ECE 586: Vector Space Methods
Chapter 4: Representation and Approximation

Henry D. Pfister
Duke University

November 4th–, 2019

4.1: Best Approximation

Let W be a subspace of a Banach space V and, for any $\underline{v} \in V$, consider finding a vector $\underline{w} \in W$ such that $\|\underline{v} - \underline{w}\|$ is as small as possible.

Definition

The vector $\underline{w} \in W$ is a best approximation of $\underline{v} \in V$ by vectors in W if

$$\|\underline{v} - \underline{w}\| \leq \|\underline{v} - \underline{w}'\|$$

for all $\underline{w}' \in W$.

Example

If W is spanned by the vectors $\underline{w}_1, \ldots, \underline{w}_n \in V$, then we can write

$$\underline{v} = \underline{w} + \underline{e} = s_1 \underline{w}_1 + \cdots + s_n \underline{w}_n + \underline{e},$$

where $\underline{e} = \underline{v} - \underline{w}$ is the approximation error.

Vector Projection Revisited

Let $\underline{u}, \underline{v}$ be vectors in an inner-product space V with inner product $\langle \cdot | \cdot \rangle$.

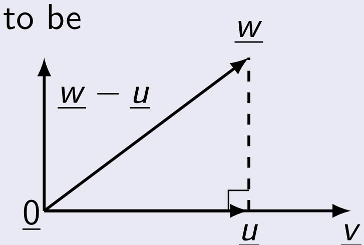
Lemma

If $\langle \underline{w} | \underline{v} \rangle = 0$, then $\|\underline{w} + \underline{v}\|^2 = \|\underline{w}\|^2 + 2\,Re\{\langle \underline{w} | \underline{v} \rangle\} + \|\underline{v}\|^2 = \|\underline{w}\|^2 + \|\underline{v}\|^2$.

Definition (Vector Projection)

The projection of $\underline{w}$ onto $\underline{v}$ is defined to be

$$\underline{u} = \frac{\langle \underline{w} | \underline{v} \rangle}{\|\underline{v}\|^2} \underline{v}$$



Lemma

Let $\underline{u}$ be the projection of $\underline{w}$ onto $\underline{v}$. If $\langle \underline{w} | \underline{v} \rangle \neq 0$, then $\|\underline{w} - \underline{u}\| < \|\underline{w}\|$.

Proof.

$\langle \underline{w} - \underline{u} | \underline{u} \rangle = 0$ implies $\|\underline{w}\|^2 = \|(\underline{w} - \underline{u}) + \underline{u}\|^2 = \|\underline{w} - \underline{u}\|^2 + \|\underline{u}\|^2$. $\square$

4.1: Orthogonal Projection

In an arbitrary Banach space, finding a best approximation can be hard.

But, if the norm $\|\cdot\|$ corresponds to the induced norm of a Hilbert space, then orthogonal projection greatly simplifies the problem.

Theorem (Subspace Projection)

Suppose W is a subspace of a Hilbert space V and $\underline{v} \in V$. Then,

1. *The vector $\underline{w} \in W$ is a best approximation of $\underline{v} \in V$ by vectors in W if and only if $\underline{v} - \underline{w}$ is orthogonal to every vector in W.*
2. *If a best approximation of $\underline{v} \in V$ by vectors in W exists, it is unique.*
3. *If W is a closed subspace with a countable orthogonal basis $\underline{w}_1, \underline{w}_2, \ldots$, then the best approximation of $\underline{v}$ by vectors in W is*

$$\underline{w} = \sum_{i=1}^{\infty} \frac{\langle \underline{v} | \underline{w}_i \rangle}{\|\underline{w}_i\|^2} \underline{w}_i.$$

Note: the implied linear mapping $E\colon V \to W$ defined by $E(\underline{v}) = \underline{w}$ is called the orthogonal projection of V onto W.

Proof on whiteboard.

4.1.1: Projections Without Orthogonality (1)

Definition

A function $F\colon X \to Y$ with $Y \subseteq X$ is idempotent if $F(F(x)) = F(x)$. When F is a linear transformation, this reduces to $F^2 = F \cdot F = F$.

Definition

Let V be a vector space and $T\colon V \to V$ be a linear transformation. If T is idempotent, then T is called a projection because $T\underline{v} = \underline{v}$ if $\underline{v} \in \mathcal{R}(T)$.

Example

The idempotent matrix A is a projection onto the first two coordinates.

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

4.1.1: Projections Without Orthogonality (2)

Theorem

Let V be a vector space and $T\colon V \to V$ be a projection operator. Then, the range $\mathcal{R}(T)$ and the $\mathcal{N}(T)$ are disjoint subspaces of V.

Proof.

For a non-zero $\underline{v} \in \mathcal{R}(T)$, there is a non-zero $\underline{w} \in V$ such that $\underline{v} = T\underline{w}$. Thus, $T\underline{v} = T^2\underline{w} = T\underline{w} = \underline{v} \neq \underline{0}$. But, if $\underline{v} \in \mathcal{N}(T)$ was also true, then one would get the contradiction $T\underline{v} = \underline{0}$. $\square$

Example

Consider the linear transform $T\colon V \to V$ defined by $T = I - P$, where P is a projection. It is easy to verify that T is a projection operator because

$$T^2 = (I-P)(I-P) = I - P - P + P^2 = I - P = T.$$

In fact, T is a projection onto $\mathcal{R}(T) = \mathcal{N}(P)$ because $P\underline{v} = \underline{0}$ (i.e., $\underline{v} \in \mathcal{N}(P)$) if and only if $(I-P)\underline{v} = \underline{v}$ (i.e., $\underline{v} \in \mathcal{R}(T)$).

4.1.1: Orthogonal Projection Operators

Definition

Let V be an inner-product space and $P\colon V \to V$ be a projection operator. If $\mathcal{R}(P) \perp \mathcal{N}(P)$, then P is called an orthogonal projection.

Example

Let V be an inner-product space and $P\colon V \to V$ be an orthogonal projection. Then, $\underline{v} = P\underline{v} + (I-P)\underline{v}$ gives an orthogonal decomposition of $\underline{v}$ because $P\underline{v} \in \mathcal{R}(P)$, $(I-P)\underline{v} \in \mathcal{N}(P)$, and $\mathcal{R}(P) \perp \mathcal{N}(P)$.

Theorem

For $V = F^n$ with standard inner product, P is an orthogonal projection matrix if it is idempotent and Hermitian (i.e. $P^2 = P$ and $P^H = P$).

Proof.

Since $\mathcal{R}(P) = \{P\underline{u} | \underline{u} \in V\}$ and $\mathcal{N}(P) = \{\underline{v} \in V | P\underline{v} = \underline{0}\}$, the general condition is $\langle P\underline{u} | (I-P)\underline{v} \rangle = 0$ for all $\underline{u}, \underline{v} \in V$. Simplifying this gives

$$\underline{v}^H (I-P)^H P \underline{u} = \underline{v}^H (P - P^H P)\underline{u} = \underline{v}^H (P - P^2)\underline{u} = 0. \qquad \square$$

4.2: Normal Equations

Let W be a subspace of a Hilbert space V that is spanned by the linearly independent (but not orthogonal) set of vectors $\underline{w}_1, \ldots, \underline{w}_n \in V$.

The projection theorem shows that $\hat{\underline{v}} \in W$ is the best approximation of $\underline{v} \in V$ if and only if $(\underline{v} - \hat{\underline{v}}) \perp \underline{w}_j$ for $j = 1, \ldots, n$. This implies that

$$\langle \underline{v} - \hat{\underline{v}} | \underline{w}_j \rangle = \left\langle \underline{v} - \sum_{i=1}^{n} s_i \underline{w}_i \middle| \underline{w}_j \right\rangle = 0$$

or, equivalently, the normal equations

$$\sum_{i=1}^{n} s_i \langle \underline{w}_i | \underline{w}_j \rangle = \langle \underline{v} | \underline{w}_j \rangle .$$

The gives a system of n linear equations in n unknowns defined by

$$\underbrace{\begin{bmatrix} \langle \underline{w}_1 | \underline{w}_1 \rangle & \langle \underline{w}_2 | \underline{w}_1 \rangle & \cdots & \langle \underline{w}_n | \underline{w}_1 \rangle \\ \langle \underline{w}_1 | \underline{w}_2 \rangle & \langle \underline{w}_2 | \underline{w}_2 \rangle & \cdots & \langle \underline{w}_n | \underline{w}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \underline{w}_1 | \underline{w}_n \rangle & \langle \underline{w}_2 | \underline{w}_n \rangle & \cdots & \langle \underline{w}_n | \underline{w}_n \rangle \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}}_{\underline{s}} = \underbrace{\begin{bmatrix} \langle \underline{v} | \underline{w}_1 \rangle \\ \langle \underline{v} | \underline{w}_2 \rangle \\ \vdots \\ \langle \underline{v} | \underline{w}_n \rangle \end{bmatrix}}_{\underline{t}} .$$

4.2: The Gramian

Definition

For $\underline{w}_1, \ldots, \underline{w}_n$, the $n \times n$ Gramian matrix is defined to be

$$G = \begin{bmatrix} \langle \underline{w}_1 | \underline{w}_1 \rangle & \langle \underline{w}_2 | \underline{w}_1 \rangle & \cdots & \langle \underline{w}_n | \underline{w}_1 \rangle \\ \langle \underline{w}_1 | \underline{w}_2 \rangle & \langle \underline{w}_2 | \underline{w}_2 \rangle & \cdots & \langle \underline{w}_n | \underline{w}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \underline{w}_1 | \underline{w}_n \rangle & \langle \underline{w}_2 | \underline{w}_n \rangle & \cdots & \langle \underline{w}_n | \underline{w}_n \rangle \end{bmatrix}$$

Since $g_{ij} = \langle \underline{w}_j | \underline{w}_i \rangle$, we see G is Hermitian symmetric (i.e. $G^H = G$).

Definition

A matrix $M \in F^{n \times n}$ is positive-semidefinite if $\underline{v}^H M \underline{v} \geq 0$ for all $\underline{v} \in F^n$.
A matrix $M \in F^{n \times n}$ is positive-definite if $\underline{v}^H M \underline{v} > 0$ for all $\underline{v} \in F^n - \{\underline{0}\}$.

Theorem

A Gramian matrix G is always positive-semidefinite. It is positive-definite if and only if the vectors $\underline{w}_1, \ldots, \underline{w}_n$ is linearly independent.

Proof on whiteboard.

4.3: Least-Squares Solution of a Linear System

For $V = F^m$, let $A \in F^{m \times n}$ be a matrix whose i-th column is $\underline{w}_i \in V$.
Then, a vector $\hat{\underline{v}} \in W = \text{colspace}(A)$ can be written as

$$\hat{\underline{v}} = A\underline{s} = \sum_{i=1}^{n} s_i \underline{w}_i.$$

Also, the best approximation of $\underline{v}$ by vectors in W is found by solving

$$\min_{\hat{\underline{v}} \in W} \|\underline{v} - \hat{\underline{v}}\| = \min_{\underline{s}} \|\underline{v} - A\underline{s}\|.$$

For the induced norm, any solution must satisfy the normal equations

$$\langle \underline{v} - \hat{\underline{v}} | \underline{w}_j \rangle = \langle \underline{v} - A\underline{s} | \underline{w}_j \rangle = 0, \quad j \in [n].$$

For the standard inner product, these equations can be expressed as

$$\underline{0} = \begin{bmatrix} \underline{w}_1^H \\ \vdots \\ \underline{w}_n^H \end{bmatrix} (\underline{v} - A\underline{s}) = A^H \underline{v} - A^H A \underline{s} = \underline{t} - G\underline{s},$$

where $G = A^H A$ is the Gramian and $\underline{t}$ is the cross-correlation vector.

4.3.2: Pseudo-Inverse and Projection

When the vectors $\underline{w}_1, \ldots, \underline{w}_n$ are linearly independent, the Gramian matrix is positive definite and hence invertible. Thus, the optimal solution for the least-squares problem is given by

$$\underline{s} = G^{-1}\underline{t} = \left(A^H A\right)^{-1} A^H \underline{v},$$

where the matrix $\left(A^H A\right)^{-1} A^H$ is the pseudoinverse of A in this case.

Using this, the best approximation of $\underline{v} \in V$ by vectors in W is equal to

$$\underline{\hat{v}} = A\underline{s} = A\left(A^H A\right)^{-1} A^H \underline{v}.$$

The matrix $P = A\left(A^H A\right)^{-1} A^H$ is the projection matrix for the range of A. It defines an orthogonal projection onto the range of A (i.e., the subspace spanned by the columns of A).

4.3.3: Weighted Least-Squares Solution of a Linear System

For the standard inner product with induced Euclidean norm $\|\cdot\|_E$ and any invertible B, consider the weighted least-squares problem

$$\min_{\underline{\hat{v}} \in W} \|B(\underline{v} - \underline{\hat{v}})\|_E = \min_{\underline{s}} \|B(\underline{v} - A\underline{s})\|_E$$

But, $\|B\underline{v}\|_E$ equals the induced norm of the non-standard inner product

$$\langle \underline{u} | \underline{v} \rangle \triangleq \underline{v}^H B^H B \underline{v}.$$

For the non-standard inner product, the normal equations look the same

$$\left\langle \underline{v} - \underline{\hat{v}} | \underline{w}_j \right\rangle = \left\langle \underline{v} - A\underline{s} | \underline{w}_j \right\rangle = 0, \quad j \in [n].$$

but they solve a different problem and they reduce to

$$\underline{0} = \begin{bmatrix} \underline{w}_1^H \\ \vdots \\ \underline{w}_n^H \end{bmatrix} B^H B\left(\underline{v} - A\underline{s}\right) = A^H B^H B \underline{v} - A^H B^H B A \underline{s}$$

4.3.4: Expression for Minimum Approximation Error

Let $\underline{\hat{v}} \in W$ be the best approximation of $\underline{v}$ by vectors in W. Then,

$$\underline{v} = \underline{\hat{v}} + \underline{e},$$

where $\underline{e} \in W^{\perp}$ is the minimum achievable error. The squared norm of the minimum error is given implicitly by

$$\left\|\underline{v}\right\|^2 = \left\|\underline{\hat{v}} + \underline{e}\right\|^2 = \langle \underline{\hat{v}} + \underline{e} | \underline{\hat{v}} + \underline{e} \rangle = \langle \underline{\hat{v}} | \underline{\hat{v}} \rangle + \langle \underline{e} | \underline{e} \rangle = \left\|\underline{\hat{v}}\right\|^2 + \left\|\underline{e}\right\|^2 .$$

For the weighted problem, let $H = B^H B$ and write

$$\begin{aligned}
\left\|\underline{e}\right\|^2 = \left\|\underline{v}\right\|^2 - \left\|\underline{\hat{v}}\right\|^2 &= \underline{v}^H H \underline{v} - \underline{\hat{v}}^H H \underline{\hat{v}} \\
&= \underline{v}^H H \underline{v} - \underline{s}^H A^H H A \underline{s} \\
&= \underline{v}^H H \underline{v} - \underline{v}^H H A \left(A^H H A\right)^{-1} A^H H \underline{v} \\
&= \underline{v}^H \left(H - H A \left(A^H H A\right)^{-1} A^H H\right) \underline{v}.
\end{aligned}$$

4.4.2: Linear Minimum Mean-Squared Error Estimation

Let $Y, X_1, \ldots, X_n$ be zero-mean random variables. Linear minimum mean-squared error (LMMSE) estimation finds $s_1, \ldots, s_n$ such that

$$\hat{Y} = s_1 X_1 + \cdots + s_n X_n$$

minimizes the mean squared-error $\mathrm{E}[|Y - \hat{Y}|^2]$. Using the inner product

$$\langle X | Y \rangle = \mathrm{E}\left[X \overline{Y}\right],$$

the normal equations for the LMMSE estimate $\hat{Y}$ are $G\underline{s} = \underline{t}$, where

$$G = \begin{bmatrix} \mathrm{E}\left[X_1\overline{X}_1\right] & \mathrm{E}\left[X_2\overline{X}_1\right] & \cdots & \mathrm{E}\left[X_n\overline{X}_1\right] \\ \mathrm{E}\left[X_1\overline{X}_2\right] & \mathrm{E}\left[X_2\overline{X}_2\right] & \cdots & \mathrm{E}\left[X_n\overline{X}_2\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}\left[X_1\overline{X}_n\right] & \mathrm{E}\left[X_2\overline{X}_n\right] & \cdots & \mathrm{E}\left[X_n\overline{X}_n\right] \end{bmatrix}, \quad \underline{t} = \begin{bmatrix} \mathrm{E}\left[Y\overline{X}_1\right] \\ \mathrm{E}\left[Y\overline{X}_2\right] \\ \vdots \\ \mathrm{E}\left[Y\overline{X}_n\right] \end{bmatrix}.$$

If the matrix G is invertible, the minimum mean-squared error is given by

$$\left\|Y - \hat{Y}\right\|^2 = \mathrm{E}\left[Y\overline{Y}\right] - \left[\hat{Y}\overline{\hat{Y}}\right] = \mathrm{E}\left[Y\overline{Y}\right] - \underline{t}^H G^{-1} \underline{t}.$$

4.5.1: Dual Approximation and Minimum-Norm Solutions

An underdetermined system of linear equations has an infinite number of solutions. It often makes sense to prefer the minimum-norm solution.

Let V be a Hilbert space and $\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n$ be a basis for subspace W. For any $\underline{v} \in V$, the best approximation of $\underline{v}$ in W can be found by solving

$$\begin{bmatrix} \langle \underline{w}_1 | \underline{w}_1 \rangle & \langle \underline{w}_2 | \underline{w}_1 \rangle & \cdots & \langle \underline{w}_n | \underline{w}_1 \rangle \\ \langle \underline{w}_1 | \underline{w}_2 \rangle & \langle \underline{w}_2 | \underline{w}_2 \rangle & \cdots & \langle \underline{w}_n | \underline{w}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \underline{w}_1 | \underline{w}_n \rangle & \langle \underline{w}_2 | \underline{w}_n \rangle & \cdots & \langle \underline{w}_n | \underline{w}_n \rangle \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} \langle \underline{v} | \underline{w}_1 \rangle \\ \langle \underline{v} | \underline{w}_2 \rangle \\ \vdots \\ \langle \underline{v} | \underline{w}_n \rangle \end{bmatrix}. \quad (1)$$

Theorem

Let V be a Hilbert space and $\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n$ be a basis for $W \subseteq V$. The dual approximation problem is to find the minimum-norm vector $\underline{w} \in V$ satisfying $\langle \underline{w} | \underline{w}_i \rangle = c_i$ for $i = 1, \ldots, n$. Then, the solution $\underline{w}$ satisfies

$$\underline{w} = \sum_{i=1}^{n} s_i \underline{w}_i \in W,$$

where $s_1, s_2, \ldots, s_n$ can be found by solving (1) with $\langle \underline{v} | \underline{w}_i \rangle = c_i$.

4.5.2: Minimum-Norm Solutions

For $A \in \mathbb{C}^{m \times n}$ with $m < n$ and $\underline{v} \in \mathbb{C}^m$, consider the underdetermined linear system $A\underline{s} = \underline{v}$. Then, the dual approximation theorem can be applied to solve the minimum-norm problem

$$\min_{\underline{s}: A\underline{s} = \underline{v}} \|\underline{s}\|.$$

To see this as a dual approximation, we can rewrite the constraint $A\underline{s} = \underline{v}$ as $B^H \underline{s} = \underline{v}$ where $B = A^H$. Then, the theorem concludes that the minimum-norm solution lies in the column space of $B = A^H$.

Using $\underline{s} \in \mathcal{R}(A^H)$, there is a $\underline{t}$ such that $\hat{\underline{s}} = A^H \underline{t}$ and the constraint gives $A(A^H \underline{t}) = \underline{v}$. If the rows of A are linearly independent, then the columns of $B = A^H$ are linearly independent and $(B^H B)^{-1} = (AA^H)^{-1}$ exists.

Thus, the solution $\hat{\underline{s}}$ can be obtained in closed form and is given by

$$\hat{\underline{s}} = A^H \left(AA^H\right)^{-1} \underline{v}.$$

The Four Fundamental Subspaces

Consider a linear transform mapping $\mathbb{R}^n \to \mathbb{R}^m$ represented by $A \in \mathbb{R}^{m \times n}$

- The four fundamental subspaces are: $\mathcal{R}(A)$, $\mathcal{N}(A)$, $\mathcal{R}(A^T)$, $\mathcal{N}(A^T)$
- Recall $A^T \in \mathbb{R}^{n \times m}$ maps $\mathbb{R}^m \to \mathbb{R}^n$ and $\mathcal{R}(A^T)$ is the row space of A
- Notice that $\underline{x} \in \mathbb{R}^n$ is in the null space of A *if and only if*

$$A\underline{x} = \begin{bmatrix} -- \text{ row } 1 -- \\ -- \text{ row } 2 -- \\ \vdots \\ -- \text{ row m } -- \end{bmatrix} \underline{x} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

if and only if all rows are orthogonal to $\underline{x}$ in standard inner product

- Thus, the null space of A is orthogonal to the column space of A^T
- Symmetry: null space of A^T is orthogonal to the column space of A
- In our notation, this means that $\mathcal{N}(A) \perp \mathcal{R}(A^T)$ and $\mathcal{N}(A^T) \perp \mathcal{R}(A)$

The Four Fundamental Subspaces: Linear Equations

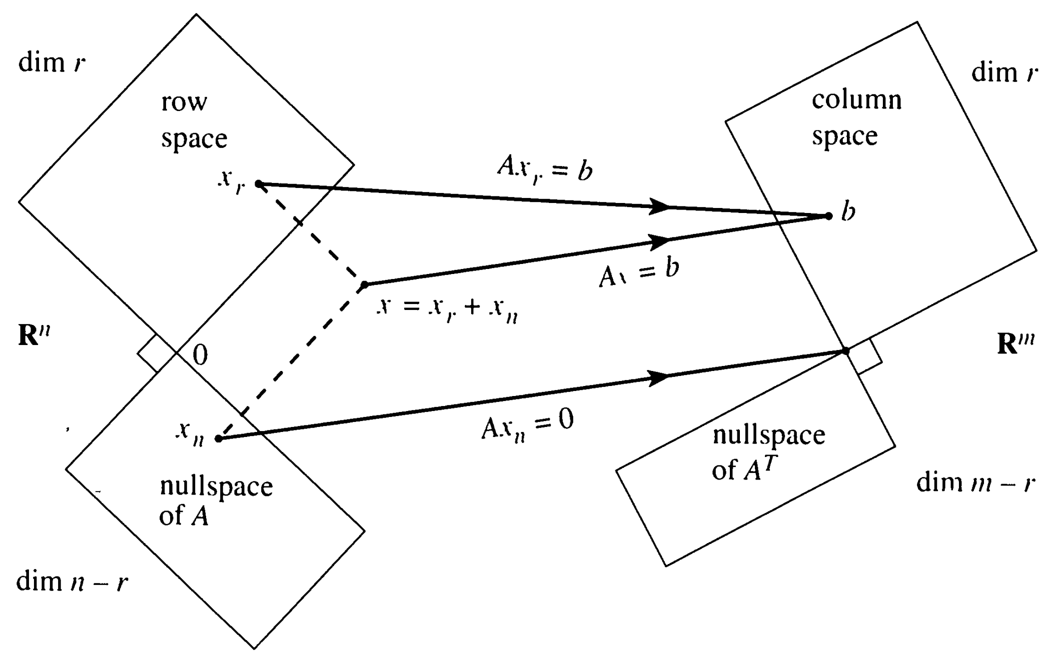


Figure 1. The action of A: Row space to column space, nullspace to zero.

$r \triangleq \dim(\mathcal{R}(A))$ implies $\dim(\mathcal{N}(A)) = n - r$ and $\dim(\mathcal{N}(A^T)) = m - r$

Figure from "The Fundamental Theorem of Linear Algebra" by Gilbert Strang, The American Mathematical Monthly, Nov. 1993

The Four Fundamental Subspaces: Least Squares

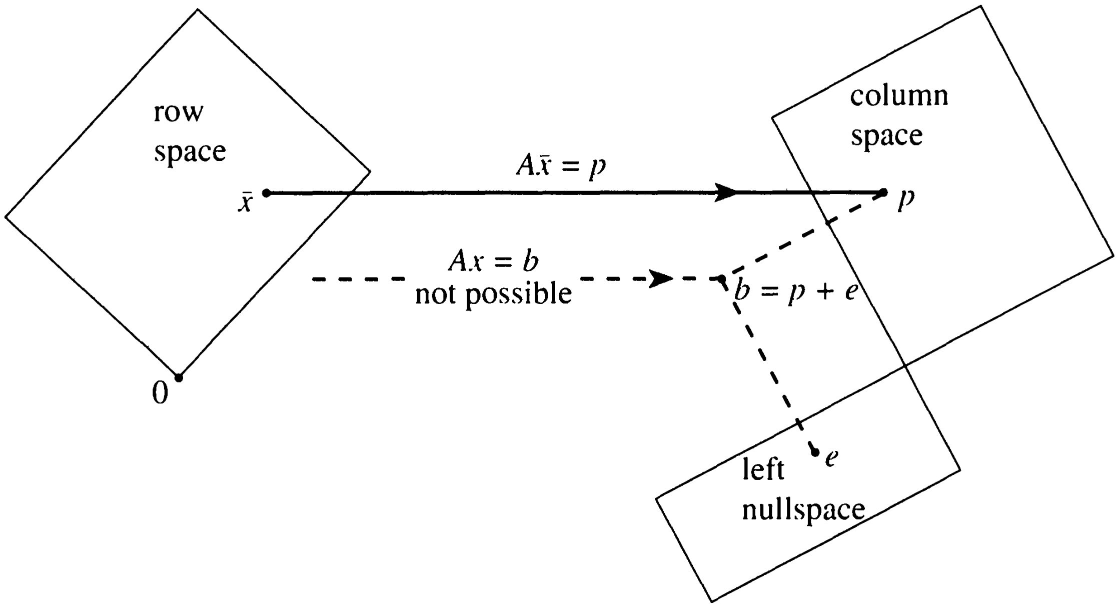


Figure 2. Least squares: $\bar{x}$ minimizes $\|b - Ax\|^2$ by solving $A^T A\bar{x} = A^T b$.

Observe $A^T A \colon \mathbb{R}^n \to \mathbb{R}^n$ is invertible if non-singular (i.e., if $n = r$)

Figure from "The Fundamental Theorem of Linear Algebra" by Gilbert Strang, The American Mathematical Monthly, Nov. 1993

Eigenvalue Decomposition

Definition

Let V be a vector space over F and let $T \colon V \to V$ be a linear operator. An eigenvalue of T is a scalar $\lambda \in F$ such that there exists a non-zero vector $\underline{v} \in V$ with $T\underline{v} = \lambda\underline{v}$. Any vector $\underline{v}$ such that $T\underline{v} = \lambda\underline{v}$ is called an eigenvector of T associated with the eigenvalue value λ.

Definition

The square matrix B is diagonalizable if there is an invertible matrix S (whose columns are eigenvectors) such that $S^{-1}BS = \Lambda$ is diagonal.

Theorem

Any Hermitian matrix B can be diagonalized by a unitary matrix U so that $U^H BU = \Lambda$ is a real-valued diagonal matrix.

Matrices $A^H A$ and AA^H are always Hermitian and positive semidefinite

Singular Value Decomposition (SVD)

Idea is to find orthonormal bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ in which A is diagonal

- Let $\underline{v}_1, \dots, \underline{v}_r$ be orthonormal eigenvectors of $A^H A$ with positive eigenvalues $\sigma_1^2, \dots, \sigma_r^2$. Then,

$$\underline{v}_i^H (A^H A \underline{v}_i) = \sigma_i^2 \underline{v}_i^H \underline{v}_i = \sigma_i^2$$

- This implies that $\|A\underline{v}_i\| = \sigma_i$. So $\underline{u}_i = \frac{1}{\sigma_i} A\underline{v}_i$ has $\|\underline{u}_i\| = 1$ and

$$AA^H \underline{u}_i = \frac{1}{\sigma_i} AA^H A \underline{v}_i = \frac{1}{\sigma_i} \sigma_i^2 A \underline{v}_i = \sigma_i^2 \underline{u}_i$$

- For $U_1 = [\underline{u}_1, \dots, \underline{u}_r]$ and $V_1 = [\underline{v}_1, \dots, \underline{v}_r]$, this gives $AV_1 = U_1 \Sigma_1$ where Σ_1 is a $r \times r$ diagonal matrix with diagonal entries $\sigma_1, \dots, \sigma_r$
- Solving for A gives the compressed SVD

$$A = U_1 \Sigma V_1^H$$

The Four Fundamental Subspaces: Orthogonal Bases

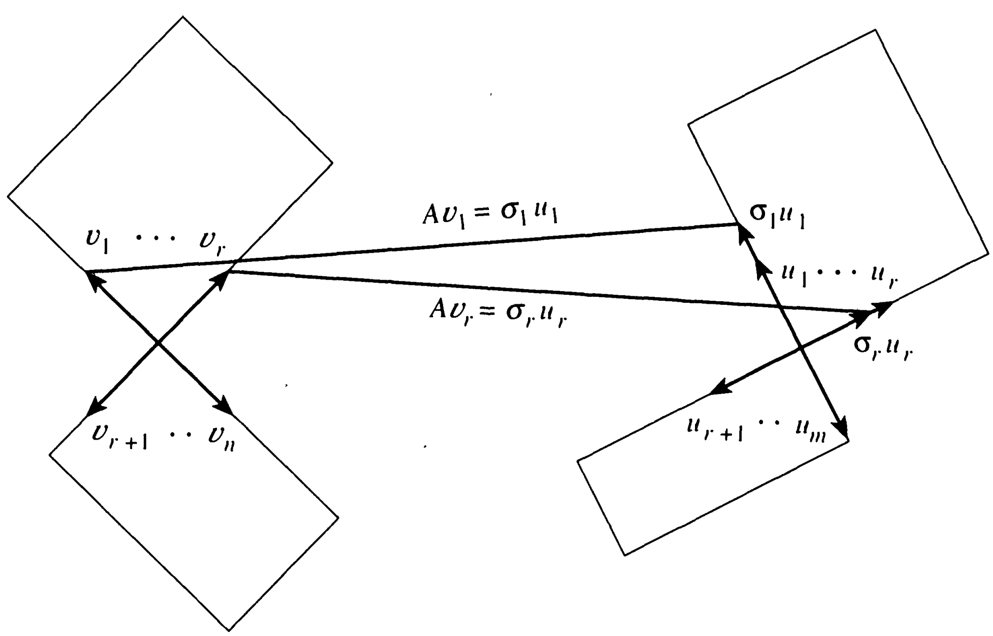


Figure 3. Orthonormal bases that diagonalize A.

For $V = [\underline{v}_1, \dots, \underline{v}_n]$ and $U = [\underline{u}_1, \dots, \underline{u}_m]$, $AV = U\Sigma$ where $\Sigma \in \mathbb{R}^{m \times n}$ has diagonal $\sigma_1, \dots, \sigma_r$. Thus, $A = U\Sigma V^H$.

Figure from "The Fundamental Theorem of Linear Algebra" by Gilbert Strang, The American Mathematical Monthly, Nov. 1993

The Four Fundamental Subspaces: The Pseudo-Inverse

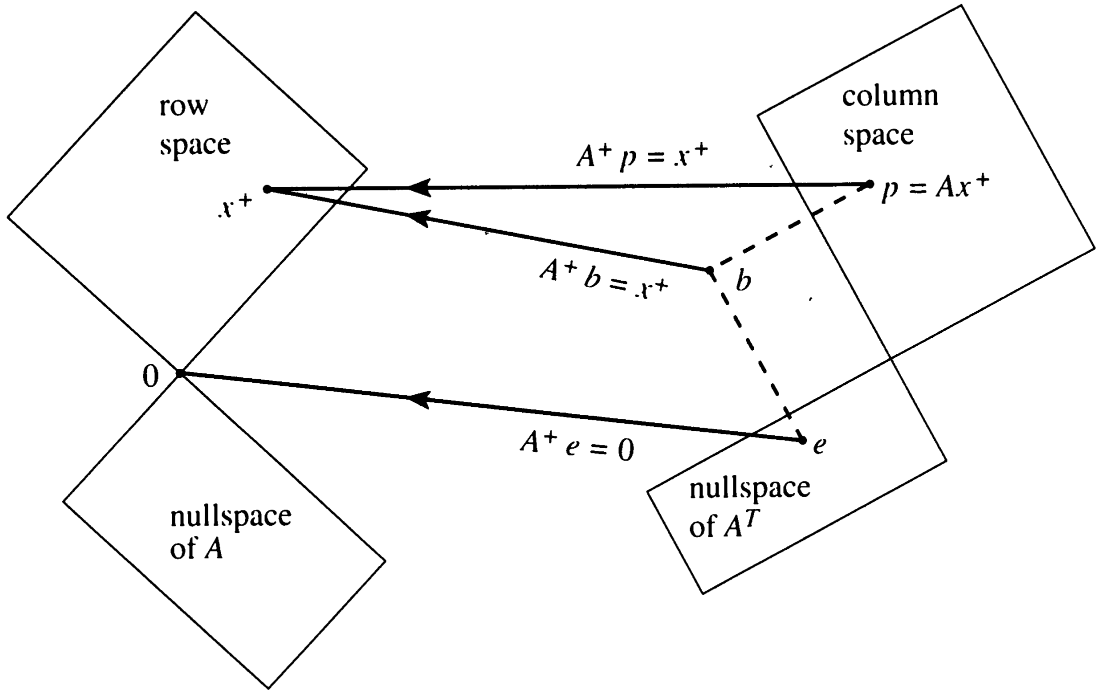


Figure 4. The inverse of A (where possible) is the pseudoinverse A^+.

Figure from "The Fundamental Theorem of Linear Algebra" by Gilbert Strang, The American Mathematical Monthly, Nov. 1993

Singular Value Decomposition Example

Consider the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 5 & -1 \\ -1 & 5 \end{bmatrix}.$$

The eigenvalue decomposition of $A^H A$ is given by

$$A^H A = \begin{bmatrix} 27 & -9 \\ -9 & 27 \end{bmatrix} = V \Lambda V^H = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right) \begin{bmatrix} 36 & 0 \\ 0 & 18 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right)$$

This implies $\Sigma_1 = \Lambda^{1/2}$ and $V_1 = V$. Thus, we find $U_1 = A V_1 \Sigma_1^{-1}$ with

$$U_1 = \begin{bmatrix} 1 & 1 \\ 5 & -1 \\ -1 & 5 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right) \begin{bmatrix} \frac{1}{\sqrt{36}} & 0 \\ 0 & \frac{1}{\sqrt{18}} \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{2}{3} \\ -\frac{1}{\sqrt{2}} & \frac{2}{3} \end{bmatrix}$$

Putting this all together, we have the compressed SVD

$$A = U_1 \Sigma_1 V_1^H = \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{2}{3} \\ -\frac{1}{\sqrt{2}} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \sqrt{36} & 0 \\ 0 & \sqrt{18} \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right).$$