LECTURE 6

- RANK-1 MATRIX APPROXIMATION THROUGH RANK-1 PCA

CONSIDER THE PROBLEM:

$$\min_{\substack{x \in \mathbb{R}^m \\ w \in \mathbb{R}^n}} \|M - xw^T\|_F^2 , \quad M \in \mathbb{R}^{m \times n} \qquad (*)$$

HOW IS THIS RELATED TO MATRIX SENSING?

$$\|M - xw^T\|_F^2 = \|\text{vec}(M) - \text{vec}(xw^T)\|_2^2$$

$$= \|y - \mathcal{A}(X)\|_2^2$$ WHERE $y \equiv \text{vec}(M)$ (FULL OBSERVATION SE

$\mathcal{A} \equiv I$ (PERFECT ISOMETRY)

$X \equiv xw^T$ (RANK-1)

BY SVD OF M: $M = \sum_{i=1}^{\min\{m,n\}} \sigma_i \cdot u_i v_i^T$, $\|u_i\|_2 = \|v_i\|_2 = 1.$

$\sigma_1 > \sigma_2 \geq \dots \geq 0.$

EXPAND THE OBJECTIVE:

$$\|M - xw^T\|_F^2 = \|M\|_F^2 - 2x^T M w + \|xw^T\|_F^2$$

$$= \|M\|_F^2 - 2x^T M w + \|x\|_2^2 \cdot \|w\|_2^2$$

THUS, (*) BECOMES:

$$\min_{\substack{x \in \mathbb{R}^m \\ w \in \mathbb{R}^n}} -2x^T M w + \|x\|_2^2 \|w\|_2^2$$

LET $f(x) = \min_{w \in \mathbb{R}^n} \left(\underbrace{-2x^T M w}_{\text{linear}} + \underbrace{\|x\|_2^2 \|w\|_2^2}_{\text{quadratic}} \right) \xrightarrow{\text{CONVEX}} \nabla_w = 0 \Rightarrow -2M^T x + 2\|x\|_2^2 \cdot w =$

$$\Rightarrow w = \frac{M^T x}{\|x\|_2^2}$$

THEN: $$f(x) = -\frac{2x^T M M^T x}{\|x\|_2^2} + \|x\|_2^2 \cdot \frac{x^T M M^T x}{\|x\|_2^4}$$

$$= -\frac{x^T M M^T x}{\|x\|_2^2}$$

THUS, THE ORIGINAL PROBLEM IS EQUIVALENT TO: ANY INTUITION WHY THIS IS NON-CONVEX?

$$\min_{x \in \mathbb{R}^m} f(x) := -\frac{x^T M M^T x}{\|x\|_2^2} \longrightarrow \text{PCA OBJECTIVE! (FOR RANK-1)}$$

REMARK 1: LENGTH OF x DOES NOT MATTER - ONLY ITS DIRECTION.

(TO SEE THIS, DEFINE $y = \frac{x}{\|x\|_2^2}$. THEN

$$\min_{x \in \mathbb{R}^m} f(x) \equiv \min_{\substack{y \in \mathbb{R}^m \\ \|y\|_2 = 1}} -y^T M M^T y$$
)

VIA THE INNER PRODUCT EXPRESSION:

$$\langle x, u_1 \rangle = \cos(\theta) \cdot \|u_1\| \cdot \|x\|_2$$

SINCE θ DEPENDS ON x, AND $\|u_1\|_2 = 1$, WE HAVE:

$$\theta(x) = \cos^{-1}\left(\frac{1}{\|x\|_2} \langle x, u_1 \rangle \right)$$

GRADIEN DESCENT ON $f(x)$:

$$x_{t+1} = x_t - \eta \cdot \nabla f(x_t)$$

WHERE: $$\nabla f(x_t) = \frac{1}{\|x\|_2^4} \cdot \left[\nabla_x \left(-x^T M M^T x \right) \cdot \|x\|_2^2 + x^T M M^T x \cdot \nabla_x \|x\|_2^2 \right]$$

(QUOTIENT RULE)

$$= \frac{1}{\|x\|_2^4} \left[-2 \|x\|_2^2 \cdot M M^T x + 2 (x^T M M^T x) \cdot x \right]$$

$$= \frac{2}{\|x\|_2^4} \left[(x^T M M^T x) \cdot x - \|x\|_2^2 \cdot M M^T x \right]$$

KEY OBSERVATION FOR GD ON PCA IS THAT:

IF $\langle x_t, u_1 \rangle = 0$, THEN $\langle x_{t+1}, u_1 \rangle = 0$.

TO SEE THIS:

$$\langle x_{t+1}, u_1 \rangle = \langle x_t - \eta \nabla f(x_t), u_1 \rangle = -\eta \langle \nabla f(x_t), u_1 \rangle$$

THEN:

$$\langle \nabla f(x_t), u_1 \rangle = \frac{2}{\|x_t\|_2^4} \left[(x_t^T M M^T x_t) \overset{0}{\cancel{x_t^T u_1}} - \|x_t\|_2^2 \, x_t^T M M^T u_1 \right]$$

$$= -\frac{2}{\|x_t\|_2^4} \cdot \|x_t\|_2^2 \; x_t^T M M^T u_\perp = -\frac{2}{\|x_t\|_2^2} \cdot x_t^T \cdot \sum_i \sigma_i^2 u_i u_i^T u_\perp$$

$$= -\frac{2}{\|x_t\|_2^2} \sigma_\perp^2 x_t^T u_\perp = 0.$$

IN WORDS : i) IF x_t IS ORTHOGONAL TO $u_\perp$, x_{t+1} IS ALSO ORTHOGONAL (NO IMPROVEMENT)

ii) THIS FURTHER MEANS THAT IF WE START FROM A POINT SUCH THAT $\langle x_0, u_\perp \rangle = 0$, WE FAIL TO RECOVER $u_\perp$.

iii) HOWEWER, MAYBE THERE IS HOPE STARTING FROM A POINT NOT ORTHOGONAL TO $u_\perp$. (TO SEE THIS, A RANDOMLY SELECTED $x_0 \in \mathbb{R}^m$ ALMOST SURELY HAS NON-ZERO COMPONENT ON THE SPAN OF $u_\perp$)

LET'S STUDY THE BEHAVIOR OF THE POTENTIAL FUNCTION:

$$\Psi_{t+1} = 1 - \frac{\langle x_{t+1}, u_\perp \rangle^2}{\|x_{t+1}\|_2^2}$$

INTUITION: IF $\Psi_{t+1} \to 0$, $\frac{\langle x_{t+1}, u_\perp \rangle^2}{\|x_{t+1}\|_2^2} \to 1$. WHICH THE OPTIMAL THING TO ACHIEVE FOR NORMALIZED VECTORS

WE HAVE THE FOLLOWING:

$$\|x_{t+1}\|_2^2 = \|x_t - \eta \nabla f(x_t)\|_2^2 = \|x_t\|_2^2 - 2\eta \, x_t^T \nabla f(x_t) + \eta^2 \cdot \|\nabla f(x_t)\|$$

(OBSERVE THAT : $x_t^T \nabla f(x_t) = \frac{2}{\|x_t\|_2^4} \left((x_t^T M M^T x_t) \cdot \|x_t\|_2^2 - \|x_t\|_2^2 (x_t^T M M^T x_t) \right) = 0.$)

$$= \|x_t\|_2^2 + \eta^2 \cdot \|\nabla f(x_t)\|_2^2$$

THEN: $\Psi_{t+1} = 1 - \frac{\langle x_{t+1}, u_\perp \rangle^2}{\|x_{t+1}\|_2^2}$

$$= 1 - \frac{\langle x_t, u_\perp \rangle^2 - 2\eta \cdot \langle x_t, u_\perp \rangle \langle \nabla f(x_t), u_\perp \rangle + \eta^2 \cdot \langle \nabla f(x_t), u_\perp \rangle^2}{\|x_t\|_2^2 + \eta^2 \cdot \|\nabla f(x_t)\|_2^2}$$

OBSERVE THAT:

$$\frac{\langle x_{t+1}, u_1\rangle^2}{\|x_{t+1}\|_2^2} = \langle \frac{x_{t+1}}{\|x_{t+1}\|_2}, u_1\rangle^2 = \cos^2(\theta(x_{t+1}))$$

ALSO: $1 - \cos^2(\theta(x_{t+1})) = \sin^2(\theta(x_{t+1}))$.

THUS:

$$\sin^2(\theta(x_{t+1})) = \frac{\|x_t\|_2^2 + \eta^2 \cdot \|\nabla f(x_t)\|_2^2 - \langle x_t, u_1\rangle^2 + 2\eta \langle x_t, u_1\rangle\langle \nabla f(x_t), u_1\rangle - \eta^2 \langle \nabla f(x_t), u_1\rangle}{\|x_t\|_2^2 + \eta^2 \cdot \|\nabla f(x_t)\|_2^2}$$

$$\leq \frac{1}{\|x_t\|_2^2} \cdot \left(\|x_t\|_2^2 + \eta^2 \cdot \|\nabla f(x_t)\|_2^2 - \langle x_t, u_1\rangle^2 + 2\eta \langle x_t, u_1\rangle\langle \nabla f(x_t), u_1\rangle - \eta^2 \langle \nabla f(x_t), u_1\rangle \right.$$

$$\leq \sin^2(\theta(x_t)) + \frac{\eta^2}{\|x_t\|_2^2} \cdot \|\nabla f(x_t)\|_2^2 + \frac{2\eta}{\|x_t\|_2^2} \cdot \langle x_t, u_1\rangle \langle \nabla f(x_t), u_1\rangle$$

- FOR THE INNER PRODUCT: $\langle \nabla f(x_t), u_1\rangle$ WE HAVE:

$$\langle \nabla f(x_t), u_1\rangle \leq -\frac{2}{\|x_t\|_2} (\sigma_1^2 - \sigma_2^2) \cdot \sin^2\theta(x_t) \cdot \cos\theta(x_t) \leq 0.$$

(PROVE IT!)

- FOR $\|\nabla f(x_t)\|_2^2$ WE HAVE:

$$\|\nabla f(x_t)\|_2^2 \leq \frac{4}{\|x_t\|_2^2} (\sigma_1^4 + \sigma_2^4) \cdot \sin^2\theta(x_t)$$

THEN:

$$\sin^2(\theta(x_{t+1})) \leq \sin^2(\theta(x_t)) + \frac{4\eta^2}{\|x_t\|_2^4} (\sigma_1^4 + \sigma_2^4) \cdot \sin^2(\theta(x_t))$$

$$- \frac{4\eta}{\|x_t\|_2^4} (\sigma_1^2 - \sigma_2^2) \cdot \sin^2(\theta(x_t)) \cdot \langle x_t, u_1\rangle^2$$

$$= \sin^2(\theta(x_t)) \left(1 + \frac{4\eta^2}{\|x_t\|_2^4} (\sigma_1^4 + \sigma_2^4) - \frac{4\eta}{\|x_t\|_2^2} \cdot (\sigma_1^2 - \sigma_2^2) \cdot \frac{\langle x_t, u_1\rangle^2}{\|x_t\|_2^2} \right)$$

WE WILL PROVIDE LOCAL CONVERGENCE GUARANTEES: GIVEN A PROPER INITIALIZATION, WE GET CONVERGENCE TO GLOBAL MINIMUM.

IN PARTICULAR, IF $\left\langle \frac{x_t}{\|x_t\|_2}, u_1 \right\rangle^2 \geq c$ (I.E., $x_t \not\perp u_1$), $0 \leq c < 1$

WE OBTAIN:

$$\sin^2\theta(x_{t+1}) \leq \sin^2\theta(x_t) \underbrace{\left(1 + \frac{4\eta^2}{\|x_t\|_2^4}(\sigma_1^4 + \sigma_2^4) - \frac{4\eta}{\|x_t\|_2^2}(\sigma_1^2 - \sigma_2^2)\cdot c\right)}_{\text{WE REQUIRE TO BE } < 1.}$$

$/\!/ \; p$

SELECT $\eta = \frac{c}{2} \cdot \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^4 + \sigma_2^4} \cdot \|x_t\|_2^2$. THEN:

$$p = 1 + c^2 \cdot \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^4 + \sigma_2^4} - 2c^2 \cdot \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^4 + \sigma_2^4} = 1 - c^2 \frac{(\sigma_1^2 - \sigma_2^2)^2}{(\sigma_1^4 + \sigma_1^4)} < 1.$$

$< 1.$

$0 \leq c < 1.$

THUS: $\sin^2\theta(x_{t+1}) \leq p \cdot \sin^2\theta(x_t)$

LINEAR CONVERGENCE! $O\left(\log \frac{1}{\varepsilon}\right)$

- PROOF FOR $\min_{U \in \mathbb{R}^{n \times r}} f(UU^T)$ WHERE f IS CONVEX, SMOOTH AND STR. CONVEX AND r IS THE RANK OF X^*.

THE ALGORITHM WE USE IS: $U_{t+1} = U_t - \eta \cdot \nabla f(U_t U_t^T) \cdot U_t$

WE DEFINE: $\text{DIST}(U, U^*) = \min_{R \in O_r} \|U - U^* R\|_F$, WHERE $X^* = U^* U^{*T}$

(REMEMBER, THERE CAN BE INFINITE U^*)

THEN, WE HAVE:

$$\text{DIST}(U_{t+1}, U^*)^2 = \min_{R \in O_r} \|U_{t+1} - U^* R\|_F^2 \leq \|U_{t+1} - U^* R_t\|_F^2$$

WHERE $R_t = \min \|U_t - U^* R\|$

$$= \|U_{t+1} - U_t + U_t - U^* R_t\|_F^2$$

$$= \|U_{t+1} - U_t\|_F^2 + \|U_t - U^* R_t\|_F^2 + 2\langle U_{t+1} - U_t, U_t - U^* R_t\rangle$$

$$= \eta^2 \|\nabla f(U_tU_t^T)U_t\|_F^2 + \|U_t - U^*R_t\|_F^2$$
$$- 2\eta \langle \nabla f(U_tU_t^T)U_t, U_t - U^*R_t \rangle$$

DOES THIS RING A BELL?

KEY RESULT IS THE FACT THAT WE CAN PROVE A REGULATORY CONDITION:

$$\langle \nabla f(UU^T)U, U - U^*R \rangle \geq \frac{2}{3}\eta \cdot \|\nabla f(UU^T)U\|_F^2 + \frac{3\mu}{20} \sigma_r(X^*) \cdot \text{DIST}(U_t, U^*)^2$$

THUS:

$$\text{DIST}(U_{t+1}, U^*)^2 \leq \text{DIST}(U_t - U^*R_t)^2 + \eta^2 \cdot \|\nabla f(U_tU_t^T)U_t\|_F^2$$
$$- \frac{4}{3}\eta^2 \|\nabla f(U_tU_t^T)U_t\|_F^2 - \frac{6\mu\eta}{20} \sigma_r(X^*) \cdot \text{DIST}(U_t, U^*)^2$$
$$\leq \left(1 - \frac{3\mu\eta}{10} \sigma_r(X^*)\right) \cdot \text{DIST}(U_t, U^*)^2$$

THIS DEFINES THE STEP SIZE η

(IN PRACTICE, THE PAPER "DROPPING CONVEXITY FOR FASTER SEMIDEFINITE OPTIMIZATION" HAS MORE SOPHISTICATED BUT MORE PRACTICAL η)

HOWEVER, IN ORDER TO PROVE THE REGULATORY CONDITION, WE REQUIRE

$$\text{DIST}(U_t, U^*) \leq \rho \cdot \sigma_r(X^*)^{1/2} \quad \text{FOR ALL } t.$$

WHICH MEANS $\text{DIST}(U_0, U^*) \leq \rho \cdot \sigma_r(X^*)^{1/2}$ $\longrightarrow$ GOOD INITIALIZATION.