ECE 273 Course Project: Blind Deconvolution

ECE 273 Convex Optimization and Applications

Sheng-Wei Chang
Department of Electrical and Computer Engineering
University of California, San Diego
s7chang@eng.ucsd.edu

I. Objective and Goal

The objective of this project is to review the paper, Blind Deconvolution Using Convex Programming, by Ahmed, Recht, and Romberg. I will answer the proposed questions in the "Paper Review" section, and tried to reproduce the experiment mentioned in the paper.

II. Background and motivation

In this paper, we discuss the problem of recovering two vectors $\boldsymbol{w}$ and $\boldsymbol{x}$ from their convolution. This problem is important in the field of signal processing and communication. If we receive some convolution signal, we can use this technique to separate the original two signals. For example, one of its applications is "Multipath channel protection using random codes." This application is mentioned in section 1.6 of the paper. This is an important problem in communication, describing a message encoded and convoluted by some unknown channel. Our goal is to decode the message from the convolution. Another example of its application is "Image Deblurring" Given a blurring picture, if we can assume that the blurring picture is the convolution of a picture and another signal, then we can use this technique to recover the original picture from the blurring picture.

III. Paper Review

In this section, I will list the problems proposed by the professor and answer them, based on the knowledge in the paper.

A. Why is the solution to the problem non-unique in general? Can you explain what kind of ambiguities exist?

Given a convolution $\boldsymbol{y}$ of $\boldsymbol{w}$ and $\boldsymbol{x}$, the solution to $\boldsymbol{w}$ and $\boldsymbol{x}$ is non-unique. Review the equation (1) in the paper:

$$\boldsymbol{y} = \boldsymbol{w} * \boldsymbol{x}, \text{ or } y[l] = \sum_{l'=1}^{L} w[l']x[l-l'+1]$$

If given $y = \begin{bmatrix} 8 \\ 22 \\ 40 \\ 26 \\ 12 \end{bmatrix}$, then $w = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $x = \begin{bmatrix} 8 \\ 6 \\ 4 \end{bmatrix}$ is a solution.

Also, $\boldsymbol{w} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}$ is another solution. Generally speaking, any $\boldsymbol{w} = k \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$, $\boldsymbol{x} = \frac{1}{k} \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}$, for all $k \in \mathbb{R}, k \neq 0$, can satisfy $\boldsymbol{y} = \boldsymbol{w} * \boldsymbol{x}$. Thus, the solution to the deconvolution problem is non-unique in general.

As we know in the last question, the solution is non-unique. If we don't have any constraint on $\boldsymbol{w}$ and $\boldsymbol{x}$, then we cannot find exact original vectors.

B. Under what conditions on the signals (sparsity, subspace structure etc), the problem can permit unique solutions?

In the "Introduction" section, it mentions that "if both $\boldsymbol{w}$ and $\boldsymbol{x}$ have length L, $\boldsymbol{w}$ lives in a fixed subspace of dimension K and is spread out in the frequency domain, and $\boldsymbol{x}$ lives in a "generic" subspace chosen at random, then $\boldsymbol{w}$ and $\boldsymbol{x}$ are separable with high probability." To be more specifically, we need more information about $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{C}}$ to realize the conditions on $\boldsymbol{w}$ and $\boldsymbol{x}$. In the paper, it also mentions that "we can guarantee the effectiveness of (9) for relatively large subspace dimensions K and N when B is incoherent in the Fourier domain, and when C is generic." About the coherence, if the signal $\boldsymbol{w}$ is more or less "flat" in the frequency domain, then μ_h will be small. Thus,these explain the conditions on the signal to permit unique solutions for the deconvolution.

C. Why is the problem non-convex? Explain the main idea in this paper under which the measurements can be expressed as a linear function of a suitable transformed variable.

After the discrete Fourier transformation($\hat{\boldsymbol{y}} = \boldsymbol{F}\boldsymbol{y}$) and the expansion, we can view the deconvolution problem as a linear inverse problem. What we want to recover is a $K \times N$ matrix from the result of a linear operator $\mathcal{A}$

$$\hat{\boldsymbol{y}} = \boldsymbol{\mathcal{A}}(\boldsymbol{X_0})$$

The measurements are showed in the equation (5) in the paper:

$$\hat{y}_l = <\hat{\boldsymbol{c}}_l, \boldsymbol{m}><\boldsymbol{h}, \hat{\boldsymbol{b}}_l> = \text{trace}(\boldsymbol{A}_l^*(\boldsymbol{h}\boldsymbol{m}^*))$$

The suitable transformed variable $\boldsymbol{X_0}$ is $\boldsymbol{h}\boldsymbol{m}^*$. With this linear operator and the transformed variable. We can recast the original problem as a matrix recovery problem. What we want to solve is described in the equation (6),

$$\min_{u,v} \|\boldsymbol{u}\|_2^2 + \|\boldsymbol{v}\|_2^2$$

$$\text{subject to } \hat{y}_l =< \hat{\boldsymbol{c}_l}, \boldsymbol{u} >< \boldsymbol{v}, \hat{\boldsymbol{b}_l} >, l = 1, ..., L$$

As we can observe, the cost function $\min_{u,v} \|\boldsymbol{u}\|_2^2 + \|\boldsymbol{v}\|_2^2$ is convex. However, the quadratic equality constraints imply that the feasible set is not convex. Therefore, this is a non-convex quadratic optimization problem.

D. How is this transformed variable related to the constituent signals, and given this transformed variable, how can one obtain the constituent signals?

$\boldsymbol{X_0}$ is the outer product of $\boldsymbol{h}$ and $\boldsymbol{m}$,

$$\boldsymbol{X_0} = \boldsymbol{hm}^*$$

where $\boldsymbol{h}$ and $\boldsymbol{m}$ have the following relation to the original constituent signals $\boldsymbol{w}$ and $\boldsymbol{x}$,

$$\boldsymbol{w} = \boldsymbol{Bh}, \boldsymbol{x} = \boldsymbol{Cm}$$

Given this variable $\boldsymbol{X_0}$, we can separate $\boldsymbol{h}$ and $\boldsymbol{m}$ because we have $\|\boldsymbol{h}\|_{\boldsymbol{2}} = 1$. Take one of the rows of $\boldsymbol{X_0}$ and normalize it, then we can get $\boldsymbol{h}$. Secondly, $\boldsymbol{m}$ can be calculated by divide $\boldsymbol{X_0}$ by $\boldsymbol{h}$. We can get $\boldsymbol{w}$ and $\boldsymbol{x}$ by multiplying $\boldsymbol{B}$ and $\boldsymbol{C}$ separately.

E. Does the variable transformation make the problem convex?

No, the variable transformation makes the original problem be the equation (6) in the paper.

$$\min_{u,v} \|\boldsymbol{u}\|_2^2 + \|\boldsymbol{v}\|_2^2$$

$$\text{subject to } \hat{y}_l =< \hat{\boldsymbol{c}_l}, \boldsymbol{u} >< \boldsymbol{v}, \hat{\boldsymbol{b}_l} >, l = 1, ..., L$$

As I answered in problem D, this is non-convex.

F. What additional relaxations are necessary to cast it as a convex problem?

Using duality, the equation (7) in the paper shows the dual of equation (6) is a semi-definite program, and the equation (8) shows the dual of this semi-definite program. Using duality twice, the equation (7) and (8) in the paper show that the problem can be cast as the following:

$$\min \|\boldsymbol{X}\|_*$$

$$\text{subject to } \hat{\boldsymbol{y}} = \boldsymbol{\mathcal{A}}(\boldsymbol{X})$$

This is dual-dual relaxation of the original non-convex least square problem, and this becomes a convex problem.

G. What are the theoretical guarantees of the convex algorithm proposed in the paper? Are they deterministic, or probabilistic?

Theorem 1 in the paper gives a theoretical guarantee that if equation (16) is satisfied, then $\boldsymbol{X_0} = \boldsymbol{hm}^*$ is the unique solution ot equation (9) with probability $1 - O(L^{-\alpha-1})$, and we can recover the signal $\boldsymbol{w}$ and $\boldsymbol{x}$ from $\boldsymbol{y}$.

Since the entries of $\boldsymbol{C}$ are random variables, this theoretical guarantee for the convex algorithm is probabilistic. If $\boldsymbol{C}$ is fixed and the constraints mentioned above hold, then the result is deterministic, i.e. we can find the unique solutions for the original problem.

H. What should be the relation of the number of measurements, sparsity and the dimension of the subspace, so that exact (or stable, in presence of noise) recovery is possible? Clearly state the mathematical guarantees and spell out the assumptions.

The number of measurements is L. Sparsity is related to μ. The dimensions of the subspace are K and M. The exact recovery is also possible in the presence of noise. **Theorem 2** gives the theoretical guarantees. We can recast the problem with noise as equation (17), and this is also a convex problem. Similar to **Theorem 1**, with probability $1 - O(L^{-\alpha-1})$, the solution $\hat{\boldsymbol{X}}$ will obey equation (18). The performance of the algorithm relies on the conditioning of $\mathcal{A}\mathcal{A}^*$.

I. Compare the performance of the blind deconvolution algorithm against non-blind deconvolution (i.e. where you are given one of the constituent signals). Do a literature survey on alternating minimization techniques for blind deconvolution (which are non-convex), and implement one algorithm of your choice. Compare the performance of convex and non-convex blind deconvolution.

Instead of solving two unknown signals, non-blind deconvolution problem is given one signal to be known. However, there still may not be a unique solution, since the circulant matrix may not be invertible. We can use the pseudo-inverse of the circulant matrix. If we call the circulant matrix as C', then we can write $\boldsymbol{y} = C'\boldsymbol{w'}$, where $\boldsymbol{w'} = \begin{bmatrix} m(1)\boldsymbol{w} \\ ... \\ m(N)\boldsymbol{w} \end{bmatrix}$. Applying the pseudo-inverse of C', we have the normal equation

$$C'^T\boldsymbol{y} = C'^T C'\boldsymbol{w'}$$

. We can get

$$w'^* = (C'^T C')^{-1} C'^T \boldsymbol{y} = C'^\dagger \boldsymbol{y}$$

IV. Method and Techniques

I implement the numerical simulation, Phase Transitions, proposed in the paper. In this part, I will describe how I recreate the figures with Python and ”CVXPY” package.

A. Parameters

The first step is to define the basic parameters. Since it is highly time-consuming to solve high dimension convex problem. I choose some lower L, and vary the subspace dimensions N and K and run 20 experiments with different random $\boldsymbol{w}$ and $\boldsymbol{x}$.

B. Functions

I implement some functions to realize the experiment. In this part, I will briefly introduce the functions I write, explaining inputs, output, and how they works.

circular convolution:
The input of this function are $\boldsymbol{w}$ and $\boldsymbol{x}$. The output is their circular convolution. This is simply following the definition of circular convolution in equation (1) in the paper.

checking y hat:
The input of this function are $\hat{\boldsymbol{B}}, \hat{\boldsymbol{C}}, \boldsymbol{h}, \boldsymbol{m}, \hat{\boldsymbol{y}}$. The object of this function is to check whether the $\hat{\boldsymbol{y}}$ that calculated from equation (3) and the $\hat{\boldsymbol{y}}$ that calculated from equation (5) are the same.

cvx:
The input of this function are $L, K, N, \hat{\boldsymbol{B}}, \hat{\boldsymbol{C}}, \hat{\boldsymbol{y}}$. The output of this function is $\boldsymbol{X}$. This function is the key part of this experiment, using CVXPY to calculate the optimal $\boldsymbol{X}$. I follow the equation (9) in the paper to define the objective function and constraints. Before I implement it, I read the tutorial on the website of CVXPY to finish this function.

error rate:
The input of this function are $\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{w}, \boldsymbol{x}$. The output of this function is the error rate. First, I separate $\boldsymbol{X}$ into the optimal $\boldsymbol{h}$ and $\boldsymbol{m}$ and multiply them by $\boldsymbol{B}$ and $\boldsymbol{C}$ relatively to get $\boldsymbol{w}$ and $\boldsymbol{x}$. Following the equation (9) in the paper, we can find the error between the real $\boldsymbol{h}$ and $\boldsymbol{m}$ and the optimal $\boldsymbol{h}$ and $\boldsymbol{m}$.

experiment:
The input of this function are L, K, N, F, R, where F is the discrete Fourier transform (DFT) matrix and R is the number of recovery with same K and N. In my experiment, I set R to be 20. The output of this function is the success rate in 20 times of recovery. We say a recovery a success if its error rate is less than 2%. I have implemented two kinds of experiment. We first form B by randomly selecting K columns from the $L \times L$ identity matrix, and form B from the first K columns of the identity. As for C, the entries of C matrix are independent and identically distributed normal random variables. Also, the elements of both $\boldsymbol{h}$ and $\boldsymbol{m}$ are random normal variables. Given these parameters, we can calculate $\boldsymbol{w}$, $\boldsymbol{x}$, $\boldsymbol{y}$, $\hat{\boldsymbol{B}}$, $\hat{\boldsymbol{C}}$, $\hat{\boldsymbol{y}}$, and call the functions, **checking y hat**, **cvx**, **error rate**. Finally, I repeat 20 times of the above steps to get a success rate among these 20 experiments.

main:
I use those functions to repeat the experiments for all K and N, then I get every success rates. The last step is to plot the figures. I will show the results in the next section.

V. Results

In this part, I show my results that recreate the figures in the paper.

I choose L as 100, K and N from 4 to 50 with step as 2 to find the success rates. In Fig. 1. w is a generic sparse vector. Its entries are chosen randomly. In Fig. 2. w is a generic short vector whose first K terms are nonzero and chosen randomly.

Also, I choose L as 60, K and N from 2 to 40 without skipping to find the success rates. In Fig. 3. w is a generic sparse vector. Its entries are chosen randomly. In Fig. 4. w is a generic short vector whose first K terms are nonzero and chosen randomly.

VI. Discussion

In the figures above, we can easily see a sharp region around which the probability changes from close to zero, to close to 1. However, since the scale of L, K, and N is smaller, thus the edges are not as sharp as the results in the paper. Also, it claims in the paper that the algorithm can deconvolution the signals with high success rate when $L > 2.7(K+N)$ approximately. This is also true in my results. The result also obeys **Theorem 1**.

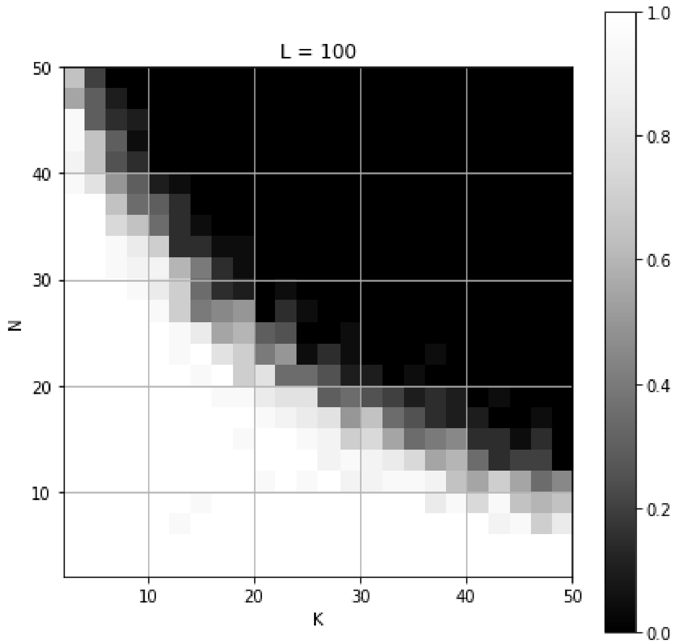

Fig. 1.

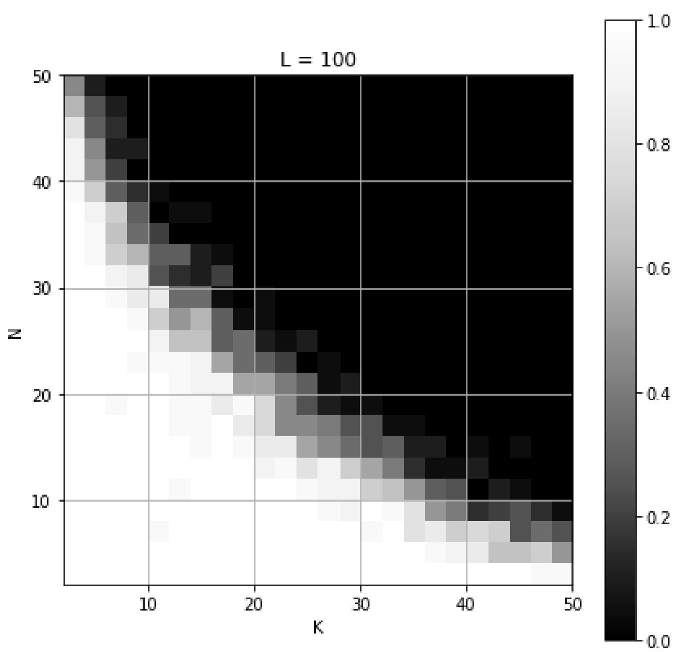

Fig. 2.

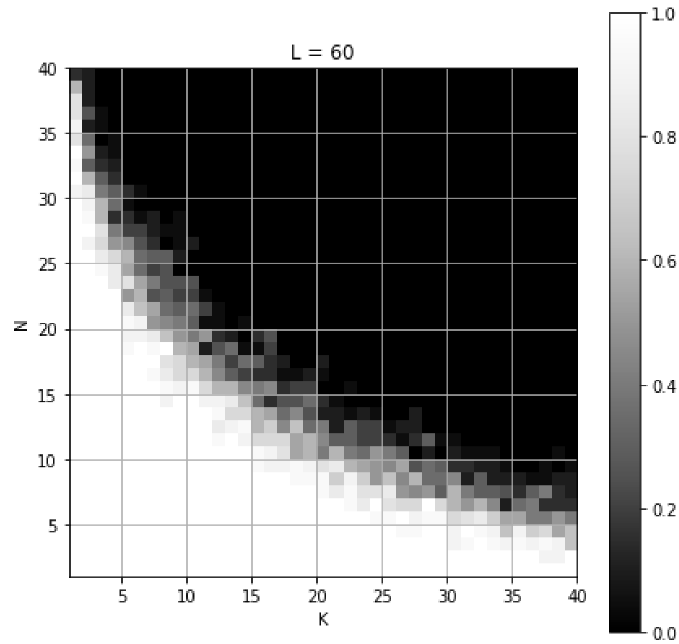

Fig. 3.

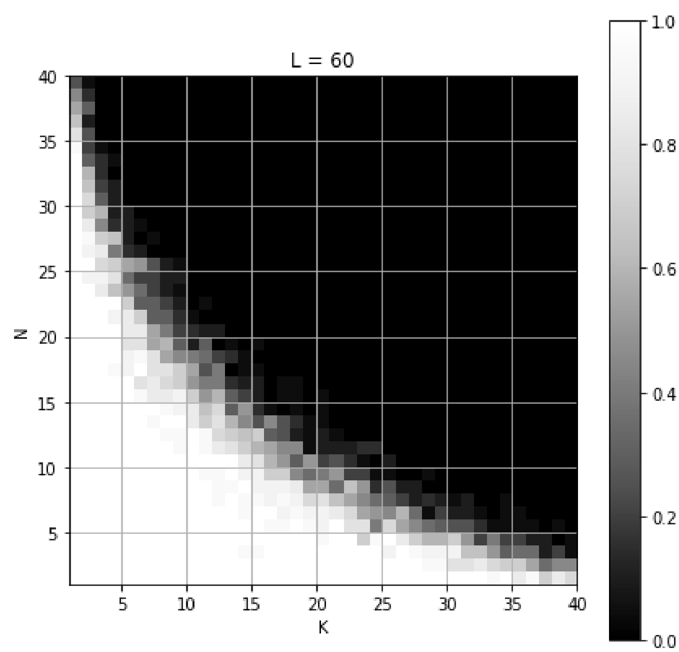

Fig. 4.