SIGIBE: Solving Random Bilinear Equations via Gradient Descent with Spectral Initialization

Antonio G. Marques[†], Gonzalo Mateos[‡], and Yonina C. Eldar   E-mail: antonio.garcia.marques@urjc.es

[†]Signal Theory & Comms, King Juan Carlos University   † Electrical & Computer Eng., University of Rochester   * Electrical Eng., Technion - Israel Institute of Tech.

EUSIPCO 2016 — 29 August - 2 September 2016, Hilton Budapest | Hungary

Abstract

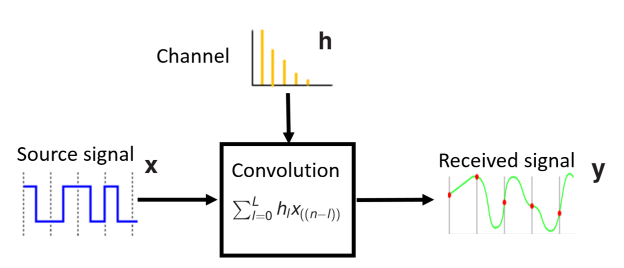
We investigate the problem of finding the real-valued vectors $\mathbf{h}$*, of size L, and* $\mathbf{x}$*, of size P, from M independent measurements* $y_m = \langle \mathbf{a}_m, \mathbf{h}\rangle\langle \mathbf{b}_m, \mathbf{x}\rangle$*, where* $\mathbf{a}_m$ *and* $\mathbf{b}_m$ *are known random vectors. Inspired by phase retrieval solvers, we propose SIGIBE an algorithm that proceeds in two steps: (i) first a spectral method is used to obtain an initial guess; which is then (ii) refined using simple and scalable gradient descent iterations to minimize a natural non-convex formulation of the recovery problem.*

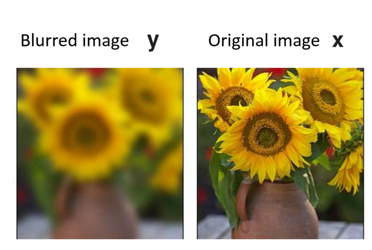
Bilinear problems

- Given a collection of M scalar measurements $y_m \in \mathbb{R}$ of the form

$$y_m = \langle \mathbf{a}_m, \mathbf{h}\rangle\langle \mathbf{b}_m, \mathbf{x}\rangle = \mathbf{a}_m^T\mathbf{h}\cdot\mathbf{b}_m^T\mathbf{x}, \quad m = 1,\dots,M \tag{1}$$

  ⇒ with $M \geq (L+P)$, $\mathbf{a} \in \mathbb{R}^L$ and $\mathbf{b} \in \mathbb{R}^P$ known
- Goal: recover the unknown $\mathbf{h} \in \mathbb{R}^L$ and $\mathbf{x} \in \mathbb{R}^P$
  - ⇒ Up to an inherent scaling ambiguity
- Challenges: non-convex, multiple solutions, scaling ambiguity
- Assumptions:
  - ⇒ $\{\mathbf{a}_m\}_{m=1}^M$ and $\{\mathbf{b}_m\}_{m=1}^M$ are random, zero-mean, identity cov
  - ⇒ Any correlation between $\mathbf{a}_m$ and $\mathbf{b}_m$; including $\mathbf{a}_m = \mathbf{b}_m$
- Many meaningful applications are inverse bilinear problems
  - ⇒ Blind deconvolution (channel equalization or image deblurring)
  - ⇒ Array self-calibration for direction-of-arrival estimation
  - ⇒ Modeling of network diffusion processes





Related problems

- Phase retrieval
  - ⇒ Measurements of the form $y_m = |\langle \mathbf{a}_m, \mathbf{x}\rangle|^2$
  - ⇒ Long history in astronomy, optics and microscopy
  - ⇒ Symmetry not present in our setup
  - ⇒ Many approaches: Phaselift, SDP-based, greedy, gradient
- Bilinear deconvolution by lifting
  - ⇒ y_m bilinear in $\mathbf{x}$ and $\mathbf{h}$, but linear in (rank-one) matrix $\mathbf{x}\mathbf{h}^T$
  - ⇒ rank minimization ⇒ convex relax with performance guarantees
  - ⇒ SDP-based solvers entail higher computational complexity

Contributions

- SIGIBE: two-step gradient-based algorithm
- Arbitrary correlation among measurement vectors
- No lifting ⇒ Smaller computational complexity

Problem formulation

- Measurements $\{y_m\}_{m=1}^M$ given, $\{\mathbf{a}_m\}_{m=1}^M$ and $\{\mathbf{b}_m\}_{m=1}^M$ known
  - ⇒ A natural criterion is to minimize the LS cost

Inverse bilinear problem

$$\{\hat{\mathbf{x}}, \hat{\mathbf{h}}\} = \arg\min_{\{\mathbf{x},\mathbf{h}\}} f(\mathbf{x},\mathbf{h}) := \frac{1}{2M}\sum_{m=1}^M \left(\mathbf{a}_m^T\mathbf{h}\cdot\mathbf{x}^T\mathbf{b}_m - y_m\right)^2. \tag{2}$$

- Problem (2) bilinear, hence non-convex optimization.
- Approach:
  - ⇒ Judicious initialization + simple gradient descent iterations
  - ⇒ Similar to recent ideas for phase retrieval

Gradient iterations

- Let i be the iteration index and $\{\mathbf{x}_0, \mathbf{h}_0\}$ the (spectral) initializations

Gradient iterations

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \mu_{i|x}\nabla_{\mathbf{x}} f(\mathbf{x}_i, \mathbf{h}_i) \tag{3}$$
$$\mathbf{h}_{i+1} = \mathbf{h}_i - \mu_{i|h}\nabla_{\mathbf{h}} f(\mathbf{x}_i, \mathbf{h}_i) \tag{4}$$

- The gradients of $f(\mathbf{x},\mathbf{h})$ are

$$\nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{h}) = \frac{1}{M}\sum_{m=1}^M \left(\mathbf{a}_m^T\mathbf{h}\cdot\mathbf{x}^T\mathbf{b}_m - y_m\right)\left(\mathbf{a}_m^T\mathbf{h}\right)\mathbf{b}_m \tag{5}$$

$$\nabla_{\mathbf{h}} f(\mathbf{x},\mathbf{h}) = \frac{1}{M}\sum_{m=1}^M \left(\mathbf{a}_m^T\mathbf{h}\cdot\mathbf{x}^T\mathbf{b}_m - y_m\right)\left(\mathbf{b}_m^T\mathbf{x}\right)\mathbf{a}_m. \tag{6}$$

- Stepsizes $\mu_{i|x}$ and $\mu_{i|h}$
  - ⇒ Different alternatives ⇒ diff. convergence and recovery
  - ⇒ Simulations will be run with $\mu_{i|x} = \mu_i/\bar{\mu}_{i|x}$
    $\mu_i = \min\{\mu_{\max}, 1 - e^{-i/(-i_{\text{thr}}\ln(1-\mu_{\max}))}\}$ and $\bar{\mu}_{i|x} = \|\mathbf{x}\|^2$
  - ⇒ $\|\mathbf{x}\|^2$ can be known, estimated, or replaced with $\|\mathbf{x}_i\|^2$.
- Computational complexity
  - ⇒ $\mathcal{O}(M(L+P)^2)$ operations per iteration

Initialization I: SVD-based for uncorrelated vectors

- Consider the *non-symmetric* $L \times P$ matrix

$$\mathbf{Y}_{NS} := \frac{1}{M}\sum_{m=1}^M y_m\mathbf{a}_m\mathbf{b}_m^T. \tag{7}$$

- Suppose that $\mathbf{a}_m$ and $\mathbf{b}_m$ are uncorrelated for each $m = 1,\dots,M$

$$\mathbb{E}[\mathbf{Y}_{NS}] = \frac{1}{M}\sum_{m=1}^M \mathbb{E}[\mathbf{a}_m\mathbf{a}_m^T]\mathbf{h}\mathbf{x}^T\mathbb{E}[\mathbf{b}_m\mathbf{b}_m^T] = \mathbf{h}\mathbf{x}^T \tag{8}$$

  - ⇒ Rank-one matrix
  - ⇒ Strong Law of Large Numbers (LLN): $\mathbf{Y}_{NS} \to \mathbb{E}[\mathbf{Y}_{NS}] = \mathbf{h}\mathbf{x}^T$
  - ⇒ If M large, dominant singular vectors align with $\mathbf{h}$ and $\mathbf{x}$
- Simple but instructive initialization based on SVD decomposition

Algorithm 1: Spectral initialization for uncorrelated data

INPUTS: $\{y_m\}_{m=1}^M$, $\{\mathbf{a}_m\}_{m=1}^M$, $\{\mathbf{b}_m\}_{m=1}^M$, and $I_{\max}^P$
OUTPUTS: initial estimates $\mathbf{h}_0$ and $\mathbf{x}_0$
Step 1. Use inputs to find $\mathbf{Y}_{NS}$ and run the iterations in Step 2 for $(i \leq I_{\max}^P)$
Step 2: power method. Generate random $\mathbf{v}_0$ with unit-norm and run $\mathbf{u}_i = \mathbf{Y}_{NS}\mathbf{v}_i/\|\mathbf{Y}_{NS}\mathbf{v}_i\|$ and $\mathbf{v}_{i+1} = \mathbf{Y}_{NS}{}^T\mathbf{u}_i/\|\mathbf{Y}_{NS}{}^T\mathbf{u}_i\|$
Step 3. Return $\mathbf{x}_0 = \sigma\mathbf{v}_{I_{\max}^P}$ and $\mathbf{h}_0 = \sigma\mathbf{u}_{I_{\max}^P}$ with $\sigma^2 = \|\mathbf{Y}_{NS}\mathbf{v}_{I_{\max}^P}\|\|\mathbf{Y}_{NS}^T\mathbf{u}_{I_{\max}^P}\|$

- Low computational complexity
  - ⇒ $\mathcal{O}(MLP)$ to form $\mathbf{Y}_{NS}$ and $\mathcal{O}(I_{\max}^P LP)$ for power method
  - ⇒ Lower than gradient operations

Initialization II: EIG-based for correlated vectors

- Form augmented vectors $\gamma_m := [\mathbf{a}_m^T, \mathbf{b}_m^T]^T \in \mathbb{R}^{L+P}$ and *symmetric* matrix

$$\mathbf{Y}_S = \frac{1}{M}\sum_{m=1}^M y_m\gamma_m\gamma_m^T. \tag{9}$$

  - ⇒ Define $\mathbf{A} \in \mathbb{R}^{(L+P)\times(L+P)}$ symmetric and rank-2

$$\mathbf{A} = \begin{bmatrix}\mathbf{h}\\ \mathbf{0}_P\end{bmatrix}\begin{bmatrix}\mathbf{0}_L^T & \mathbf{x}^T\end{bmatrix} + \begin{bmatrix}\mathbf{0}_L\\ \mathbf{x}\end{bmatrix}\begin{bmatrix}\mathbf{h}^T & \mathbf{0}_P^T\end{bmatrix} = \begin{bmatrix}\mathbf{0}_{L\times L} & \mathbf{h}\mathbf{x}^T\\ \mathbf{x}\mathbf{h}^T & \mathbf{0}_{P\times P}\end{bmatrix}.$$

  - ⇒ It follows that $y_m = (1/2)\gamma_m^T\mathbf{A}\gamma_m$, taking expectations in (9)

$$\mathbb{E}[\mathbf{Y}_S] = \frac{1}{2}\mathbb{E}[\gamma_1\gamma_1^T\mathbf{A}\gamma_1\gamma_1^T]. \tag{10}$$

- Measurement vecs: $\mathbf{a}_m$ and $\mathbf{b}_m$ white, but correlated with $\mathbf{C} = \mathbb{E}[\mathbf{a}_m\mathbf{b}_m^T]$

$$\mathbf{S} = \begin{bmatrix}\mathbf{I}_L & \mathbf{C}\\ \mathbf{C}^T & \mathbf{I}_P\end{bmatrix}. \tag{11}$$

  - ⇒ Fourth order moment of a Gaussian yields

$$\mathbb{E}[\mathbf{Y}_S] = \mathbf{SAS} + (1/2)\text{tr}[\mathbf{AS}]\,\mathbf{S} \tag{12}$$

- Left multiply by $\mathbf{S}^{-1}$ to simplify the second term
  - ⇒ Then it follows that the expected value of $\tilde{\mathbf{Y}} := \mathbf{S}^{-1}\mathbf{Y}_S$ is

$$\mathbb{E}[\tilde{\mathbf{Y}}] = \mathbf{AS} + (1/2)\text{tr}[\mathbf{AS}]\,\mathbf{I}_{L+P} = \begin{bmatrix}\mathbf{h}\mathbf{x}^T\mathbf{C}^T & \mathbf{h}\mathbf{x}^T\\ \mathbf{x}\mathbf{h}^T & \mathbf{x}\mathbf{h}^T\mathbf{C}\end{bmatrix} + (\mathbf{x}^T\mathbf{C}^T\mathbf{h})\mathbf{I}_{L+P}. \tag{13}$$

- Two eigenvectors: $\mathbf{v}_1 = \left[\mathbf{h}^T/\|\mathbf{h}\|,\ \mathbf{x}^T/\|\mathbf{x}\|\right]^T$, $\mathbf{v}_2 = \left[-\mathbf{h}^T/\|\mathbf{h}\|,\ \mathbf{x}^T/\|\mathbf{x}\|\right]^T$
- Simple EIG-based initialization (power method)

Algorithm 2: Spectral initialization for correlated data

INPUTS: $\{y_m\}_{m=1}^M$, $\{\mathbf{a}_m\}_{m=1}^M$, $\{\mathbf{b}_m\}_{m=1}^M$, $\mathbf{C}$, and $I_{\max}^P$
OUTPUTS: initial estimates $\mathbf{h}_0$ and $\mathbf{x}_0$
Step 1: Finding $\mathbf{z}^*$. Use inputs to find $\tilde{\mathbf{Y}}$ and get eigenvector $\mathbf{z}^* = \mathbf{z}_{I_{\max}^P}$ using a power method $\mathbf{z}_i = \tilde{\mathbf{Y}}\mathbf{z}_{i-1}/\|\tilde{\mathbf{Y}}\mathbf{z}_{i-1}\|$ for $i = 1,\dots,I_{\max}^P$
Step 2: Finding the initializations $\tilde{\mathbf{h}}_0$ and $\tilde{\mathbf{x}}_0$ using $\mathbf{z}^*$.
Extract $\bar{\mathbf{z}}^{top} := [z_1^*,\dots,z_L^*]^T$, $\bar{\mathbf{z}}^{bot} := [z_{L+1}^*,\dots,z_{L+P}^*]^T$ from $\mathbf{z}^*$
Normalize $\bar{\mathbf{z}}_h := \bar{\mathbf{z}}^{top}/\|\bar{\mathbf{z}}^{top}\|$, $\bar{\mathbf{z}}_x := \bar{\mathbf{z}}^{bot}/\|\bar{\mathbf{z}}^{bot}\|$
Stack $\bar{\mathbf{z}}_h$ and $\bar{\mathbf{z}}_x$ in $\mathbf{v}_A := \frac{1}{\sqrt{2}}[\bar{\mathbf{z}}_h^T, \bar{\mathbf{z}}_x^T]^T$, $\mathbf{v}_B := \frac{1}{\sqrt{2}}[-\bar{\mathbf{z}}_h^T, \bar{\mathbf{z}}_x^T]^T$
Compute $\lambda_A = \|\tilde{\mathbf{Y}}\mathbf{v}_A\|$, $\lambda_B = \|\tilde{\mathbf{Y}}\mathbf{v}_B\|$ and $\lambda_{xh} = (\lambda_A + \lambda_B)/2$
Set $\tilde{\mathbf{h}}_0 = \sqrt{\lambda_{xh}}\bar{\mathbf{z}}_h$ and $\tilde{\mathbf{x}}_0 = \sqrt{\lambda_{xh}}\bar{\mathbf{z}}_x$.
Step 3: Fixing the sign of the initializations.
If $\text{sign}([\tilde{\mathbf{Y}}]_{1,L+1}) = \text{sign}([\tilde{\mathbf{h}}_0\tilde{\mathbf{x}}_0^T]_{1,1})$, return $\mathbf{h}_0 = \tilde{\mathbf{h}}_0$ and $\mathbf{x}_0 = \tilde{\mathbf{x}}_0$.
If not, return $\mathbf{h}_0 = -\tilde{\mathbf{h}}_0$ and $\mathbf{x}_0 = \tilde{\mathbf{x}}_0$.

  - ⇒ As $M \to \infty$: p1) $\|\mathbf{h}_0\| = \|\mathbf{x}_0\| = \sqrt{\|\mathbf{h}\|\|\mathbf{x}\|}$ and p2) $\mathbf{h}_0\mathbf{x}_0^T = \mathbf{h}\mathbf{x}^T$
  - ⇒ $\mathbf{S}^{-1}$ pre-whitening
- Computational complexity higher than that for Algorithm 1
  - ⇒ Larger matrix and $\tilde{\mathbf{Y}}$ requires inverting block matrix $\mathbf{S}$
  - ⇒ Cost dominated by Step 1:
    - a) $\mathcal{O}(I_{\max}(L+P)^2)$ for power method, and
    - b) $\mathcal{O}((L+P)^3)$ for $\mathbf{S}^{-1}$ and $\mathcal{O}(M(L+P)^2)$ for $\tilde{\mathbf{Y}}$
  - ⇒ Overall cost still dominated by gradient step $\mathcal{O}(I_{\max}^G M(L+P)^2)$

Initialization II: Special cases

Fully uncorrelated: $\mathbf{a}_m \perp \mathbf{b}_m$ **for** $m = 1,\dots,M$
- $\mathbf{C} = \mathbf{0}_{L\times P}$ and $\mathbf{S} = \mathbf{I}_{L+P}$
- Simplified (12): $\mathbb{E}[\mathbf{Y}_S] = \mathbf{A}$ (notice that $\text{tr}[\mathbf{A}] = 0$).
- EIGs $\mathbb{E}[\mathbf{Y}_S] = \mathbf{A}$:
  1) $\mathbf{v}_1 = \frac{1}{\sqrt{2}}\left[\mathbf{h}^T/\|\mathbf{h}\|,\ \mathbf{x}^T/\|\mathbf{x}\|\right]^T$ with $\lambda_1 = \|\mathbf{x}\|\|\mathbf{h}\|$; 2) $\mathbf{v}_2 = \frac{1}{\sqrt{2}}\left[-\mathbf{h}^T/\|\mathbf{h}\|,\ \mathbf{x}^T/\|\mathbf{x}\|\right]^T$ with $\lambda_2 = -\|\mathbf{x}\|\|\mathbf{h}\|$; and 3) $\lambda_n = 0$ for $n > 2$.

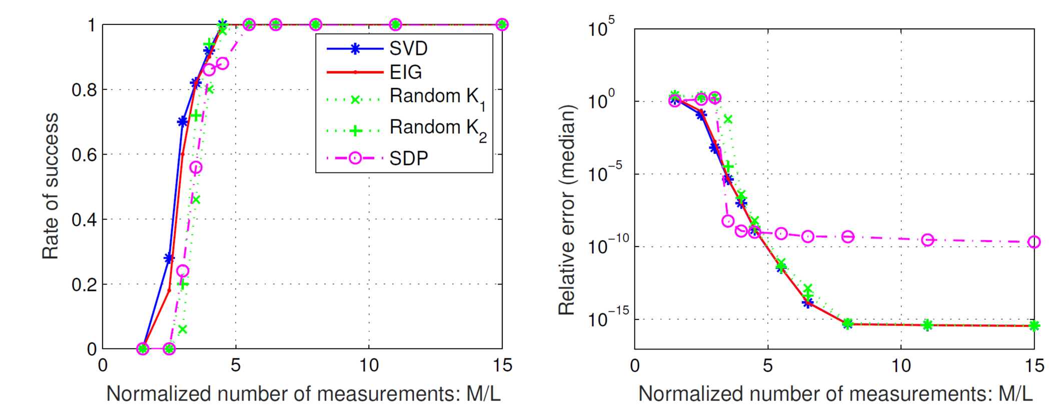
Fully correlated: $\mathbf{a}_m = \mathbf{b}_m$
- $\mathbf{C} = \mathbf{I}_P \Rightarrow \mathbf{S} = \mathbf{I}_{2P}$
- Simplified (12): each of the four blocks in $\mathbb{E}[\mathbf{Y}_S]$ are identical
- EIG of block $\mathbf{B} = [\mathbb{E}[\mathbf{Y}_S]]_{1:P,1:P} = \mathbf{h}\mathbf{x}^T + \mathbf{x}\mathbf{h}^T + (\mathbf{x}^T\mathbf{h})\mathbf{I}_P$:
  1) $\mathbf{v}_1 = \mathbf{x}/\|\mathbf{x}\| + \mathbf{h}/\|\mathbf{h}\|$ with $\lambda_1 = 2\mathbf{x}^T\mathbf{h} + \|\mathbf{x}\|\|\mathbf{h}\|$; 2) $\mathbf{v}_2 = \mathbf{x}/\|\mathbf{x}\| - \mathbf{h}/\|\mathbf{h}\|$ with $\lambda_2 = 2\mathbf{x}^T\mathbf{h} - \|\mathbf{x}\|\|\mathbf{h}\|$; and 3) $\lambda_n = \mathbf{x}^T\mathbf{h}$ for $n > 2$
- Useful to simplify Alg. 2: smaller matrix and no $\mathbf{S}^{-1}$ pre-whitening

Numerical experiments: setup

- Setup: $\mathbf{x}$ and $\mathbf{h}$ zero-mean Gaussians with $\sigma_x^2 = 4^2$ and $\sigma_h^2 = 1^2$; $I_{\max}^G = 500$; $\mu_{\max} = 0.4$, $i_{\text{thr}} = 75$, $\bar{\mu}_{i|x} = \|\mathbf{x}_i\|^2$ and $\bar{\mu}_{i|h} = \|\mathbf{h}_i\|^2$
- Five algorithms (results are averaged across 100 trials):
  - ⇒ A1) SIGIBE using Algorithm 1 and A2) using Algorithm 2 for $\mathbf{C} = \mathbf{0}$
  - ⇒ A3) Random initializ. with $K_1 = 5$ seeds and A4) $K_2 = 15$ seeds
  - ⇒ A5) SDP relaxation based on matrix lifting
- Metric: $\text{err} = \|\mathbf{x}\mathbf{h}^T - \hat{\mathbf{x}}\hat{\mathbf{h}}^T\|_F/\|\mathbf{x}\mathbf{h}^T\|_F$

Numerical experiments I: Uncorrelated case

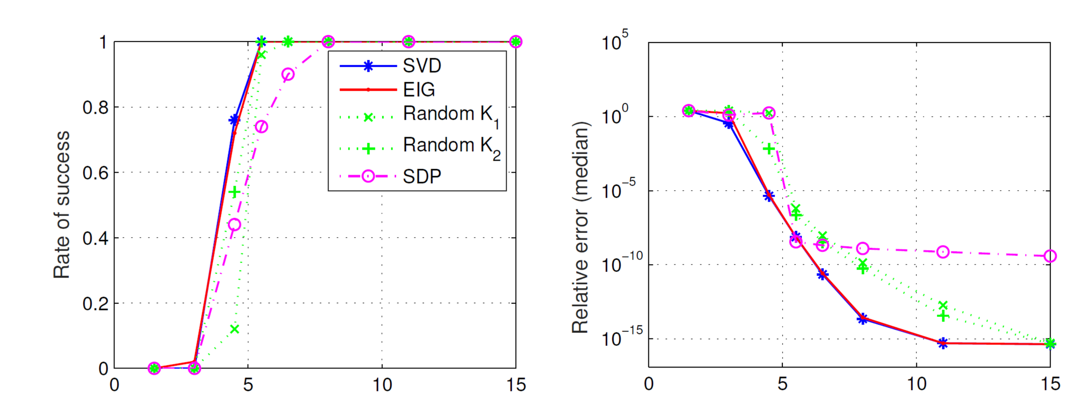
- No correlation: $\mathbf{C} = \mathbf{0}$, $P = 64$, $L = 2P = 128$



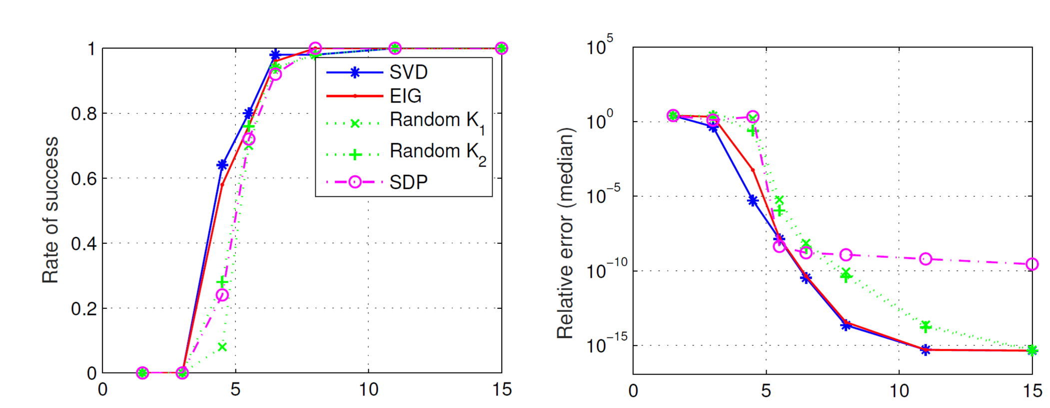
- Observations
  - ⇒ If $M \leq L + S = 1.5L$ all fail
  - ⇒ If $M \geq 8L$ all work
  - ⇒ A1 best performance
  - ⇒ Speed: A1, 1.1A1, 5.0A1, 14.9A1, 20.4A1

Numerical experiments II: Correlated case

- Slight correlation: $\mathbf{C} = 0.25\mathbf{I}$, $P = L = 128$



- Strong correlation: $\mathbf{C} = 0.75\mathbf{I}$, $P = L = 128$



- Observations
  - ⇒ The higher the correlation, the more difficult
  - ⇒ For $\rho = 0.25$, $M = 5.5L$
  - ⇒ For $\rho = 0.75$, $M = 6.5L$
  - ⇒ A1 works well even in the correlated case

Conclusions and future work

- Non-convex algorithm for inverse bilinear problems
  - ⇒ Gradient descent plus spectral initializations
  - ⇒ Different forms of correlation among measurement vectors
- Develop theoretical recovery guarantees
- Extension to the complex case
- Explore the fact that SVD works well for the correlated case

References

- A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," IEEE Trans. Info. Theory, vol. 60, no. 3, pp. 1711–1732, 2014.
- T. Strohmer and S. Ling, "Self-calibration and biconvex compressive sensing," arXiv preprint arXiv:1501.06864 [cs.IT], 2015.
- Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," IEEE Signal Process. Mag., vol. 32, no. 3, pp. 87–109, 2015.
- E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," IEEE Trans. Info. Theory, vol. 61, no. 4, pp. 1985–2007, 2015.
- E. J. Candes Y. Chen, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," arXiv preprint arXiv:1505.05114 [cs.IT], 2015.

Financial support

- [†]Spanish MINECO grant TEC2013-41604-R; [‡]NSF CCF-1217963; and [*]EU's Horizon 2020 programme under grant agreement ERC-BNYQ, ISF under grant no. 335/14, and ICore: the Israeli Excellence Center "Circle of Light".