Metodi Numerici dell'Informatica

Orthogonality and QR decomposition

Emanuele Rodolà
rodola@di.uniroma1.it

SAPIENZA
UNIVERSITÀ DI ROMA

2nd semester a.y. 2022/2023 · March 28, 2023

Motivation

Column space

For any $\mathbf{A} \in \mathbb{R}^{n \times m}$, the linear problem:

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

can be seen as:

"write $\mathbf{b}$ as a linear combination of the columns of $\mathbf{A}$ with coefficients stored in $\mathbf{x}$".

Column space

For any $\mathbf{A} \in \mathbb{R}^{n \times m}$, the linear problem:

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

can be seen as:

"write $\mathbf{b}$ as a linear combination of the columns of $\mathbf{A}$ with coefficients stored in $\mathbf{x}$".

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = x_1 \begin{pmatrix} | \\ \mathbf{a}_1 \\ | \end{pmatrix} + x_2 \begin{pmatrix} | \\ \mathbf{a}_2 \\ | \end{pmatrix} + \cdots$$

Numerical instability

Consider the following situation:

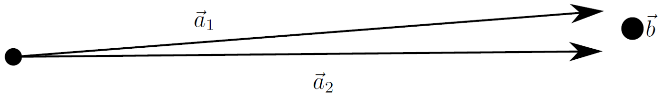
$$\mathbf{A} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{pmatrix} = \begin{pmatrix} 0 & 0.0001 & \\ \vdots & \vdots & \dots \\ 1 & 1.0001 & \end{pmatrix}$$

Numerical instability

Consider the following situation:

$$\mathbf{A} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{pmatrix} = \begin{pmatrix} 0 & 0.0001 & \\ \vdots & \vdots & \dots \\ 1 & 1.0001 & \end{pmatrix}$$

The columns are almost linearly dependent:



This problem is poorly conditioned, since we can write either:

$$\mathbf{b} \approx x_1\mathbf{a}_1 + x_2\mathbf{a}_2$$

$$\mathbf{b} \approx x_2\mathbf{a}_1 + x_1\mathbf{a}_2$$

Column space

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = x_1 \begin{pmatrix} | \\ \mathbf{a}_1 \\ | \end{pmatrix} + x_2 \begin{pmatrix} | \\ \mathbf{a}_2 \\ | \end{pmatrix} + \cdots$$

- A solution might not exist
- The solution might be unique

Column space

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = x_1 \begin{pmatrix} | \\ \mathbf{a}_1 \\ | \end{pmatrix} + x_2 \begin{pmatrix} | \\ \mathbf{a}_2 \\ | \end{pmatrix} + \cdots$$

- A solution might not exist
- The solution might be unique

It ultimately depends on the structure of the column space of $\mathbf{A}$, which we define as $\text{col } \mathbf{A} = \text{span}(\mathbf{a}_1, \mathbf{a}_2, \dots)$.

Column space

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = x_1 \begin{pmatrix} | \\ \mathbf{a}_1 \\ | \end{pmatrix} + x_2 \begin{pmatrix} | \\ \mathbf{a}_2 \\ | \end{pmatrix} + \cdots$$

- A solution might not exist
- The solution might be unique

It ultimately depends on the structure of the column space of $\mathbf{A}$, which we define as $\mathrm{col}\ \mathbf{A} = \mathrm{span}(\mathbf{a}_1, \mathbf{a}_2, \dots)$.

Can we find a simpler representation for the column space?

Column space

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = x_1 \begin{pmatrix} | \\ \mathbf{a}_1 \\ | \end{pmatrix} + x_2 \begin{pmatrix} | \\ \mathbf{a}_2 \\ | \end{pmatrix} + \cdots$$

- A solution might not exist
- The solution might be unique

It ultimately depends on the structure of the column space of $\mathbf{A}$, which we define as $\mathrm{col}\ \mathbf{A} = \mathrm{span}(\mathbf{a}_1, \mathbf{a}_2, \dots)$.

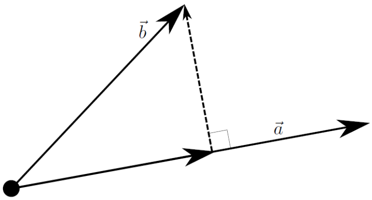
Can we find a simpler representation for the column space?

Can we deal with numerical instability explicitly?

Orthogonality

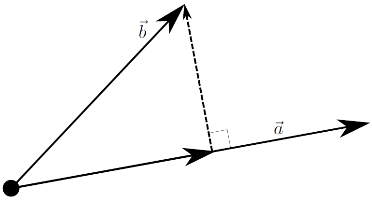
Projections

Consider two vectors $\mathbf{a}$ and $\mathbf{b}$. Which multiple of $\mathbf{a}$ is closest to $\mathbf{b}$?



Projections

Consider two vectors $\mathbf{a}$ and $\mathbf{b}$. Which multiple of $\mathbf{a}$ is closest to $\mathbf{b}$?

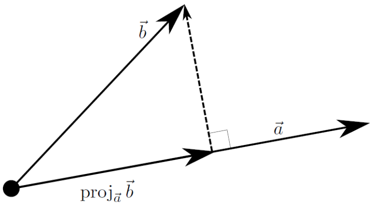


This can be phrased as a projection problem:

$$\min_{c} \|c\,\mathbf{a} - \mathbf{b}\|_2^2$$

Projections

Consider two vectors $\mathbf{a}$ and $\mathbf{b}$. Which multiple of $\mathbf{a}$ is closest to $\mathbf{b}$?



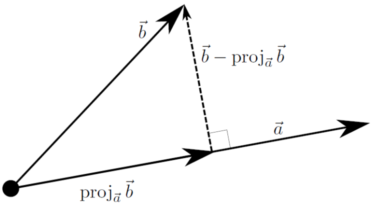
This can be phrased as a projection problem:

$$\min_c \|c\,\mathbf{a} - \mathbf{b}\|_2^2$$

The optimal c^* gives us the projection $\mathrm{proj}_{\mathbf{a}}\mathbf{b} \equiv c^*\mathbf{a}$.

Projections

Consider two vectors $\mathbf{a}$ and $\mathbf{b}$. Which multiple of $\mathbf{a}$ is closest to $\mathbf{b}$?



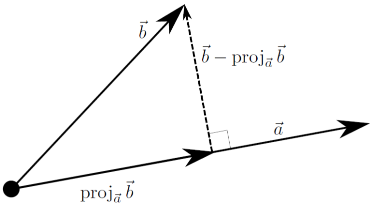
This can be phrased as a projection problem:

$$\min_c \|c\,\mathbf{a} - \mathbf{b}\|_2^2$$

The optimal c^* gives us the projection $\mathrm{proj}_{\mathbf{a}}\mathbf{b} \equiv c^*\mathbf{a}$.

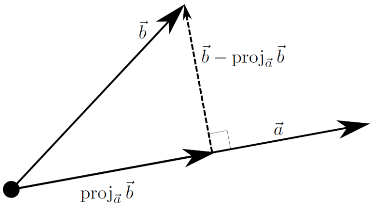
Further, the complement $\mathbf{b} - \mathrm{proj}_{\mathbf{a}}\mathbf{b}$ is orthogonal to $\mathrm{proj}_{\mathbf{a}}\mathbf{b}$.

Projections



The projection problem is a least-squares problem where $\mathbf{a}\,c \approx \mathbf{b}$.

Projections

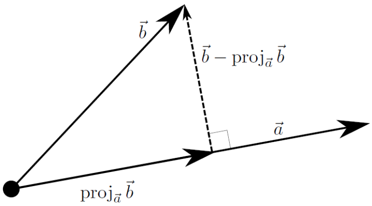


The projection problem is a least-squares problem where $\mathbf{a}\,c \approx \mathbf{b}$.

We have the normal equations $\mathbf{a}^\top \mathbf{a}\,c = \mathbf{a}^\top \mathbf{b}$, hence

$$c = \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}}$$

Projections

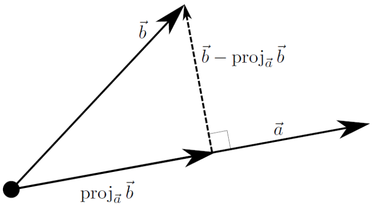


The projection problem is a least-squares problem where $\mathbf{a}\,c \approx \mathbf{b}$.

We have the normal equations $\mathbf{a}^\top \mathbf{a}\,c = \mathbf{a}^\top \mathbf{b}$, hence

$$c = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|_2^2}$$

Projections



The projection problem is a least-squares problem where $\mathbf{a}\,c \approx \mathbf{b}$.

We have the normal equations $\mathbf{a}^\top \mathbf{a}\,c = \mathbf{a}^\top \mathbf{b}$, hence

$$c = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|_2^2}$$

Therefore:

$$\text{proj}_{\mathbf{a}}\mathbf{b} = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|_2^2}\mathbf{a}$$

Projections

Similarly, we can project $\mathbf{b}$ onto an entire space.

Projections

Similarly, we can project $\mathbf{b}$ onto an entire space.

Assume we are given an orthonormal basis $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$; then, we can project onto its span by solving the problem:

$$\min_{c_1,\ldots,c_k} \|c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k - \mathbf{b}\|_2^2$$

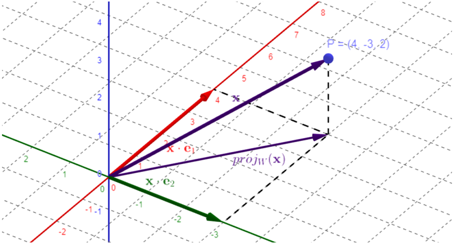
Projections

Similarly, we can project $\mathbf{b}$ onto an entire space.

Assume we are given an orthonormal basis $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$; then, we can project onto its span by solving the problem:

$$\min_{c_1,\ldots,c_k} \|c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k - \mathbf{b}\|_2^2$$

Example in $\mathbb{R}^3$ and $k = 2$:



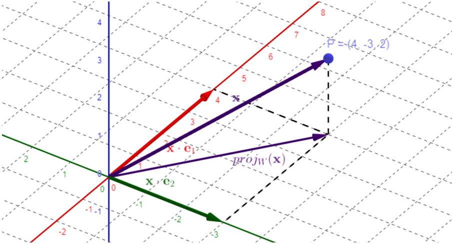
Projections

Similarly, we can project $\mathbf{b}$ onto an entire space.

Assume we are given an orthonormal basis $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$; then, we can project onto its span by solving the problem:

$$\min_{c_1,\dots,c_k} \|c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k - \mathbf{b}\|_2^2$$

Example in $\mathbb{R}^3$ and $k = 2$:



The projection is: $\mathrm{proj}_{\mathrm{span}(\mathbf{a}_1,\dots,\mathbf{a}_k)}\mathbf{b} = (\mathbf{a}_1^\top\mathbf{b})\mathbf{a}_1 + \cdots + (\mathbf{a}_k^\top\mathbf{b})\mathbf{a}_k$

Orthogonal matrices

Consider the trivial linear problem:

$$\mathbf{I}\mathbf{x} = \mathbf{b}$$

where $\mathbf{I}$ is the identity matrix. The obvious solution is $\mathbf{x} = \mathbf{b}$.

Orthogonal matrices

Consider the trivial linear problem:

$$\mathbf{I}\mathbf{x} = \mathbf{b}$$

where $\mathbf{I}$ is the identity matrix. The obvious solution is $\mathbf{x} = \mathbf{b}$.

From the past lectures, we also know that the least-squares problem

$$\mathbf{A}\mathbf{x} \approx \mathbf{b}$$

corresponds to a linear problem

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$$

Orthogonal matrices

Consider the trivial linear problem:

$$\mathbf{I}\mathbf{x} = \mathbf{b}$$

where $\mathbf{I}$ is the identity matrix. The obvious solution is $\mathbf{x} = \mathbf{b}$.

From the past lectures, we also know that the least-squares problem

$$\mathbf{A}\mathbf{x} \approx \mathbf{b}$$

corresponds to a linear problem

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$$

In the lucky case where $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$ (i.e. matrix $\mathbf{A}$ is orthogonal), we boil down to the trivial case. Further, this implies that $\mathbf{A}^{-1} = \mathbf{A}^\top$.

Let's analyze this more carefully.

Orthogonal matrices

For clarity, let us call $\mathbf{Q}$ the orthogonal matrices.

The product $\mathbf{Q}^\top\mathbf{Q}$ has the structure:

$$Q^\top Q = \begin{pmatrix} - & \vec{q}_1^\top & - \\ - & \vec{q}_2^\top & - \\ & \vdots & \\ - & \vec{q}_n^\top & \end{pmatrix} \begin{pmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \vec{q}_1 \cdot \vec{q}_1 & \vec{q}_1 \cdot \vec{q}_2 & \cdots & \vec{q}_1 \cdot \vec{q}_n \\ \vec{q}_2 \cdot \vec{q}_1 & \vec{q}_2 \cdot \vec{q}_2 & \cdots & \vec{q}_2 \cdot \vec{q}_n \\ \vdots & \vdots & \cdots & \vdots \\ \vec{q}_n \cdot \vec{q}_1 & \vec{q}_n \cdot \vec{q}_2 & \cdots & \vec{q}_n \cdot \vec{q}_n \end{pmatrix}$$

Orthogonal matrices

For clarity, let us call $\mathbf{Q}$ the orthogonal matrices.

The product $\mathbf{Q}^\top\mathbf{Q}$ has the structure:

$$Q^\top Q = \begin{pmatrix} - & \vec{q}_1^\top & - \\ - & \vec{q}_2^\top & - \\ & \vdots & \\ - & \vec{q}_n^\top & \end{pmatrix} \begin{pmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \vec{q}_1 \cdot \vec{q}_1 & \vec{q}_1 \cdot \vec{q}_2 & \cdots & \vec{q}_1 \cdot \vec{q}_n \\ \vec{q}_2 \cdot \vec{q}_1 & \vec{q}_2 \cdot \vec{q}_2 & \cdots & \vec{q}_2 \cdot \vec{q}_n \\ \vdots & \vdots & \cdots & \vdots \\ \vec{q}_n \cdot \vec{q}_1 & \vec{q}_n \cdot \vec{q}_2 & \cdots & \vec{q}_n \cdot \vec{q}_n \end{pmatrix}$$

Then, $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ means that the columns of $\mathbf{Q}$ have unit norm and they are orthogonal to one another.

Orthogonal matrices

For clarity, let us call $\mathbf{Q}$ the orthogonal matrices.

The product $\mathbf{Q}^\top\mathbf{Q}$ has the structure:

$$Q^\top Q = \begin{pmatrix} - & \vec{q}_1^\top & - \\ - & \vec{q}_2^\top & - \\ & \vdots & \\ - & \vec{q}_n^\top & \end{pmatrix} \begin{pmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \vec{q}_1 \cdot \vec{q}_1 & \vec{q}_1 \cdot \vec{q}_2 & \cdots & \vec{q}_1 \cdot \vec{q}_n \\ \vec{q}_2 \cdot \vec{q}_1 & \vec{q}_2 \cdot \vec{q}_2 & \cdots & \vec{q}_2 \cdot \vec{q}_n \\ \vdots & \vdots & \cdots & \vdots \\ \vec{q}_n \cdot \vec{q}_1 & \vec{q}_n \cdot \vec{q}_2 & \cdots & \vec{q}_n \cdot \vec{q}_n \end{pmatrix}$$

Then, $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ means that the columns of $\mathbf{Q}$ have unit norm and they are orthogonal to one another.

The columns of $\mathbf{Q}$ form an orthonormal basis for the column space.

Orthogonal matrices

For clarity, let us call $\mathbf{Q}$ the orthogonal matrices.

The product $\mathbf{Q}^\top \mathbf{Q}$ has the structure:

$$Q^\top Q = \begin{pmatrix} - & \vec{q}_1^\top & - \\ - & \vec{q}_2^\top & - \\ & \vdots & \\ - & \vec{q}_n^\top & \end{pmatrix} \begin{pmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \vec{q}_1 \cdot \vec{q}_1 & \vec{q}_1 \cdot \vec{q}_2 & \cdots & \vec{q}_1 \cdot \vec{q}_n \\ \vec{q}_2 \cdot \vec{q}_1 & \vec{q}_2 \cdot \vec{q}_2 & \cdots & \vec{q}_2 \cdot \vec{q}_n \\ \vdots & \vdots & \cdots & \vdots \\ \vec{q}_n \cdot \vec{q}_1 & \vec{q}_n \cdot \vec{q}_2 & \cdots & \vec{q}_n \cdot \vec{q}_n \end{pmatrix}$$

Then, $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}$ means that the columns of $\mathbf{Q}$ have unit norm and they are orthogonal to one another.

The columns of $\mathbf{Q}$ form an orthonormal basis for the column space.

Examples:

- The standard basis of indicator vectors is an orthonormal basis

Orthogonal matrices

For clarity, let us call $\mathbf{Q}$ the orthogonal matrices.

The product $\mathbf{Q}^\top\mathbf{Q}$ has the structure:

$$Q^\top Q = \begin{pmatrix} - & \vec{q}_1^\top & - \\ - & \vec{q}_2^\top & - \\ & \vdots & \\ - & \vec{q}_n^\top & \end{pmatrix} \begin{pmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \vec{q}_1 \cdot \vec{q}_1 & \vec{q}_1 \cdot \vec{q}_2 & \cdots & \vec{q}_1 \cdot \vec{q}_n \\ \vec{q}_2 \cdot \vec{q}_1 & \vec{q}_2 \cdot \vec{q}_2 & \cdots & \vec{q}_2 \cdot \vec{q}_n \\ \vdots & \vdots & \cdots & \vdots \\ \vec{q}_n \cdot \vec{q}_1 & \vec{q}_n \cdot \vec{q}_2 & \cdots & \vec{q}_n \cdot \vec{q}_n \end{pmatrix}$$

Then, $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ means that the columns of $\mathbf{Q}$ have unit norm and they are orthogonal to one another.

The columns of $\mathbf{Q}$ form an orthonormal basis for the column space.

Examples:

- The standard basis of indicator vectors is an orthonormal basis
- The identity matrix $\mathbf{I}$ is orthogonal

Orthogonal matrices

For clarity, let us call $\mathbf{Q}$ the orthogonal matrices.

The product $\mathbf{Q}^\top \mathbf{Q}$ has the structure:

$$Q^\top Q = \begin{pmatrix} - & \vec{q}_1^\top & - \\ - & \vec{q}_2^\top & - \\ & \vdots & \\ - & \vec{q}_n^\top & \end{pmatrix} \begin{pmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \vec{q}_1 \cdot \vec{q}_1 & \vec{q}_1 \cdot \vec{q}_2 & \cdots & \vec{q}_1 \cdot \vec{q}_n \\ \vec{q}_2 \cdot \vec{q}_1 & \vec{q}_2 \cdot \vec{q}_2 & \cdots & \vec{q}_2 \cdot \vec{q}_n \\ \vdots & \vdots & \cdots & \vdots \\ \vec{q}_n \cdot \vec{q}_1 & \vec{q}_n \cdot \vec{q}_2 & \cdots & \vec{q}_n \cdot \vec{q}_n \end{pmatrix}$$

Then, $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}$ means that the columns of $\mathbf{Q}$ have unit norm and they are orthogonal to one another.

The columns of $\mathbf{Q}$ form an orthonormal basis for the column space.

Examples:

- The standard basis of indicator vectors is an orthonormal basis
- The identity matrix $\mathbf{I}$ is orthogonal
- Any permutation matrix is orthogonal

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Q}\mathbf{x}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Q}\mathbf{x}$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Q}\mathbf{x}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Q}\mathbf{x} = \mathbf{x}^\top \mathbf{I}\mathbf{x}$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Q}\mathbf{x}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Q}\mathbf{x} = \mathbf{x}^\top \mathbf{I}\mathbf{x} = \mathbf{x}^\top \mathbf{x}$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2$$

Orthogonal matrices also preserve angles (i.e. inner products):

$$\langle \mathbf{Qx}, \mathbf{Qy} \rangle = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qy} = \mathbf{x}^\top \mathbf{Iy} = \mathbf{x}^\top \mathbf{y}$$

Isometries

Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2$$

Orthogonal matrices also preserve angles (i.e. inner products):

$$\langle \mathbf{Qx}, \mathbf{Qy} \rangle = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qy} = \mathbf{x}^\top \mathbf{Iy} = \mathbf{x}^\top \mathbf{y}$$

By these properties, the map $\mathbf{x} \mapsto \mathbf{Qx}$ is an isometry of $\mathbb{R}^n$.

Isometries

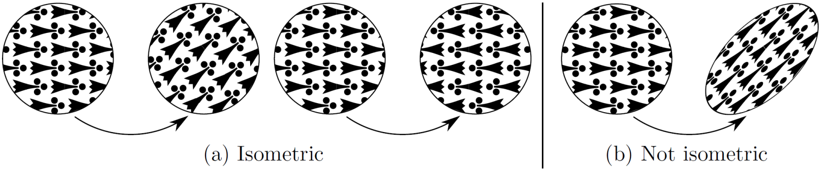
Orthogonal matrices preserve lengths:

$$\|\mathbf{Qx}\|_2^2 = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx} = \mathbf{x}^\top \mathbf{Ix} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2$$

Orthogonal matrices also preserve angles (i.e. inner products):

$$\langle \mathbf{Qx}, \mathbf{Qy}\rangle = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qy} = \mathbf{x}^\top \mathbf{Iy} = \mathbf{x}^\top \mathbf{y}$$

By these properties, the map $\mathbf{x} \mapsto \mathbf{Qx}$ is an isometry of $\mathbb{R}^n$.



(a) Isometric (b) Not isometric

Non-orthogonal matrices

When solving $\mathbf{A}\mathbf{x} = \mathbf{b}$, most of the times $\mathbf{A}$ is not orthogonal.

However, orthogonal matrices can still help us to simplify the problem.

Non-orthogonal matrices

When solving $\mathbf{A}\mathbf{x} = \mathbf{b}$, most of the times $\mathbf{A}$ is not orthogonal.

However, orthogonal matrices can still help us to simplify the problem.

General idea Rewrite $\mathbf{A}$ as a product of two matrices, so that $\mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{A}^\top\mathbf{b}$ becomes easier to solve.

Non-orthogonal matrices

When solving $\mathbf{A}\mathbf{x} = \mathbf{b}$, most of the times $\mathbf{A}$ is not orthogonal.

However, orthogonal matrices can still help us to simplify the problem.

General idea Rewrite $\mathbf{A}$ as a product of two matrices, so that $\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$ becomes easier to solve.

In particular, suppose we can write:

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is invertible.

Non-orthogonal matrices

When solving $\mathbf{A}\mathbf{x} = \mathbf{b}$, most of the times $\mathbf{A}$ is not orthogonal.

However, orthogonal matrices can still help us to simplify the problem.

General idea Rewrite $\mathbf{A}$ as a product of two matrices, so that $\mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{A}^\top\mathbf{b}$ becomes easier to solve.

In particular, suppose we can write:

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is invertible.

Then, the operation

$$\mathbf{A}\mathbf{R}^{-1} = \mathbf{Q}$$

can be seen as transforming the columns of $\mathbf{A}$ to make them orthogonal.

QR decomposition

QR decomposition

It turns out that any matrix $\mathbf{A}$ admits the factorization:

$$\mathbf{A} = \mathbf{QR}$$

which means that we can always "orthogonalize" the columns of any $\mathbf{A}$.

QR decomposition

It turns out that any matrix $\mathbf{A}$ admits the factorization:

$$\mathbf{A} = \mathbf{QR}$$

which means that we can always "orthogonalize" the columns of any $\mathbf{A}$.

Further, it can be shown that

$$\text{col } \mathbf{AR}^{-1} = \text{col } \mathbf{A}$$

QR decomposition

It turns out that any matrix $\mathbf{A}$ admits the factorization:

$$\mathbf{A} = \mathbf{QR}$$

which means that we can always "orthogonalize" the columns of any $\mathbf{A}$.

Further, it can be shown that

$$\text{col } \mathbf{AR}^{-1} = \text{col } \mathbf{A}$$

i.e., orthogonalization does not change the column space of $\mathbf{A}$.

QR decomposition

It turns out that any matrix $\mathbf{A}$ admits the factorization:

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

which means that we can always "orthogonalize" the columns of any $\mathbf{A}$.

Further, it can be shown that

$$\text{col } \mathbf{A}\mathbf{R}^{-1} = \text{col } \mathbf{A}$$

i.e., orthogonalization does not change the column space of $\mathbf{A}$.

Since $\mathbf{Q} = \mathbf{A}\mathbf{R}^{-1}$, then

$$\text{col } \mathbf{Q} = \text{col } \mathbf{A}$$

means that the columns of $\mathbf{Q}$ are an orthonormal basis for $\text{col } \mathbf{A}$.

Better conditioning

How does this help us?

Recall the "almost parallel" numerical issue we encountered before:

$$\mathbf{A} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{pmatrix} = \begin{pmatrix} 0 & 0.0001 & \\ \vdots & \vdots & \dots \\ 1 & 1.0001 & \end{pmatrix}$$

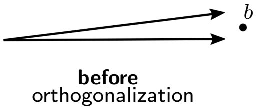
Better conditioning

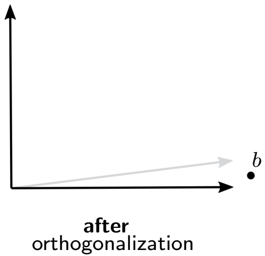
How does this help us?

Recall the "almost parallel" numerical issue we encountered before:

$$\mathbf{A} = \begin{pmatrix} | & | & \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots \\ | & | & \end{pmatrix} = \begin{pmatrix} 0 & 0.0001 & \\ \vdots & \vdots & \dots \\ 1 & 1.0001 & \end{pmatrix}$$

With orthogonalization, we make the problem better conditioned by simply applying invertible transformations to the columns of $\mathbf{A}$.





QR for least squares

Using QR factorization, the normal equations:

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$$

become:

$$(\mathbf{QR})^\top \mathbf{QR}\mathbf{x} = (\mathbf{QR})^\top \mathbf{b}$$

QR for least squares

Using QR factorization, the normal equations:

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$$

become:

$$\mathbf{R}^\top \mathbf{Q}^\top \mathbf{Q}\mathbf{R}\mathbf{x} = \mathbf{R}^\top \mathbf{Q}^\top \mathbf{b}$$

QR for least squares

Using QR factorization, the normal equations:

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$$

become:

$$\mathbf{R}^\top \mathbf{I}\mathbf{R}\mathbf{x} = \mathbf{R}^\top \mathbf{Q}^\top \mathbf{b}$$

QR for least squares

Using QR factorization, the normal equations:

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$$

become:

$$\mathbf{R}^\top \mathbf{R}\mathbf{x} = \mathbf{R}^\top \mathbf{Q}^\top \mathbf{b}$$

QR for least squares

Using QR factorization, the normal equations:

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$$

become:

$$\mathbf{R}^\top \mathbf{R}\mathbf{x} = \mathbf{R}^\top \mathbf{Q}^\top \mathbf{b}$$

And since $\mathbf{R}$ is invertible:

$$\mathbf{R}\mathbf{x} = \mathbf{Q}^\top \mathbf{b}$$

QR for least squares

Using QR factorization, the normal equations:

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$$

become:

$$\mathbf{R}^\top \mathbf{R}\mathbf{x} = \mathbf{R}^\top \mathbf{Q}^\top \mathbf{b}$$

And since $\mathbf{R}$ is invertible:

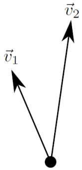
$$\mathbf{R}\mathbf{x} = \mathbf{Q}^\top \mathbf{b}$$

One additional property of matrix $\mathbf{R}$ in QR decomposition, is that it is upper triangular.

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} & \dots & u_{1,n} \\ & u_{2,2} & u_{2,3} & \dots & u_{2,n} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & u_{n-1,n} \\ 0 & & & & u_{n,n} \end{bmatrix}$$

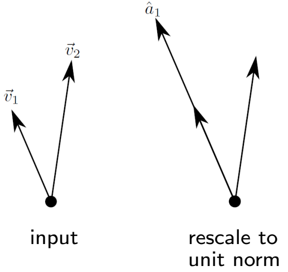
This makes the system immediate to solve by back-substitution.

Gram-Schmidt algorithm

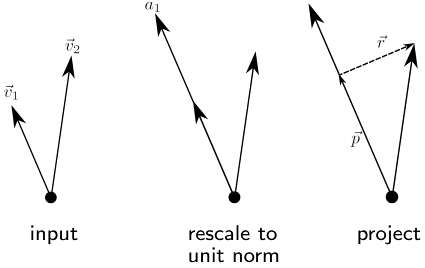


input

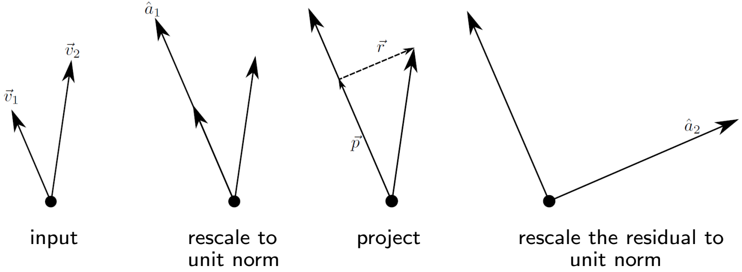
Gram-Schmidt algorithm



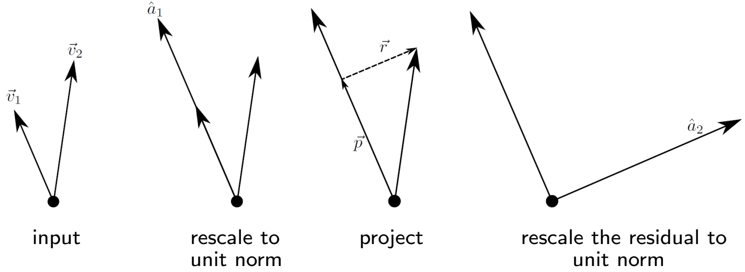
Gram-Schmidt algorithm



Gram-Schmidt algorithm



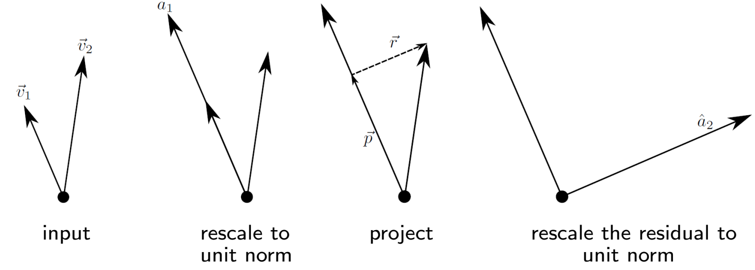
Gram-Schmidt algorithm



Observe the following:

- The orthogonalized vectors span the same space of $\mathbf{v}_1, \mathbf{v}_2$

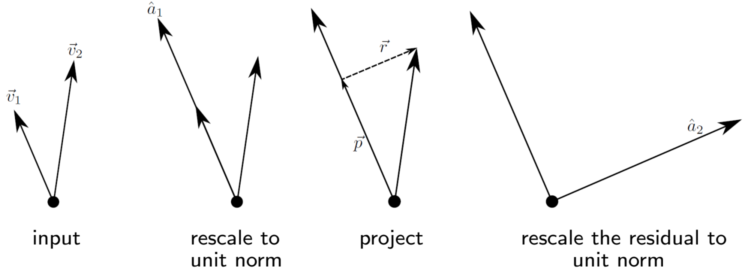
Gram-Schmidt algorithm



Observe the following:

- The orthogonalized vectors span the same space of $\mathbf{v}_1, \mathbf{v}_2$
- We can introduce more input vectors one at a time, and orthogonalize them with respect to the previous ones

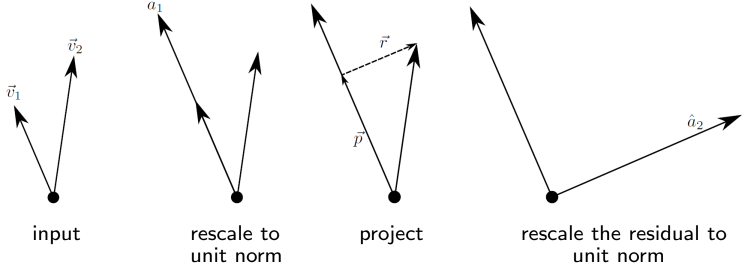
Gram-Schmidt algorithm



Observe the following:

- The orthogonalized vectors span the same space of $\mathbf{v}_1, \mathbf{v}_2$
- We can introduce more input vectors one at a time, and orthogonalize them with respect to the previous ones
- If the input vectors are columns of $\mathbf{A}$, the orthogonalized vectors are the columns of $\mathbf{Q}$. .

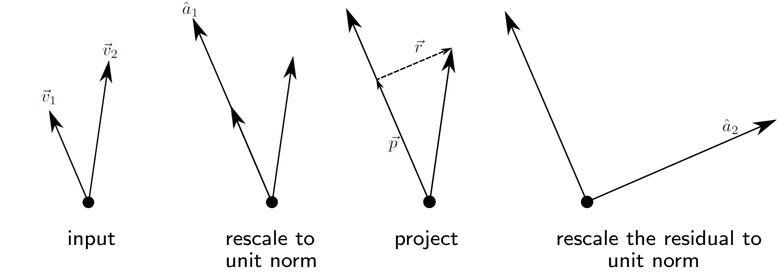
Gram-Schmidt algorithm



Observe the following:

- The orthogonalized vectors span the same space of $\mathbf{v}_1, \mathbf{v}_2$
- We can introduce more input vectors one at a time, and orthogonalize them with respect to the previous ones
- If the input vectors are columns of $\mathbf{A}$, the orthogonalized vectors are the columns of $\mathbf{Q}$. Further, we can compute $\mathbf{R} = \mathbf{Q}^\top \mathbf{A}$.

Gram-Schmidt algorithm



function GRAM-SCHMIDT$(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k)$

  ▷ Computes an orthonormal basis $\hat{a}_1, \ldots, \hat{a}_k$ for span $\{\vec{v}_1, \ldots, \vec{v}_k\}$

  ▷ Assumes $\vec{v}_1, \ldots, \vec{v}_k$ are linearly independent.

  $\hat{a}_1 \leftarrow \vec{v}_1 / \|\vec{v}_1\|_2$ ▷ Nothing to project out of the first vector

  for $i \leftarrow 2, 3, \ldots, k$

    $\vec{p} \leftarrow \vec{0}$ ▷ Projection of $\vec{v}_i$ onto span $\{\hat{a}_1, \ldots, \hat{a}_{i-1}\}$

    for $j \leftarrow 1, 2, \ldots, i-1$

      $\vec{p} \leftarrow \vec{p} + (\vec{v}_i \cdot \hat{a}_j)\hat{a}_j$ ▷ Projecting onto orthonormal basis

    $\vec{r} \leftarrow \vec{v}_i - \vec{p}$ ▷ Residual is orthogonal to current basis

    $\hat{a}_i \leftarrow \vec{r} / \|\vec{r}\|_2$ ▷ Normalize this residual and add it to the basis

  return $\{\hat{a}_1, \ldots, \hat{a}_k\}$

Reduced QR

Let us take a closer look at the dimensions:

$$\mathbf{Ax} = \mathbf{b}$$

In general, $\mathbf{A}$ will be tall with size $m \times n$.

Reduced QR

Let us take a closer look at the dimensions:

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

In general, $\mathbf{A}$ will be tall with size $m \times n$. Then it can be:

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times n}\underbrace{\mathbf{R}}_{n\times n}$$

or

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times m}\underbrace{\mathbf{R}}_{m\times n}$$

Reduced QR

Let us take a closer look at the dimensions:

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

In general, $\mathbf{A}$ will be tall with size $m \times n$. Then it can be:

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times n}\underbrace{\mathbf{R}}_{n\times n}$$

or

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times m}\underbrace{\mathbf{R}}_{m\times n}$$

- Gram-Schmidt can be modified to output either case

Reduced QR

Let us take a closer look at the dimensions:

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

In general, $\mathbf{A}$ will be tall with size $m \times n$. Then it can be:

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times n}\underbrace{\mathbf{R}}_{n\times n}$$

or

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times m}\underbrace{\mathbf{R}}_{m\times n}$$

- Gram-Schmidt can be modified to output either case
- $\mathbf{A}$ has $k \leq n$ linearly independent columns...

Reduced QR

Let us take a closer look at the dimensions:

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

In general, $\mathbf{A}$ will be tall with size $m \times n$. Then it can be:

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times n}\underbrace{\mathbf{R}}_{n\times n}$$

or

$$\underbrace{\mathbf{A}}_{m\times n} = \underbrace{\mathbf{Q}}_{m\times m}\underbrace{\mathbf{R}}_{m\times n}$$

- Gram-Schmidt can be modified to output either case
- $\mathbf{A}$ has $k \leq n$ linearly independent columns...
- ...therefore, we can discard the last $n - k$ columns of $\mathbf{Q}$

Suggested reading

Read Sections 5.1 – 5.4 of the book:

J. Solomon, "Numerical Algorithms"