Lesson 2
Least Squares
• Last lecture we reviewed uniform convergence of Fourier series

\[ f(\theta) = \sum_{k=-\infty}^{\infty} \hat{f}_k e^{i k \theta} \]
• Last lecture we reviewed uniform convergence of Fourier series

\[ f(\theta) = \sum_{k=-\infty}^{\infty} \hat{f}_k e^{ik\theta} \]

• We proved that it converges uniformly if

\[ f \in C^2[\mathbb{T}] \]
• Last lecture we reviewed uniform convergence of Fourier series

\[ f(\theta) = \sum_{k=-\infty}^{\infty} \hat{f}_k e^{ik\theta} \]

• We proved that it converges uniformly if

\[ f \in C^2[\mathbb{T}] \]

• We also showed that the convergence rate increases the more differentiable \( f \)

• Namely,

\[ f \in C^{\lambda+2} \Rightarrow \mathcal{F} f \in \ell^1_\lambda \]

\[ \Rightarrow \left\| f - \sum_{k=-n}^{n} \hat{f}_k e^{ik\theta} \right\|_{L_\infty} \leq (n + 2)^{-\lambda} \| \mathcal{F} f \|_{\ell^1_\lambda} \]
Define the $2$-norm

$$\|f\|_2 = \sqrt{\int_{-\pi}^{\pi} |f(\theta)|^2 \, d\theta}$$

over the vector space $L^2[\mathbb{T}]$ of functions with bounded $2$-norm
• Define the 2-norm

\[ \|f\|_2 = \sqrt{\int_{-\pi}^{\pi} |f(\theta)|^2 \, d\theta} \]

over the vector space \( L^2[\mathbb{T}] \) of functions with bounded 2-norm

• In the next few lectures, we will establish that

\[ \left\| f - \sum_{k=-n}^{n} \hat{f}_k e^{ik\theta} \right\|_2 \to 0 \]

for all \( f \in L^2[\mathbb{T}] \)
• Define the 2-norm

\[ \| f \|_2 = \sqrt{\int_{-\pi}^{\pi} |f(\theta)|^2 \, d\theta} \]

over the vector space \( L^2[\mathbb{T}] \) of functions with bounded 2-norm

• In the next few lectures, we will establish that

\[ \| f - \sum_{k=-n}^{n} \hat{f}_k e^{ik\theta} \|_2 \rightarrow 0 \]

for all \( f \in L^2[\mathbb{T}] \)

• Why do we care? After all 2-norm convergence is weaker to \( \infty \)-norm convergence, and the functions we are interested in are smooth, hence have \( \infty \)-norm convergence
• Define the 2-norm

\[ \|f\|_2 = \sqrt{\int_{-\pi}^{\pi} |f(\theta)|^2 \, d\theta} \]

over the vector space \( L^2[\mathbb{T}] \) of functions with bounded 2-norm

• In the next few lectures, we will establish that

\[ \left\| f - \sum_{k=-n}^{n} \hat{f}_k e^{ik\theta} \right\|_2 \to 0 \]

for all \( f \in L^2[\mathbb{T}] \)

• Why do we care? After all 2-norm convergence is weaker to \( \infty \)-norm convergence, and the functions we are interested in are smooth, hence have \( \infty \)-norm convergence

• The answer is that it is much easier to prove convergence of algorithms in 2-norm than \( \infty \)-norm

• This lies on the fact that \( L^2[\mathbb{T}] \) is not just a normed space, it is also a inner product space
\( e^\theta \)

\[ n = 5 \]

Thursday, 1 August 13
$n = 100$

$e^\theta$

Thursday, 1 August 13
\[ e^{\theta} \]
Finite-dimensional least squares
Let

\[ A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \]

be an \( n \times n \) matrix with complex entries (i.e., \( A \in \mathbb{C}^{n \times n} \))
• Let

\[ A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \]

be an \( n \times n \) matrix with complex entries (i.e., \( A \in \mathbb{C}^{n \times n} \)).

- It is helpful to view \( A \) as a row-vector whose columns are in \( \mathbb{C}^n \):

\[ A = (a_1 | \cdots | a_n) \]
• Let

\[
A = \begin{pmatrix}
a_{11} & \cdots & a_{1n} \\
\vdots & \ddots & \vdots \\
a_{n1} & \cdots & a_{nn}
\end{pmatrix}
\]

be an $n \times n$ matrix with complex entries (i.e., $A \in \mathbb{C}^{n \times n}$)

– It is helpful to view $A$ as a row-vector whose columns are in $\mathbb{C}^n$:

\[
A = (a_1 | \cdots | a_n)
\]

• Recall that if $A$ is nonsingular, then we can always solve the linear system

\[
Ac = b, \quad \text{for} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{C}^n
\]
Let
\[
A = \begin{pmatrix}
a_{11} & \cdots & a_{1n} \\
\vdots & \ddots & \vdots \\
a_{n1} & \cdots & a_{nn}
\end{pmatrix}
\]
be an \( n \times n \) matrix with complex entries (i.e., \( A \in \mathbb{C}^{n \times n} \)).

It is helpful to view \( A \) as a row-vector whose columns are in \( \mathbb{C}^n \):
\[
A = (a_1 | \cdots | a_n)
\]

Recall that if \( A \) is nonsingular, then we can always solve the linear system
\[
Ac = b, \quad \text{for} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{C}^n
\]

What if \( A \) is singular? Can we find \( c \) so that \( Ac \) is "close" to \( b \)?
• Let

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

be an $n \times n$ matrix with complex entries (i.e., $A \in \mathbb{C}^{n \times n}$)

− It is helpful to view $A$ as a row-vector whose columns are in $\mathbb{C}^n$: 

$$A = (\begin{array}{c} a_1 \\ \vdots \\ a_n \end{array})$$

• Recall that if $A$ is nonsingular, then we can always solve the linear system

$$Ac = b,$$  \hspace{0.5cm} \text{for} \hspace{0.5cm} b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{C}^n$$

• What if $A$ is singular? Can we find $c$ so that $Ac$ is "close" to $b$?

• In other words, for $c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$, we want

$$c_1 a_1 + \cdots + c_n a_n \approx b$$
• More generally, let $A \in \mathbb{C}^{m \times n}$:

$$A = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \quad \text{for} \quad a_k \in \mathbb{C}^m$$

• Can we **numerically** compute $c_1, \ldots, c_n$ so that

$$c_1 a_1 + \cdots c_n a_n \approx b$$

• More precisely, we find $c$ such that

$$\|Ac - b\|_{\ell^2}$$

takes its minimal value
• Let's review the real $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$

• Minimizing $\|Ac - b\|$ is equivalent to minimizing $\|Ac - b\|^2$
• Let's review the real $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$

• Minimizing $\|Ac - b\|$ is equivalent to minimizing $\|Ac - b\|^2$

• We simplify

$$\|Ac - b\|^2 = (Ac - b)^\top (Ac - b)$$
• Let's review the real $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$

• Minimizing $\|Ac - b\|$ is equivalent to minimizing $\|Ac - b\|^2$

• We simplify

$$\|Ac - b\|^2 = (Ac - b)^T (Ac - b)$$

$$= \|Ac\|^2 - (Ac)^T b - b^T Ac + \|b\|^2$$
• Let's review the real $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$

• Minimizing $\|Ac - b\|$ is equivalent to minimizing $\|Ac - b\|^2$

• We simplify

\[
\|Ac - b\|^2 = (Ac - b)^\top (Ac - b) \\
= \|Ac\|^2 - (Ac)^\top b - b^\top Ac + \|b\|^2 \\
= c^\top A^\top Ac - 2c^\top A^\top b + \|b\|^2
\]
• Let's review the real $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$

• Minimizing $\|Ac - b\|$ is equivalent to minimizing $\|Ac - b\|^2$

• We simplify

$$\|Ac - b\|^2 = (Ac - b)^\top(Ac - b)$$
$$= \|Ac\|^2 - (Ac)^\top b - b^\top Ac + \|b\|^2$$
$$= c^\top A^\top Ac - 2c^\top A^\top b + \|b\|^2$$

• We can heuristically assume that the minimum is a stationary point of this equation; i.e., we want

$$0 = \nabla_c \|Ac - b\|^2$$
Theorem: Suppose $A$ has linearly independent columns. The vector $c = (A^T A)^{-1} A^T b$ is the unique minimizer of

$$\|Ac - b\|$$
\textbf{Theorem}: Suppose \( A \) has linearly independent columns. The vector \( c = (A^T A)^{-1} A^T b \) is the unique minimizer of
\[
\| Ac - b \|
\]

\textbf{Proof}:

- We first remark that \( A^T A \) is positive definite, i.e., \( x^T A^T A x > 0 \) for all (real) \( x \).
  (Why?)
**Theorem:** Suppose $A$ has linearly independent columns. The vector $c = (A^T A)^{-1} A^T b$ is the unique minimizer of 
\[ \|Ac - b\| \]

**Proof:**

- We first remark that $A^T A$ is positive definite, i.e., $x^T A^T A x > 0$ for all (real) $x$. (Why?)

- Minimizing $\|Ac - b\|$ is equivalent to minimizing $\|Ac - b\|^2$.

- For all $x$, we have
  \[ \|A(c + x) - b\|^2 = (c + x)^T A^T A(c + x) - 2(c + x)^T A^T b + \|b\|^2 \]
**Theorem:** Suppose $A$ has linearly independent columns. The vector $c = (A^T A)^{-1} A^T b$ is the unique minimizer of

$$\|Ac - b\|$$

**Proof:**

- We first remark that $A^T A$ is positive definite, i.e., $x^T A^T Ax > 0$ for all (real) $x$. (Why?)
- Minimizing $\|Ac - b\|$ is equivalent to minimizing $\|Ac - b\|^2$.
- For all $x$, we have

$$\|A(c + x) - b\|^2 = (c + x)^T A^T A(c + x) - 2(c + x)^T A^T b + \|b\|^2$$

$$= c^T A^T A(c + x) + x^T A^T A(c + x) - 2(c + x)^T b + \|b\|^2$$
**Theorem:** Suppose $A$ has linearly independent columns. The vector $c = (A^TA)^{-1}A^Tb$ is the unique minimizer of

$$\|Ac - b\|$$

**Proof:**

- We first remark that $A^TA$ is positive definite, i.e., $x^TA^TAx > 0$ for all (real) $x$. (Why?)

- Minimizing $\|Ac - b\|$ is equivalent to minimizing $\|Ac - b\|^2$.

- For all $x$, we have

$$\|A(c + x) - b\|^2 = (c + x)^TA^TA(c + x) - 2(c + x)^TA^Tb + \|b\|^2$$

$$= c^TA^TA(c + x) + x^TA^TA(c + x) - 2(c + x)^Tb + \|b\|^2$$

$$= x^TA^TAx + c^TA^TAc + 2x^TA^TAc - 2(c + x)^TA^Tb + \|b\|^2$$
**Theorem:** Suppose $A$ has linearly independent columns. The vector $c = (A^T A)^{-1} A^T b$ is the unique minimizer of

$$\|Ac - b\|$$

**Proof:**

- We first remark that $A^T A$ is positive definite, i.e., $x^T A^T A x > 0$ for all (real) $x$. (Why?)
- Minimizing $\|Ac - b\|$ is equivalent to minimizing $\|Ac - b\|^2$.
- For all $x$, we have

$$\|A(c + x) - b\|^2 = (c + x)^T A^T A(c + x) - 2(c + x)^T A^T b + \|b\|^2$$

$$= c^T A^T A(c + x) + x^T A^T A(c + x) - 2(c + x)^T A^T b + \|b\|^2$$

$$= x^T A^T A x + c^T A^T A c + 2x^T A^T A c - 2(c + x)^T A^T b + \|b\|^2$$

$$= x^T A^T A x + c^T A^T A c + 2x^T A^T A (A^T A)^{-1} A^T b - 2(c + x)^T A^T b + \|b\|^2$$
**Theorem:** Suppose $A$ has linearly independent columns. The vector $c = (A^\top A)^{-1} A^\top b$ is the unique minimizer of

$$\|A c - b\|$$

**Proof:**

- We first remark that $A^\top A$ is positive definite, i.e., $x^\top A^\top A x > 0$ for all (real) $x$. (Why?)

- Minimizing $\|A c - b\|$ is equivalent to minimizing $\|A c - b\|^2$.

- For all $x$, we have

\[
\|A(c + x) - b\|^2 = (c + x)^\top A^\top A(c + x) - 2(c + x)^\top A^\top b + \|b\|^2
\]

\[
= c^\top A^\top A(c + x) + x^\top A^\top A(c + x) - 2(c + x)^\top A^\top b + \|b\|^2
\]

\[
= x^\top A^\top A x + c^\top A^\top A c + 2x^\top A^\top A c - 2(c + x)^\top A^\top b + \|b\|^2
\]

\[
= x^\top A^\top A x + c^\top A^\top A c + 2x^\top A^\top A (A^\top A)^{-1} A^\top b - 2(c + x)^\top A^\top b + \|b\|^2
\]

\[
= x^\top A^\top A x + c^\top A^\top A c + 2x^\top A^\top b - 2(c + x)^\top A^\top b + \|b\|^2
\]
**Theorem:** Suppose $A$ has linearly independent columns. The vector $c = (A^T A)^{-1} A^T b$ is the unique minimizer of

$$\|A c - b\|$$

**Proof:**

- We first remark that $A^T A$ is positive definite, i.e., $x^T A^T A x > 0$ for all (real) $x$. (Why?)

- Minimizing $\|A c - b\|$ is equivalent to minimizing $\|A c - b\|^2$.

- For all $x$, we have

$$\|A(c + x) - b\|^2 = (c + x)^T A^T A(c + x) - 2(c + x)^T A^T b + \|b\|^2$$

$$= c^T A^T A(c + x) + x^T A^T A(c + x) - 2(c + x)^T A^T b + \|b\|^2$$

$$= x^T A^T A x + c^T A^T A c + 2x^T A^T A c - 2(c + x)^T A^T b + \|b\|^2$$

$$= x^T A^T A x + c^T A^T A c + 2x^T A^T A (A^T A)^{-1} A^T b - 2(c + x)^T A^T b + \|b\|^2$$

$$= x^T A^T A x + c^T A^T A c + 2x^T A^T b - 2(c + x)^T A^T b + \|b\|^2$$

$$= x^T A^T A x + c^T A^T A c - 2c^T A^T b + \|b\|^2$$
\[ \|Ac - b\| \]

**Proof:**

- We first remark that \(A^TA\) is positive definite, i.e., \(x^TA^TAx > 0\) for all (real) \(x\). (Why?)
- Minimizing \(\|Ac - b\|\) is equivalent to minimizing \(\|Ac - b\|^2\).
- For all \(x\), we have

\[
\|A(c + x) - b\|^2 = (c + x)^TA^TA(c + x) - 2(c + x)^TA^Tb + \|b\|^2 \\
= c^TA^TA(c + x) + x^TA^TA(c + x) - 2(c + x)^TA^Tb + \|b\|^2 \\
= x^TA^TAx + c^TA^TAc + 2x^TA^TAc - 2(c + x)^TA^Tb + \|b\|^2 \\
= x^TA^TAx + c^TA^TAc + 2x^TA^Tb - 2(c + x)^TA^Tb + \|b\|^2 \\
= x^TA^TAx + c^TA^TAc - 2c^TA^Tb + \|b\|^2
\]

Minimized when \(x\) is zero! Independent of \(x\)!
General inner product spaces
Normed space

• Recall: a norm $\|f\|$ on a vector space $V$ over the complex numbers satisfies, for $f, g \in V$ and $c \in \mathbb{C}$:
  
  $- \|cf\| = |c| \|f\|

  $- \|f + g\| \leq \|f\| + \|g\|

  $- \text{If } \|f\| = 0 \text{ then } f = 0$

• With a norm attached, $V$ is referred to as a normed space

• Exercise: verify the spaces $\ell_\chi^p$ and $L^p$ are normed spaces
• Let $V$ be a vector space, such as $\mathbb{C}^n$, over the field of complex numbers

• Let $\langle u, v \rangle$ denote an inner product defined for $u, v \in V$, turning it into an inner product space
• Let $V$ be a vector space, such as $\mathbb{C}^n$, over the field of complex numbers

• Let $\langle u, v \rangle$ denote an inner product defined for $u, v \in V$, turning it into an inner product space

• For example, on $\mathbb{C}^n$ the usual inner product is $\langle u, v \rangle = u^* v$

• On $\mathbb{R}^n$ (whose field is the real numbers) it is the dot product $\langle u, v \rangle = u \cdot v = u^T v$
• Let \( V \) be a vector space, such as \( \mathbb{C}^n \), over the field of complex numbers

• Let \( \langle u, v \rangle \) denote an inner product defined for \( u, v \in V \), turning it into an inner product space

• For example, on \( \mathbb{C}^n \) the usual inner product is \( \langle u, v \rangle = u^*v \)

• On \( \mathbb{R}^n \) (whose field is the real numbers) it is the dot product \( \langle u, v \rangle = u \cdot v = u^T v \)

• More generally, an inner product is a function of two vectors in \( V \) that satisfies (for \( u, v, w \in V \) and \( c \in \mathbb{C} \)):

  - Conjugate symmetry: \( \langle u, v \rangle = \overline{\langle u, v \rangle} \)
  - Linearity: \( \langle u, cv \rangle = c \langle u, v \rangle \) and \( \langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle \)
  - Positive definiteness: \( \|u\| \geq 0 \), with \( \|u\| = 0 \) only if \( u = 0 \)
• Let $V$ be a vector space, such as $\mathbb{C}^n$, over the field of complex numbers

• Let $\langle u, v \rangle$ denote an inner product defined for $u, v \in V$, turning it into an inner product space

• For example, on $\mathbb{C}^n$ the usual inner product is $\langle u, v \rangle = u^* v$

• On $\mathbb{R}^n$ (whose field is the real numbers) it is the dot product $\langle u, v \rangle = u \cdot v = u^T v$

• More generally, an inner product is a function of two vectors in $V$ that satisfies (for $u, v, w \in V$ and $c \in \mathbb{C}$):
  
  - Conjugate symmetry: $\langle u, v \rangle = \overline{\langle u, v \rangle}$
  - Linearity: $\langle u, cv \rangle = c \langle u, v \rangle$ and $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$
  - Positive definiteness: $\|u\| \geq 0$, with $\|u\| = 0$ only if $u = 0$

• Here, $\|u\| = \sqrt{\langle u, u \rangle}$ is the norm associated with the inner product, so $V$ is also a normed space
• Consider a row vector of elements $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$:

$$A = ( a_1 | \cdots | a_n )$$
• Consider a row vector of elements $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$:

$$A = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$$

• We can associate with $A$ a **Gram matrix**

$$K = \begin{pmatrix}
\langle a_1, a_1 \rangle & \cdots & \langle a_1, a_n \rangle \\
\vdots & \ddots & \vdots \\
\langle a_1, a_n \rangle & \cdots & \langle a_n, a_n \rangle
\end{pmatrix}$$
• Consider a row vector of elements $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$:

$$A = (a_1 | \cdots | a_n)$$

• We can associate with $A$ a **Gram matrix**

$$K = \begin{pmatrix} \langle a_1, a_1 \rangle & \cdots & \langle a_1, a_n \rangle \\ \vdots & \ddots & \vdots \\ \langle a_1, a_n \rangle & \cdots & \langle a_n, a_n \rangle \end{pmatrix}$$

• In the case where $V = \mathbb{R}^m$, the Gram matrix is precisely the matrix we used in least squares

$$K = A^T A$$

• In the case where $V = \mathbb{C}^m$, we get the similar

$$K = A^* A$$
**Lemma:** The Gram matrix $K$ is Hermitian: $K^* = K$. 
**Lemma**: The Gram matrix $K$ is Hermitian: $K^* = K$.

**Proof**:

- Follows from the fact that $\langle u, v \rangle = \overline{\langle u, v \rangle}$:
Lemma: The Gram matrix $K$ is Hermitian: $K^* = K$.

Proof:

- Follows from the fact that $\langle u, v \rangle = \overline{\langle u, v \rangle}$:

$$
K^* = \begin{pmatrix}
\langle a_1, a_1 \rangle & \cdots & \overline{\langle a_n, a_1 \rangle} \\
\vdots & \ddots & \vdots \\
\overline{\langle a_1, a_n \rangle} & \cdots & \langle a_n, a_n \rangle
\end{pmatrix}
$$
**Lemma**: The Gram matrix $K$ is Hermitian: $K^* = K$.

**Proof**: 

- Follows from the fact that $\langle u, v \rangle = \overline{\langle u, v \rangle}$:

$$K^* = \begin{pmatrix}
\langle a_1, a_1 \rangle & \cdots & \langle a_n, a_1 \rangle \\
\vdots & \ddots & \vdots \\
\langle a_1, a_n \rangle & \cdots & \langle a_n, a_n \rangle 
\end{pmatrix} = \begin{pmatrix}
\overline{\langle a_1, a_1 \rangle} & \cdots & \overline{\langle a_n, a_1 \rangle} \\
\vdots & \ddots & \vdots \\
\overline{\langle a_1, a_n \rangle} & \cdots & \overline{\langle a_n, a_n \rangle} 
\end{pmatrix} = K$$
**Lemma:** Let $A = (a_1 | \cdots | a_n)$ and $K$ denote the associated Gram matrix. For $x, y \in \mathbb{C}^n$ we have

$$\langle Ay, Ax \rangle = y^* K x.$$
**Lemma**: Let $A = (a_1 \mid \cdots \mid a_n)$ and $K$ denote the associated Gram matrix. For $x, y \in \mathbb{C}^n$ we have

$$\langle Ay, Ax \rangle = y^* K x.$$ 

**Proof**: 

$$y^* K x = y^* \begin{pmatrix} \langle a_1, a_1 \rangle & \cdots & \langle a_1, a_n \rangle \\ \vdots & \ddots & \vdots \\ \langle a_n, a_1 \rangle & \cdots & \langle a_n, a_n \rangle \end{pmatrix} x$$
Lemma: Let $A = (a_1 | \cdots | a_n)$ and $K$ denote the associated Gram matrix. For $x, y \in \mathbb{C}^n$ we have

$$\langle Ay, Ax \rangle = y^* K x.$$ 

Proof:

$$y^* K x = y^* \begin{pmatrix} \langle a_1, a_1 \rangle & \cdots & \langle a_1, a_n \rangle \\ \vdots & \ddots & \vdots \\ \langle a_n, a_1 \rangle & \cdots & \langle a_n, a_n \rangle \end{pmatrix} x$$

$$= y^* \begin{pmatrix} \langle a_1, a_1 x_1 + \cdots + a_n x_n \rangle \\ \vdots \\ \langle a_n, a_1 x_1 + \cdots + a_n x_n \rangle \end{pmatrix}$$
Lemma: Let $A = (a_1 | \cdots | a_n)$ and $K$ denote the associated Gram matrix. For $x, y \in \mathbb{C}^n$ we have

$$\langle Ay, Ax \rangle = y^* K x.$$  

Proof:

$$y^* K x = y^* \begin{pmatrix} 
\langle a_1, a_1 \rangle & \cdots & \langle a_1, a_n \rangle \\
\vdots & \ddots & \vdots \\
\langle a_n, a_1 \rangle & \cdots & \langle a_n, a_n \rangle
\end{pmatrix} x$$

$$= y^* \begin{pmatrix} 
\langle a_1, a_1 x_1 + \cdots + a_n x_n \rangle \\
\vdots \\
\langle a_n, a_1 x_1 + \cdots + a_n x_n \rangle
\end{pmatrix}$$

$$= \langle a_1 y_1 + \cdots + a_n y_n, a_1 x_1 + \cdots + a_n x_n \rangle$$

$$= \langle Ay, Ax \rangle$$
Lemma: The Gram matrix $K$ is positive semi-definite. If the vectors $a_1, \ldots, a_n$ are linearly independent, then the Gram matrix is positive definite.
Lemma: The Gram matrix $K$ is positive semi-definite. If the vectors $a_1, \ldots, a_n$ are linearly independent, then the Gram matrix is positive definite.

Proof:

\[ x^* K x = \langle Ax, Ax \rangle = \|Ax\|^2 \geq 0 \]
**Lemma:** The Gram matrix $K$ is positive semi-definite. If the vectors $a_1, \ldots, a_n$ are linearly independent, then the Gram matrix is positive definite.

**Proof:**

- $$x^* K x = \langle Ax, Ax \rangle = \|Ax\|^2 \geq 0$$

- Linear independence of $a_1, \ldots, a_n$ shows that $Ax = a_1 x_1 + \cdots + a_n x_n = 0$ if and only if $x = 0$
Two more useful relationships:
Two more useful relationships:

\[
\langle Ax, b \rangle = \left\langle \sum_{k=1}^{n} x_k a_k, b \right\rangle = \sum_{k=1}^{n} \bar{x}_k \langle a_k, b \rangle = x^* \left( \begin{array}{c}
\langle a_1, b \rangle \\
\vdots \\
\langle a_n, b \rangle 
\end{array} \right)
\]
Two more useful relationships:

\[
\langle Ax, b \rangle = \left\langle \sum_{k=1}^{n} x_k a_k, b \right\rangle = \sum_{k=1}^{n} \bar{x}_k \langle a_k, b \rangle = x^* \begin{pmatrix} \langle a_1, b \rangle \\ \vdots \\ \langle a_n, b \rangle \end{pmatrix}
\]

\[
\langle b, Ax \rangle = \left\langle b, \sum_{k=1}^{n} x_k a_k \right\rangle = \sum_{k=1}^{n} x_k \langle b, a_k \rangle = \sum_{k=1}^{n} x_k \langle a_k, b \rangle = \left( \begin{pmatrix} \langle a_1, b \rangle \\ \vdots \\ \langle a_n, b \rangle \end{pmatrix} \right)^* x
\]
**Theorem:** Suppose the columns of $A = (a_1 | \ldots | a_n)$ are linearly independent in an inner product space $V$. Let $K$ be the associated Gram matrix. The vector

$$c = K^{-1} \begin{pmatrix} \langle a_1, b \rangle \\ \vdots \\ \langle a_n, b \rangle \end{pmatrix}$$

is the unique minimizer of

$$\| Ac - b \|$$

(where the norm is the norm associated with the inner product)

**Proof:**
Orthonormal vectors and calculating least squares approximations
A set of nonzero vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are called *orthogonal* if

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \text{ whenever } i \neq k.$$ 

They are called *orthonormal* if they are orthogonal and all vectors are of unit norm:

$$1 = ||\mathbf{v}_i||,$$ or equivalently, $$\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1.$$

The Gram matrix of orthonormal vectors is the identity!

$$K = \begin{pmatrix} 
\langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_1, \mathbf{v}_n \rangle \\
\vdots & \ddots & \vdots \\
\langle \mathbf{v}_n, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_n, \mathbf{v}_n \rangle 
\end{pmatrix} = \begin{pmatrix} 
1 & \cdots & 1 \\
\vdots & \ddots & \vdots \\
1 & \cdots & 1 
\end{pmatrix} = I$$