MATH 532: Linear Algebra

Chapter 5: Norms, Inner Products and Orthogonality

Greg Fasshauer

Department of Applied Mathematics
Illinois Institute of Technology

Spring 2015

Outline

Vector Norms[0]

Definition

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ ($\mathbb{C}^n$). Then

$$\boldsymbol{x}^T\boldsymbol{y} = \sum_{i=1}^n x_i y_i \quad \in \mathbb{R}$$

$$\boldsymbol{x}^*\boldsymbol{y} = \sum_{i=1}^n \bar{x}_i y_i \quad \in \mathbb{C}$$

is called the standard inner product for $\mathbb{R}^n$ ($\mathbb{C}^n$).

Definition

Let $\mathcal{V}$ be a vector space. A function $\|\cdot\|: \mathcal{V} \to \mathbb{R}_{\geq 0}$ is called a norm provided for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$ and $\alpha \in \mathbb{R}$

1. $\|\boldsymbol{x}\| \geq 0$ and $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$,
2. $\|\alpha \boldsymbol{x}\| = |\alpha| \, \|\boldsymbol{x}\|$,
3. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

Remark

The inequality in (3) is known as the triangle inequality.

Remark

- *Any inner product $\langle \cdot, \cdot \rangle$ induces a norm via (more later)*

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}.$$

- *We will show that the standard inner product induces the Euclidean norm (cf. length of a vector).*

Remark

Inner products let us define angles via

$$\cos\theta = \frac{\boldsymbol{x}^T \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}.$$

In particular, $\boldsymbol{x}, \boldsymbol{y}$ are orthogonal if and only if $\boldsymbol{x}^T \boldsymbol{y} = 0$.

Example

Let $\boldsymbol{x} \in \mathbb{R}^n$ and consider the Euclidean norm

$$\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^T \boldsymbol{x}} = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

We show that $\| \cdot \|_2$ is a norm. We do this for the real case, but the complex case goes analogously.

1. Clearly, $\|\boldsymbol{x}\|_2 \geq 0$. Also,

$$\begin{aligned} \|\boldsymbol{x}\|_2 = 0 \quad &\iff \quad \|\boldsymbol{x}\|_2^2 = 0 \\ \iff \quad \sum_{i=1}^{n} x_i^2 = 0 \quad &\iff \quad x_i = 0, \ i = 1, \ldots, n, \\ \iff \quad \boldsymbol{x} = 0. \end{aligned}$$

Example (cont.)

2. We have

$$\|\alpha \boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} (\alpha x_i)^2\right)^{1/2} = |\alpha| \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} = |\alpha| \, \|\boldsymbol{x}\|_2.$$

3. To establish (3) we need

Lemma

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Then

$$|\boldsymbol{x}^T \boldsymbol{y}| \le \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2. \quad \textit{(Cauchy–Schwarz–Bunyakovsky)}$$

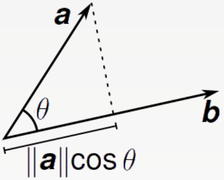
Moreover, equality holds if and only if $\boldsymbol{y} = \alpha \boldsymbol{x}$ with

$$\alpha = \frac{\boldsymbol{x}^T \boldsymbol{y}}{\|\boldsymbol{x}\|_2^2}.$$

Motivation for Proof of Cauchy–Schwarz–Bunyakovsky

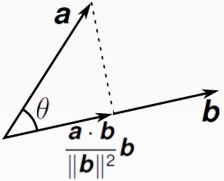
As already alluded to above, the angle θ between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is related to the inner product by

$$\cos\theta = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}.$$



Using trigonometry as in the figure, the projection of $\boldsymbol{a}$ onto $\boldsymbol{b}$ is then

$$\|\boldsymbol{a}\|\cos\theta\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|} = \|\boldsymbol{a}\|\frac{\boldsymbol{a}^T\boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|} = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\|\boldsymbol{b}\|^2}\boldsymbol{b}.$$



Now, we let $\boldsymbol{y} = \boldsymbol{a}$ and $\boldsymbol{x} = \boldsymbol{b}$, so that the projection of $\boldsymbol{y}$ onto $\boldsymbol{x}$ is given by

$$\alpha\boldsymbol{x}, \qquad \text{where } \alpha = \frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{x}\|^2}.$$

Proof of Cauchy–Schwarz–Bunyakovsky

We know that $\|\boldsymbol{y} - \alpha \boldsymbol{x}\|_2^2 \geq 0$ since it's (the square of) a norm. Therefore,

$$
\begin{aligned}
0 \leq \|\boldsymbol{y} - \alpha \boldsymbol{x}\|_2^2 &= (\boldsymbol{y} - \alpha \boldsymbol{x})^T (\boldsymbol{y} - \alpha \boldsymbol{x}) \\
&= \boldsymbol{y}^T \boldsymbol{y} - 2\alpha \boldsymbol{x}^T \boldsymbol{y} + \alpha^2 \boldsymbol{x}^T \boldsymbol{x} \\
&= \boldsymbol{y}^T \boldsymbol{y} - 2 \frac{\boldsymbol{x}^T \boldsymbol{y}}{\|\boldsymbol{x}\|^2} \boldsymbol{x}^T \boldsymbol{y} + \frac{\left(\boldsymbol{x}^T \boldsymbol{y}\right)^2}{\|\boldsymbol{x}\|^4} \underbrace{\boldsymbol{x}^T \boldsymbol{x}}_{=\|\boldsymbol{x}\|_2^2} \\
&= \|\boldsymbol{y}\|_2^2 - \frac{\left(\boldsymbol{x}^T \boldsymbol{y}\right)^2}{\|\boldsymbol{x}\|_2^2}.
\end{aligned}
$$

This implies

$$
\left(\boldsymbol{x}^T \boldsymbol{y}\right)^2 \leq \|\boldsymbol{x}\|_2^2 \|\boldsymbol{y}\|_2^2,
$$

and the Cauchy–Schwarz–Bunyakovsky inequality follows by taking square roots.

Proof (cont.)

Now we look at the equality claim.

"$\Longrightarrow$": Let's assume that $\left|\boldsymbol{x}^T\boldsymbol{y}\right| = \|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2$. But then the first part of the proof shows that

$$\|\boldsymbol{y} - \alpha\boldsymbol{x}\|_2 = 0$$

so that $\boldsymbol{y} = \alpha\boldsymbol{x}$.

"$\Longleftarrow$": Let's assume $\boldsymbol{y} = \alpha\boldsymbol{x}$. Then

$$\left|\boldsymbol{x}^T\boldsymbol{y}\right| = \left|\boldsymbol{x}^T(\alpha\boldsymbol{x})\right| = |\alpha|\|\boldsymbol{x}\|_2^2$$
$$\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 = \|\boldsymbol{x}\|_2\|\alpha\boldsymbol{x}\|_2 = |\alpha|\|\boldsymbol{x}\|_2^2,$$

so that we have equality. $\square$

Example (cont.)

3. Now we can show that $\|\cdot\|_2$ satisfies the triangle inequality:

$$
\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|_2^2 &= (\boldsymbol{x}+\boldsymbol{y})^T(\boldsymbol{x}+\boldsymbol{y}) \\
&= \underbrace{\boldsymbol{x}^T\boldsymbol{x}}_{=\|\boldsymbol{x}\|_2^2} + \boldsymbol{x}^T\boldsymbol{y} + \underbrace{\boldsymbol{y}^T\boldsymbol{x}}_{=\boldsymbol{x}^T\boldsymbol{y}} + \underbrace{\boldsymbol{y}^T\boldsymbol{y}}_{=\|\boldsymbol{y}\|_2^2} \\
&= \|\boldsymbol{x}\|_2^2 + 2\boldsymbol{x}^T\boldsymbol{y} + \|\boldsymbol{y}\|_2^2 \\
&\le \|\boldsymbol{x}\|_2^2 + 2|\boldsymbol{x}^T\boldsymbol{y}| + \|\boldsymbol{y}\|_2^2 \\
&\overset{\text{CSB}}{\le} \|\boldsymbol{x}\|_2^2 + 2\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 + \|\boldsymbol{y}\|_2^2 \\
&= (\|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2)^2 .
\end{aligned}
$$

Now we just need to take square roots to have the triangle inequality.

Lemma

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then we have the backward triangle inequality

$$\big| \|\mathbf{x}\| - \|\mathbf{y}\| \big| \le \|\mathbf{x} - \mathbf{y}\|.$$

Proof

We write

$$\|\mathbf{x}\| = \|\mathbf{x} - \mathbf{y} + \mathbf{y}\| \overset{\text{tri.ineq.}}{\le} \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\|.$$

But this implies

$$\|\mathbf{x}\| - \|\mathbf{y}\| \le \|\mathbf{x} - \mathbf{y}\|.$$

Proof (cont.)

Switch the roles of $\boldsymbol{x}$ and $\boldsymbol{y}$ to get

$$\|\boldsymbol{y}\| - \|\boldsymbol{x}\| \leq \|\boldsymbol{y} - \boldsymbol{x}\| \quad \Longleftrightarrow \quad -(\|\boldsymbol{x}\| - \|\boldsymbol{y}\|) \leq \|\boldsymbol{x} - \boldsymbol{y}\|.$$

Together with the previous inequality we have

$$|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \leq \|\boldsymbol{x} - \boldsymbol{y}\|.$$

□

Other common norms

- ℓ_1-norm (or taxi-cab norm, Manhattan norm):

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

- ℓ_∞-norm (or maximum norm, Chebyshev norm):

$$\|\boldsymbol{x}\|_\infty = \max_{1 \le i \le n} |x_i|$$

- ℓ_p-norm:

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Remark

In the homework you will use Hölder's and Minkowski's inequalities to show that the p-norm is a norm.

Remark

We now show that

$$\|\boldsymbol{x}\|_\infty = \lim_{p\to\infty} \|\boldsymbol{x}\|_p.$$

Let's use tildes to mark all components of $\boldsymbol{x}$ *that are maximal, i.e..*

$$\tilde{x}_1 = \tilde{x}_2 = \ldots = \tilde{x}_k = \max_{1\le i\le n} |x_i|.$$

The remaining components are then $\tilde{x}_{k+1}, \ldots, \tilde{x}_n$*.*
This implies that

$$\frac{\tilde{x}_i}{\tilde{x}_1} < 1, \quad \textit{for } i = k+1, \ldots, n.$$

Remark (cont.)

Now

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |\tilde{x}_i|^p\right)^{1/p}$$

$$= |\tilde{x}_1| \left(k + \underbrace{\left|\frac{\tilde{x}_{k+1}}{\tilde{x}_1}\right|^p}_{<1} + \ldots + \underbrace{\left|\frac{\tilde{x}_n}{\tilde{x}_1}\right|^p}_{<1} \right)^{1/p}.$$

Since the terms inside the parentheses — except for k — go to 0 for $p \to \infty$, $(\cdot)^{1/p} \to 1$ for $p \to \infty$.

And so

$$\|\boldsymbol{x}\|_p \to |\tilde{x}_1| = \max_{1 \le i \le n} |x_i| = \|\boldsymbol{x}\|_\infty.$$

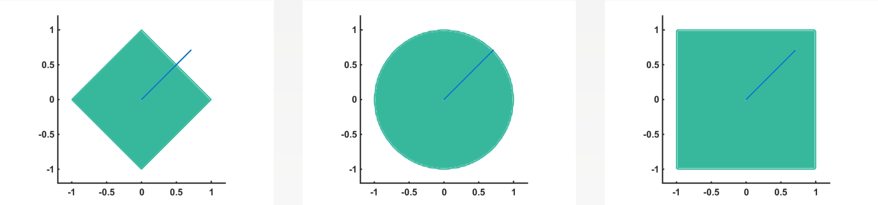


Figure: Unit "balls" in $\mathbb{R}^2$ for the ℓ_1, ℓ_2 and ℓ_∞ norms.

Note that $B_1 \subseteq B_2 \subseteq B_\infty$ since, e.g.,

$$\left\| \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \right\|_1 = \sqrt{2}, \quad \left\| \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \right\|_2 = \sqrt{\frac{1}{2} + \frac{1}{2}} = 1, \quad \left\| \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \right\|_\infty = \frac{\sqrt{2}}{2},$$

so that $$\left\| \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \right\|_1 \geq \left\| \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \right\|_2 \geq \left\| \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \right\|_\infty.$$

In fact, we have in general (similar to HW)

$$\|\boldsymbol{x}\|_1 \geq \|\boldsymbol{x}\|_2 \geq \|\boldsymbol{x}\|_\infty, \quad \text{for any } \boldsymbol{x} \in \mathbb{R}^n.$$

Norm equivalence

Definition

Two norms $\|\cdot\|$ and $\|\cdot\|'$ on a vector space $\mathcal{V}$ are called equivalent if there exist constants α, β such that

$$\alpha \leq \frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|'} \leq \beta \quad \text{for all } \boldsymbol{x}(\neq \boldsymbol{0}) \in \mathcal{V}.$$

Example

$\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent since from above $\|\boldsymbol{x}\|_1 \geq \|\boldsymbol{x}\|_2$ and also $\|\boldsymbol{x}\|_1 \leq \sqrt{n}\|\boldsymbol{x}\|_2$ (see HW) so that

$$\alpha = 1 \leq \frac{\|\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_2} \leq \sqrt{n} = \beta.$$

Remark

In fact, all norms on finite-dimensional vector spaces are equivalent.

[0]

Matrix norms are special norms — they will satisfy one additional property.

This property should help us measure $\|AB\|$ for two matrices A, B of appropriate sizes.

We look at the simplest matrix norm, the Frobenius norm, defined for $A \in \mathbb{R}^{m,n}$:

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} = \left(\sum_{i=1}^{m}\|A_{i*}\|_2^2\right)^{1/2}$$

$$= \left(\sum_{j=1}^{n}\|A_{*j}\|_2^2\right)^{1/2} = \sqrt{\operatorname{trace}(A^TA)},$$

i.e., the Frobenius norm is just a 2-norm for the vector that contains all elements of the matrix.

Now

$$\|\mathsf{A}\boldsymbol{x}\|_2^2 = \sum_{i=1}^{m} |\mathsf{A}_{i*}\boldsymbol{x}|^2$$

$$\stackrel{\text{CSB}}{\leq} \underbrace{\sum_{i=1}^{m} \|\mathsf{A}_{i*}\|_2^2}_{=\|\mathsf{A}\|_F^2} \|\boldsymbol{x}\|_2^2$$

so that

$$\|\mathsf{A}\boldsymbol{x}\|_2 \leq \|\mathsf{A}\|_F \|\boldsymbol{x}\|_2.$$

We can generalize this to matrices, i.e., we have

$$\|\mathsf{AB}\|_F \leq \|\mathsf{A}\|_F \|\mathsf{B}\|_F,$$

which motivates us to require this submultiplicativity for any matrix norm.

Definition

A matrix norm is a function $\|\cdot\|$ from the set of all real (or complex) matrices of finite size into $\mathbb{R}_{\geq 0}$ that satisfies

1. $\|A\| \geq 0$ and $\|A\| = 0$ if and only if $A = O$ (a matrix of all zeros).
2. $\|\alpha A\| = |\alpha|\|A\|$ for all $\alpha \in \mathbb{R}$.
3. $\|A + B\| \leq \|A\| + \|B\|$ (requires A, B to be of same size).
4. $\|AB\| \leq \|A\|\|B\|$ (requires A, B to have appropriate sizes).

Remark

This definition is usually too general. In addition to the Frobenius norm, most useful matrix norms are induced by a vector norm.

Induced matrix norms

Theorem

Let $\|\cdot\|_{(m)}$ and $\|\cdot\|_{(n)}$ be vector norms on $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, and let A *be an $m \times n$ matrix. Then*

$$\|\mathsf{A}\| = \max_{\|\boldsymbol{x}\|_{(n)}=1} \|\mathsf{A}\boldsymbol{x}\|_{(m)}$$

is a matrix norm called the induced matrix norm.

Remark

Here the vector norm could be any vector norm. In particular, any p-norm. For example, we could have

$$\|\mathsf{A}\|_2 = \max_{\|\boldsymbol{x}\|_{2,(n)}=1} \|\mathsf{A}\boldsymbol{x}\|_{2,(m)}.$$

To keep notation simple we often drop indices.

Proof

1. $\|A\| \geq 0$ is obvious since this holds for the vector norm.
   It remains to show that $\|A\| = 0$ if and only if $A = O$.
   Assume $A = O$, then

   $$\|A\| = \max_{\|\boldsymbol{x}\|=1} \| \underbrace{A\boldsymbol{x}}_{=\mathbf{0}} \| = 0.$$

   So now consider $A \neq O$. We need to show that $\|A\| > 0$.
   There must exist a column of A that is not $\mathbf{0}$. We call this column A_{*k} and take $\boldsymbol{x} = \boldsymbol{e}_k$.
   Then

   $$\|A\| = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| \overset{\|\boldsymbol{e}_k\|=1}{\geq} \|A\boldsymbol{e}_k\| = \|A_{*k}\| > 0$$

   since $A_{*k} \neq \mathbf{0}$.

Proof (cont.)

2. Using the corresponding property for the vector norm we have

$$\|\alpha \mathsf{A}\| = \max \|\alpha \mathsf{A}\boldsymbol{x}\| = |\alpha| \max \|\mathsf{A}\boldsymbol{x}\| = |\alpha|\|\mathsf{A}\|.$$

3. Also straightforward (based on the triangle inequality for the vector norm)

$$\begin{aligned}\|\mathsf{A}+\mathsf{B}\| &= \max \|(\mathsf{A}+\mathsf{B})\boldsymbol{x}\| = \max \|\mathsf{A}\boldsymbol{x}+\mathsf{B}\boldsymbol{x}\| \\ &\le \max \left(\|\mathsf{A}\boldsymbol{x}\| + \|\mathsf{B}\boldsymbol{x}\|\right) \\ &= \max \|\mathsf{A}\boldsymbol{x}\| + \max \|\mathsf{B}\boldsymbol{x}\| = \|\mathsf{A}\| + \|\mathsf{B}\|.\end{aligned}$$

Proof (cont.)

4. First note that

$$\max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$$

and so

$$\|A\| = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \geq \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}.$$

Therefore

$$\|A\boldsymbol{x}\| \leq \|A\|\|\boldsymbol{x}\|. \tag{1}$$

But then we also have $\|AB\| \leq \|A\|\|B\|$ since

$$\|AB\| = \max_{\|\boldsymbol{x}\|=1} \|AB\boldsymbol{x}\| = \|AB\boldsymbol{y}\| \quad (\text{for some } \boldsymbol{y} \text{ with } \|\boldsymbol{y}\| = 1)$$

$$\stackrel{(1)}{\leq} \|A\|\|B\boldsymbol{y}\| \stackrel{(1)}{\leq} \|A\|\|B\| \underbrace{\|\boldsymbol{y}\|}_{=1}.$$

□

Remark

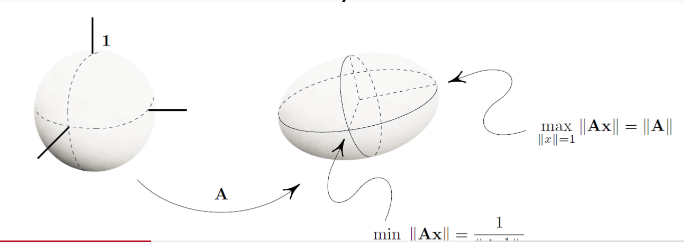
- *One can show (see HW) that — if* A *is invertible —*

$$\min_{\|\boldsymbol{x}\|=1} \|\mathsf{A}\boldsymbol{x}\| = \frac{1}{\|\mathsf{A}^{-1}\|}.$$

- *The induced matrix norm can be interpreted geometrically:*

  $\|\mathsf{A}\|$: *the most a vector on the unit sphere can be stretched when transformed by* A.

  $\frac{1}{\|\mathsf{A}^{-1}\|}$: *the most a vector on the unit sphere can be shrunk when transformed by* A.



Matrix 2-norm

Theorem

Let A *be an* $m \times n$ *matrix. Then*

1. $\|A\|_2 = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|_2 = \sqrt{\lambda_{max}}$.
2. $\|A^{-1}\|_2 = \dfrac{1}{\min_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|_2} = \dfrac{1}{\sqrt{\lambda_{min}}}$.

where λ_{max} *and* λ_{min} *are the largest and smallest eigenvalues of* $A^T A$*, respectively.*

Remark

We also have

$$\sqrt{\lambda_{max}} = \sigma_1, \quad \text{the largest singular value of } A,$$
$$\sqrt{\lambda_{min}} = \sigma_n, \quad \text{the smallest singular value of } A.$$

Proof

We will show only (1), the largest singular value ((2) goes similarly).

The idea is to solve a constrained optimization problem (as in calculus), i.e.,

$$\begin{aligned} \text{maximize} \quad & f(\boldsymbol{x}) = \|\mathsf{A}\boldsymbol{x}\|_2^2 = (\mathsf{A}\boldsymbol{x})^T\mathsf{A}\boldsymbol{x} \\ \text{subject to} \quad & g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2 = \boldsymbol{x}^T\boldsymbol{x} = 1. \end{aligned}$$

We do this by introducing a Lagrange multiplier λ and define

$$h(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda g(\boldsymbol{x}) = \boldsymbol{x}^T\mathsf{A}^T\mathsf{A}\boldsymbol{x} - \lambda\boldsymbol{x}^T\boldsymbol{x}.$$

Proof (cont.)

Necessary and sufficient (since quadratic) condition for maximum:
$\frac{\partial h}{\partial x_i} = 0$, $i = 1, \ldots, n$, $g(\boldsymbol{x}) = 1$

$$\begin{aligned}
\frac{\partial}{\partial x_i}\left(\boldsymbol{x}^T \mathsf{A}^T \mathsf{A}\boldsymbol{x} - \lambda \boldsymbol{x}^T \boldsymbol{x}\right) &= \frac{\partial \boldsymbol{x}^T}{\partial x_i}\mathsf{A}^T \mathsf{A}\boldsymbol{x} + \boldsymbol{x}^T \mathsf{A}^T \mathsf{A}\frac{\partial \boldsymbol{x}}{\partial x_i} - \lambda \frac{\partial \boldsymbol{x}^T}{\partial x_i}\boldsymbol{x} - \lambda \boldsymbol{x}^T \frac{\partial \boldsymbol{x}}{\partial x_i} \\
&= 2\boldsymbol{e}_i^T \mathsf{A}^T \mathsf{A}\boldsymbol{x} - 2\lambda \boldsymbol{e}_i^T \boldsymbol{x} \\
&= 2\left((\mathsf{A}^T \mathsf{A}\boldsymbol{x})_i - (\lambda \boldsymbol{x})_i\right), \quad i = 1, \ldots, n.
\end{aligned}$$

Together this yields

$$\mathsf{A}^T \mathsf{A}\boldsymbol{x} - \lambda \boldsymbol{x} = \boldsymbol{0} \quad \Longleftrightarrow \quad \left(\mathsf{A}^T \mathsf{A} - \lambda \mathsf{I}\right)\boldsymbol{x} = \boldsymbol{0},$$

so that λ must be an eigenvalue of $\mathsf{A}^T\mathsf{A}$ (since $g(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x} = 1$ ensures $\boldsymbol{x} \neq \boldsymbol{0}$).

Proof (cont.)

In fact, as we now show, λ is the maximal eigenvalue.
First,

$$A^T A\boldsymbol{x} = \lambda \boldsymbol{x} \quad \Longrightarrow \quad \boldsymbol{x}^T A^T A\boldsymbol{x} = \lambda \boldsymbol{x}^T \boldsymbol{x} = \lambda$$

so that

$$\|A\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^T A^T A\boldsymbol{x}} = \sqrt{\lambda}.$$

And then

$$\|A\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|A\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{x}\|_2^2=1} \|A\boldsymbol{x}\|_2$$
$$= \max \sqrt{\lambda} = \sqrt{\lambda_{max}}.$$

□

Special properties of the 2-norm

1. $\|A\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \max_{\|\boldsymbol{y}\|_2=1} |\boldsymbol{y}^T A \boldsymbol{x}|$
2. $\|A\|_2 = \|A^T\|_2$
3. $\|A^T A\|_2 = \|A\|_2^2 = \|AA^T\|_2$
4. $\left\| \begin{pmatrix} A & O \\ O & B \end{pmatrix} \right\| = \max\{\|A\|_2, \|B\|_2\}$
5. $\|U^T A V\|_2 = \|A\|_2$ provided $UU^T = I$ and $V^T V = I$ (orthogonal matrices).

Remark

The proof is a HW problem.

Matrix 1-norm and ∞-norm

Theorem

Let A be an $m \times n$ matrix. Then we have

1. *the column sum norm*

$$\|A\|_1 = \max_{\|\boldsymbol{x}\|_1=1} \|A\boldsymbol{x}\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}|,$$

2. *and the row sum norm*

$$\|A\|_\infty = \max_{\|\boldsymbol{x}\|_\infty=1} \|A\boldsymbol{x}\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}|.$$

Remark

We know these are norms, so what we need to do is verify that the formulas hold. We will show (1).

Proof

First we look at $\|\mathsf{A}\boldsymbol{x}\|_1$.

$$\|\mathsf{A}\boldsymbol{x}\|_1 = \sum_{i=1}^{m} |(\mathsf{A}\boldsymbol{x})_i| = \sum_{i=1}^{m} |\mathsf{A}_{i*}\boldsymbol{x}| = \sum_{i=1}^{m} \left|\sum_{j=1}^{n} a_{ij}x_j\right|$$

$$\overset{\text{reg.}\Delta}{\leq} \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|\,|x_j|$$

$$= \sum_{j=1}^{n}\left[|x_j|\sum_{i=1}^{m}|a_{ij}|\right] \leq \left[\max_{j=1,\ldots,n}\sum_{i=1}^{m}|a_{ij}|\right]\sum_{j=1}^{n}|x_j|.$$

Since we actually need to look at $\|\mathsf{A}\boldsymbol{x}\|_1$ for $\|\boldsymbol{x}\|_1 = 1$ we note that $\|\boldsymbol{x}\|_1 = \sum_{j=1}^{n}|x_j|$ and therefore have

$$\|\mathsf{A}\boldsymbol{x}\|_1 \leq \max_{j=1,\ldots,n}\sum_{i=1}^{m}|a_{ij}|.$$

Proof (cont.)

We even have equality since for $\boldsymbol{x} = \boldsymbol{e}_k$, where k is the index such that A_{*k} has maximum column sum, we get

$$\|\mathsf{A}\boldsymbol{x}\|_1 = \|\mathsf{A}\boldsymbol{e}_k\|_1 = \|\mathsf{A}_{*k}\|_1 = \sum_{i=1}^{m} |a_{ik}|$$
$$= \max_{j=1\ldots,n} \sum_{i=1}^{m} |a_{ik}|$$

due to our choice of k.

Since $\|\boldsymbol{e}_k\|_1 = 1$ we indeed have the desired formula. □

Definition

[0]

A general inner product in a real (complex) vector space $\mathcal{V}$ is a symmetric (Hermitian) bilinear form $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ ($\mathbb{C}$), i.e.,

1. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \in \mathbb{R}_{\geq 0}$ with $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$.
2. $\langle \boldsymbol{x}, \alpha \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ for all scalars α.
3. $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$.
4. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ (or $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}$ if complex).

Remark

The following two properties (providing bilinearity) are implied (see HW)

$$\langle \alpha \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\alpha} \langle \boldsymbol{x}, \boldsymbol{y} \rangle$$
$$\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle.$$

As before, any inner product induces a norm via

$$\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}.$$

One can show (analogous to the Euclidean case) that $\|\cdot\|$ is a norm.

In particular, we have a general Cauchy–Schwarz–Bunyakovsky inequality

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|.$$

Example

1. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{y}$ (or $\boldsymbol{x}^* \boldsymbol{y}$), the standard inner product for $\mathbb{R}^n$ ($\mathbb{C}^n$).
2. For nonsingular matrices A we get the A-inner product on $\mathbb{R}^n$, i.e.,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T \mathsf{A}^T \mathsf{A} \boldsymbol{y}$$

   with

$$\|\boldsymbol{x}\|_{\mathsf{A}} = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{\boldsymbol{x}^T \mathsf{A}^T \mathsf{A} \boldsymbol{x}} = \|\mathsf{A}\boldsymbol{x}\|_2.$$

3. If $\mathcal{V} = \mathbb{R}^{m \times n}$ (or $\mathbb{C}^{m \times n}$) then we get the standard inner product for matrices, i.e.,

$$\langle \mathsf{A}, \mathsf{B} \rangle = \operatorname{trace}(\mathsf{A}^T \mathsf{B}) \quad (\text{or } \operatorname{trace}(\mathsf{A}^* \mathsf{B}))$$

   with

$$\|\mathsf{A}\| = \sqrt{\langle \mathsf{A}, \mathsf{A} \rangle} = \sqrt{\operatorname{trace}(\mathsf{A}^T \mathsf{A})} = \|\mathsf{A}\|_F.$$

Remark

In the infinite-dimensional setting we have, e.g., for f, g continuous functions on (a, b)

$$\langle f, g \rangle = \int_a^b f(t)g(t)\mathrm{d}t$$

with

$$\|f\| = \sqrt{\langle f, f \rangle} = \left(\int_a^b (f(t))^2 \,\mathrm{d}t \right)^{1/2}.$$

Parallelogram identity

In any inner product space the so-called parallelogram identity holds, i.e.,

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2+\|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\left(\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2\right). \tag{2}$$

This is true since

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2+\|\boldsymbol{x}-\boldsymbol{y}\|^2 &= \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}\rangle + \langle \boldsymbol{x}-\boldsymbol{y}, \boldsymbol{x}-\boldsymbol{y}\rangle \\
&= \langle \boldsymbol{x}, \boldsymbol{x}\rangle + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{y}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle \\
&\quad + \langle \boldsymbol{x}, \boldsymbol{x}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle - \langle \boldsymbol{y}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle \\
&= 2\langle \boldsymbol{x}, \boldsymbol{x}\rangle + 2\langle \boldsymbol{y}, \boldsymbol{y}\rangle = 2\left(\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2\right).
\end{aligned}$$

Polarization identity

The following theorem shows that we

- not only get a norm from an inner product (i.e., every Hilbert space is a Banach space),
- but — if the parallelogram identity holds — then we can get an inner product from a norm (i.e., a Banach space becomes a Hilbert space).

Theorem

Let $\mathcal{V}$ be a real vector space with norm $\|\cdot\|$. If the parallelogram identity (2) holds then

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2\right) \tag{3}$$

is an inner product on $\mathcal{V}$.

Proof

We need to show that all four properties of a general inner product hold.

1. Nonnegativity:

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{x}\|^2 - \|\boldsymbol{x}-\boldsymbol{x}\|^2 \right) = \frac{1}{4}\|2\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2 \geq 0.$$

Moreover, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$ since $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \|\boldsymbol{x}\|^2$.

4. Symmetry:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$$

is clear since $\|\boldsymbol{x}-\boldsymbol{y}\| = \|\boldsymbol{y}-\boldsymbol{x}\|$.

Proof (cont.)

3. Additivity: The parallelogram identity implies

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2+\|\boldsymbol{x}+\boldsymbol{z}\|^2=\frac{1}{2}\left(\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{x}+\boldsymbol{z}\|^2+\|\boldsymbol{y}-\boldsymbol{z}\|^2\right). \quad (4)$$

and

$$\|\boldsymbol{x}-\boldsymbol{y}\|^2+\|\boldsymbol{x}-\boldsymbol{z}\|^2=\frac{1}{2}\left(\|\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{x}-\boldsymbol{z}\|^2+\|\boldsymbol{z}-\boldsymbol{y}\|^2\right). \quad (5)$$

Subtracting (5) from (4) we get

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2-\|\boldsymbol{x}-\boldsymbol{y}\|^2+\|\boldsymbol{x}+\boldsymbol{z}\|^2-\|\boldsymbol{x}-\boldsymbol{z}\|^2 = \frac{1}{2}\left(\|2\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|^2-\|2\boldsymbol{x}-\boldsymbol{y}-\boldsymbol{z}\|^2\right). \quad (6)$$

Proof (cont.)

The specific form of the polarized inner product implies

$$\begin{aligned}
\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle &= \frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2 + \|\boldsymbol{x}+\boldsymbol{z}\|^2 - \|\boldsymbol{x}-\boldsymbol{z}\|^2\right) \\
&\overset{(6)}{=} \frac{1}{8}\left(\|2\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|^2 - \|2\boldsymbol{x}-\boldsymbol{y}-\boldsymbol{z}\|^2\right) \\
&= \frac{1}{2}\left(\left\|\boldsymbol{x}+\frac{\boldsymbol{y}+\boldsymbol{z}}{2}\right\|^2 - \left\|\boldsymbol{x}-\frac{\boldsymbol{y}+\boldsymbol{z}}{2}\right\|^2\right) \\
&\overset{\text{polarization}}{=} 2\langle \boldsymbol{x}, \frac{\boldsymbol{y}+\boldsymbol{z}}{2}\rangle. \qquad (7)
\end{aligned}$$

Setting $\boldsymbol{z} = \boldsymbol{0}$ in (7) yields

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 2\langle \boldsymbol{x}, \frac{\boldsymbol{y}}{2}\rangle \qquad (8)$$

since $\langle \boldsymbol{x}, \boldsymbol{z}\rangle = 0$.

Proof (cont.)

To summarize, we have

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle = 2\langle \boldsymbol{x}, \frac{\boldsymbol{y}+\boldsymbol{z}}{2} \rangle. \tag{7}$$

and

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 2\langle \boldsymbol{x}, \frac{\boldsymbol{y}}{2} \rangle. \tag{8}$$

Since (8) is true for any $\boldsymbol{y} \in \mathcal{V}$ we can, in particular, set $\boldsymbol{y} = \boldsymbol{y} + \boldsymbol{z}$ so that we have

$$\langle \boldsymbol{x}, \boldsymbol{y}+\boldsymbol{z} \rangle = 2\langle \boldsymbol{x}, \frac{\boldsymbol{y}+\boldsymbol{z}}{2} \rangle.$$

This, however, is the right-hand side of (7) so that we end up with

$$\langle \boldsymbol{x}, \boldsymbol{y}+\boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle,$$

as desired.

Proof (cont.)

2. Scalar multiplication:
To show $\langle \boldsymbol{x}, \alpha \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ for integer α we can just repeatedly apply the additivity property just proved.

From this we can get the property for rational α as follows.
We let $\alpha = \frac{\beta}{\gamma}$ with integer $\beta, \gamma \neq 0$ so that

$$\beta\gamma \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \gamma \boldsymbol{x}, \beta \boldsymbol{y} \rangle = \gamma^2 \langle \boldsymbol{x}, \frac{\beta}{\gamma} \boldsymbol{y} \rangle.$$

Dividing by γ^2 we get

$$\frac{\beta}{\gamma} \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \frac{\beta}{\gamma} \boldsymbol{y} \rangle.$$

Proof (cont.)

Finally, for real α we use the continuity of the norm function (see HW) which implies that our inner product $\langle \cdot, \cdot \rangle$ also is continuous.

Now we take a sequence $\{\alpha_n\}$ of rational numbers such that $\alpha_n \to \alpha$ for $n \to \infty$ and have — by continuity

$$\langle \boldsymbol{x}, \alpha_n \boldsymbol{y} \rangle \to \langle \boldsymbol{x}, \alpha \boldsymbol{y} \rangle \quad \text{as } n \to \infty.$$

□

Theorem

The only vector p-norm induced by an inner product is the 2-norm.

Remark

Since many problems are more easily dealt with in inner product spaces (since we then have lengths and angles, see next section) the 2-norm has a clear advantage over other p-norms.

Proof

We know that the 2-norm does induce an inner product, i.e.,

$$\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}.$$

Therefore we need to show that it doesn't work for $p \neq 2$.
We do this by showing that the parallelogram identity

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\right)$$

fails for $p \neq 2$.
We will do this for $1 \leq p < \infty$. You will work out the case $p = \infty$ in a HW problem.

Proof (cont.)

All we need is a counterexample, so we take $\boldsymbol{x} = \boldsymbol{e}_1$ and $\boldsymbol{y} = \boldsymbol{e}_2$ so that

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p^2 = \|\boldsymbol{e}_1+\boldsymbol{e}_2\|_p^2 = \left(\sum_{i=1}^{n} |[\boldsymbol{e}_1+\boldsymbol{e}_2]_i|^p\right)^{2/p} = 2^{2/p}$$

and, similarly

$$\|\boldsymbol{x}-\boldsymbol{y}\|_p^2 = \|\boldsymbol{e}_1-\boldsymbol{e}_2\|_p^2 = 2^{2/p}.$$

Together, the left-hand side of the parallelogram identity is $2\left(2^{2/p}\right) = 2^{2/p+1}$.

Proof (cont.)

For the right-hand side of the parallelogram identity we calculate

$$\|\boldsymbol{x}\|_p^2 = \|\boldsymbol{e}_1\|_p^2 = 1 = \|\boldsymbol{e}_2\|_p^2 = \|\boldsymbol{y}\|_p^2,$$

so that the right-hand side comes out to 4.

Finally, we have

$$2^{2/p+1} = 4 \quad \Longleftrightarrow \quad \frac{2}{p} + 1 = 2 \quad \Longleftrightarrow \quad \frac{2}{p} = 1 \text{ or } p = 2.$$

□

Orthogonal Vectors

[0]

We will now work in a general inner product space $\mathcal{V}$ with induced norm

$$\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}.$$

Definition

Two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$ are called orthogonal if

$$\langle\boldsymbol{x}, \boldsymbol{y}\rangle = 0.$$

We often use the notation $\boldsymbol{x} \perp \boldsymbol{y}$.

In the HW you will prove the Pythagorean theorem for the 2-norm and standard inner product $\boldsymbol{x}^T\boldsymbol{y}$, i.e.,
$\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 = \|\boldsymbol{x}-\boldsymbol{y}\|^2 \iff \boldsymbol{x}^T\boldsymbol{y} = 0.$
Moreover, the law of cosines states

$$\|\boldsymbol{x}-\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\|\boldsymbol{x}\|\|\boldsymbol{y}\|\cos\theta,$$

so that

$$\cos\theta = \frac{\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2}{2\|\boldsymbol{x}\|\|\boldsymbol{y}\|} \overset{\text{Pythagoras}}{=} \frac{2\boldsymbol{x}^T\boldsymbol{y}}{2\|\boldsymbol{x}\|\|\boldsymbol{y}\|}.$$

This motivates our general definition of angles:

Definition

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$. The angle between $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined via

$$\cos\theta = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}, \qquad \theta \in [0, \pi].$$

Orthonormal sets

Definition

A set $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\} \subseteq \mathcal{V}$ is called orthonormal if

$$\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \delta_{ij} \quad \text{(Kronecker delta)}.$$

Theorem

Every orthonormal set is linearly independent.

Corollary

Every orthonormal set of n vectors from an n-dimensional vector space $\mathcal{V}$ is an orthonormal basis for $\mathcal{V}$.

Proof (of the theorem)

We want to show linear independence, i.e., that

$$\sum_{j=1}^{n} \alpha_j \boldsymbol{u}_j = \boldsymbol{0} \quad \Longrightarrow \quad \alpha_j = 0,\ j = 1, \ldots, n.$$

To see this is true we take the inner product with $\boldsymbol{u}_i$:

$$\langle \boldsymbol{u}_i, \sum_{j=1}^{n} \alpha_j \boldsymbol{u}_j \rangle = \langle \boldsymbol{u}_i, \boldsymbol{0} \rangle$$

$$\Longleftrightarrow \quad \sum_{j=1}^{n} \alpha_j \underbrace{\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle}_{=\delta_{ij}} = 0 \quad \Longleftrightarrow \quad \alpha_i = 0.$$

Since i was arbitrary this holds for all $i = 1, \ldots, n$, and we have linear independence. □

Example

The standard orthonormal basis of $\mathbb{R}^n$ is given by

$$\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}.$$

Using this basis we can express any $\boldsymbol{x} \in \mathbb{R}^n$ as

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \ldots + x_n\boldsymbol{e}_n,$$

we get a coordinate expansion of $\boldsymbol{x}$.

In fact, *any* other orthonormal basis provides just as simple a representation of $\boldsymbol{x}$;

Consider the orthonormal basis $\mathcal{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ and assume

$$\boldsymbol{x} = \sum_{j=1}^{n} \alpha_j \boldsymbol{u}_j$$

for some appropriate scalars α_j.
To find these expansion coefficients α_j we take the inner product with $\boldsymbol{u}_i$, i.e.,

$$\langle \boldsymbol{u}_i, \boldsymbol{x} \rangle = \sum_{j=1}^{n} \alpha_j \underbrace{\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle}_{=\delta_{ij}} = \alpha_i.$$

We therefore have proved

Theorem

Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ be an orthonormal basis for an inner product space $\mathcal{V}$. Then any $\boldsymbol{x} \in \mathcal{V}$ can be written as

$$\boldsymbol{x} = \sum_{j=1}^{n} \langle \boldsymbol{x}, \boldsymbol{u}_i \rangle \boldsymbol{u}_i.$$

This is a (finite) Fourier expansion with Fourier coefficients $\langle \boldsymbol{x}, \boldsymbol{u}_i \rangle$.

Remark

The classical (infinite-dimensional) Fourier series for continuous functions on $(-\pi,\pi)$ *uses the orthogonal (but not yet orthonormal) basis*

$$\{1, \sin t, \cos t, \sin 2t, \cos 2t, \ldots, \}$$

and the inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(t)g(t)\mathrm{d}t.$$

Example

Consider the basis

$$\mathcal{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}.$$

It is clear by inspection that $\mathcal{B}$ is an orthogonal subset of $\mathbb{R}^3$, i.e., using the Euclidean inner product, we have $\boldsymbol{u}_i^T \boldsymbol{u}_j = 0$, $i, j = 1, 2, 3$, $i \neq j$.

We can obtain an orthonormal basis by normalizing the vectors, i.e., by computing $\boldsymbol{v}_i = \frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|_2}$, $i = 1, 2, 3$.

This yields

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{v}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Example (cont.)

The Fourier expansion of $\boldsymbol{x} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}^T$ is given by

$$\begin{aligned}
\boldsymbol{x} &= \sum_{i=1}^{3} \left(\boldsymbol{x}^T \boldsymbol{v}_i\right) \boldsymbol{v}_i \\
&= \frac{4}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \frac{2}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \\
&= \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.
\end{aligned}$$

[0]

We want to convert an arbitrary basis $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ of $\mathcal{V}$ to an orthonormal basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$.

Idea: construct $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ successively so that $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ is an ON basis for $\operatorname{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$, $k = 1, \ldots, n$.

Construction

$k = 1$:

$$\boldsymbol{u}_1 = \frac{\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|}.$$

$k = 2$: Consider the projection of $\boldsymbol{x}_2$ onto $\boldsymbol{u}_1$, i.e.,

$$\langle \boldsymbol{u}_1, \boldsymbol{x}_2 \rangle \boldsymbol{u}_1.$$

Then

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - \langle \boldsymbol{u}_1, \boldsymbol{x}_2 \rangle \boldsymbol{u}_1$$

and

$$\boldsymbol{u}_2 = \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}.$$

In general, consider $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ as a given ON basis for $\text{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$.
Use the Fourier expansion to express $\boldsymbol{x}_{k+1}$ with respect to $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{k+1}\}$:

$$\begin{aligned}
& \boldsymbol{x}_{k+1} = \sum_{i=1}^{k+1} \langle \boldsymbol{u}_i, \boldsymbol{x}_{k+1} \rangle \boldsymbol{u}_i \\
\iff \quad & \boldsymbol{x}_{k+1} = \sum_{i=1}^{k} \langle \boldsymbol{u}_i, \boldsymbol{x}_{k+1} \rangle \boldsymbol{u}_i + \langle \boldsymbol{u}_{k+1}, \boldsymbol{x}_{k+1} \rangle \boldsymbol{u}_{k+1} \\
\iff \quad & \boldsymbol{u}_{k+1} = \frac{\boldsymbol{x}_{k+1} - \sum_{i=1}^{k} \langle \boldsymbol{u}_i, \boldsymbol{x}_{k+1} \rangle \boldsymbol{u}_i}{\langle \boldsymbol{u}_{k+1}, \boldsymbol{x}_{k+1} \rangle} = \frac{\boldsymbol{v}_{k+1}}{\langle \boldsymbol{u}_{k+1}, \boldsymbol{x}_{k+1} \rangle}
\end{aligned}$$

This vector, however is not yet normalized.

We now want $\|\mathbf{u}_{k+1}\| = 1$, i.e.,

$$\sqrt{\langle \frac{\mathbf{v}_{k+1}}{\langle \mathbf{u}_{k+1}, \mathbf{x}_{k+1} \rangle}, \frac{\mathbf{v}_{k+1}}{\langle \mathbf{u}_{k+1}, \mathbf{x}_{k+1} \rangle} \rangle} = \frac{1}{|\langle \mathbf{u}_{k+1}, \mathbf{x}_{k+1} \rangle|} \|\mathbf{v}_{k+1}\| = 1$$

$$\Longrightarrow \quad \|\mathbf{v}_{k+1}\| = \|\mathbf{x}_{k+1} - \sum_{i=1}^{k} \langle \mathbf{u}_i, \mathbf{x}_{k+1} \rangle \mathbf{u}_i\| = |\langle \mathbf{u}_{k+1}, \mathbf{x}_{k+1} \rangle|.$$

Therefore

$$\langle \mathbf{u}_{k+1}, \mathbf{x}_{k+1} \rangle = \pm \|\mathbf{x}_{k+1} - \sum_{i=1}^{k} \langle \mathbf{u}_i, \mathbf{x}_{k+1} \rangle \mathbf{u}_i\|.$$

Since the factor ± 1 does not change the span, nor orthogonality, nor normalization we can pick the positive sign.

Gram–Schmidt algorithm

Summarizing, we have

$$\boldsymbol{u}_1 = \frac{\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|},$$

$$\boldsymbol{v}_k = \boldsymbol{x}_k - \sum_{i=1}^{k-1} \langle \boldsymbol{u}_i, \boldsymbol{x}_k \rangle \boldsymbol{u}_i, \quad k = 2, \ldots, n,$$

$$\boldsymbol{u}_k = \frac{\boldsymbol{v}_k}{\|\boldsymbol{v}_k\|}.$$

Using matrix notation to describe Gram–Schmidt

We will assume $\mathcal{V} \subseteq \mathbb{R}^m$ (but this also works in the complex case).
Let

$$\mathsf{U}_1 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^m$$

and for $k = 2, 3, \ldots, n$ let

$$\mathsf{U}_k = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_{k-1} \end{pmatrix} \in \mathbb{R}^{m \times k-1}.$$

Then

$$\mathsf{U}_k^T \boldsymbol{x}_k = \begin{pmatrix} \boldsymbol{u}_1^T \boldsymbol{x}_k \\ \boldsymbol{u}_2^T \boldsymbol{x}_k \\ \vdots \\ \boldsymbol{u}_{k-1}^T \boldsymbol{x}_k \end{pmatrix}$$

and

$$\begin{aligned} \mathsf{U}_k \mathsf{U}_k^T \boldsymbol{x}_k &= \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_{k-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_1^T \boldsymbol{x}_k \\ \boldsymbol{u}_2^T \boldsymbol{x}_k \\ \vdots \\ \boldsymbol{u}_{k-1}^T \boldsymbol{x}_k \end{pmatrix} \\ &= \sum_{i=1}^{k-1} \boldsymbol{u}_i (\boldsymbol{u}_i^T \boldsymbol{x}_k) = \sum_{i=1}^{k-1} (\boldsymbol{u}_i^T \boldsymbol{x}_k) \boldsymbol{u}_i. \end{aligned}$$

Now, Gram–Schmidt says

$$\boldsymbol{v}_k = \boldsymbol{x}_k - \sum_{i=1}^{k-1} (\boldsymbol{u}_i^T \boldsymbol{x}_k) \boldsymbol{u}_i = \boldsymbol{x}_k - \mathsf{U}_k \mathsf{U}_k^T \boldsymbol{x}_k$$
$$= \left(\mathsf{I} - \mathsf{U}_k \mathsf{U}_k^T \right) \boldsymbol{x}_k, \qquad k = 1, 2, \ldots, n,$$

where the case $k = 1$ is also covered by the special definition of U_1.

Remark

$\mathsf{U}_k \mathsf{U}_k^T$ *is a projection matrix, and* $\mathsf{I} - \mathsf{U}_k \mathsf{U}_k^T$ *is a complementary projection. We will cover these later.*

QR Factorization (via Gram–Schmidt)

Consider an $m \times n$ matrix A with rank(A) $= n$.

We want to convert the set of columns of A, $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ to an ON basis $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n\}$ of $R(\mathsf{A})$.

From our discussion of Gram–Schmidt we know

$$\begin{aligned}
\boldsymbol{q}_1 &= \frac{\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|}, \\
\boldsymbol{v}_k &= \boldsymbol{a}_k - \sum_{i=1}^{k-1} \langle \boldsymbol{q}_i, \boldsymbol{a}_k \rangle \boldsymbol{q}_i, \quad k = 2, \ldots, n, \\
\boldsymbol{q}_k &= \frac{\boldsymbol{v}_k}{\|\boldsymbol{v}_k\|}.
\end{aligned}$$

We now rewrite as follows:

$$\begin{aligned} \boldsymbol{a}_1 &= \|\boldsymbol{a}_1\|\boldsymbol{q}_1 \\ \boldsymbol{a}_k &= \langle \boldsymbol{q}_1, \boldsymbol{a}_2\rangle \boldsymbol{q}_1 + \ldots + \langle \boldsymbol{q}_{k-1}, \boldsymbol{a}_k\rangle \boldsymbol{q}_{k-1} + \|\boldsymbol{v}_k\|\boldsymbol{q}_k, \quad k = 2, \ldots, n. \end{aligned}$$

We also introduce the new notation

$$r_{11} = \|\boldsymbol{a}_1\|, \quad r_{kk} = \|\boldsymbol{v}_k\|, \; k = 2, \ldots, n.$$

Then

$$\begin{aligned} \mathsf{A} &= \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{pmatrix}}_{=\mathsf{Q}} \underbrace{\begin{pmatrix} r_{11} & \langle \boldsymbol{q}_1, \boldsymbol{a}_2\rangle & \cdots & \langle \boldsymbol{q}_1, \boldsymbol{a}_n\rangle \\ & r_{22} & \cdots & \langle \boldsymbol{q}_2, \boldsymbol{a}_n\rangle \\ & & \ddots & \vdots \\ \mathsf{O} & & & r_{nn} \end{pmatrix}}_{=\mathsf{R}} \end{aligned}$$

and we have the reduced QR factorization of A.

Remark

- *The matrix* Q *is* $m \times n$ *with orthonormal columns*
- *The matrix* R *is* $n \times n$ *upper triangular with positive diagonal entries.*
- *The reduced QR factorization is unique (see HW).*

Example

Find the QR factorization of the matrix $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$.

$$\boldsymbol{q}_1 = \frac{\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \qquad r_{11} = \|\boldsymbol{a}_1\| = \sqrt{2}$$

$$\boldsymbol{v}_2 = \boldsymbol{a}_2 - \left(\boldsymbol{q}_1^T \boldsymbol{a}_2\right) \boldsymbol{q}_1, \qquad \boldsymbol{q}_1^T \boldsymbol{a}_2 = \frac{2}{\sqrt{2}} = \sqrt{2} = r_{12}$$

$$= \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} - \frac{\sqrt{2}}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \qquad \|\boldsymbol{v}_2\| = \sqrt{3} = r_{22}$$

$$\Longrightarrow \quad \boldsymbol{q}_2 = \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$$

Example (cont.)

$$\boldsymbol{v}_3 = \boldsymbol{a}_3 - \left(\boldsymbol{q}_1^T \boldsymbol{a}_3\right) \boldsymbol{q}_1 - \left(\boldsymbol{q}_2^T \boldsymbol{a}_3\right) \boldsymbol{q}_2$$

with

$$\boldsymbol{q}_1^T \boldsymbol{a}_3 = \frac{1}{\sqrt{2}} = r_{13}, \qquad \boldsymbol{q}_2^T \boldsymbol{a}_3 = 0 = r_{23}$$

Thus

$$\boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{\sqrt{2}\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - 0 = \frac{1}{2} \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \qquad \|\boldsymbol{v}_3\| = \frac{\sqrt{6}}{2} = r_{33}$$

So

$$\boldsymbol{q}_3 = \frac{\boldsymbol{v}_3}{\|\boldsymbol{v}_3\|} = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$$

Example (cont.)

Together we have

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{pmatrix}, \qquad R = \begin{pmatrix} \sqrt{2} & \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{6} \end{pmatrix}$$

Solving linear systems with the QR factorization

Recall the use of the LU factorization to solve $A\boldsymbol{x} = \boldsymbol{b}$.
Now, $A = QR$ implies

$$A\boldsymbol{x} = \boldsymbol{b} \quad \Longleftrightarrow \quad QR\boldsymbol{x} = \boldsymbol{b}.$$

In the special case of a nonsingular $n \times n$ matrix A the matrix Q is also $n \times n$ with ON columns so that

$$Q^{-1} = Q^T \qquad (\text{since } Q^TQ = I)$$

and

$$QR\boldsymbol{x} = \boldsymbol{b} \quad \Longleftrightarrow \quad R\boldsymbol{x} = Q^T\boldsymbol{b}.$$

Therefore we solve $A\boldsymbol{x} = \boldsymbol{b}$ by the following steps:

1. Compute $A = QR$.
2. Compute $\boldsymbol{y} = Q^T\boldsymbol{b}$.
3. Solve the upper triangular system $R\boldsymbol{x} = \boldsymbol{y}$.

Remark

This procedure is comparable to the three-step LU solution procedure.

The real advantage of the QR factorization lies in the solution of least squares problems.
Consider $A\boldsymbol{x} = \boldsymbol{b}$ with $A \in \mathbb{R}^{m\times n}$ and $\text{rank}(A) = n$ (so that a unique least squares solution exists).
We know that the least squares solution is given by the solution of the normal equations

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b}.$$

Using the QR factorization of A this becomes

$$\begin{aligned}
& (QR)^T QR\boldsymbol{x} = (QR)^T\boldsymbol{b} \\
\iff\quad & R^T \underbrace{Q^T Q}_{=I} R\boldsymbol{x} = R^T Q^T\boldsymbol{b} \\
\iff\quad & R^T R\boldsymbol{x} = R^T Q^T\boldsymbol{b}.
\end{aligned}$$

Now R is upper triangular with positive diagonal and therefore invertible. Therefore solving the normal equations corresponds to solving (cf. the previous discussion)

$$R\boldsymbol{x} = Q^T\boldsymbol{b}.$$

Remark

This is the same as the QR factorization applied to a square and consistent system $A\boldsymbol{x} = \boldsymbol{b}$.

Summary

The QR factorization provides a simple and efficient way to solve least squares problems.

The ill-conditioned matrix $A^T A$ is never computed.

If it is required, then it can be computed from R as $R^T R$ (in fact, this is the Cholesky factorization) of $A^T A$.

Modified Gram–Schmidt

There is still a problem with the QR factorization via Gram–Schmidt:

it is not numerically stable (see HW).

A better — but still not ideal — approach is provided by the modified Gram–Schmidt algorithm.

Idea: rearrange the order of calculation, i.e., write the projection matrices

$$\mathsf{U}_k \mathsf{U}_k^T = \sum_{i=1}^{k-1} \boldsymbol{u}_i \boldsymbol{u}_i^T$$

as a sum of rank-1 projections.

MGS Algorithm

k=1: $\boldsymbol{u}_1 \leftarrow \frac{\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|}, \qquad \boldsymbol{u}_j \leftarrow \boldsymbol{x}_j,\ j = 2, \ldots, n$

for $k = 2 : n$

$$\begin{aligned}
&\mathsf{E}_k = \mathsf{I} - \boldsymbol{u}_{k-1}\boldsymbol{u}_{k-1}^T \\
&\text{for } j = k, \ldots, n \\
&\qquad \boldsymbol{u}_j \leftarrow \mathsf{E}_k \boldsymbol{u}_j \\
&\boldsymbol{u}_k \leftarrow \frac{\boldsymbol{u}_k}{\|\boldsymbol{u}_k\|}
\end{aligned}$$

Remark

- *The MGS algorithm is theoretically equivalent to the GS algorithm, i.e., in exact arithmetic, but in practice it preserves orthogonality better.*
- *Most stable implementations of the QR factorization use Householder reflections or Givens rotations (more later).*
- *Householder reflections are also more efficient than MGS.*

Unitary and Orthogonal Matrices

Definition

A real (complex) $n \times n$ matrix is called orthogonal (unitary) if its columns form an orthonormal basis for $\mathbb{R}^n$ ($\mathbb{C}^n$).

Theorem

Let U *be an orthogonal $n \times n$ matrix. Then*

1. U *has orthonormal rows.*
2. $U^{-1} = U^T$.
3. $\|U\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$*, i.e.,* U *is an isometry.*

Remark

Analogous properties for unitary matrices are formulated and proved in [Mey00].

Proof

2. By definition $\mathsf{U} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{pmatrix}$ has orthonormal columns, i.e.,

$$\begin{aligned} \boldsymbol{u}_i \perp \boldsymbol{u}_j \quad &\iff \quad \boldsymbol{u}_i^T \boldsymbol{u}_j = \delta_{ij} \\ &\iff \quad \left(\mathsf{U}^T\mathsf{U}\right)_{ij} = \delta_{ij} \\ &\iff \quad \mathsf{U}^T\mathsf{U} = \mathsf{I}. \end{aligned}$$

   But $\mathsf{U}^T\mathsf{U} = \mathsf{I}$ implies $\mathsf{U}^T = \mathsf{U}^{-1}$.

1. Therefore the statement about orthonormal rows follows from

$$\mathsf{U}\mathsf{U}^{-1} = \mathsf{U}\mathsf{U}^T = \mathsf{I}.$$

Proof (cont.)

3. To show that U is an isometry we assume U is orthogonal. Then, for any $\boldsymbol{x} \in \mathbb{R}^n$

$$\begin{aligned}\|\mathsf{U}\boldsymbol{x}\|_2^2 &= (\mathsf{U}\boldsymbol{x})^T(\mathsf{U}\boldsymbol{x}) \\ &= \boldsymbol{x}^T \underbrace{\mathsf{U}^T\mathsf{U}}_{=\mathsf{I}} \boldsymbol{x} \\ &= \|\boldsymbol{x}\|_2^2.\end{aligned}$$

□

Remark

The converse of (3) is also true, i.e., if $\|\mathsf{U}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$ *then* U *must be orthogonal. Consider* $\boldsymbol{x} = \boldsymbol{e}_i$*. Then*

$$\|\mathsf{U}\boldsymbol{e}_i\|_2^2 = \boldsymbol{u}_i^T\boldsymbol{u}_i \stackrel{(3)}{=} \|\boldsymbol{e}_i\|_2^2 = 1,$$

so the columns of U *have norm 1.*

Moreover, for $\boldsymbol{x} = \boldsymbol{e}_i + \boldsymbol{e}_j$ $(i \neq j)$ *we get*

$$\|\mathsf{U}(\boldsymbol{e}_i + \boldsymbol{e}_j)\|_2^2 = \underbrace{\boldsymbol{u}_i^T\boldsymbol{u}_i}_{=1} + \boldsymbol{u}_i^T\boldsymbol{u}_j + \boldsymbol{u}_j^T\boldsymbol{u}_i + \underbrace{\boldsymbol{u}_j^T\boldsymbol{u}_j}_{=1} \stackrel{(3)}{=} \|\boldsymbol{e}_i + \boldsymbol{e}_j\|_2^2 = 2,$$

so that $\boldsymbol{u}_i^T\boldsymbol{u}_j = 0$ *for* $i \neq j$ *and the columns of* U *are orthogonal.*

Example

- The simplest orthogonal matrix is the identity matrix I.
- Permutation matrices are orthogonal, e.g.,

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

  In fact, for permutation matrices we even have $P^T = P$ so that $P^T P = P^2 = I$. Such matrices are called involutary (see pretest).
- An orthogonal matrix can be viewed as a unitary matrix, but a unitary matrix may not be orthogonal. For example for

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ -1 & i \end{pmatrix}$$

  we have $A^* A = AA^* = I$, but $A^T A \neq I \neq AA^T$.

Elementary Orthogonal Projectors

Definition

A matrix Q of the form

$$Q = I - \boldsymbol{u}\boldsymbol{u}^T, \qquad \boldsymbol{u} \in \mathbb{R}^n, \ \|\boldsymbol{u}\|_2 = 1,$$

is called an elementary orthogonal projection.

Remark

Note that Q *is not an orthogonal matrix:*

$$Q^T = (I - \boldsymbol{u}\boldsymbol{u}^T)^T = I - \boldsymbol{u}\boldsymbol{u}^T = Q.$$

All projectors are idempotent, i.e., $Q^2 = Q$:

$$\begin{aligned} Q^T Q \overset{\text{above}}{=} Q^2 &= (I - \boldsymbol{u}\boldsymbol{u}^T)(I - \boldsymbol{u}\boldsymbol{u}^T) \\ &= I - 2\boldsymbol{u}\boldsymbol{u}^T + \boldsymbol{u}\underbrace{\boldsymbol{u}^T\boldsymbol{u}}_{=1}\boldsymbol{u}^T \\ &= (I - \boldsymbol{u}\boldsymbol{u}^T) \\ &= Q. \end{aligned}$$

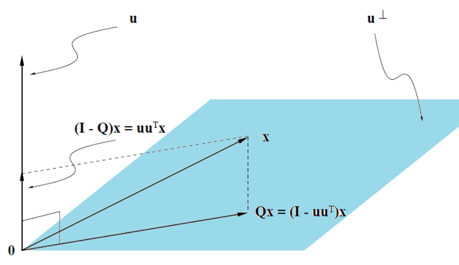
Geometric interpretation

Consider

$$\boldsymbol{x} = (\mathsf{I} - \mathsf{Q})\boldsymbol{x} + \mathsf{Q}\boldsymbol{x}$$

and observe that $(\mathsf{I} - \mathsf{Q})\boldsymbol{x} \perp \mathsf{Q}\boldsymbol{x}$:

$$\begin{aligned} ((\mathsf{I} - \mathsf{Q})\boldsymbol{x})^T \mathsf{Q}\boldsymbol{x} &= \boldsymbol{x}^T(\mathsf{I} - \mathsf{Q}^T)\mathsf{Q}\boldsymbol{x} \\ &= \boldsymbol{x}^T(\mathsf{Q} - \underbrace{\mathsf{Q}^T\mathsf{Q}}_{=\mathsf{Q}})\boldsymbol{x} = 0. \end{aligned}$$



Also,

$$(\mathsf{I} - \mathsf{Q})\boldsymbol{x} = (\boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x} = \boldsymbol{u}(\boldsymbol{u}^T\boldsymbol{x}) \in \operatorname{span}\{\boldsymbol{u}\}.$$

Therefore $\mathsf{Q}\boldsymbol{x} \in \boldsymbol{u}^\perp$, the orthogonal complement of $\boldsymbol{u}$.
Also note that $\|(\boldsymbol{u}^T\boldsymbol{x})\boldsymbol{u}\| = |\boldsymbol{u}^T\boldsymbol{x}| \underbrace{\|\boldsymbol{u}\|_2}_{=1}$, so that $|\boldsymbol{u}^T\boldsymbol{x}|$ is the length of the orthogonal projection of $\boldsymbol{x}$ onto $\operatorname{span}\{\boldsymbol{u}\}$.

Summary

- $(\mathsf{I} - \mathsf{Q})\boldsymbol{x} \in \text{span}\{\boldsymbol{u}\}$, so
$$\mathsf{I} - \mathsf{Q} = \boldsymbol{u}\boldsymbol{u}^T = P_{\boldsymbol{u}}$$
is a projection onto $\text{span}\{\boldsymbol{u}\}$.
- $\mathsf{Q}\boldsymbol{x} \in \boldsymbol{u}^\perp$, so
$$\mathsf{Q} = \mathsf{I} - \boldsymbol{u}\boldsymbol{u}^T = P_{\boldsymbol{u}^\perp}$$
is a projection onto $\boldsymbol{u}^\perp$.

Remark

Above we assumed that $\|\boldsymbol{u}\|_2 = 1$.

For an arbitrary vector $\boldsymbol{v}$ *we get a unit vector* $\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_2} = \frac{\boldsymbol{v}}{\sqrt{\boldsymbol{v}^T\boldsymbol{v}}}$.

Therefore, for general $\boldsymbol{v}$

- $P_{\boldsymbol{v}} = \frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}$ *is a projection onto* $\text{span}\{\boldsymbol{v}\}$.
- $P_{\boldsymbol{v}^\perp} = \mathsf{I} - P_{\boldsymbol{v}} = \mathsf{I} - \frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}$ *is a projection onto* $\boldsymbol{v}^\perp$.

Elementary Reflections

Definition

Let $\boldsymbol{v}(\neq \mathbf{0}) \in \mathbb{R}^n$. Then

$$\mathsf{R} = \mathsf{I} - 2\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}$$

is called the elementary (or Householder) reflector about $\boldsymbol{v}^\perp$.

Remark

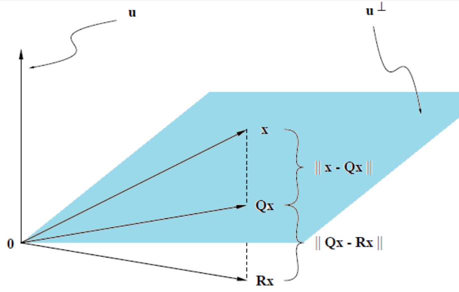
For $\boldsymbol{u} \in \mathbb{R}^n$ with $\|\boldsymbol{u}\|_2 = 1$ we have

$$\mathsf{R} = \mathsf{I} - 2\boldsymbol{u}\boldsymbol{u}^T.$$

Geometric interpretation

Consider $\|\boldsymbol{u}\|_2 = 1$, and note that $\mathsf{Q}\boldsymbol{x} = (\mathsf{I} - \boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x}$ is the orthogonal projection of $\boldsymbol{x}$ onto $\boldsymbol{u}^\perp$ as above. Also,

$$\begin{aligned} \mathsf{Q}(\mathsf{R}\boldsymbol{x}) &= \mathsf{Q}(\mathsf{I} - 2\boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x} \\ &= \mathsf{Q}\left(\mathsf{I} - 2(\mathsf{I} - \mathsf{Q})\right)\boldsymbol{x} \\ &= (\mathsf{Q} - 2\mathsf{Q} + 2\underbrace{\mathsf{Q}^2}_{=\mathsf{Q}})\boldsymbol{x} = \mathsf{Q}\boldsymbol{x}, \end{aligned}$$



so that $\mathsf{Q}\boldsymbol{x}$ is also the orthogonal projection of $\mathsf{R}\boldsymbol{x}$ onto $\boldsymbol{u}^\perp$.

Moreover, $\|\boldsymbol{x} - \mathsf{Q}\boldsymbol{x}\| = \|\boldsymbol{x} - (\mathsf{I} - \boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x}\| = |\boldsymbol{u}^T\boldsymbol{x}|\,\|\boldsymbol{u}\| = |\boldsymbol{u}^T\boldsymbol{x}|$ and

$$\begin{aligned} \|\mathsf{Q}\boldsymbol{x} - \mathsf{R}\boldsymbol{x}\| = \|(\mathsf{Q} - \mathsf{R})\boldsymbol{x}\| &= \left\| \left(\mathsf{I} - \boldsymbol{u}\boldsymbol{u}^T - (\mathsf{I} - 2\boldsymbol{u}\boldsymbol{u}^T)\right)\boldsymbol{x}\right\| \\ &= \|\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{x}\| = |\boldsymbol{u}^T\boldsymbol{x}|. \end{aligned}$$

Together, $\mathsf{R}\boldsymbol{x}$ is the reflection of $\boldsymbol{x}$ about $\boldsymbol{u}^\perp$.

Properties of elementary reflections

Theorem

Let R *be an elementary reflector. Then*

$$\mathsf{R}^{-1} = \mathsf{R}^{T} = \mathsf{R},$$

i.e., R *is orthogonal, symmetric, and involutary.*

Remark

However, these properties do not characterize a reflection, i.e., an orthogonal, symmetric and involutary matrix is not necessarily a reflection (see HW).

Proof.

$$\mathsf{R}^T = (\mathsf{I} - 2\boldsymbol{u}\boldsymbol{u}^T)^T = \mathsf{I} - 2\boldsymbol{u}\boldsymbol{u}^T = \mathsf{R}.$$

Also,

$$\begin{aligned}\mathsf{R}^2 &= (\mathsf{I} - 2\boldsymbol{u}\boldsymbol{u}^T)(\mathsf{I} - 2\boldsymbol{u}\boldsymbol{u}^T) \\ &= \mathsf{I} - 4\boldsymbol{u}\boldsymbol{u}^T + 4\boldsymbol{u}\underbrace{\boldsymbol{u}^T\boldsymbol{u}}_{=1}\boldsymbol{u}^T = \mathsf{I},\end{aligned}$$

so that $\mathsf{R}^{-1} = \mathsf{R}$. □

□

Reflection of $\boldsymbol{x}$ onto $\boldsymbol{e}_1$

If we can construct a matrix R such that $\mathsf{R}\boldsymbol{x} = \alpha \boldsymbol{e}_1$, then we can use R to zero out entries in (the first column of) a matrix.
To this end consider

$$\boldsymbol{v} = \boldsymbol{x} \pm \mu \|\boldsymbol{x}\|_2 \boldsymbol{e}_1, \quad \text{where } \mu = \begin{cases} 1 & \text{if } x_1 \text{ real,} \\ \frac{x_1}{|x_1|} & \text{if } x_1 \text{ complex,} \end{cases}$$

and note

$$\begin{aligned} \boldsymbol{v}^T \boldsymbol{v} &= (\boldsymbol{x} \pm \mu \|\boldsymbol{x}\|_2 \boldsymbol{e}_1)^T (\boldsymbol{x} \pm \mu \|\boldsymbol{x}\|_2 \boldsymbol{e}_1) \\ &= \boldsymbol{x}^T \boldsymbol{x} \pm 2\mu \|\boldsymbol{x}\|_2 \boldsymbol{e}_1^T \boldsymbol{x} + \underbrace{\mu^2}_{=1} \|\boldsymbol{x}\|_2^2 \\ &= 2(\boldsymbol{x}^T \boldsymbol{x} \pm \mu \|\boldsymbol{x}\| \boldsymbol{e}_1^T \boldsymbol{x}) = 2\boldsymbol{v}^T \boldsymbol{x}. \end{aligned} \tag{9}$$

Our Householder reflection was defined as

$$\mathsf{R} = \mathsf{I} - 2\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}$$

so that

$$\begin{aligned}\mathsf{R}\boldsymbol{x} &= \boldsymbol{x} - 2\frac{\boldsymbol{v}\boldsymbol{v}^T\boldsymbol{x}}{\boldsymbol{v}^T\boldsymbol{v}} = \boldsymbol{x} - \underbrace{\frac{2\boldsymbol{v}^T\boldsymbol{x}}{\boldsymbol{v}^T\boldsymbol{v}}}_{\stackrel{(9)}{=}1}\boldsymbol{v} \\ &= \boldsymbol{x} - \boldsymbol{v} \\ &= \underbrace{\mp\mu\|\boldsymbol{x}\|_2}_{=\alpha}\boldsymbol{e}_1.\end{aligned}$$

Remark

These special reflections are used in the Householder variant of the QR factorization. For optimal numerical stability of real matrices one lets $\mp\mu = \operatorname{sign}(x_1)$.

Remark

Since $\mathsf{R}^2 = \mathsf{I}$ ($\mathsf{R}^{-1} = \mathsf{R}$) *we have — whenever* $\|\boldsymbol{x}\|_2 = 1$ —

$$\mathsf{R}\boldsymbol{x} = \mp\mu\boldsymbol{e}_1 \quad \Longrightarrow \quad \mathsf{R}^2\boldsymbol{x} = \mp\mu\mathsf{R}\boldsymbol{e}_1 \quad \Longleftrightarrow \quad \boldsymbol{x} = \mp\mu\mathsf{R}_{*1}.$$

Therefore the matrix $\mathsf{U} = \mp\mathsf{R}$ *(taking* $|\mu| = 1$*) is orthogonal (since* R *is) and contains* $\boldsymbol{x}$ *as its first column.*

Thus, this allows us to construct an ON basis for $\mathbb{R}^n$ *that contains* $\boldsymbol{x}$ *(see example in [Mey00].*

Rotations

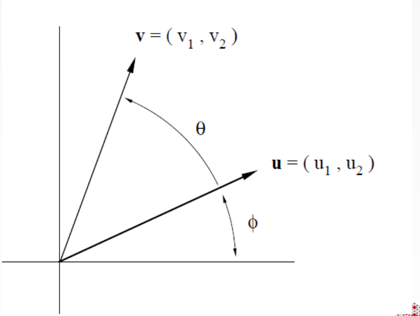
We give only a brief overview (more details can be found in [Mey00]).

We begin in $\mathbb{R}^2$ and look for a matrix representation of the rotation of a vector $\boldsymbol{u}$ into another vector $\boldsymbol{v}$, counterclockwise by an angle θ:

Here

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \|\boldsymbol{u}\| \cos\phi \\ \|\boldsymbol{u}\| \sin\phi \end{pmatrix} \qquad (10)$$

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \|\boldsymbol{v}\| \cos(\phi+\theta) \\ \|\boldsymbol{v}\| \sin(\phi+\theta) \end{pmatrix} \qquad (11)$$



We use the trigonometric identities

$$\cos(A+B) = \cos A \cos B - \sin A \sin B$$
$$\sin(A+B) = \sin A \cos B + \sin B \cos A$$

with $A = \phi$, $B = \theta$ and $\|\boldsymbol{v}\| = \|\boldsymbol{u}\|$ to get

$$\begin{aligned}
\boldsymbol{v} &\overset{(11)}{=} \begin{pmatrix} \|\boldsymbol{v}\| \cos(\phi+\theta) \\ \|\boldsymbol{v}\| \sin(\phi+\theta) \end{pmatrix} \\
&= \begin{pmatrix} \|\boldsymbol{u}\| \left(\cos\phi\cos\theta - \sin\phi\sin\theta\right) \\ \|\boldsymbol{u}\| \left(\sin\phi\cos\theta + \sin\theta\cos\phi\right) \end{pmatrix} \\
&\overset{(10)}{=} \begin{pmatrix} u_1\cos\theta - u_2\sin\theta \\ u_2\cos\theta + u_1\sin\theta \end{pmatrix} \\
&= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \boldsymbol{u} = \mathsf{P}\boldsymbol{u},
\end{aligned}$$

where P is the rotation matrix.

Remark

- *Note that*

$$\begin{aligned}\mathsf{P}^T\mathsf{P} &= \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\\ &= \begin{pmatrix}\cos^2\theta+\sin^2\theta & -\cos\theta\sin\theta+\cos\theta\sin\theta\\ -\cos\theta\sin\theta+\cos\theta\sin\theta & \sin^2\theta+\cos^2\theta\end{pmatrix}\\ &= \mathsf{I},\end{aligned}$$

  so that P *is an orthogonal matrix.*

- P^T *is also a rotation matrix (by an angle* $-\theta$*).*

Rotations about a coordinate axis in $\mathbb{R}^3$ are very similar. Such rotations are referred to a plane rotations.

For example, rotation about the x-axis (in the yz-plane) is accomplished with

$$P_{yz} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$$

Rotation about the y and z-axes is done analogously.

We can use the same ideas for plane rotations in higher dimensions.

Definition

An orthogonal matrix of the form

$$P_{ij} = \begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & c & & & s & & \\ & & & 1 & & & & \\ & & & & \ddots & & & \\ & & -s & & & c & & \\ & & & & & & 1 & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \leftarrow i \\ \\ \\ \leftarrow j \\ \\ \\ \\ \end{matrix}$$

$$\begin{matrix} \uparrow & \quad & \uparrow \\ i & & j \end{matrix}$$

with $c^2 + s^2 = 1$ is called a plane rotation (or Givens rotation).

Note that the orientation is reversed from the earlier discussion.

Usually we set

$$c = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}, \quad s = \frac{x_j}{\sqrt{x_i^2 + x_j^2}}$$

since then for $\boldsymbol{x} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}^T$

$$\mathsf{P}_{ij}\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ cx_i + sx_j \\ \vdots \\ -sx_i + cx_j \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ \frac{x_i^2+x_j^2}{\sqrt{x_i^2+x_j^2}} \\ \vdots \\ 0 \\ \vdots \\ x_n \end{pmatrix}$$

This shows that P_{ij} zeros the j^{th} component of $\boldsymbol{x}$.

Note that $\frac{x_i^2+x_j^2}{\sqrt{x_i^2+x_j^2}} = \sqrt{x_i^2 + x_j^2}$ so that repeatedly applying Givens rotations P_{ij} with the same i, but different values of j will zero out all but the i^{th} component of $\boldsymbol{x}$, and that component will become $\sqrt{x_1^2 + \ldots + x_n^2} = \|\boldsymbol{x}\|_2$.

Therefore, the sequence

$$P = P_{in} \cdots P_{i,i+1} P_{i,i-1} \cdots P_{i1}$$

of Givens rotations rotates the vector $\boldsymbol{x} \in \mathbb{R}^n$ onto $\boldsymbol{e}_i$, i.e.,

$$P\boldsymbol{x} = \|\boldsymbol{x}\|_2 \boldsymbol{e}_i.$$

Moreover, the matrix P is orthogonal.

Remark

- *Givens rotations can be used as an alternative to Householder reflections to construct a QR factorization.*
- *Householder reflections are in general more efficient, but for sparse matrices Givens rotations are more efficient because they can be applied more selectively.*

Orthogonal Reduction

[0]

Recall the form of LU factorization (Gaussian elimination):

$$T_{n-1} \cdots T_2 T_1 A = U,$$

where T_k are lower triangular and U is upper triangular, i.e., we have a triangular reduction.

For the QR factorization we will use orthogonal Householder reflectors R_k to get

$$R_{n-1} \cdots R_2 R_1 A = T,$$

where T is upper triangular, i.e., we have an orthogonal reduction.

Recall Householder reflectors

$$\mathsf{R} = \mathsf{I} - 2\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}, \qquad \text{with } \boldsymbol{v} = \boldsymbol{x} \pm \mu\|\boldsymbol{x}\|\boldsymbol{e}_1,$$

so that

$$\mathsf{R}\boldsymbol{x} = \mp\mu\|\boldsymbol{x}\|\boldsymbol{e}_1$$

and $\mu = 1$ for $\boldsymbol{x}$ real.

Now we explain how to use these Householder reflectors to convert an $m \times n$ matrix A to an upper triangular matrix of the same size, i.e., how to do a full QR factorization.

Apply Householder reflector to the first column of A:

$$\mathsf{R}_1\mathsf{A}_{*1} = \left(\mathsf{I} - 2\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}\right)\mathsf{A}_{*1} \qquad \text{with } \boldsymbol{v} = \mathsf{A}_{*1} \pm \|\mathsf{A}_{*1}\|\boldsymbol{e}_1$$

$$= \mp\|\mathsf{A}_{*1}\|\boldsymbol{e}_1 = \begin{pmatrix} t_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Then, R_1 applied to all of A yields

$$\mathsf{R}_1\mathsf{A} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \cdots & * \end{pmatrix} = \begin{pmatrix} t_{11} & \boldsymbol{t}_1^T \\ \boldsymbol{0} & \mathsf{A}_2 \end{pmatrix}$$

Next, we apply the same idea to A_2, i.e., we let

$$R_2 = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \hat{R}_2 \end{pmatrix}$$

Then

$$R_2 R_1 A = \begin{pmatrix} t_{11} & \boldsymbol{t}_1^T \\ \mathbf{0} & \hat{R}_2 A_2 \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ & t_{22} & & \boldsymbol{t}_2^T \\ \mathbf{0} & \mathbf{0} & & A_3 \end{pmatrix}$$

We continue the process until we get an upper triangular matrix, i.e.,

$$\underbrace{\mathsf{R}_n \cdots \mathsf{R}_2 \mathsf{R}_1}_{=\mathsf{P}} \mathsf{A} = \begin{pmatrix} t_{11} & & * \\ & \ddots & \vdots \\ \mathsf{O} & & t_{nn} \\ & \mathsf{O} & \end{pmatrix} \quad \text{whenever } m > n$$

or

$$\underbrace{\mathsf{R}_m \cdots \mathsf{R}_2 \mathsf{R}_1}_{=\mathsf{P}} \mathsf{A} = \begin{pmatrix} t_{11} & & * & \\ & \ddots & \vdots & * \\ \mathsf{O} & & t_{mm} & \end{pmatrix} \quad \text{whenever } n > m$$

Since each R_k is orthogonal (unitary for complex A) we have

$$\mathsf{PA} = \mathsf{T}$$

with P $m \times m$ orthogonal and T $m \times n$ upper triangular, i.e.,

$$\mathsf{A} = \mathsf{QR} \qquad (\mathsf{Q} = \mathsf{P}^T, \; \mathsf{R} = \mathsf{T})$$

Remark

- *This is similar to obtaining the QR factorization via MGS, but now* Q *is orthogonal (square) and* R *is rectangular.*
- *This gives us the full QR factorization, whereas MGS gave us the reduced QR factorization (with* $m \times n$ Q *and* $n \times n$ R*).*

Example

We use Householder reflections to find the QR factorization (where R has positive diagonal elements) of

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

$$R_1 = I - 2\frac{\boldsymbol{v}_1 \boldsymbol{v}_1^T}{\boldsymbol{v}_1^T \boldsymbol{v}_1}, \qquad \text{with } \boldsymbol{v}_1 = A_{*1} \pm \|A_{*1}\| \boldsymbol{e}_1$$

so that

$$R_1 A = \mp \|A_{*1}\| \boldsymbol{e}_1 = \mp \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix}.$$

Thus we take the $\pm$ sign as "$-$" so that $t_{11} = \sqrt{2} > 0$.

Example ((cont.))

To find $R_1 A$ we can either compute R_1 using the formula above and then compute the matrix-matrix product, or — more cheaply — note that

$$R_1 \boldsymbol{x} = \left(I - 2\frac{\boldsymbol{v}_1 \boldsymbol{v}_1^T}{\boldsymbol{v}_1^T \boldsymbol{v}_1} \right) \boldsymbol{x} = \boldsymbol{x} - 2\boldsymbol{v}_1^T \boldsymbol{x} \frac{\boldsymbol{v}_1}{\boldsymbol{v}_1^T \boldsymbol{v}_1},$$

so that we can compute $\boldsymbol{v}_1^T A_{*j}$, $j = 2, 3$, instead of the full R_1.

$$\boldsymbol{v}_1^T A_{*2} = (1 - \sqrt{2}) \cdot 2 + 0 \cdot 1 + 1 \cdot 0 = 2 - 2\sqrt{2}$$
$$\boldsymbol{v}_1^T A_{*3} = (1 - \sqrt{2}) \cdot 0 + 0 \cdot 1 + 1 \cdot 1 = 1$$

Also

$$2\frac{\boldsymbol{v}_1}{\boldsymbol{v}_1^T \boldsymbol{v}_1} = \frac{1}{2 - \sqrt{2}} \begin{pmatrix} 1 - \sqrt{2} \\ 0 \\ 1 \end{pmatrix}$$

Example ((cont.))

Therefore

$$R_1 A_{*2} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} - \underbrace{\frac{2-2\sqrt{2}}{2-\sqrt{2}}}_{=-\sqrt{2}} \begin{pmatrix} 1-\sqrt{2} \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \sqrt{2} \\ 1\sqrt{2} \end{pmatrix}$$

$$R_1 A_{*3} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 1 \\ -\frac{\sqrt{2}}{2} \end{pmatrix}$$

so that

$$R_1 A = \begin{pmatrix} \sqrt{2} & \sqrt{2} & \frac{\sqrt{2}}{2} \\ 0 & 1 & 1 \\ 0 & \sqrt{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 1 & 1 \\ \sqrt{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}$$

Example ((cont.))

Next

$$\hat{R}_2 \boldsymbol{x} = \boldsymbol{x} - 2\boldsymbol{v}_2^T \boldsymbol{x} \frac{\boldsymbol{v}_2}{\boldsymbol{v}_2^T \boldsymbol{v}_2} \quad \text{with } \boldsymbol{v}_2 = (A_2)_{*1} - \|(A_2)_{*1}\|\boldsymbol{e}_1 = \begin{pmatrix} 1-\sqrt{3} \\ \sqrt{2} \end{pmatrix}$$

$$\boldsymbol{v}_2^T (A_2)_{*1} = 3\sqrt{3}, \quad \boldsymbol{v}_2^T (A_2)_{*2} = -\sqrt{3}, \quad 2\frac{\boldsymbol{v}_2}{\boldsymbol{v}_2^T \boldsymbol{v}_2} = \frac{1}{3-\sqrt{3}} \begin{pmatrix} 1-\sqrt{3} \\ \sqrt{2} \end{pmatrix}$$

so

$$\hat{R}_2 (A_2)_{*1} = \begin{pmatrix} \sqrt{3} \\ 0 \end{pmatrix}, \qquad \hat{R}_2 (A_2)_{*2} = \begin{pmatrix} 0 \\ \frac{\sqrt{6}}{2} \end{pmatrix}$$

Using $R_2 = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \hat{R}_2 \end{pmatrix}$ we get

$$\underbrace{R_2 R_1}_{=P} A = \begin{pmatrix} \sqrt{2} & \sqrt{2} & \frac{\sqrt{2}}{2} \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \frac{\sqrt{6}}{2} \end{pmatrix} = T$$

Remark

- *As mentioned earlier, the factor* R *of the QR factorization is given by the matrix* T*.*
- *The factor* $\mathsf{Q} = \mathsf{P}^T$ *is not explicitly given in the example.*
- *One could also obtain the same answer using Givens rotations (compare [Mey00, Example 5.7.2]).*

Theorem

Let A be an $n \times n$ nonsingular real matrix. Then the factorization

$$A = QR$$

with $n \times n$ orthogonal matrix Q and $n \times n$ upper triangular matrix R with positive diagonal entries is unique.

Remark

In this $n \times n$ case the reduced and full QR factorizations coincide, i.e., the results obtained via Gram–Schmidt, Householder and Givens should be identical.

Proof

Assume we have two QR factorizations

$$A = Q_1 R_1 = Q_2 R_2 \quad \Longleftrightarrow \quad Q_2^T Q_1 = R_2 R_1^{-1} = U.$$

Now, $R_2 R_1^{-1}$ is upper triangular with positive diagonal (since each factor is) and $Q_2^T Q_1$ is orthogonal. Therefore U has all of these properties.

Since U is upper triangular

$$U_{*1} = \begin{pmatrix} u_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Moreover, since U is orthogonal $u_{11} = 1$.

Proof (cont.)

Next,

$$U_{*1}^T U_{*2} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} u_{12} \\ u_{22} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = u_{12} = 0$$

since the columns of U are orthogonal, and the fact that $\|U_{*2}\| = 1$ implies $u_{22} = 1$.

Comparing all the other pairs of columns of U shows that $U = I$, and therefore $Q_1 = Q_2$ and $R_1 = R_2$. $\square$

Recommendations (so far) for solution of $A\boldsymbol{x} = \boldsymbol{b}$

1. If A is square and nonsingular, then use LU factorization with partial pivoting. This is stable for most practical problems and requires $\mathcal{O}(\frac{n^3}{3})$ operations.
2. To find a least square solution, use QR factorization:

$$A\boldsymbol{x} = \boldsymbol{b} \iff QR\boldsymbol{x} = \boldsymbol{b} \iff R\boldsymbol{x} = Q^T\boldsymbol{b}.$$

Usually the reduced QR factorization is all that's needed.

Even though (for square nonsingular A) the Gram–Schmidt, Householder and Givens versions of the QR factorization are equivalent (due to the uniqueness theorem), we have — for general A — that

- classical GS is not stable,
- modified GS is stable for least squares, but unstable for QR (since it has problems maintaining orthogonality),
- Householder and Givens are stable, both for least squares and QR

Computational cost (for $n \times n$ matrices)

- LU with partial pivoting: $\mathcal{O}(\frac{n^3}{3})$
- Gram–Schmidt: $\mathcal{O}(n^3)$
- Householder: $\mathcal{O}(\frac{2n^3}{3})$
- Givens: $\mathcal{O}(\frac{4n^3}{3})$

Householder reflections are often the preferred method since they provide both stability and also decent efficiency.

Complementary Subspaces

[0]

Definition

Let $\mathcal{V}$ be a vector space and $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{V}$ be subspaces. $\mathcal{X}$ and $\mathcal{Y}$ are called complementary provided

$$\mathcal{V} = \mathcal{X} + \mathcal{Y} \quad \text{and} \quad \mathcal{X} \cap \mathcal{Y} = \{\mathbf{0}\}.$$

In this case, $\mathcal{V}$ is also called the direct sum of $\mathcal{X}$ and $\mathcal{Y}$, and we write

$$\mathcal{V} = \mathcal{X} \oplus \mathcal{Y}.$$

Example

- Any two lines through the origin in $\mathbb{R}^2$ are complementary.
- Any plane through the origin in $\mathbb{R}^3$ is complementary to any line through the origin not contained in the plane.
- Two planes through the origin in $\mathbb{R}^3$ are not complementary since they must intersect in a line.

Theorem

Let $\mathcal{V}$ be a vector space, and $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{V}$ be subspaces with bases $\mathcal{B}_{\mathcal{X}}$ and $\mathcal{B}_{\mathcal{Y}}$. The following are equivalent:

1. $\mathcal{V} = \mathcal{X} \oplus \mathcal{Y}$.
2. *For every $\boldsymbol{v} \in \mathcal{V}$ there exist unique $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$ such that $\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{y}$.*
3. *$\mathcal{B}_{\mathcal{X}} \cap \mathcal{B}_{\mathcal{Y}} = \{\}$ and $\mathcal{B}_{\mathcal{X}} \cup \mathcal{B}_{\mathcal{Y}}$ is a basis for $\mathcal{V}$.*

Proof.

See [Mey00]. □

Definition

Suppose $\mathcal{V} = \mathcal{X} \oplus \mathcal{Y}$, i.e., any $\boldsymbol{v} \in \mathcal{V}$ can be uniquely decomposed as $\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{y}$. Then

1. $\boldsymbol{x}$ is called the projection of $\boldsymbol{v}$ onto $\mathcal{X}$ along $\mathcal{Y}$.
2. $\boldsymbol{y}$ is called the projection of $\boldsymbol{v}$ onto $\mathcal{Y}$ along $\mathcal{X}$.

Properties of projectors

Theorem

Let $\mathcal{X}, \mathcal{Y}$ be complementary subspaces of $\mathcal{V}$. Let P*, defined by* $\mathsf{P}\boldsymbol{v} = \boldsymbol{x}$*, be the projector onto $\mathcal{X}$ along $\mathcal{Y}$. Then*

1. P *is unique.*
2. $\mathsf{P}^2 = \mathsf{P}$*, i.e.,* P *is idempotent.*
3. $\mathsf{I} - \mathsf{P}$ *is the complementary projector (onto $\mathcal{Y}$ along $\mathcal{X}$).*
4. $R(\mathsf{P}) = \{\boldsymbol{x} : \mathsf{P}\boldsymbol{x} = \boldsymbol{x}\} = \mathcal{X}$ *("fixed points" for* P*).*
5. $N(\mathsf{I} - \mathsf{P}) = \mathcal{X} = R(\mathsf{P})$ *and* $R(\mathsf{I} - \mathsf{P}) = N(\mathsf{P}) = \mathcal{Y}$*.*
6. *If $\mathcal{V} = \mathbb{R}^n$ (or $\mathbb{C}^n$), then*

$$\mathsf{P} = \begin{pmatrix} \mathsf{X} & \mathsf{O} \end{pmatrix} \begin{pmatrix} \mathsf{X} & \mathsf{Y} \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} \mathsf{X} & \mathsf{Y} \end{pmatrix} \begin{pmatrix} \mathsf{I} & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix} \begin{pmatrix} \mathsf{X} & \mathsf{Y} \end{pmatrix}^{-1},$$

where the columns of X *and* Y *are bases for $\mathcal{X}$ and $\mathcal{Y}$.*

Proof

1. Assume $\mathsf{P}_1 \boldsymbol{v} = \boldsymbol{x} = \mathsf{P}_2 \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathcal{V}$. But then $\mathsf{P}_1 = \mathsf{P}_2$.
2. We know
$$\mathsf{P}\boldsymbol{v} = \boldsymbol{x} \quad \text{for every } \boldsymbol{v} \in \mathcal{V}$$
so that
$$\mathsf{P}^2\boldsymbol{v} = \mathsf{P}(\mathsf{P}\boldsymbol{v}) = \mathsf{P}\boldsymbol{x} = \boldsymbol{x}.$$
Together we therefore have $\mathsf{P}^2 = \mathsf{P}$.
3. Using the unique decomposition of $\boldsymbol{v}$ we can write
$$\begin{aligned} & \boldsymbol{v} = \boldsymbol{x} + \boldsymbol{y} = \mathsf{P}\boldsymbol{v} + \boldsymbol{y} \\ \iff \quad & (\mathsf{I} - \mathsf{P})\boldsymbol{v} = \boldsymbol{y}, \end{aligned}$$
the projection of $\boldsymbol{v}$ onto $\mathcal{Y}$ along $\mathcal{X}$.

Proof (cont.)

4. Note that $\boldsymbol{x} \in R(\mathsf{P})$ if and only if $\boldsymbol{x} = \mathsf{P}\boldsymbol{x}$. This is true since if $\boldsymbol{x} = \mathsf{P}\boldsymbol{x}$ then $\boldsymbol{x}$ obviously in $R(\mathsf{P})$. On the other hand, if $\boldsymbol{x} \in R(\mathsf{P})$ then $\boldsymbol{x} = \mathsf{P}\boldsymbol{v}$ for some $\boldsymbol{v} \in \mathcal{V}$ and so

$$\mathsf{P}\boldsymbol{x} = \mathsf{P}^2\boldsymbol{v} \stackrel{(2)}{=} \mathsf{P}\boldsymbol{v} = \boldsymbol{x}.$$

Therefore

$$\begin{aligned} R(\mathsf{P}) &= \{\boldsymbol{x} : \ \boldsymbol{x} = \mathsf{P}\boldsymbol{v}, \ \boldsymbol{v} \in \mathcal{V}\} = \mathcal{X} \\ &= \{\boldsymbol{x} : \ \mathsf{P}\boldsymbol{x} = \boldsymbol{x}\}. \end{aligned}$$

5. Since $N(\mathsf{I}-\mathsf{P}) = \{\boldsymbol{x} : \ (\mathsf{I}-\mathsf{P})\boldsymbol{x} = \boldsymbol{0}\}$, and

$$(\mathsf{I}-\mathsf{P})\boldsymbol{x} = \boldsymbol{0} \quad \Longleftrightarrow \quad \mathsf{P}\boldsymbol{x} = \boldsymbol{x}$$

we have $N(\mathsf{I}-\mathsf{P}) = \mathcal{X} = R(\mathsf{P})$.
The claim $R(\mathsf{I}-\mathsf{P}) = \mathcal{Y} = N(\mathsf{P})$ is shown similarly.

Proof (cont.)

6. Take $B = \begin{pmatrix} X & Y \end{pmatrix}$, where the columns of X and Y form a basis for $\mathcal{X}$ and $\mathcal{Y}$, respectively.
Then the columns of B form a basis for $\mathcal{V}$ and B is nonsingular.
From above we have $P\boldsymbol{x} = \boldsymbol{x}$, where $\boldsymbol{x}$ can be any column of X.
Also, $P\boldsymbol{y} = \mathbf{0}$, where $\boldsymbol{y}$ is any column of Y.
So

$$PB = P\begin{pmatrix} X & Y \end{pmatrix} = \begin{pmatrix} X & O \end{pmatrix}$$

or

$$P = \begin{pmatrix} X & O \end{pmatrix} B^{-1} = \begin{pmatrix} X & Y \end{pmatrix}^{-1}.$$

This establishes the first part of (6).
The second part follows by noting that

$$B\begin{pmatrix} I & O \\ O & O \end{pmatrix} = \begin{pmatrix} X & Y \end{pmatrix}\begin{pmatrix} I & O \\ O & O \end{pmatrix} = \begin{pmatrix} X & O \end{pmatrix}.$$

□

We just saw that any projector is idempotent, i.e., $P^2 = P$. In fact,

Theorem

A matrix P is a projector if and only if $P^2 = P$.

Proof.

One direction is given above. For the other see [Mey00]. □

Remark

This theorem is sometimes used to define projectors.

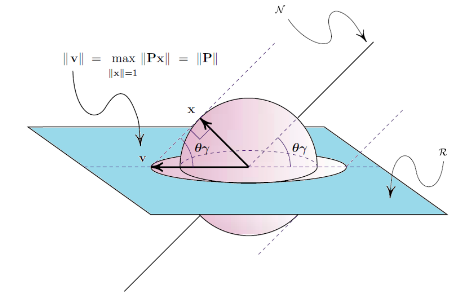
Angle between subspaces

In some applications, e.g., when determining the convergence rates of iterative algorithms, it is useful to know the angle between subspaces.

If $\mathcal{R}, \mathcal{N}$ are complementary then

$$\sin\theta = \frac{1}{\|P\|_2} = \frac{1}{\lambda_{\max}} = \frac{1}{\sigma_1},$$

where P is the projector onto $\mathcal{R}$ along $\mathcal{N}$, $\lambda_{\max}$ is the largest eigenvalue of $P^T P$ and σ_1 is the largest singular value of P.



See [Mey00, Example 5.9.2] for more details.

Remark

We will skip [Mey00, Section 5.10] on the range–nullspace decomposition.

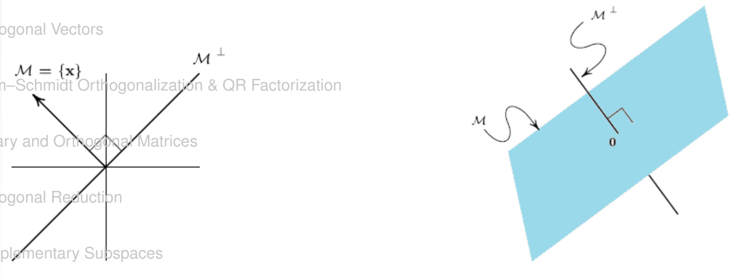
While the range–nullspace decomposition is theoretically important, its practical usefulness is limited because computation is very unstable due to lack of orthogonality.

This also means we will not discuss nilpotent matrices and — later on — the Jordan normal form.

Definition
[0]

Let $\mathcal{V}$ be an inner product space and $\mathcal{M} \subseteq \mathcal{V}$. The orthogonal complement $\mathcal{M}^{\perp}$ of $\mathcal{M}$ is

$$\mathcal{M}^{\perp} = \{\boldsymbol{x} \in \mathcal{V} : \langle \boldsymbol{m}, \boldsymbol{x} \rangle = 0 \text{ for all } \boldsymbol{m} \in \mathcal{M}\}.$$



Remark

Even if $\mathcal{M}$ is not a subspace of $\mathcal{V}$ (i.e., only a subset), $\mathcal{M}^{\perp}$ is (see HW).

Theorem

Let $\mathcal{V}$ be an inner product space and $\mathcal{M} \subseteq \mathcal{V}$. If $\mathcal{M}$ is a subspace of $\mathcal{V}$, then

$$\mathcal{V} = \mathcal{M} \oplus \mathcal{M}^{\perp}.$$

Proof

According to the definition of complementary subspaces we need to show

1. $\mathcal{M} \cap \mathcal{M}^{\perp} = \{\mathbf{0}\}$,
2. $\mathcal{M} + \mathcal{M}^{\perp} = \mathcal{V}$.

Proof (cont.)

1. Let's assume there exists an $\boldsymbol{x} \in \mathcal{M} \cap \mathcal{M}^\perp$, i.e., $\boldsymbol{x} \in \mathcal{M}$ and $\boldsymbol{x} \in \mathcal{M}^\perp$.

   The definition of $\mathcal{M}^\perp$ implies

   $$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0.$$

   But then the definition of an inner product implies $\boldsymbol{x} = \boldsymbol{0}$.

   This is true for any $\boldsymbol{x} \in \mathcal{M} \cap \mathcal{M}^\perp$, so $\boldsymbol{x} = \boldsymbol{0}$ is the only such vector.

Proof (cont.)

2. We let $\mathcal{B}_{\mathcal{M}}$ and $\mathcal{B}_{\mathcal{M}^\perp}$ be ON bases for $\mathcal{M}$ and $\mathcal{M}^\perp$, respectively.

   Since $\mathcal{M} \cap \mathcal{M}^\perp = \{\mathbf{0}\}$ we know that $\mathcal{B}_{\mathcal{M}} \cup \mathcal{B}_{\mathcal{M}^\perp}$ is an ON basis for some $\mathcal{S} \subseteq \mathcal{V}$.

   In fact, $\mathcal{S} = \mathcal{V}$ since otherwise we could extend $\mathcal{B}_{\mathcal{M}} \cup \mathcal{B}_{\mathcal{M}^\perp}$ to an ON basis of $\mathcal{V}$ (using the extension theorem and GS).

   However, any vector in the extension must be orthogonal to $\mathcal{M}$, i.e., in $\mathcal{M}^\perp$, but this is not possible since the extended basis must be linearly independent.

   Therefore, the extension set is empty.

□

Theorem

Let $\mathcal{V}$ be an inner product space with $\dim(\mathcal{V}) = n$ and $\mathcal{M}$ be a subspace of $\mathcal{V}$. Then

1. $\dim \mathcal{M}^{\perp} = n - \dim \mathcal{M}$,
2. $\mathcal{M}^{\perp\perp} = \mathcal{M}$.

Proof

For (1) recall our dimension formula from Chapter 4

$$\dim(\mathcal{X} + \mathcal{Y}) = \dim \mathcal{X} + \dim \mathcal{Y} - \dim(\mathcal{X} \cap \mathcal{Y}).$$

Here $\mathcal{M} \cap \mathcal{M}^{\perp} = \{\mathbf{0}\}$, so that $\dim(\mathcal{M} \cap \mathcal{M}^{\perp}) = 0$.

Also, since $\mathcal{M}$ is a subspace of $\mathcal{V}$ we have $\mathcal{V} = \mathcal{M} + \mathcal{M}^{\perp}$ and the dimension formula implies (1).

Proof (cont.)

2. Instead of directly establishing equality we first show that $\mathcal{M}^{\perp\perp} \subseteq \mathcal{M}$.
Since $\mathcal{M} \oplus \mathcal{M}^{\perp} = \mathcal{V}$ any $\boldsymbol{x} \in \mathcal{V}$ can be uniquely decomposed into

$$\boldsymbol{x} = \boldsymbol{m} + \boldsymbol{n} \quad \text{with } \boldsymbol{m} \in \mathcal{M},\; \boldsymbol{n} \in \mathcal{M}^{\perp}.$$

Now we take $\boldsymbol{x} \in \mathcal{M}^{\perp\perp}$ so that $\langle \boldsymbol{x}, \boldsymbol{n} \rangle = 0$ for all $\boldsymbol{n} \in \mathcal{M}^{\perp}$, and therefore

$$0 = \langle \boldsymbol{x}, \boldsymbol{n} \rangle = \langle \boldsymbol{m} + \boldsymbol{n}, \boldsymbol{n} \rangle = \underbrace{\langle \boldsymbol{m}, \boldsymbol{n} \rangle}_{=0} + \langle \boldsymbol{n}, \boldsymbol{n} \rangle.$$

But

$$\langle \boldsymbol{n}, \boldsymbol{n} \rangle = 0 \quad \Longleftrightarrow \quad \boldsymbol{n} = \boldsymbol{0},$$

and therefore $\boldsymbol{x} = \boldsymbol{m}$ is in $\mathcal{M}$.

Proof (cont.)

Now, recall from Chapter 4 that for subspaces $\mathcal{X} \subseteq \mathcal{Y}$

$$\dim \mathcal{X} = \dim \mathcal{Y} \quad \Longrightarrow \quad \mathcal{X} = \mathcal{Y}.$$

We take $\mathcal{X} = \mathcal{M}^{\perp^\perp}$ and $\mathcal{Y} = \mathcal{M}$ (and know from the work just performed that $\mathcal{M}^{\perp^\perp}$ is a subspace of $\subseteq \mathcal{M}$).
From (1) we know

$$\begin{aligned}\dim \mathcal{M}^{\perp} &= n - \dim \mathcal{M} \\ \dim \mathcal{M}^{\perp^\perp} &= n - \dim \mathcal{M}^{\perp} \\ &= n - (n - \dim \mathcal{M}) = \dim \mathcal{M}.\end{aligned}$$

But then $\mathcal{M}^{\perp^\perp} = \mathcal{M}$. □

Back to Fundamental Subspaces

Theorem

Let A *be a real* $m \times n$ *matrix. Then*

1. $R(\mathsf{A})^\perp = N(\mathsf{A}^T)$,
2. $N(\mathsf{A})^\perp = R(\mathsf{A}^T)$.

Corollary

$$\mathbb{R}^m = \underbrace{R(\mathsf{A})}_{\subseteq \mathbb{R}^m} \oplus R(\mathsf{A})^\perp = R(\mathsf{A}) \oplus N(\mathsf{A}^T),$$

$$\mathbb{R}^n = \underbrace{N(\mathsf{A})}_{\subseteq \mathbb{R}^n} \oplus N(\mathsf{A})^\perp = N(\mathsf{A}) \oplus R(\mathsf{A}^T).$$

Proof (of Theorem)

1. We show that $\boldsymbol{x} \in R(\mathsf{A})^\perp$ implies $\boldsymbol{x} \in N(\mathsf{A}^T)$ and vice versa.

$$\begin{aligned}
\boldsymbol{x} \in R(\mathsf{A})^\perp \quad &\iff \quad \langle \mathsf{A}\boldsymbol{y}, \boldsymbol{x} \rangle = 0 \quad \text{for any } \boldsymbol{y} \in \mathbb{R}^n \\
&\iff \quad \boldsymbol{y}^T \mathsf{A}^T \boldsymbol{x} = 0 \quad \text{for any } \boldsymbol{y} \in \mathbb{R}^n \\
&\iff \quad \langle \boldsymbol{y}, \mathsf{A}^T \boldsymbol{x} \rangle = 0 \quad \text{for any } \boldsymbol{y} \in \mathbb{R}^n \\
&\iff \quad \mathsf{A}^T \boldsymbol{x} = \boldsymbol{0} \quad \iff \quad \boldsymbol{x} \in N(\mathsf{A}^T)
\end{aligned}$$

by the definitions of these subspaces and of an inner product.

2. Using (1), we have

$$\begin{aligned}
R(\mathsf{A})^\perp \stackrel{(1)}{=} N(\mathsf{A}^T) \quad &\stackrel{\perp}{\iff} \quad R(\mathsf{A}) = N(\mathsf{A}^T)^\perp \\
&\stackrel{\mathsf{A} \to \mathsf{A}^T}{\iff} \quad R(\mathsf{A}^T) = N(\mathsf{A})^\perp.
\end{aligned}$$

□

Starting to think about the SVD

The decompositions of $\mathbb{R}^m$ and $\mathbb{R}^n$ from the corollary help prepare for the SVD of an $m \times n$ matrix A.

Assume rank(A) $= r$ and let

$$\begin{aligned}
\mathcal{B}_{R(\mathsf{A})} &= \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\} && \text{ON basis for } R(\mathsf{A}) \subseteq \mathbb{R}^m,\\
\mathcal{B}_{N(\mathsf{A}^T)} &= \{\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m\} && \text{ON basis for } N(\mathsf{A}^T) \subseteq \mathbb{R}^m,\\
\mathcal{B}_{R(\mathsf{A}^T)} &= \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\} && \text{ON basis for } R(\mathsf{A}^T) \subseteq \mathbb{R}^n,\\
\mathcal{B}_{N(\mathsf{A})} &= \{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\} && \text{ON basis for } N(\mathsf{A}) \subseteq \mathbb{R}^n.
\end{aligned}$$

By the corollary

$$\begin{aligned}
&\mathcal{B}_{R(\mathsf{A})} \cup \mathcal{B}_{N(\mathsf{A}^T)} && \text{ON basis for } \mathbb{R}^m,\\
&\mathcal{B}_{R(\mathsf{A}^T)} \cup \mathcal{B}_{N(\mathsf{A})} && \text{ON basis for } \mathbb{R}^n,
\end{aligned}$$

and therefore the following are orthogonal matrices

$$\begin{aligned}
\mathsf{U} &= \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m \end{pmatrix}\\
\mathsf{V} &= \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{pmatrix}.
\end{aligned}$$

Consider

$$R = U^T A V = \left(\boldsymbol{u}_i^T A \boldsymbol{v}_j \right)_{i,j=1}^{m,n}.$$

Note that

$$\begin{aligned} A\boldsymbol{v}_j &= \boldsymbol{0}, \qquad j = r+1, \ldots, n, \\ \boldsymbol{u}_i^T A = \boldsymbol{0} \quad &\Longleftrightarrow \quad A^T \boldsymbol{u}_i = \boldsymbol{0}, \qquad i = r+1, \ldots, m, \end{aligned}$$

so

$$R = \begin{pmatrix} \boldsymbol{u}_1^T A \boldsymbol{v}_1 & \cdots & \boldsymbol{u}_1^T A \boldsymbol{v}_r & \\ \vdots & & \vdots & O \\ \boldsymbol{u}_r^T A \boldsymbol{v}_1 & \cdots & \boldsymbol{u}_r^T A \boldsymbol{v}_r & \\ & O & & O \end{pmatrix} = \begin{pmatrix} C_{r \times r} & O \\ O & O \end{pmatrix}.$$

Thus

$$\mathsf{R} = \mathsf{U}^T\mathsf{A}\mathsf{V} = \begin{pmatrix} \mathsf{C}_{r\times r} & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix}$$

$$\iff \quad \mathsf{A} = \mathsf{U}\mathsf{R}\mathsf{V}^T = \mathsf{U}\begin{pmatrix} \mathsf{C}_{r\times r} & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix}\mathsf{V}^T,$$

the URV factorization of A.

Remark

The matrix $\mathsf{C}_{r\times r}$ is nonsingular since

$$\text{rank}(\mathsf{C}) = \text{rank}(\mathsf{U}^T\mathsf{A}\mathsf{V}) = \text{rank}(\mathsf{A}) = r$$

because multiplication by the orthogonal (and therefore nonsingular) matrices U^T and V does not change the rank of A.

We have now shown that the ON bases for the fundamental subspaces of A yield the URV factorization.

As we show next, the converse is also true, i.e., any URV factorization of A yields a ON bases for the fundamental subspaces of A.

However, the URV factorization is not unique. Different ON bases result in different factorizations.

Consider $A = URV^T$ with U, V orthogonal $m \times m$ and $n \times n$ matrices, respectively, and $R = \begin{pmatrix} C & O \\ O & O \end{pmatrix}$ with C nonsingular.

We partition

$$U = \begin{pmatrix} \underbrace{U_1}_{m\times r} & \underbrace{U_2}_{m\times m-r} \end{pmatrix}, \quad V = \begin{pmatrix} \underbrace{V_1}_{n\times r} & \underbrace{V_2}_{n\times n-r} \end{pmatrix}$$

Then V (and therefore also V^T) is nonsingular and we see that

$$\begin{aligned} R(A) &= R(URV^T) \\ &= R(UR) \\ &= R\left(\begin{pmatrix} U_1C & O \end{pmatrix}\right) = R(\underbrace{U_1C}_{m\times r}) \overset{\text{rank}(C)=r}{=} R(U_1) \end{aligned} \tag{12}$$

so that the columns of U_1 are an ON basis for $R(A)$.

Moreover,

$$N(\mathsf{A}^T) \overset{\text{prev. thm}}{=} R(\mathsf{A})^\perp \overset{(12)}{=} R(\mathsf{U}_1)^\perp = R(\mathsf{U}_2)$$

since U is orthogonal and $\mathbb{R}^m = R(\mathsf{U}_1) \oplus R(\mathsf{U}_2)$.

This implies that the columns of U_2 are an ON basis for $N(\mathsf{A}^T)$.

The other two cases can be argued similarly using $N(\mathsf{AB}) = N(\mathsf{B})$ provided $\text{rank}(\mathsf{A}) = n$.

The main difference between a URV factorization and the SVD is that the SVD will contain a diagonal matrix Σ with r nonzero singular values, while R contains the full $r \times r$ block C.

As a first step in this direction, we can easily obtain a URV factorization of A with a lower triangular matrix C.

Idea: use Householder reflections (or Givens rotations)

Consider an $m \times n$ matrix A.
We apply an $m \times m$ orthogonal (Householder reflection) matrix P so that

$$A \quad \longrightarrow \quad PA = \begin{pmatrix} B \\ O \end{pmatrix}, \quad \text{with } r \times m \text{ matrix } B, \; \text{rank}(B) = r.$$

Next, use $n \times n$ orthogonal Q as follows:

$$B^T \quad \longrightarrow \quad QB^T = \begin{pmatrix} T \\ O \end{pmatrix}, \quad \text{with } r \times r \text{ upper triangular } T, \; \text{rank}(T) = r.$$

Then

$$BQ^T = \begin{pmatrix} T^T & O \end{pmatrix} \quad \Longleftrightarrow \quad B = \begin{pmatrix} T^T & O \end{pmatrix} Q$$

and

$$\begin{pmatrix} B \\ O \end{pmatrix} = \begin{pmatrix} T^T & O \\ O & O \end{pmatrix} Q.$$

Together,

$$\begin{aligned} \mathsf{PA} &= \begin{pmatrix} \mathsf{T}^T & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix} \mathsf{Q} \\ \iff \quad \mathsf{A} &= \mathsf{P}^T \begin{pmatrix} \mathsf{T}^T & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix} \mathsf{Q}, \end{aligned}$$

a URV factorization with lower triangular block T^T.

Remark

See HW for an example of this process with numbers.

Singular Value Decomposition

[0]

We know

$$A = URV^T = U \begin{pmatrix} C & O \\ O & O \end{pmatrix} V^T,$$

where C is upper triangular and U, V are orthogonal.

Now we want to establish that C can even be made diagonal.

Note that

$$\|A\|_2 = \|C\|_2 =: \sigma_1$$

since multiplication by an orthogonal matrix does not change the 2-norm (see HW).

Also,

$$\|C\|_2 = \max_{\|\boldsymbol{z}\|_2=1} \|C\boldsymbol{z}\|_2$$

so that

$$\|C\|_2 = \|C\boldsymbol{x}\|_2 \quad \text{for some } \boldsymbol{x}, \ \|\boldsymbol{x}\|_2 = 1.$$

In fact (see Sect.5.2), $\boldsymbol{x}$ is such that $(C^T C - \lambda I)\boldsymbol{x} = \boldsymbol{0}$, i.e., $\boldsymbol{x}$ is an eigenvector of $C^T C$ so that

$$\|C\|_2 = \sigma_1 = \sqrt{\lambda} = \sqrt{\boldsymbol{x}^T C^T C \boldsymbol{x}}. \tag{13}$$

Since $\boldsymbol{x}$ is a unit vector we can extend it to an orthogonal matrix

$$\mathsf{R}_{\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{x} & \mathsf{X} \end{pmatrix},$$

e.g., using Householder reflectors as discussed at the end of Sect.5.6.

Similarly, let

$$\boldsymbol{y} = \frac{\mathsf{C}\boldsymbol{x}}{\|\mathsf{C}\boldsymbol{x}\|_2} = \frac{\mathsf{C}\boldsymbol{x}}{\sigma_1}. \tag{14}$$

Then

$$\mathsf{R}_{\boldsymbol{y}} = \begin{pmatrix} \boldsymbol{y} & \mathsf{Y} \end{pmatrix}$$

is also orthogonal (and Hermitian/symmetric) since it's a Householder reflector.

Now

$$\underbrace{R_{\boldsymbol{y}}^T}_{=R_{\boldsymbol{y}}} C R_{\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{y}^T \\ Y^T \end{pmatrix} C \begin{pmatrix} \boldsymbol{x} & X \end{pmatrix} = \begin{pmatrix} \boldsymbol{y}^T C \boldsymbol{x} & \boldsymbol{y}^T C X \\ Y^T C \boldsymbol{x} & Y^T C X \end{pmatrix}.$$

From above

$$\sigma_1^2 = \lambda \overset{(13)}{=} \boldsymbol{x}^T C^T C \boldsymbol{x} \overset{(14)}{=} \sigma_1 \boldsymbol{y}^T C \boldsymbol{x}$$

$$\Longrightarrow \quad \boldsymbol{y}^T C \boldsymbol{x} = \sigma_1.$$

Also,

$$Y^T C \boldsymbol{x} \overset{(14)}{=} Y^T (\sigma_1 \boldsymbol{y}) = \boldsymbol{0}$$

since $R_{\boldsymbol{y}}$ is orthogonal, i.e., $\boldsymbol{y}$ is orthogonal to the columns of Y.

Let $Y^T C X = C_2$ and $\boldsymbol{y}^T C X = \boldsymbol{c}^T$ so that

$$R_{\boldsymbol{y}} C R_{\boldsymbol{x}} = \begin{pmatrix} \sigma_1 & \boldsymbol{c}^T \\ \boldsymbol{0} & C_2 \end{pmatrix}.$$

To show that $\boldsymbol{c}^T = \boldsymbol{0}^T$ consider

$$\begin{aligned} \boldsymbol{c}^T &= \boldsymbol{y}^T C X \overset{(14)}{=} \left(\frac{C\boldsymbol{x}}{\sigma_1}\right)^T C X \\ &= \frac{\boldsymbol{x}^T C^T C X}{\sigma_1}. \end{aligned} \tag{15}$$

From (13) $\boldsymbol{x}$ is an eigenvector of $C^T C$, i.e.,

$$C^T C \boldsymbol{x} = \lambda \boldsymbol{x} = \sigma_1^2 \boldsymbol{x} \quad \Longleftrightarrow \quad \boldsymbol{x}^T C^T C = \sigma_1^2 \boldsymbol{x}^T.$$

Plugging this into (15) yields

$$\boldsymbol{c}^T = \sigma_1 \boldsymbol{x}^T X = \boldsymbol{0}$$

since $R_{\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{x} & X \end{pmatrix}$ is orthogonal.

Moreover, $\sigma_1 \geq \|C_2\|_2$ since

$$\sigma_1 = \|C\|_2 \overset{\text{HW}}{=} \|R_{\boldsymbol{y}} C R_{\boldsymbol{x}}\|_2 = \max\{\sigma_1, \|C_2\|_2\}.$$

Next, we repeat this process for C_2, i.e.,

$$S_{\boldsymbol{y}} C_2 S_{\boldsymbol{x}} = \begin{pmatrix} \sigma_2 & \mathbf{0}^T \\ \mathbf{0} & C_3 \end{pmatrix} \quad \text{with} \quad \sigma_2 \geq \|C_3\|_2.$$

Let

$$P_2 = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & S_{\boldsymbol{y}}^T \end{pmatrix} R_{\boldsymbol{y}}^T, \quad Q_2 = R_{\boldsymbol{x}} \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & S_{\boldsymbol{x}} \end{pmatrix}.$$

Then

$$P_2 C Q_2 = \begin{pmatrix} \sigma_1 & 0 & \mathbf{0}^T \\ 0 & \sigma_2 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0} & C_3 \end{pmatrix} \quad \text{with } \sigma_1 \geq \sigma_2 \geq \|C_3\|_2.$$

We continue this until

$$\mathsf{P}_{r-1}\mathsf{C}\mathsf{Q}_{r-1} = \begin{pmatrix} \sigma_1 & & & \mathsf{O} \\ & \sigma_2 & & \\ & & \ddots & \\ \mathsf{O} & & & \sigma_r \end{pmatrix} = \mathsf{D}, \quad \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r.$$

Finally, let

$$\widetilde{\mathsf{U}}^T = \begin{pmatrix} \mathsf{P}_{r-1} & \mathsf{O} \\ \mathsf{O} & \mathsf{I} \end{pmatrix} \mathsf{U}^T, \quad \text{and} \quad \widetilde{\mathsf{V}} = \begin{pmatrix} \mathsf{Q}_{r-1} & \mathsf{O} \\ \mathsf{O} & \mathsf{I} \end{pmatrix}.$$

Together,

$$\widetilde{\mathsf{U}}^T \mathsf{A} \widetilde{\mathsf{V}} = \begin{pmatrix} \mathsf{D} & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix}$$

or — without the tildes — the singular value decomposition (SVD) of A

$$\mathsf{A} = \mathsf{U} \begin{pmatrix} \mathsf{D} & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix} \mathsf{V}^T,$$

where A is $m \times n$, U is $m \times m$, $\mathsf{D} = r \times r$ and $\mathsf{V} = n \times n$.

We use the following terminology:

singular values: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$,

left singular vectors: columns of U,

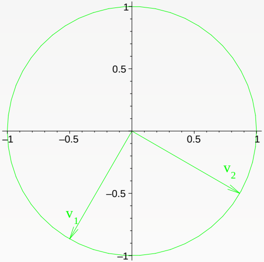
right singular vectors: columns of V.

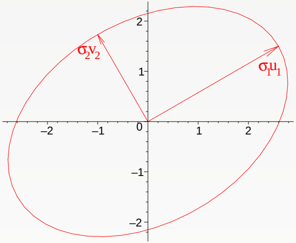
Remark

In Chapter 7 we will see that the columns of U *and* V *are also special eigenvectors of* $A^T A$.

Geometric interpretation of SVD

For the following we assume $A \in \mathbb{R}^{n\times n}$, $n = 2$.





This picture is true since

$$A = UDV^T \quad \iff \quad AV = UD$$

and σ_1, σ_2 are the lengths of the semi-axes of the ellipse because $\|\boldsymbol{u}_1\| = \|\boldsymbol{u}_2\| = 1$.

Remark

See [Mey00] for more details.

For general n, A transforms the 2-norm unit sphere to an ellipsoid whose semi-axes have lengths

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n.$$

Therefore,

$$\kappa_2(\mathsf{A}) = \frac{\sigma_1}{\sigma_n}$$

is the distortion ratio of the transformation A.
Moreover,

$$\sigma_1 = \|\mathsf{A}\|_2, \qquad \sigma_n = \frac{1}{\|\mathsf{A}^{-1}\|_2}$$

so that

$$\kappa_2(\mathsf{A}) = \|\mathsf{A}\|_2 \|\mathsf{A}^{-1}\|_2$$

is the 2-norm condition number of A ($\in \mathbb{R}^{n \times n}$).

Remark

The relations for σ_1 and σ_n hold because

$$\|A\|_2 = \|UDV^T\|_2 \stackrel{HW}{=} \|D\|_2 = \sigma_1$$
$$\|A^{-1}\|_2 = \|VD^{-1}U^T\|_2 \stackrel{HW}{=} \|D^{-1}\|_2 = \frac{1}{\sigma_n}$$

Remark

We always have $\kappa_2(A) \geq 1$, and $\kappa_2(A) = 1$ if and only if A is a multiple of an orthogonal matrix (typo in [Mey00], see proof on next slide).

Proof

"$\Longleftarrow$": Assume $A = \alpha Q$ with $\alpha > 0$, Q orthogonal, i.e.,

$$\|A\|_2 = \alpha \|Q\|_2 = \alpha \max_{\|\boldsymbol{x}\|_2=1} \|Q\boldsymbol{x}\|_2 \stackrel{\text{invariance}}{=} \alpha \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{x}\|_2 = \alpha.$$

Also

$$A^T A = \alpha^2 Q^T Q = \alpha^2 I \quad \Longrightarrow \quad A^{-1} = \frac{1}{\alpha^2} A^T \quad \text{and} \quad \|A^T\|_2 = \|A\|_2$$

so that $\|A^{-1}\|_2 = \frac{1}{\alpha}$ and

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \alpha \frac{1}{\alpha} = 1.$$

Proof (cont.)

"$\Longrightarrow$": Assume $\kappa_2(A) = \frac{\sigma_1}{\sigma_n} = 1$ so that $\sigma_1 = \sigma_n$ and therefore

$$D = \sigma_1 I.$$

Thus

$$A = UDV^T = \sigma_1 UV^T$$

and

$$\begin{aligned} A^T A &= \sigma_1^2 (UV^T)^T UV^T \\ &= \sigma_1^2 VU^T UV^T = \sigma_1^2 I. \end{aligned}$$

□

Applications of the Condition Number

Let $\tilde{\boldsymbol{x}}$ be the answer obtained by solving $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$ with $\mathsf{A} \in \mathbb{R}^{n \times n}$.

Is a small residual

$$\boldsymbol{r} = \boldsymbol{b} - \mathsf{A}\tilde{\boldsymbol{x}}$$

a good indicator for the accuracy of $\tilde{\boldsymbol{x}}$?

Since $\boldsymbol{x}$ is the exact answer, and $\tilde{\boldsymbol{x}}$ the computed answer we have the relative error

$$\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|}.$$

Now

$$\begin{aligned}\|\boldsymbol{r}\| &= \|\boldsymbol{b} - \mathsf{A}\tilde{\boldsymbol{x}}\| = \|\mathsf{A}\boldsymbol{x} - \mathsf{A}\tilde{\boldsymbol{x}}\| \\ &= \|\mathsf{A}(\boldsymbol{x} - \tilde{\boldsymbol{x}})\| \le \|\mathsf{A}\|\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|.\end{aligned}$$

To get the relative error we multiply by $\frac{\|\mathsf{A}^{-1}\boldsymbol{b}\|}{\|\boldsymbol{x}\|} = 1$.
Then

$$\begin{aligned}\|\boldsymbol{r}\| &\le \|\mathsf{A}\|\|\mathsf{A}^{-1}\boldsymbol{b}\|\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \\ \frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} &\le \kappa(\mathsf{A})\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|}. \qquad (16)\end{aligned}$$

Moreover, using $\boldsymbol{r} = \boldsymbol{b} - \mathsf{A}\tilde{\boldsymbol{x}} = \boldsymbol{b} - \tilde{\boldsymbol{b}}$,

$$\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\| = \|\mathsf{A}^{-1}(\boldsymbol{b} - \tilde{\boldsymbol{b}})\| \le \|\mathsf{A}^{-1}\|\|\boldsymbol{r}\|.$$

Multiplying by $\frac{\|\mathsf{A}\boldsymbol{x}\|}{\|\boldsymbol{b}\|} = 1$ we have

$$\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \le \kappa(\mathsf{A})\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}. \tag{17}$$

Combining (16) and (17) yields

$$\frac{1}{\kappa(\mathsf{A})}\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \le \frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \le \kappa(\mathsf{A})\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}.$$

Therefore, the relative residual $\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}$ is a good indicator of relative error if and only if A is well conditioned, i.e., $\kappa(\mathsf{A})$ is small (close to 1).

Applications of the SVD

1. Determination of "numerical rank(A)":

   rank(A) $\approx$ index of smallest singular value greater or equal a desired threshold

2. Low-rank approximation of A:

   The Eckart–Young theorem states that

$$\mathsf{A}_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

   is the best rank k approximation to A in the 2-norm (also the Frobenius norm), i.e.,

$$\|\mathsf{A} - \mathsf{A}_k\|_2 = \min_{\text{rank}(\mathsf{B})=k} \|\mathsf{A} - \mathsf{B}\|_2.$$

   Moreover,

$$\|\mathsf{A} - \mathsf{A}_k\|_2 = \sigma_{k+1}.$$

   Run `SVD_movie.m`

3. Stable solution of least squares problems:
Use Moore–Penrose pseudoinverse

Definition

Let $\mathsf{A} \in \mathbb{R}^{m\times n}$ and

$$\mathsf{A} = \mathsf{U}\begin{pmatrix} \mathsf{D} & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix}\mathsf{V}^T$$

be the SVD of A. Then

$$\mathsf{A}^\dagger = \mathsf{V}\begin{pmatrix} \mathsf{D}^{-1} & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix}\mathsf{U}^T$$

is called the Moore–Penrose pseudoinverse of A.

Remark

Note that $\mathsf{A}^\dagger \in \mathbb{R}^{n\times m}$ *and*

$$\mathsf{A}^\dagger = \sum_{i=1}^{r} \frac{\boldsymbol{v}_i \boldsymbol{u}_i^T}{\sigma_i}, \qquad r = \operatorname{rank}(\mathsf{A}).$$

We now show that the least squares solution of

$$A\boldsymbol{x} = \boldsymbol{b}$$

is given by

$$\boldsymbol{x} = A^{\dagger}\boldsymbol{b}.$$

Start with normal equations and use

$$\mathsf{A} = \mathsf{U}\begin{pmatrix} \mathsf{D} & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix}\mathsf{V}^T = \tilde{\mathsf{U}}\mathsf{D}\tilde{\mathsf{V}}^T,$$

the reduced SVD of A, i.e., $\tilde{\mathsf{U}} \in \mathbb{R}^{m\times r}, \tilde{\mathsf{V}} \in \mathbb{R}^{n\times r}$.

$$\mathsf{A}^T\mathsf{A}\boldsymbol{x} = \mathsf{A}^T\boldsymbol{b} \quad \Longleftrightarrow \quad \tilde{\mathsf{V}}\mathsf{D}\underbrace{\tilde{\mathsf{U}}^T\tilde{\mathsf{U}}}_{=\mathsf{I}}\mathsf{D}\tilde{\mathsf{V}}^T\boldsymbol{x} = \tilde{\mathsf{V}}\mathsf{D}\tilde{\mathsf{U}}^T\boldsymbol{b}$$

$$\Longleftrightarrow \quad \tilde{\mathsf{V}}\mathsf{D}^2\tilde{\mathsf{V}}^T\boldsymbol{x} = \tilde{\mathsf{V}}\mathsf{D}\tilde{\mathsf{U}}^T\boldsymbol{b}$$

Multiplication by $\mathsf{D}^{-1}\tilde{\mathsf{V}}^T$ yields

$$\mathsf{D}\tilde{\mathsf{V}}^T\boldsymbol{x} = \tilde{\mathsf{U}}^T\boldsymbol{b}.$$

Thus

$$\mathsf{D}\widetilde{\mathsf{V}}^T \boldsymbol{x} = \widetilde{\mathsf{U}}^T \boldsymbol{b}$$

implies

$$\begin{aligned} & \boldsymbol{x} = \widetilde{\mathsf{V}}\mathsf{D}^{-1}\widetilde{\mathsf{U}}^T \boldsymbol{b} \\ \iff \quad & \boldsymbol{x} = \mathsf{V}\begin{pmatrix} \mathsf{D}^{-1} & \mathsf{O} \\ \mathsf{O} & \mathsf{O} \end{pmatrix}\mathsf{U}^T \boldsymbol{b} \\ \iff \quad & \boldsymbol{x} = \mathsf{A}^\dagger \boldsymbol{b}. \end{aligned}$$

Remark

- *If* A *is nonsingular then* $\mathsf{A}^\dagger = \mathsf{A}^{-1}$ *(see HW).*
- *If* $\operatorname{rank}(\mathsf{A}) < n$ *(i.e., the least squares solution is not unique), then* $\boldsymbol{x} = \mathsf{A}^\dagger \boldsymbol{b}$ *provides the unique solution with minimum 2-norm (see justification on following slide).*

Minimum norm solution of underdetermined systems

Note that the general solution of $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$ is given by

$$\boldsymbol{z} = \mathsf{A}^{\dagger}\boldsymbol{b} + \boldsymbol{n}, \qquad \boldsymbol{n} \in N(\mathsf{A}).$$

Then

$$\begin{aligned}\|\boldsymbol{z}\|_2^2 &= \|\mathsf{A}^{\dagger}\boldsymbol{b} + \boldsymbol{n}\|_2^2 \\ &\overset{\text{Pythag. thm}}{=} \|\mathsf{A}^{\dagger}\boldsymbol{b}\|_2^2 + \|\boldsymbol{n}\|_2^2 \geq \|\mathsf{A}^{\dagger}\boldsymbol{b}\|_2^2.\end{aligned}$$

The Pythagorean theorem applies since (see HW)

$$\mathsf{A}^{\dagger}\boldsymbol{b} \in R(\mathsf{A}^{\dagger}) = R(\mathsf{A}^T)$$

so that, using $R(\mathsf{A}^T) = N(\mathsf{A})^{\perp}$,

$$\mathsf{A}^{\dagger}\boldsymbol{b} \perp \boldsymbol{n}.$$

Remark

Explicit use of the pseudoinverse is usually not recommended.
Instead we solve $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$, $\mathsf{A} \in \mathbb{R}^{m \times n}$, *by*

1. $\mathsf{A} = \widetilde{\mathsf{U}}\mathsf{D}\widetilde{\mathsf{V}}^T$ *(reduced SVD)*
2. $\mathsf{A}\boldsymbol{x} = \boldsymbol{b} \quad \Longleftrightarrow \quad \mathsf{D}\widetilde{\mathsf{V}}^T\boldsymbol{x} = \widetilde{\mathsf{U}}^T\boldsymbol{b}$, *so*
   1. *Solve* $\mathsf{D}\boldsymbol{y} = \widetilde{\mathsf{U}}^T\boldsymbol{b}$ *for* $\boldsymbol{y}$
   2. *Compute* $\boldsymbol{x} = \widetilde{\mathsf{V}}\boldsymbol{y}$

Other Applications

Also known as principal component analysis (PCA), (discrete) Karhunen-Loève (KL) transformation, Hotelling transform, or proper orthogonal decomposition (POD)

- Data compression
- Noise filtering
- Regularization of inverse problems
  - Tomography
  - Image deblurring
  - Seismology
- Information retrieval and data mining (latent semantic analysis)
- Bioinformatics and computational biology
  - Immunology
  - Molecular dynamics
  - Microarray data analysis

Orthogonal Projections

Earlier we discussed orthogonal complementary subspaces of an inner product space $\mathcal{V}$, i.e.,

$$\mathcal{V} = \mathcal{M} \oplus \mathcal{M}^{\perp}.$$

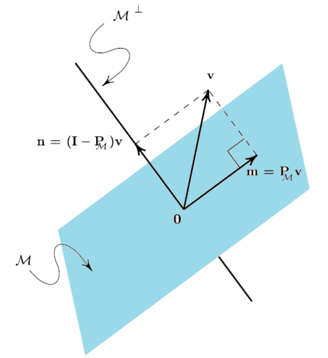
Definition

Consider $\mathcal{V} = \mathcal{M} \oplus \mathcal{M}^{\perp}$ so that for every $\boldsymbol{v} \in \mathcal{V}$ there exist unique vectors $\boldsymbol{m} \in \mathcal{M}$, $\boldsymbol{n} \in \mathcal{M}^{\perp}$ such that

$$\boldsymbol{v} = \boldsymbol{m} + \boldsymbol{n}.$$

Then $\boldsymbol{m}$ is called the orthogonal projection of $\boldsymbol{v}$ onto $\mathcal{M}$.

The matrix $P_{\mathcal{M}}$ such that $P_{\mathcal{M}}\boldsymbol{v} = \boldsymbol{m}$ is the orthogonal projector onto $\mathcal{M}$ along $\mathcal{M}^{\perp}$.



For arbitrary complementary subspaces $\mathcal{X}, \mathcal{Y}$ we showed earlier that the projector onto $\mathcal{X}$ along $\mathcal{Y}$ is given by

$$\begin{aligned} P &= \begin{pmatrix} X & O \end{pmatrix} \begin{pmatrix} X & Y \end{pmatrix}^{-1} \\ &= \begin{pmatrix} X & Y \end{pmatrix} \begin{pmatrix} I & O \\ O & O \end{pmatrix} \begin{pmatrix} X & Y \end{pmatrix}^{-1}, \end{aligned}$$

where the columns of X and Y are bases for $\mathcal{X}$ and $\mathcal{Y}$.

Now we let $\mathcal{X} = \mathcal{M}$ and $\mathcal{Y} = \mathcal{M}^\perp$ be orthogonal complementary subspaces, where M and N contain the basis vectors of $\mathcal{M}$ and $\mathcal{M}^\perp$ in their columns.
Then

$$\mathsf{P} = \begin{pmatrix} \mathsf{M} & \mathsf{O} \end{pmatrix} \begin{pmatrix} \mathsf{M} & \mathsf{N} \end{pmatrix}^{-1}. \tag{18}$$

To find $\begin{pmatrix} \mathsf{M} & \mathsf{N} \end{pmatrix}^{-1}$ we note that

$$\mathsf{M}^T \mathsf{N} = \mathsf{N}^T \mathsf{M} = \mathsf{O}$$

and if N is an orthogonal matrix (i.e., contains an ON basis), then

$$\begin{pmatrix} (\mathsf{M}^T\mathsf{M})^{-1}\mathsf{M}^T \\ \mathsf{N}^T \end{pmatrix} \begin{pmatrix} \mathsf{M} & \mathsf{N} \end{pmatrix} = \begin{pmatrix} \mathsf{I} & \mathsf{O} \\ \mathsf{O} & \mathsf{I} \end{pmatrix}$$

(note that $\mathsf{M}^T\mathsf{M}$ is invertible since M is full rank because its columns form a basis of $\mathcal{M}$).

Thus

$$\begin{pmatrix} \mathsf{M} & \mathsf{N} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathsf{M}^T\mathsf{M})^{-1}\mathsf{M}^T \\ \mathsf{N}^T \end{pmatrix}. \tag{19}$$

Inserting (19) into (18) yields

$$\begin{aligned} \mathsf{P}_{\mathcal{M}} &= \begin{pmatrix} \mathsf{M} & \mathsf{O} \end{pmatrix} \begin{pmatrix} (\mathsf{M}^T\mathsf{M})^{-1}\mathsf{M}^T \\ \mathsf{N}^T \end{pmatrix} \\ &= \mathsf{M}(\mathsf{M}^T\mathsf{M})^{-1}\mathsf{M}^T. \end{aligned}$$

Remark

Note that $\mathsf{P}_{\mathcal{M}}$ is unique so that this formula holds for an arbitrary basis of $\mathcal{M}$ (collected in M).
In particular, if M contains an ON basis for $\mathcal{M}$, then

$$\mathsf{P}_{\mathcal{M}} = \mathsf{M}\mathsf{M}^T.$$

Similarly,

$$\mathsf{P}_{\mathcal{M}^\perp} = \mathsf{N}(\mathsf{N}^T\mathsf{N})^{-1}\mathsf{N}^T \quad \text{(arbitrary basis for } \mathcal{N}\text{)}$$
$$\mathsf{P}_{\mathcal{M}^\perp} = \mathsf{N}\mathsf{N}^T \quad \text{ON basis}$$

As before,

$$\mathsf{P}_{\mathcal{M}} = \mathsf{I} - \mathsf{P}_{\mathcal{M}^\perp}.$$

Example

If $\mathcal{M} = \text{span}\{\boldsymbol{u}\}$, $\|\boldsymbol{u}\| = 1$ then

$$\mathsf{P}_{\mathcal{M}} = \mathsf{P}_{\boldsymbol{u}} = \boldsymbol{u}\boldsymbol{u}^T$$

and

$$\mathsf{P}_{\boldsymbol{u}^T} = \mathsf{I} - \mathsf{P}_{\boldsymbol{u}} = \mathsf{I} - \boldsymbol{u}\boldsymbol{u}^T$$

(cf. elementary orthogonal projectors earlier).

Properties of orthogonal projectors

Theorem

Let $\mathsf{P} \in \mathbb{R}^{n \times n}$ *be a projector, i.e.,* $\mathsf{P}^2 = \mathsf{P}$*. Then the matrix* P *is an orthogonal projector if*

1. $R(\mathsf{P}) \perp N(\mathsf{P})$,
2. $\mathsf{P}^T = \mathsf{P}$,
3. $\|\mathsf{P}\|_2 = 1$.

Proof

1. Follows directly from the definition.
2. "$\Longrightarrow$": Assume P is an orthogonal projector, i.e.,

$$\mathsf{P} = \mathsf{M}(\mathsf{M}^T\mathsf{M})^{-1}\mathsf{M}^T \quad \text{and} \quad \mathsf{P}^T = \mathsf{M}\underbrace{(\mathsf{M}^T\mathsf{M})^{-T}}_{=(\mathsf{M}^T\mathsf{M})^{-1}}\mathsf{M}^T = \mathsf{P}.$$

   "$\Longleftarrow$": Assume $\mathsf{P} = \mathsf{P}^T$. Then

$$R(\mathsf{P}) = R(\mathsf{P}^T) \overset{\text{Orth.decomp.}}{=} N(\mathsf{P})^\perp$$

   so that P is an orthogonal projector via (1).

Proof (cont.)

3. For complementary subspaces $\mathcal{X}, \mathcal{Y}$ we know the angle between $\mathcal{X}$ and $\mathcal{Y}$ is given by

$$\|\mathsf{P}\|_2 = \frac{1}{\sin\theta}, \quad \theta \in \left[0, \frac{\pi}{2}\right].$$

Assume P is an orthogonal projector, then $\theta = \frac{\pi}{2}$ so that $\|\mathsf{P}\|_2 = 1$.

Conversely, if $\|\mathsf{P}\|_2 = 1$, then $\theta = \frac{\pi}{2}$ and $\mathcal{X}, \mathcal{Y}$ are orthogonal complements, i.e., P is an orthogonal projector.

□

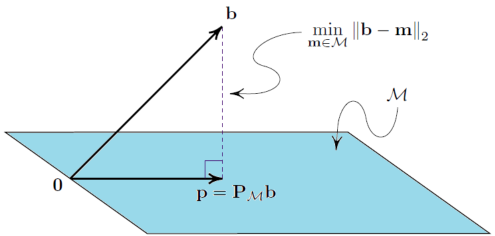
Why is orthogonal projection so important?

Theorem

Let $\mathcal{V}$ be an inner product space with subspace $\mathcal{M}$, and let $\boldsymbol{b} \in \mathcal{V}$. Then

$$dist(\boldsymbol{b}, \mathcal{M}) = \min_{\boldsymbol{n} \in \mathcal{M}} \|\boldsymbol{b} - \boldsymbol{m}\|_2 = \|\boldsymbol{b} - \mathsf{P}_{\mathcal{M}}\boldsymbol{b}\|_2,$$

i.e., $\mathsf{P}_{\mathcal{M}}\boldsymbol{b}$ is the unique vector in $\mathcal{M}$ closest to $\boldsymbol{b}$. The quantity $dist(\boldsymbol{b}, \mathcal{M})$ is called the (orthogonal) distance from $\boldsymbol{b}$ to $\mathcal{M}$.



Proof

Let $\boldsymbol{p} = \mathsf{P}_{\mathcal{M}}\boldsymbol{b}$. Then $\boldsymbol{p} \in \mathcal{M}$ and $\boldsymbol{p} - \boldsymbol{m} \in \mathcal{M}$ for every $\boldsymbol{m} \in \mathcal{M}$.

Moreover,

$$\boldsymbol{b} - \boldsymbol{p} = (\mathsf{I} - \mathsf{P}_{\mathcal{M}})\boldsymbol{b} \in \mathcal{M}^{\perp},$$

so that

$$(\boldsymbol{p} - \boldsymbol{m}) \perp (\boldsymbol{b} - \boldsymbol{p}).$$

Then

$$\begin{aligned} \|\boldsymbol{b} - \boldsymbol{m}\|_2^2 &= \|\boldsymbol{b} - \boldsymbol{p} + \boldsymbol{p} - \boldsymbol{m}\|_2^2 \\ &\stackrel{\text{Pythag.}}{=} \|\boldsymbol{b} - \boldsymbol{p}\|_2^2 + \|\boldsymbol{p} - \boldsymbol{m}\|_2^2 \\ &\geq \|\boldsymbol{b} - \boldsymbol{p}\|_2^2. \end{aligned}$$

Therefore $\min_{\boldsymbol{m} \in \mathcal{M}} \|\boldsymbol{b} - \boldsymbol{m}\|_2 = \|\boldsymbol{b} - \boldsymbol{p}\|_2$.

Proof (cont.)

Uniqueness: Assume there exists a $\boldsymbol{q} \in \mathcal{M}$ such that

$$\|\boldsymbol{b} - \boldsymbol{q}\|_2 = \|\boldsymbol{b} - \boldsymbol{p}\|_2. \tag{20}$$

Then

$$\|\boldsymbol{b} - \boldsymbol{q}\|_2^2 = \|\underbrace{\boldsymbol{b} - \boldsymbol{p}}_{\in \mathcal{M}^\perp} + \underbrace{\boldsymbol{p} - \boldsymbol{q}}_{\in \mathcal{M}}\|_2^2$$

$$\overset{\text{Pythag.}}{=} \|\boldsymbol{b} - \boldsymbol{p}\|_2^2 + \|\boldsymbol{p} - \boldsymbol{q}\|_2^2.$$

But then (20) implies that $\|\boldsymbol{p} - \boldsymbol{q}\|_2^2 = 0$ and therefore $\boldsymbol{p} = \boldsymbol{q}$. $\square$

Least squares approximation revisited

Now we give a "modern" derivation of the normal equations (without calculus), and note that much of this remains true for best L_2 approximation.

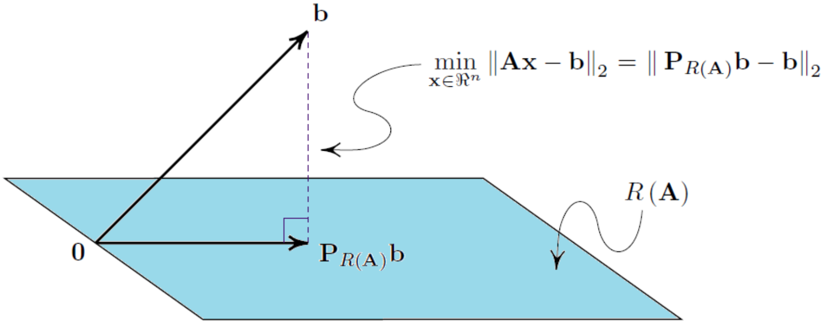
Goal of least squares: For $\mathsf{A} \in \mathbb{R}^{m\times n}$, find

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \sqrt{\sum_{i=1}^{m} \left((\mathsf{A}\boldsymbol{x})_i - b_i\right)^2} \quad \Longleftrightarrow \quad \min_{\boldsymbol{x}\in\mathbb{R}^n} \|\mathsf{A}\boldsymbol{x} - \boldsymbol{b}\|_2.$$

Now $\mathsf{A}\boldsymbol{x} \in R(\mathsf{A})$, so the least squares error is

$$\begin{aligned}\operatorname{dist}(\boldsymbol{b}, R(\mathsf{A})) &= \min_{\mathsf{A}\boldsymbol{x}\in R(\mathsf{A})} \|\boldsymbol{b} - \mathsf{A}\boldsymbol{x}\|_2 \\ &= \|\boldsymbol{b} - \mathsf{P}_{R(\mathsf{A})}\boldsymbol{b}\|_2\end{aligned}$$

with $\mathsf{P}_{R(\mathsf{A})}$ the orthogonal projector onto $R(\mathsf{A})$.


b
$\min_{\mathbf{x}\in\Re^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 = \|\mathbf{P}_{R(\mathbf{A})}\mathbf{b} - \mathbf{b}\|_2$
$R(\mathbf{A})$
0
$\mathbf{P}_{R(\mathbf{A})}\mathbf{b}$

Moreover, the least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ is given by that $\boldsymbol{x}$ for which

$$A\boldsymbol{x} = P_{R(A)}\boldsymbol{b}.$$

The following argument shows that this is equivalent to the normal equations:

$$\begin{aligned}
& A\boldsymbol{x} = P_{R(A)}\boldsymbol{b} \\
\iff\ & P_{R(A)}A\boldsymbol{x} = P^2_{R(A)}\boldsymbol{b} = P_{R(A)}\boldsymbol{b} \\
\iff\ & P_{R(A)}(A\boldsymbol{x} - \boldsymbol{b}) = \mathbf{0} \\
\iff\ & A\boldsymbol{x} - \boldsymbol{b} \in N(P_{R(A)}) = R(A)^{\perp} \quad (\text{P orth. proj. onto } R(A)) \\
\overset{\text{Orth.decomp.}}{\iff}\ & A\boldsymbol{x} - \boldsymbol{b} \in N(A^T) \\
\iff\ & A^T(A\boldsymbol{x} - \boldsymbol{b}) = \mathbf{0} \\
\iff\ & A^T A\boldsymbol{x} = A^T\boldsymbol{b}.
\end{aligned}$$

Remark

No we are no longer limited to the real case.

References I

[Mey00] Carl D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, PA, 2000.