EECS 16B Designing Information Devices and Systems II
Summer 2020 UC Berkeley

Note 17

1 Overview

In this note, we will be taking a look the **Singular Value Decomposition** or SVD. It is sometimes referred to as the "Swiss Army Knife" or Lienar Algebra. To build the SVD, we will heavily rely on the results from the Spectral Theorem and specifically the matrix $A^T A$.

The SVD lets us look at a matrix A as a weighted sum of rank 1 matrices sometimes referred to as "features." It is an extremely useful tool that is applied in many fields such as Statistics, Image Processing, Machine Learning, and even for Control Systems. At the end of the note, we will take a look at some applications in Image Processing and see how we can use the SVD to control our system with **minimum norm.**

2 Singular Value Decomposition

We will start by defining the SVD and then prove its existence.

2.1 Definition

The Singular Value Decomposition of an $m \times n$ matrix A of rank k is

$$A = U\Sigma V^T = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T \tag{1}$$

$$U = \begin{bmatrix} \vec{u}_1 & \dots & \vec{u}_m \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix} \qquad V = \begin{bmatrix} \vec{v}_1 & \dots & \vec{v}_n \end{bmatrix} \tag{2}$$

The vectors $\vec{u}_i$ are orthonormal and are called the **left singular vectors**. The vectors $\vec{v}_i$ are also orthonormal and are called the **right singular vectors**. The scalars σ_i are the **singular values** of A. For values of i greater than k, $\sigma_i = 0$.

Note that Σ is an $m \times n$ matrix and will change shape based on the shape of A.

A is a tall matrix $m > n$

$$\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix}$$

A is a wide matrix $n > m$

$$\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_m & \end{bmatrix}$$

2.2 Understanding the SVD

The matrix A is a linear transformation that sends vectors in $\mathbb{R}^n$ to $\mathbb{R}^m$. Therefore, the right singular vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ form a basis for $\mathbb{R}^n$ while the right singular vectors $\{\vec{v}_1, \ldots, \vec{v}_m\}$ form a basis for $\mathbb{R}^m$.

In addition to this, the choices of $\vec{v}_i$ and $\vec{u}_i$ are special in that

$$A\vec{v}_i = \sigma_i \vec{u}_i \tag{3}$$

We will now prove the existence of the SVD and its connection to the eigenvectors of the matrix $A^T A$.

2.2.1 Positive Definiteness

The Spectral Theorem tells us that the symmetric matrix $A^T A$ has a set of orthonormal eigenvectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$. We can show that the eigenvalues of $A^T A$ are all greater than or equal to 0.

$$\|A\vec{v}\|^2 = \langle A\vec{v}, A\vec{v}\rangle = \vec{v}^T A^T A\vec{v} = \lambda \vec{v}^T \vec{v} = \lambda \|\vec{v}\|^2 \implies \lambda = \frac{\|A\vec{v}\|^2}{\|\vec{v}\|^2} \geq 0 \tag{4}$$

The last inequality follows from the positive-definiteness of inner products and norms.

2.2.2 Eigenspaces

Now we show the relation between the eigenvectors of $A^T A$ and AA^T. If $\vec{v}$ is an eigenvector of $A^T A$ with nonzero eigenvalue, we show that $\vec{w} = A\vec{v}$ must be an eigenvector of AA^T.

$$A^T A\vec{v} = \lambda \vec{v} \implies A(A^T A\vec{v}) = A(\lambda \vec{v}) \implies AA^T(A\vec{v}) = \lambda(A\vec{v}) \tag{5}$$

If $\vec{v}$ were an eigenvector of zero eigenvalue, then $\vec{v}$ is in the null-space of A since $\mathrm{Nul}(A) = \mathrm{Nul}(A^T A)$.

Now recalling our results from the last section, let's compute the norm of the vector $\vec{w} = A\vec{v}$

$$\|\vec{w}\| = \|A\vec{v}\| = \sqrt{\lambda}\|\vec{v}\| \tag{6}$$

This means if we were to normalize $\vec{w}$ as a unit vector $\vec{u}$, it would follow that

$$\vec{u} = \frac{\vec{w}}{\|\vec{w}\|} = \frac{A\vec{v}}{\sqrt{\lambda}} \implies A\vec{v} = \sqrt{\lambda}\vec{u} \tag{7}$$

As a result, we define $\sigma = \sqrt{\lambda}$ and call a singular value of A.

2.2.3 Summary

If A is an $m \times n$ matrix of rank k, then $A^T A$ will have k orthonormal eigenvectors $\{\vec{v}_1, \ldots, \vec{v}_k\}$. For each of these k eigenvectors, there exist orthonormal eiegenvectors $\{\vec{u}_1, \ldots, \vec{u}_k\}$ of AA^T.

As a result, for $i = 1, \ldots, k$ we can say that

$$A\vec{v}_i = \sigma_i \vec{u}_i \tag{8}$$

the remaining vectors $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ form the $\mathrm{Nul}(A)$ and can be picked through the Gram-Schmidt process. Similarly the remaining vectors $\{\vec{u}_{k+1}, \ldots, \vec{u}_m\}$ form the $\mathrm{Nul}(A^T)$ can also be picked through Gram-Schmidt.

2.3 SVD From the Other Side

While we have defined the left-singular vectors $\vec{u}$ from the right-singular vectors $\vec{v}$, we could have done the entire process from the other side. This perspective will be very useful when computing the SVD.

Starting with an eigenvector $\vec{u}$ of AA^T, we can show that $\vec{w} = A^T\vec{u}$ is an eigenvector of A^TA.

$$AA^T\vec{u} = \lambda\vec{u} \implies A^T(AA^T\vec{u}) = A^T(\lambda\vec{u}) \implies A^TA(A^T\vec{u}) = \lambda(A^T\vec{u}) \tag{9}$$

The norm of $\|\vec{w}\| = \sqrt{\lambda} = \sigma$ and we can again define a relation between $\vec{u}$ and $\vec{v}$.

$$\vec{v} = \frac{\vec{w}}{\|\vec{w}\|} = \frac{A^T\vec{u}}{\sigma} \implies A^T\vec{u} = \sigma\vec{v} \tag{10}$$

2.4 Computing the SVD

We will show two examples of computing the SVD for tall and wide matrices but note that in practice, we will always compute the SVD of large matrices using numerical tools.

2.4.1 Tall Matrix

Let's first look at a 3×2 matrix A and compute its SVD

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Step 1: Compute the symmetric matrix A^TA

$$A^TA = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Step 2: Find orthonormal eigenpairs $(\lambda_i, \vec{v}_i)$ of A^TA for $i = 1, \ldots, k$ and order them from largest to smallest

$$\lambda^2 - 4\lambda + 3 = 0 \implies \lambda_1 = 3, \lambda_2 = 1$$

$$\vec{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Step 3: Compute the singular values $\sigma_i = \sqrt{\lambda_i}$

$$\sigma_1 = \sqrt{3} \qquad \sigma_2 = \sqrt{1}$$

Step 4: Compute the right singular vectors $\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}$ for $i = 1, \ldots, k$.

$$\vec{u}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \sqrt{6}/3 \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \qquad \vec{u}_2 = \frac{1}{1} \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Step 5: Complete the bases U and V through Gram-Schmidt or by computing the appropriate null spaces.

Since $\text{Rank}(A) = 2$ we don't need to add any move vectors to V. However, we will need to add a third vector to U. We can find $\vec{u}_3$ by finding a basis for $\text{Nul}(A^T)$.

$$\vec{u}_3 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

To summarize, the SVD of the matrix A can be written as

$$A = U\Sigma V^T = \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

2.4.2 Wide Matrix

Now let us look at a 2×3 matrix A and compute its SVD

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Step 1: Compute the symmetric matrix AA^T [1]

$$AA^T = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Step 2: Find orthonormal eigenpairs $(\lambda_i, \vec{u}_i)$ of AA^T for $i = 1, \dots, k$ and order them from largest to smallest

$$\lambda^2 - 4\lambda + 3 = 0 \implies \lambda_1 = 3, \lambda_2 = 1$$

$$\vec{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \vec{u}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Step 3: Compute the singular values $\sigma_i = \sqrt{\lambda_i}$

$$\sigma_1 = \sqrt{3} \qquad \sigma_2 = \sqrt{1}$$

Step 4: Compute the left singular vectors $\vec{v}_i = \frac{A^T \vec{u}_i}{\sigma_i}$ for $i = 1, \dots, k$.

$$\vec{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{\sqrt{6}}{3} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \qquad \vec{v}_2 = \frac{1}{1} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Step 5: Complete the bases U and V through Gram-Schmidt or by computing the appropriate null spaces. Since $\text{Rank}(A) = 2$ we don't need to add any move vectors to U. However, we will need to add a third vector

[1] We compute AA^T instead of A^TA since it is a smaller, 2×2 matrix. In general, it will be easier to diagonalize a smaller matrix so we pick A^TA for tall matrices and AA^T for wide matrices

to V. We can find $\vec{v}_3$ by finding a basis for $\text{Nul}(A)$.

$$\vec{v}_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

To summarize, the SVD of the matrix A can be written as

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{\sqrt{6}}{3} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

3 Fundamental Theorem of Linear Algebra

The results from the SVD can be summarized by the **Fundamental Theorem of Linear Algebra** which states that for an $m \times n$ matrix A

$$\text{Col}(A) \perp \text{Nul}(A^T) \tag{11}$$

$$\text{Nul}(A) \perp \text{Col}(A^T) \tag{12}$$

In otherwords, the $\text{Col}(A)$ is orthogonal to the $\text{Nul}(A^T)$ and the $\text{Nul}(A)$ is orthogonal to the $\text{Col}(A^T)$.

3.1 Proof

3.1.1 Basis for Col(A^T)

If A is of rank k, then the first k right-singular vectors $\{\vec{v}_1, \ldots, \vec{v}_k\}$ form a basis for the $\text{Col}(A^T)$. To see this, recall that the right-singular vectors are eigenvectors of $A^T A$.

$$A^T(A\vec{v}_i) = \lambda \vec{v}_i \implies \vec{v}_i \in \text{Col}(A^T) \tag{13}$$

Since $\text{Rank}(A) = \text{Rank}(A^T) = k$ and $\{\vec{v}_1, \ldots, \vec{v}_k\}$ are all in $\text{Col}(A^T)$, they must form a basis.

3.1.2 Dimension of Nul(A)

By the Rank-Nullity Theorem,

$$\text{Rank}(A) + \dim \text{Nul}(A) = n \tag{14}$$

Since $\text{Rank}(A) = k$, it follows that $\dim \text{Nul}(A) = n - k$.

3.1.3 Basis for Nul(A)

The last $n-k$ right-singular vectors $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ are all eigenvectors of eigenvalue 0. Hence, they form a basis for $\text{Nul}(A^T A)$. Since, $\text{Nul}(A^T A) = \text{Nul}(A)$, we can say that $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ forms a basis for $\text{Nul}(A)$.

3.1.4 Orthogonality

From the Spectral Theorem, we can pick an orthonormal set of eigenvectors for the matrix A^TA. Therefore, since all of the vectors $\{\vec{v}_1,\dots,\vec{v}_n\}$ are orthonormal, the individual bases for $\mathrm{Col}(A^T)$ and $\mathrm{Nul}(A)$ must also be orthogonal. Since the bases for two vector spaces are orthogonal, we conclude by saying every vector in $\mathrm{Col}(A^T)$ must be orthogonal to $\mathrm{Nul}(A)$.

We can use a similar argument using the eigenvectors of AA^T to show that $\mathrm{Col}(A) \perp \mathrm{Nul}(A^T)$.

4 Pseudoinverse of a Matrix

Now that we have developed the SVD and some intuition about its relation to the four fundamental subspaces of a matrix A, we will be looking at some applications of the SVD. The first application is the Pseudoinverse of a matrix A.

Given a system of equations $A\vec{x} = \vec{y}$, we have developed multiple ways to solve this problem

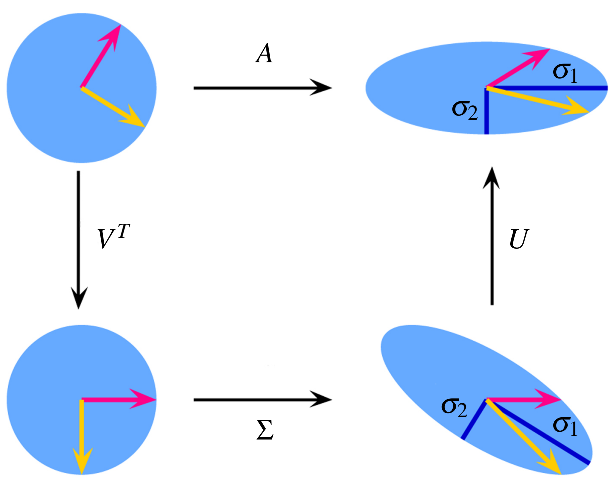
- If A is square and invertible, we can solve $\vec{x} = A^{-1}\vec{y}$
- If A is tall and full rank, we can use Least-Squares to say $\vec{x}^* = (A^TA)A^T\vec{y}$.
- If A is wide and has a spare solution, we can use Orthogonal Matching Pursuit to estimate $\vec{x}$.

We will now develop a solution for $A\vec{x} = \vec{y}$ making no assumptions about the sparsity of the solution. Similar to how we performed matrix inversion, we will define a matrix $A^\dagger$ called the **pseudoinverse**.

4.1 SVD as Rotations

To derive the pseudoinverse, we will decompose the matrix A into a series of rotations and scaling. Since U and V are orthonormal matrices, they do not change the norm of a vector and will only rotate it.

$$
\begin{aligned}
\vec{y} &= A\vec{x} = U\Sigma V^T\vec{x} & &A \text{ sends } \vec{x} \in \mathbb{R}^n \text{ to } \vec{y} \in \mathbb{R}^m. \\
\vec{z} &= V^T\vec{x} & &V^T \text{ rotates the vector } \vec{x} \text{ to a new vector } \vec{z}. \\
\vec{w} &= \Sigma\vec{z} & &\Sigma \text{ scales } \vec{z} \text{ and sends it to a vector } \vec{w} \in \mathbb{R}^m. \\
\vec{y} &= U\vec{w} & &U \text{ rotates the vector } \vec{w} \text{ to a new vector } \vec{y}.
\end{aligned}
$$



4.2 Undoing the Matrices

In order to undo the effect of the matrix A, the plan is to undo each rotation and scaling one by one.

$$\begin{aligned} \vec{y} = A\vec{x} &= U\Sigma V^T\vec{x} && A \text{ sends } \vec{x} \in \mathbb{R}^n \text{ to } \vec{y} \in \mathbb{R}^m. \\ U^T\vec{y} &= \Sigma V^T\vec{x} && \text{Undoing the rotation } U. \\ \Sigma^\dagger U^T\vec{y} &= V^T\vec{x} && \text{“Unscaling” the matrix } \Sigma. \\ \vec{x} &= V\Sigma^\dagger U^T\vec{y} && \text{Undoing the rotation } V^T \end{aligned}$$

The matrices U and V^T are invertible but the matrix Σ is not even square.

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \Large 0 \\ & & \sigma_m & \end{bmatrix}$$

If $\Sigma\vec{z} = \vec{w}$, the best we can do is divide by the singular values where division is possible. Therefore, if A is a wide matrix of rank k, we can undo σ_i for $i = 1, \dots, k$.

$$\sigma_i z_i = w_i \implies z_i = \frac{w_i}{\sigma_i}$$

For the remaining singular values $i > k+1$, we will fill in the matrix with zeros. As a result, we define $\Sigma^\dagger$ as the following

$$\Sigma^\dagger = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ \vdots & \ddots & \dots & 0 \\ 0 & \dots & \frac{1}{\sigma_k} & \vdots \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \tag{15}$$

To summarize, the pseudoinverse of a matrix A can be written as

$$\boxed{A^\dagger = V\Sigma^\dagger U^T} \tag{16}$$

4.3 Compact SVD

Sometimes, we like to write out the SVD in its **compact** form since $\sigma_i = 0$ for $i > k$

$$A = U_c\Sigma_c V_c^T = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T \tag{17}$$

$$U_c = \begin{bmatrix} \vec{u}_1 & \dots & \vec{u}_k \end{bmatrix} \qquad \Sigma_c = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \qquad V_c = \begin{bmatrix} \vec{v}_1 & \dots & \vec{v}_k \end{bmatrix} \tag{18}$$

Here Σ_c is a $k \times k$ diagonal matrix with the singular values on its diagonal.

In fact, we can write the pseudoinverse of a matrix $A^\dagger$ using the compact SVD.

$$\boxed{A^\dagger = V_c\Sigma_c^{-1}U_c^T} \tag{19}$$

4.4 Minimum Norm Property

If the system of equations $A\vec{x} = \vec{y}$ has infinite solutions, we claim the pseudoinverse gives the solution with minimum norm

$$\vec{x} = A^{\dagger}\vec{y} = V_c\Sigma_c^{-1}U_c^T\vec{y} \tag{20}$$

As an optimization problem, we can phrase this as the following[2]

$$\min_{\vec{x}\in\mathbb{R}^n} \|\vec{x}\|^2 \quad \text{subject to } A\vec{x} = \vec{y} \tag{21}$$

We will now prove that the pseudoinverse gives the solution to this optimization problem.

4.5 Proof:

The compact SVD helps us understand the matrix A in terms of its fundamental subspaces. We can write out the matrix A in terms of its SVD but also break it down into its compact and null-space terms.

$$A = U\Sigma V^T = \begin{bmatrix} U_c \end{bmatrix} \begin{bmatrix} \Sigma_c & 0 \end{bmatrix} \begin{bmatrix} V_c^T \\ V_n^T \end{bmatrix} = U_c\Sigma_c V_c^T \tag{22}$$

Here $U = U_c$ since the matrix A has rank m.

The constraint of our optimization problem can we rewritten by expanding out the compact SVD of A.

$$A\vec{x} = \vec{y} \implies U_c\Sigma_c V_c^T\vec{x} = \vec{y} \tag{23}$$

4.5.1 Change of Coordinates

The vectors of the V matrix form an orthonormal basis for $\mathbb{R}^n$. Recall that the last k columns of V form a basis for $\text{Nul}(A)$. Now let $\vec{z}$ be the coordinates of $\vec{x}$ using the basis V. We can break up these coordinates into two components: $\vec{z}_c$ representing the first k coordinates and $\vec{z}_n$ representing the last $n-k$.

$$\vec{x} = V\vec{w} = \begin{bmatrix} V_c & V_n \end{bmatrix} \begin{bmatrix} \vec{z}_c \\ \vec{z}_n \end{bmatrix} = V_c\vec{z}_c + V_n\vec{z}_n \tag{24}$$

Plugging our coordinate representation $\vec{z}$, we see that

$$A\vec{x} = U_c\Sigma_c V_c^T\vec{x} = U_c\Sigma_c V_c^T V_c\vec{w}_c + U_c\Sigma_c V_c^T V_n\vec{w}_n = U_c\Sigma_c\vec{w}_c = \vec{y} \tag{25}$$

The last equality comes from the fact that the columns V_c are orthonormal to V_n.

The objective $\|\vec{x}\|^2$ can be rewritten as follows using the coordinates $\vec{w}$.

$$\|\vec{x}\|^2 = \|V_c\vec{z}_c + V_n\vec{z}_n\|^2 = \|\vec{z}_c\|^2 + \|\vec{z}_n\|^2 \tag{26}$$

Since the constraint $A\vec{x} = \vec{y}$ does not depend on $\vec{z}_n$, we can pick $\vec{z}_n = \vec{0}$. The optimization problem can then be rewritten as

$$\min_{\vec{z}_c\in\mathbb{R}^k} \|\vec{z}_c\|^2 \quad \text{subject to } U_c\Sigma_c\vec{z}_c = \vec{y} \tag{27}$$

Since U_c and Σ_c are invertible matrices, there must be a unique solution $\vec{z}_c = \Sigma_c^{-1}U_c^T\vec{y}$. Lastly, converting back to standard basis coordinates, it follows that

$$\vec{x} = V\vec{z} = \begin{bmatrix} V_c & V_n \end{bmatrix} \begin{bmatrix} \Sigma_c^{-1}U_c^T\vec{y} \\ 0 \end{bmatrix} = V_c\Sigma_c^{-1}U_c^T\vec{y} = A^{\dagger}\vec{y} \tag{28}$$

[2]Minimizing the norm of $\vec{z}$ is equivalent to minimizing the squared norm since norms are positive definite and $f(x) = x^2$ is a monotonic transform.

5 Optimal Control

Now that we have equipped ourselves with the pseudoinverse and have proved its minimum norm property, let us take a look at an application to control systems. Suppose we have a discrete-time system with the following dynamics

$$\vec{x}[n+1] = A\vec{x}[n] + B\vec{u}[n]$$

Let us assume that this system is controllable meaning we can reach any target state $\vec{t} \in \mathbb{R}^n$ in at most n time-steps. While it may be desirable to reach our target $\vec{t}$ as quickly as possible, we may be limited by the physical constraints of the system which prevents high valued inputs.

Therefore, we can try to increase the number of time-steps in order to relax our system's constraints. We will be using the squared norm as a measure of the amount of "energy" it takes to move our system to a target state. By relxaing our system to reach our target in $m > n$ states, we can write out $\vec{x}[m]$ as

$$\begin{aligned}
\vec{x}[m] &= A^n\vec{x}[0] + A^{m-1}B\vec{u}[0] + A^{n-2}B\vec{u}[1] + \ldots + AB\vec{u}[m-2] + B\vec{u}[m-1] && (29)\\
&= A^n\vec{x}[0] + \begin{bmatrix} B & AB & \cdots & A^{m-1}B \end{bmatrix} \begin{bmatrix} \vec{u}[m-1] \\ \vec{u}[m-2] \\ \vdots \\ \vec{u}[0] \end{bmatrix} && (30)\\
&= A^n\vec{x}[0] + H\vec{w} && (31)
\end{aligned}$$

Reaching $\vec{t}$ in m time-steps with minimum energy can be phrased as the following optimization problem

$$\min_{\vec{w} \in \mathbb{R}^{m \cdot p}} \|\vec{w}\|^2 \quad \text{subject to } H\vec{w} = \vec{t} - A^n\vec{x}[0]$$

If the matrix B is of size $n \times p$ then H will be a wide matrix of size $n \times mp$. The system $H\vec{w} = \vec{y}$ will have infinite solutions, but as we saw in the previous section, the minimum norm solution comes from the pseudoinverse.

$$\vec{w}^* = H^\dagger(\vec{t} - A^n\vec{x}[0])$$

5.1 Car Example

Let us take a look at the car model once more represented by following dynamics

$$\frac{d}{dt}\begin{bmatrix} p(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p(t) \\ v(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{M} \end{bmatrix} u(t)$$

Suppose $M = 1\,\text{kg}$ and we discretized our car model with rate $T = 0.1\,\text{s}$. The discretized model would be

$$\begin{bmatrix} p[n+1] \\ v[n+1] \end{bmatrix} = \begin{bmatrix} 1 & 0.1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p[n] \\ v[n] \end{bmatrix} + \begin{bmatrix} 0.005 \\ 0.05 \end{bmatrix}$$

Assuming the car starts at rest $\vec{x}[0] = \vec{0}$ let us try to move the car to $p = 10\,\text{m}$ with no velocity in two time-steps. [3]

$$\vec{x}[2] = \begin{bmatrix} 10 \\ 0 \end{bmatrix} = A^2\vec{x}[0] + ABu[0] + Bu[1] = \begin{bmatrix} B & AB \end{bmatrix} \begin{bmatrix} u[1] \\ u[0] \end{bmatrix}$$

[3] Warning: Do not try this at home. Moving 10 m in 0.02 would be traveling at an average of 1100 miles per hour.

The inputs $u[0]$ and $u[1]$ can be computed through a matrix inverse as follows

$$\vec{w} = \begin{bmatrix} u[1] \\ u[0] \end{bmatrix} = \mathscr{C}^{-1}\vec{x}[2] = \begin{bmatrix} -2000 \\ 2000 \end{bmatrix}$$

To reach our target in two time-steps, we would need to apply forces of 20 kN which would break our car. Therefore, let's try to set up a minimum norm problem to reach our target in ten time-steps.

$$H = \begin{bmatrix} B & AB & \cdots & A^9 B \end{bmatrix} \qquad \vec{w} = \begin{bmatrix} u[9] & u[8] & \cdots & u[0] \end{bmatrix}^T$$
$$\min_{\vec{w}\in\mathbb{R}^{10}} \|\vec{w}\|^2 \quad \text{subject to } H\vec{w} = \vec{t}$$

Again we compute the minimum norm solution by applying the pseudoinverse

$$\vec{w}^* = H^\dagger \vec{t} \qquad \|\vec{w}\| = 220$$

Notice how the norm drops significantly. As we add more time-steps, the norm $\|\vec{w}\|$ will continue to drop off. The results are plotted below.