MTH 995-003: Intro to CS and Big Data Spring 2014

Lecture 3 — Jan 14, 2014

Inst. Mark Iwen *Scribe: Islam Badreldin*

Contents

1 Overview

In the last lecture an outline and motivation for the course was given. In this lecture we review Singular Value Decomposition (SVD) and its perturbation bounds. Additionally, we introduce the Semi-definite Programming (SDP) and the topic of Convexity.

2 Singular Value Decomposition (SVD)

Theorem 1. *Let* $\mathbf{A} \in \mathbb{C}^{m\times N}$. *Then* $\mathbf{A}$ *has a "unique" SVD:*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* \tag{1}$$

where

i) $\mathbf{U} \in \mathbb{C}^{m\times m}$ *is* unitary, *i.e.* $\mathbf{U}^{-1} = \mathbf{U}^*$

ii) $\mathbf{V} \in \mathbb{C}^{N\times N}$ *is* unitary *iii)* $\mathbf{\Sigma} \in \mathbb{R}^{m \times N}$ *is diagonal, i.e.*

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_q \end{pmatrix}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_q$ *and* $q = \min(m, N)$.

The "unique" diagonal entries $\sigma_1, \sigma_2, \cdots, \sigma_q$ *are called the singular values of* $\mathbf{A}$.

Proof: The proof follows from the next 2 lemmas. □

Before presenting the 2 lemmas, we begin by some definitions.

Definition 1. *Let* $\{\mathbf{w}_1, \cdots, \mathbf{w}_N\}$ *be an orthonormal basis for* $\mathbb{C}^N$. *Define* $s_j = ||\mathbf{A}\mathbf{w}_j||$, *and*

$$\mathbf{h}_j = \begin{cases} \mathbf{0} & \text{if } s_j = 0 \\ \frac{1}{s_j}\mathbf{A}\mathbf{w}_j \in \mathbb{C}^m & \text{if } s_j \neq 0 \end{cases}$$

and define $\mathbf{W}$ *to be the unitary matrix* $(\mathbf{w}_1 \cdots \mathbf{w}_N) \in \mathbb{C}^{N \times N}$.

Lemma 1.

$$\mathbf{A} = (\mathbf{h}_1 \cdots \mathbf{h}_N) \begin{pmatrix} s_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_N \end{pmatrix} \mathbf{W}^*$$

Proof:

$$\mathbf{A}\mathbf{W} = \begin{pmatrix} s_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_N \end{pmatrix} (\mathbf{h}_1 \cdots \mathbf{h}_N)$$

and $\mathbf{W}^{-1} = \mathbf{W}^*$. □

Lemma 2. *Let* $\mathbf{A} \in \mathbb{C}^{m \times N}$. *Let* $\mathbf{w}_1, \cdots, \mathbf{w}_N$ *be the eigenvectors of* $\mathbf{A}^*\mathbf{A}$ *(which is a Hermitian). Then* $\langle \mathbf{h}_j, \mathbf{h}_l \rangle = 0$ *if* $j \neq l$.

Proof:

$$\begin{aligned} \langle \mathbf{A}\mathbf{w}_j, \mathbf{A}\mathbf{w}_l \rangle &= (\mathbf{A}\mathbf{w}_j)^* \mathbf{A}\mathbf{w}_l \\ &= \mathbf{w}_j^* \mathbf{A}^* \mathbf{A}\mathbf{w}_l \\ &= \mathbf{w}_j^* (\lambda_l \mathbf{w}_l) \\ &= 0 \quad \text{if } j \neq l. \end{aligned}$$

□

2.1 Calculation of the SVD

Notice that

$$\begin{aligned}\mathbf{A}^*\mathbf{A} &= (\mathbf{U\Sigma V}^*)^*(\mathbf{U\Sigma V}^*)\\ &= \mathbf{V\Sigma}^*\mathbf{U}^*\mathbf{U\Sigma V}^*\\ &= \mathbf{V\Sigma}^2\mathbf{V}^*.\end{aligned}$$

Then, $\mathbf{V}$ contains the eigenvectors of $\mathbf{A}^*\mathbf{A}$ as columns, and $\sigma_1, \sigma_2, \cdots, \sigma_q$ are the squared eigenvalues of $\mathbf{A}^*\mathbf{A}$.

Numerically, we can use, e.g., the QR algorithm to find the eigenvalues of $\mathbf{A}^*\mathbf{A}$ to get the singular values of $\mathbf{A}$. The shifted inverse power method, e.g., can be used to calculate $\mathbf{V}$. Similarly, from $\mathbf{AA}^*$ we can find $\mathbf{U}$.

3 Perturbation Bounds

Theorem 2. *Weyl's Bounds*
Let $\mathbf{A}, \mathbf{E} \in \mathbb{C}^{m\times N}$ and $q = \min(m, N)$. Then the following inequalities hold (Notice that the singular values are assumed to be ordered)

a) $\sigma_{j+i-1}(\mathbf{A} + \mathbf{E}) \leq \sigma_i(A) + \sigma_j(E)$

b) $\sigma_{j+i-1}(\mathbf{AE}^*) \leq \sigma_i(A)\sigma_j(E)$

for $1 \leq i, j \leq q$ *such that* $i + j \leq q + 1$.

Proof: For a proof, see [HJ94]. □

There are perturbation results for the singular vectors as well. (See Stewart's notes.)

3.1 Homework Problems

Problem 1 Choose scribe dates.

Problem 2 Prove using Theorem 2 that

$$|\sigma_i(\mathbf{A} + \mathbf{E}) - \sigma_i(A)| \leq \sigma_1(E).$$

Problem 3 Suppose every entry of $\mathbf{A} \in \mathbb{C}^{m\times N}$ is corrupted with an additive error of magnitude $\leq \varepsilon$. How much is $\sigma_1(\mathbf{A})$ changed in terms of ε?

4 Semi-definite Programming (SDP)

Standard Form (See [VB'96].)
Given $\mathbf{c} \in \mathbb{R}^m$ and $\mathbf{F}_0, \ldots, \mathbf{F}_m \in \mathbb{S}^N$ that are fixed. (Note: $\mathbb{S}^N$ is the space of real symmetric $N \times N$ matrices.)

And given variables $\mathbf{x} \in \mathbb{R}^m$. Minimize $\mathbf{c}^{\mathrm{T}}\mathbf{x}$ such that

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}_0 + \sum_{j=1}^{m} x_j \mathbf{F}_j \geq 0$$

which means the matrix $\mathbf{F}(\mathbf{x})$ is positive semi-definite.

Notes

- $\mathbf{F}(\mathbf{x}) \in \mathbb{S}^N$ and so $\mathbf{F}(\mathbf{x})$ has N eigenvalues (with possible repetitions).
- $\mathbf{F}(\mathbf{x}) \geq 0$ tells us that we want all the N eigenvalues to remain *non-negative* (within machine precision).
- $\mathbf{y}^{\mathrm{T}}\mathbf{F}(\mathbf{x})\mathbf{y} \geq 0 \quad \forall \mathbf{y} \in \mathbb{R}^N$.
- SDPs can be solved computationally in *polynomial time* in mN. (See "Interior Point Methods".)
- SDPs are a subset of the more general convex optimization problems.

5 Convexity

(See [BV04] and Appendix B of [FR13]).

5.1 Definitions

Definition 2. *A function $\mathbf{f} : \mathbb{R}^D \to \mathbb{R}^d$ is **convex** if*

$$\mathbf{f}(\alpha\mathbf{x} + \beta\mathbf{y}) \leq^{(*)} \alpha\mathbf{f}(\mathbf{x}) + \beta\mathbf{f}(\mathbf{y})$$

$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^D$ and $\forall \alpha, \beta \in (0, 1)$ such that $\alpha + \beta = 1$.

$^{(*)}$ The inequality is coordinate-wise.

- $f(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x}$ is convex since it is linear.

Definition 3. *A function $\mathbf{F} : \mathbb{R}^m \to \mathbb{R}^{N \times N}$ is **convex** if*

$$\mathbf{F}(\alpha\mathbf{x} + \beta\mathbf{y}) \leq^{(*)} \alpha\mathbf{F}(\mathbf{x}) + \beta\mathbf{F}(\mathbf{y})$$

$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ and $\forall \alpha, \beta \in \mathbb{R}^+$ such that $\alpha + \beta = 1$.

$^{(*)}$ For matrices, $\mathbf{A} \leq \mathbf{B}$ if $\mathbf{B} - \mathbf{A}$ is positive semi-definite.

- $\mathbf{F}(\mathbf{x}) = \mathbf{F}_0 + \sum_{j=1}^{m} x_j \mathbf{F}_j$ is convex since it is linear.

Definition 4. *A set* $K \subseteq \mathbb{R}^m$ *is* ***convex*** *if*

$$(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \in K$$

$\forall \mathbf{x}, \mathbf{y} \in K$ *and* $\forall \alpha \in (0,1)$.

- The set $K = \{\mathbf{x}|\mathbf{F}(\mathbf{x}) \geq 0\}$ used in the SDP constraint is a convex set.

Definition 5. $K \subseteq \mathbb{R}^m$ *is a* ***convex cone*** *if it is both* ***convex*** *and a* ***cone****.*

Definition 6. $K \subseteq \mathbb{R}^m$ *is a* ***cone*** *if* $\alpha\mathbf{x} \in K, \quad \forall \mathbf{x} \in K$ *and* $\forall \alpha \in \mathbb{R}^+$.

5.2 Homework Problems

Problem 4 Show that $K = \{\mathbf{x}|\mathbf{F}(\mathbf{x}) = \mathbf{F}_0 + \sum_{j=1}^m x_j \mathbf{F}_j\}$ (used in the SDP constraint) is convex.

References

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing.* Birkhäuser Basel, 2013.

[HJ94] Roger A. Horn and Charles R. Johnson. *Topics in matrix analysis.* Cambridge University Press, 1994.