ME 7247: Advanced Control Systems **Fall 2022–23**

Lecture 06: Applications of SVD

Tuesday September 27, 2022

Lecturer: Laurent Lessard *Scribe: Matthew Wallace*

This lecture begins with a review of the SVD and the its usage in solving linear equations. Least squares is then rederived with this approach. SVD is then compared with eigenvalue decomposition, and diagonalizability and invertibility is discussed. The lecture concludes with the application of SVD to sensitivity of matrices in both the controls and estimation problem.

1 Linear equations

Let $A \in \mathbb{R}^{m \times n}$. Consider the linear equations $Ax = b$. Let $A = U\Sigma V^\mathsf{T}$ be the full SVD of A. rearrange the equation to obtain:

$$\begin{aligned} Ax &= b \\ u\Sigma V^\mathsf{T} x &= b \\ \Sigma(V^\mathsf{T} x) &= (U^\mathsf{T} v) \\ \implies \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} &= \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{bmatrix} \end{aligned}$$

where we defined the tilde vectors $\tilde{x}_1, \tilde{x}_2, \tilde{b}_1, \tilde{b}_2$ as:

$$V^\mathsf{T} x = \begin{bmatrix} V_1^\mathsf{T} x \\ V_2^\mathsf{T} x \end{bmatrix} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} \qquad \text{and} \qquad U^\mathsf{T} b = \begin{bmatrix} U_1^\mathsf{T} b \\ U_2^\mathsf{T} b \end{bmatrix} = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{bmatrix}$$

Recall that these are simply extracting the coordinates in a given basis. For example, to find the coordinates of $x \in \mathbb{R}^n$ expressed in the orthonormal basis $\{v_1, \dots, \boldsymbol{v}_n\}$, we let $V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and we write:

$$x = VV^\mathsf{T} x = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} v_1^\mathsf{T} x \\ \vdots \\ v_n^\mathsf{T} x \end{bmatrix} = v_1(v_1^\mathsf{T} x) + \cdots + v_n(v_n^\mathsf{T} x) = \tilde{v}_1 v_1 + \cdots + \tilde{v}_n v_n = V\tilde{v}.$$

Our equations then reduce to: $\Sigma_1 \tilde{x}_1 = \tilde{b}_1$ and $\tilde{b}_2 = 0$. A solution exists if and only if $\tilde{b}_2 = 0$, which means that $U_2^\mathsf{T} b = 0$, i.e. $b \in \text{range}(U_1) = \text{range}(A)$. The solution is unique if and only if $\tilde{x}_2$ has dimension zero (since we can make $\tilde{x}_2$ anything and it will satisfy the equations. This occurs when V_2 has no columns, i.e. $\text{null}(A) = \{0\}$. In these cases, the unique solution is $\tilde{x}_1 = \Sigma_1^{-1}\tilde{b}_1$. Returning to the original coordinates, $x = V_1\Sigma_1^{-1}U_1^\mathsf{T} b + V_2\tilde{x}_2 = V_1\Sigma_1^{-1}U_1^\mathsf{T} b + w$, where $w \in \text{null}(A)$.

2 Least Squares with SVD

We seek to solve the minimization problem:

$$\begin{aligned}
\min_x \|Ax - b\|^2 &= \min_x \|U\Sigma V^\top x - b\|^2 \\
&= \min_x \|U(\Sigma V^\top x - U^\top b)\|^2 \\
&= \min_x \|\Sigma V^\top x - U^\top b\|^2 \\
&= \min_x \|\Sigma \tilde{x} - \tilde{b}\|^2 \\
&= \min_x \left\| \begin{bmatrix} \Sigma_1 \tilde{x}_1 - \tilde{b}_1 \\ -\tilde{b}_2 \end{bmatrix} \right\|^2 \\
&= \min_x \|\Sigma_1 \tilde{x}_1 - \tilde{b}_1\|^2 + \|\tilde{b}_2\|^2 \\
\implies \tilde{x}_1 &= \Sigma_1^{-1} \tilde{b}_1
\end{aligned}$$

With the third equality following from U being an orthogonal matrix and thus preserving length. The minimum residual squared is $\|\tilde{b}_2\|^2 = \|U_2^\top b\|^2$. Note that this follows from projecting b onto range(A). We can write b by expanding as

$$b = UU^\top b = \underbrace{U_1 U_1^\top b}_{\in \text{range}(A)} + \underbrace{U_2 U_2^\top b}_{\in \text{range}(A)^\perp}$$

So the norm of the part in the perp space of the range of A is $\|U_2 U_2^\top b\| = \|U_2^\top b\| = \|\tilde{b}_2\|$.

If A is tall and full rank $m > n$:

$$A = \begin{bmatrix} \\ U_1 \\ \\ \end{bmatrix} \begin{bmatrix} \Sigma_1 \end{bmatrix} \begin{bmatrix} V_1^\top \end{bmatrix}$$

Then:

$$\begin{aligned}
(A^\top A)^{-1} A^\top b &= (V_1 \Sigma_1^\top U_1^\top U_1 \Sigma V_1^\top)^{-1} V_1 \Sigma_1 U_1^\top b \\
&= (U_1 \Sigma_1^2 V_1^\top)^{-1} V_1 \Sigma_1 U_1^\top b \\
&= V_1 \Sigma_1^{-2} V_1^\top V_1 \Sigma_1 U_1^\top b \\
&= V_1 \Sigma_1^{-1} U_1^\top b
\end{aligned}$$

So if A is tall and full rank, $A^\dagger = (A^\top A)^{-1} A^\top = V_1 \Sigma_1^{-1} U_1^\top$.

Likewise, we can similarly show that if A is wide and full rank, $A^\dagger = A^\top (AA^\top)^{-1} = V_1 \Sigma_1^{-1} U_1^\top$.

In general (when A is not necessarily full-rank), we use $A^\dagger = V_1 \Sigma_1^{-1} U_1^\top$ as the universal definition of the pseudoinverse. This definition generalizes all the cases discussed above.

3 Eigenvalues and singular values

Eigenvalue decomposition	Singular value decomposition
Matrix must be square	Matrix can be any size
Eigenvalues can be real or complex	Singular values are always real and positive
Eigenvectors can be real complex	Singular vectors are always real
$Ax = \lambda x$	$Av = \sigma u$ and $A^\mathsf{T} u = \sigma v$
In matrix form: $AX = X\Lambda$	In matrix form: $AV = U\Sigma$
X may not be invertible (non-diagonalizable A)	U and V are always orthogonal

Important note: For a square matrix, *invertibility* and *diagonalizability* are unrelated concepts. A matrix is invertible if all of its *eigenvalues* are nonzero. A matrix is diagonalizable if it has a set of *eigenvectors* that are linearly independent. Here are some examples of matrices of each type:

	Diagonalizable	Not diagonalizable
invertible	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$
not invertible	$\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$	$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$

Symmetric matrices. If $A = A^\mathsf{T}$ (A is square and symmetric), then:

(i) All eigenvalues of A are real.

(ii) A is always diagonalizable (we can find linearly independent eigenvectors)

(iii) Even stronger: we can find mutually orthogonal eigenvalues. So A is actually *orthogonally diagonalizable*. This means $A = U\Lambda U^\mathsf{T}$ for some orthogonal matrix U.

Positive definite matrices. Given a symmetric matrix A, the following are equivalent.

(i) All eigenvalues of A are *strictly positive* ($\lambda_i > 0$).

(ii) $x^\mathsf{T} A x > 0$ for all $x \neq 0$.

When either of these properties hold, we say A is *positive definite*, written as $A \succ 0$.

Positive semidefinite matrices. Given a symmetric matrix A, the following are equivalent.

(i) All eigenvalues of A are *nonnegative* ($\lambda_i \geq 0$).

(ii) $x^\mathsf{T} A x \geq 0$ for all x.

When either of these properties hold, we say A is *positive semidefinite*, written as $A \succeq 0$.

We also define negative (semi)definite as follows: $A \prec 0$ means that $-A \succ 0$. We can also write something like $P \succ Q$, which is short-hand for $P - Q \succ 0$. Note that unlike inequalities with real numbers, where $a > 0$ or $a < 0$ or $a = 0$, with symmetric matrices, a fourth option is possible, where e.g. some eigenvalues of A are positive and others are negative. In this case, we say that A is *indefinite*.

The relationship between eigenvalues and quadratic forms can be derived by changing coordinates. If A is symmetric, it is orthogonally diagonalizable, so we can write $A = U\Lambda U^\top$. Then letting $U^\top x = \tilde{x}$, we have:

$$x^\top Ax = x^\top U\Lambda U^\top x = \tilde{x}^\top \Lambda \tilde{x} = \sum_{i=1}^{n} \lambda_i \|\tilde{x}_i\|^2$$

Therefore, positivity of all λ_i is equivalent to positivity of $x^\top Ax$ for all x.

4 SVD and sensitivity

Suppose all numbers have uncertainty ± 0.001:

$$\begin{aligned} 2x = 1 \quad &\implies \quad 0.499 \leq x \leq 0.5008 \\ 0.2x = 1 \quad &\implies \quad 4.97 \leq x \leq 5.03 \\ 0.02x = 1 \quad &\implies \quad 47.65 \leq x \leq 52.7 \\ 0.002x = 1 \quad &\implies \quad 333 \leq x \leq 1001 \end{aligned}$$

Uncertainty grows because we are inverting a number close to zero.

Matrices are more subtle. Consider solving $Ax = b$ in the following settings:

$$\begin{aligned} A = \begin{bmatrix} 1 & 1 \\ 1 & 1.001 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \qquad &\implies \qquad A^{-1}b = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \\ A = \begin{bmatrix} 1 & 1 \\ 1 & 1.001 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 2.001 \end{bmatrix} \qquad &\implies \qquad A^{-1}b = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

In this example, we only changed b a bit (we didn't change the A matrix, which is the one being inverted), and it still had a large effect on the solution $A^{-1}b$.

In general, x and b will be perturbed to $x + \delta x$ and $b + \delta b$, respectively. So our governing equation will actually be $A(x + \delta x) = (b + \delta b)$. We consider two settings:

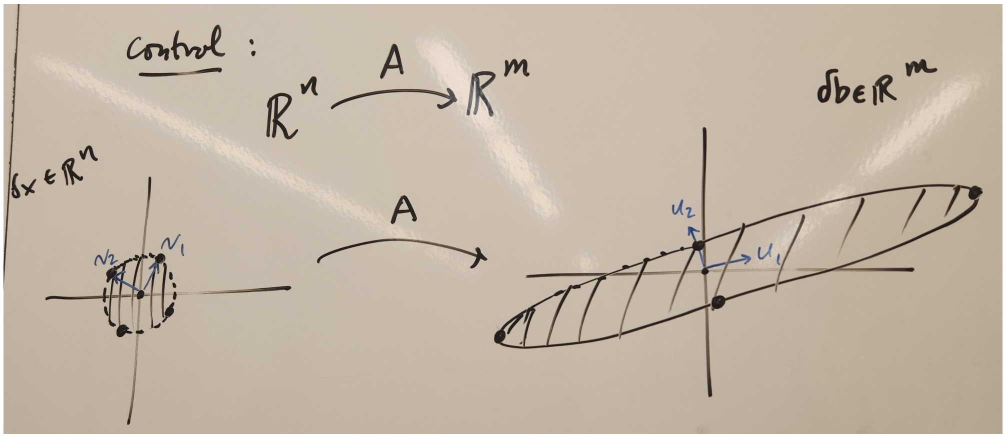
- Control problems: For a given perturbation of the control input δx, how large will the perturbation in the end effect δb be?
- Estimation problems: For a given perturbation δb, how large will the perturbation in the estimate δx be?

Since $Ax = b$ and $A(x + \delta x) = (b + \delta b)$, we have $A\delta x = \delta b$. We can think of A as a map between *perturbations of* x to *perturbations of* b.

Control. In the control setting, we assume we know the perturbation in δx, and we want to know how large the perturbation δb will be. Recall that we have the equations:

$$Av_i = \sigma_i u_i$$

So we should express δx in the basis of the *right singular vectors* $\{v_i\}$. The perturbation observed in δb will then be aligned with the corresponding *left singular vectors* $\{u_i\}$, and amplified according to the corresponding singular value σ_i. For example, if $A \in \mathbb{R}^{2\times 2}$, we get $Av_1 = \sigma_1 u_1$ and $Av_2 = \sigma_2 u_2$, as shown below. The most sensitive perturbation is $\delta x = v_1$, which leads to $\delta b = \sigma_1 u_1$.



In the estimation problem, we instead get that $A\left(\frac{1}{\sigma_i}\right) = u_1$, so we can reverse the above argument, and we see that the perturbations in δx are scaled by the inverses of the corresponding singular values, $\frac{1}{\sigma_i}$. The most sensitive perturbation is $\delta b = u_2$, which leads to $\delta x = \frac{1}{\sigma_2} v_2$.