Blind Deconvolution Using Convex Programming

Jiaming Cao

Problem Statement – The basic problem

- Consider that the received signal is the circular convolution of two vectors $\boldsymbol{w}$ and $\boldsymbol{x}$, both of length L.

- How can we recover the vectors $\boldsymbol{w}$ and $\boldsymbol{x}$ from the single received signal?

$$\boldsymbol{y} = \boldsymbol{w} * x \quad (\text{or } \boldsymbol{y}[\ell] = \textstyle\sum_{\ell'}^{L} \boldsymbol{w}[\ell']\boldsymbol{x}[\ell - \ell' + 1])$$

$\boldsymbol{w} =?$

$\boldsymbol{x} =?$

Problem Statement – Structural assumptions

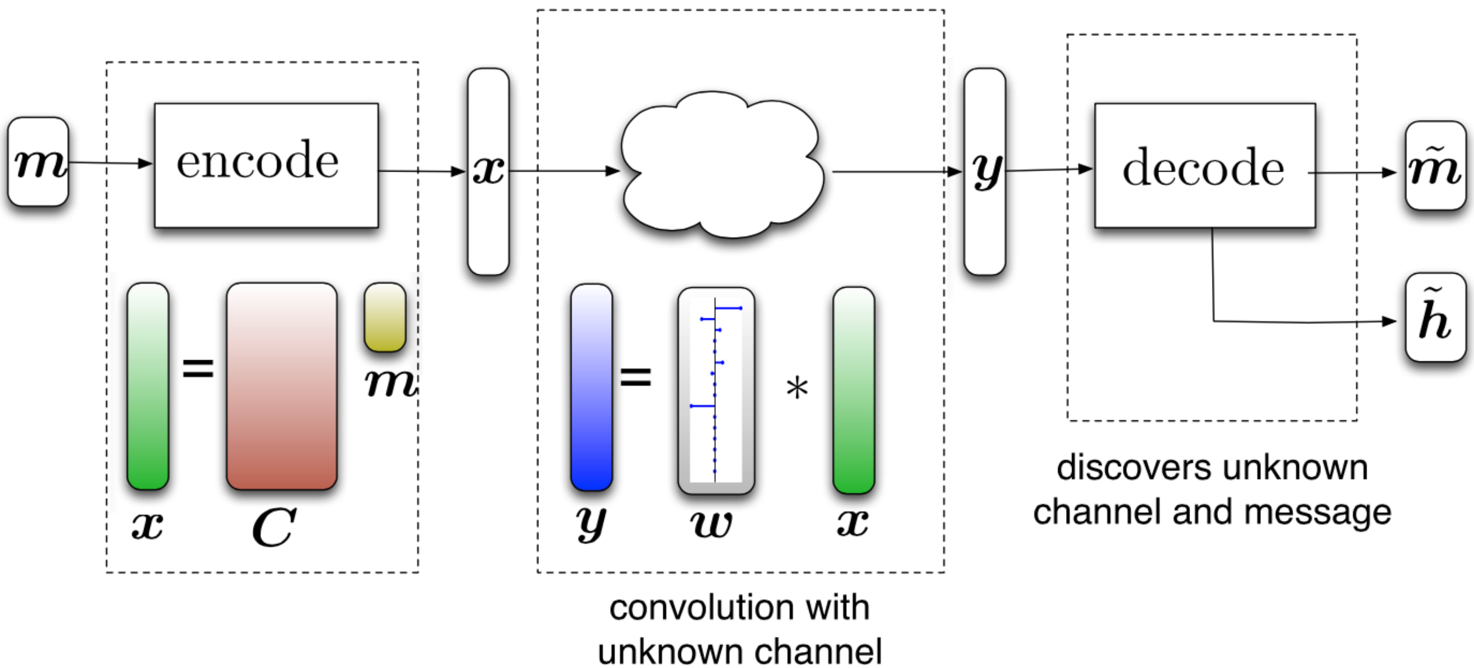
- Assume that w and x live in subspaces with dimensions K and N respectively, *i.e.*

$$\mathbf{w} = \mathbf{B}\mathbf{h}, \quad \mathbf{h} \in \mathbb{R}^K$$
$$\mathbf{x} = \mathbf{C}\mathbf{m}, \quad \mathbf{m} \in \mathbb{R}^N$$

where $\boldsymbol{B}$ is a $L \times K$ matrix, and $\boldsymbol{C}$ is a $L \times N$ matrix.

Knowing matrices $\boldsymbol{B}$ and $\boldsymbol{C}$, reconstructing w and x is equivalent to reconstructing m and h.

Problem Statement – An intuition



Ahmed et al (2014)

Proposed Algorithm – Matrix Observation

Expand the convolution equation using the structural assumption,

$$y = m(1)\boldsymbol{w} * \boldsymbol{C}_1 + \cdots + m(N)\boldsymbol{w} * \boldsymbol{C}_N$$

$$= [circ(\boldsymbol{C}_1) \cdots circ(\boldsymbol{C}_N)] \begin{bmatrix} m(1)\boldsymbol{w} \\ \vdots \\ m(N)\boldsymbol{w} \end{bmatrix}$$

where $circ(\boldsymbol{C}_n)$ denotes the $L \times L$ circulant matrix constructed be nth column of matrix $\boldsymbol{C}$.

Proposed Algorithm – Matrix Observation

Take the Fourier transform, let the DFT matrix by $\boldsymbol{F}$. Then use $\widehat{\boldsymbol{C}} = \boldsymbol{F}\boldsymbol{C}$, $\widehat{\boldsymbol{B}} = \boldsymbol{F}\boldsymbol{B}$, and,

$$\widehat{\boldsymbol{y}} = \boldsymbol{F}\boldsymbol{y} = [\Delta_1 \widehat{\boldsymbol{B}} \quad \cdots \quad \Delta_N \widehat{\boldsymbol{B}}] \boxed{\begin{bmatrix} m(1)\boldsymbol{h} \\ \vdots \\ m(N)\boldsymbol{h} \end{bmatrix}}$$

where $\Delta_n = diag(\sqrt{L}\widehat{\boldsymbol{C}}_n)$

Related to outer product of $\boldsymbol{h}$ and $\boldsymbol{m}$,
$\boldsymbol{h}\boldsymbol{m}^* = [m(1)\boldsymbol{h} \quad \cdots \quad m(N)\boldsymbol{h}]$

Proposed Algorithm – Matrix Observation

$$\widehat{\boldsymbol{y}} = \boldsymbol{F}\boldsymbol{y} = [\Delta_1 \widehat{\boldsymbol{B}} \quad \cdots \quad \Delta_N \widehat{\boldsymbol{B}}] \begin{bmatrix} m(1)\boldsymbol{h} \\ \vdots \\ m(N)\boldsymbol{h} \end{bmatrix}$$

Let $\boldsymbol{X}_0 = \boldsymbol{h}\boldsymbol{m}^*$, using the observation that the operation to get $\widehat{\boldsymbol{y}}$ is linear, we can note the expression as,

$$\widehat{\boldsymbol{y}} = \mathcal{A}(\boldsymbol{X}_0)$$

Further, $\boldsymbol{X}_0$ is a rank 1 matrix by definition. Now we have a way to formulate the recovery of $\boldsymbol{X}_0$.

Proposed Algorithm – Formulation

$$\tilde{\boldsymbol{X}} = \quad \begin{array}{ll} \arg\min & rank(\boldsymbol{X}) \\ s.t. & \hat{\boldsymbol{y}} = \mathcal{A}(\boldsymbol{X}) \end{array}$$

$\Downarrow$ Convex relaxation

$$\begin{array}{ll} \arg\min & \|\boldsymbol{X}\|_* \\ s.t. & \hat{\boldsymbol{y}} = \mathcal{A}(\boldsymbol{X}) \end{array}$$

Let $\tilde{\sigma}\tilde{u}_1\tilde{v}_1$ be the best rank 1 approximation to $\tilde{X}$, then set $\tilde{h} = \sqrt{\tilde{\sigma}}\tilde{u}_1$ and $\widetilde{m} = \sqrt{\tilde{\sigma}}\tilde{v}_1$

Performance Guarantee – Definitions and Assumptions

WOLG, assume columns in B to be orthonormal, such that,

$$B^*B = \hat{B}^*\hat{B} = \sum_{l=1}^{L} \hat{b}_l \hat{b}_l^* = I$$

Define,

$$\begin{cases} \mu_{max}^2 = \dfrac{L}{K} \max\limits_{1 \le l \le L} \left\| \hat{b}_l \right\|_2^2 \in [1, L/K] \\ \mu_{min}^2 = \dfrac{L}{K} \min\limits_{1 \le l \le L} \left\| \hat{b}_l \right\|_2^2 \in [0,1] \\ \mu_h^2 = L \max\limits_{1 \le l \le L} \left| \langle h, \hat{b}_l \rangle \right|^2 \in [1, K\mu_{max}^2] (h \text{ unity norm}) \end{cases}$$

Let,

$$C[l,n] \sim N(0, L^{-1})$$

Performance Guarantee – Theorem 1

Under the above assumptions, fix $\alpha \geq 1$. Then there exists a constant $C_\alpha = O(\alpha)$ depending only on α such that if,

$$\max\left(\mu_{max}^2 K, \mu_h^2 N\right) \leq \frac{L}{C_\alpha\,(\log L)^3}$$

then $\boldsymbol{X}_0 = \boldsymbol{hm}^*$ is the unique solution to the neuclear norm minimization problem with probability $1 - O(L^{-\alpha+1})$, and we can recover both w and x within a scalar multiple from $y = w * x$.

When the coherences are low (*i.e.* μ_{max} and μ_h are on the order of a constant), the inequality is tight to within a logarithmic factor, as we always have $\max\,(K, N) \leq L$

Performance Guarantee – Theorem 1

$$\max\left(\mu_{max}^2 K, \mu_h^2 N\right) \leq \frac{L}{C_\alpha (\log L)^3}$$

As we would like to have the lower bound low, small μ_{max} and μ_h (*i.e.* $\boldsymbol{B}$ "spread out" in frequency domain, or "incoherent") are preferred.

Eg. when $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{I_K} \\ \boldsymbol{0} \end{bmatrix}$, $\mu_{max}^2 = \mu_{min}^2 = 1$

Performance Guarantee – Theorem 2 (stability in presence of noise)

Let the noisy observation be,

$$\widehat{\boldsymbol{y}} = \mathcal{A}(\boldsymbol{X}_0) + \boldsymbol{z}$$

where z is an unknown noise vector with $\|z\|_2 \leq \delta$.

The optimization problem is now,

$$\begin{array}{ll} \arg\min & \|\boldsymbol{X}\|_* \\ s.t. & \|\widehat{\boldsymbol{y}} - \mathcal{A}(\boldsymbol{X})\|_2 \leq \delta \end{array}$$

Performance Guarantee – Theorem 2 (stability in presence of noise)

Let λ_{min} and λ_{max} be the smallest and largest non-zero eigenvalues of $\mathcal{A}\mathcal{A}^*$, then with probability $1 - L^{-\alpha+1}$, the solution to the modified optimization problem will obey,

$$\left\|\tilde{X} - X_0\right\|_F \leq C\,\frac{\lambda_{max}}{\lambda_{min}}\sqrt{\min\left(K, N\right)}\delta$$

for a fixed constant C.

When $\mathcal{A}$ is sufficiently underdetermined, $NK \geq \frac{C_\alpha}{\mu_{min}^2} L(\log L)^2$, then with high probability,

$$\frac{\lambda_{max}}{\lambda_{min}} \sim \frac{\mu_{max}}{\mu_{min}}$$

Performance Guarantee – Theorem 2 (stability in presence of noise)

Set $\tilde{\delta} = \left\|\tilde{X} - X_0\right\|_F$, the there exists a constant C such that,

$$\left\|h - \alpha\tilde{h}\right\|_2 \leq C\min\left(\frac{\tilde{\delta}}{\|h\|_2}, \|h\|_2\right)$$
$$\left\|m - \frac{1}{\alpha}\widetilde{m}\right\|_2 \leq C\min\left(\frac{\tilde{\delta}}{\|m\|_2}, \|m\|_2\right)$$

for some scalar multiple α.

Numerical Simulations – Phase Transition

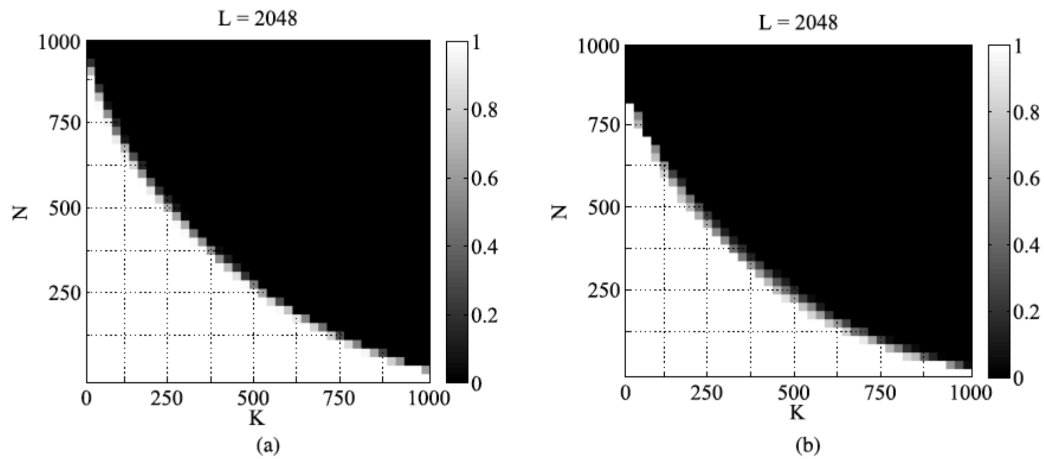


Fig. 3. Empirical success rate for the deconvolution of two vectors $\boldsymbol{x}$ and $\boldsymbol{w}$. In these experiments, $\boldsymbol{x}$ is a random vector in the subspace spanned by the columns of an $L \times N$ matrix whose entries are independent and identically distributed Gaussian random variables. In part (a), $\boldsymbol{w}$ is a generic sparse vector, with support and nonzero entries chosen randomly. In part (b) $\boldsymbol{w}$ is a generic short vector whose first K terms are nonzero and chosen randomly.

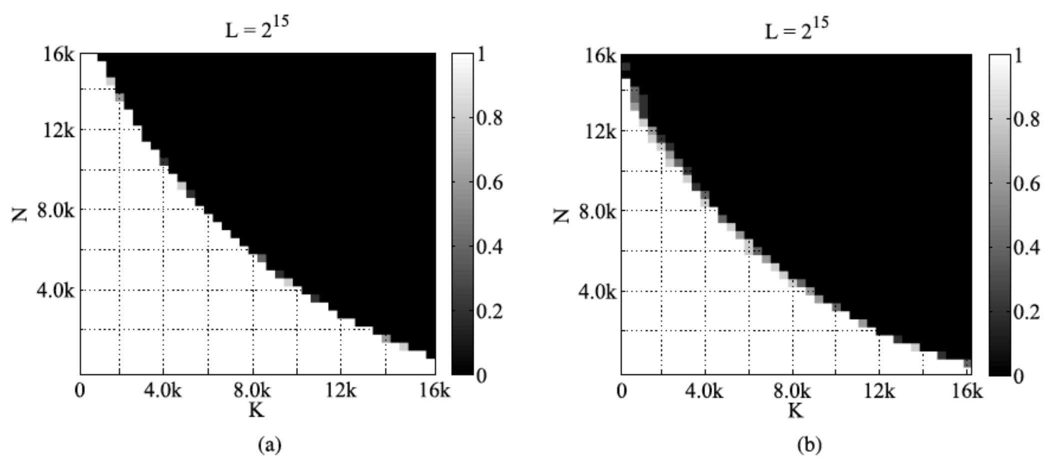
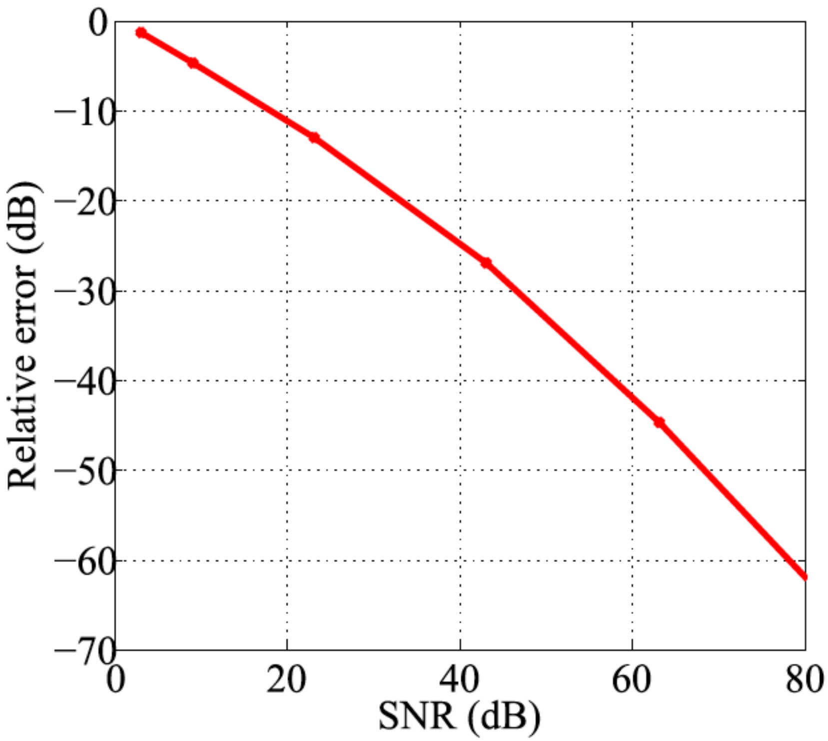
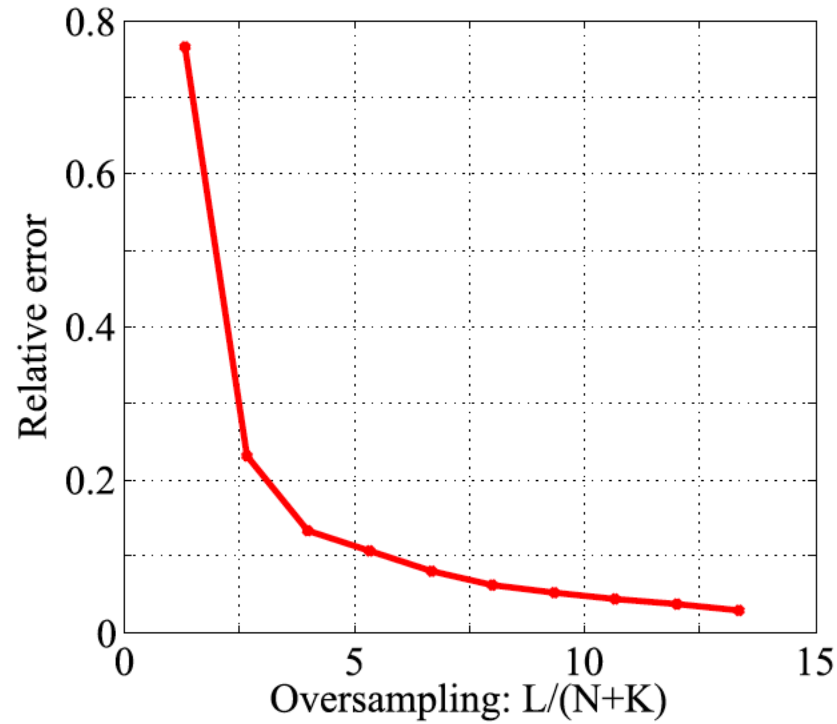


Fig. 4. Empirical success rate for the deconvolution of two vectors $\boldsymbol{x}$ and $\boldsymbol{w}$. In these experiments, $\boldsymbol{x}$ is a random sparse vector whose support and N non-zero values on that support are chosen at random. In part (a), $\boldsymbol{w}$ is a generic sparse vector, with support and K nonzero entries chosen randomly. In part (b) $\boldsymbol{w}$ is a generic short vector whose first K terms are nonzero and chosen randomly.

Ahmed et al (2014)

Numerical Simulations – In Presence of Noise





Ahmed et al (2014)

Toy Example – Image Deblurring

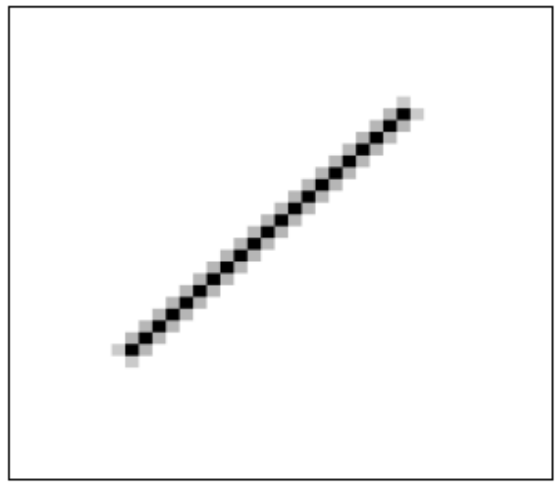
- $x \in R^L$ represents an image of 256x256 pixels, and $w \in R^L$ represents a blur kernel with the same dimension. Therefore, $L = 256 \times 256 = 65536$.

- Let C be a set wavelet basis, and m be the active coefficients in wavelet domain.

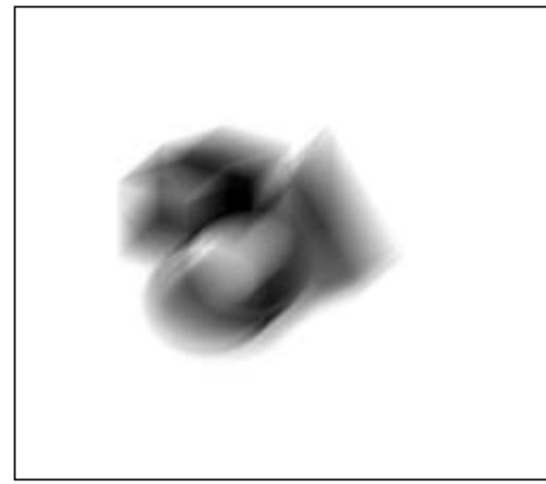
- Let B be formed by a subset of columns in I matrix, and h is an unknown short vector.

Toy Example – Image Deblurring

(a)

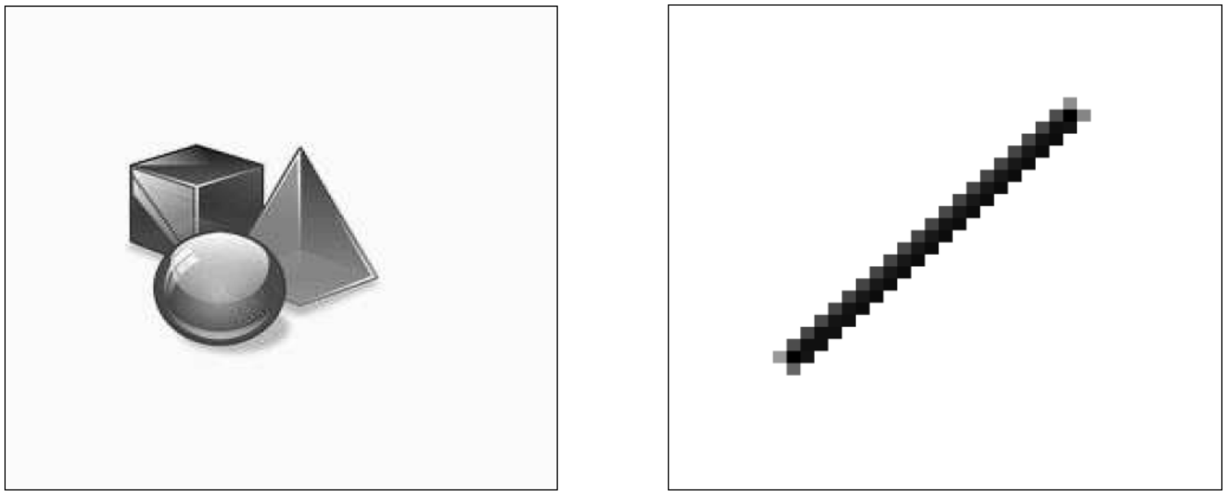


(b)

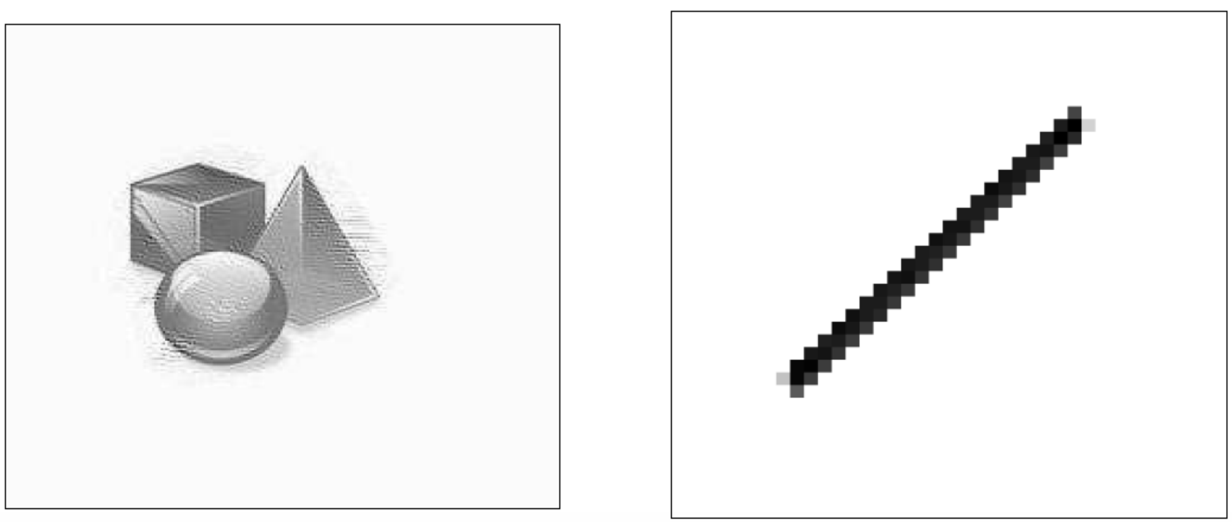


(c)

Toy Example – Image Deblurring



Knowing the support of the original image in wavelet domain



Not knowing the support of the original image in wavelet domain

Comments and Related Works

- Novelty: casting blind deconvolution to a low-rank matrix recovery problem

- Drawback: it is known that SDP is feasible but very expensive, esp. at large scales

- (Li et al.) Speed up using non-convex methods (in presence of noise):

$$\min_{m,h} \|\hat{y} - \mathcal{A}(mh^*)\|^2$$

Sketch of Proof (Theorem 2)

Let $\tilde{X} = X_0 + h, \mathcal{P}_{\mathcal{A}}$ be the projection operator onto the row space of $\mathcal{A}$. By triangular inequality and definition,

$$\|\mathcal{A}(h)\|_2 \leq \|\hat{y} - \mathcal{A}(X_0)\|_2 + \left\|\mathcal{A}\left(\tilde{X}\right) - \hat{y}\right\|_2 \leq 2\delta$$

Recovery error can be decomposed as,

$$\|h\|_F^2 = \|\mathcal{P}_{\mathcal{A}}(h)\|_F^2 + \left\|\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_F^2$$

It can be shown that (details not included, *Proposition 1*) since $\mathcal{P}_{\mathcal{A}^\perp}(h)$ lies in the null space of $\mathcal{A}$,

$$\left\|X_0 + \mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_* - \|X_0\|_* \geq C\left\|\mathcal{P}_{T^\perp}\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_*$$

By triangular inequality, after rearranging,

$$\left\|\mathcal{P}_{T^\perp}\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_* \leq C\left\|\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_* \leq C\sqrt{\min\left(K, N\right)}\|\mathcal{P}_{\mathcal{A}}(h)\|_F$$

It can be shown that (details not included) since $\mathcal{P}_{\mathcal{A}^\perp}(h)$ lies in the null space of $\mathcal{A}$,

$$\left\|\mathcal{P}_T\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_F^2 \leq 2\lambda_{max}^2\left\|\mathcal{P}_{T^\perp}\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_F^2$$

Therefore,

$$\left\|\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_F^2 = \left\|\mathcal{P}_T\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_F^2 + \left\|\mathcal{P}_{T^\perp}\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_F^2 \leq (2\lambda_{max}^2 + 1)\left\|\mathcal{P}_{T^\perp}\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_F^2$$

Sketch of Proof (Theorem 2)

Plug into the expression for recovery error, we get,

$$\|h\|_F^2 \leq \|\mathcal{P}_{\mathcal{A}}(h)\|_F^2 + (2\lambda_{max}^2 + 1)\left\|\mathcal{P}_{T^\perp}\mathcal{P}_{\mathcal{A}^\perp}(h)\right\|_F^2$$

Knowing that Frobenius norm is no greater than nuclear norm, by applying the previous bound on nuclear norm, we have,

$$\|\mathcal{A}(h)\|_2 \leq \|\mathcal{P}_{\mathcal{A}}(h)\|_F^2 + C(2\lambda_{max}^2 + 1)\min(K, N)\|\mathcal{P}_{\mathcal{A}}(h)\|_F^2$$

Absorbing all constants into C,

$$\|h\|_F^2 \leq C\lambda_{max}\sqrt{\min(K, N)}\|\mathcal{P}_{\mathcal{A}}(h)\|_F \leq C\lambda_{max}\sqrt{\min(K, N)}\left\|\mathcal{A}^\dagger\right\|\|\mathcal{A}(h)\|_2$$

where $\mathcal{A}^\dagger$ is the pseudoinverse of $\mathcal{A}$, whose norm is λ_{min}. Also use the previously established inequality on $\|\mathcal{A}(h)\|_2$, the conclusion follows.

Discussion

- What is the benefit of viewing blind deconvolution as a low rank recovery problem?

References

1. Ahmed, Ali, Benjamin Recht, and Justin Romberg. "Blind deconvolution using convex programming." *IEEE Transactions on Information Theory* 60.3 (2014): 1711-1732.
2. Li, Xiaodong, et al. "Rapid, robust, and reliable blind deconvolution via nonconvex optimization." *arXiv preprint arXiv:1606.04933* (2016).