Nuclear Norm stuff

William J. Maxwell

Portland State University, 2016

ABSTRACT

some cool stuff about the nuclear norm

Nuclear Norm stuff

William J. Maxwell

B.S. Computer Science

501 Project

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

at

Portland State University

2016

APPROVAL PAGE

Master of Science 501 Project

Nuclear Norm stuff

Presented by

William J. Maxwell, ,

Advisor ____
Mau Nam Nguyen

Second reader ____

Portland State University

2016

Contents

Matrix Decompositions

Consider a real matrix $A \in \mathbb{R}^{m\times n}$. For any such matrix there are four fundamental subspaces associated with it. These are $null(A) \subset \mathbb{R}^n$, $null(A^T) \subset \mathbb{R}^m$, $col(A) \subset \mathbb{R}^m$, and $col(A^T) \subset \mathbb{R}^n$. It can easily be shown that $null(A) \perp col(A^T)$ and $null(A^T) \perp col(A)$ and furthermore are orthogonal complements of each other.

Lemma 0.0.1. *$null(A)$ and $col(A^T)$ are orthogonal complements in $\mathbb{R}^n$.*

Proof. Let $v \in null(A)$ and $u \in col(A^T)$. Then $\langle u, v\rangle = \langle A^T b, v\rangle$ for some $b \in \mathbb{R}^m$, so $\langle A^T b, v\rangle = (A^T b)^T v = b^T A v = b^T 0 = 0$. Thus, $null(A) \perp col(A^T)$. Since $dim(col(A^T)) = rank(A)$ the rank-nullity theorem tells us that the sum of the dimensions of $col(A^T)$ and $null(A)$ is n. So, every vector in $\mathbb{R}^n$ is in $null(A)$ or $col(A^T)$. Thus they are orthogonal complements. $\square$

Theorem 1. Let $A \in \mathbb{R}^{m\times n}$. Then there exists $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, $\Sigma \in \mathbb{R}^{m\times n}$ such that U, V are orthogonal and Σ is diagonal-like with $A = U\Sigma V^T$.

Proof.

Let $A \in \mathbb{R}^{m\times n}$ then define $L = A^T A \in R^{n\times n}$. Note that L is positive semi-definite since for $x \in \mathbb{R}^n$ we have $\langle Lx, x\rangle = \langle A^T Ax, x\rangle = \langle Ax, Ax\rangle = \|Ax\|^2 \geq 0$. Similarly, define $M = AA^T \in \mathbb{R}^{m\times m}$ which is also positive semi-definite.

Let $x \in null(A)$ then $Ax = 0 \implies A^T Ax = A^T 0 = 0 \implies Lx = 0 \implies x \in null(L)$.

Let $x \in null(L)$ then $Lx = 0 \implies \langle Lx, x\rangle = \langle 0, x\rangle = 0 \implies \langle A^T Ax, x\rangle = 0 \implies \langle Ax, Ax\rangle = 0 \implies \|Ax\|^2 = 0 \implies Ax = 0$, so $x \in null(A)$ and $null(A) = null(L)$.

Consequently, we have $range(L) = range(L^T) = null(L)^\perp = null(A)^\perp = range(A^T)$.

Similarly, $range(M) = range(A)$ and $null(A^T) = null(M)$.

Since $rank(A) = rank(A^T)$ we have $rank(L) = rank(A^T) = rank(A) = rank(M) = \rho$.

Now we need to find orthonormal bases for $null(A) = null(L)$, $null(A^T) = null(M)$, $range(A^T) = range(L)$, and $range(A) = range(M)$.

Since L is positive semi-definite there exists an orthonormal basis for $null(L) = null(A)$ given by $\phi_1, ..., \phi_{N_A}$ which are eigenvectors of L corresponding to the eigenvalue 0.

Since M is positive semi-definite there exists an orthonormal basis for $null(M) = null(A^T)$ given by $\psi_1, ..., \psi_{N_A}$ which are eigenvectors of M corresponding to the eigenvalue 0.

Since L is positive semi-definite there exist positive eigenvalues $\lambda_1 \geq ... \geq \lambda_\rho > 0$ and their eigenvectors $v_1, ..., v_\rho$ form an orthonormal basis for $range(L) = range(A^T)$.

We have an orthonormal basis of $range(L) = range(A^T)$ consisting of eigenvectors $v_1, ..., v_\rho$ corresponding to eigenvalues $\lambda_1 \geq ... \geq \lambda_\rho > 0$. So, $Lv_j = \lambda_j v_j$. Define $w_j = Av_j$. Then $\langle w_i, w_j \rangle = \langle Av_i, Av_j \rangle = \langle v_i, A^T Av_j \rangle = \langle v_i, Lv_j \rangle = \langle v_i, \lambda_j v_j \rangle$. This equals 0 when $i \neq j$ and λ_j when $i = j$. So, $w_1, ... w_\rho$ are orthogonal, but $\|w_j\|^2 = \lambda_j$. Define $u_j = \frac{w_j}{\|w_j\|} = \frac{w_j}{\sqrt{\lambda_j}}$ then $u_1, ..., u_\rho$ form an orthonormal basis for $range(M) = range(A)$. Denote $\sigma_j = \sqrt{\lambda_j}$ we call these the singular values of A. Note that $u_j = \frac{w_j}{\sigma_j} = \frac{Av_j}{\sigma_j} \implies Av_j = \sigma_j u_j$ and $A^T u_j = \frac{A^T Av_j}{\sigma_j} = \frac{Lv_j}{\sigma_j} = \frac{\lambda_j v_j}{\sigma_j} = \sigma_j v_j$. This is the fundamental relation of singular values.

Now that we have orthonormal bases, we can construct the matrices $U = [u_1 \, ... \, u_\rho \, \psi_1 \, ... \, \psi_{N_{A^T}}] \in \mathbb{R}^{m \times m}$ and $V = [v_1 \, ... \, v_\rho \, \phi_1 \, ... \, \phi_{N_A}] \in \mathbb{R}^{n \times n}$. Note that both U and V are orthogonal since they are constructed from orthonormal column vectors. Then we have $AV = [Av_1 \, ... \, Av_\rho \, A\phi_1 \, ... \, A\phi_{N_A}] = [\sigma_1 u_1 \, ... \, \sigma_\rho u_\rho \, 0 \, ... \, 0]$. It follows that

$$U^T AV = \begin{bmatrix} u_1^T \\ \vdots \\ u_\rho^T \\ \psi_1^T \\ \vdots \\ \psi_{N_{A^T}}^T \end{bmatrix} \begin{bmatrix} \sigma_1 u_1 & ... & \sigma_\rho u_\rho & 0 & ... & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 & ... & 0 & 0 & ... & 0 \\ 0 & \sigma_2 & ... & 0 & 0 & ... & 0 \\ \vdots & & & & & & \\ 0 & 0 & ... & \sigma_\rho & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & ... & 0 \\ \vdots & & & & & & \\ 0 & 0 & ... & 0 & 0 & ... & 0 \end{bmatrix} = \Sigma \tag{0.0.1}$$

And by the orthogonality of U^T and V we have $A = U\Sigma V^T$.

□

Theorem 2.

Let X be an $n \times n$ symmetric matrix. Then $X = U^T \lambda(X) U$ where $\lambda(X)$ is an $n \times n$ matrix containing the n real eigenvalues of X across the diagonal and U is an orthogonal matrix.

Proof.

Let $\lambda_1, ..., \lambda_n$ be the eigenvalues of X corresponding to the eigenvectors $U_1, ..., U_n$ and let $U = [U_1, ..., U_n]$. Then we have the following

$$\begin{aligned}
XU &= X[U_1, ..., U_n] \\
&= [XU_1, ..., XU_n] \\
&= [\lambda_1 U_1, ..., \lambda_n U_n] \\
&= \begin{bmatrix} \lambda_1 u_{11} & ... & \lambda_n u_{n1} \\ \lambda_1 u_{12} & ... & \lambda_n u_{n2} \\ \vdots & & \\ \lambda_1 u_{1n} & ... & \lambda_n u_{nn} \end{bmatrix} \\
&= \begin{bmatrix} u_{11} & ... & u_{n1} \\ u_{12} & ... & u_{n2} \\ \vdots & & \\ u_{1n} & ... & u_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ \vdots & & & \\ 0 & 0 & ... & \lambda_n \end{bmatrix} \\
&= U(\lambda(X))
\end{aligned}$$

Since U consists of eigenvectors it is an orthogonal matrix and we have $X = U(\lambda(X))U^T$ and since X is a symmetric matrix we have $X = U^T(\lambda(X))U$. $\square$

Fenchel Conjugate

Definition 1. For any function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ the **Fenchel conjugate** $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined by

$$
\begin{aligned}
f^*(v) &= \sup\{\langle v, x\rangle - f(x) \mid x \in \mathbb{R}^n\} \\
f^*(v) &= \sup\{\langle v, x\rangle - f(x) \mid x \in \mathrm{dom}(f)\}
\end{aligned}
$$

Note that the Fenchel conjugate is always convex as it is the supremum of a family of affine functions, and it is order reversing in the sense that $f(x) \leq g(x)$ for all $x \in \mathbb{R}^n$ implies that $f^*(v) \geq g^*(v)$ since for all $v \in \mathbb{R}^n$.

$$f^*(v) = \sup\{\langle v, x\rangle - f(x) \mid x \in \mathbb{R}^n\} \geq \sup\{\langle v, x\rangle - g(x) \mid x \in \mathbb{R}^n\} = g^*(v).$$

Lemma 0.0.2. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function with $\mathrm{dom}(f) \neq \emptyset$. Then we have $\langle v, x\rangle \leq f(x) + f^*(v)$ for all $x, v \in \mathbb{R}^n$.*

Proof. From the definition of the Fenchel conjugate it follows

$$
\begin{aligned}
f^*(v) &= \sup\{\langle v, x\rangle - f(x) \mid \forall x \in \mathbb{R}^n\} \\
f^*(v) &\geq \langle v, x\rangle - f(x) \\
f^*(v) + f(x) &\geq \langle v, x\rangle
\end{aligned}
$$

□

Theorem 3. For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and any $\overline{x} \in \mathrm{dom}(f)$, we have $v \in \partial f(\overline{x})$ if and only if $f(\overline{x}) + f^*(v) = \langle v, \overline{x}\rangle$.

Proof.

By the previous lemma we have $\langle v, \overline{x}\rangle \leq f(\overline{x}) + f^*(v)$. Since $v \in \partial f(\overline{x})$ we have for all

$x \in \mathbb{R}^n$

$$\begin{aligned} f(\overline{x}) + \langle v, x \rangle - f(x) &\leq \langle v, \overline{x} \rangle \\ f(\overline{x}) + \sup\{\langle v, x \rangle - f(x)\} &\leq \langle v, \overline{x} \rangle \\ f(\overline{x}) + f^*(v) &\leq \langle v, \overline{x} \rangle \end{aligned}$$

Conversely, we have

$$\begin{aligned} f(\overline{x}) + f^*(v) &= \langle v, \overline{x} \rangle \\ f(\overline{x}) + \sup\{\langle v, x - f(x) \mid \forall x \in \mathbb{R}^n\} &= \langle v, \overline{x} \rangle \\ f(\overline{x}) + \langle v, x \rangle - f(x) &\leq \langle v, \overline{x} \rangle \\ \langle v, x - \overline{x} \rangle &\leq f(x) - f(\overline{x}) \end{aligned}$$

Thus, $v \in \partial f(\overline{x})$. $\square$

Symmetric Functions

We define the function $[\cdot] : \mathbb{R}^n \to \mathbb{R}^n$ to map x to the vector consisting of the components of x in nonincreasing order. We call a function $f : \mathbb{R}^n \to \mathbb{R}^n$ symmetric if $f(x) = f([x])$. Note that for $x, y \in \mathbb{R}^n$ we have $x^T y \leq [x]^T [y]$ since $x^T y = x_1 y_1 + ... + x_n y_n$ and $[x]^T [y] = [x]_1 [y]_1 + ... + [x]_n [y]_n$ and for each $i = 1, ..., n$ we have $x_i \leq [x]_i$ and $y_i \leq [y]_i$.

Definition 2. The spectral function is defined as $\lambda : \mathbb{S}^n \to \mathbb{R}^n$ assigning $X \in \mathbb{S}^n$ to the vector $x \in \mathbb{R}^n$ such that $x = (\lambda_1(X), \lambda_2(X), ..., \lambda_n(X))$ where each $\lambda_i(X)$ is an eigenvalue of X and $\lambda_1(X) \geq \lambda_2(X) \geq ... \geq \lambda_n(X)$.

Definition 3. Two matrices $X, Y \in \mathbb{S}^n$ have a **simultaneous ordered spectral** decomposition if there exists a matrix $U \in \mathbb{O}^n$ with $X = U^T(\mathrm{Diag}(\lambda(X)))\mathrm{U}$ and $Y = U^T(\mathrm{Diag}(\lambda(Y)))\mathrm{U}$.

Theorem 4. Every double stochastic matrix can be represented as a convex combination of permutation matrices.

Proof.

It is sufficient to show that permutation matrices are the extreme points of $\lneqq^n$. Let P be a permutation matrix such that $P = \lambda Q + (1 - \lambda)R$ for $Q, R \in \lneqq^n$ and $\lambda \in (0, 1)$. Let p_{ij} be an entry in P. If $p_{ij} = 1$ then $q_{ij} = r_{ij} = 1$ and if $p_{ij} = 0$ then $q_{ij} = r_{ij} = 0$. It follows that $P = Q = R$ so the convex combination is trivial and P is an extreme point. $\square$

Theorem 5. Any matrices X and Y in $\mathbb{S}^n$ satisfy the inequality

$$\mathrm{tr(XY)} \leq \lambda(\mathrm{X})^{\mathrm{T}} \lambda(\mathrm{Y})$$

Equality holds if and only if X and Y have a simultaneous ordered spectral decomposition

Theorem 6. If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a symmetric function, then $(f \circ \lambda)^* = (f^* \circ \lambda)$.

Proof. By Fan's inequality for $Y \in \mathbb{S}^n$ it follows that

$$\begin{aligned}(f \circ \lambda)^*(Y) &= \sup_{X \in \mathbb{S}^n} \{\operatorname{tr}(XY) - f(\lambda(X))\} \\ &\leq \sup_{X \in \mathbb{S}^n} \{\lambda(X)^T \lambda(Y) - f(\lambda(X))\} \\ &\leq \sup_{x \in \mathbb{R}^n} \{x^T \lambda(Y) - f(x)\} \\ &= f^*(\lambda(Y))\end{aligned}$$

Then by using the spectral decomposition $Y = U^T(\operatorname{Diag}(\lambda(Y)))U$ for some $U \in \mathbb{O}^n$ we have the inequality

$$\begin{aligned}f^*(\lambda(Y)) &= \sup_{x \in \mathbb{R}^n} \{x^T \lambda(Y) - f(x)\} \\ &= \sup_x \{\operatorname{tr}(\operatorname{Diag}(x) UYU^T) - f(x)\} \\ &= \sup_x \{\operatorname{tr}(U^T(\operatorname{Diag}(x)UY)) - f(\lambda(U^T \operatorname{Diag}(x)U))\} \\ &\leq \sup_{X \in \mathbb{S}^n} \{\operatorname{tr}(XY) - f(\lambda(X))\} \\ &= (f \circ \lambda)^*(Y)\end{aligned}$$

□

Theorem 7. If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a symmetric function, then for any matrices $X, Y \in \mathbb{S}^n$ the following are equivalent:
(i) $Y \in \partial(f \circ \lambda)(X)$.
(ii) X and Y have a simultaneous ordered spectral decomposition and satisfy $\lambda(Y) \in \partial f(\lambda(X))$.
(iii) $X = U^T(\operatorname{Diag}(x))U$ and $Y = U^T(\operatorname{Diag}(y))U$ for some matrix $U \in \mathbb{O}^n$ and vectors $x, y \in \mathbb{R}^n$ with $y \in \partial f(x)$.

Proof.

($i \implies ii$) Let $Y \in \partial(f \circ \lambda)(X)$. By Fan's inequality we have $\operatorname{tr}(XY) \leq \lambda(Y)^T \lambda(X)$. Since Y is in the subdifferential it follows from the preceeding theorem and a lemma about the Fenchel conjugate that:

$$\begin{aligned}(f \circ \lambda)(X) + (f \circ \lambda)^*(Y) &= \langle X, Y \rangle = \operatorname{tr}(XY) \\ f(\lambda(X)) + f^*(\lambda(Y)) &= \operatorname{tr}(XY) \\ \langle \lambda(Y), \lambda(X) \rangle &\leq \operatorname{tr}(XY) \\ \lambda(Y)^T \lambda(X) &\leq \operatorname{tr}(XY)\end{aligned}$$

Since $\operatorname{tr}(XY) = \lambda(Y)^T\lambda(X)$ it follows that X and Y have a simultaneous ordered spectral decomposition. Using the fact that X and Y have a simultaneous ordered spectral decomposition we get

$$\begin{aligned}(f\circ\lambda)(X) + (f\circ\lambda)^*(Y) &= \langle X, Y\rangle = \operatorname{tr}(XY)\\ (f\circ\lambda)(X) + (f\circ\lambda)^*(Y) &= \lambda(X)^T\lambda(Y)\\ f(\lambda(X)) + f^*(\lambda(Y)) &= \langle \lambda(X), \lambda(Y)\rangle\end{aligned}$$

So, by the lemma we have $\lambda(Y) \in \partial f(\lambda(X))$.

($ii \implies i$) Since X and Y have a simultaneous ordered spectral decomposition it follows from the definition of $\lambda(Y) \in \partial f(\lambda(X))$ that

$$\begin{aligned}\langle \lambda(Y), \lambda(\overline{X}) - \lambda(X)\rangle &\le f(\lambda(Y)) - f(\lambda(X))\\ \langle \lambda(Y), \lambda(\overline{X})\rangle - \langle \lambda(Y), \lambda(X)\rangle &\le (f\circ\lambda)(Y) - (f\circ\lambda)(X)\\ \operatorname{tr}(Y\overline{X}) - \operatorname{tr}(YX) &\le (f\circ\lambda)(Y) - (f\circ\lambda)(X)\\ \operatorname{tr}(Y(\overline{X} - X) &\le (f\circ\lambda)(Y) - (f\circ\lambda)(X)\\ \langle Y, \overline{X} - X\rangle &\le (f\circ\lambda)(Y) - (f\circ\lambda)(X)\end{aligned}$$

Thus, $Y \in \partial(f\circ\lambda)(X)$.

($ii \implies iii$) Since X and Y have a simultaneous ordered spectral decomposition we have $X = U^T(\operatorname{Diag}(\lambda(X)))U$ and $Y = U^T(\operatorname{Diag}(\lambda(Y)))U$ for some $U \in \mathbb{O}^n$. Let $x = \lambda(X)$ and $y = \lambda(Y)$ then $X = U^T(\operatorname{Diag}(x))U$ and $Y = U^T(\operatorname{Diag}(y))U$. Since $\lambda(Y) \in \partial f(\lambda(X))$ it immediately follows that $y \in \partial f(x)$.

($iii \implies ii$) Let $X = U^T(\operatorname{Diag}(x))U$ and $Y = U^T(\operatorname{Diag}(y))U$ for some matrix $U \in \mathbb{O}^n$ with $x, y \in \mathbb{R}^n$ and $y \in \partial f(x)$. Then by taking determinants we have

$$\begin{aligned}\operatorname{Det}(X - \lambda I) &= \operatorname{Det}(U^T(\operatorname{Diag}(x))U - \lambda I)\\ &= \operatorname{Det}(U^T(\operatorname{Diag}(x))U) - U^T(\lambda I)U)\\ &= \operatorname{Det}(U^T(\operatorname{Diag}(x) - \lambda I)U)\\ &= \operatorname{Det}(U^T)\operatorname{Det}(\operatorname{Diag}(x) - \lambda I)\operatorname{Det}(U)\\ &= \operatorname{Det}(\operatorname{Diag}(x) - \lambda I)\\ &= (x_1 - \lambda)(x_2 - \lambda)\ldots(x_n - \lambda)\end{aligned}$$

It follows that the eigenvalues of X are the components of the vector x, thus $x = \lambda(X)$ and $X = U^T(\lambda(x))U$. A similar argument shows that $y = \lambda(y)$ and $Y = U^T(\lambda(y))U$. Then $y \in \partial f(x)$ implies that $\lambda(y) \in \partial f(\lambda(x))$. $\square$