to the choice of a prior and describing a method to directly compute $\mathbb{E}_{\bar{\rho}}[\boldsymbol{X}]$ from a solution of (3.10).

3.3.3 Probabilistic Interpretation of Dual Problem

If no information is known about the original image, the prior μ 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 $\boldsymbol{X} = [X_1, X_2, \dots]$ 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 ℓ, one can enforce this constraint by taking the random variable X_k to be distributed uniformly on $[\ell - \epsilon, \ell + \epsilon]$. Constructing μ in this fashion guarantees that its support (and hence Ω) is compact.

To deal with the integrals in (3.10) and (3.11) it is convenient to note that (cf. [31, Sec. 4.4])

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

the moment-generating function of $\boldsymbol{X}$ evaluated at $C^T\lambda$. Since the X_i are independently distributed, $\mathbb{M}_{\boldsymbol{X}}[t] = \Pi_{i=1}^d \mathbb{M}_{X_i}[t]$ [31, Sec. 4.4], and since the X_i are uniformly distributed on $[u_i, v_i]$ one has

$$\mathbb{M}_{\boldsymbol{X}}[t] = \prod_{i=1}^d \frac{e^{t_i v_i} - e^{t_i u_i}}{t_i(v_i - u_i)},$$

and therefore the dual problem (3.10) with this choice of prior can be written as

$$\max_{\lambda \in \mathbb{R}^d} \left\{ \langle b, \lambda \rangle - \frac{1}{2\alpha} ||\lambda||_2^2 - \sum_{i=1}^{d} \log \left(\frac{e^{C_i^T \lambda v_i} - e^{C_i^T \lambda u_i}}{C_i^T \lambda (v_i - u_i)} \right) \right\}, \tag{3.13}$$

with C_i^T denoting the transpose of the i-th column of C. A solution of (3.13) can be determined using a number of standard numerical solvers. We opted for the implementation [6] of the L-BFGS algorithm due to its speed and efficiency.

Since only the expectation of the optimal probability measure for (3.3) is of interest, we compute the i^{th} component of the expectation $(\mathbb{E}_{\bar{\rho}}[\boldsymbol{X}])_i$ of the optimizer provided by the primal-dual recovery formula (3.11) via

$$\frac{\int_\Omega x_i e^{\langle C^T \bar{\lambda}, x \rangle} \mathrm{d}\mu}{\int_\Omega e^{\langle C^T \bar{\lambda}, x \rangle} \mathrm{d}\mu} = \partial_{t_i} \log \left(\int_\Omega e^{\langle t, x \rangle} \, \mathrm{d}\mu \right) \Big|_{t = C^T \bar{\lambda}}.$$

Using the independence assumption on the prior, we obtain

$$\mathbb{E}_{\bar{\rho}}[\boldsymbol{X}] = \nabla_t \sum_{i=1}^{d} \log \left(\mathbb{M}_{X_i}[t] \right) \Big|_{t = C^T \bar{\lambda}}$$

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

$$(\mathbb{E}_{\bar{\rho}}[\boldsymbol{X}])_i = \frac{v_i e^{C_i^T \bar{\lambda} v_i} - u_i e^{C_i^T \bar{\lambda} u_i}}{e^{C_i^T \bar{\lambda} v_i} - e^{C_i^T \bar{\lambda} u_i}} - \frac{1}{C_i^T \bar{\lambda}}. \tag{3.14}$$

With (3.13) and (3.14) in hand, our entropic method for deconvolution can be implemented.

3.3.4 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 (3.3) 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

$$\inf_{\eta \in \mathcal{P}(\Omega)} \left\{ \frac{\gamma}{2} \left|\left| \mathbb{E}_\eta[\boldsymbol{X}] * \tilde{x} - \tilde{b} \right|\right|_2^2 + \mathcal{K}(\eta, \nu) \right\}. \tag{3.15}$$

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 (3.3), the expectation of the optimizer of (3.15) 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 (3.15) can be determined by passing to the dual problem (which is of the same form as (3.13)), solving the dual problem numerically and using the primal-dual recovery formula (3.14). A summary of the blind deblurring algorithm is compiled in Algorithm 2. 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 2 Entropic Blind Deblurring

Require: Blurred image b, symbology $\tilde{x}$, prior μ, kernel width k, fidelity parameters γ, α;
Ensure: Deblurred image $\bar{x}$
$\nu \leftarrow$ density of k^2 uniformly distributed independent random variables
$\lambda_{\bar{c}} \leftarrow$ argmax of analog of (3.13) for kernel estimate.
$\bar{c} \leftarrow$ expectation of argmin of (3.15) computed via analog of (3.14) for kernel estimate evaluated at $\lambda_{\bar{c}}$
$\lambda_{\bar{x}} \leftarrow$ argmax of (3.13)
$\bar{x} \leftarrow$ expectation of argmin of (3.3) with kernel $\bar{c}$ computed via (3.14) evaluated at $\lambda_{\bar{x}}$
return $\bar{x}$

3.4 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*

$$\rho_i = \operatorname{argmin}_{\rho \in \mathcal{P}(\Omega)} \left\{ \mathcal{K}(\rho, \mu) + \frac{\alpha}{2} \, ||C\mathbb{E}_\rho[\boldsymbol{X}] - b_i||_2^2 \right\} \quad (i = 1, 2),$$

then

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

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 (3.3), which we denote $v : \mathbb{R}^d \to \mathbb{R}$, as

$$v(b) := \inf_{\rho \in \mathcal{P}(\Omega)} \left\{ \mathcal{K}(\rho, \mu) + \frac{\alpha}{2} ||C\mathbb{E}_\rho[\boldsymbol{X}] - b||_2^2 \right\} = \inf_{\rho \in \mathcal{P}(\Omega)} \{k(\rho, b) + h \circ L(\rho, b)\} , \tag{3.16}$$

where

$$k : (\rho, b) \in \mathcal{M}(\Omega) \times \mathbb{R}^d \mapsto \mathcal{K}(\rho, \mu), \quad h = \frac{\alpha}{2} ||\cdot||_2^2 , \quad L(\rho, b) = C\mathbb{E}_\rho[\boldsymbol{X}] - b. \tag{3.17}$$

The following results will allow us to conclude that ∇v is (globally) α-Lipschitz.

Lemma 5. *The operator* L *in* (3.17) *is linear and continuous in the product topology, its adjoint is the map* $z \mapsto (\left\langle C^T z, \cdot \right\rangle, -z) \in C(\Omega) \times \mathbb{R}^d$*.*

Proof. Linearity and continuity of this operator follow from the linearity and weak* continuity of the expectation operator (cf. Lemma 2). The adjoint is obtained as in Lemma 2,

$$\langle C\mathbb{E}_\rho[\boldsymbol{X}] - b, z \rangle = (\left\langle C^T z, \cdot \right\rangle, \rho) + \langle b, -z \rangle .$$

$\square$

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

Lemma 6. *The conjugate of $k + h \circ L$ defined in* (3.17) *is the function*

$$(\phi, y) \in C(\Omega) \times \mathbb{R}^d \mapsto (\mathcal{K}(\cdot, \mu))^*(\phi + \left\langle C^T y, \cdot \right\rangle) + \frac{1}{2\alpha} \left|\left|y\right|\right|_2^2, \tag{3.18}$$

where $(\mathcal{K}(\cdot, \mu))^*$ is the conjugate computed in Lemma 3.

Proof. Since $\operatorname{dom} h = \mathbb{R}^d$, h is continuous and k is proper, there exists $x \in L \operatorname{dom} k \cap \operatorname{dom} h$ such that h is continuous at x. The previous condition guarantees that, [39, Thm. 2.8.3]

$$(k + h \circ L)^*(\phi, y) = \min_{z \in \mathbb{R}^d} \left\{ k^*((\phi, y) - L^*(z)) + h^*(z) \right\}. \tag{3.19}$$

The conjugate of k is given by

$$k^*(\phi, y) = \sup_{\substack{\rho \in \mathcal{M}(\Omega) \\ b \in \mathbb{R}^d}} \left\{ (\phi, \rho) + \langle y, b \rangle - \mathcal{K}(\rho, \mu) \right\}.$$

For $y \neq 0$, $\sup_{\mathbb{R}^d} \langle y, \cdot \rangle = +\infty$. Thus,

$$k^*(\phi, y) = \sup_{\rho \in \mathcal{M}(\Omega)} \left\{ (\phi, \rho) - \mathcal{K}(\rho, \mu) \right\} + \delta_{\{0\}}(y) = (\mathcal{K}(\cdot, \rho))^*(\phi) + \delta_{\{0\}}(y).$$

The conjugate of h was established in Lemma 4 and the adjoint of L is given in Lemma 5. Substituting these expressions into (3.19) yields,

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

□

The conjugate computed in the previous lemma can be used to establish that of the optimal value function.

Lemma 7. *The conjugate of v in* (3.16) *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,

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

In light of (3.18), $v^*(y) = (\mathcal{K}(\cdot,\mu))^*(\langle C^T y,\cdot\rangle) + \frac{1}{2\alpha}\,||y||_2^2$ which is the sum of a convex function and a $\frac{1}{\alpha}$-strongly convex function and is thus $\frac{1}{\alpha}$-strongly convex. □

Remark 4. *Theorem 2 establishes attainment for the problem defining v in* (3.16), *so* $\operatorname{dom} v = \mathbb{R}^d$ *and v is proper. Moreover, [4, Prop. 2.152] and [4, Prop. 2.143] establish, respectively, continuity and convexity of v. Consequently,* $(v^*)^* = v$ *[39, Thm. 2.3.3] and since v^* is $\frac{1}{\alpha}$-strongly convex, v is Gâteaux differentiable with globally α-Lipschitz derivative [39, Rmk. 3.5.3].*

We now compute the derivative of v.

Lemma 8. *The derivative of v is the map* $b \mapsto \alpha\,(b - C\mathbb{E}_{\bar\rho}[\boldsymbol{X}])$, *where $\bar\rho$ is the solution of the primal problem* (3.3), *which is given in* (3.11).

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 [39, 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

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

and

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

Thus

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

The first order optimality conditions for (3.10) imply that $\bar{\lambda} = \alpha(b - C\mathbb{E}_{\bar{\rho}}[\boldsymbol{X}])$, so

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

Thus, $\bar{\lambda}$ is the unique vector satisfying $v(b) + v^*(\cdot) = \langle b, \cdot\rangle$ and thus $\nabla v(b) = \bar{\lambda} = \alpha(b - C\mathbb{E}_{\bar{\rho}}[\boldsymbol{X}])$. $\square$

We now prove Theorem 3.

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

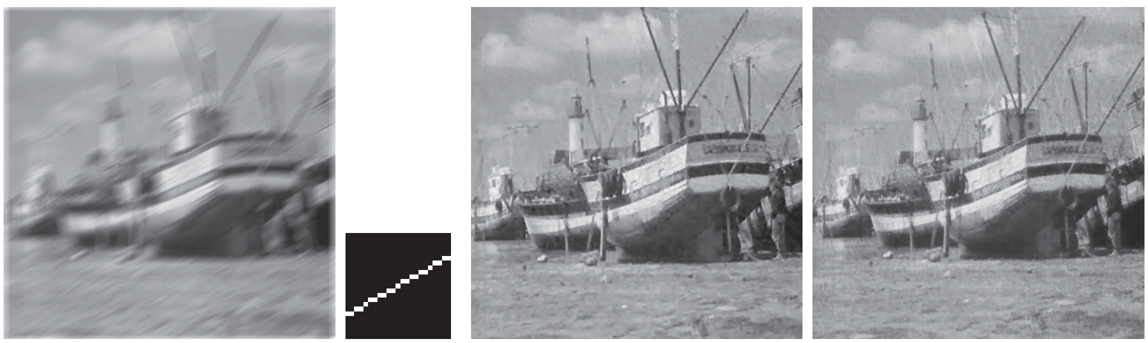
$$\left|\left|\nabla v(b_1) - \nabla v(b_2)\right|\right|_2 \le \alpha \left|\left|b_1 - b_2\right|\right|_2$$

and

$$
\begin{aligned}
\left|\left|\nabla v(b_1) - \nabla v(b_2)\right|\right|_2 &= \alpha \left|\left|b_1 - b_2 + C(\mathbb{E}_{\rho_2}[\boldsymbol{X}] - \mathbb{E}_{\rho_1}[\boldsymbol{X}])\right|\right|_2 \\
&\geq \alpha \left|\left|C\left(\mathbb{E}_{\rho_2}[\boldsymbol{X}] - \mathbb{E}_{\rho_1}[\boldsymbol{X}]\right)\right|\right|_2 - \alpha \left|\left|b_2 - b_1\right|\right|_2 \\
&\geq \alpha \sigma_{\min}(C) \left|\left|\mathbb{E}_{\rho_1}[\boldsymbol{X}] - \mathbb{E}_{\rho_2}[\boldsymbol{X}]\right|\right|_2 - \alpha \left|\left|b_1 - b_2\right|\right|_2 .
\end{aligned}
$$

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

3.5 Numerical Results



(a) PSNR: 20.83 dB

(b) Cho et al[7], PSNR: 27.50 dB

(c) Ours, PSNR: 26.68 dB

Figure 3–1: **Deconvolution with noise:** Original image is 512×512 pixels. (a) is the blurred image which is further degraded with 1% Gaussian noise along with the 23 pixel wide convolution kernel. (b) is the result obtained using Cho *et al*'s deconvolution method [7]. (c) is the result obtained from the blurred image via our non-blind deblurring method.

We present results obtained using our method on certain simulated images. We begin with deconvolution, i.e. when the blurring kernel c is known. Figure 3–1 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 when a uniform prior is used, so a preprocessing step consisting of expected patch log likelihood (EPLL) denoising [40] is first performed. For the sake of consistency, the same preprocessing step is applied prior to using Cho *et al*'s deconvolution method [7] (this step also improves the quality of the restoration for this method). The resulting image is subsequently deblurred and finally TV denoising [32] is used to smooth the image in our case (this step is unnecessary for the other method as it already results in a smooth restoration). Note that for binary images such as text, TV denoising can be replaced by a thresholding step (see figure 3–3).

3.5.1 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 $b = c * x$ is replaced by $b = c * x + p$ 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 are incompatible.

However, Figures 3–1 and 3–2 show that our method yields good results in both deconvolution and blind deblurring when a denoising preprocessing step (the other methods are applied to the preprocessed version of the image as well for the sake of consistency) 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.

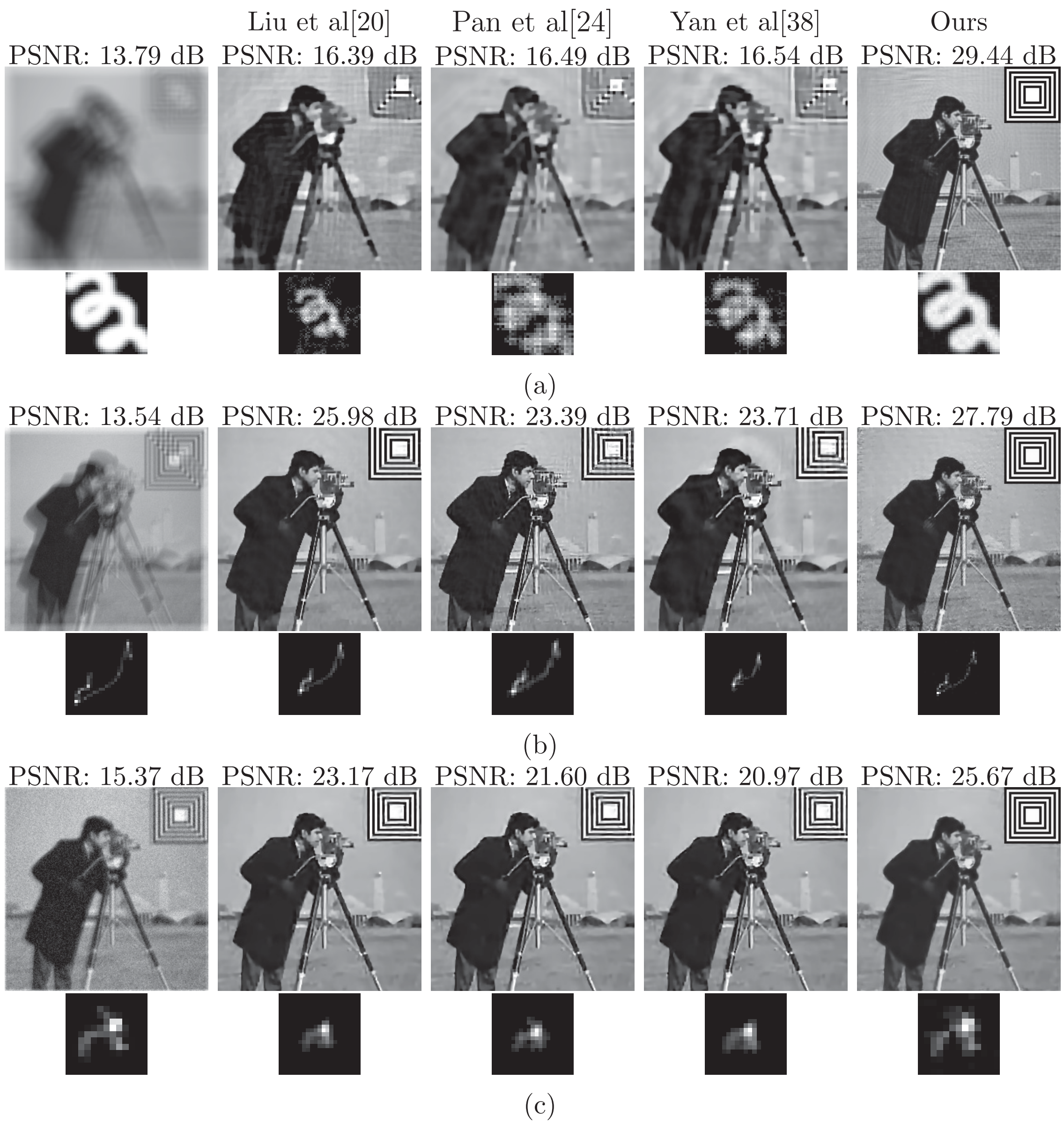


Figure 3–2: **Blind deblurring with and without noise:** Original image is 256×256 pixels. The performance, with varying amounts of noise and different blurring kernels, of our blind deblurring method with EPLL denoising preprocessing and TV denoising postprocessing to that of other contemporary methods with the EPLL denoising preprocessing step. The blurred and noisy image is on the left with the original convolution kernel below it. (a) is noiseless with a 33 pixel wide kernel. (b) has 1% Gaussian noise with a 27 pixel wide kernel. (c) has 5% Gaussian Noise with a 13 pixel wide kernel.

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.

(a) (b)

Figure 3–3: **Blind text deblurring with and without noise:** Original image is 500×155 pixels. Top: Blurred and noisy image. Middle: Original convolution kernel on the left and estimated kernel on the right. Bottom: Deblurred image obtained using our method with an EPLL denoising preprocessing step and a thresholding postprocessing step. (a) is noiseless with a 57 pixel kernel. (b) has 1% Gaussian noise with a 45 pixel wide kernel.

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

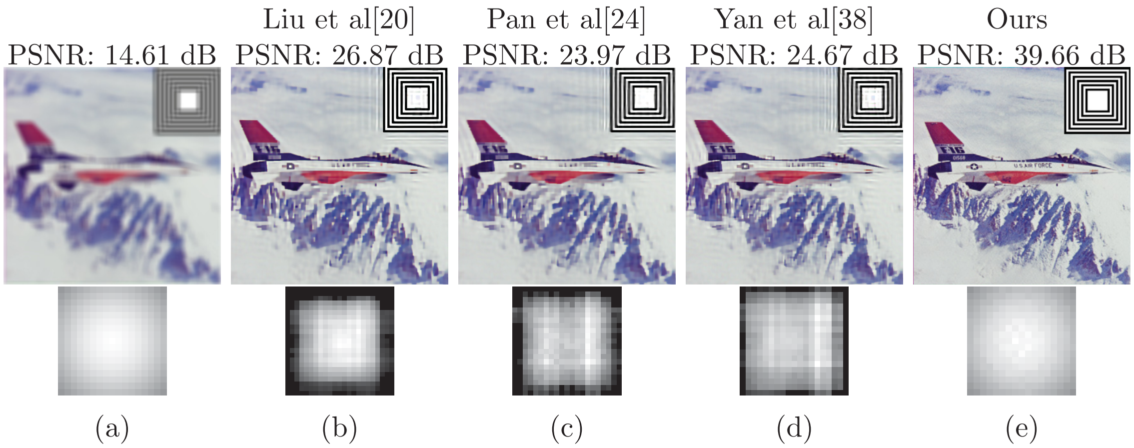


Figure 3–4: **Blind deblurring in color:** Original image is 512×512 pixels. (a) is the image which has been blurred with a 17×17 kernel. (b)-(e) are the latent image and estimated kernel obtained with different methods.

estimate for the results in Figure 3–2 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.

3.6 The Role of the Prior

Our method is based upon the premise that a priori the probability density ρ 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 μ. Let us first recall the role of the prior 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:

1. μ has a tractable moment-generating function (so that the dual problem can be solved via gradient-based methods such as L-BFGS),
2. 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. The moment-generating functions for these priors are

$$\mathbb{M}_{X_i}[t] = 1 - p_i + p_i e^{t_i}, \text{ and } \mathbb{M}_{X_i}[t] = e^{\lambda_i (e^{t_i} - 1)},$$

respectively with p_i denoting the probability that the i-th pixel is white and λ_i denoting the mean of the Poisson distribution. Thus to deblur images with these priors, one simply substitutes these expressions into the dual problem and the recovery formula, as in (3.13) and (3.14). Figure 3–5 presents a comparison of deblurring a binary text image using different priors with the same choice of $\alpha = 2 \times 10^{11}$. 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.

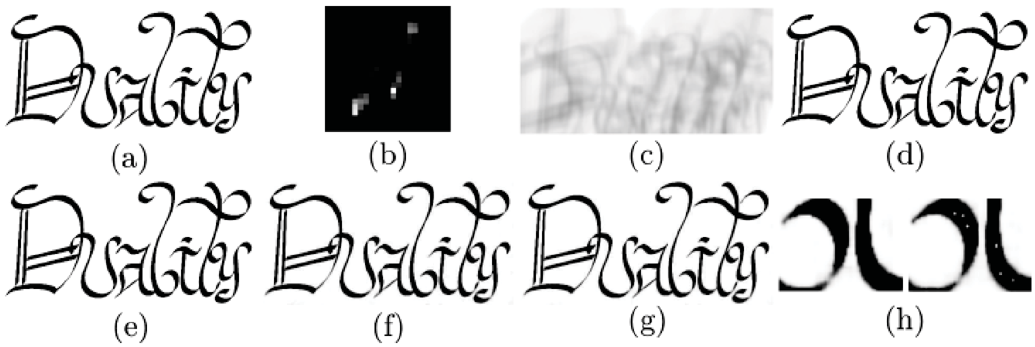


Figure 3–5: **Deconvolution with different priors:** (a) is the original 480×240 binary text image. (b) is the 207×207 pixel convolution kernel. (c) is obtained by blurring the text with the convolution kernel. (d) and (e) are the results of performing deconvolution on the previous blurred image using Bernoulli and Poisson priors respectively using $\alpha = 10^{10}$. (f) and (g) were obtained by deconvolving with $\alpha = 10^6$ with the two priors. (e) presents the fine detail of the deconvolution with $\alpha = 10^6$ 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. [14]) 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. [17]) 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 [37].

3.7 Discussion

It is surprising that inference schemes of the type considered in this paper have only seen limited popularity in image deblurring (see [2, 3, 21]). 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. [11]).

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 [13])

Finally, our method is flexible with respect to the choice of prior and as we briefly discussed in Section 3.6, 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 [19] as well as Gaussian blur kernels. 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 α) 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.

Acknowledgements

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. [28]).

REFERENCES

[1] H. Attouch, G. Buttazzo, and G. Michaille. *Variational Analysis in Sobolev and BV Spaces.* Society for Industrial and Applied Mathematics, 2014.

[2] J. F. Bercher, G. L. Besnerais, and G. Demoment. The Maximum Entropy on the Mean Method, Noise and Sensitivity. In Skilling, John and Sibisi, Sibusiso, editor, *Maximum Entropy and Bayesian Methods*, pages 223–232. Springer Netherlands, 1996.

[3] G. L. Besnerais, J. F. Bercher, and G. Demoment. A New Look at Entropy for Solving Linear Inverse Problems. *IEEE Trans. Inf. Theory*, 45(5), 1999.

[4] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, 2000.

[5] J. Borwein and A. Lewis. Partially finite convex programming, Part I: Quasi relative interiors and duality theory. *Math.l Program.*, 57:15–48, 1992.

[6] R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.*, 16:1190–1208, 1995.

[7] S. Cho, J. Wang, and S. Lee. Handling outliers in non-blind image deconvolution. In *2011 IEEE International Conference on Computer Vision*, pages 1–8, 2011.

[8] J. B. Conway. *A Course in Functional Analysis*. Springer, 2007.

[9] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2006.

[10] J.-D. Deuschel and D. W. Stroock. *Large Deviations*. Academic Press, 1989.

[11] M. Dudik, S. Philips, and R. Schapire. Maximum Entropy Density Estimation with Generalized Regularization and an Applicationto Species Distribution Modeling. *Journal of Machine Learning Research*, 8:1217–1260, 2007.

[12] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Wiley, 1999.

[13] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. Hengel, and Q. Shi. From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur. *CPVR 2016*, 12 2016.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. *Proceedings of the International Conference on Neural Information Processing Systems (NIPS)*, pages 2672–2680, 2014.

[15] E. T. Jaynes. Information Theory and Statistical Mechanics. *Phys. Rev.*, 106: 620–630, May 1957.

[16] E. T. Jaynes. Information Theory and Statistical Mechanics. II. *Phys. Rev.*, 108:171–190, Oct 1957.

[17] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4396–4405, 2019.

[18] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin., and J. Matasi. Deblur-GAN: Blind Motion Deblurring Using Conditional Adversarial Networks. *CVPR 2018*, pages 8183–8192, 2018.

[19] A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding and Evaluating Deconvolution Algorithms. *CVPR 2009*, pages 1964–1971, 2009.

[20] J. Liu, M. Yan, and T. Zeng. Surface-aware blind image deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2019.

[21] P. Maréchal and A. Lannes. Unification of some Deterministic and Probabilistic Methods for the Solution of Linear Inverse Problems via the Principle of Maximum Entropy on the Mean. *Inverse Problems*, 78(1), 1997.

[22] S. Nah, T. Kim, and K. Lee. Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring. *CVPR 2017*, pages 3883–3891, 2017.

[23] M. Noroozi, P. Chandramouli, and P. Favaro. Motion Deblurring in the Wild. *GCPR 2017*, pages 65–77, 2017.

[24] J. Pan, D. Sun, H. Pfister, and M.-H. Yang. Blind Image Deblurring Using Dark Channel Prior. *CVPR 2016*, pages 1628–1636, 2016.

[25] J. Pan, Z. Hu, Z. Su, and M.-H. Yang. L_0-Regularized Intensity and Gradient Prior for Deblurring Text Images and Beyond. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(2):342–355, 2017.

[26] S. Ramakrishnan, S. Pachori, A. Gangopadhyay, and S. Raman. Deep Generative Filter for Motion Deblurring. *ICCV 2017*, pages 2993–3000, 2017.

[27] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M.-H. Yang. Image Deblurring via Enhanced Low-Rank Prior. *IEEE Trans. Image Process.*, 25(7):3426–3437, 2016.

[28] G. Rioux, C. Scarvelis, R. Choksi, T. Hoheisel, and P. Maréchal. Blind Deblurring of Barcodes via Kullback-Leibler Divergence. *IEEE Trans. Pattern Anal. Mach. Intell., in press*, 2020. doi: `10.1109/TPAMI.2019.2927311`.

[29] R. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, 2009.

[30] R. T. Rockafellar. *Convex analysis*. Princeton University Press, 1970.

[31] V. Rohatgi and A. M. E. Saleh. *An Introduction to Probability and Statistics*. Wiley, 2001.

[32] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[33] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1987.

[34] C. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf. Learning to Deblur. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(7):1439–1451, 2015.

[35] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia. Scale-Recurrent Network for Deep Image Deblurring. *CVPR 2018*, pages 8174–8182, 2018.

[36] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Winston and Sons, 1977.

[37] M. Wang, J. Liu, and E. Fang. Accelerating Stochastic Composition Optimization. *Journal of Machine Learning Research*, 18:1–23, 2017.

[38] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao. Image deblurring via extreme channels prior. In *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 6978–6986, 2017.

[39] C. Zălinescu. *Convex Analysis in General Vector Spaces*. World Scientific, 2002.

[40] D. Zoran and Y. Weiss. From Learning Models of Natural Image Patches to Whole Image Restoration. *ICCV 2011*, pages 479–486, 2011.

CHAPTER 4
Conclusion

In summary, this manuscript has outlined a novel method for both binary and generic image deblurring. This approach is based upon the principle of maximum entropy on the mean, yielding a problem which is a priori intractable due to its immense dimensionality. The Fenchel-Rockafellar duality theorem affords a finite-dimensional dual problem (with dimensionality equal to that of the image being deblurred), along with a formula to recover solutions of the primal problem upon solving its dual analogue. The resolution of the dual problem is made tractable by noting that the convex conjugate of the Kullback-Leibler divergence evaluated at the adjoint of the expectation operator is simply the moment-generating function of the prior measure being used. Hence, for many choices of prior, the dual problem can be solved numerically using black-box optimization software (the L-BFGS algorithm was used throughout). Once this solution is acquired, a formula is provided to recover the expectation of the solution of the primal problem.

In the case of barcode deblurring, the maximum entropy on the mean deblurring method performs exceedingly well in both blind and non-blind deblurring with and without noise. This performance is aided by the embedded finder patterns, which are used to great effect in both the kernel estimation and deconvolution steps.

For general images, the deconvolution results obtained by this method are particularly striking. However, the supplementary assumptions posited to enable the kernel estimation in this context are admittedly rather restrictive. Indeed, the number of situations in which symbology is naturally embedded in an image is limited. However, it should be stressed that these results should be taken as a proof of concept for the potential of this method. In effect, additional study of the method could reveal prior measures which are better adapted to the kernel estimation step and remove the need for symbolic constraints.

To conclude, the maximum entropy on the mean method provides a novel and mathematically elegant framework for blind and non-blind image deblurring. The potential of this methodology is demonstrated through multiple examples, however additional study will be necessary in order to fully characterize its strengths and weaknesses.

REFERENCES

[1] J. M. Anderson, B. A. Mair, M. Rao, and C.-H. Wu. Weighted Least-Squares Reconstruction Methods for Positron Emission Tomography. *IEEE Trans. Med. Imaging*, 16(2): 159–165, 1997.

[2] H. H. Bauschke, D. Noll, A. Celler, and J. M. Borwein. An EM Algorithm for Dynamic SPECT. *IEEE Trans. Med. Imaging*, 18(3): 252–261, 1999.

[3] J. F. Bercher, G. Le Besnerais, and G. Demoment. The Maximum Entropy on the Mean Method, Noise and Sensitivity. In J. Skilling, and S. Sibisi, editors, *Maximum Entropy and Bayesian Methods. Fundamental Theories of Physics*, Springer, 223–232, 1996.

[4] G. Le Besnerais, J. F. Bercher, and G. Demoment. A New Look at Entropy for Solving Linear Inverse Problems. *IEEE Trans. Inf. Theory*, 45(5): 1565–1578, 1999.

[5] J. Borwein and A.S. Lewis, Partially Finite Convex Programming, part I: Quasi relative interiors and duality theory, *Mathematical Programming*, vol. 57, 15–48, 1992.

[6] J. Borwein and A.S. Lewis, Partially Finite Convex Programming, part II: Explicit lattice models, *Mathematical Programming*, vol. 57, 49–83, 1992.

[7] J. F. Cai, H. Ji, C. Liu, and Z. Shen. Blind Motion Deblurring from a Single Image Using Sparse Approximation. In *CVPR 2009*, 104–111, 2009.

[8] S. H. Chan and T. Q. Nguyen. Single Image Spatially Variant Out-of-Focus Blur Removal. In *ICIP 2011*, 677–680, 2011.

[9] T.-J. Chen, K.-S. Chuang, J.-H. Chang, Y.-H. Shiao and C.-C. Chuang. A Blurring Index for Medical Images. *J. Digit. Imaging*, 19(2): 118–125, 2005.

[10] S. Cho and S. Lee. Fast Motion Deblurring. *ACM Trans. Graph.*, 28(5): 1–8, 2009.

[11] S. Cho, Y. Matsushita, and S. Lee. Removing Non-Uniform Motion Blur from Images. In *ICCV 2007*, 1–8, 2007.

[12] F. Couzinié-Devy, J. Sun, K. Alahari, and J. Ponce. Learning to Estimate and Remove Non-uniform Image Blur. In *CVPR 2013*, 1075–1082, 2013.

[13] L. Denis, E. Thiébaut, and F. Soulez. Fast Model of Space-variant Blurring and its Application to Deconvolution in Astronomy. In *ICIP 2011*, 2817–2820, 2011.

[14] EHT Collaboration et al. First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole. *Astrophys. J.*, 870,(1), 2019.

[15] S. Esedoglu. Blind Deconvolution of Bar Code Signals. *Inverse Problems*, 20: 121–135 2003.

[16] "Fftconvolve," `https://docs.scipy.org/doc/scipy-0.16.1/reference/generated/scipy.signal.fftconvolve.htm`

[17] Y. van Gennip, P. Athavale, J. Gilles and R. Choksi. A Regularization Approach to Blind Deblurring and Denoising of Barcodes. *IEE Trans. Image Process.* 24,(9): 2864–2873 2015.

[18] S. Harmeling, M. Hirsch, S. Sra, and B. Schölkopf. Online Blind Deconvolution for Astronomical Imaging. In *ICCP 2009*, 1–7, 2009.

[19] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Schölkopf. Fast removal of non-uniform camera shake. In *ICCP 2011*, 463–470, 2011.

[20] Y. Huang, M. K. Ng, and Y.-W. Wen. A Fast Total Variation Minimization Method for Image Restoration. *Multiscale Model. Simul.* 7,(2): 774–795 2008.

[21] E. T. Jaynes. Information Theory and Statistical Mechanics. *Phys. Rev.*, 106: 620–630, May 1957.

[22] E. T. Jaynes. Information Theory and Statistical Mechanics. II. *Phys. Rev.*, 108:171–190, Oct 1957.

[23] J. Jia. Mathematical Models and Practical Solvers for Uniform Motion Deblurring. In Rajagopalan, and R. Chellappa, editors, *Motion Deblurring: Algorithms and Systems*, Cambridge University Press, 1–30, 2014.

[24] W. C. Karl. Regularization in Image Restoration and Reconstruction. In A. Bovik, editor, *Handbook of Image and Video Processing*, Academic Press, 183 – 202, 2005.

[25] D. Krishnan, T. Tay, and R. Fergus. Blind Deconvolution Using a Normalized Sparsity Measure. In *CVPR 2011*, 233–240, 2011.

[26] S. Kullback and R. Leibler. On Information and Sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951.

[27] L. D. Landau, L. E. M., L. P. Pitaevskii, J. B. Sykes, and M. J. Kearsley. *Statistical Physics Volume 5: Statistical Physics*. Pergamon, 1993.

[28] H. Liu, J. Gu, and C. Huang. Image Deblurring by Generalized Total Variation Regularization and Least Squares Fidelity. In *ICIA 2013*, 1945–1949, 2016.

[29] N. Liu, X. Zheng, H. Sun, and X. Tan, "Two-dimensional bar code out-of-focus deblurring via the Increment Constrained Least Squares filter," *Pattern Recogni. Lett.*, vol. 34, 124–130, 2013.

[30] N. Liu, Y. Du, and Y. Xu "QR Codes Blind deconvolution Algorithm Based on Binary Characteristic and L0 Norm Minimization" *Pattern Recognit. Lett.*, vol. 111, 117–123, 2018.

[31] X.-G. Lu, T.-V. Huang, Z.-B. Xu, and X.-L. Zhao. A Special Hermitian and Skew-Hermitian Splitting Method for Image Restoration. *Applied Mathematical Modeling*, 37:1069–1082, 2013.

[32] P. Maréchal, and A. Lannes. Unification of some Deterministic and Probabilistic Methods for the Solution of Linear Inverse Problems via the Principle of Maximum Entropy on the Mean. *Inverse Problems*, 13:135–151, 1997.

[33] R. Mourya, L. Denis, J.-M. Becker, and E. Thiébaut. A Blind Deblurring and Image Decomposition Approach for Astronomical Image Restoration. In *EUSIPCO 2015*, 1636–1640, 2015.

[34] J. Nagy and D. O'Leary. Restoring Images Degraded by Spatially Variant Blur. *SIAM J. Sci. Comput.*, 19,(4): 1063–1082, 1998.

[35] J. Nagy, K. Palmer, and L. Perrone. Iterative Methods for Image Deblurring: A Matlab Object-Oriented Approach. *Numerical Algorithms*, 36:73–93, 2004.

[36] S. Nah, T. Kim, and K. Lee. Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring. In *CVPR 2017*, 3883–3891, 2017.

[37] D. O'Connor, and L. Vandenberghe. *Total Variation Image Deblurring with Space-Varying Kernel*. *Computational Optimization and Applications*, 67:

521–541, 2017.

[38] D. O'Leary, J. Nagy, and P. C. Hansen. *Deblurring Images: Matrices, Spectra, and Filtering*. SIAM, 2007.

[39] J. Pan, Z. Hu, Z. Su, and M.-H. Yang. Deblurring Images via L^0-Regularized Intensity and Gradient Prior. In *CVPR 2014*, 2091–2908, 2014.

[40] J. Pan, Z. Hu, Z. Su, and M.-H. Yang. L^0-Regularized Intensity and Gradient Prior for Deblurring Text Images and Beyond. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39,(2): 342–355, 2017.

[41] H. Pu, M. Fan, J. Yang, and F. Lian. Quick Response Barcode Deblurring via Doubly Convolutional Neural Network. *Multimedia Tools and Applications*, 78 (2): 897–912, 2018.

[42] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60:259–268, 1992.

[43] J.L. Sans, F. Argëso, L. Cayón, E. Martínez-González, R.B. Barreiro, and L. Toffolatti. Wavelets Applied to Cosmic Microwave Background Maps: a Multiresolution Analysis for Denoising. *Mon. Not. R. Astron. Soc.*, 309(3): 672–680, 1999.

[44] C. J. Schuler, H. C. Burger, S. Harmeling, and B. Schölkopf. A Machine Learning Approach for Non-Blind Image Deblurring. *CVPR 2013*, 1067–1074, 2013.

[45] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf. Learning to Deblur. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(7): 1439–1451, 2016.

[46] Q. Shan, J. Jia, and A. Agarwala. High-Quality Motion Deblurring from a Single Image. In *SIGGRAPH 2008*, 1–10, 2008.

[47] G. Sörös, S. Semmler, L. Humair, and O. Hilliges. Fast Blur Removal for Wearable QR Code Scanners. In *ISWC 2015*, 117–124, 2015.

[48] J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a Convolutional Neural Network for Non-uniform Motion Blur Removal. In *CVPR 2015*, 769–777, 2015.

[49] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Winston and Sons, 1977.

[50] G. Wang, D. Snyder, J. O'Sullivan, and M. Vannier. Iterative Deblurring for Metal Artifact Reduction. *IEEE Trans. Med. Imaging*, 15,(5): 657–664, 1999.

[51] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. *SIAM J. Imaging Sci.*, 1:248–272, 2008.

[52] W. Wen, M. K. Ng, and Y. M. Huang. Efficient total variation minimization methods for color image restoration. *IEEE Trans. Image Process.*, 17,(11): 2081–2088, 2008.

[53] W. Xu and S. McCloskey, "2D barcode localization and motion deblurring using a flutter shutter camera," In *WACV 2011*, 159–165, 2011.

[54] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," In *NeurIPS 2014*, 1790–1798, 2014.

[55] J. Yang, W. Yin, Y. Zhang, and Y. Wang. A Fast Algorithm for Edge-Preserving Variational Multichannel Image Restoration. *SIAM J. Imaging*

Sci., 2, (2): 569–592, 2009.

[56] "ZBar bar code reader," `http://zbar.sourceforge.net/`

[57] H. Zhang, J. Yang, Y. Zhang and T. Huang. Sparse Representation Based Blind Image Deblurring. In *ICME 2011*, 1–6, 2011.

[58] X. Zhang, R. Wang, X. Jiang, W. Wang, and W. Gao. Spatially variant defocus blur map estimation and deblurring from a single image. *J. Vis. Commun. Image Represent.*, 35:257 – 264, 2016.

[59] H. Zhao, Z. Ke, N. Chen, S. Wang, K. Li, L. Wang, X. Gong, W. Zheng, L. Song, Z. Liu, D. Liang and C. Liu. A new Deep Learning Method for Image Deblurring in Optical Microscopic Systems. *J. Biophotonics*, 13:(3), 2020.

[60] A. Zifan, and P. Liatsis. Medical Image Deblurring via Lagrangian Pursuit in Frame Dictionaries. In *DESE 2011*, 86–91, 2011.

[61] D. Zoran, and Y. Weiss. From Learning Models of Natural Image Patches to whole Image Restoration. In *ICCV 2011*, 479–486, 2011.

[62] H.-O. Georgii. *Gibbs Measures and Phase Transitions*. De Gruyter, 2011.

[63] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[64] J.-W. Park, C. U. Kim, and W. Isard. Permit allocation in emissions trading using the Boltzmann distribution. *Physica A*, 391:(20), 2012.