Student Learning Targets, and Objectives

**SLOs: QR-Factorization Least Squares Problems**

**Target** The QR-Factorization
- **Objective** How to compute using the Gram-Schmidt Orthogonalization Method

**Target** Building Blocks
- **Objective** Projectors, Idempotent Matrices, Complementary Projectors
- **Objective** Characterization of the SVD using Orthogonal Projectors

**Projectors**
- Idempotent Matrices; Range & Nullspace; Complementary
- Orthogonal Projectors
- Orthonormal and Non-Orthonormal Basis

**The QR-Factorization**
- The Full and Reduced QR-Factorizations
- Gram-Schmidt Orthogonalization
- QR: Existence and Uniqueness

A Quick Check of the Roadmap Rear-view Mirror

So far we have reviewed (or quickly introduced) basic linear algebra concepts, e.g.
- Vector and Matrix operations, including norms.
- Matrix properties (vocabulary): rank, range, nullspace, domain, Hermitian conjugate (adjoint), unitary...

Then we introduced the idea — from a geometrical perspective — of the Singular Value Decomposition \( A = U \Sigma V^* \) of a matrix.

Finally, we connected the SVD and its properties to the majority of the concepts introduced.

In a sense, with the SVD we have extracted all information from the matrix \( A \) and we are "done."
Problem#1: We do not have a stable algorithm to compute the SVD. (We don’t even know what “stable” means!)

Problem#2: Even when we have such an algorithm (later in the semester), it will turn out to be quite computationally expensive.

The Approach: We will now start building our computational toolbox so that in the end we can implement a stable, effective algorithm for the SVD.

Along the way we will study other decompositions which may not be as complete as the SVD, but are cheaper to compute and are quite useful in certain applications.

- **Projectors**: Orthogonal and non-orthogonal projection matrices.
- **The QR-Factorization**
  - As an idea...
  - Computed using Gram-Schmidt orthogonalization
  - Computed using Householder triangularization
  - Alternative **not** discussed: Computed using Givens rotation ($\approx 50\%$ more expensive than Householder, with no additional benefit.)
  - Solving least-squares problems using the QR-factorization

**Projection, Projections, Everywhere!!!**

**Figure**: **Left**— Projecting a geometrical shape onto different planes (the figure itself is a 2D projection of this 3D-to-2D projection!); **Right**— Map projections; $S^2 \mapsto \mathbb{R}^2$, and $S^2 \mapsto \mathbb{R} \times [-\pi, \pi]$.

**Definition (Projector)**

A **projector** is a square matrix $P$ that satisfies

\[ P^2 = P. \]

Think, for instance of

\[
P = \begin{bmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & 0
\end{bmatrix},
\]

as the projection of a vector in $\mathbb{R}^3$ onto the $x$-$y$ plane in $\mathbb{R}^3$: — find a corner in the room; put a broom-stick in the corner and let it point into the room; observe the shadow on the floor. (This is making the assumption that the lighting is laser-based and arranged so that all light-rays go straight from ceiling-to-floor...).
Complementary Projectors

We notice that a projector $P$ separates $\mathbb{C}^m$ into two spaces. Conversely, if $S_1, S_2 \subseteq \mathbb{C}^m$ such that $S_1 \cap S_2 = \{0\}$, and $S_1 + S_2 = \mathbb{C}^m$, then $S_1$ and $S_2$ are complementary subspaces and there exists a projector $P$ onto $S_1$ along $S_2$ such that $\text{range}(P) = S_1$, and $\text{null}(P) = S_2$.

An orthogonal projector is a projector that projects onto a subspace $S_1$ along a space $S_2$, where $S_1$ and $S_2$ are orthogonal.

Orthogonal Projectors

An orthogonal projector is a projector that is also Hermitian, i.e.

$$P^* = P, \quad P^2 = P$$

If $P = P^*$, then the inner product of $P\tilde{x} \in S_1$ and $(I - P)\tilde{y} \in S_2$ is zero:

$$\langle P\tilde{x}, (I - P)\tilde{y} \rangle = \tilde{x}^*P^* (I - P)\tilde{y} = \tilde{x}^*(P - P^2)\tilde{y} = 0$$
Orthogonal Projectors

We now show that if $P$ projects onto $S_1$ along $S_2$ ($S_1 \perp S_2$, and $S_1$ has dimension $n$), then $P = P^*$ — the construction will give us a very simple characterization of the projector in terms of the SVD!

**Projection with an Orthonormal Basis**

Since some singular values (in $\Sigma$) are zero, we can use the reduced SVD instead, i.e. we only keep the first $n$ columns in $Q$, and we end up with

$$P = \hat{Q} \hat{Q}^*$$

where the columns of $Q \in \mathbb{C}^{m \times n}$ are orthonormal.

There is nothing magic about orthonormal vectors associated with the SVD — as long as the columns, $\vec{q}_j \in \mathbb{C}^m$, of $\hat{Q}$ are orthonormal, the matrix $P = QQ^*$ defines an orthogonal projection onto $S_1 = \text{range}(Q)$.

**Recap**

Orthogonal Projectors

With this construction we have

$$PQ = \begin{bmatrix} \vec{q}_1 & \ldots & \vec{q}_n & 0 & \ldots \end{bmatrix}$$

and, multiplying by $Q^*$ from the left:

$$Q^*PQ = \text{diag}(1, \ldots, 1, 0, \ldots) = \Sigma$$

Thus we have constructed an SVD of $P$:

$$P = Q\Sigma Q^*$$

and clearly $P$ is Hermitian

$$P^* = (Q\Sigma Q^*)^* = (Q^*)^*\Sigma^*Q^* = Q\Sigma Q^* = P. \Box$$

**Projection with an Orthonormal Basis**

The projection

$$\vec{v} \mapsto P\vec{v} \quad \text{defined by} \quad QQ^*\vec{v} = \sum_{i=1}^{n} (\vec{q}_i \vec{q}_i^*) \vec{v}$$

can be viewed as a sum on $n$ rank-one projections,

$$P_i = \vec{q}_i \vec{q}_i^*$$

where each such projection isolates the component in a single direction given by $\vec{q}_i$. These rank-one projectors will show up as building blocks in future algorithms.

For completeness, we note that the complement of a rank-one projector is a rank-$(m-1)$ projector that eliminates the component in the direction of $\vec{q}_i$

$$P_{\perp \vec{q}_i} = (I - \vec{q}_i \vec{q}_i^*)$$
The Reduced QR-Factorization

The Idea

Let \( S_1 \) be the subspace spanned by the linearly independent vectors \( \{\vec{a}_1, \ldots, \vec{a}_n\} \) and let \( A \) be the matrix with columns \( \vec{a}_j \).

\[
\vec{v} \xrightarrow{P} \vec{y} \in \text{range}(A), \quad \vec{y} = A\vec{x}, \text{ some } \vec{x} \in \mathbb{C}^n
\]

\[
\vec{y} - \vec{v} \perp \text{range}(A)
\]

\[
\Rightarrow \vec{a}_j^*(\vec{y} - \vec{v}) = 0, \ \forall j
\]

\[
\Rightarrow \vec{a}_j^*(A\vec{x} - \vec{v}) = 0, \ \forall j
\]

\[
\Rightarrow A^*(A\vec{x} - \vec{v}) = 0
\]

\[
\Rightarrow A^*A\vec{x} = A^*\vec{v}
\]

\[
\Rightarrow \vec{x} = (A^*A)^{-1}A^*\vec{v}
\]

\[
\Rightarrow \vec{y} = \begin{pmatrix} A(A^*A)^{-1}A^* \end{pmatrix} \vec{v} = P\vec{v}
\]

We can build an orthogonal projector from an arbitrary (not necessarily orthogonal) basis.

Let \( S_1 \) be the subspace spanned by the linearly independent vectors \( \{\vec{a}_1, \ldots, \vec{a}_n\} \) and let \( A \) be the matrix with columns \( \vec{a}_j \).
The Full QR-Factorization

As for the SVD, we can extend the QR-factorization by padding \( \hat{Q} \) with an additional \((m-n)\) orthonormal columns, and zero-padding \( \hat{R} \) with an additional \((m-n)\) rows of zeros:

\[
\begin{bmatrix}
\hat{Q} \\
\hat{R}
\end{bmatrix} =
\begin{bmatrix}
Q \\
R
\end{bmatrix} =
\begin{bmatrix}
Q \\
R
\end{bmatrix} =
\begin{bmatrix}
\hat{Q} \\
\hat{R}
\end{bmatrix}
\]

**Figure:** The Reduced QR-Factorization, \( A = Q\hat{R} \)

**Figure:** The Full QR-Factorization, \( A = QR \)

In the full QR-factorization, the columns \( \hat{a}_j, j > n \) are orthogonal to \( \text{range}(A) \). If \( \text{rank}(A) = n \), they are an orthonormal basis for \( \text{range}(A)^\perp = \text{null}(A^*) \), the space orthogonal to \( \text{range}(A) \).

Algorithm: Classical Gram-Schmidt

We summarize our findings:

Algorithm (Classical Gram-Schmidt)

1. for \( k \in \{1, \ldots, n\} \) do
2. \( \hat{v}_k \leftarrow \hat{a}_k \)
3. for \( i \in \{1, \ldots, k-1\} \) do
4. \( r_{ik} \leftarrow \hat{q}_i^* \hat{a}_k \)
5. \( \hat{v}_k \leftarrow \hat{v}_k - r_{ik} \hat{q}_i \)
6. end for
7. \( r_{kk} \leftarrow \|\hat{v}_k\|_2 \)
8. \( \hat{q}_k \leftarrow \hat{v}_k / r_{kk} \)
9. end for

Mathematically, we are done. Numerically, however, we can run into trouble due to roundoff errors.

Building the QR-Factorization — Gram-Schmidt Orthogonalization

The equations on slide 20 outline a method for computing reduced QR-factorizations.

At the \( k \)th step, we are looking to construct \( \hat{q}_k \in \text{span}(\hat{a}_1, \ldots, \hat{a}_k) \) such that \( \hat{q}_k \perp \text{span}(\hat{a}_1, \ldots, \hat{a}_{k-1}) \)

We simply take \( \hat{a}_k \), and subtract all the projections onto the directions \( \hat{a}_1, \ldots, \hat{a}_{k-1} \), and then normalize the resulting vector

\[
(*) \quad \hat{v}_k = \hat{a}_k - (\hat{q}_1^* \hat{a}_1)\hat{q}_1 - \cdots - (\hat{q}_{k-1}^* \hat{a}_{k-1})\hat{q}_{k-1}
\]

\[
\hat{q}_k = \hat{v}_k / \|\hat{v}_k\|_2
\]

Computationally, it is more efficient to compute

\[
(*') \quad \hat{v}_k = \hat{a}_k - \hat{q}_1(\hat{q}_1^* \hat{a}_k) - \cdots - \hat{q}_{k-1}(\hat{q}_{k-1}^* \hat{a}_k)
\]

The QR-Factorization: Existence and Uniqueness

**Theorem (Existence of the QR-Factorization)**

Every \( A \in \mathbb{C}^{m \times n} (m \geq n) \) has a full QR-factorization, hence also a reduced QR-factorization.

**Theorem (Uniqueness of the QR-Factorization)**

Every \( A \in \mathbb{C}^{m \times n} (m \geq n) \) of full rank has a unique reduced QR-factorization \( A = \hat{Q}\hat{R} \), with \( r_{kk} > 0 \).
1. If $A$ is full rank, the Gram-Schmidt algorithm gives the unique reduced QR-factorization.

2. If $A$ does not have full rank, then $\vec{v}_k = 0$ can occur during the iteration; if it does set $\vec{q}_k$ to be an arbitrary vector orthogonal to $\text{span}(\vec{q}_1, \ldots, \vec{q}_{k-1})$, and proceed.

3. If $m > n$, follow Gram-Schmidt as described until $j = n$, then take an addition ($m - n$) steps, introducing arbitrary orthogonal $\vec{q}_k$ in each step.

* Column pivoting (exchanges) may be necessary.

---

### Homework #3

Implement the reduced QR-factorization by **Classical Gram-Schmidt**.

1. Write a function which given $A \in \mathbb{C}^{m \times n}$ computes $Q \in \mathbb{C}^{m \times n}$, and $R \in \mathbb{C}^{n \times n}$ — in MATLAB/Python you want to start something like this (e.g. file: qr_cgs.m, or qr_cgs.py):

   ```matlab
   function [Q,R] = qr_cgs(A)
   % Indentation does not matter...
   % Implicit Return of Results
   ```

   ```python
   def qr_cgs(A):
       # Indentation matters
       # Explicit Return of Results
       return Q, R
   ```

   See `help function` in MATLAB, or `python_functions` (clickable) for help on writing functions.

2. Validate your function — test that (i) $(A - QR) \approx 0$; (ii) $Q$ is unitary; and (iii) $R$ upper triangular. Show 3 test cases for $(3 \times 3)$, $(5 \times 5)$, and $(251 \times 251)$ matrices.

3. Compare the result for the $(3 \times 3)$, $(5 \times 5)$ cases with the built-in (“library”) version of the QR-factorization; comment on the similarities/differences. See `help qr` in MATLAB, or `numpy.linalg.qr` (clickable).


   - Hand in your code, and your validation/test-cases.
   - Infinity. Appropriately “tag” all pages with the corresponding question(s) in Gradescope.