Generic analysis of local convergence

Cong Ma

University of Chicago, Autumn 2021

Outline

- Low-rank matrix sensing
- Phase retrieval
- Low-rank matrix completion

Low-rank matrix sensing

Low-rank matrix sensing

- Groundtruth: rank-r matrix $\boldsymbol{M}^\star \in \mathbb{R}^{n_1 \times n_2}$
- Observations:

$$y_i = \langle \boldsymbol{A}_i, \boldsymbol{M}^\star \rangle, \qquad \text{for } 1 \le i \le m$$

- Goal: recover $\boldsymbol{M}^\star$ based on linear measurements $\{\boldsymbol{A}_i, y_i\}_{1 \le i \le m}$

How many measurements are needed

- $m \geq n_1 n_2$ "generic" measurements suffice given theory of solving linear equations
- But $\boldsymbol{M}^{\star}$ only has $O((n_1 + n_2)r)$ degrees of freedom. Ideally, one hope for using only $O((n_1 + n_2)r)$ measurements

Recovery is possible if $\{A_i\}$'s satisfy restricted isometry property

Restricted isometry property (RIP)

Define linear operator $\mathcal{A}: \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$ t obe

$$\mathcal{A}(\boldsymbol{M}) = [m^{-1/2} \langle \boldsymbol{A}_i, \boldsymbol{M} \rangle]_{1 \leq i \leq m}$$

Definition 8.1

The operator $\mathcal{A}$ is said to satisfy r-RIP with RIP constant $\delta_r < 1$ if

$$(1 - \delta_r) \|\boldsymbol{M}\|_{\mathsf{F}}^2 \leq \|\mathcal{A}(\boldsymbol{M})\|_2^2 \leq (1 + \delta_r) \|\boldsymbol{M}\|_{\mathsf{F}}^2$$

holds simultaneously for all $\boldsymbol{M}$ of rank at most r.

- Many random designs satisfy RIP with high probability
- For instance, when $\boldsymbol{A}_i$ is composed of i.i.d. $\mathcal{N}(0,1)$ entries, $\mathcal{A}$ obeys r-RIP with constant δ_r as soon as $m \gtrsim (n_1 + n_2) r / \delta_r^2$

An optimization-based method

Consider the simple case when $\boldsymbol{M}^\star$ is psd and rank 1, i.e.,

$$\boldsymbol{M}^\star = \boldsymbol{x}^\star \boldsymbol{x}^{\star\top}$$

Then least-squares estimation yields

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \qquad f(\boldsymbol{x}) = \frac{1}{4m}\sum_{i=1}^{m}\Big(\langle \boldsymbol{A}_i, \boldsymbol{x}\boldsymbol{x}^\top\rangle - y_i\Big)^2$$

Gradient descent

Starting from $\boldsymbol{x}^0$, one proceeds by

$$\begin{aligned}\boldsymbol{x}^{t+1} &= \boldsymbol{x}^t - \eta \nabla f(\boldsymbol{x}^t) \\ &= \boldsymbol{x}^t - \frac{\eta}{m} \sum_{i=1}^{m} \Big(\langle \boldsymbol{A}_i, \boldsymbol{x}^t \boldsymbol{x}^{t\top} \rangle - y_i \Big) \boldsymbol{A}_i \boldsymbol{x}^t\end{aligned}$$

Here we made simplifying assumption that A_i is symmetric

- Under random design, when $m \to \infty$, this mirrors PCA problem with loss $\frac{1}{4}\|\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{x}^\star \boldsymbol{x}^{\star\top}\|_{\mathrm{F}}^2$; GD works locally
- How about finite-sample case?

— *RIP helps*

Local convergence of gradient descent

Theorem 8.2

Suppose that $\mathcal{A}$ *obeys 4-RIP with constant* $\delta_4 \leq 1/44$. *If* $\|\boldsymbol{x}^0 - \boldsymbol{x}^\star\|_2 \leq \|\boldsymbol{x}^\star\|_2/12$, *then GD with* $\eta = 1/(3\|\boldsymbol{x}^\star\|_2^2)$ *obeys*

$$\|\boldsymbol{x}^t - \boldsymbol{x}^\star\|_2 \leq \left(\tfrac{11}{12}\right)^t \|\boldsymbol{x}^0 - \boldsymbol{x}^\star\|_2, \qquad \text{for } t = 0, 1, 2, \ldots$$

- local linear convergence within basin of attraction $\{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \leq \|\boldsymbol{x}^\star\|_2/12\}$
- how do we initialize GD? spectral method

Proof of Theorem 8.2

In view of theory of gradient descent for locally strongly convex and smooth functions, it suffices to prove that

$$0.25\|\boldsymbol{x}^\star\|_2^2 \boldsymbol{I}_n \preceq \nabla^2 f(\boldsymbol{x}) \preceq 3\|\boldsymbol{x}^\star\|_2^2 \boldsymbol{I}_n$$

holds for all

$$\{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \le \|\boldsymbol{x}^\star\|_2/12\}$$

To analyze spectral properties of $\nabla^2 f(\boldsymbol{x})$, we focus on quadratic forms

$$\boldsymbol{z}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{z}$$

Proof of Theorem 8.2 (cont.)

Simple calculations show

$$\boldsymbol{z}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{z} = \frac{1}{m} \sum_{i=1}^{m} \langle \boldsymbol{A}_i, \boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{x}^\star \boldsymbol{x}^{\star\top} \rangle (\boldsymbol{z}^\top \boldsymbol{A}_i \boldsymbol{z}) + 2(\boldsymbol{z}^\top \boldsymbol{A}_i \boldsymbol{x})^2,$$

which admits a more "compact" expression

$$\begin{aligned} \boldsymbol{z}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{z} &= \langle \mathcal{A}(\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{x}^\star \boldsymbol{x}^{\star\top}), \mathcal{A}(\boldsymbol{z}\boldsymbol{z}^\top) \rangle \\ &\quad + \frac{1}{2} \langle \mathcal{A}(\boldsymbol{z}\boldsymbol{x}^\top + \boldsymbol{x}\boldsymbol{z}^\top), \mathcal{A}(\boldsymbol{z}\boldsymbol{x}^\top + \boldsymbol{x}\boldsymbol{z}^\top) \rangle \end{aligned}$$

— *requires analyzing* $\langle \mathcal{A}(\boldsymbol{X}), \mathcal{A}(\boldsymbol{Y}) \rangle$

RIP preserves inner product

A consequence of RIP

Lemma 8.3

Suppose that $\mathcal{A}$ satisfies 2r-RIP with constant $\delta_{2r} < 1$, then

$$|\langle \mathcal{A}(\boldsymbol{X}), \mathcal{A}(\boldsymbol{Y}) \rangle - \langle \boldsymbol{X}, \boldsymbol{Y} \rangle| \leq \delta_{2r} \|\boldsymbol{X}\|_{\mathrm{F}} \|\boldsymbol{Y}\|_{\mathrm{F}}$$

holds for any $\boldsymbol{X}, \boldsymbol{Y}$ of rank no more than r

Proof of Theorem 8.2 (cont.)

Apply Lemma 8.3 to obtain

$$\begin{aligned}&\left|\langle \mathcal{A}(\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{x}^\star\boldsymbol{x}^{\star\top}), \mathcal{A}(\boldsymbol{z}\boldsymbol{z}^\top)\rangle - \langle \boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{x}^\star\boldsymbol{x}^{\star\top}, \boldsymbol{z}\boldsymbol{z}^\top\rangle\right| \\ &\quad \le \delta_4 \|\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{x}^\star\boldsymbol{x}^{\star\top}\|_{\mathrm{F}} \|\boldsymbol{z}\boldsymbol{z}^\top\|_{\mathrm{F}} \le 3\delta_4 \|\boldsymbol{x}^\star\|_2^2 \|\boldsymbol{z}\|_2^2,\end{aligned}$$

while last relation uses $\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \le \|\boldsymbol{x}^\star\|_2$.

Similarly, one has

$$\begin{aligned}&\left|\langle \mathcal{A}(\boldsymbol{z}\boldsymbol{x}^\top + \boldsymbol{x}\boldsymbol{z}^\top), \mathcal{A}(\boldsymbol{z}\boldsymbol{x}^\top + \boldsymbol{x}\boldsymbol{z}^\top)\rangle - \|\boldsymbol{z}\boldsymbol{x}^\top + \boldsymbol{x}\boldsymbol{z}^\top\|_{\mathrm{F}}^2\right| \\ &\quad \le \delta_4 \|\boldsymbol{z}\boldsymbol{x}^\top + \boldsymbol{x}\boldsymbol{z}^\top\|_{\mathrm{F}}^2 \le 4\delta_4 \|\boldsymbol{x}\|_2^2 \|\boldsymbol{z}\|_2^2 \le 16\delta_4 \|\boldsymbol{x}^\star\|_2^2 \|\boldsymbol{z}\|_2^2\end{aligned}$$

Proof of Theorem 8.2 (cont.)

Define

$$g(\boldsymbol{x}, \boldsymbol{z}) := \langle \boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{x}^\star \boldsymbol{x}^{\star\top}, \boldsymbol{z}\boldsymbol{z}^\top \rangle + \frac{1}{2} \|\boldsymbol{z}\boldsymbol{x}^\top + \boldsymbol{x}\boldsymbol{z}^\top\|_{\mathrm{F}}^2$$

Key conclusion so far: when $\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \leq \|\boldsymbol{x}^\star\|_2$, $\boldsymbol{z}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{z}$ is close to $g(\boldsymbol{x}, \boldsymbol{z})$

It boils down to upper and lower bounding $g(\boldsymbol{x}, \boldsymbol{z})$—a much easier task

Proof of Lemma 8.3

Without loss of generality, assume that $\|\boldsymbol{X}\|_{\mathrm{F}} = \|\boldsymbol{Y}\|_{\mathrm{F}} = 1$
Since $\boldsymbol{X}+\boldsymbol{Y}$ and $\boldsymbol{X}-\boldsymbol{Y}$ have rank at most $2r$, we can leverage $2r$-RIP to obtain

$$
\begin{aligned}
(1-\delta_{2r})\|\boldsymbol{X}+\boldsymbol{Y}\|_{\mathsf{F}}^2 &\overset{(1)}{\leq} \|\mathcal{A}(\boldsymbol{X}+\boldsymbol{Y})\|_2^2 \overset{(2)}{\leq} (1+\delta_{2r})\|\boldsymbol{X}+\boldsymbol{Y}\|_{\mathsf{F}}^2 \\
(1-\delta_{2r})\|\boldsymbol{X}-\boldsymbol{Y}\|_{\mathsf{F}}^2 &\overset{(3)}{\leq} \|\mathcal{A}(\boldsymbol{X}-\boldsymbol{Y})\|_2^2 \overset{(4)}{\leq} (1+\delta_{2r})\|\boldsymbol{X}-\boldsymbol{Y}\|_{\mathsf{F}}^2
\end{aligned}
$$

Combine (2) and (3) to see

$$
\begin{aligned}
4\langle \mathcal{A}(\boldsymbol{X}), \mathcal{A}(\boldsymbol{Y})\rangle &= \|\mathcal{A}(\boldsymbol{X}+\boldsymbol{Y})\|_2^2 - \|\mathcal{A}(\boldsymbol{X}-\boldsymbol{Y})\|_2^2 \\
&\leq (1+\delta_{2r})\|\boldsymbol{X}+\boldsymbol{Y}\|_{\mathsf{F}}^2 - (1-\delta_{2r})\|\boldsymbol{X}-\boldsymbol{Y}\|_{\mathsf{F}}^2 \\
&= 4\delta_{2r} + 4\langle \boldsymbol{X}, \boldsymbol{Y}\rangle
\end{aligned}
$$

Combine (1) and (4) to finish the proof

Spectral initialization

Construct a surrogate matrix

$$\boldsymbol{M} = \frac{1}{m}\sum_{i=1}^{m} y_i \boldsymbol{A}_i$$

Define adjoint operator of $\mathcal{A}$: $\mathcal{A}^*(\cdot) : \mathbb{R}^m \to \mathbb{R}^{n_1 \times n_2}$

$$\mathcal{A}^*(\boldsymbol{v}) = \frac{1}{\sqrt{m}}\sum_{i=1}^{m} v_i \boldsymbol{A}_i$$

As a result, one has $\boldsymbol{M} = \mathcal{A}^*(\mathcal{A}(\boldsymbol{M}^\star))$

- Let $\lambda_1 \boldsymbol{u}_1 \boldsymbol{u}_1^\top$ be the top eigendecomposition of $\boldsymbol{M}$; return $\boldsymbol{x}^0 = \sqrt{\lambda_1}\boldsymbol{u}_1$

Performance guarantee of spectral initialization

Lemma 8.4

Suppose that $\mathcal{A}$ obeys 2-RIP with RIP constant $\delta_2 \leq 1/4$. Then one has

$$\|\boldsymbol{x}^0 - \boldsymbol{x}^\star\|_2 \lesssim \delta_2 \|\boldsymbol{x}^\star\|_2.$$

- as long as δ_4 is small enough, spectral initialization + GD works for low-rank matrix sensing since $\delta_2 \leq \delta_4$
- under Gaussian design, we only need $O((n_1 + n_2)r)$ linear measurements

Proof of Lemma 8.4

By definition, one has

$$\begin{aligned}\|\boldsymbol{M}-\boldsymbol{M}^\star\| &= \|\mathcal{A}^*(\mathcal{A}(\boldsymbol{M}^\star))-\boldsymbol{M}^\star\| \\ &= \sup_{\boldsymbol{v}:\|\boldsymbol{v}\|_2=1} \boldsymbol{v}^\top\left(\mathcal{A}^*(\mathcal{A}(\boldsymbol{M}^\star))-\boldsymbol{M}^\star\right)\boldsymbol{v} \\ &= \sup_{\boldsymbol{v}:\|\boldsymbol{v}\|_2=1} \langle \mathcal{A}^*(\mathcal{A}(\boldsymbol{M}^\star))-\boldsymbol{M}^\star, \boldsymbol{v}\boldsymbol{v}^\top\rangle \\ &\le \sup_{\boldsymbol{v}:\|\boldsymbol{v}\|_2=1} \delta_2\|\boldsymbol{M}^\star\|_{\mathsf{F}}\|\boldsymbol{v}\boldsymbol{v}^\top\|_{\mathsf{F}} \\ &\le \delta_2\|\boldsymbol{x}^\star\|_2^2\end{aligned}$$

Consequently, by Wely's inequality and Davis-Kahan's theorem, we have

$$\begin{aligned}\lambda_1-\lambda_1^\star &\le \|\boldsymbol{M}-\boldsymbol{M}^\star\| \le \delta_2\|\boldsymbol{x}^\star\|_2^2 \\ \|\boldsymbol{u}_1-\boldsymbol{u}_1^\star\|_2 &\lesssim \frac{\|\boldsymbol{M}-\boldsymbol{M}^\star\|}{\|\boldsymbol{M}^\star\|} \lesssim \delta_2\end{aligned}$$

Proof of Lemma 8.4

Note that

$$\begin{aligned}
\left\|\sqrt{\lambda_1}\boldsymbol{u}_1 - \sqrt{\lambda_1^\star}\boldsymbol{u}_1^\star\right\|_2 &\leq \left\|\left(\sqrt{\lambda_1} - \sqrt{\lambda_1^\star}\right)\boldsymbol{u}_1\right\|_2 + \left\|\sqrt{\lambda_1^\star}\left(\boldsymbol{u}_1 - \boldsymbol{u}_1^\star\right)\right\|_2 \\
&= \left(\sqrt{\lambda_1} - \sqrt{\lambda_1^\star}\right) + \|\boldsymbol{x}^\star\|_2 \cdot \|\boldsymbol{u}_1 - \boldsymbol{u}_1^\star\|_2 \\
&= \frac{\lambda_1 - \lambda_1^\star}{\sqrt{\lambda_1} + \sqrt{\lambda_1^\star}} + \|\boldsymbol{x}^\star\|_2 \cdot \|\boldsymbol{u}_1 - \boldsymbol{u}_1^\star\|_2 \\
&\lesssim \delta_2 \|\boldsymbol{x}^\star\|_2
\end{aligned}$$

Sampling operators that do NOT satisfy RIP

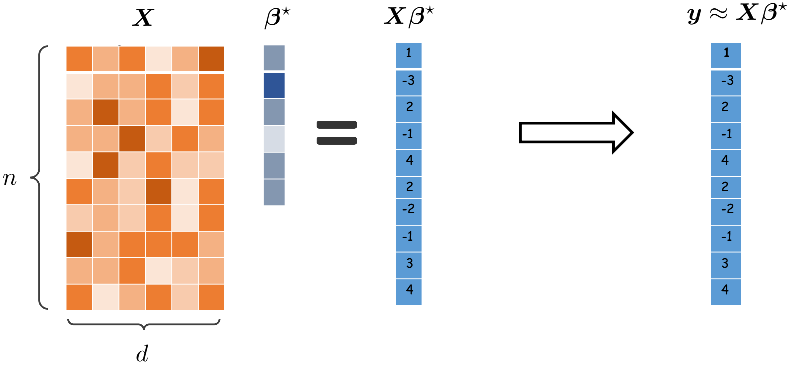
Unfortunately, many sampling operators fail to satisfy RIP

Two important examples:

- Phase retrieval / solving random quadratic systems of equations
- Matrix completion

Phase retrieval / solving random quadratic systems of equations

Solving linear systems (linear regression)

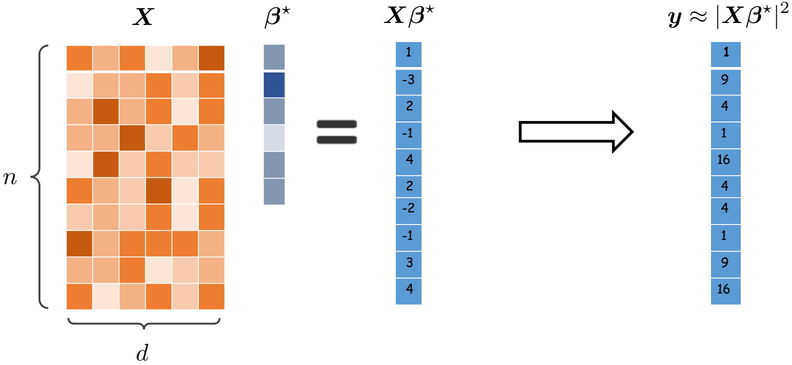


Estimate $\boldsymbol{\beta}^\star \in \mathbb{R}^d$ from n linear samples

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^\star + \varepsilon_i, \qquad i = 1, \ldots, n$$

— *assume w.l.o.g.* $\|\boldsymbol{\beta}^\star\|_2 = 1$

Solving quadratic systems of equations

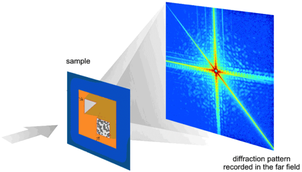


Estimate $\boldsymbol{\beta}^\star \in \mathbb{R}^d$ from n quadratic samples

$$y_i = (\boldsymbol{x}_i^\top \boldsymbol{\beta}^\star)^2, \qquad i = 1, \ldots, n$$

Motivation: phase retrieval

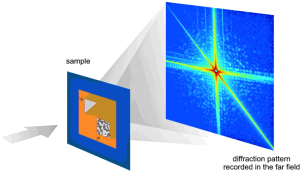
- electric field $\beta^\star(t_1, t_2) \longrightarrow$ Fourier transform $\mathcal{F}\beta^\star(f_1, f_2)$



- detectors record intensities $\left|\mathcal{F}\beta^\star(f_1, f_2)\right|^2$ of Fourier transform

Motivation: phase retrieval

- electric field $\beta^\star(t_1, t_2) \longrightarrow$ Fourier transform $\mathcal{F}\beta^\star(f_1, f_2)$

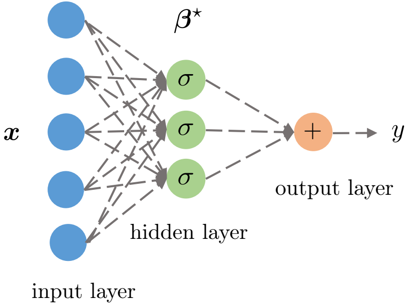


- detectors record intensities $\left|\mathcal{F}\beta^\star(f_1, f_2)\right|^2$ of Fourier transform

Phase retrieval: recover signal $\beta^\star(t_1, t_2)$ from $\left|\mathcal{F}\beta^\star(f_1, f_2)\right|^2$

Motivation: learning neural nets with quadratic activations

— *Soltanolkotabi, Javanmard, Lee '17, Li, Ma, Zhang '17*



input features: $\boldsymbol{x}$ weights: $\boldsymbol{\beta}^\star = [\boldsymbol{\beta}_1^\star, \cdots, \boldsymbol{\beta}_r^\star]$

output: $y = \sum_{k=1}^r \sigma(\boldsymbol{x}^\top \boldsymbol{\beta}_k^\star) + \varepsilon \overset{\sigma(z)=z^2}{\Longrightarrow} \sum_{k=1}^r (\boldsymbol{x}^\top \boldsymbol{\beta}_k^\star)^2 + \varepsilon$

Rank-one measurements in matrix space

Equivalent representation for measurements:

$$y_i = \boldsymbol{a}_i^\top \underbrace{\boldsymbol{x}^\star \boldsymbol{x}^{\star\top}}_{:=\boldsymbol{M}^\star} \boldsymbol{a}_i = \langle \underbrace{\boldsymbol{a}_i \boldsymbol{a}_i^\top}_{:=\boldsymbol{A}_i}, \boldsymbol{M}^\star \rangle, \qquad 1 \leq i \leq m$$

Using operator notation

$$\mathcal{A}(\boldsymbol{X}) = \begin{bmatrix} \langle \boldsymbol{A}_1, \boldsymbol{X} \rangle \\ \langle \boldsymbol{A}_2, \boldsymbol{X} \rangle \\ \vdots \\ \langle \boldsymbol{A}_m, \boldsymbol{X} \rangle \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{a}_1 \boldsymbol{a}_1^\top, \boldsymbol{X} \rangle \\ \langle \boldsymbol{a}_2 \boldsymbol{a}_2^\top, \boldsymbol{X} \rangle \\ \vdots \\ \langle \boldsymbol{a}_m \boldsymbol{a}_m^\top, \boldsymbol{X} \rangle \end{bmatrix}$$

Does $\mathcal{A}$ obey RIP?

Suppose $\boldsymbol{a}_i \overset{\text{ind.}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$

- If $\boldsymbol{x}$ is independent of $\{\boldsymbol{a}_i\}$, then

$$\langle \boldsymbol{a}_i \boldsymbol{a}_i^\top, \boldsymbol{x}\boldsymbol{x}^\top \rangle = \left|\boldsymbol{a}_i^\top \boldsymbol{x}\right|^2 \asymp \|\boldsymbol{x}\|_2^2 \;\Rightarrow\; \frac{1}{\sqrt{m}} \left\|\mathcal{A}(\boldsymbol{x}\boldsymbol{x}^\top)\right\|_{\mathrm{F}} \asymp \|\boldsymbol{x}\boldsymbol{x}^\top\|_{\mathrm{F}}$$

- Consider $\boldsymbol{A}_i = \boldsymbol{a}_i \boldsymbol{a}_i^\top$: with high prob.,

$$\langle \boldsymbol{a}_i \boldsymbol{a}_i^\top, \boldsymbol{A}_i \rangle = \|\boldsymbol{a}_i\|_2^4 \approx n \|\boldsymbol{a}_i \boldsymbol{a}_i^\top\|_{\mathrm{F}}$$

$$\Longrightarrow \quad \frac{1}{\sqrt{m}} \|\mathcal{A}(\boldsymbol{A}_i)\|_{\mathrm{F}} \geq \frac{1}{\sqrt{m}} |\langle \boldsymbol{a}_i \boldsymbol{a}_i^\top, \boldsymbol{A}_i \rangle| \approx \frac{n}{\sqrt{m}} \|\boldsymbol{A}_i\|_{\mathrm{F}}$$

— *fails to obey RIP when $m \approx n$*

Why do we lose RIP?

- Some low-rank matrices $\boldsymbol{X}$ (e.g. $\boldsymbol{a}_i \boldsymbol{a}_i^\top$) might be too aligned with some (rank-1) measurement matrices
  - loss of "incoherence" in some measurements

A natural least-squares formulation

$$\text{given:} \qquad y_k \;=\; (\boldsymbol{a}_k^\top \boldsymbol{x}^\star)^2, \quad 1 \le k \le m$$

$$\Downarrow$$

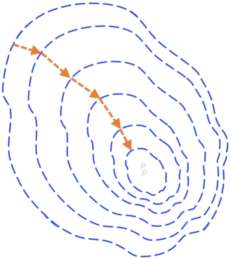
$$\text{minimize}_{\boldsymbol{x} \in \mathbb{R}^n} \quad f(\boldsymbol{x}) = \frac{1}{4m} \sum_{k=1}^{m} \Big[\big(\boldsymbol{a}_k^\top \boldsymbol{x} \big)^2 - y_k \Big]^2$$

Wirtinger flow (Candès, Li, Soltanolkotabi '14)

$$\text{minimize}_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = \frac{1}{4m}\sum_{k=1}^{m}\left[\left(\boldsymbol{a}_k^\top \boldsymbol{x}\right)^2 - y_k\right]^2$$

Wirtinger flow (Candès, Li, Soltanolkotabi '14)

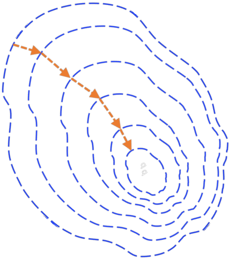
$$\text{minimize}_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = \frac{1}{4m} \sum_{k=1}^{m} \left[\left(\boldsymbol{a}_k^\top \boldsymbol{x} \right)^2 - y_k \right]^2$$



- **spectral initialization:** $\boldsymbol{x}^0 \leftarrow$ leading eigenvector of certain data matrix

Wirtinger flow (Candès, Li, Soltanolkotabi '14)

$$\text{minimize}_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = \frac{1}{4m}\sum_{k=1}^{m}\left[\left(\boldsymbol{a}_k^\top \boldsymbol{x}\right)^2 - y_k\right]^2$$



- **spectral initialization:** $\boldsymbol{x}^0 \leftarrow$ leading eigenvector of certain data matrix
- **gradient descent:**

$$\boldsymbol{x}^{t+1} = \boldsymbol{x}^t - \eta\,\nabla f(\boldsymbol{x}^t), \qquad t = 0, 1, \ldots$$

Spectral initialization

—cf. homework 1

$\lambda^0, \boldsymbol{u}^0 \longleftarrow$ leading eigenvalue, eigenvector of

$$\boldsymbol{M} := \frac{1}{m} \sum_{k=1}^{m} y_k \, \boldsymbol{a}_k \boldsymbol{a}_k^\top$$

Then set $\boldsymbol{x}^0 = \sqrt{\lambda_0} \, \boldsymbol{u}^0$

Rationale: under random Gaussian design $\boldsymbol{a}_i \overset{\text{ind.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$,

$$\mathbb{E}[\boldsymbol{M}] := \mathbb{E}\left[\frac{1}{m} \sum_{k=1}^{m} \boldsymbol{y}_k \boldsymbol{a}_k \boldsymbol{a}_k^\top\right] = \underbrace{\|\boldsymbol{x}^\star\|_2^2 \, \boldsymbol{I} + 2\boldsymbol{x}^\star \boldsymbol{x}^{\star\top}}_{\text{leading eigenvector: } \pm \boldsymbol{x}^\star}$$

First theory of WF

$$\text{dist}(\boldsymbol{x}^t, \boldsymbol{x}^\star) := \min\{\|\boldsymbol{x}^t \pm \boldsymbol{x}^\star\|_2\}$$

Theorem 8.5 (Candès, Li, Soltanolkotabi '14)

Under i.i.d. Gaussian design, WF with spectral initialization achieves

$$\text{dist}(\boldsymbol{x}^t, \boldsymbol{x}^\star) \lesssim \left(1 - \frac{\eta}{4}\right)^{t/2} \|\boldsymbol{x}^\star\|_2,$$

with high prob., provided that step size $\eta \lesssim 1/n$ *and sample size:* $m \gtrsim n \log n$.

- Iteration complexity: $O(n \log \frac{1}{\epsilon})$
- Sample complexity: $O(n \log n)$
- Derived based on (worst-case) local geometry

First theory of WF

$$\text{dist}(\boldsymbol{x}^t, \boldsymbol{x}^\star) := \min\{\|\boldsymbol{x}^t \pm \boldsymbol{x}^\star\|_2\}$$

Theorem 8.5 (Candès, Li, Soltanolkotabi '14)

Under i.i.d. Gaussian design, WF with spectral initialization achieves

$$\text{dist}(\boldsymbol{x}^t, \boldsymbol{x}^\star) \lesssim \left(1 - \frac{\eta}{4}\right)^{t/2} \|\boldsymbol{x}^\star\|_2,$$

with high prob., provided that step size $\eta \lesssim 1/n$ *and sample size:* $m \gtrsim n \log n$.

- Iteration complexity: $O(n \log \frac{1}{\epsilon})$
- Sample complexity: $O(n \log n)$
- Derived based on (worst-case) local geometry

First theory of WF

$$\text{dist}(\boldsymbol{x}^t, \boldsymbol{x}^\star) := \min\{\|\boldsymbol{x}^t \pm \boldsymbol{x}^\star\|_2\}$$

Theorem 8.5 (Candès, Li, Soltanolkotabi '14)

Under i.i.d. Gaussian design, WF with spectral initialization achieves

$$\text{dist}(\boldsymbol{x}^t, \boldsymbol{x}^\star) \lesssim \left(1 - \frac{\eta}{4}\right)^{t/2} \|\boldsymbol{x}^\star\|_2,$$

with high prob., provided that step size $\eta \lesssim 1/n$ *and sample size:* $m \gtrsim n \log n$.

- Iteration complexity: $O(n \log \frac{1}{\epsilon})$
- Sample complexity: $O(n \log n)$
- Derived based on (worst-case) local geometry

Spectral initialization for phase retrieval

Key: control

$$\left\| \frac{1}{m} \sum_{k=1}^{m} y_k \, \boldsymbol{a}_k \boldsymbol{a}_k^\top - \left(\|\boldsymbol{x}^\star\|_2^2 \boldsymbol{I}_n + 2\boldsymbol{x}^\star \boldsymbol{x}^{\star\top} \right) \right\|$$

Lemma 8.6

Fix any small constant $\delta > 0$. As long as $m \geq c_\delta n \log n$ for some sufficiently large constant c_δ (which potentially depends on δ), the following holds with high probability

$$\left\| \frac{1}{m} \sum_{k=1}^{m} y_k \, \boldsymbol{a}_k \boldsymbol{a}_k^\top - \left(\|\boldsymbol{x}^\star\|_2^2 \boldsymbol{I}_n + 2\boldsymbol{x}^\star \boldsymbol{x}^{\star\top} \right) \right\| \leq \delta \|\boldsymbol{x}^\star\|_2^2$$

- Proof: truncation-based matrix Bernstein or ε-net argument

Spectral initialization

Since

$$\left\| \frac{1}{m} \sum_{k=1}^{m} y_k \, \boldsymbol{a}_k \boldsymbol{a}_k^\top - \left(\|\boldsymbol{x}^\star\|_2^2 \boldsymbol{I}_n + 2\boldsymbol{x}^\star \boldsymbol{x}^{\star\top} \right) \right\|$$

is small, by Weyl's inequality and Davis-Kahan's theorem, we know

- $\lambda^0 - \lambda^\star$ is small
- $\|\boldsymbol{u}^0 - \boldsymbol{u}^\star\|_2$ is small

Consequently, $\boldsymbol{x}^0 = \sqrt{\lambda^0}\boldsymbol{u}^0$ is close to $\boldsymbol{x}^\star = \sqrt{\lambda^\star}\boldsymbol{u}^\star$ in the sense that

$$\|\boldsymbol{x}^0 - \boldsymbol{x}^\star\|_2 \ll \|\boldsymbol{x}^\star\|_2$$

Local geometry for phase retrieval

Now we move on to local convergence of GD, which boils down to characterizing local geometry of $f(\cdot)$

Lemma 8.7

Assume that $m \geq c_0 n \log n$*. Then with high probability,*

$$0.5\boldsymbol{I}_n \preceq \nabla^2 f(\boldsymbol{x}) \preceq c_2 n \boldsymbol{I}_n$$

holds simultaneously for all $\boldsymbol{x}$ *obeying* $\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \leq c_1 \|\boldsymbol{x}^\star\|_2$*. Here* $c_0, c_1, c_2 > 0$ *are some universal constants.*

Proof of Lemma 8.7

First, write Hessian as

$$\nabla^2 f(\boldsymbol{x}) = \frac{1}{m}\sum_{i=1}^{m}(3(\boldsymbol{a}_i^\top \boldsymbol{x})^2 - y_i)\boldsymbol{a}_i\boldsymbol{a}_i^\top$$

When $\boldsymbol{x} = \boldsymbol{x}^\star$, one has

$$\begin{aligned}\nabla^2 f(\boldsymbol{x}^\star) &= \frac{1}{m}\sum_{i=1}^{m} 2(\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2\boldsymbol{a}_i\boldsymbol{a}_i^\top \\ &\approx 2(\|\boldsymbol{x}^\star\|_2^2\boldsymbol{I}_n + 2\boldsymbol{x}^\star\boldsymbol{x}^{\star\top})\end{aligned}$$

Therefore at minimizer $\boldsymbol{x}^\star$, $f(\cdot)$ is strongly convex and smooth; how about nearby points $\boldsymbol{x}$

Local strong convexity

Recall Hessian

$$\begin{aligned}\nabla^2 f(\boldsymbol{x}) &= \frac{1}{m}\sum_{i=1}^{m}\left(3(\boldsymbol{a}_i^\top \boldsymbol{x})^2 - (\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2\right)\boldsymbol{a}_i\boldsymbol{a}_i^\top \\ &= \frac{1}{m}\sum_{i=1}^{m} 3(\boldsymbol{a}_i^\top \boldsymbol{x})^2 \boldsymbol{a}_i\boldsymbol{a}_i^\top - \left(\|\boldsymbol{x}^\star\|_2^2 \boldsymbol{I}_n + 2\boldsymbol{x}^\star\boldsymbol{x}^{\star\top}\right) \\ &\quad + \|\boldsymbol{x}^\star\|_2^2 \boldsymbol{I}_n + 2\boldsymbol{x}^\star\boldsymbol{x}^{\star\top} - \frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2 \boldsymbol{a}_i\boldsymbol{a}_i^\top\end{aligned}$$

- Lemma 8.6 guarantees that if $m \geq c_0 n \log n$, then whp.,

$$\left\| \|\boldsymbol{x}^\star\|_2^2 \boldsymbol{I}_n + 2\boldsymbol{x}^\star\boldsymbol{x}^{\star\top} - \frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2 \boldsymbol{a}_i\boldsymbol{a}_i^\top \right\| \leq 0.001\|\boldsymbol{x}^\star\|_2^2$$

Local strong convexity (cont.)

Now we turn to a uniform lower bound over $\boldsymbol{x}$

$$\frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{a}_i^\top \boldsymbol{x})^2 \boldsymbol{a}_i \boldsymbol{a}_i^\top$$

Observe that for any constant $C > 0$

$$\frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{a}_i^\top \boldsymbol{x})^2 \boldsymbol{a}_i \boldsymbol{a}_i^\top \succeq \frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{a}_i^\top \boldsymbol{x})^2 \mathbb{1}\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \leq C\} \boldsymbol{a}_i \boldsymbol{a}_i^\top$$

- Intuition: truncation helps concentration due to better tail behavior

Local strong convexity (cont.)

Using covering argument, it is seen that with high probability

$$\left\| \frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{a}_i^\top \boldsymbol{x})^2 \mathbb{1}\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \le C\} \boldsymbol{a}_i \boldsymbol{a}_i^\top - 3(\beta_1 \boldsymbol{x}\boldsymbol{x}^\top + \beta_2 \|\boldsymbol{x}\|_2^2 \boldsymbol{I}_n) \right\| \ll \|\boldsymbol{x}\|_2^2,$$

for all $\boldsymbol{x}$, where

$$\begin{aligned} \beta_1 &:= \mathbb{E}[\xi^4 \mathbb{1}\{|\xi| \le C\}] - \mathbb{E}[\xi^2 \mathbb{1}\{|\xi| \le C\}], \\ \beta_2 &:= \mathbb{E}[\xi^2 \mathbb{1}\{|\xi| \le C\}] \end{aligned}$$

Observe that $\beta_1 \overset{C\to\infty}{\to} 2$, and $\beta_2 \overset{C\to\infty}{\to} 1$

Local smoothness

Decompose Hessian as

$$\nabla^2 f(\boldsymbol{x}) = \frac{1}{m}\sum_{i=1}^{m}\left(3(\boldsymbol{a}_i^\top \boldsymbol{x})^2 - (\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2\right)\boldsymbol{a}_i\boldsymbol{a}_i^\top$$

$$= \frac{3}{m}\sum_{i=1}^{m}[(\boldsymbol{a}_i^\top \boldsymbol{x})^2 - (\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2]\boldsymbol{a}_i\boldsymbol{a}_i^\top := \boldsymbol{\Lambda}_1$$

$$+ \frac{2}{m}\sum_{i=1}^{m}(\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2\boldsymbol{a}_i\boldsymbol{a}_i^\top - 2(\|\boldsymbol{x}^\star\|_2^2\boldsymbol{I}_n + 2\boldsymbol{x}^\star\boldsymbol{x}^{\star\top}) := \boldsymbol{\Lambda}_2$$

$$+ 2(\|\boldsymbol{x}^\star\|_2^2\boldsymbol{I}_n + 2\boldsymbol{x}^\star\boldsymbol{x}^{\star\top}) := \boldsymbol{\Lambda}_3$$

Our goal is to upper bound $\|\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2 + \boldsymbol{\Lambda}_3\|$

Local smoothness (cont.)

- Term $\|\mathbf{\Lambda}_3\|$ is easy to control
- By Lemma 8.6, term $\|\mathbf{\Lambda}_2\|$ is also small
- We are left with first term, which can be controlled as

$$\begin{aligned}\|\Lambda_1\| &\leq 3\left\|\frac{1}{m}\sum_{i=1}^{m}[(\boldsymbol{a}_i^\top \boldsymbol{x})^2 - (\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2]\boldsymbol{a}_i\boldsymbol{a}_i^\top\right\| \\ &\leq 3\left\|\frac{1}{m}\sum_{i=1}^{m}\left|(\boldsymbol{a}_i^\top \boldsymbol{x})^2 - (\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2\right|\boldsymbol{a}_i\boldsymbol{a}_i^\top\right\| \\ &= 3\left\|\frac{1}{m}\sum_{i=1}^{m}\left|\boldsymbol{a}_i^\top(\boldsymbol{x} - \boldsymbol{x}^\star)\right|\left|\boldsymbol{a}_i^\top(\boldsymbol{x} + \boldsymbol{x}^\star)\right|\boldsymbol{a}_i\boldsymbol{a}_i^\top\right\|\end{aligned}$$

Control Λ_1

By Cauchy–Schwarz, we have

$$\left| \boldsymbol{a}_i^\top (\boldsymbol{x} - \boldsymbol{x}^\star) \right| \leq \|\boldsymbol{a}_i\|_2 \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \lesssim \sqrt{n} \|\boldsymbol{x}^\star\|_2,$$

where we have used the fact that $\|\boldsymbol{a}_i\|_2 \lesssim \sqrt{n}$ with high probability, and the assumption that $\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \lesssim \|\boldsymbol{x}^\star\|_2$

As a result, we obtain

$$\|\Lambda_1\| \lesssim n \|\boldsymbol{x}^\star\|_2^2 \left\| \frac{1}{m} \sum_{i=1}^m \boldsymbol{a}_i \boldsymbol{a}_i^\top \right\| \asymp n$$

A closer look at smoothness

- We obtain $O(n)$ smoothness parameter for coherent points $\boldsymbol{x}$ such that $|\boldsymbol{a}_i^\top \boldsymbol{x}| \asymp \sqrt{n}$
- Our prediction of local smoothness is tight; take

$$\boldsymbol{x} = \boldsymbol{x}^\star + \delta \frac{\boldsymbol{a}_i}{\|\boldsymbol{a}_i\|_2}$$

consider $\boldsymbol{x}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{x}$

Low-rank matrix completion

Low-rank matrix completion

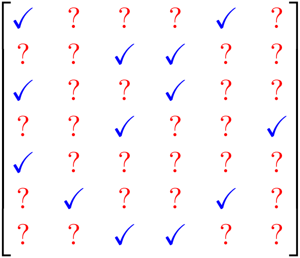
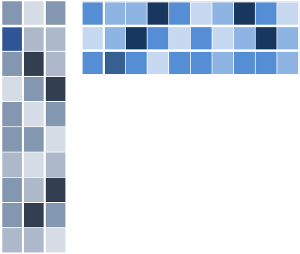


figure credit: Candès

- consider a low-rank matrix $\boldsymbol{M}^{\star} = \boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{U}^{\star\top}$
- each entry $M_{i,j}^{\star}$ is observed independently with prob. p
- **Goal:** estimate $\boldsymbol{M}^{\star}$

A natural least-squares loss

Represent low-rank matrix by $\boldsymbol{X}\boldsymbol{X}^\top$ with $\underbrace{\boldsymbol{X} \in \mathbb{R}^{n \times r}}_{\text{low-rank factor}}$



$$\underset{\boldsymbol{X} \in \mathbb{R}^{n \times r}}{\text{minimize}} \quad f(\boldsymbol{X}) = \sum_{(i,j) \in \Omega} \left[\left(\boldsymbol{X}\boldsymbol{X}^\top \right)_{i,j} - M^\star_{i,j} \right]^2$$

—how does local geometry look like?

Local geometry of $f(\cdot)$

Lemma 8.8

Suppose that $n^2 p \geq C\kappa^2 \mu r n \log n$ *for some sufficiently large constant* $C > 0$. *Then with high probability, the Hessian* $\nabla^2 f(\boldsymbol{X})$ *obeys*

$$\mathrm{vec}\left(\boldsymbol{V}\right)^\top \nabla^2 f\left(\boldsymbol{X}\right) \mathrm{vec}\left(\boldsymbol{V}\right) \geq \frac{\sigma_{\min}}{2} \left\|\boldsymbol{V}\right\|_{\mathrm{F}}^2$$

$$\left\|\nabla^2 f\left(\boldsymbol{X}\right)\right\| \leq \frac{5}{2}\sigma_{\max}$$

for all $\boldsymbol{X}$, $\boldsymbol{V} = \boldsymbol{Y}\boldsymbol{H}_Y - \boldsymbol{X}^\star$ *s.t.*
$\boldsymbol{H}_Y := \arg\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{Y}\boldsymbol{R} - \boldsymbol{X}^\star\right\|_{\mathrm{F}}$,

$$\left\|\boldsymbol{X} - \boldsymbol{X}^\star\right\|_{2,\infty} \leq \epsilon \left\|\boldsymbol{X}^\star\right\|_{2,\infty},$$

where $\epsilon \ll 1/\sqrt{\kappa^3 \mu r \log^2 n}$.

Restricted local strong convexity

- Due to rotation ambiguity, $f(\cdot)$ cannot be strongly convex along every direction; it is strongly convex along specific directions $\boldsymbol{V} = \boldsymbol{Y}\boldsymbol{H}_Y - \boldsymbol{X}^\star$
- Instead of ℓ_{F} ball, $f(X)$ is strongly convex in a local $\ell_{2,\infty}$ ball; $\boldsymbol{X}$ needs to be incoherent in the sense that

$$\|\boldsymbol{X}\|_{2,\infty} \lesssim \sqrt{\frac{\mu r}{n}}\|\boldsymbol{X}^\star\|$$

Revisit Incoherence

Definition 8.9

Fix an orthonormal matrix $\boldsymbol{U}^\star \in \mathbb{R}^{n\times r}$. Define its incoherence to be

$$\mu(\boldsymbol{U}^\star) := \frac{n\|\boldsymbol{U}^\star\|_{2,\infty}^2}{r}$$

—*recover incoherence of eigenvector when* $r = 1$

- For $\boldsymbol{M}^\star = \boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{U}^{\star\top}$, define $\mu(\boldsymbol{M}^\star) := \mu(\boldsymbol{U}^\star)$

Projected gradient descent for matrix completion

(1) **Projected spectral initialization**: let $\boldsymbol{U}^0\boldsymbol{\Sigma}^0\boldsymbol{U}^{0\top}$ be rank-r eigendecomposition of

$$\frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{Y}).$$

and set $\boldsymbol{Z}^0 = \boldsymbol{U}^0\left(\boldsymbol{\Sigma}^0\right)^{1/2}$, and incoherence set

$$\mathcal{C} := \{\boldsymbol{X} \mid \|\boldsymbol{X}\|_{2,\infty} \leq \sqrt{\frac{2\mu r}{n}}\|\boldsymbol{Z}^0\|\}$$

let $\boldsymbol{X}^0 = \mathcal{P}_{\mathcal{C}}(\boldsymbol{Z}^0)$

(2) **Projected gradient descent updates:**

$$\boldsymbol{X}^{t+1} = \mathcal{P}_{\mathcal{C}}(\boldsymbol{X}^t - \eta_t \nabla f(\boldsymbol{X}^t)), \qquad t = 0, 1, \cdots$$

Projection operator

Projection onto can be implemented via a row-wise "clipping operation"

$$[\mathcal{P}_{\mathcal{C}}(\boldsymbol{X})]_{i,\cdot} = \min\left\{1, \sqrt{\frac{2\mu r}{n}}\frac{\|\boldsymbol{Z}^0\|}{\|\boldsymbol{X}_{i,\cdot}\|_2}\right\} \cdot \boldsymbol{X}_{i,\cdot}$$

Performance guarantees

Theorem 8.10

Suppose that $n^2 p \geq c_0 \mu^2 r^2 \kappa^2 n \log n$ *for some large constant* $c_0 > 0$. *With high probability, one has for all* $t \geq 0$

$$\|\boldsymbol{X}^t \boldsymbol{Q}^t\|_{\mathrm{F}}^2 \leq \left(1 - \frac{c_1}{\mu^2 r^2 \kappa^2}\right)^t \sigma_r(\boldsymbol{M}^\star),$$

provided that step size is chosen as $\eta \asymp \frac{1}{\mu^2 r^2 \kappa \sigma_1(\boldsymbol{M}^\star)}$

Here $\boldsymbol{Q}^t$ is the optimal alignment matrix between $\boldsymbol{X}^t$ and $\boldsymbol{X}^\star$

$$\boldsymbol{Q}^t := \mathsf{argmin}_{\boldsymbol{R} \in \mathcal{O}^{r \times r}} \left\| \boldsymbol{X}^t \boldsymbol{R} - \boldsymbol{X}^\star \right\|_{\mathrm{F}}$$

Regularity condition

Key to prove convergence is the following regularity condition

Lemma 8.11

Suppose that $n^2 p \geq \mu^2 r^2 \kappa^2 n \log n$*. Then with high probability, for all* $\boldsymbol{X} \in \mathcal{C}$*, and* $\|\boldsymbol{X} - \boldsymbol{X}^\star \boldsymbol{H}\|_{\mathrm{F}}^2 \leq \frac{1}{16}\sigma_r(\boldsymbol{M}^\star)$ f *obeys*

$$\begin{aligned}\langle \nabla f(\boldsymbol{X}), \boldsymbol{X} - \boldsymbol{X}^\star \boldsymbol{H} \rangle \geq \frac{99}{512}\sigma_r(\boldsymbol{M}^\star)\|\boldsymbol{X} - \boldsymbol{X}^\star \boldsymbol{H}\|_{\mathrm{F}}^2 \\ + \frac{1}{13196\mu^2 r^2 \kappa \sigma_1(\boldsymbol{M}^\star)}\|\nabla f(\boldsymbol{X})\|_{\mathrm{F}}^2\end{aligned}$$

Here $\boldsymbol{H}$ is optimal alignment matrix

Complete the proof