Mini-Course 4: Matrix Completion has No Spurious Local Minimum

Yang Yuan

Computer Science Department
Cornell University

Today:

Today:

- “Guaranteed matrix completion via nonconvex factorization”, Ruoyu Sun and Zhi-Quan Luo, FOCS’15
  - For every global minimum, its neighborhood is strongly convex.

Today:

- "Guaranteed matrix completion via nonconvex factorization", Ruoyu Sun and Zhi-Quan Luo, FOCS'15
  - For every global minimum, its neighborhood is strongly convex.
- "Matrix Completion has No Spurious Local Minimum", Rong Ge, Jason Lee, Tengyu Ma, NIPS'16 (**Best Student Paper**)
  - Every local minimum is close to one global minimum.

Today:

- “Guaranteed matrix completion via nonconvex factorization”, Ruoyu Sun and Zhi-Quan Luo, FOCS’15
  - For every global minimum, its neighborhood is strongly convex.
- “Matrix Completion has No Spurious Local Minimum”, Rong Ge, Jason Lee, Tengyu Ma, NIPS’16 (**Best Student Paper**)
  - Every local minimum is close to one global minimum.
- Therefore, all local minima are global minima

What is matrix completion?

- $M = ZZ^\top, M \in \mathbb{R}^{d\times d}, Z \in \mathbb{R}^{d\times r}$

What is matrix completion?

- $M = ZZ^\top, M \in \mathbb{R}^{d \times d}, Z \in \mathbb{R}^{d \times r}$
- M is partially observable. $\Omega = \{(i,j) : M_{i,j} \text{ is observed}\}$.
  - Each entry is observed with probability p.

What is matrix completion?

- $M = ZZ^\top, M \in \mathbb{R}^{d\times d}, Z \in \mathbb{R}^{d\times r}$
- M is partially observable. $\Omega = \{(i,j) : M_{i,j} \text{ is observed}\}$.
  - Each entry is observed with probability p.
- Loss:

$$f(X) = \frac{1}{2}\sum_{(i,j)\in\Omega}\left[M_{i,j} - (XX^\top)_{i,j}\right]^2$$

What is matrix completion?

- $M = ZZ^\top, M \in \mathbb{R}^{d\times d}, Z \in \mathbb{R}^{d\times r}$
- M is partially observable. $\Omega = \{(i,j) : M_{i,j} \text{ is observed}\}$.
  - Each entry is observed with probability p.
- Loss:

$$f(X) = \frac{1}{2}\sum_{(i,j)\in\Omega}\left[M_{i,j} - (XX^\top)_{i,j}\right]^2$$

What is matrix completion?

- $M = ZZ^\top, M \in \mathbb{R}^{d\times d}, Z \in \mathbb{R}^{d\times r}$
- M is partially observable. $\Omega = \{(i,j) : M_{i,j} \text{ is observed}\}$.
  - Each entry is observed with probability p.
- Loss:

$$f(X) = \frac{1}{2} \sum_{(i,j)\in\Omega} \left[M_{i,j} - (XX^\top)_{i,j}\right]^2$$

- Z is not unique: $M = (ZA)(ZA)^\top$ for any $AA^\top = \mathbf{I}$.
  - Suffices to find one of them.

What is matrix completion?

- $M = ZZ^\top, M \in \mathbb{R}^{d\times d}, Z \in \mathbb{R}^{d\times r}$
- M is partially observable. $\Omega = \{(i,j) : M_{i,j} \text{ is observed}\}$.
  - Each entry is observed with probability p.
- Loss:

$$f(X) = \frac{1}{2}\sum_{(i,j)\in\Omega}\left[M_{i,j} - (XX^\top)_{i,j}\right]^2$$

- Z is not unique: $M = (ZA)(ZA)^\top$ for any $AA^\top = \mathbf{I}$.
  - Suffices to find one of them.
- Works very well in practice
  - Collaborative filtering
  - Recommendation systems
  - Dimension reduction
  - Multi-class learning.

What is matrix completion?

- $M = ZZ^\top, M \in \mathbb{R}^{d \times d}, Z \in \mathbb{R}^{d \times r}$
- M is partially observable. $\Omega = \{(i,j) : M_{i,j} \text{ is observed}\}$.
  - Each entry is observed with probability p.
- Loss:

$$f(X) = \frac{1}{2} \sum_{(i,j) \in \Omega} \left[M_{i,j} - (XX^\top)_{i,j} \right]^2$$

- Z is not unique: $M = (ZA)(ZA)^\top$ for any $AA^\top = \mathbf{I}$.
  - Suffices to find one of them.
- Works very well in practice
  - Collaborative filtering
  - Recommendation systems
  - Dimension reduction
  - Multi-class learning.
- Much faster than its convex relaxation

What is matrix completion?

- $M = ZZ^\top, M \in \mathbb{R}^{d \times d}, Z \in \mathbb{R}^{d \times r}$
- M is partially observable. $\Omega = \{(i,j) : M_{i,j} \text{ is observed}\}$.
  - Each entry is observed with probability p.
- Loss:

$$f(X) = \frac{1}{2} \sum_{(i,j) \in \Omega} \left[M_{i,j} - (XX^\top)_{i,j} \right]^2 + \lambda R(X)$$

- Z is not unique: $M = (ZA)(ZA)^\top$ for any $AA^\top = \mathbf{I}$.
  - Suffices to find one of them.
- Works very well in practice
  - Collaborative filtering
  - Recommendation systems
  - Dimension reduction
  - Multi-class learning.
- Much faster than its convex relaxation

What is matrix completion?

- $M = ZZ^\top, M \in \mathbb{R}^{d \times d}, Z \in \mathbb{R}^{d \times r}$
- M is partially observable. $\Omega = \{(i,j) : M_{i,j} \text{ is observed}\}$.
  - Each entry is observed with probability p.
- Loss:

$$f(X) = \frac{1}{2}\|P_\Omega(M - XX^\top)\|_F^2 + \lambda R(X)$$

- Z is not unique: $M = (ZA)(ZA)^\top$ for any $AA^\top = \mathbf{I}$.
  - Suffices to find one of them.
- Works very well in practice
  - Collaborative filtering
  - Recommender systems
  - Dimension reduction
  - Multi-class learning.
- Much faster than its convex relaxation

Assumption

- If $M = \mathbf{I}_r$, i.e., only has r non-zero entries.
    - May fail to observe any non-zero entries!

Assumption

- If $M = \mathbf{I}_r$, i.e., only has r non-zero entries.
  - May fail to observe any non-zero entries!

Assumption

For any row Z_i, we have

$$\|Z_i\| \leq \mu/\sqrt{d} \cdot \|Z\|_F$$

Moreover, $\sigma_{\max}(Z)/\sigma_{\min}(Z) = \kappa$. μ, κ are constants.

Assumption

- If $M = \mathbf{I}_r$, i.e., only has r non-zero entries.
  - May fail to observe any non-zero entries!

Assumption

For any row Z_i, we have

$$\|Z_i\| \leq \mu/\sqrt{d} \cdot \|Z\|_F$$

Moreover, $\sigma_{\max}(Z)/\sigma_{\min}(Z) = \kappa$. μ, κ are constants.

- $\|Z_i\|^2 \leq \mu^2/d \cdot \|Z\|_F^2$

Assumption

- If $M = \mathbf{I}_r$, i.e., only has r non-zero entries.
  - May fail to observe any non-zero entries!

Assumption

For any row Z_i, we have

$$\|Z_i\| \leq \mu/\sqrt{d} \cdot \|Z\|_F$$

Moreover, $\sigma_{\max}(Z)/\sigma_{\min}(Z) = \kappa$. μ, κ are constants.

- $\|Z_i\|^2 \leq \mu^2/d \cdot \|Z\|_F^2$
- $\|Z\|_F, \|Z_i\|$ are rotation invariant.

Assumption

- If $M = \mathbf{I}_r$, i.e., only has r non-zero entries.
  - May fail to observe any non-zero entries!

Assumption

For any row Z_i, we have

$$\|Z_i\| \leq \mu/\sqrt{d} \cdot \|Z\|_F$$

Moreover, $\sigma_{\max}(Z)/\sigma_{\min}(Z) = \kappa$. μ, κ are constants.

- $\|Z_i\|^2 \leq \mu^2/d \cdot \|Z\|_F^2$
- $\|Z\|_F, \|Z_i\|$ are rotation invariant.
  - $\|Z_i\|^2 = (ZZ^\top)_{i,i} = M_{i,i}$

Proof strategy in [Ge et al. 2016]

Proof strategy in [Ge et al. 2016]

- Use first order and second order conditions.

$$\nabla f(x) = 0, \qquad \nabla^2 f(x) \succeq 0$$

Proof strategy in [Ge et al. 2016]

- Use first order and second order conditions.

$$\nabla f(x) = 0, \qquad \nabla^2 f(x) \succeq 0$$

- **Linear properties** are preserved by concentration inequality:

$$\sum_{(i,j)\in\Omega} M_{i,j} \approx p \times \sum_{(i,j)\in d\times d} M_{i,j}$$

Concentration inequalities for linear properties

Theorem (Random sampling$\rightarrow$ good spectral approximation)

If $\|M\|_\infty \leq \frac{\|M\|_F}{d}$, *whp,*

$$\|P_\Omega(M) - pM\| \leq \epsilon p \|M\|_F$$

where $\epsilon = O(\sqrt{\log d/(pd)})$.

Concentration inequalities for linear properties

Theorem (Random sampling→ good spectral approximation)

If $\|M\|_\infty \leq \frac{\|M\|_F}{d}$, *whp,*

$$\|P_\Omega(M) - pM\| \leq \epsilon p \|M\|_F$$

where $\epsilon = O(\sqrt{\log d/(pd)})$.

Proof.

- $P_\Omega(M) = \sum_{i,j} s_{i,j} M_{i,j} \delta_{i,j}$, Where $\delta_{i,j} \in \mathbb{R}^{d\times d}$ is indicator matrix, $s_{i,j} \in \{0, 1\}$ are Bernoulli variable with probability p.

Concentration inequalities for linear properties

Theorem (Random sampling→ good spectral approximation)

If $\|M\|_\infty \leq \frac{\|M\|_F}{d}$, whp,

$$\|P_\Omega(M) - pM\| \leq \epsilon p \|M\|_F$$

where $\epsilon = O(\sqrt{\log d/(pd)})$.

Proof.

- $P_\Omega(M) = \sum_{i,j} s_{i,j} M_{i,j} \delta_{i,j}$, Where $\delta_{i,j} \in \mathbb{R}^{d \times d}$ is indicator matrix, $s_{i,j} \in \{0, 1\}$ are Bernoulli variable with probability p.
- Expectation: $\mathbb{E}[P_\Omega(M)] = pM$

Concentration inequalities for linear properties

Theorem (Random sampling→ good spectral approximation)

If $\|M\|_\infty \leq \frac{\|M\|_F}{d}$, *whp,*

$$\|P_\Omega(M) - pM\| \leq \epsilon p \|M\|_F$$

where $\epsilon = O(\sqrt{\log d/(pd)})$.

Proof.

- $P_\Omega(M) = \sum_{i,j} s_{i,j} M_{i,j} \delta_{i,j}$, Where $\delta_{i,j} \in \mathbb{R}^{d\times d}$ is indicator matrix, $s_{i,j} \in \{0,1\}$ are Bernoulli variable with probability p.
- Expectation: $\mathbb{E}[P_\Omega(M)] = pM$
- Variance: $\|\sum_{i,j} \mathbb{E}[s_{i,j}^2 M_{i,j}^2 \delta_{i,j}^\top \delta_{i,j}]\| \leq \frac{p}{d}$

Concentration inequalities for linear properties

Theorem (Random sampling→ good spectral approximation)

If $\|M\|_\infty \leq \frac{\|M\|_F}{d}$, *whp,*

$$\|P_\Omega(M) - pM\| \leq \epsilon p \|M\|_F$$

where $\epsilon = O(\sqrt{\log d/(pd)})$.

Proof.

- $P_\Omega(M) = \sum_{i,j} s_{i,j} M_{i,j} \delta_{i,j}$, Where $\delta_{i,j} \in \mathbb{R}^{d\times d}$ is indicator matrix, $s_{i,j} \in \{0,1\}$ are Bernoulli variable with probability p.
- Expectation: $\mathbb{E}[P_\Omega(M)] = pM$
- Variance: $\|\sum_{i,j} \mathbb{E}[s_{i,j}^2 M_{i,j}^2 \delta_{i,j}^\top \delta_{i,j}]\| \leq \frac{p}{d}$
- Apply Bernstein inequality. □

Concentration inequalities for “linear properties”

Concentration inequalities for "linear properties"

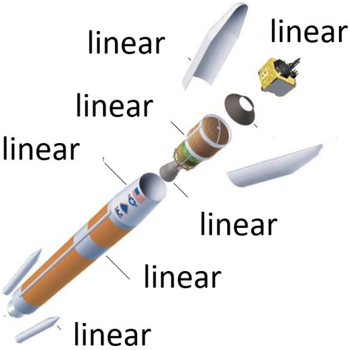
- $|\langle P_\Omega(M), P_\Omega(Z)\rangle - p\langle M, Z\rangle| \leq \tilde{O}(|M|_\infty |Z|_\infty dr)$.
  - $\langle M, Z\rangle \triangleq \mathrm{tr}(M^\top Z)$, entry-wise inner product.

Concentration inequalities for "linear properties"

- $|\langle P_\Omega(M), P_\Omega(Z)\rangle - p\langle M, Z\rangle| \le \tilde{O}(|M|_\infty |Z|_\infty dr)$.
  - $\langle M, Z\rangle \triangleq \mathrm{tr}(M^\top Z)$, entry-wise inner product.
- $\|P_\Omega(XX^\top)X - pXX^\top X\|_F \le p\epsilon\|X\|_F^3$

Concentration inequalities for "linear properties"

- $|\langle P_\Omega(M), P_\Omega(Z)\rangle - p\langle M, Z\rangle| \le \tilde{O}(|M|_\infty |Z|_\infty dr)$.
  - $\langle M, Z\rangle \triangleq \mathrm{tr}(M^\top Z)$, entry-wise inner product.
- $\|P_\Omega(XX^\top)X - pXX^\top X\|_F \le p\epsilon\|X\|_F^3$
- $\cdots$



Rank 1 case

Rank 1 case

- We will stick to rank 1 case today.

Rank 1 case

- We will stick to rank 1 case today.
- Assume $M = zz^\top$.

$$f(x) = \frac{1}{2}\|P_\Omega(M - xx^\top)\|_F^2 + \lambda R(x)$$

where $R(x) = \sum_{i=1}^{d}(|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$. $\alpha = \frac{10\mu}{\sqrt{d}}$.

Rank 1 case

- We will stick to rank 1 case today.
- Assume $M = zz^\top$.

$$f(x) = \frac{1}{2}\|P_\Omega(M - xx^\top)\|_F^2 + \lambda R(x)$$

where $R(x) = \sum_{i=1}^d (|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$. $\alpha = \frac{10\mu}{\sqrt{d}}$.

- We assume $\|z\| = 1, \|z\|_\infty \leq \frac{\mu}{\sqrt{d}}$. (Incoherence)

Rank 1 case

- We will stick to rank 1 case today.
- Assume $M = zz^\top$.

$$f(x) = \frac{1}{2}\|P_\Omega(M - xx^\top)\|_F^2 + \lambda R(x)$$

where $R(x) = \sum_{i=1}^{d}(|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$. $\alpha = \frac{10\mu}{\sqrt{d}}$.

- We assume $\|z\| = 1, \|z\|_\infty \leq \frac{\mu}{\sqrt{d}}$. (Incoherence)
- Why do we need $R(x)$?

Rank 1 case

- We will stick to rank 1 case today.
- Assume $M = zz^\top$.

$$f(x) = \frac{1}{2}\|P_\Omega(M - xx^\top)\|_F^2 + \lambda R(x)$$

where $R(x) = \sum_{i=1}^{d}(|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$. $\alpha = \frac{10\mu}{\sqrt{d}}$.

- We assume $\|z\| = 1, \|z\|_\infty \leq \frac{\mu}{\sqrt{d}}$. (Incoherence)
- Why do we need $R(x)$?
  - The proof does not go through when x has large coordinates.

Rank 1 case

- We will stick to rank 1 case today.
- Assume $M = zz^\top$.

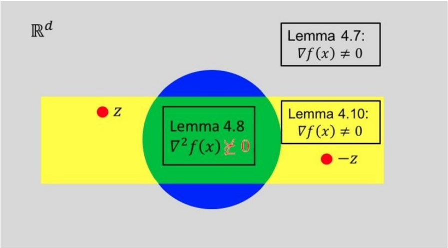
$$f(x) = \frac{1}{2}\|P_\Omega(M - xx^\top)\|_F^2 + \lambda R(x)$$

where $R(x) = \sum_{i=1}^{d}(|x_i| - \alpha)^4 \mathbb{1}_{|x_i|\geq\alpha}$. $\alpha = \frac{10\mu}{\sqrt{d}}$.

- We assume $\|z\| = 1, \|z\|_\infty \leq \frac{\mu}{\sqrt{d}}$. (Incoherence)
- Why do we need $R(x)$?
  - The proof does not go through when x has large coordinates.
  - Add $R(x)$ to make sure x couldn't have large coordinates.

Proof sketch

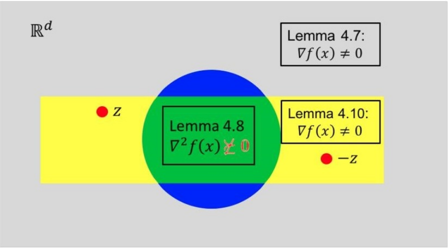
Proof sketch



Proof sketch

$\nabla f(x) = 0$ implies

- $\|x\|_\infty \leq 4\alpha$. i.e., inside yellow.
- $\|zz^\top x - xx^\top x - \gamma \nabla R(x)\|_F \leq O(\epsilon)$.
  i.e., $xx^\top \approx zz^\top$ under "scaling".
  x could be 0.



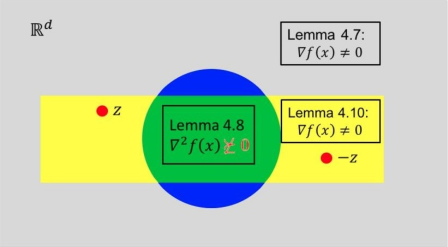
Proof sketch

$\nabla f(x) = 0$ implies

- $\|x\|_\infty \leq 4\alpha$. i.e., inside yellow.
- $\|zz^\top x - xx^\top x - \gamma \nabla R(x)\|_F \leq O(\epsilon)$. i.e., $xx^\top \approx zz^\top$ under "scaling". x could be 0.

$\nabla^2 f(x) \succeq 0$ implies

- $\|x\|^2 \geq \frac{1}{8}$. i.e., outside blue



Proof sketch

$\nabla f(x) = 0$ implies

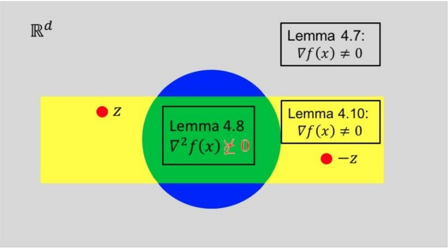
- $\|x\|_\infty \leq 4\alpha$. i.e., inside yellow.
- $\|zz^\top x - xx^\top x - \gamma \nabla R(x)\|_F \leq O(\epsilon)$.
  i.e., $xx^\top \approx zz^\top$ under "scaling".
  x could be 0.

$\nabla^2 f(x) \succeq 0$ implies

- $\|x\|^2 \geq \frac{1}{8}$. i.e., outside blue

Combining them together:

- $\|xx^\top - zz^\top\|_F^2 \leq O(\epsilon)$.



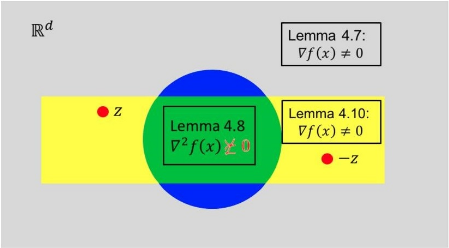
Proof sketch

$\nabla f(x) = 0$ implies

- $\|x\|_\infty \leq 4\alpha$. i.e., inside yellow.
- $\|zz^\top x - xx^\top x - \gamma \nabla R(x)\|_F \leq O(\epsilon)$.
  i.e., $xx^\top \approx zz^\top$ under "scaling".
  x could be 0.

$\nabla^2 f(x) \succeq 0$ implies

- $\|x\|^2 \geq \frac{1}{8}$. i.e., outside blue



Combining them together:

- $\|xx^\top - zz^\top\|_F^2 \leq O(\epsilon)$.

Handy Lemma:
If $\|xx^\top - zz^\top\|_F^2 \leq O(\epsilon)$, then $xx^\top = zz^\top = M$. [Sun and Luo 2015]

First thing first: optimality conditions

$$
\begin{aligned}
& f(x+\delta) \\
=& \frac{1}{2}\|P_\Omega(M-(x+\delta)(x+\delta)^\top\|_F^2 + \lambda R(x+\delta) + o(\|\delta\|^2) \\
=& \frac{1}{2}\|P_\Omega(M-xx^\top-(x\delta^\top+\delta x^\top)-\delta\delta^\top\|_F^2 \\
& + \lambda\left(R(x)+\langle\nabla R(x),\delta\rangle+\frac{1}{2}\delta^\top\nabla^2 R(x)\delta\right)+o(\|\delta\|^2) \\
=& f(x) - \langle P_\Omega(M-xx^\top), x\delta^\top+\delta x^\top\rangle + \langle\lambda\nabla R(x),\delta\rangle + o(\|\delta\|^2) \\
& -\langle P_\Omega(M-xx^\top),\delta\delta^\top\rangle + \frac{1}{2}\|P_\Omega(x\delta^\top+\delta x^\top)\|_F^2 + \frac{1}{2}\lambda\delta\nabla^2 r(x)\delta
\end{aligned}
$$

First thing first: optimality conditions

$$
\begin{aligned}
& f(x+\delta) \\
=& \frac{1}{2}\|P_\Omega(M-(x+\delta)(x+\delta)^\top\|_F^2 + \lambda R(x+\delta) + o(\|\delta\|^2) \\
=& \frac{1}{2}\|P_\Omega(M - xx^\top - (x\delta^\top + \delta x^\top) - \delta\delta^\top\|_F^2 \\
& + \lambda\left(R(x) + \langle \nabla R(x), \delta\rangle + \frac{1}{2}\delta^\top \nabla^2 R(x)\delta\right) + o(\|\delta\|^2) \\
=& f(x) - \langle P_\Omega(M - xx^\top), x\delta^\top + \delta x^\top\rangle + \langle \lambda \nabla R(x), \delta\rangle + o(\|\delta\|^2) \\
& - \langle P_\Omega(M - xx^\top), \delta\delta^\top\rangle + \frac{1}{2}\|P_\Omega(x\delta^\top + \delta x^\top)\|_F^2 + \frac{1}{2}\lambda\delta\nabla^2 r(x)\delta
\end{aligned}
$$

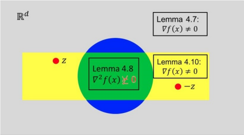
- First order condition: $2P_\Omega(M - xx^\top)x = \lambda \nabla R(x)$

First thing first: optimality conditions

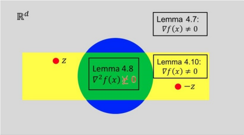
$$
\begin{aligned}
& f(x+\delta) \\
=& \frac{1}{2}\|P_\Omega(M-(x+\delta)(x+\delta)^\top\|_F^2 + \lambda R(x+\delta) + o(\|\delta\|^2) \\
=& \frac{1}{2}\|P_\Omega(M - xx^\top - (x\delta^\top + \delta x^\top) - \delta\delta^\top\|_F^2 \\
& + \lambda\left(R(x) + \langle \nabla R(x), \delta\rangle + \frac{1}{2}\delta^\top \nabla^2 R(x)\delta\right) + o(\|\delta\|^2) \\
=& f(x) - \langle P_\Omega(M - xx^\top), x\delta^\top + \delta x^\top\rangle + \langle \lambda \nabla R(x), \delta\rangle + o(\|\delta\|^2) \\
& - \langle P_\Omega(M - xx^\top), \delta\delta^\top\rangle + \frac{1}{2}\|P_\Omega(x\delta^\top + \delta x^\top)\|_F^2 + \frac{1}{2}\lambda\delta\nabla^2 r(x)\delta
\end{aligned}
$$

- First order condition: $2P_\Omega(M - xx^\top)x = \lambda \nabla R(x)$
- Second order condition:
  $\forall \delta, \|P_\Omega(\delta x^\top + x\delta^\top)\|_F^2 + \lambda \delta^\top \nabla^2 R(x)\delta \geq 2\delta^\top P_\Omega(M - xx^\top)\delta$

First order condition: yellow box



First order condition: yellow box



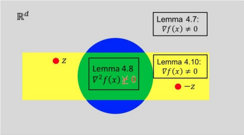
Theorem

Let $p \geq \frac{c \log d}{d}, \alpha = 10\mu/\sqrt{d}$, $\lambda \geq \mu^2 p/\alpha^2$. *Whp over* Ω, *for any* x *s.t.* $\nabla f(x) = 0$, *we have*

$$\|x\|_\infty \leq 4 \max\{\alpha, \mu\sqrt{p/\lambda}\} = 4\alpha$$

$R(x)$ helps in this case.

First order condition: yellow box



Theorem

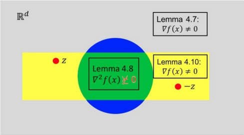
Let $p \geq \frac{c \log d}{d}, \alpha = 10\mu/\sqrt{d}$, $\lambda \geq \mu^2 p/\alpha^2$. Whp over Ω, for any x s.t. $\nabla f(x) = 0$, we have

$$\|x\|_\infty \leq 4 \max\{\alpha, \mu\sqrt{p/\lambda}\} = 4\alpha$$

$R(x)$ helps in this case.

- $i^* = \max_i |x_i|$. WLOG, $x_{i^*} \geq 0$.

First order condition: yellow box



Theorem

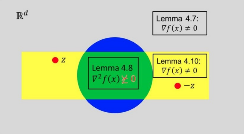
Let $p \geq \frac{c \log d}{d}, \alpha = 10\mu/\sqrt{d}$, $\lambda \geq \mu^2 p/\alpha^2$. Whp over Ω, for any x s.t. $\nabla f(x) = 0$, we have

$$\|x\|_\infty \leq 4 \max\{\alpha, \mu\sqrt{p/\lambda}\} = 4\alpha$$

$R(x)$ helps in this case.

- $i^* = \max_i |x_i|$. WLOG, $x_{i^*} \geq 0$.
- If $x_{i^*} \leq 2\alpha$, done. Otherwise $x_{i^*} > 2\alpha$

First order condition: yellow box



Theorem

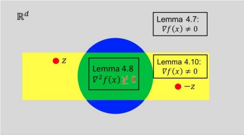
Let $p \geq \frac{c \log d}{d}$, $\alpha = 10\mu/\sqrt{d}$, $\lambda \geq \mu^2 p/\alpha^2$. Whp over Ω, for any x s.t. $\nabla f(x) = 0$, we have

$$\|x\|_\infty \leq 4 \max\{\alpha, \mu\sqrt{p/\lambda}\} = 4\alpha$$

$R(x)$ helps in this case.

- $i^* = \max_i |x_i|$. WLOG, $x_{i^*} \geq 0$.
- If $x_{i^*} \leq 2\alpha$, done. Otherwise $x_{i^*} > 2\alpha$
- Assume Ω contains $|S_{i^*}|$ entries in row i^*.

First order condition: yellow box



Theorem

Let $p \geq \frac{c \log d}{d}, \alpha = 10\mu/\sqrt{d}$, $\lambda \geq \mu^2 p/\alpha^2$. *Whp over* Ω, *for any* x *s.t.* $\nabla f(x) = 0$, *we have*

$$\|x\|_\infty \leq 4 \max\{\alpha, \mu\sqrt{p/\lambda}\} = 4\alpha$$

$R(x)$ helps in this case.

- $i^* = \max_i |x_i|$. WLOG, $x_{i^*} \geq 0$.
- If $x_{i^*} \leq 2\alpha$, done. Otherwise $x_{i^*} > 2\alpha$
- Assume Ω contains $|S_{i^*}|$ entries in row i^*.
- whp, $|S_{i^*}| \leq 2pd$.

First order condition: yellow box

Recall first order condition: $2(P_\Omega(M - xx^\top)x)_{i*} = (\lambda \nabla R(x))_{i*}$

First order condition: yellow box

Recall first order condition: $2(P_\Omega(M - xx^\top)x)_{i^*} = (\lambda \nabla R(x))_{i^*}$

$$|(P_\Omega(M)x)_{i^*}| = |(P_\Omega(zz^\top)x)_{i^*}| = \left| \sum_{j \in S_{i^*}} z_{i^*} z_j x_j \right|$$

$$\leq |x_{i^*}| \sum_{j \in S_{i^*}} |z_{i^*} z_j| \leq |x_{i^*}| \cdot \mu^2/d \cdot |S_{i^*}| \leq 2|x_{i^*}| p\mu^2$$

First order condition: yellow box

Recall first order condition: $2(P_\Omega(M - xx^\top)x)_{i^*} = (\lambda \nabla R(x))_{i^*}$

$$
\begin{aligned}
|(P_\Omega(M)x)_{i^*}| = |(P_\Omega(zz^\top)x)_{i^*}| &= \left| \sum_{j \in S_{i^*}} z_{i^*} z_j x_j \right| \\
&\leq |x_{i^*}| \sum_{j \in S_{i^*}} |z_{i^*} z_j| \leq |x_{i^*}| \cdot \mu^2/d \cdot |S_{i^*}| \leq 2|x_{i^*}| p\mu^2
\end{aligned}
$$

Moreover, $(P_\Omega(xx^\top)x)_{i^*} = \sum_{j \in S_{i^*}} x_{i^*} x_j^2 \geq 0$

First order condition: yellow box

Recall first order condition: $2(P_\Omega(M - xx^\top)x)_{i^*} = (\lambda \nabla R(x))_{i^*}$

$$
\begin{aligned}
|(P_\Omega(M)x)_{i^*}| = |(P_\Omega(zz^\top)x)_{i^*}| &= \left| \sum_{j \in S_{i^*}} z_{i^*} z_j x_j \right| \\
&\leq |x_{i^*}| \sum_{j \in S_{i^*}} |z_{i^*} z_j| \leq |x_{i^*}| \cdot \mu^2/d \cdot |S_{i^*}| \leq 2|x_{i^*}| p\mu^2
\end{aligned}
$$

Moreover, $(P_\Omega(xx^\top)x)_{i^*} = \sum_{j \in S_{i^*}} x_{i^*} x_j^2 \geq 0$
And

$$
(\lambda \nabla R(x))_{i^*} = 4\lambda(|x_{i^*}| - \alpha)^3 \geq \frac{\lambda}{2}|x_{i^*}|^3 \quad (\text{since } x_{i^*} \geq 2\alpha)
$$

First order condition: yellow box

Recall first order condition: $2(P_\Omega(M - xx^\top)x)_{i^*} = (\lambda \nabla R(x))_{i^*}$

$$|(P_\Omega(M)x)_{i^*}| = |(P_\Omega(zz^\top)x)_{i^*}| = \left|\sum_{j \in S_{i^*}} z_{i^*} z_j x_j\right|$$

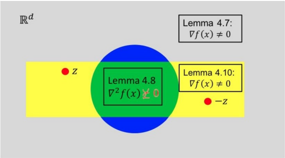
$$\leq |x_{i^*}| \sum_{j \in S_{i^*}} |z_{i^*} z_j| \leq |x_{i^*}| \cdot \mu^2/d \cdot |S_{i^*}| \leq 2|x_{i^*}|p\mu^2$$

Moreover, $(P_\Omega(xx^\top)x)_{i^*} = \sum_{j \in S_{i^*}} x_{i^*} x_j^2 \geq 0$
And

$$(\lambda \nabla R(x))_{i^*} = 4\lambda(|x_{i^*}| - \alpha)^3 \geq \frac{\lambda}{2}|x_{i^*}|^3 \quad (\text{since } x_{i^*} \geq 2\alpha)$$

So $4|x_{i^*}|p\mu^2 \geq 2(P_\Omega(M - xx^\top)x)_{i^*} = (\lambda \nabla R(x))_{i^*} \geq \frac{\lambda}{2}|\mathbf{x}_{\mathbf{i}^*}|^3$
Therefore $|x_{i^*}| \leq 4\sqrt{p\mu^2/\lambda} \leq 4\alpha$.

Second order condition: outside blue ball



Theorem

Whp over Ω, for any $x \in$ yellow s.t. $\nabla^2 f(x) \succeq 0$, we have $\|x\|^2 \geq 1/8$.

Bound $R(x)$ in this case.

Second order condition: outside blue ball

Theorem

Whp over Ω, for any $x \in$ yellow s.t. $\nabla^2 f(x) \succeq 0$, we have $\|x\|^2 \geq 1/8$.

Second order condition: outside blue ball

Theorem

Whp over Ω, for any $x \in$ yellow s.t. $\nabla^2 f(x) \succeq 0$, we have $\|x\|^2 \geq 1/8$.

- WLOG, assume $\|x\| < 1$. Let $J = \{i : |x_i| \leq \alpha\}$, so

Second order condition: outside blue ball

Theorem

Whp over Ω, for any $x \in$ yellow s.t. $\nabla^2 f(x) \succeq 0$, we have $\|x\|^2 \geq 1/8$.

- WLOG, assume $\|x\| < 1$. Let $J = \{i : |x_i| \leq \alpha\}$, so
  - $R(x) = \sum_{i=1}^{d}(|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$, so $\nabla^2 R(x)$ is diagonal.

Second order condition: outside blue ball

Theorem

Whp over Ω, for any $x \in$ yellow s.t. $\nabla^2 f(x) \succeq 0$, we have $\|x\|^2 \geq 1/8$.

- WLOG, assume $\|x\| < 1$. Let $J = \{i : |x_i| \leq \alpha\}$, so
  - $R(x) = \sum_{i=1}^{d} (|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$, so $\nabla^2 R(x)$ is diagonal.
  - z_J is z mapped to J

Second order condition: outside blue ball

Theorem

Whp over Ω, for any $x \in$ yellow s.t. $\nabla^2 f(x) \succeq 0$, we have $\|x\|^2 \geq 1/8$.

- WLOG, assume $\|x\| < 1$. Let $J = \{i : |x_i| \leq \alpha\}$, so
  - $R(x) = \sum_{i=1}^{d}(|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$, so $\nabla^2 R(x)$ is diagonal.
  - z_J is z mapped to J
  - for $i \in J$, $(\nabla^2 R(x))_{i,i} = 0$.

Second order condition: outside blue ball

Theorem

Whp over Ω, for any $x \in$ yellow s.t. $\nabla^2 f(x) \succeq 0$, we have $\|x\|^2 \geq 1/8$.

- WLOG, assume $\|x\| < 1$. Let $J = \{i : |x_i| \leq \alpha\}$, so
  - $R(x) = \sum_{i=1}^{d}(|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$, so $\nabla^2 R(x)$ is diagonal.
  - z_J is z mapped to J
  - for $i \in J$, $(\nabla^2 R(x))_{i,i} = 0$.
  - $z_J^\top \nabla^2 R(x) z_J = 0$.

Second order condition: outside blue ball

Theorem

Whp over Ω, for any $x \in$ yellow s.t. $\nabla^2 f(x) \succeq 0$, we have $\|x\|^2 \geq 1/8$.

- WLOG, assume $\|x\| < 1$. Let $J = \{i : |x_i| \leq \alpha\}$, so
  - $R(x) = \sum_{i=1}^{d} (|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$, so $\nabla^2 R(x)$ is diagonal.
  - z_J is z mapped to J
  - for $i \in J$, $(\nabla^2 R(x))_{i,i} = 0$.
  - $z_J^\top \nabla^2 R(x) z_J = 0$.
- **Second order optimality**:
  $\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 \geq 2 z_J^\top P_\Omega(M - xx^\top) z_J$

Second order condition: outside blue ball

Theorem

Whp over Ω, for any $x \in$ yellow s.t. $\nabla^2 f(x) \succeq 0$, we have $\|x\|^2 \geq 1/8$.

- WLOG, assume $\|x\| < 1$. Let $J = \{i : |x_i| \leq \alpha\}$, so
  - $R(x) = \sum_{i=1}^{d} (|x_i| - \alpha)^4 \mathbb{1}_{|x_i| \geq \alpha}$, so $\nabla^2 R(x)$ is diagonal.
  - z_J is z mapped to J
  - for $i \in J$, $(\nabla^2 R(x))_{i,i} = 0$.
  - $z_J^\top \nabla^2 R(x) z_J = 0$.
- **Second order optimality**: $\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 \geq 2 z_J^\top P_\Omega(M - xx^\top) z_J$
- $\|x\|_\infty \leq 4\alpha, \|z_J\|_\infty \leq \frac{\mu}{\sqrt{d}}, \|z_J\| \leq 1$

Second order condition: outside blue ball

- **Linear property**:
  $\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 = 2\langle P_\Omega(z_J x^\top), P_\Omega(z_J x^\top)\rangle + 2\langle P_\Omega(z_J x^\top), P_\Omega(x z_J^\top)\rangle \approx 2p\|x\|^2\|z_J\|^2 + 2p\langle x, z_J\rangle^2$

Second order condition: outside blue ball

- **Linear property**:
  $\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 = 2\langle P_\Omega(z_J x^\top), P_\Omega(z_J x^\top)\rangle + 2\langle P_\Omega(z_J x^\top), P_\Omega(x z_J^\top)\rangle \approx 2p\|x\|^2\|z_J\|^2 + 2p\langle x, z_J\rangle^2$
- **Linear property**:
  $2z_J^\top P_\Omega(M - xx^\top)z_J = 2\left(\langle P_\Omega(zz^\top), P_\Omega(z_J z_J^\top)\rangle - \langle P_\Omega(xx^\top), P_\Omega(z_j z_J^\top)\rangle\right) \approx 2p\|z_J\|^4 - 2p\langle x, z_J\rangle^2$

Second order condition: outside blue ball

- **Linear property**: $\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 = 2\langle P_\Omega(z_J x^\top), P_\Omega(z_J x^\top)\rangle + 2\langle P_\Omega(z_J x^\top), P_\Omega(x z_J^\top)\rangle \approx 2p\|x\|^2\|z_J\|^2 + 2p\langle x, z_J\rangle^2$
- **Linear property**: $2z_J^\top P_\Omega(M - xx^\top)z_J = 2\left(\langle P_\Omega(zz^\top), P_\Omega(z_J z_J^\top)\rangle - \langle P_\Omega(xx^\top), P_\Omega(z_j z_J^\top)\rangle\right) \approx 2p\|z_J\|^4 - 2p\langle x, z_J\rangle^2$
- Thus, $p\|x\|^2\|z_J\|^2 + p\langle x, z_j\rangle^2 \geq p\|z_J\|^4 - p\langle x, z_J\rangle^2$

Second order condition: outside blue ball

- **Linear property**:
  $\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 = 2\langle P_\Omega(z_J x^\top), P_\Omega(z_J x^\top)\rangle + 2\langle P_\Omega(z_J x^\top), P_\Omega(x z_J^\top)\rangle \approx 2p\|x\|^2\|z_J\|^2 + 2p\langle x, z_J\rangle^2$
- **Linear property**:
  $2z_J^\top P_\Omega(M - xx^\top)z_J = 2\left(\langle P_\Omega(zz^\top), P_\Omega(z_J z_J^\top)\rangle - \langle P_\Omega(xx^\top), P_\Omega(z_j z_J^\top)\rangle\right) \approx 2p\|z_J\|^4 - 2p\langle x, z_J\rangle^2$
- Thus, $p\|x\|^2\|z_J\|^2 + p\langle x, z_j\rangle^2 \geq p\|z_J\|^4 - p\langle x, z_J\rangle^2$
- Cauchy-Schwartz, $3p\|x\|^2\|z_J\|^2 \geq p\|z_J\|^4$

Second order condition: outside blue ball

- **Linear property**:
$\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 = 2\langle P_\Omega(z_J x^\top), P_\Omega(z_J x^\top)\rangle + 2\langle P_\Omega(z_J x^\top), P_\Omega(x z_J^\top)\rangle \approx 2p\|x\|^2\|z_J\|^2 + 2p\langle x, z_J\rangle^2$
- **Linear property**:
$2z_J^\top P_\Omega(M - xx^\top)z_J = 2\left(\langle P_\Omega(zz^\top), P_\Omega(z_J z_J^\top)\rangle - \langle P_\Omega(xx^\top), P_\Omega(z_j z_J^\top)\rangle\right) \approx 2p\|z_J\|^4 - 2p\langle x, z_J\rangle^2$
- Thus, $p\|x\|^2\|z_J\|^2 + p\langle x, z_j\rangle^2 \geq p\|z_J\|^4 - p\langle x, z_J\rangle^2$
- Cauchy-Schwartz, $3p\|x\|^2\|z_J\|^2 \geq p\|z_J\|^4$
- $\|x\|^2 \geq \frac{1}{3}\|z_J\|^2$

Second order condition: outside blue ball

- **Linear property**:
  $\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 = 2\langle P_\Omega(z_J x^\top), P_\Omega(z_J x^\top)\rangle + 2\langle P_\Omega(z_J x^\top), P_\Omega(x z_J^\top)\rangle \approx 2p\|x\|^2\|z_J\|^2 + 2p\langle x, z_J\rangle^2$
- **Linear property**:
  $2z_J^\top P_\Omega(M - xx^\top)z_J = 2\left(\langle P_\Omega(zz^\top), P_\Omega(z_J z_J^\top)\rangle - \langle P_\Omega(xx^\top), P_\Omega(z_j z_J^\top)\rangle\right) \approx 2p\|z_J\|^4 - 2p\langle x, z_J\rangle^2$
- Thus, $p\|x\|^2\|z_J\|^2 + p\langle x, z_j\rangle^2 \geq p\|z_J\|^4 - p\langle x, z_J\rangle^2$
- Cauchy-Schwartz, $3p\|x\|^2\|z_J\|^2 \geq p\|z_J\|^4$
- $\|x\|^2 \geq \frac{1}{3}\|z_J\|^2$
- Since $\|x\| \leq 1$, $|J^c| \leq 1/\alpha^2$. So $\|z_{J^c}\|^2 \leq \mu^2/(d\alpha^2)$

Second order condition: outside blue ball

- **Linear property**:
  $\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 = 2\langle P_\Omega(z_J x^\top), P_\Omega(z_J x^\top)\rangle + 2\langle P_\Omega(z_J x^\top), P_\Omega(x z_J^\top)\rangle \approx 2p\|x\|^2\|z_J\|^2 + 2p\langle x, z_J\rangle^2$
- **Linear property**:
  $2z_J^\top P_\Omega(M - xx^\top)z_J = 2\left(\langle P_\Omega(zz^\top), P_\Omega(z_J z_J^\top)\rangle - \langle P_\Omega(xx^\top), P_\Omega(z_j z_J^\top)\rangle\right) \approx 2p\|z_J\|^4 - 2p\langle x, z_J\rangle^2$
- Thus, $p\|x\|^2\|z_J\|^2 + p\langle x, z_j\rangle^2 \geq p\|z_J\|^4 - p\langle x, z_J\rangle^2$
- Cauchy-Schwartz, $3p\|x\|^2\|z_J\|^2 \geq p\|z_J\|^4$
- $\|x\|^2 \geq \frac{1}{3}\|z_J\|^2$
- Since $\|x\| \leq 1$, $|J^c| \leq 1/\alpha^2$. So $\|z_{J^c}\|^2 \leq \mu^2/(d\alpha^2)$
- $\|z_J\|^2 = \|z\|^2 - \|z_{J^c}\|^2 = 1 - \mu^2/(d\alpha^2) \geq \frac{1}{2}$.

Second order condition: outside blue ball

- **Linear property**:
  $\|P_\Omega(z_J x^\top + x z_J^\top)\|_F^2 = 2\langle P_\Omega(z_J x^\top), P_\Omega(z_J x^\top)\rangle + 2\langle P_\Omega(z_J x^\top), P_\Omega(x z_J^\top)\rangle \approx 2p\|x\|^2\|z_J\|^2 + 2p\langle x, z_J\rangle^2$
- **Linear property**:
  $2z_J^\top P_\Omega(M - xx^\top)z_J = 2\left(\langle P_\Omega(zz^\top), P_\Omega(z_J z_J^\top)\rangle - \langle P_\Omega(xx^\top), P_\Omega(z_j z_J^\top)\rangle\right) \approx 2p\|z_J\|^4 - 2p\langle x, z_J\rangle^2$
- Thus, $p\|x\|^2\|z_J\|^2 + p\langle x, z_j\rangle^2 \geq p\|z_J\|^4 - p\langle x, z_J\rangle^2$
- Cauchy-Schwartz, $3p\|x\|^2\|z_J\|^2 \geq p\|z_J\|^4$
- $\|x\|^2 \geq \frac{1}{3}\|z_J\|^2$
- Since $\|x\| \leq 1$, $|J^c| \leq 1/\alpha^2$. So $\|z_{J^c}\|^2 \leq \mu^2/(d\alpha^2)$
- $\|z_J\|^2 = \|z\|^2 - \|z_{J^c}\|^2 = 1 - \mu^2/(d\alpha^2) \geq \frac{1}{2}$.
- Therefore $\|x\|^2 \geq \frac{1}{3} \cdot \frac{1}{2} > \frac{1}{8}$

First order condition: $x \approx z$ after scaling

Theorem

Whp over Ω, for any x s.t. $\nabla f(x) = 0$, we have

$$\|zz^\top x - xx^\top x - \frac{\lambda}{2p} \cdot \nabla R(x)\| \leq O(\epsilon)$$

where $\epsilon = \tilde{O}(\mu^6 (pd)^{-1/2})$.
Bound $R(x)$ in this case.

First order condition: $x \approx z$ after scaling

Theorem

Whp over Ω, for any x s.t. $\nabla f(x) = 0$, we have

$$\|zz^\top x - xx^\top x - \frac{\lambda}{2p} \cdot \nabla R(x)\| \leq O(\epsilon)$$

where $\epsilon = \tilde{O}(\mu^6 (pd)^{-1/2})$.
Bound $R(x)$ in this case.

- $\nabla f(x) = 0 \Rightarrow 2P_\Omega(M - xx^\top)x = \lambda \nabla R(x)$

First order condition: $x \approx z$ after scaling

Theorem

Whp over Ω, for any x s.t. $\nabla f(x) = 0$, we have

$$\|zz^\top x - xx^\top x - \frac{\lambda}{2p} \cdot \nabla R(x)\| \leq O(\epsilon)$$

where $\epsilon = \tilde{O}(\mu^6 (pd)^{-1/2})$.
Bound $R(x)$ in this case.

- $\nabla f(x) = 0 \Rightarrow 2P_\Omega(M - xx^\top)x = \lambda \nabla R(x)$
- $2P_\Omega(zz^\top)x - 2P_\Omega(xx^\top)x - \lambda \nabla R(x) = 0$

First order condition: $x \approx z$ after scaling

Theorem

Whp over Ω, for any x s.t. $\nabla f(x) = 0$, we have

$$\|zz^\top x - xx^\top x - \frac{\lambda}{2p} \cdot \nabla R(x)\| \le O(\epsilon)$$

where $\epsilon = \tilde{O}(\mu^6 (pd)^{-1/2})$.
Bound $R(x)$ in this case.

- $\nabla f(x) = 0 \Rightarrow 2P_\Omega(M - xx^\top)x = \lambda \nabla R(x)$
- $2P_\Omega(zz^\top)x - 2P_\Omega(xx^\top)x - \lambda \nabla R(x) = 0$
- **Linear property**:

$$\|2P_\Omega(zz^\top)x - 2pzz^\top x\| \le p\epsilon$$
$$\|2P_\Omega(xx^\top)x - 2pxx^\top x\| \le p\epsilon$$

First order condition: $x \approx z$ after scaling

Theorem

Whp over Ω, for any x s.t. $\nabla f(x) = 0$, we have

$$\|zz^\top x - xx^\top x - \frac{\lambda}{2p} \cdot \nabla R(x)\| \leq O(\epsilon)$$

where $\epsilon = \tilde{O}(\mu^6 (pd)^{-1/2})$.
Bound $R(x)$ in this case.

- $\nabla f(x) = 0 \Rightarrow 2P_\Omega(M - xx^\top)x = \lambda \nabla R(x)$
- $2P_\Omega(zz^\top)x - 2P_\Omega(xx^\top)x - \lambda \nabla R(x) = 0$
- **Linear property**:

$$\|2P_\Omega(zz^\top)x - 2pzz^\top x\| \leq p\epsilon$$

$$\|2P_\Omega(xx^\top)x - 2pxx^\top x\| \leq p\epsilon$$

- Therefore $\|zz^\top x - xx^\top x - \frac{\lambda}{2p} \cdot \nabla R(x)\| \leq O(\epsilon)$.

Safely ignore $\nabla R(x)$..

Theorem

Whp over Ω, Suppose $x \in$ yellow s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\nabla f(x) = 0$, we have $\|zz^\top x - xx^\top x\| \leq O(\epsilon)$.

Safely ignore $\nabla R(x)$..

Theorem

Whp over Ω, Suppose $x \in$ yellow s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\nabla f(x) = 0$, we have $\|zz^\top x - xx^\top x\| \leq O(\epsilon)$.

- Recall $J = \{i : |x_i| \leq \alpha\}$. For $i \in J$, $(\nabla R(x))_i = 0$

Safely ignore $\nabla R(x)$..

Theorem

Whp over Ω, Suppose $x \in$ yellow s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\nabla f(x) = 0$, we have $\|zz^\top x - xx^\top x\| \leq O(\epsilon)$.

- Recall $J = \{i : |x_i| \leq \alpha\}$. For $i \in J$, $(\nabla R(x))_i = 0$
- For $i \in J^c$, we want
  $(z_i z^\top x - x_i x^\top x)^2 \leq (z_i z^\top x - x_i x^\top x - (\gamma \nabla R(x))_i)^2$

Safely ignore $\nabla R(x)$..

Theorem

Whp over Ω, Suppose $x \in$ yellow s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\nabla f(x) = 0$, we have $\|zz^\top x - xx^\top x\| \leq O(\epsilon)$.

- Recall $J = \{i : |x_i| \leq \alpha\}$. For $i \in J$, $(\nabla R(x))_i = 0$
- For $i \in J^c$, we want $(z_i z^\top x - x_i x^\top x)^2 \leq (z_i z^\top x - x_i x^\top x - (\gamma \nabla R(x))_i)^2$
- I.e., $(z_i z^\top x - x_i x^\top x)(\nabla R(x))_i \leq 0$

Safely ignore $\nabla R(x)$..

Theorem

Whp over Ω, Suppose $x \in$ yellow s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\nabla f(x) = 0$, we have $\|zz^\top x - xx^\top x\| \leq O(\epsilon)$.

- Recall $J = \{i : |x_i| \leq \alpha\}$. For $i \in J$, $(\nabla R(x))_i = 0$
- For $i \in J^c$, we want $(z_i z^\top x - x_i x^\top x)^2 \leq (z_i z^\top x - x_i x^\top x - (\gamma \nabla R(x))_i)^2$
- I.e., $(z_i z^\top x - x_i x^\top x)(\nabla R(x))_i \leq 0$
- By definition, this is true because

Safely ignore $\nabla R(x)$..

Theorem

Whp over Ω, Suppose $x \in$ yellow s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\nabla f(x) = 0$, we have $\|zz^\top x - xx^\top x\| \leq O(\epsilon)$.

- Recall $J = \{i : |x_i| \leq \alpha\}$. For $i \in J$, $(\nabla R(x))_i = 0$
- For $i \in J^c$, we want
  $(z_i z^\top x - x_i x^\top x)^2 \leq (z_i z^\top x - x_i x^\top x - (\gamma \nabla R(x))_i)^2$
- I.e., $(z_i z^\top x - x_i x^\top x)(\nabla R(x))_i \leq 0$
- By definition, this is true because
  - $(\nabla R(x))_i = \gamma x_i$

Safely ignore $\nabla R(x)$..

Theorem

Whp over Ω, Suppose $x \in$ yellow s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\nabla f(x) = 0$, we have $\|zz^\top x - xx^\top x\| \leq O(\epsilon)$.

- Recall $J = \{i : |x_i| \leq \alpha\}$. For $i \in J$, $(\nabla R(x))_i = 0$
- For $i \in J^c$, we want $(z_i z^\top x - x_i x^\top x)^2 \leq (z_i z^\top x - x_i x^\top x - (\gamma \nabla R(x))_i)^2$
- I.e., $(z_i z^\top x - x_i x^\top x)(\nabla R(x))_i \leq 0$
- By definition, this is true because
  - $(\nabla R(x))_i = \gamma x_i$
  - $\|x\|^2 \geq \frac{1}{8}$

Safely ignore $\nabla R(x)$..

Theorem

Whp over Ω, Suppose $x \in$ yellow s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\nabla f(x) = 0$, we have $\|zz^\top x - xx^\top x\| \leq O(\epsilon)$.

- Recall $J = \{i : |x_i| \leq \alpha\}$. For $i \in J$, $(\nabla R(x))_i = 0$
- For $i \in J^c$, we want $(z_i z^\top x - x_i x^\top x)^2 \leq (z_i z^\top x - x_i x^\top x - (\gamma \nabla R(x))_i)^2$
- I.e., $(z_i z^\top x - x_i x^\top x)(\nabla R(x))_i \leq 0$
- By definition, this is true because
  - $(\nabla R(x))_i = \gamma x_i$
  - $\|x\|^2 \geq \frac{1}{8}$
  - $|x_i| \geq \alpha \geq 4|z_i|$

Safely ignore $\nabla R(x)$..

Theorem

Whp over Ω, Suppose $x \in$ yellow s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\nabla f(x) = 0$, we have $\|zz^\top x - xx^\top x\| \leq O(\epsilon)$.

- Recall $J = \{i : |x_i| \leq \alpha\}$. For $i \in J$, $(\nabla R(x))_i = 0$
- For $i \in J^c$, we want $(z_i z^\top x - x_i x^\top x)^2 \leq (z_i z^\top x - x_i x^\top x - (\gamma \nabla R(x))_i)^2$
- I.e., $(z_i z^\top x - x_i x^\top x)(\nabla R(x))_i \leq 0$
- By definition, this is true because
  - $(\nabla R(x))_i = \gamma x_i$
  - $\|x\|^2 \geq \frac{1}{8}$
  - $|x_i| \geq \alpha \geq 4|z_i|$
  - $\|z\| = 1$.

Combine everything

Theorem

Suppose x s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\|zz^\top x - xx^\top x\| \leq \delta$. Then for $\delta \in (0, 0.1)$, $\|xx^\top - zz^\top\|_F^2 \leq O(\delta)$.

Combine everything

Theorem

Suppose x s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\|zz^\top x - xx^\top x\| \leq \delta$. Then for $\delta \in (0, 0.1)$, $\|xx^\top - zz^\top\|_F^2 \leq O(\delta)$.

- $z = ux + v$, where $\langle v, x \rangle = 0$.

Combine everything

Theorem

Suppose x s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\|zz^\top x - xx^\top x\| \leq \delta$. Then for $\delta \in (0, 0.1)$, $\|xx^\top - zz^\top\|_F^2 \leq O(\delta)$.

- $z = ux + v$, where $\langle v, x \rangle = 0$.
- $zz^\top x = u^2\|x\|^2 x + u\|x\|^2 v$

Combine everything

Theorem

Suppose x s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\|zz^\top x - xx^\top x\| \leq \delta$. Then for $\delta \in (0, 0.1)$, $\|xx^\top - zz^\top\|_F^2 \leq O(\delta)$.

- $z = ux + v$, where $\langle v, x \rangle = 0$.
- $zz^\top x = u^2\|x\|^2 x + u\|x\|^2 v$
- $\delta \geq \|zz^\top x - xx^\top x\| = \|x\|^2 \sqrt{u^2\|v\|^2 + (1 - u^2)^2}$

Combine everything

Theorem

Suppose x s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\|zz^\top x - xx^\top x\| \leq \delta$. Then for $\delta \in (0, 0.1)$, $\|xx^\top - zz^\top\|_F^2 \leq O(\delta)$.

- $z = ux + v$, where $\langle v, x \rangle = 0$.
- $zz^\top x = u^2\|x\|^2 x + u\|x\|^2 v$
- $\delta \geq \|zz^\top x - xx^\top x\| = \|x\|^2\sqrt{u^2\|v\|^2 + (1-u^2)^2}$
- So $|1 - u^2| \leq 8\delta$, $u\|v\| \leq 8\delta$.

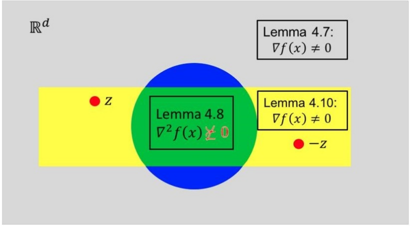
Combine everything

Theorem

Suppose x s.t. $\|x\|^2 \geq \frac{1}{8}$, and $\|zz^\top x - xx^\top x\| \leq \delta$. Then for $\delta \in (0, 0.1)$, $\|xx^\top - zz^\top\|_F^2 \leq O(\delta)$.

- $z = ux + v$, where $\langle v, x \rangle = 0$.
- $zz^\top x = u^2\|x\|^2 x + u\|x\|^2 v$
- $\delta \geq \|zz^\top x - xx^\top x\| = \|x\|^2 \sqrt{u^2\|v\|^2 + (1 - u^2)^2}$
- So $|1 - u^2| \leq 8\delta$, $u\|v\| \leq 8\delta$.
- $|u| \approx 1, \|v\| \approx 0 \Rightarrow z \approx x$.

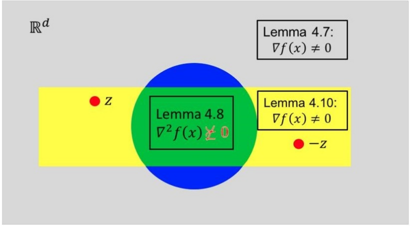
Rank k case...



$\mathbb{R}^d$
Lemma 4.7:
$\nabla f(x) \neq 0$
z
Lemma 4.8
$\nabla^2 f(x) \not\succeq 0$
Lemma 4.10:
$\nabla f(x) \neq 0$
$-z$

Rank k case...

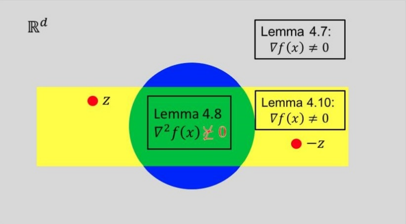
$\nabla f(x) = 0 \Rightarrow \nabla f(X) = 0$



Rank k case...

$\nabla f(x) = 0 \Rightarrow \nabla f(X) = 0$

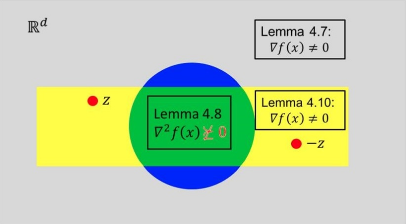
- $\|x\|_\infty \leq 4\alpha \Rightarrow \max_i \|X_i\| \leq 4\alpha$



Rank k case...

$\nabla f(x) = 0 \Rightarrow \nabla f(X) = 0$

- $\|x\|_\infty \leq 4\alpha \Rightarrow \max_i \|X_i\| \leq 4\alpha$
- $\|zz^\top x - xx^\top x - \gamma \nabla R(x)\|_F \leq O(\epsilon) \Rightarrow$ $\|ZZ^\top X - XX^\top X - \gamma \nabla R(X)\|_\leq O(\epsilon)$.

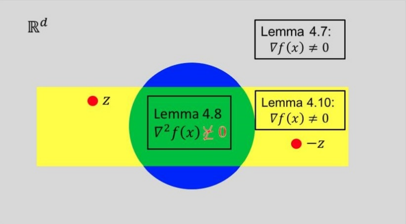


Rank k case...

$\nabla f(x) = 0 \Rightarrow \nabla f(X) = 0$

- $\|x\|_\infty \leq 4\alpha \Rightarrow \max_i \|X_i\| \leq 4\alpha$
- $\|zz^\top x - xx^\top x - \gamma \nabla R(x)\|_F \leq O(\epsilon) \Rightarrow \|ZZ^\top X - XX^\top X - \gamma \nabla R(X)\|_{\leq} O(\epsilon)$.

$\nabla^2 f(x) \succeq 0 \Rightarrow \nabla^2 f(X) \succeq 0$



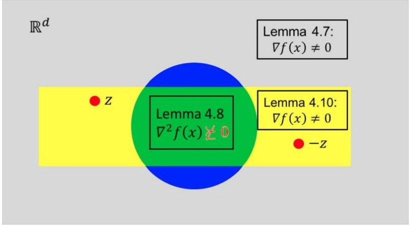
Rank k case...

$\nabla f(x) = 0 \Rightarrow \nabla f(X) = 0$

- $\|x\|_\infty \leq 4\alpha \Rightarrow \max_i \|X_i\| \leq 4\alpha$
- $\|zz^\top x - xx^\top x - \gamma \nabla R(x)\|_F \leq O(\epsilon) \Rightarrow$ $\|ZZ^\top X - XX^\top X - \gamma \nabla R(X)\|_{\leq} O(\epsilon)$.

$\nabla^2 f(x) \succeq 0 \Rightarrow \nabla^2 f(X) \succeq 0$

- $\|x\|^2 \geq \frac{1}{8} \Rightarrow \sigma_{\min}(X) \geq \frac{1}{4}\sigma_{\min}(Z)$



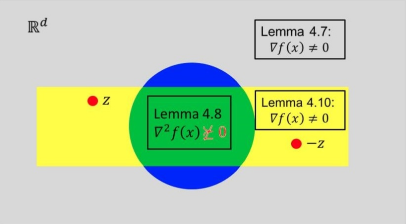
Rank k case...

$\nabla f(x) = 0 \Rightarrow \nabla f(X) = 0$

- $\|x\|_\infty \leq 4\alpha \Rightarrow \max_i \|X_i\| \leq 4\alpha$
- $\|zz^\top x - xx^\top x - \gamma \nabla R(x)\|_F \leq O(\epsilon) \Rightarrow$ $\|ZZ^\top X - XX^\top X - \gamma \nabla R(X)\|_\leq O(\epsilon)$.

$\nabla^2 f(x) \succeq 0 \Rightarrow \nabla^2 f(X) \succeq 0$

- $\|x\|^2 \geq \frac{1}{8} \Rightarrow \sigma_{\min}(X) \geq \frac{1}{4}\sigma_{\min}(Z)$



Combining them together:

- $\|xx^\top - zz^\top\|_F^2 \leq O(\epsilon) \Rightarrow \|XX^\top - ZZ^\top\|_F^2 \leq O(\epsilon)$

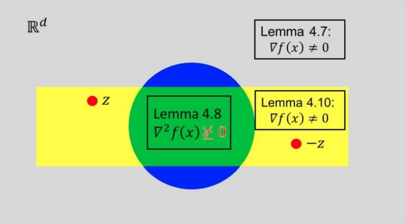
Rank k case...

$\nabla f(x) = 0 \Rightarrow \nabla f(X) = 0$

- $\|x\|_\infty \leq 4\alpha \Rightarrow \max_i \|X_i\| \leq 4\alpha$
- $\|zz^\top x - xx^\top x - \gamma \nabla R(x)\|_F \leq O(\epsilon) \Rightarrow$ $\|ZZ^\top X - XX^\top X - \gamma \nabla R(X)\|_{\leq} O(\epsilon)$.

$\nabla^2 f(x) \succeq 0 \Rightarrow \nabla^2 f(X) \succeq 0$

- $\|x\|^2 \geq \frac{1}{8} \Rightarrow \sigma_{\min}(X) \geq \frac{1}{4}\sigma_{\min}(Z)$



Combining them together:

- $\|xx^\top - zz^\top\|_F^2 \leq O(\epsilon) \Rightarrow \|XX^\top - ZZ^\top\|_F^2 \leq O(\epsilon)$

Handy Lemma:

If $\|XX^\top - ZZ^\top\|_F^2 \leq O(\epsilon)$, $XX^\top = ZZ^\top = M$. [Sun and Luo 2015]

The lemma from Luo and Sun

Theorem (Locally one point strongly convex)

Suppose $p \geq O(\frac{\log d}{d})$, *whp over* Ω, *for any point* X *in* $B_\epsilon = \{X \in \mathbb{R}^{d \times r} : \|XX^\top - ZZ^\top\|_F \leq \epsilon\}$, *there exists* U *such that* $UU^\top = ZZ^\top$, *and*

$$\langle -\nabla f(X), U - X \rangle \geq \frac{p}{4} \|M - XX^\top\|_F^2$$

The lemma from Luo and Sun

Theorem (Locally one point strongly convex)

Suppose $p \geq O(\frac{\log d}{d})$, *whp over* Ω, *for any point* X *in* $B_\epsilon = \{X \in \mathbb{R}^{d\times r} : \|XX^\top - ZZ^\top\|_F \leq \epsilon\}$, *there exists* U *such that* $UU^\top = ZZ^\top$, *and*

$$\langle -\nabla f(X), U - X\rangle \geq \frac{p}{4}\|M - XX^\top\|_F^2$$

- One point strongly convex with respect to U.

The lemma from Luo and Sun

Theorem (Locally one point strongly convex)

Suppose $p \geq O(\frac{\log d}{d})$*, whp over* Ω*, for any point* X *in* $B_\epsilon = \{X \in \mathbb{R}^{d\times r} : \|XX^\top - ZZ^\top\|_F \leq \epsilon\}$*, there exists* U *such that* $UU^\top = ZZ^\top$*, and*

$$\langle -\nabla f(X), U - X\rangle \geq \frac{p}{4}\|M - XX^\top\|_F^2$$

- One point strongly convex with respect to U.
  - Slightly different, here it's using function value $\|M - XX^\top\|_F^2$.

The lemma from Luo and Sun

Theorem (Locally one point strongly convex)

Suppose $p \geq O(\frac{\log d}{d})$*, whp over* Ω*, for any point* X *in* $B_\epsilon = \{X \in \mathbb{R}^{d\times r} : \|XX^\top - ZZ^\top\|_F \leq \epsilon\}$*, there exists* U *such that* $UU^\top = ZZ^\top$*, and*

$$\langle -\nabla f(X), U - X\rangle \geq \frac{p}{4}\|M - XX^\top\|_F^2$$

- One point strongly convex with respect to U.
  - Slightly different, here it's using function value $\|M - XX^\top\|_F^2$.
  - Instead of $\|x_t - x^*\|_2^2$.

The lemma from Luo and Sun

Theorem (Locally one point strongly convex)

Suppose $p \geq O(\frac{\log d}{d})$, whp over Ω, for any point X in $B_\epsilon = \{X \in \mathbb{R}^{d\times r} : \|XX^\top - ZZ^\top\|_F \leq \epsilon\}$, there exists U such that $UU^\top = ZZ^\top$, and

$$\langle -\nabla f(X), U - X\rangle \geq \frac{p}{4}\|M - XX^\top\|_F^2$$

- One point strongly convex with respect to U.
  - Slightly different, here it's using function value $\|M - XX^\top\|_F^2$.
  - Instead of $\|x_t - x^*\|_2^2$.
- If $\nabla f(X) = 0$, we have

$$0 = \langle -\nabla f(X), U - X\rangle \geq \frac{p}{4}\|M - XX^\top\|_F^2$$

So $XX^\top = M$.

Proof step 1

Theorem

If $\|XX^\top - M\|_F = \epsilon$, *there exists* U *such that* $UU^\top = M$ *and* $\|X - U\|_F \leq O(\epsilon\sqrt{r})$

Proof step 1

Theorem

If $\|XX^\top - M\|_F = \epsilon$*, there exists* U *such that* $UU^\top = M$ *and* $\|X - U\|_F \leq O(\epsilon\sqrt{r})$

- WLOG, $M = \mathrm{Diag}\{\sigma_1(Z)^2, \sigma_2(Z)^2, \cdots, \sigma_r(Z)^2, 0, \cdots 0\}$.

Proof step 1

Theorem

If $\|XX^\top - M\|_F = \epsilon$*, there exists* U *such that* $UU^\top = M$ *and* $\|X - U\|_F \leq O(\epsilon\sqrt{r})$

- WLOG, $M = \mathrm{Diag}\{\sigma_1(Z)^2, \sigma_2(Z)^2, \cdots, \sigma_r(Z)^2, 0, \cdots 0\}$.
- Let $M' = \mathrm{Diag}\{\sigma_1(Z)^2, \sigma_2(Z)^2, \cdots, \sigma_r(Z)^2\}$.

Proof step 1

Theorem

If $\|XX^\top - M\|_F = \epsilon$*, there exists* U *such that* $UU^\top = M$ *and* $\|X - U\|_F \leq O(\epsilon\sqrt{r})$

- WLOG, $M = \mathrm{Diag}\{\sigma_1(Z)^2, \sigma_2(Z)^2, \cdots, \sigma_r(Z)^2, 0, \cdots 0\}$.
- Let $M' = \mathrm{Diag}\{\sigma_1(Z)^2, \sigma_2(Z)^2, \cdots, \sigma_r(Z)^2\}$.
- $X = [V, W]$, $V \in r \times r$.

Proof step 1

Theorem

If $\|XX^\top - M\|_F = \epsilon$*, there exists* U *such that* $UU^\top = M$ *and* $\|X - U\|_F \le O(\epsilon\sqrt{r})$

- WLOG, $M = \mathrm{Diag}\{\sigma_1(Z)^2, \sigma_2(Z)^2, \cdots, \sigma_r(Z)^2, 0, \cdots 0\}$.
- Let $M' = \mathrm{Diag}\{\sigma_1(Z)^2, \sigma_2(Z)^2, \cdots, \sigma_r(Z)^2\}$.
- $X = [V, W]$, $V \in r \times r$.
- Let $U = [P, 0]$, $P = V(V^\top (M')^{-1} V)^{-1/2} \in r \times r$.

Proof step 1

Theorem

If $\|XX^\top - M\|_F = \epsilon$, *there exists* U *such that* $UU^\top = M$ *and* $\|X - U\|_F \leq O(\epsilon\sqrt{r})$

- WLOG, $M = \mathrm{Diag}\{\sigma_1(Z)^2, \sigma_2(Z)^2, \cdots, \sigma_r(Z)^2, 0, \cdots 0\}$.
- Let $M' = \mathrm{Diag}\{\sigma_1(Z)^2, \sigma_2(Z)^2, \cdots, \sigma_r(Z)^2\}$.
- $X = [V, W]$, $V \in r \times r$.
- Let $U = [P, 0]$, $P = V(V^\top (M')^{-1} V)^{-1/2} \in r \times r$.
- $PP^\top = V(V^\top (M')^{-1} V)^{-1} V^\top = M'$

Proof step 2

- $\langle -\nabla f(X), U - X\rangle =$
  $\langle -\nabla \left(\frac{1}{2}\|P_\Omega(M - XX^\top)\|_F^2\right) - \nabla R(x), U - X\rangle$

Proof step 2

- $\langle -\nabla f(X), U - X \rangle = \langle -\nabla \left(\frac{1}{2} \| P_\Omega (M - XX^\top) \|_F^2 \right) - \nabla R(x), U - X \rangle$
- $R(x)$ is helping us: $\langle -\nabla R(x), U - X \rangle \geq 0$.

Proof step 2

- $\langle -\nabla f(X), U - X\rangle = \langle -\nabla\left(\frac{1}{2}\|P_\Omega(M - XX^\top)\|_F^2\right) - \nabla R(x), U - X\rangle$
- $R(x)$ is helping us: $\langle -\nabla R(x), U - X\rangle \geq 0$.

$$\langle -\nabla\left(\frac{1}{2}\|P_\Omega(M - XX^\top)\|_F^2\right), U - X\rangle = 2\langle P_\Omega(XX^\top - M)X, X - U\rangle$$

$$= \langle P_\Omega(XX^\top - M), (X - U)X^\top + X(X - U)^\top\rangle$$

Proof step 2

- $\langle -\nabla f(X), U-X\rangle = \langle -\nabla\left(\frac{1}{2}\|P_\Omega(M-XX^\top)\|_F^2\right) - \nabla R(x), U-X\rangle$
- $R(x)$ is helping us: $\langle -\nabla R(x), U-X\rangle \geq 0$.

$$\langle -\nabla\left(\frac{1}{2}\|P_\Omega(M-XX^\top)\|_F^2\right), U-X\rangle = 2\langle P_\Omega(XX^\top - M)X, X-U\rangle$$
$$= \langle P_\Omega(XX^\top - M), (X-U)X^\top + X(X-U)^\top\rangle$$

- $a = U(X-U)^\top + (X-U)U^\top, b = (U-X)(U-X)^\top$

Proof step 2

- $\langle -\nabla f(X), U - X \rangle =$
  $\langle -\nabla \left(\frac{1}{2}\|P_\Omega(M - XX^\top)\|_F^2\right) - \nabla R(x), U - X \rangle$
- $R(x)$ is helping us: $\langle -\nabla R(x), U - X \rangle \geq 0$.

$$\langle -\nabla \left(\frac{1}{2}\|P_\Omega(M - XX^\top)\|_F^2\right), U - X \rangle = 2\langle P_\Omega(XX^\top - M)X, X - U \rangle$$

$$= \langle P_\Omega(XX^\top - M), (X - U)X^\top + X(X - U)^\top \rangle$$

- $a = U(X - U)^\top + (X - U)U^\top, b = (U - X)(U - X)^\top$
- $XX^\top - M = a + b$, $(X - U)X^\top + X(X - U)^\top = a + 2b$.

Proof step 2

- $\langle -\nabla f(X), U - X\rangle =$
  $\langle -\nabla\left(\frac{1}{2}\|P_\Omega(M - XX^\top)\|_F^2\right) - \nabla R(x), U - X\rangle$
- $R(x)$ is helping us: $\langle -\nabla R(x), U - X\rangle \geq 0$.

$$\langle -\nabla\left(\frac{1}{2}\|P_\Omega(M - XX^\top)\|_F^2\right), U - X\rangle = 2\langle P_\Omega(XX^\top - M)X, X - U\rangle$$

$$= \langle P_\Omega(XX^\top - M), (X - U)X^\top + X(X - U)^\top\rangle$$

- $a = U(X - U)^\top + (X - U)U^\top, b = (U - X)(U - X)^\top$
- $XX^\top - M = a + b$, $(X - U)X^\top + X(X - U)^\top = a + 2b$.

Intuitively, a is first order, b is second order.

$$\langle P_\Omega(XX^\top - M), (X - U)X^\top + X(X - U)^\top\rangle$$

$$\geq \|P_\Omega(a)\|_F^2 + \|P_\Omega(b)\|_F^2 - 3\|P_\Omega(a)\|_F\|P_\Omega(b)\|_F \geq \frac{p}{4}\|M - XX^\top\|_F^2$$