Linear Analysis
Lecture 13

Singular Value Decomposition (SVD)

If $A \in \mathbb{C}^{m\times n}$, then there exists unitary matrices

$$U \in \mathbb{C}^{m\times m} \quad \text{and} \quad V \in \mathbb{C}^{n\times n}$$

such that

$$A = U\Sigma V^H,$$

where $\Sigma \in \mathbb{C}^{m\times n}$ is the diagonal matrix of singular values.
In particular, if

$$\sigma_1 \geq \sigma_2 \geq \ldots \sigma_p \qquad (p = \min(m, n))$$

are the singular values of A with

$$\mathrm{diag}\,(\Sigma) = [\sigma_1, \sigma_2, \sigma_3, \ldots]$$

and

$$U = [u_1,\ u_2, \ldots,\ u_m] \quad \text{and} \quad V = [v_1,\ v_2, \ldots,\ v_n]\ ,$$

then

$$\sigma_j u_j = A v_j \quad j = 1, 2, \ldots, p\ .$$

Singular Value Decomposition: Proof

As in the square case, $\|A\|^2 = \|A^{\mathrm{H}}A\|$. But

$$\|A^{\mathrm{H}}A\| = \lambda_1 = \sigma_1^2, \quad \text{so} \quad \|A\| = \sigma_1.$$

So $\exists\, x \in \mathbb{C}^n$ with $\|x\| = 1$ and $\|Ax\| = \sigma_1$, and write $Ax = \sigma_1 y$ where $\|y\| = 1$. Complete x and y to unitary matrices

$$V_1 = [x, \widetilde{v}_2, \cdots, \widetilde{v}_n] \in \mathbb{C}^{n\times n} \quad \text{and} \quad U_1 = [y, \widetilde{u}_2, \cdots, \widetilde{u}_m] \in \mathbb{C}^{m\times m}.$$

Since $U_1^H A V_1 =: A_1$ is the matrix of A in these bases it follows that

$$A_1 = \begin{bmatrix} \sigma_1 & w^H \\ 0 & B \end{bmatrix}$$

for some $w \in \mathbb{C}^{n-1}$ and $B \in \mathbb{C}^{(m-1)\times(n-1)}$. Observe that

$$\begin{aligned} \sigma_1^2 + w^* w &\le \left\| \begin{bmatrix} \sigma_1^2 + w^* w \\ Bw \end{bmatrix} \right\| = \left\| A_1 \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\| \\ &\le \|A_1\| \cdot \left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\| = \sigma_1 (\sigma_1^2 + w^* w)^{\frac{1}{2}} \end{aligned}$$

since $\|A_1\| = \|A\| = \sigma_1$ by the invariance of $\|\cdot\|$ under unitary multiplication. It follows that $(\sigma_1^2 + w^* w)^{\frac{1}{2}} \le \sigma_1$, so $w = 0$, and thus

$$A_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix}.$$

Singular Value Decomposition: Proof

Thus far we have

$$U_1^H A V_1 = A_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix}.$$

Apply the same argument to B and repeat to get the result. For this, observe that

$$\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & B^H B \end{bmatrix} = A_1^H A_1 = V_1^H A^{\mathrm{H}} A V_1$$

is unitarily similar to $A^{\mathrm{H}} A$, so the eigenvalues of $B^H B$ are

$$\lambda_2 \geq \cdots \geq \lambda_n \geq 0.$$

Observe also that the same argument shows that if $A \in \mathbb{R}^{m \times n}$, then U and V can be taken to be real orthogonal matrices. $\square$

Alternative Proof of SVD

Although short, this proof masks some of the key ideas. An alternative proof revealing more of the structure of the SVD is given below.

Alternative Proof: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $A^{\mathrm{H}}A$ associated with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, respectively, and let $V = [v_1 \cdots v_n] \in \mathbb{C}^{n\times n}$. Then V is unitary, and

$$V^H A^{\mathrm{H}} A V = \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n\times n}.$$

For $1 \leq i \leq n$, $\|Av_i\|^2 = e_i^H V^H A^{\mathrm{H}} A V e_i = \lambda_i = \sigma_i^2$.
Choose r so that $\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$ ($r = \operatorname{rank} A$).
Then, for $1 \leq i \leq r$, $Av_i = \sigma_i u_i$ for a unique $u_i \in \mathbb{C}^m$ with $\|u_i\| = 1$.
Moreover, for $1 \leq i, j \leq r$,

$$u_i^H u_j = \frac{1}{\sigma_i \sigma_j} v_i^H A^{\mathrm{H}} A v_j = \frac{1}{\sigma_i \sigma_j} e_i^H \Lambda e_j = \delta_{ij}.$$

So we can append vectors $u_{r+1}, \ldots, u_m \in \mathbb{C}^m$ (if necessary) so that $U = [u_1 \cdots u_m] \in \mathbb{C}^{m\times m}$ is unitary. It follows easily that

$$AV = U\Sigma, \quad \text{so} \quad A = U\Sigma V^H.$$

□

Insight from the Alternative Proof of SVD

The key insite in the alternative proof is the relation

$$AV = U\Sigma\ .$$

Interpreting this equation columnwise gives

$$(*) \qquad Av_i = \sigma_i u_i \quad (1 \le i \le p),$$

and

$$Av_i = 0 \quad \text{for } i > m \text{ if } n > m,$$

where $\{v_1, \ldots, v_n\}$ are the columns of V and $\{u_1, \ldots, u_m\}$ are the columns of U. So A maps the orthonormal vectors $\{v_1, \ldots, v_p\}$ into the orthogonal directions $\{u_1, \ldots, u_p\}$ with the singular values $\sigma_1 \ge \cdots \ge \sigma_p$ as scale factors.

Next, multiply the equations $(*)$ through by A^H to get

$$\sigma_i^2 v_i = A^H A v_i = \sigma_i A^H u_i \quad (1 \le i \le p)$$

yielding

$$(**) \qquad A^H u_i = \sigma_i v_i \quad (1 \le i \le p).$$

That is, A^H maps the orthonormal vectors $\{u_1, \ldots, u_p\}$ into the orthogonal directions $\{v_1, \ldots, v_p\}$ with the singular values $\sigma_1 \ge \cdots \ge \sigma_p$ as scale factors.

The vectors $v_1, \ldots, v_n$ are called the **right singular vectors** of A, and $u_1, \ldots, u_m$ are called the **left singular vectors** of A.

Observe that

$$A^{\mathrm{H}} A = V \Sigma^H \Sigma V^H \quad \text{and} \quad \Sigma^H \Sigma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2) \in \mathbb{R}^{n \times n}$$

even if $m < n$. So

$$V^H A^{\mathrm{H}} A V = \Lambda = \operatorname{diag}(\lambda_1, \ldots \lambda_n),$$

and thus the columns of V form an orthonormal basis consisting of eigenvectors of $A^{\mathrm{H}} A \in \mathbb{C}^{n \times n}$. Similarly $A A^{\mathrm{H}} = U \Sigma \Sigma^H U^H$, so

$$U^H A A^{\mathrm{H}} U = \Sigma \Sigma^H = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_p^2, \overbrace{0, \ \ldots, \ 0}^{(m-n \text{ zeroes if } m>n)}) \in \mathbb{R}^{m \times m},$$

and thus the columns of U form an orthonormal basis of $\mathbb{C}^m$ consisting of eigenvectors of $A A^{\mathrm{H}} \in \mathbb{C}^{m \times m}$.

Reduced Singular Value Decomposition (Reduced SVD)

Let $A \in \mathbb{C}^{m\times n}$ have SVD $A = U\Sigma V^H$, where

$$U = [u^1, u^2, \ldots, u^m] \in \mathbb{C}^{m\times m} \quad \text{and} \quad V = [v^1, v^2, \ldots, v^m] \in \mathbb{C}^{n\times n}$$

are unitary and $\Sigma \in \mathbb{C}^{m\times n}$ is diagonal with the first $p = \min\{n, m\}$ diagonal entries being the singular values of A ordered largest to smallest. Let $1 \le k \le p$ be such that $\sigma_k > 0$ and $\sigma_{k+1} = 0$. The $\mathrm{rank}\, A = k$ and

$$A = \hat{U}\hat{\Sigma}\hat{V}^H,$$

where $\hat{\Sigma} = \mathrm{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_k) \in \mathbb{R}^{k\times k}$,

$$\hat{U} = [u^1, u^2, \ldots, u^k] \in \mathbb{C}^{m\times k} \quad \text{and} \quad \hat{V} = [v^1, v^2, \ldots, v^k] \in \mathbb{C}^{n\times k}.$$

Moreover, UU^H is the orthogonal projector onto $\mathrm{Ran}(A)$ and VV^H is the orthogonal projector onto $\mathrm{Nul}(A)^\perp$.

Proposition. Let $A \in \mathbb{C}^{n\times n}$ be normal, and order the eigenvalues of A as $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Then the singular values of A are

$$\sigma_i = |\lambda_i|, \quad 1 \leq i \leq n \ .$$

Proof: By the Spectral Theorem for normal operators,

$$A = V\Lambda V^H \ ,$$

where $V \in \mathbb{C}^{n\times n}$ is unitary. For $1 \leq i \leq n$, choose $d_i \in \mathbb{C}$ for which

$$\bar{d}_i\lambda_i = |\lambda_i| \quad \text{and} \quad |d_i| = 1,$$

and let $D = \operatorname{diag}(d_1, \ldots, d_n)$. Then D is unitary, and

$$A = (VD)(D^H\Lambda)V^H \equiv U\Sigma V^H,$$

where $U = VD$ is unitary and

$$\Sigma = D^H\Lambda = \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_n|)$$

is diagonal with decreasing nonnegative diagonal entries. □

The SVD and Matrix Norms

The Frobenius and Euclidean operator norms of $A \in \mathbb{C}^{m\times n}$ are easily expressed in terms of the singular values of A: set $t := \min\{m, n\}$,

$$\|A\|_F = \left(\sum_{i=1}^{t} \sigma_i^2\right)^{\frac{1}{2}} = \left\| \begin{matrix} \sigma_1 \\ \vdots \\ \sigma_t \end{matrix} \right\|_2$$

and

$$\|A\| = \sigma_1 = \sqrt{p(A^{\mathrm{H}}A)} = \left\| \begin{matrix} \sigma_1 \\ \vdots \\ \sigma_t \end{matrix} \right\|_\infty ,$$

as follows from the unitary invariance of these norms.

There are no such simple expressions (in general) for these norms in terms of the eigenvalues of A if A is square (but not normal).

The Schatten-p Norms

Extending these expressions for the Frobenius and Euclidean operator norms of $A \in \mathbb{C}^{m\times n}$, we obtain the Schatten-p Norms:

$$\|A\|_{(p)} = \left\| \begin{array}{c} \sigma_1 \\ \vdots \\ \sigma_t \end{array} \right\|_p ,$$

for $1 \leq p < \infty$.

The case $p = 1$ is also called the **trace** norm or the **nuclear** norm:

$$\|A\|_{(1)} = \sum_{i=1}^{t} \sigma_i \ .$$

The SVD and Rank

The SVD is useful computationally for questions involving rank.

The rank of $A \in \mathbb{C}^{m \times n}$ is the number of nonzero singular values of A since rank is invariant under pre- and post-multiplication by invertible matrices.

There are stable numerical algorithms for computing SVD (try `matlab`).

In the presence of round-off error, row-reduction to echelon form usually fails to find the rank of A when its rank is $< \min(m, n)$.

For such a matrix, the computed SVD has the zero singular values computed to be on the order of machine ϵ, and these are often identifiable as "numerical zeroes."

For example, if the computed singular values of A are

$$10^2, 10, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-15}, 10^{-15}, 10^{-16}$$

with machine $\epsilon \approx 10^{-16}$, one can safely expect $\operatorname{rank}(A) = 7$.

The SVD and Polar Form

We now consider the matrix analogue of the polar form $z = re^{i\theta}$.

Proposition: Every $A \in \mathbb{C}^{n \times n}$ may be written as $A = PU$, where P is positive semi-definite Hermitian and U is unitary.

Proof: Let

$$A = U\Sigma V^H$$

be a SVD for A, and write

$$A = (U\Sigma U^H)(UV^H).$$

Then

$$U\Sigma U^H$$

is positive semi-definite Hermitian and

$$UV^H$$

is unitary. $\square$

Linear Least Squares Problems

If $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$, and consider the system $Ax = b$. This system may not be solvable, especially if $m > n$.
Instead solve

$$\inf_{x\in\mathbb{C}^n} \frac{1}{2}\|Ax - b\|_2^2 \ . \qquad (*)$$

This is called a least-squares problem since the square of the Euclidean norm is a sum of squares.
Set $\varphi(x) = \frac{1}{2}\|Ax - b\|^2$, then at the solution to (*) $\nabla\varphi(x) = 0$, or equivalently $\varphi'(x; v) = 0 \quad \forall\, v \in \mathbb{C}^n$, where $\varphi'(x; v) = \frac{d}{dt}\varphi(x + tv)\Big|_{t=0}$ is the directional derivative. If $y(t)$ is a differentiable curve in $\mathbb{C}^m$, then

$$\frac{d}{dt}(\frac{1}{2}\|y(t)\|^2) = \frac{1}{2}(\langle y'(t), y(t)\rangle + \langle y(t), y'(t)\rangle) = \mathcal{R}e\langle y(t), y'(t)\rangle.$$

Taking $y(t) = A(x + tv) - b$, we obtain that

$$\nabla\varphi(x) = 0 \Leftrightarrow (\forall\, v \in \mathbb{C}^n)\ \mathcal{R}e\langle Ax - b, Av\rangle = 0 \Leftrightarrow A^H(Ax - b) = 0,$$

i.e.,

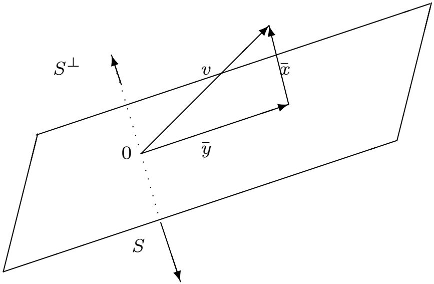
$$A^{\mathrm{H}}Ax \ = \ A^H b \ .$$

These are called the **normal equations** (they say $(Ax - b) \perp \mathcal{R}(A)$).

The Projection Theorem

Let $S \subset V$ be a subspace of the Euclidean space V.

1. $V = S \oplus S^{\perp}$, i.e., given $v \in V$, $\exists$ unique $\bar{y} \in S$ and $\bar{z} \in S^{\perp}$ for which $v = \bar{y} + \bar{z}$ (so $\bar{y} = Pv$ and $\bar{z} = (I - P)v$, where P is the orth. proj. onto S and $(I - P)$ is the orth. proj. onto $S^{\perp}$).
2. Given $v \in V$, $\bar{y}$ is the unique element of S which satisfies $\langle v - \bar{y}, y \rangle = 0 \; \forall \, y \in S$.
3. $\hat{y} = \bar{y}$ if and only if $\hat{y}$ is the unique element of S solving the minimization problem $\quad \min \left\{ \|v - y\|^2 \;\; : \; y \in S \right\}.$



Proof of The Projection Theorem

(1) $V = S \oplus S^{\perp}$, i.e., given $v \in V$, $\exists$ unique $\bar{y} \in S$ and $\bar{z} \in S^{\perp}$ for which $v = \bar{y} + \bar{z}$ so
$$\bar{y} = Pv \quad \text{and} \quad \bar{z} = (I - P)v,$$
where P is the orthogonal projection of V onto S.

Proof: Let $\{\psi_1, \ldots, \psi_r\}$ be an orthonormal basis of S. Given $v \in V$, let

$$\bar{y} = \sum_{j=1}^{r} \langle v, \psi_j \rangle \psi_j \quad \text{and} \quad \bar{z} = v - \bar{y}.$$

Then $v = \bar{y} + \bar{z}$ and $\bar{y} \in S$. For $1 \leq k \leq r$,

$$\langle \bar{z}, \psi_k \rangle = \langle v, \psi_k \rangle - \langle \bar{y}, \psi_k \rangle = \langle v, \psi_k \rangle - \langle v, \psi_k \rangle = 0,$$

so $\bar{z} \in S^{\perp}$.
Uniqueness follows from the fact that $S \cap S^{\perp} = \{0\}$. □

Proof of The Projection Theorem

(2) Given $v \in V$, the $\bar{y}$ in (1) is the unique element of S which satisfies

$$(\forall\, y \in S) \qquad \langle v - \bar{y}, y \rangle = 0.$$

Proof
Since $\bar{z} = v - \bar{y}$, this is just a restatement of $\bar{z} \in S^{\perp}$. $\square$

Proof of The Projection Theorem

(3) Given $v \in V$ let $\bar{y} = P_V y$. Then $\hat{y} = \bar{y}$ if and only if $\hat{y}$ is the unique element of S solving the minimization problem

$$\underset{y \in S}{\text{minimize}} \ \|v - y\|^2 \ .$$

Proof: For any $y \in S$,

$$v - y = \underbrace{\bar{y} - y}_{\in S} + \underbrace{\bar{z}}_{\in S^\perp},$$

so by the Pythagorean Theorem

$(p \perp q \Leftrightarrow \|p \pm q\|^2 = \|p\|^2 + \|q\|^2)$, and so $\|v - y\|^2 = \|\bar{y} - y\|^2 + \|\bar{z}\|^2$.

Therefore, $\|v - y\|^2$ is minimized iff $y = \bar{y}$, and $\|v - \bar{y}\|^2 = \|\bar{z}\|^2$. $\square$