Chapter 10: Matrix Estimation with rank constraints:

1. $$\hat{\Theta}_{RR} = \underset{\substack{\Theta \in \mathbb{R}^{p \times r} \\ \text{rank}(\Theta) \le r}}{\text{argmin}} \left\{ \frac{1}{2n} |||Y - Z\Theta|||_F^2 \right\}$$ "Reduced rank regression estimate"

$$\hat{\Sigma}_{ZZ} = \frac{1}{n} Z^T Z$$ "Sample Covariance Matrix"

$$\hat{\Sigma}_{ZY} = \frac{1}{n} Z^T Y$$ "Sample cross Covariance Matrix"

$$\hat{\Theta}_{RR} = \underset{\substack{\Theta \in \mathbb{R}^{p \times r} \\ \text{rank}(\Theta) \le r}}{\text{argmin}} \left\{ \frac{1}{2n} |||Y - Z\Theta|||_F^2 \right\}$$

$$= \underset{\text{rank}(\theta) \le r}{\text{argmin}} \left\{ Z \cdot |Y - Z\Theta| \right\}$$

With all the craft, argmin about rank requires a smaller matrix.

Method #1:

$$= \underset{\text{rank}(\theta) \le r}{\text{argmin}} \left\{ Z |Y - Z\Theta| \right\}$$

$$= \left(Z \cdot |\underset{(T \times T)}{Y} - Z \Theta_{RR}| \right) \circ \underset{(T \times r)}{V}$$

$$= Z^T | \underset{(T \times T)}{Y} \circ \underset{(T \times r)}{V} - Z \circ \Theta_{RR} \circ V |$$

$$= Z^T | \underset{(T \times r)}{X} - Z \cdot \Theta_{RRR} \cdot V |$$

$$= 0$$

$$\Theta_{RRR} = \frac{Z^T X}{Z^T Z} \circ V^T = \frac{\hat{\Sigma}_{ZX}}{\hat{\Sigma}_{ZZ}} \circ V^T$$

Method #2:

$$= \underset{\text{rank}(\theta) \le r}{\text{argmin}} \left\{ Z |Y - Z\Theta| = 0 \right\}$$

$$= \underset{\text{rank}(\theta) \le r}{\text{argmin}} \left\{ \Theta^* = \frac{ZY}{ZZ} \right\}$$

$$= \Theta^* \cdot V^T \cdot V^T$$

$$\boxed{V^T V = \sum_{i=1}^{T} \sum_{j=1}^{r} V_{ij} \circ V_{ij} = 1}$$

$$= \frac{Z \underset{(T \times T)}{Y}}{ZZ} \underset{(T \times r)}{V^T} \cdot V^T$$

$$= \frac{Z \underset{(T \times r)}{X} V^T}{ZZ}$$

$$= \frac{\hat{\Sigma}_{ZX}}{\hat{\Sigma}_{ZZ}} V^T$$

10.2. From Example 10.5,

$$z^{t+1} = \theta^* z^t \theta^* + w^t \text{ for } t=1,\dots N-1$$ "Vector Autoregressive Process"

$$\Sigma^t = cov(z^t)$$ "covariance"

$$\Sigma^{t+1} = \theta^* \cdot \Sigma^t \cdot (\theta^*)^T + T$$ "Autoregressive covariance"

a) Proof by Induction:

Base Case: $\{z^1\} = N(0, \Sigma^1) = \frac{1}{\sqrt{2\pi\Sigma^1}} e^{\frac{-x^2}{2\Sigma^1}}$

$= \text{constant}$

Next Step: $\{z^2\} = N(0, \Sigma^2) = \frac{1}{\sqrt{2\pi\Sigma^2}} e^{\frac{-x^2}{2\Sigma^2}}$

$= \frac{1}{\sqrt{2\pi(\theta^*\Sigma^1\theta^* + T)}} e^{\frac{-x^2}{2(\theta^*\Sigma^1\theta^* + T)}}$

$= \text{constant}$

Inductive Step: $\{z^t\} = N(0, \Sigma^t) = \frac{1}{\sqrt{2\pi\Sigma^t}} e^{\frac{-x^2}{2\Sigma^t}}$

$= \frac{1}{\sqrt{2\pi(\theta^*\Sigma^{t-1}\theta^* + T)}} e^{\frac{-x^2}{2(\theta^*\Sigma^{t-1}\theta^* + T)}}$

b) $cov(z^{t+1}) = \Sigma^{t+1}$

$= \theta^*\Sigma^t\theta^* + T$

$\Sigma^{t+1} = \theta^*\Sigma^t\theta^* + T$ when stationary Σ^{t+1} equals Σ^t

$\Sigma^t = \theta^*\Sigma^t\theta^* + T$

$\Sigma^t = \frac{T}{1-\theta^*\theta^*}$, since no covariance is negative $0 < \theta^*\theta^* < 1$

$0 < \|\theta^*\| < 1$

$|||\theta^*||| < 1$

Suppose $|||\hat{\Theta}|||_{NUC} \leq |||\Theta^*|||_{NUC}$ ← "from the problem"

$|||\hat{\Theta}|||_{NUC} \leq |||\Theta^*_M|||_{NUC}$ if $\hat{\Delta} = \hat{\Theta} - \Theta^*_M$

$\leq |||\Theta^*_M + \hat{\Delta}|||_{NUC}$ "Decomposability"

$\leq |||\Theta^*_M + \hat{\Delta}_{\bar{M}^\perp} + \hat{\Delta}_M|||_{NUC}$

$\leq |||\Theta^*_M + \hat{\Delta}_{\bar{M}^\perp}||| - |||\hat{\Delta}_M|||_{NUC}$

$\leq |||\Theta^*_M||| + |||\hat{\Delta}_{\hat{M}^\perp}||| - |||\hat{\Delta}_M|||_{NUC}$

$|||\hat{\Theta}|||_{NUC} - |||\Theta^*_M||| \leq 0$ "from line one"

$\leq |||\hat{\Delta}_{M^\perp}||| - |||\hat{\Delta}_M|||_{NUC}$

$|||\Delta_{M^\perp}||| \leq |||\Delta_M|||_{NUC}$ ... "with absolute symbols"

(Equation 9.22) The cost function satisfies a ϕ^*-norm curvature

$$\phi^*(\nabla\mathcal{L}(\Theta^*+\Delta) - \nabla\mathcal{L}(\Theta^*)) \geq \underbrace{\kappa\,\phi^*(\Delta)}_{\text{"curvature"}} - \underbrace{\tau\,\phi(\Delta)}_{\text{"tolerance"}}$$

10.5 (Equation 9.36) $\mathcal{E}_n(\Delta) := \mathcal{L}_n(\Theta^*+\Delta) - \mathcal{L}_n(\Theta^*) - \langle \nabla\mathcal{L}_n(\Theta^*), \Delta \rangle$

$$\phi^*(\nabla\mathcal{L}(\Theta^*+\Delta) - \nabla\mathcal{L}(\Theta^*)) = \phi^*(\mathcal{E}_n(\Delta) + \langle \nabla\mathcal{L}_n(\Theta^*), \Delta \rangle)$$

$$\geq \phi^*(\mathcal{E}_n(\Delta)) + \phi^*(\langle \nabla\mathcal{L}(\Theta^*), \Delta \rangle)$$

$$\geq \kappa\,\phi^*(\Delta) - \tau\,\phi(\Delta)$$

(Equation 10.20) $|||\frac{1}{n}\mathcal{X}_n^*\mathcal{X}_n(n)|||_2 \geq \kappa|||\Delta|||_2 - \tau|||\Delta|||_{nuc}$

The equation above derives similarly to Equation 9.36 when $\nabla\mathcal{L}(\Theta) = \frac{1}{n}\sum_{i=1}^{n} X_i^T(y_i - \langle\langle X_i, \Theta \rangle\rangle)$

$$= \frac{1}{n}\mathcal{X}_n^*(y - \mathcal{X}_n(\Theta))$$

c) In part a, $T > 0$

In part b, $\Sigma = \frac{T}{1-\theta^* \theta^*}$ where $|\theta^* \theta^*| < 1$

For every case, $\frac{T}{1-\theta^* \theta^*} = \frac{\text{positive}}{\text{positive}} = \text{positive} = \Sigma$

10.3 (Equation 10.3.7) $\theta^{bad} = e_1 \otimes e_1 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$

$$\mathbb{P}[X(\theta^{bad}) = 0] = \mathbb{P}[\lambda X(\theta^{bad}) = 0]$$

$$:= \mathbb{P}[e^{\lambda(1 + \mathbb{E}[X(\theta^{bad})] + \frac{\lambda^2 \mathbb{E}[X(\theta^{bad})]}{2} + \cdots} = e^0]$$

$$\cong \mathbb{P}[e^{\lambda^2 \mathbb{E}[\lambda(\theta^{bad})]} = e^0]$$

$$\underset{\lambda}{\operatorname{argmin}} \left\{ \frac{\lambda^2 \mathbb{E}[X(\theta^{bad})]}{2} - 0 \right\} = 0$$

$$\lambda^* = 0$$

$$\mathbb{P}[X(\theta^{bad}) = 0] := \mathbb{P}[e^{\frac{\lambda^{*2} \mathbb{E}[X(\theta^{bad})]}{2}} = e^0]$$

$$= e^0$$

$$= 1$$

10.4. (Equation 10.14) $A \in \mathbb{M}(U, V)$ $\quad B \in \mathbb{M}^{\perp}(U, V)$

$$A = U \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & 0 \end{bmatrix} V^T \qquad B = U \begin{bmatrix} 0 & T_{12} \\ T_{21} & T_{22} \end{bmatrix} V^T$$

b) The problem describes a $\mathbb{M}$-space, from the book:

$$\mathbb{M}(\mathbb{U},\mathbb{V}) := \{\Theta \in \mathbb{R}^{r_1 \times r_2} \mid \text{rowspan}(\Theta) \subseteq \mathbb{V}, \text{colspan} \subseteq \mathbb{U}\}$$

$$\bar{\mathbb{M}}^{\perp}(\mathbb{U},\mathbb{V}) := \{\Theta \in \mathbb{R}^{r_1 \times r_2} \mid \text{colspan}(\Theta) \subseteq \mathbb{V}, \text{colspan} \subseteq \mathbb{U}\}$$

where $A_{\mathbb{M}} = \begin{bmatrix} T_{11} & 0 \\ 0 & 0 \end{bmatrix}$ and $B_{\bar{\mathbb{M}}^{\perp}} = \begin{bmatrix} 0 & 0 \\ 0 & T_{22} \end{bmatrix}$

In this $\mathbb{M}$-space, a Frobenius Norm is the l-2 Norm.

$$|||\hat{\Theta} - \Theta|||_F = \sqrt{\sum_{i}^{r_1} \sum_{j=i}^{r_2} |\hat{\Theta} - \Theta|^2_{ij}} = |||\hat{\Theta} - \Theta|||_2$$

Outside the $\mathbb{M}$space, a similar coefficient appears, as:

$$|||\hat{\Theta} - \Theta|||_F = \sqrt{\sum_{i=1}^{r_1} \sum_{j=i}^{r_2} |\hat{\Theta} - \Theta|^2_{ij}} \leq \sqrt{\sum_{i}^{r_1} |\hat{\Theta} - \Theta|^2_{\mathbb{M}}} + \sqrt{\sum_{j}^{r_2} |\hat{\Theta} - \Theta|^2_{\bar{\mathbb{M}}^{\perp}}}$$

$$\leq 2 \cdot \sqrt{\sum_{i}^{r_1} |\hat{\Theta} - \Theta|^2}$$

$$\leq 2 \cdot \sqrt{2r \, |\hat{\Theta} - \Theta|^2}$$

$$\leq 2\sqrt{2r} \sqrt{|\hat{\Theta} - \Theta|^2}$$

$$\leq 2\sqrt{2r} \, |||\hat{\Theta} - \Theta|||_2$$

A matrix test occurred because the questions:

$$|||\hat{\Theta} - \Theta|||_F = \sqrt{\sum_{i=1}^{r_1} \sum_{j=i}^{r_2} |\hat{\Theta} - \Theta|^2_{ij}} = \sqrt{\sum_{i=1}^{r_1} \sum_{j=i}^{r_2} |A|^2_{ij}}$$

$$= \sqrt{\sum_{i=1}^{r_1} \sum_{j=i}^{r_2} \left| \begin{bmatrix} T_{11} & 0 \\ 0 & 0 \end{bmatrix} \right|^2}$$

$$\leq \sqrt{\sum_{i=1}^{r_1} \left| \begin{bmatrix} T_{11} & 0 \\ 0 & 0 \end{bmatrix}_{\mathbb{M}} \right|^2} + \sqrt{\sum_{j=1}^{r_2} \left| \begin{bmatrix} T_{11} & 0 \\ 0 & 0 \end{bmatrix}_{\bar{\mathbb{M}}^{\perp}} \right|^2}$$

$$\leq \sqrt{2 \cdot r \cdot T_{11}^2} + \sqrt{2 \cdot r \cdot T_{11}^2}$$

$$\leq 2 \cdot \sqrt{2r} \sqrt{T_{11}^2} = 2\sqrt{2r} \, |||\hat{\Theta} - \Theta|||_2$$

<u>Note:</u> Part c is not in the document because an introspective derivation in part b.

(Theorem 10.8) "Lower bound for norms"

$$\frac{\|X_n(\Delta)\|_2^2}{n} \geq c_1 \|\sqrt{\Sigma}\, vec(\Delta)\|_2^2 - c_2 \rho^2(\Sigma) \left\{\frac{d_1+d_2}{n}\right\} |||\Delta|||_{Nuc}^2 \quad \forall \Delta \in \mathbb{R}^{d_1 \times d_2}$$

10.6. If $\Sigma = I_{dd}$ where $D = d_1 d_2$

and $B(t) := \{\Delta \in \mathbb{R}^{d_1 \times d_2} \mid |||\Delta|||_F = 1, |||\Delta|||_{Nuc} \leq t\}$

$$\inf_{\Delta \in B(t)} \sqrt{\frac{1}{n}\sum \langle\langle X_i, \Delta\rangle\rangle^2} = \inf_{\Delta \in B(t)} \sqrt{\frac{\|X_n(\Delta)\|^2}{n}}$$

$$\geq \sqrt{c_1 \|\sqrt{\Sigma}\, vec(\Delta)\|_2^2 - c_2 \rho^2(\Sigma)\left\{\frac{d_1+d_2}{n}\right\} |||\Delta|||_{Nuc}^2}$$

$$\geq \sqrt{c_1 \|\sqrt{\Sigma}\, vec(\Delta)\|_2^2} - \sqrt{c_2 \rho^2(\Sigma)\left\{\frac{d_1+d_2}{n}\right\} |||\Delta|||_{Nuc}^2}$$

The chapter defines $\rho^2(\Sigma) = \sup var(\langle\langle X, uv^T\rangle\rangle)$

$= 1$ when $\Sigma = I_d$

Also, a reference to later chapters helps:

(Theorem 14.12) "Euclidean norm bounds"

$$\|f\|_n^2 \geq \frac{1}{2}\|f\|_2^2$$

$$\inf_{\Delta \in B(t)} \sqrt{\frac{1}{n}\sum \langle\langle X_i, \Delta\rangle\rangle^2} = \sqrt{c_1 \|\sqrt{\Sigma}\, vec(\Delta)\|_2^2} - \sqrt{c_2 \rho^2(\Sigma)\left\{\frac{d_1+d_2}{n}\right\}} |||\Delta|||_{Nuc}$$

$$= \sqrt{c_1} \cdot \underbrace{(1)}_{\text{"}\sqrt{\Sigma}=\sqrt{I}\text{"}} \cdot \underbrace{\left(\frac{1}{2} - \delta\right)}_{\text{"Theorem 14.12"}} - \sqrt{c_2} \cdot \underbrace{(1)}_{\text{"}\rho^2(\Sigma)=1\text{"}} \cdot \left\{\sqrt{\frac{d_1}{n}} + \sqrt{\frac{d_2}{n}}\right\} \cdot \underbrace{t}_{\text{"set } B(t)\text{"}}$$

$$= \frac{1}{2} - \delta - 2\left(\sqrt{\frac{d_1}{n}} + \sqrt{\frac{d_2}{n}}\right) \cdot t$$

when $\sqrt{c_1} = 1, \sqrt{c_2} = 2$

(Equation 10.26)

$$\mathbb{B}_q(R_q) := \{\Theta \in \mathbb{R}^{d_1 \times d_2} \mid \sum_{j=1}^{d} \sigma(\Theta)^q \leq R\}.$$

Represents all values

that belong to

Real numbers

such that

this property holds for all numbers in the set

10.7. From page 322-323, $|||\hat{\Theta}|||_{nuc} \geq |||\Theta^*_{M}||| + |||\hat{\Delta}_{M^\perp}||| - |||\hat{\Delta}_{\bar{M}}|||_{nuc}$

a)

$$|||\hat{\Delta}_{\bar{M}^\perp}|||_{nuc} \leq |||\hat{\Delta}_{\bar{M}}|||_{nuc} + |||\hat{\Theta}|||_{nuc} - |||\Theta^*_{M}|||$$

$$\leq 2\sqrt{2r}\,|||\hat{\Delta}|||_F + \underbrace{2|||\Theta|||}_{\text{From the set}}$$

$$\leq 2\sqrt{2r}\,|||\hat{\Delta}|||_F + 2\sum_{j=r+1}^{d} \sigma_j(\Theta)$$

b) $$\frac{||\mathfrak{X}_n(\Delta)||_2^2}{n} = |||\Delta|||_F^2 \geq c_1 ||\sqrt{\Sigma}\, vec(\Delta)||_2^2 - c_2 \rho^2(\Sigma)\left\{\frac{d_1+d_2}{n}\right\} |||\Delta|||_{nuc}^2$$

The problem gave two relationships: $T_1(r) = \sigma\sqrt{\frac{d}{n}} \sum_{j=r+1}^{d} \sigma_j(\Theta^*)$

$$|||\tfrac{1}{n}\sum_i w_i X_i|||_2 \leq \sigma\sqrt{\frac{d}{n}}$$

$$\frac{|||\mathfrak{X}_n(\Delta)|||_2^2}{n} = |||\Delta|||_F^2 \geq c_1 ||\sqrt{\Sigma}\, vec(\Delta)||_2^2 - c_2 \rho^2(\Sigma)\left\{\frac{d_1+d_2}{n}\right\} |||\Delta|||_{nuc}^2$$

from Equation 10.24; $|||\hat{\Delta}|||_{nuc} \leq 2\sqrt{2r}\,|||\hat{\Delta}|||_F$

$$\geq c_1 \gamma_{min}(\Sigma)|||\hat{\Delta}|||_F - 8c_2\rho^2(\Sigma)\frac{r(d_1+d_2)}{n}|||\hat{\Delta}|||_F$$

$$\geq c_1 \sum_{j=r+1}^{d} \sigma_j(\Theta^*)|||\tfrac{1}{n}\sum_i w_i X_i|||_2^2 - 8c_2 |||\hat{\Delta}|||_F\, |||\tfrac{1}{n}\sum_i w_i X_i|||_2$$

$$\geq c_1 \cdot T_1^2(r) + \sigma\sqrt{\frac{rd}{n}}\,|||\Delta|||_F \quad \text{when } c_1 = 1, c_2 = -1/8$$

$$\geq \max\{T_1(r), T_1^2(r)\} + \sigma\sqrt{\frac{rd}{n}}\,|||\hat{\Delta}|||_F$$

c) A similar coefficient comes from an r around one for both the terms on the right-side.

10.8. Corollary 10.8:

$$|||\hat{\theta}-\theta^*|||_2 \leq 30\sqrt{2}\,\frac{\sigma\sqrt{\gamma_{max}(\hat{\Sigma})}}{\gamma_{min}(\hat{\Sigma})}\left(\sqrt{\frac{p+T}{n}}+\delta\right)$$

$$\mathbb{P}\left[|||\tfrac{1}{n}Z^TW|||_2 \geq 5\sigma\sqrt{\gamma_{max}(\hat{\Sigma})}\left(\sqrt{\frac{d+T}{n}}+\delta\right)\right]$$

$$\simeq \mathbb{P}\left[|||\tfrac{1}{n}Z^TW|||_2 \geq 2\delta\right] \simeq \mathbb{P}\left[|||Z^TW|||_2 \geq \sqrt{n}\,2\delta\right]$$

$$= \mathbb{P}\left[\lambda|||Z^TW|||_2 \geq \lambda\cdot\sqrt{n}\cdot 2\delta\right]$$

$$\simeq \mathbb{P}\left[e^{1+\lambda\mathbb{E}[|||Z^TW|||_2]+\frac{\lambda^2\mathbb{E}[|||Z^TW|||_2^2]^2}{2}+\cdots} \geq e^{\lambda\cdot\sqrt{n}\,2\delta}\right]$$

$$\leq \mathbb{P}\left[e^{\frac{\lambda^2\mathbb{E}[|||Z^TW|||^2]^2}{2}} \geq e^{\lambda\sqrt{n}\,2\delta}\right]$$

$$\underset{\lambda}{\operatorname{argmin}}\left\{\frac{\lambda^2\mathbb{E}[|||Z^TW|||_2^2]^2}{2}-\lambda\cdot\sqrt{n}\cdot 2\delta\right\}=0$$

$$\lambda^* = \frac{\sqrt{n}\cdot 2\delta}{\mathbb{E}[|||Z^TW|||]^2}$$

$$\mathbb{P}\left[|||\tfrac{1}{n}Z^TW|||_2 \geq 5\sigma\sqrt{\gamma_{max}(\hat{\Sigma})}\left(\sqrt{\frac{d+T}{n}}+\delta\right)\right] \leq \mathbb{P}\left[e^{\frac{\lambda^{*2}\cdot\mathbb{E}[|||Z^TW|||_2^2]^2}{2}} \geq e^{\lambda^*\sqrt{n}\,2\delta}\right]$$

$$\leq 2\cdot\mathbb{P}\left[e^{\frac{4\cdot n\delta^2}{2}} \geq e^{4\cdot n\delta}\right]$$

$$\leq 2\cdot e^{-2n\delta^2}$$

10.9.

a) $F_\theta(X) = \langle\langle X, \Theta\rangle\rangle$ for a random matrix $X = x \otimes x$ with $x \sim N(0, I_n)$

A decomposition of $f_\theta(X)$ has a similar distribution with $f_D(X)$, in the case $\Theta = U^TDU$. With a nonsingular matrix U, the eigenvalues in Θ and D coincide, along with distributions.

b) $\mathbb{E}[f_\Theta^2(X)] = \mathbb{E}[\langle\langle X, \Theta \rangle\rangle]$

$= \mathbb{E}\left[\sum X \circ \Theta\right]$

$= \mathbb{E}\left[\sum X \otimes X\, \Theta\right]$

$= |||\sum \Theta^2||| + |||\Theta|||^2$

$= |||\Theta|||_F^2 + 2(\text{trace}(\Theta))^2$

Gaussian Moments:

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} X^2 \circ \Theta(X)\, dX$$

$$= \mu^2 + \sigma^2$$

10.10 (Corollary 10.18):

$$|||\hat{\Theta} - \Theta^*|||_F^2 \leq c_1 \max\{\sigma^2, \alpha^2\} r \left\{\frac{d \log d}{n} + \delta^2\right\}$$

where $|||\Theta^*|||_{max} \leq \alpha / \sqrt{d_1 d_2}$

$\delta \in (0, \sigma^2/2b)$ in the situation Bernstein's condition with parameters (σ, b).

a) $|||\hat{\Theta}|||_{NUC} \leq |||\Theta^*|||_{NUC}$

$\leq |||\Theta_M^*|||_{NUC}$ if $\hat{\Delta} = \hat{\Theta} - \Theta_M^*$

$\leq |||\Theta_M^* + \hat{\Delta}|||_{NUC}$ "Decomposability"

$\leq |||\Theta_M^* + \hat{\Delta}_{\bar{M}^\perp} + \hat{\Delta}_M|||$

$\leq |||\Theta_M^* + \hat{\Delta}_{\bar{M}^\perp}||| - |||\hat{\Delta}_M|||_{NUC}$

$\leq |||\Theta_M^*||| + |||\hat{\Delta}_{\bar{M}^\perp}||| - |||\hat{\Delta}_M|||_{NUC}$

$|||\hat{\Delta}_{\hat{M}^\perp}|||_{NUC} \leq |||\hat{\Delta}_M|||_{nuc} + \underbrace{|||\hat{\Theta}|||_{nuc} - |||\Theta_M^*|||}$

$\leq 2\sqrt{2r}\, |||\hat{\Delta}|||_F$ at max $= 0$

$$\frac{|||\hat{\Delta}|||_{NUC}}{|||\hat{\Delta}|||_F} \leq 2\sqrt{r}$$

The second portion derives from $\tilde{y}_i = \Theta + \frac{w_i}{\sqrt{d_1 d_2}}$ (Equation 10.6)

$\Theta^* = |||\Theta|||_{max} + \frac{w_i}{\sqrt{d_1 d_2}}$

$$|||\hat{\Delta}|||_{max} = \frac{w_i}{\sqrt{d_1 d_2}} \leftarrow \text{"Gaussian Noise"}$$

$$= \frac{\alpha}{\sqrt{d_1 d_2}} \leftarrow \text{"spikiness in Equation 10.4"}$$

$$|||\hat{\Delta}||_{max} \leq 2 \circ \frac{\alpha}{\sqrt{d_1 d_2}}$$

$$\leq 2 \circ \alpha$$

Another derivation, from bibliography:

$$\underbrace{||\hat{\Delta}||_{max} \leq ||\Delta||_{max} + ||\Delta^*||_{max}}_{\text{triangle inequality}}$$

$$\leq \frac{\alpha}{\sqrt{d_1 d_2}} + \frac{\alpha}{\sqrt{d_1 d_2}}$$

$$\leq \frac{2\alpha}{\sqrt{d_1 d_2}}$$

b) From Equation 10.44, "Strong convexity" condition:

$$\frac{||\mathfrak{X}_n(\hat{\Delta})||_2^2}{n} \geq |||\hat{\Delta}|||_F^2 - 8\sqrt{2}\, c_1 \alpha \sqrt{\frac{rd\log d}{n}}\, |||\hat{\Delta}|||_F - 4c_2\alpha^2\left(\sqrt{\frac{d\log d}{n}} + \delta\right)^2$$

$$\geq |||\hat{\Delta}|||_F \underbrace{\left\{|||\hat{\Delta}|||_F - 8\sqrt{2}\, c_1 \cdot \alpha\sqrt{\frac{rd\log d}{n}}\right\}}_{\geq \frac{|||\hat{\Delta}|||_F}{2}} - \underbrace{8c_2\alpha^2\left(\frac{d\log d}{n} + \delta^2\right)}_{< \frac{|||\hat{\Delta}|||_F^2}{4}}$$

$$\geq \frac{1}{2}|||\hat{\Delta}|||_F - \frac{1}{4}|||\hat{\Delta}|||_F^2$$

$$\geq \frac{1}{4}|||\hat{\Delta}|||_F^2$$

10.11.

a) $$\hat{\Theta} = \arg\min\left\{\frac{1}{2}|||\Theta|||_F^2 - \langle\langle \Theta, \frac{1}{n}\sum y_i X_i\rangle\rangle + \lambda|||\Theta|||_{nuc}\right\}$$

$$= \arg\min\left\{\max \frac{1}{2}\left(|||\Theta|||_F - \langle\Theta, \frac{1}{n}\sum_{i=1}^{n} y_i X\rangle\right)^2 + \lambda|||\Theta|||_{nuc}\right\}$$

$$= \operatorname{argmin}\left\{ \max \tfrac{1}{2}\left(|||\Theta|||_F - \langle \Theta, M\rangle\right)^2 + \lambda |||\Theta|||_{nuc} \right\}$$

$$= \operatorname{argmin}\left\{ U \circ D \circ V^T + \lambda\, U \circ D \circ V^T \right\}$$

$$= \operatorname{argmin}\left\{ U \circ (D - D_\lambda) \circ V^T \right\}$$

$$= U \circ T_\lambda(D) \circ V^T \qquad \text{where } T_\lambda(D) \begin{cases} D & D - D_\lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

b) From Equation 10.16, $\hat{\Theta} \in \operatorname{argmin}\left\{ \frac{1}{2} \|y - \mathfrak{X}(\Theta)\|_2^2 + \lambda_n |||\Theta|||_{nuc} \right\}$

$$\in \operatorname{argmin}\left\{ \max \tfrac{1}{2}\left(|||\Theta|||_F - \langle \Theta, M\rangle\right)^2 + \lambda |||\Theta|||_{nuc} \right\}$$

Proposition 10.6
when Equation 10.16 and 10.17 are true.

$$\frac{|||\hat{\Theta} - \Theta^*|||_F^2}{2n} \le \frac{9}{2}\frac{\lambda_n^2}{\kappa^2} r + \frac{1}{\kappa}\left\{ 2\lambda_n \sum_{j=r+1}^{d} \sigma(\Theta^*) + \frac{32 c_0 (d_1 + d_2)}{n}\left[\sum_{j=r+1}^{d} \sigma(\Theta^*)\right]^2 \right\}$$

when $\sum_{j=r+1}^{d} \sigma(\Theta^*) = 0$

$$\frac{|||\hat{\Theta} - \Theta^*|||_F^2}{2n} \le \frac{9}{2}\frac{\lambda_n^2}{\kappa^2} r$$

$$\frac{|||\hat{\Theta} - \Theta^*|||_F}{2n} \le \frac{3}{\sqrt{2}}\frac{\lambda}{\kappa^2} r$$

c) $|||\hat{\Theta} - \Theta^*|||_F \le \frac{3}{\sqrt{2}} r \lambda_n$

$$\le \frac{3}{\sqrt{2}} r \left(2 \max_{|||\Theta|||_{nuc} \le 1} \left| \frac{1}{n}\sum_{i=1}^{n} \langle\langle U, X_i \rangle\rangle \langle\langle X_i, \Theta^* \rangle\rangle - \langle\langle U, \Theta^* \rangle\rangle \right| + 2 \left|\left|\left| \frac{1}{n}\sum w_i X_i \right|\right|\right|_2 \right)$$

$$\le 3\sqrt{2}\, r \left(\nabla \mathcal{L}(\Theta^* + \Delta) - \nabla \mathcal{L}(\Theta^*) \right)$$