Blind Deconvolution Problems: A Literature Review and Practitioner’s Guide

Michael R. Kellman

April 2017

1 Introduction

Blind deconvolution is the process of deconvolving a known acquired signal with an unknown point spread function (PSF) or channel encoding, by jointly estimating the system’s input as well as the system’s PSF. Models for this problem often present as the convolution or product of the underlying signal and the system’s PSF. The acquired signal is bi-linear in these two unknown signals (i.e. linear in one if the other is constant), which results in an inverse problem that is non-convex. In addition, the number of observations to uniquely reconstruct the signal and the system’s PSF is often unattainable as the number of degrees of freedom grows as the product of the size of the signal with the size of the PSF, thus the problem is often underdetermined.

While the task seems daunting, there has been a plethora of work over the past twenty years with the aim of performing this task with exact recovery, similar to that of sparse vector recovery in compressed sensing and of low rank matrix recovery in matrix completion. There are a variety of works in the area of exact recovery to pose convex and non-convex problem alike as well as theory regarding under what conditions this is possible. We shall see that advances in these fields compliment the blind deconvolution problem as it can be posed as a linear inverse problem with rank constraint and relaxed to a convex program.

In this literature review the point of view as a practitioner in the areas of signal processing and computational imaging (or computational sensing) is taken. This review is laid out as follows: section 2 goes over into several applications and importance, section 3 reviews the technique of lifting as well as several convex and non-convex problem formulations, section 4 explains some of the theoretical guarantees that accompany convex and non-convex problem formulations, and finally section 5 discusses a broader picture regarding computation and memory for large scale problems.
2 Motivation

Blind deconvolution problems arise in many signal processing and imaging applications due to unknown system models. Measurement and characterization of some system models are infeasible as it is possible they are time-varying or inaccessible and cannot be isolated and measured apriori. Highlighted below are two applications in magnetic resonance imaging (MRI) and photography, but many more exist.

MRI parallel imaging techniques utilize multiple receive coils to spatially encode anatomical information, which allows one to accelerate the scan procedure due to each receive coil’s limited field of view. Because each coil’s spatial encoding is unknown and the image is unknown, this requires each channel to be blindly deconvolved to reconstruct the image. There is much research surrounding the combination of these methods with compressed sensing to enable further acceleration.

Photography can be susceptible to artifacts during longer acquisitions due to
the subject’s motion or due to the camera’s motion. Above illustrates the later, where an image is captured, but the clumsy cameraman moves the camera. To reconstruct the original image, one must also jointly estimate the motion kernel, h. A similar system could characterize deconvolving a blurry image if the kernel were a low pass filter.

3 Methods

The technique of lifting is applied to the blind deconvolution problem to pose it as a linear inverse problem with rank-1 constraint, a non-convex program. Using duality the problem can be relaxed to a convex program and has been shown to be equivalent to a nuclear norm minimization problem.

3.1 Lifting Formulation

From Ahmed et al. [1] the technique of lifting is a method of manipulating unknown quantities from a bi-linear problem to a more convenient and linear problem with a rank constraint. We wish to recover \( x \in \mathbb{R}^L \) and \( w \in \mathbb{R}^L \) from an observed signal \( y \in \mathbb{R}^L \), where \( y \) is the circular convolution of \( x, w \).

\[
y[l] = w \ast x = \sum_{g=0}^{L} w[g]x_L[l - g - 1]
\]  

(1)

Where the subscript \( L \) denotes that the index of the signal is modulo \( L \) making this a circular convolution. We will assume that \( x \) and \( w \) live in specific linear subspaces \( B \in \mathbb{R}^{L \times N} \) and \( C \in \mathbb{R}^{L \times K} \), respectively. Certain specifications will be placed on these subspaces later in section 4 as conditions for recovery.

\[
w = Bh, \quad x = Cm
\]  

(2)

\[
y = \sum_i (C_i m_i) \ast w
\]  

(3)

Equivalently the circular convolution can be represented as a sum of circulant matrix vector products, specifically with a matrix \( \text{circ}(C_i) \), where the \( C_i \) are the columns of \( C \), and a vector, \( w \).

\[
y = \sum_i \text{circ}(C_i) m_i w
\]  

(4)

Expanding \( w \),

\[
y = \sum_i \text{circ}(C_i) B(m_i h)
\]  

(5)

Then expressing the sum as a matrix vector product,
\[ y = \left[ \text{circ}(C_1)B \quad \text{circ}(C_2)B \quad \ldots \quad \text{circ}(C_L)B \right] \begin{bmatrix} m_1h \\ m_2h \\ \vdots \\ m_Lh \end{bmatrix} \]  

From the above equation it is clear our observations are linear measurements of the individual element-wise products of \( m \) and \( h \). These element-wise products can be succinctly ordered in matrix form by representing them as the outer product of \( m \) and \( h \).

\[ X = mh^T \]

Which allows for the forward equation, below, to be represented with \( A \), observed linear measurements of \( X \),

\[ y = A(X) \]

### 3.2 Inverse Problem Formulation

This problem is inherently non-convex because the feasible set is over rank-1 matrices. The inverse problem can then expressed as,

\[ \min_X \frac{1}{2} \| A(X) - y \|_2^2 \]

subject to \( \text{rank}(X) = 1 \)

or an unconstrained problem that is bi-linear,

\[ \min_{u,v} \frac{1}{2} \| A(uv^T) - y \|_2^2 \]

Alternatively, the problem can be written as a least-square problem. This can be thought of as an \( \ell_2 \)-norm regularized version, which is a natural way of solve an underdetermined problem with multiple feasible points.

\[ \min_{u,v} \| u \|_2^2 + \| v \|_2^2 \]

subject to \( A(uv^T) = y \)

### 3.3 Convex Program

Starting with the non-convex formulation (equation 12), duality can be applied to relax the problem to a convex program, specifically a semidefinite program (SDP) [1].
\[
\min_{W_1, W_2, X} \frac{1}{2}(\text{trace}(W_1) + \text{trace}(W_2)) \\
\text{subject to } \begin{bmatrix} W_1 & X \\ X' & W_2 \end{bmatrix} \succeq 0 \\
A(X) = y
\]

Standard SDP solvers such as SeDuMi [2] and SDPT3 [3] can be used to solve this problem.

### 3.4 Nuclear Norm Minimization Problem

From the SDP form, Recht et al. [4] shows that this can be rewritten as a nuclear norm minimization problem. This is accomplished by rewriting the nuclear norm as the dual norm of the operator norm and by applying strong duality.

\[
\min_X \|X\|_* \\
\text{subject to } A(X) = y
\]

This is quite convenient as nuclear norm is a common heuristic for minimizing the rank of a matrix [5]. This heuristic also used for the popular problem of low rank matrix completion [6]. A large body of work surrounds techniques that minimize the nuclear norm, most of which have parallels drawn from minimizing the \(\ell_1\)-norm in lieu of the \(\ell_0\)-norm in compressive sensing.

The above problem can be written as an unconstrained problem using Lagrangian duality.

\[
\min_X \frac{1}{2} \|A(X) - y\|_2^2 + \lambda \|X\|_*
\]

This can be minimized using proximal gradient descent [7, 8] and results in a soft thresholding operation, similar to least-square problems with \(\ell_1\) regularization, with the update,

\[
X^{(k+1)} = \text{prox}_{\lambda \|\cdot\|_*} \left( X^{(k)} - \alpha^{(k)}(A^*(A(X) - y)) \right)
\]

Where the proximal operator for nuclear norm is a soft singular value thresholding, soft thresholding on the singular values. This is derived in Parikh et al. [8]. Similar ideas are presented in Cai et al. [7] for the low-rank matrix completion problem.

Along very similar lines to that of proximal gradient descent, Meka et al. [9] method, singular value projection, proposes solving the non-convex problem, stated below, by projecting on the set of matrices with rank less than \(k\). In our case, the projection operator performs a hard thresholding of the top \(k = 1\) singular values.
\[
\min_{X} \frac{1}{2} \|A(uv^T) - y\|_2^2
\]

subject to \( X \in C(k) = \{Y : \text{rank}(Y) \leq k\} \) \hspace{1cm} (22)

This can be solved iteratively with the update,

\[
X^{(k+1)} = P_{C(k)}(X^{(k)} - \alpha^{(k)}(A^*(A(X) - y)))
\] \hspace{1cm} (23)

Alternatively, an iterative reweighted least-squares (IRLS) problem equivalent to the nuclear norm minimization problem can be solved. This again draws parallels from minimizing the \( \ell_1 \)-norm, which was proposed by Daubechies et al. [10] and proposed by Mohan et al. [11] for nuclear norm minimization. This method involves iterating between solving a least-square problem and updating a weighting matrix,

\[
X^{(k)} = \arg\min_X \text{trace}(W^{(k-1)}X^TX)
\] \hspace{1cm} (24)

subject to \( A(X) = b \) \hspace{1cm} (25)

\[
W^{(k)} = (X^{(k)T}X^{(k)} + \gamma I)^{-\frac{1}{2}}
\] \hspace{1cm} (26)

3.5 Non-Convex Methods

The problem presented in equation 11 is a special form of the matrix factorization problem, where a matrix is represented as the outer product of two rank \( r \) matrices. In our case, \( r = 1 \), we will factor our matrix into the outer product of two vectors. This is similar to the Burer-Monteiro heuristic [4, 12, 13] that uses nonlinear programming to solve an SDP. This adaptation of the heuristic results in the original formalization (equation 11 or 12). A variety of method could be used to solve these formulations such as alternating minimization [14, 15],

\[
u^{(k+1)} = \arg\min_u \frac{1}{2} \|A(u^{(k+1)}v^T) - y\|_2^2
\] \hspace{1cm} (27)

\[
v^{(k+1)} = \arg\min_v \frac{1}{2} \|A(u^{(k+1)}v^T) - y\|_2^2
\] \hspace{1cm} (28)

or with the primal-dual method augmented Lagrangian as presented in [4].

4 Theoretical Guarantees

Most of these methods provide some guarantees for exact recovery if certain requirements are fulfilled. The requirements usually come in either of two similar forms: (1) the sensing matrix must fulfill some restricted isometry property
(RIP), (2) on the number of samples observed. The definition of RIP for matrices appears similar to that of RIP for vectors [16], as used in compressed sensing. The RIP condition requires that the sensing matrix, $A$, is well conditioned. As stated in Recht et al.[4],

**Definition 4.1** Let $A : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$ be a linear map. Without loss of generality, assume $m \leq n$. For every integer $r$ with $1 \leq r \leq m$, define the $r$-restricted isometry constant to be the smallest number $\delta_r(A)$ such that

$$
(1 - \delta_r(A))\|X\|_F \leq \|A(X)\| \leq (1 + \delta_r(A))\|X\|_F
$$

holds for all matrices $X$ of rank at most $r$.

### 4.1 Convex Problems

In Ahmed et al. [1], conditions for exact recovery are placed on the minimum number of observations required, $L$, and this is lower bounded by properties of the aforementioned subspaces $B$ and $C$ and signals $h$ and $m$.

- B is orthonormal
- $B$ is diffused or incoherent in the Fourier domain (i.e. the energy is evenly spread along each column of $B$, denoted $b_l$ for $0 \leq l \leq L$). The quantitative amount of $B$'s incoherence is define by $\mu_1$ and $\mu_L$, the spread of energy in the columns of $B$.

$$
\mu_1^2 = \frac{L}{K} \max_l \|b_l\|_2^2
$$

$$
\mu_L^2 = \frac{L}{K} \min_l \|b_l\|_2^2
$$

- $w$ is incoherent in the Fourier domain. The metric of this is similar to that of $B$.

$$
\mu_h^2 = \frac{L}{K} \max_l \|\langle b_l, h \rangle\|_2^2
$$

- $C$ is drawn isotropically from a zero mean fixed variance Gaussian distribution.

$$
C_{ij} \sim \text{Normal}(0, \sigma^2)
$$

As stated from Ahmed et al. [1],

**Theorem 4.1** Suppose the bases $B, C$ and expansion coefficients $h \in \mathbb{R}^K, m \in \mathbb{R}^N$ satisfy the aforementioned conditions. Then for fixed $\alpha \geq 1$ there exists a constant $C(\alpha) = O(\alpha)$, such that if
\[ L \geq C_{\alpha} \max(\mu_1^2 K, \mu_2^2 N) \log^3(NK) \]  

then \( \mathbf{X} = \mathbf{h} \mathbf{m}^T \) is the unique solution to a nuclear norm minimization problem \( \text{wp. } 1 - O(L(NK)^{-\alpha}) \).

This implies that the if the coherence of \( \mathbf{B} \) is low, then the number of observations required is within a logarithmic factor of the degrees of freedom. This is a nice result as observing \( O(NK) \) measurements might be infeasible to uniquely determine \( O(NK) \) unknowns.

4.2 Non-convex Problems

Jain et al.\cite{14} presents theoretical results for using alternating minimization to solve the non-convex problem based on the observation matrix, \( \mathbf{A} \), fulfilling RIP conditions. As stated for rank one matrices,

**Theorem 4.2** Let \( \mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \) be a rank-1 matrix with non-negative singular value \( \sigma_1 \). If \( \mathbf{A} \) be 2-RIP with constant \( \delta_2 = \frac{1}{100} \), then using alternating minimization, for all \( T > 2 \log(\frac{\|\mathbf{X}\|_F}{\epsilon}) \), iterates \( \mathbf{U}, \mathbf{V} \) satisfy,

\[ \|\mathbf{X} - \mathbf{U} \mathbf{V}^T\|_F \leq \epsilon \]  

This says that alternating minimization with good initialization will give linear convergence \( O(\log(\frac{1}{\epsilon})) \). Some nuclear norm minimization algorithms achieve linear convergence \( \cite{9} \), but require SVD to performed each iteration and storage of the full matrix.

5 Discussion: Storage and Computation

While iteration complexity theory proves that convergence and exact recovery can be achieved with a variety of formulations and algorithms, the complexity of each iteration and storage requirements vary and must be considered when applying these methods. From a signal processing and computational imaging perspective, the size of data sets can be inhibitive large and often one is attempting to reconstruct an even larger structure (compressive imaging reconstruction). If in addition, one needs to solve a blind deconvolution problem the memory requirements might be infeasible.

In the nuclear norm minimization and SDP formulations, the matrix \( \mathbf{X} \) is required to be held in \( O(NK) \) memory. When minimizing nuclear norm, algorithms like projected gradient methods, singular value projection \( \cite{9} \), achieve linear convergence \( \cite{14} \), but require an SVD at every iteration to perform the projection. For SDP interior point methods, again linear convergence can be achieved, but a Newton step is usually taken at each iteration and requires a large amount of computation to calculate.
In contrast using either the bi-linear formulation or the adapted Burer-Monteiro heuristic, the full matrix does not have to be stored in memory. Both of these methods exploit the problem’s natural rank-1 structure to efficiently represent iterates with $O(N+K)$ memory. In addition, alternating methods that aim to minimize the bi-linear problem can be reduced to solving two least-square problems each iteration.

When incorporating additional priors the bi-linear framework is the most flexible for this. For example, if one wishes to incorporate sparsity as a prior of the signal, then this can be simply added as a regularization penalty. While this adds some complexity to solving each iteration, the overall algorithm should still converge at the same rate. This is particularly interesting useful when attempting compressive sensing reconstructions or deconvolutions when the system’s PSF is not known.

6 Conclusion

From the view of the imaging practitioner, computation and memory must be considered when solving blind deconvolution problems. In practice I would recommend using non-convex formulations to solve the problem, as they provide efficiency in both memory and computation. In addition, some theoretical results accompany these methods thanks to Jain et al. [14] and the Burer-Monteiro heuristic [12, 13].

References


