EE/ACM 150 - Applications of Convex Optimization in Signal Processing and Communications
Lecture 3

Andre Tkacenko

Signal Processing Research Group
Jet Propulsion Laboratory

April 10, 2012

Caltech

Outline

The Singular Value Decomposition

Suppose that $\mathbf{A} \in \mathbb{C}^{m \times n}$ is any matrix with $\mathrm{rank}(\mathbf{A}) = \rho$. Then $\mathbf{A}$ admits the following factorization, known as the *singular value decomposition* (SVD):

$$\mathbf{A} = \underbrace{\begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0}_{\rho \times (n-\rho)} \\ \mathbf{0}_{(m-\rho) \times \rho} & \mathbf{0}_{(m-\rho) \times (n-\rho)} \end{bmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{bmatrix} \mathbf{V}_1^\dagger \\ \mathbf{V}_2^\dagger \end{bmatrix}}_{\mathbf{V}^\dagger} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^\dagger \,.$$

Here $\mathbf{U}_1 \in \mathbb{C}^{m \times \rho}$, $\mathbf{U}_2 \in \mathbb{C}^{m \times (m-\rho)}$, $\mathbf{U} \in \mathbb{C}^{m \times m}$ is unitary, $\mathbf{\Sigma}_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_\rho)$ is a $\rho \times \rho$ diagonal matrix of singular values of $\mathbf{A}$ (where $\sigma_k > 0$ for all $1 \leq k \leq \rho$), $\mathbf{\Sigma} \in \mathbb{R}_+^{m \times n}$, $\mathbf{V}_1 \in \mathbb{C}^{n \times \rho}$, $\mathbf{V}_2 \in \mathbb{C}^{n \times (n-\rho)}$, and $\mathbf{V} \in \mathbb{C}^{n \times n}$ is unitary.

If $\mathbf{A} \in \mathbb{R}^{m \times n}$, then there always exists an SVD factorization such that $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$, and so we have

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T \,.$$

Common SVD Types

Full and Compact

- **Full SVD:**
  The SVD representation given by
  $$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\dagger}$$
  from above is said to be the *full* SVD, since all columns of $\mathbf{U}$ and $\mathbf{V}$ are generated, even though they may not be appear explicitly in the expression for $\mathbf{A}$. This representation is neither computationally efficient nor economical from a memory storage perspective, but is useful for analysis.
- **Compact SVD:**
  The SVD representation given by
  $$\mathbf{A} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^{\dagger}$$
  from above is said to be the *compact* SVD, since only the ρ columns of $\mathbf{U}$ and $\mathbf{V}$ contributing to the expression for $\mathbf{A}$ are computed. This representation is both computationally efficient as well as economical in terms of memory, although not as useful for analysis as the full SVD.

Relation to Eigendecomposition

Note that we have

$$\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^{\dagger}\mathbf{U}^{\dagger} = \mathbf{U}\mathbf{\Lambda}_L\mathbf{U}^{\dagger}\,,\ \mathbf{A}^{\dagger}\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^{\dagger}\mathbf{\Sigma}\mathbf{V}^{\dagger} = \mathbf{V}\mathbf{\Lambda}_R\mathbf{V}^{\dagger}\,,$$

where $\mathbf{\Lambda}_L \in \mathbb{R}_+^{m\times m}$ and $\mathbf{\Lambda}_R \in \mathbb{R}_+^{n\times n}$ are square diagonal matrices given by

$$\mathbf{\Lambda}_L = \begin{bmatrix} \mathbf{\Sigma}_1^2 & \mathbf{0}_{\rho\times(m-\rho)} \\ \mathbf{0}_{(m-\rho)\times\rho} & \mathbf{0}_{(m-\rho)\times(m-\rho)} \end{bmatrix}\,,\ \mathbf{\Lambda}_R = \begin{bmatrix} \mathbf{\Sigma}_1^2 & \mathbf{0}_{\rho\times(n-\rho)} \\ \mathbf{0}_{(n-\rho)\times\rho} & \mathbf{0}_{(n-\rho)\times(n-\rho)} \end{bmatrix}\,.$$

Thus, the SVD of $\mathbf{A}$ is closely related to the eigendecomposition of $\mathbf{A}\mathbf{A}^{\dagger}$ and $\mathbf{A}^{\dagger}\mathbf{A}$.

- The left singular vectors (columns of $\mathbf{U}$) are eigenvectors of $\mathbf{A}\mathbf{A}^{\dagger}$.
- The right singular vectors (columns of $\mathbf{V}$) are eigenvectors of $\mathbf{A}^{\dagger}\mathbf{A}$.
- The nonzero singular values are the square roots of the eigenvalues of both $\mathbf{A}\mathbf{A}^{\dagger}$ and $\mathbf{A}^{\dagger}\mathbf{A}$, i.e.,

$$\sigma_k = \sqrt{\lambda_k(\mathbf{A}\mathbf{A}^{\dagger})} = \sqrt{\lambda_k(\mathbf{A}^{\dagger}\mathbf{A})}\,,\ 1 \le k \le \rho\,.$$

Relation to Fundamental Theorem of Linear Algebra

Every matrix $\mathbf{A} \in \mathbb{F}^{m \times n}$ induces *four fundamental subspaces* which are elegantly described in terms of the full SVD given above. These are described below.

Subspace	Description	Underlying Space	Dimension	Basis
range space, column space, or image	$\mathcal{R}(\mathbf{A})$	$\mathbb{F}^m$	ρ	$\mathbf{U}_1$
null space, right null space, or kernel	$\mathcal{N}(\mathbf{A})$	$\mathbb{F}^n$	$n - \rho$	$\mathbf{V}_2$
row space or co-image	$\mathcal{R}\left(\mathbf{A}^\dagger\right)$	$\mathbb{F}^n$	ρ	$\mathbf{V}_1$
left null space or co-kernel	$\mathcal{N}\left(\mathbf{A}^\dagger\right)$	$\mathbb{F}^m$	$m - \rho$	$\mathbf{U}_2$

In addition, we have the following properties:

- For the vector space $\mathbb{F}^n$, we have $\mathcal{N}(\mathbf{A}) = \left(\mathcal{R}\left(\mathbf{A}^\dagger\right)\right)^\perp$. In other words, the (right) null space is the *orthogonal complement* of the row space.
- For the vector space $\mathbb{F}^m$, we have $\mathcal{N}\left(\mathbf{A}^\dagger\right) = \left(\mathcal{R}(\mathbf{A})\right)^\perp$. In other words, the left null space is the *orthogonal complement* of the column space.

Here, the *orthogonal complement* $\mathcal{W}^\perp$ of a subspace $\mathcal{W}$ in an inner product space $\mathcal{V}$ is defined as

$$\mathcal{W}^\perp \triangleq \{\mathbf{x} \in \mathcal{V} : \langle \mathbf{x}, \mathbf{y} \rangle = 0 \ \forall \ \mathbf{y} \in \mathcal{W}\} .$$

Unitarily Invariant Norms and the SVD

An $m \times n$ matrix norm for which

$$||\mathbf{A}|| = ||\mathbf{UAV}|| \ ,$$

for any $m \times n$ matrix $\mathbf{A}$, $m \times m$ unitary matrix $\mathbf{U}$, and $n \times n$ unitary matrix $\mathbf{V}$ is said to be a *unitarily invariant norm*.

Any matrix norm which is only a function of the singular values of a matrix is unitarily invariant. The *Schatten ℓ_p-norm* is one such example. If $\mathbf{A}$ has nonzero singular values $\sigma_1, \ldots, \sigma_\rho$, where $\rho = \mathrm{rank}(\mathbf{A})$ the Schatten ℓ_p-norm of $\mathbf{A}$ is defined as

$$||\mathbf{A}||_p \triangleq \left(\sum_{k=1}^{\rho} \sigma_k^p \right)^{\frac{1}{p}} \ .$$

Special Cases:

- $p = 1$ (*nuclear norm* or *Ky Fan ρ-norm*)

$$||\mathbf{A}||_1 = \mathrm{tr}\left(\sqrt{\mathbf{A}^\dagger \mathbf{A}} \right)$$

- $p = 2$ (*Frobenius norm*)

$$||\mathbf{A}||_2 = \sqrt{\mathrm{tr}(\mathbf{A}^\dagger \mathbf{A})}$$

- $p = \infty$ (*spectral norm*)

$$||\mathbf{A}||_\infty = \max_{1 \le k \le \rho} \{\sigma_k\} = \sqrt{\lambda_{\max}(\mathbf{A}^\dagger \mathbf{A})}$$

Low-Rank Approximation Problem

The SVD can be used to find *globally optimal* solutions to certain *nonconvex* problems. One such example is the *low-rank matrix approximation problem*. If $\mathbf{A}$ is some $m \times n$ matrix with rank ρ, the problem is to find an $m \times n$ matrix $\widehat{\mathbf{A}}$ with rank $\widehat{\rho} \leq \rho$ that minimizes the norm of the error matrix $\left(\mathbf{A} - \widehat{\mathbf{A}}\right)$.

Low-Rank Approximation Problem:

$$\begin{array}{ll} \text{minimize} & \left|\left|\mathbf{A} - \widehat{\mathbf{A}}\right|\right| , \\ \text{subject to} & \operatorname{rank}\left(\widehat{\mathbf{A}}\right) = \widehat{\rho} \leq \rho = \operatorname{rank}(\mathbf{A}) \ . \end{array}$$

If $\mathbf{A}$ has an SVD as given above with singular values arranged in descending order (i.e., $\sigma_1 \geq \cdots \geq \sigma_\rho > 0$), the *globally optimal* choice of $\widehat{\mathbf{A}}$ for *any unitarily invariant norm* is $\widehat{\mathbf{A}}^\star = \mathbf{U}\widehat{\mathbf{\Sigma}}^\star\mathbf{V}^\dagger$, where $\widehat{\mathbf{\Sigma}}^\star$ is given by

$$\widehat{\mathbf{\Sigma}}^\star = \begin{bmatrix} \widehat{\mathbf{\Sigma}}_1^\star & \mathbf{0}_{\widehat{\rho} \times (n - \widehat{\rho})} \\ \mathbf{0}_{(m - \widehat{\rho}) \times \widehat{\rho}} & \mathbf{0}_{(m - \widehat{\rho}) \times (n - \widehat{\rho})} \end{bmatrix} ,$$

where $\widehat{\mathbf{\Sigma}}_1^\star = \operatorname{diag}(\sigma_1, \ldots, \sigma_{\widehat{\rho}})$.

Orthogonal Procrustes Problem

Another nonconvex problem for which the SVD can be used to obtain a globally optimal solution is the *orthogonal Procrustes problem*. This problem involves approximating, in the Frobenius norm sense, a given $m \times n$ matrix $\mathbf{A}$ by a right sided rotation of another given $m \times n$ matrix $\mathbf{B}$. This right sided rotation is carried out via an $n \times n$ unitary matrix $\mathbf{W}$ to produce the rotation $\mathbf{B}\mathbf{W}^\dagger$.

Orthogonal Procrustes Problem:

$$\begin{aligned} \text{minimize} \quad & \left|\left|\mathbf{A} - \mathbf{B}\mathbf{W}^\dagger\right|\right|_F \,, \\ \text{subject to} \quad & \mathbf{W}^\dagger \mathbf{W} = \mathbf{I}_n \,. \end{aligned}$$

If the $n \times n$ product matrix $\mathbf{A}^\dagger \mathbf{B}$ has an SVD of $\mathbf{A}^\dagger \mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$, then a globally optimal solution $\mathbf{W}^\star$ to the orthogonal Procrustes problem is given by

$$\mathbf{W}^\star = \mathbf{U}\mathbf{V}^\dagger \,.$$

The globally optimal objective function value is given by

$$\sqrt{||\mathbf{A}||_F^2 + ||\mathbf{B}||_F^2 - 2\mathrm{tr}(\mathbf{\Sigma})} \,.$$

Moore-Penrose Pseudoinverse: Defining Properties

Suppose $\mathbf{A}$ is any $m \times n$ matrix. Then, if an $n \times m$ matrix $\mathbf{A}^{-}$ satisfies

$$\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}\,,$$

then $\mathbf{A}^{-}$ is said to be a *generalized inverse* of $\mathbf{A}$. If, in addition, $\mathbf{A}^{-}$ satisfies

$$\mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-} = \mathbf{A}^{-}\,,$$

then $\mathbf{A}^{-}$ is said to be a *reflexive generalized inverse* of $\mathbf{A}$. Generalized inverses and reflexive generalized inverses always exist, but are not unique, in general. If we include two additional conditions, we obtain a *Moore-Penrose pseudoinverse*, which is unique.

Moore-Penrose Pseudoinverse Definition:

If $\mathbf{A}$ is any $m \times n$ matrix, then an $n \times m$ matrix $\mathbf{A}^{\#}$ is said to be a pseudoinverse of $\mathbf{A}$ if the following conditions hold:

1. $\mathbf{A}\mathbf{A}^{\#}\mathbf{A} = \mathbf{A}$: ($\mathbf{A}^{\#}$ is a generalized inverse of $\mathbf{A}$).
2. $\mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#} = \mathbf{A}^{\#}$: ($\mathbf{A}$ is a generalized inverse of $\mathbf{A}^{\#}$).
3. $\left(\mathbf{A}\mathbf{A}^{\#}\right)^{\dagger} = \mathbf{A}\mathbf{A}^{\#}$: ($\mathbf{A}\mathbf{A}^{\#}$ is Hermitian).
4. $\left(\mathbf{A}^{\#}\mathbf{A}\right)^{\dagger} = \mathbf{A}^{\#}\mathbf{A}$: ($\mathbf{A}^{\#}\mathbf{A}$ is Hermitian).

Properties of the Pseudoinverse

For all properties stated here, we assume $\mathbf{A}$ is some $m \times n$ matrix and $\mathbf{A}^{\#}$ is the $n \times m$ pseudoinverse of $\mathbf{A}$.

- **Existence and uniqueness:**
  As stated above, $\mathbf{A}^{\#}$ always exists and is unique.
- **Basic Properties:**
  - If $\mathbf{A} \in \mathbb{R}^{m \times n}$, then $\mathbf{A}^{\#} \in \mathbb{R}^{n \times m}$.
  - If $\mathbf{A}$ is invertible, then $\mathbf{A}^{\#} = \mathbf{A}^{-1}$.
  - If $\mathbf{A} = \mathbf{0}_{m \times n}$, then $\mathbf{A}^{\#} = \mathbf{0}_{n \times m}$.
  - $\left(\mathbf{A}^{\#}\right)^{\#} = \mathbf{A}$.
  - $(\mathbf{A}^{*})^{\#} = \left(\mathbf{A}^{\#}\right)^{*}$, $\left(\mathbf{A}^{T}\right)^{\#} = \left(\mathbf{A}^{\#}\right)^{T}$, $\left(\mathbf{A}^{\dagger}\right)^{\#} = \left(\mathbf{A}^{\#}\right)^{\dagger}$.
  - $(\alpha \mathbf{A})^{\#} = \frac{1}{\alpha} \mathbf{A}^{\#}$ for any scalar α such that $\alpha \neq 0$.
- **Identities:**

$$
\begin{aligned}
\mathbf{A}^{\#} &= \mathbf{A}^{\#} \left(\mathbf{A}^{\#}\right)^{\dagger} \mathbf{A}^{\dagger} . \\
\mathbf{A}^{\#} &= \mathbf{A}^{\dagger} \left(\mathbf{A}^{\#}\right)^{\dagger} \mathbf{A}^{\#} . \\
\mathbf{A} &= \left(\mathbf{A}^{\#}\right)^{\dagger} \mathbf{A}^{\dagger} \mathbf{A} . \\
\mathbf{A} &= \mathbf{A} \mathbf{A}^{\dagger} \left(\mathbf{A}^{\#}\right)^{\dagger} . \\
\mathbf{A}^{\dagger} &= \mathbf{A}^{\dagger} \mathbf{A} \mathbf{A}^{\#} . \\
\mathbf{A}^{\dagger} &= \mathbf{A}^{\#} \mathbf{A} \mathbf{A}^{\dagger} .
\end{aligned}
$$

Connection Between the Pseudoinverse and SVD

There is an intimate connection between the SVD of a matrix and its pseudoinverse. Specifically, if $\mathbf{A}$ is an $m \times n$ matrix with an SVD given by

$$\mathbf{A} = \underbrace{\left[\begin{array}{cc} \mathbf{U}_1 & \mathbf{U}_2 \end{array}\right]}_{\mathbf{U}} \underbrace{\left[\begin{array}{cc} \mathbf{\Sigma}_1 & \mathbf{0}_{\rho\times(n-\rho)} \\ \mathbf{0}_{(m-\rho)\times\rho} & \mathbf{0}_{(m-\rho)\times(n-\rho)} \end{array}\right]}_{\mathbf{\Sigma}} \underbrace{\left[\begin{array}{c} \mathbf{V}_1^\dagger \\ \mathbf{V}_2^\dagger \end{array}\right]}_{\mathbf{V}^\dagger} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^\dagger \,.$$

then the $n \times m$ pseudoinverse $\mathbf{A}^\#$ is given by

$$\mathbf{A}^\# \triangleq \underbrace{\left[\begin{array}{cc} \mathbf{V}_1 & \mathbf{V}_2 \end{array}\right]}_{\mathbf{V}} \underbrace{\left[\begin{array}{cc} \mathbf{\Sigma}_1^{-1} & \mathbf{0}_{\rho\times(m-\rho)} \\ \mathbf{0}_{(n-\rho)\times\rho} & \mathbf{0}_{(n-\rho)\times(m-\rho)} \end{array}\right]}_{\mathbf{\Sigma}^\#} \underbrace{\left[\begin{array}{c} \mathbf{U}_1^\dagger \\ \mathbf{U}_2^\dagger \end{array}\right]}_{\mathbf{U}^\dagger} = \mathbf{V}_1\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^\dagger \,.$$

Note that the matrix $\mathbf{\Sigma}^\#$ is the pseudoinverse of the matrix $\mathbf{\Sigma}$.

Pseudoinverse and Linear Least-Squares Problems

The pseudoinverse arises in solutions to *linear least-squares problems*, in which the aim is to find the best linear model fit to a set of observed data in a mean-squared sense.

If $\mathbf{A}$ is an $m \times n$ matrix characterizing the linear model, $\mathbf{b}$ is an $m \times 1$ vector characterizing the observed data, and $\mathbf{x}$ is an $n \times 1$ vector characterizing the fitting parameters, then the traditional linear least-squares problem is simply the following.

Linear Least-Squares Problem:

$$\text{minimize} \quad \xi^2 \triangleq ||\mathbf{Ax} - \mathbf{b}||_2^2 \,.$$

This problem is always convex and a globally optimizing solution $\mathbf{x}^\star$ and corresponding optimal objective value $\left(\xi^2\right)^\star$ are given by

$$\mathbf{x}^\star = \mathbf{A}^\# \mathbf{b} \,, \ \left(\xi^2\right)^\star = \left|\left|\left(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\#\right) \mathbf{b}\right|\right|_2^2 \,.$$

Completing the Square

The optimizing solution $\mathbf{x} = \mathbf{A}^{\#}\mathbf{b}$ to the linear least-squares problem can be obtained through a variety of ways. However, one technique simultaneously yields this solution and shows its global optimality. This involves using the trick of *completing the square*.

We will show this for the special case in which $m \geq n$ (i.e., $\mathbf{A}$ is a *tall* matrix) and $\mathbf{A}$ is full rank (i.e., $\text{rank}(\mathbf{A}) = n$). Expanding ξ^2 from above, we get

$$\begin{aligned} \xi^2 &= ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 = \mathbf{x}^\dagger \mathbf{A}^\dagger \mathbf{A}\mathbf{x} - \mathbf{x}^\dagger \mathbf{A}^\dagger \mathbf{b} - \mathbf{b}^\dagger \mathbf{A}\mathbf{x} + \mathbf{b}^\dagger \mathbf{b}\,, \\ &= \left(\mathbf{x}^\dagger - \mathbf{b}^\dagger \mathbf{A}\left(\mathbf{A}^\dagger \mathbf{A}\right)^{-1}\right)\left(\mathbf{A}^\dagger \mathbf{A}\right)\left(\mathbf{x} - \left(\mathbf{A}^\dagger \mathbf{A}\right)^{-1} \mathbf{A}^\dagger \mathbf{b}\right) \\ &\quad + \mathbf{b}^\dagger \left(\mathbf{I}_m - \mathbf{A}\left(\mathbf{A}^\dagger \mathbf{A}\right)^{-1} \mathbf{A}^\dagger\right)\mathbf{b}\,, \\ &= \left|\left|\mathbf{A}\left(\mathbf{x} - \left(\mathbf{A}^\dagger \mathbf{A}\right)^{-1} \mathbf{A}^\dagger \mathbf{b}\right)\right|\right|_2^2 + \left|\left|\left(\mathbf{I}_m - \mathbf{A}\left(\mathbf{A}^\dagger \mathbf{A}\right)^{-1} \mathbf{A}^\dagger\right)\mathbf{b}\right|\right|_2^2 . \end{aligned}$$

Note that only the first term from above depends upon $\mathbf{x}$. This term is nonnegative and zero if and only if its argument is zero, which occurs if we have

$$\mathbf{x} = \mathbf{x}^\star = \underbrace{\left(\mathbf{A}^\dagger \mathbf{A}\right)^{-1} \mathbf{A}^\dagger}_{\mathbf{A}^{\#}} \mathbf{b} \implies \xi^2 = \left(\xi^2\right)^\star = \left|\left|\left(\mathbf{I}_m - \underbrace{\mathbf{A}\left(\mathbf{A}^\dagger \mathbf{A}\right)^{-1} \mathbf{A}^\dagger}_{\mathbf{A}\mathbf{A}^{\#}}\right)\mathbf{b}\right|\right|_2^2 .$$