Notes on Solving Linear Least-Squares Problems

Robert A. van de Geijn
The University of Texas at Austin
Austin, TX 78712

October 1, 2014

NOTE: I have not thoroughly proof-read these notes!!!

1 Motivation

For a motivation of the linear least-squares problem, read Week 10 (Sections 10.3-10.5) of Linear Algebra: Foundations to Frontiers - Notes to LAFF With.

2 The Linear Least-Squares Problem

Let $A \in \mathbb{C}^{m \times n}$ and $y \in \mathbb{C}^m$. Then the linear least-square problem (LLS) is given by

$$\text{Find } x \text{ s.t. } \|Ax - y\|_2 = \min_{z \in \mathbb{C}^n} \|Az - y\|_2.$$

In other words, x is the vector that minimizes the expression $\|Ax - y\|_2$. Equivalently, we can solve

$$\text{Find } x \text{ s.t. } \|Ax - y\|_2^2 = \min_{z \in \mathbb{C}^n} \|Az - y\|_2^2.$$

If x solves the linear least-squares problem, then Ax is the vector in $\mathcal{C}(A)$ (the column space of A) closest to the vector y.

3 Method of Normal Equations

Let $A \in \mathbb{R}^{m \times n}$ have linearly independent columns (which implies $m \geq n$). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be defined by

$$\begin{aligned} f(x) &= \|Ax - y\|_2^2 = (Ax - y)^T (Ax - y) = x^T A^T Ax - x^T A^T y - y^T Ax + y^T y \\ &= x^T A^T Ax - 2x^T A^T y + y^T y. \end{aligned}$$

This function is minimized when the gradient is zero, $\bigtriangledown f(x) = 0$. Now,

$$\bigtriangledown f(x) = 2A^T Ax - 2A^T y.$$

If A has linearly independent columns then $A^T A$ is nonsingular. Hence, the x that minimizes $\|Ax - y\|_2$ solves $A^T Ax = A^T y$. This is known as the method of normal equations. Notice that then

$$x = \underbrace{(A^T A)^{-1} A^T}_{A^\dagger} y,$$

where $A^\dagger$ is known as the *pseudo inverse* or *Moore-Penrose pseudo inverse*.

In practice, one performs the following steps:

- Form $B = A^T A$, a symmetric positive-definite (SPD) matrix.
  Cost: approximately mn^2 floating point operations (flops), if one takes advantage of symmetry.

- Compute the Cholesky factor L, a lower triangular matrix, so that $B = LL^T$.
  This factorization, discussed in Week 8 (Section 8.4.2) of Linear Algebra: Foundations to Frontiers - Notes to LAFF With and to be revisited later in this course, exists since B is SPD.
  Cost: approximately $\frac{1}{3}n^3$ flops.

- Compute $\hat{y} = A^T y$.
  Cost: $2mn$ flops.

- Solve $Lz = \hat{y}$ and $L^T x = z$.
  Cost: n^2 flops each.

Thus, the total cost of solving the LLS problem via normal equations is approximately $mn^2 + \frac{1}{3}n^3$ flops.

Remark 1. *We will* **later** *discuss that if A is not well-conditioned (its columns are nearly linearly dependent), the Method of Normal Equations is numerically unstable because $A^T A$ is ill-conditioned.*

The above discussion can be generalized to the case where $A \in \mathbb{C}^{m\times n}$. In that case, x must solve $A^H Ax = A^H y$.

A geometric explanation of the method of normal equations (for the case where A is real valued) can be found in Week 10 (Sections 10.3-10.5) of Linear Algebra: Foundations to Frontiers - Notes to LAFF With.

4 Solving the LLS Problem Via the QR Factorization

Assume $A \in \mathbb{C}^{m\times n}$ has linearly independent columns and let $A = Q_L R_{TL}$ be its QR factorization. We wish to compute the solution to the LLS problem: Find $x \in \mathbb{C}^n$ such that

$$\|Ax - y\|_2^2 = \min_{z\in\mathbb{C}^n} \|Az - y\|_2^2.$$

4.1 Simple derivation of the solution

Notice that we know that, if A has linearly independent columns, the solution is given by $x = (A^H A)^{-1} A^H y$ (the solution to the normal equations). Now,

$$\begin{array}{rcll}
x & = & [A^H A]^{-1} A^H y & \text{Solution to the Normal Equations} \\
 & = & \left[(Q_L R_{TL})^H (Q_L R_{TL})\right]^{-1} (Q_L R_{TL})^H y & A = Q_L R_{TL} \\
 & = & \left[R_{TL}^H Q_L^H Q_L R_{TL}\right]^{-1} R_{TL}^H Q_L^H y & (BC)^H = (C^H B^H) \\
 & = & \left[R_{TL}^H R_{TL}\right]^{-1} R_{TL}^H Q_L^H y & Q_L^H Q_L = I \\
 & = & R_{TL}^{-1} R_{TL}^{-H} R_{TL}^H Q_L^H y & (BC)^{-1} = C^{-1} B^{-1} \\
 & = & R_{TL}^{-1} Q_L^H y & R_{TL}^{-H} R_{TL}^H = I
\end{array}$$

Thus, the x that solves $R_{TL} x = Q_L^H y$ solves the LLS problem.

4.2 Alternative derivation of the solution

We know that then there exists a matrix Q_R such that $Q = \left(\begin{array}{c|c} Q_L & Q_R \end{array} \right)$ is unitary. Now,

$$
\begin{array}{lll}
\multicolumn{2}{l}{\min_{z \in \mathbb{C}^n} \|Az - y\|_2^2} & \\
= & \min_{z \in \mathbb{C}^n} \|Q_L R_{TL} z - y\|_2^2 & \text{(substitute } A = Q_L R_{TL}\text{)} \\
= & \min_{z \in \mathbb{C}^n} \|Q^H (Q_L R_{TL} z - y)\|_2^2 & \text{(two-norm is preserved since } Q^H \text{ is unitary)} \\
= & \min_{z \in \mathbb{C}^n} \left\| \left(\begin{array}{c} Q_L^H \\ \hline Q_R^H \end{array} \right) Q_L R_{TL} z - \left(\begin{array}{c} Q_L^H \\ \hline Q_R^H \end{array} \right) y \right\|_2^2 & \text{(partitioning, distributing)} \\
= & \min_{z \in \mathbb{C}^n} \left\| \left(\begin{array}{c} R_{TL} z \\ \hline 0 \end{array} \right) - \left(\begin{array}{c} Q_L^H y \\ \hline Q_R^H y \end{array} \right) \right\|_2^2 & \text{(partitioned matrix-matrix multiplication)} \\
= & \min_{z \in \mathbb{C}^n} \left\| \left(\begin{array}{c} R_{TL} z - Q_L^H y \\ \hline -Q_R^H y \end{array} \right) \right\|_2^2 & \text{(partitioned matrix addition)} \\
= & \min_{z \in \mathbb{C}^n} \left(\left\| R_{TL} z - Q_L^H y \right\|_2^2 + \|Q_R^H y\|_2^2 \right) & \text{(property of the 2-norm:} \\
& & \left\| \left(\begin{array}{c} x \\ \hline y \end{array} \right) \right\|_2^2 = \|x\|_2^2 + \|y\|_2^2\text{)} \\
= & \left(\min_{z \in \mathbb{C}^n} \left\| R_{TL} z - Q_L^H y \right\|_2^2 \right) + \|Q_R^H y\|_2^2 & \text{(} Q_R^H y \text{ is independent of } z\text{)} \\
= & \|Q_R^H y\|_2^2 & \text{(minimized by } x \text{ that satisfies } R_{TL} x = Q_L^H y\text{)}
\end{array}
$$

Thus, the desired x that minimizes the linear least-squares problem solves $R_{TL} x = Q_L^H y$. The solution is unique because R_{TL} is nonsingular (because A has linearly independent columns).

In practice, one performs the following steps:

- Compute the QR factorization $A = Q_L R_{TL}$.
  If Gram-Schmidt or Modified Gram-Schmidt are used, this costs $2mn^2$ flops.
- Form $\hat{y} = Q_L^H y$.
  Cost: $2mn$ flops.
- Solve $R_{TL} x = \hat{y}$ (triangular solve).
  Cost: n^2 flops.

Thus, the total cost of solving the LLS problem via (Modified) Gram-Schmidt QR factorization is approximately $2mn^2$ flops.

Notice that the solution computed by the Method of Normal Equations (generalized to the complex case) is given by

$$
\begin{array}{rcl}
(A^H A)^{-1} A^H y & = & ((Q_L R_{TL})^H (Q_L R_{TL}))^{-1} (Q_L R_{TL})^H y = (R_{TL}^H Q_L^H Q_L R_{TL})^{-1} R_{TL}^H Q_L^H y \\
& = & (R_{TL}^H R_{TL})^{-1} R_{TL}^H Q_L^H y = R_{TL}^{-1} R_{TL}^{-H} R_{TL}^H Q_L^H y = R_{TL}^{-1} Q_L^H y = R_{TL}^{-1} \hat{y} = x
\end{array}
$$

where $R_{TL} x = \hat{y}$. This shows that the two approaches compute the same solution, generalizes the Method of Normal Equations to complex valued problems, and shows that the Method of Normal Equations computes the desired result without requiring multivariate calculus.

5 Via Householder QR Factorization

Given $A \in \mathbb{C}^{m\times n}$ with linearly independent columns, the Householder QR factorization yields n Householder transformations, $H_0, \ldots, H_{n-1}$, so that

$$\underbrace{H_{n-1}\cdots H_0}_{Q = \left(Q_L \mid Q_R \right)^H} A = \left(\begin{array}{c} R_{TL} \\ \hline 0 \end{array} \right).$$

We wish to solve $R_{TL}x = \underbrace{Q_L^H y}_{\hat{y}}$. But

$$\begin{aligned} \hat{y} = Q_L^H y &= \left[\left(I \mid 0 \right) \left(\begin{array}{c} Q_L^H \\ \hline Q_R^H \end{array} \right) \right] y = \left(I \mid 0 \right) \left(Q_L \mid Q_R \right)^H y = \left(I \mid 0 \right) Q^H y \\ &= \left(I \mid 0 \right) (H_{n-1}\cdots H_0) y. = \left(I \mid 0 \right) (\underbrace{H_{n-1}\cdots H_0 y}_{w = \left(\begin{array}{c} w_T \\ \hline w_B \end{array} \right)}) = w_T. \end{aligned}$$

This suggests the following approach:

- Compute $H_0, \ldots, H_{n-1}$ so that $H_{n-1}\cdots H_0 A = \left(\begin{array}{c} R_{TL} \\ \hline 0 \end{array} \right)$, storing the Householder vectors that define $H_0, \ldots, H_{n-1}$ over the elements in A that they zero out (see "Notes on Householder QR Factorization").
  Cost: $2mn^2 - \frac{2}{3}n^3$ flops.
- Form $w = (H_{n-1}(\cdots(H_0 y)\cdots))$ (see "Notes on Householder QR Factorization"). Partition $w = \left(\begin{array}{c} w_T \\ \hline w_B \end{array} \right)$ where $w_T \in \mathbb{C}^n$. Then $\hat{y} = w_T$.
  Cost: $4m^2 - 2n^2$ flops. (See "Notes on Householder QR Factorization" regarding this.)
- Solve $R_{TL}x = \hat{y}$.
  Cost: n^2 flops.

Thus, the total cost of solving the LLS problem via Householder QR factorization is approximately $2mn^2 - \frac{2}{3}n^3$ flops. This is cheaper than using (Modified) Gram-Schmidt QR factorization, and hence preferred (because it is also numerically more stable, as we will discuss later in the course).

6 Via the Singular Value Decomposition

Given $A \in \mathbb{C}^{m\times n}$ with linearly independent columns, let $A = U\Sigma V^H$ be its SVD decomposition. Partition

$$U = \left(U_L \mid U_R \right) \quad \text{and} \quad \Sigma = \left(\begin{array}{c} \Sigma_{TL} \\ \hline 0 \end{array} \right),$$

where $U_L \in \mathbb{C}^{m\times n}$ and $\Sigma_{TL} \in \mathbb{R}^{n\times n}$ so that

$$A = \left(U_L \mid U_R \right) \left(\begin{array}{c} \Sigma_{TL} \\ \hline 0 \end{array} \right) V^H = U_L \Sigma_{TL} V^H.$$

We wish to compute the solution to the LLS problem: Find $x \in \mathbb{C}^n$ such that

$$\|Ax - y\|_2^2 = \min_{z \in \mathbb{C}^n} \|Az - y\|_2^2.$$

6.1 Simple derivation of the solution

Notice that we know that, if A has linearly independent columns, the solution is given by $x = (A^H A)^{-1} A^H y$ (the solution to the normal equations). Now,

$$\begin{array}{rcll}
x & = & [A^H A]^{-1} A^H y & \text{Solution to the Normal Equations} \\
 & = & \left[(U_L \Sigma_{TL} V^H)^H (U_L \Sigma_{TL} V^H)\right]^{-1} (U_L \Sigma_{TL} V^H)^H y & A = U_L \Sigma_{TL} V^H \\
 & = & \left[(V \Sigma_{TL} U_L^H)(U_L \Sigma_{TL} V^H)\right]^{-1} (V \Sigma_{TL} U_L^H) y & (BCD)^H = (D^H C^H B^H) \text{ and } \Sigma_{TL}^H = \Sigma_{TL} \\
 & = & \left[V \Sigma_{TL} \Sigma_{TL} V^H\right]^{-1} V \Sigma_{TL} U_L^H y & U_L^H U_L = I \\
 & = & V \Sigma_{TL}^{-1} \Sigma_{TL}^{-1} V^H V \Sigma_{TL} U_L^H y & V^{-1} = V^H \text{ and } (BCD)^{-1} = D^{-1} C^{-1} B^{-1} \\
 & = & V \Sigma_{TL}^{-1} U_L^H y & V^H V = I \text{ and } \Sigma_{TL}^{-1} \Sigma_{TL} = I
\end{array}$$

6.2 Alternative derivation of the solution

We now discuss a derivation of the result that does not depend on the Normal Equations, in preparation for the more general case discussed in the next section.

$$\begin{array}{cll}
\multicolumn{3}{l}{\min_{z \in \mathbb{C}^n} \|Az - y\|_2^2} \\
= & \min_{z \in \mathbb{C}^n} \|U \Sigma V^H z - y\|_2^2 & \text{(substitute } A = U \Sigma V^H\text{)} \\
= & \min_{z \in \mathbb{C}^n} \|U(\Sigma V^H z - U^H y)\|_2^2 & \text{(substitute } U U^H = I \text{ and factor out } U\text{)} \\
= & \min_{z \in \mathbb{C}^n} \|\Sigma V^H z - U^H y\|_2^2 & \text{(multiplication by a unitary matrix} \\
 & & \text{preserves two-norm)} \\
= & \min_{z \in \mathbb{C}^n} \left\| \left(\frac{\Sigma_{TL}}{0} \right) V^H z - \left(\frac{U_L^H y}{U_R^H y} \right) \right\|_2^2 & \text{(partition, partitioned matrix-matrix multiplication)} \\
= & \min_{z \in \mathbb{C}^n} \left\| \left(\frac{\Sigma_{TL} V^H z - U_L^H y}{-U_R^H y} \right) \right\|_2^2 & \text{(partitioned matrix-matrix multiplication and addition)} \\
= & \min_{z \in \mathbb{C}^n} \left\| \Sigma_{TL} V^H z - U_L^H y \right\|_2^2 + \left\| U_R^H y \right\|_2^2 & \left(\left\| \left(\frac{v_T}{v_B} \right) \right\|_2^2 = \|v_T\|_2^2 + \|v_B\|_2^2 \right)
\end{array}$$

The x that solves $\Sigma_{TL} V^H x = U_L^H y$ minimizes the expression. That x is given by $x = V \Sigma_{TL}^{-1} U_L^H y$.

This suggests the following approach:

- Compute the reduced SVD: $A = U_L \Sigma_{TL} V^H$.
  Cost: Greater than computing the QR factorization! We will discuss this in a future note.
- Form $\hat{y} = \Sigma_{TL}^{-1} U_L^H y$.
  Cost: approx. $2mn$ flops.
- Compute $z = V \hat{y}$.
  Cost: approx. $2mn$ flops.

7 What If A Does Not Have Linearly Independent Columns?

In the above discussions we assume that A has linearly independent columns. Things get a bit trickier if A does not have linearly independent columns. There is a variant of the QR factorization known as the QR factorization with column pivoting that can be used to find the solution. We instead focus on using the SVD.

Given $A \in \mathbb{C}^{m\times n}$ with $\text{rank}(A) = r < n$, let $A = U\Sigma V^H$ be its SVD decomposition. Partition

$$U = \left(\begin{array}{c|c} U_L & U_R \end{array} \right), \quad V = \left(\begin{array}{c|c} V_L & V_R \end{array} \right) \quad \text{and} \quad \Sigma = \left(\begin{array}{c|c} \Sigma_{TL} & 0 \\ \hline 0 & 0 \end{array} \right),$$

where $U_L \in \mathbb{C}^{m\times r}$, $V_L \in \mathbb{C}^{n\times r}$ and $\Sigma_{TL} \in \mathbb{R}^{r\times r}$ so that

$$A = \left(\begin{array}{c|c} U_L & U_R \end{array} \right) \left(\begin{array}{c|c} \Sigma_{TL} & 0 \\ \hline 0 & 0 \end{array} \right) \left(\begin{array}{c|c} V_L & V_R \end{array} \right)^H = U_L \Sigma_{TL} V_L^H.$$

Now,

$$\begin{array}{lll}
\multicolumn{2}{l}{\min_{z\in\mathbb{C}^n} \|Az - y\|_2^2} & \\
= & \min_{z\in\mathbb{C}^n} \|U\Sigma V^H z - y\|_2^2 & (\text{substitute } A = U\Sigma V^H) \\
= & \min_{z\in\mathbb{C}^n} \|U\Sigma V^H z - UU^H y\|_2^2 & (UU^H = I) \\
= & \min_{\left\{\begin{array}{c} w\in\mathbb{C}^n \\ z = Vw \end{array}\right\}} \|U\Sigma V^H V w - UU^H y\|_2^2 & \begin{array}{l}(\text{choosing the max over } w \in \mathbb{C}^n \text{ with } z = Vw \\ \text{is the same as choosing the max over } z \in \mathbb{C}^n.)\end{array} \\
= & \min_{\left\{\begin{array}{c} w\in\mathbb{C}^n \\ z = Vw \end{array}\right\}} \|U(\Sigma w - U^H y)\|_2^2 & (\text{factor out } U \text{ and } V^H V = I) \\
= & \min_{\left\{\begin{array}{c} w\in\mathbb{C}^n \\ z = Vw \end{array}\right\}} \|\Sigma w - U^H y\|_2^2 & (\|Uv\|_2 = \|v\|_2) \\
= & \min_{\left\{\begin{array}{c} w\in\mathbb{C}^n \\ z = Vw \end{array}\right\}} \| \left(\begin{array}{c|c} \Sigma_{TL} & 0 \\ \hline 0 & 0 \end{array} \right) \left(\begin{array}{c} w_T \\ \hline w_B \end{array} \right) - \left(\begin{array}{c} U_L^H \\ \hline U_R^H \end{array} \right) y \|_2^2 & (\text{partition } \Sigma,\ w, \text{ and } U) \\
= & \min_{\left\{\begin{array}{c} w\in\mathbb{C}^n \\ z = Vw \end{array}\right\}} \left\| \left(\begin{array}{c} \Sigma_{TL} w_T - U_L^H y \\ \hline -U_R^H y \end{array} \right) \right\|_2^2 & (\text{partitioned matrix-matrix multiplication}) \\
= & \min_{\left\{\begin{array}{c} w\in\mathbb{C}^n \\ z = Vw \end{array}\right\}} \left\|\Sigma_{TL} w_T - U_L^H y\right\|_2^2 + \left\|U_R^H y\right\|_2^2 & \left(\left\| \left(\begin{array}{c} v_T \\ \hline v_B \end{array} \right) \right\|_2^2 = \|v_T\|_2^2 + \|v_B\|_2^2 \right)
\end{array}$$

Since Σ_{TL} is a diagonal with no zeroes on its diagonal, we know that Σ_{TL}^{-1} exists. Choosing $w_T = \Sigma_{TL}^{-1} U_L^H y$ means that

$$\min_{\left\{\begin{array}{c} w\in\mathbb{C}^n \\ z = Vw \end{array}\right\}} \left\|\Sigma_{TL} w_T - U_L^H y\right\|_2^2 = 0,$$

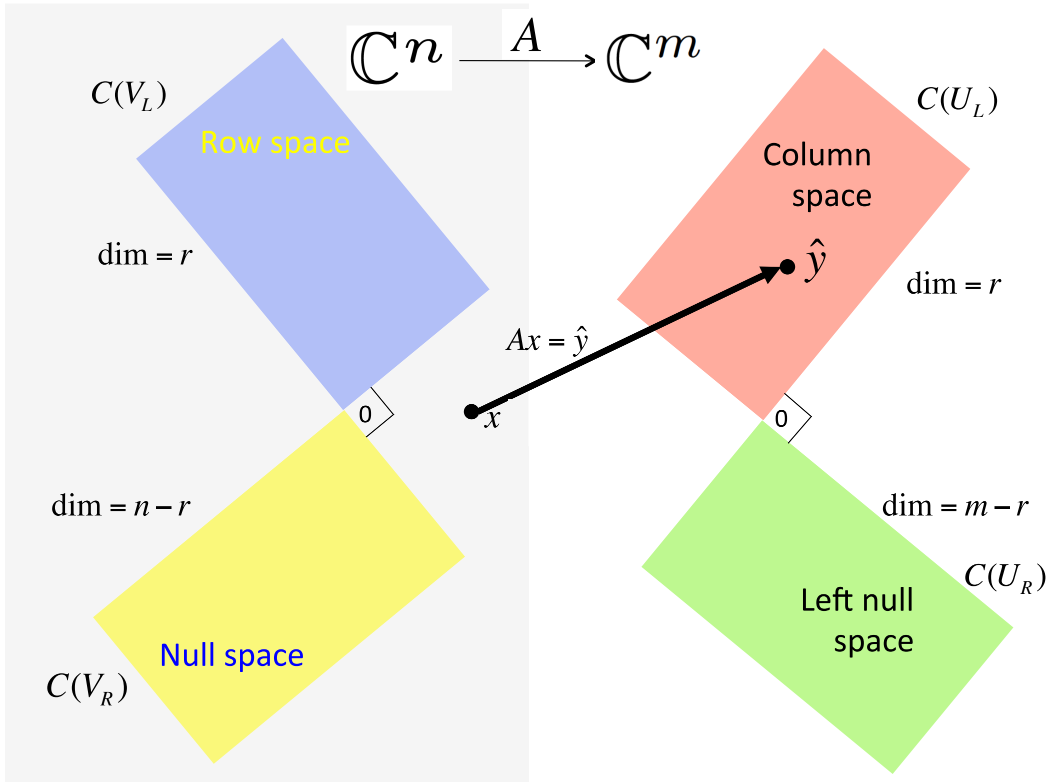
which obviously minimizes the entire expression. We conclude that

$$x = Vw = \left(\begin{array}{c|c} V_L & V_R \end{array} \right) \left(\begin{array}{c} \Sigma_{TL}^{-1} U_L^H y \\ \hline w_B \end{array} \right) = V_L \Sigma_{TL}^{-1} U_L^H y + V_R w_B$$

characterizes all solutions to the linear least-squares problem, where w_B can be chosen to be any vector of size $n - r$. By choosing $w_B = 0$ and hence $x = V_L \Sigma_{TL}^{-1} U_L^H y$ we choose the vector x that itself has minimal 2-norm.

The sequence of pictures on the following pages reasons through the insights that we gained so far (in "Notes on the Singular Value Decomposition" and this note). These pictures can be downloaded as a PowerPoint presentation from

http://www.cs.utexas.edu/users/flame/Notes/Spaces.pptx



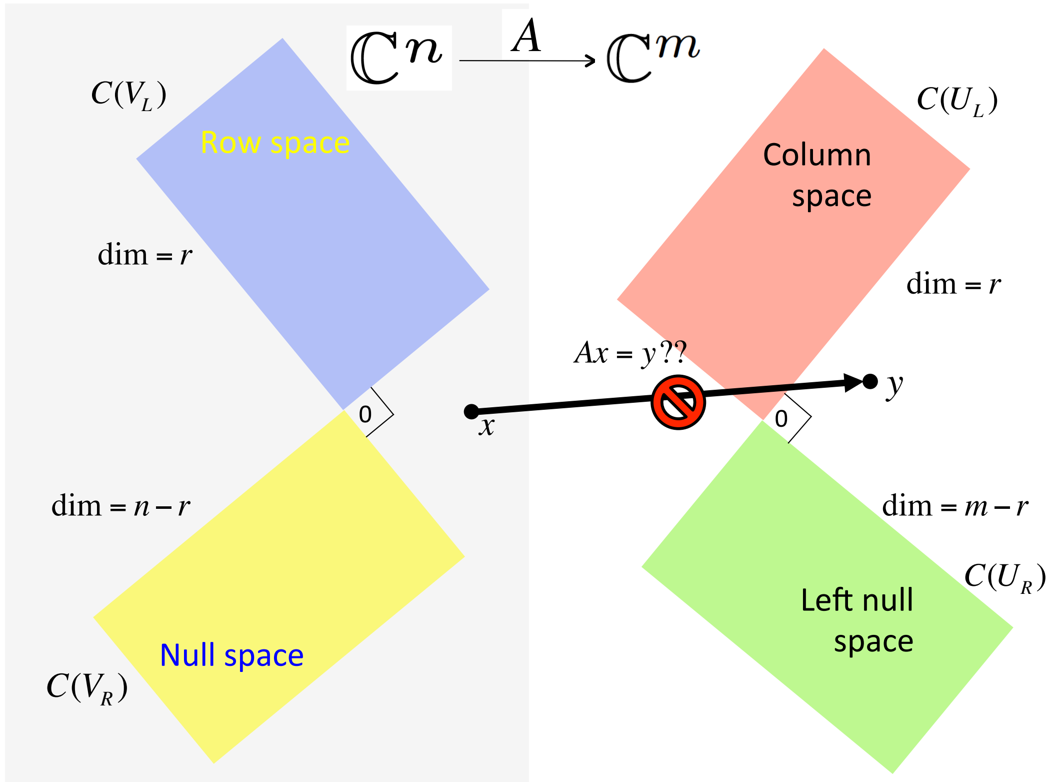
If $A \in \mathbb{C}^{m \times n}$ and

$$A = \left(\begin{array}{c|c} U_L & U_R \end{array} \right) \left(\begin{array}{c|c} \Sigma_{TL} & 0 \\ \hline 0 & 0 \end{array} \right) \left(\begin{array}{c|c} V_L & V_R \end{array} \right)^H = U_L \Sigma_{TL} V_L^H$$

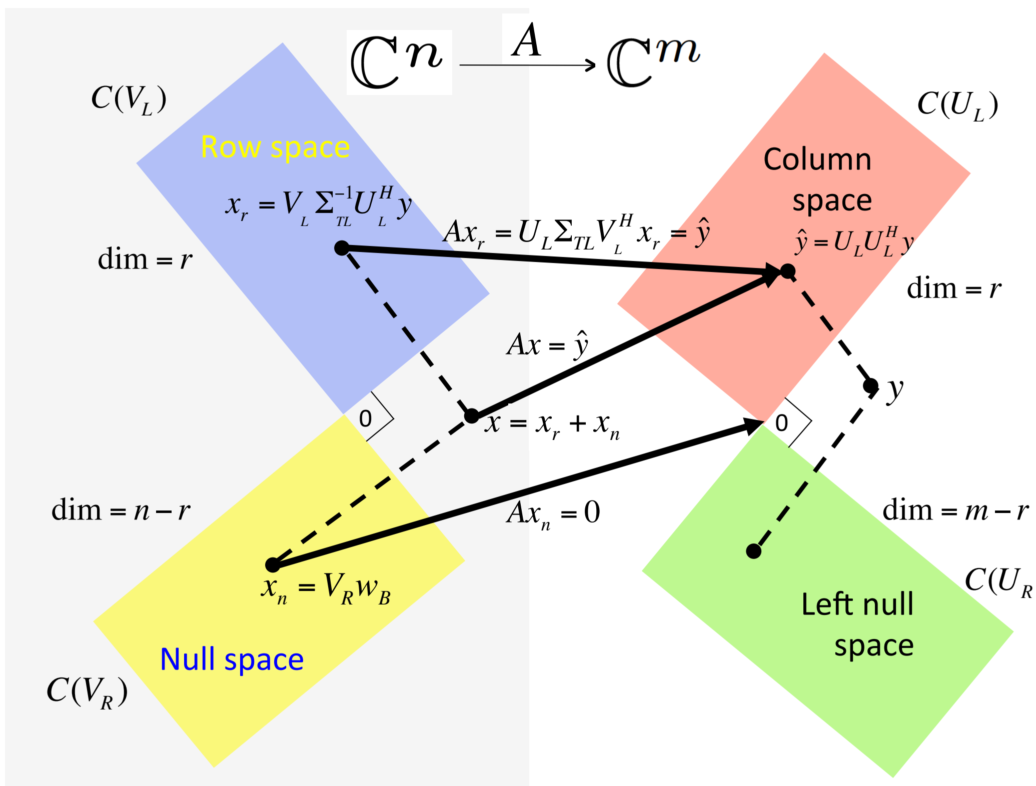
equals the SVD, where $U_L \in \mathbb{C}^{m \times r}$, $V_L \in \mathbb{C}^{n \times r}$, and $\Sigma_{TL} \in \mathbb{C}^{r \times r}$, then

- The row space of A equals $\mathcal{C}(V_L)$, the column space of V_L;
- The null space of A equals $\mathcal{C}(V_R)$, the column space of V_R;
- The column space of A equals $\mathcal{C}(U_L)$; and
- The left null space of A equals $\mathcal{C}(U_R)$.

Also, given a vector $x \in \mathbb{C}^n$, the matrix A maps x to $\widehat{y} = Ax$, which must be in $\mathcal{C}(A) = \mathcal{C}(U_L)$.



Given an arbitrary $y \in \mathbb{C}^m$ not in $\mathcal{C}(A) = \mathcal{C}(U_L)$, there cannot exist a vector $x \in \mathbb{C}^n$ such that $Ax = y$.



The solution to the Linear Least-Squares problem, x, equals any vector that is mapped by A to the projection of y onto the column space of A: $\widehat{y} = A(A^H A)^{-1} A^H y = QQ^H y$. This solution can be written as the sum of a (unique) vector in the row space of A and any vector in the null space of A. The vector in the row space of A is given by

$$x_r = V_L \Sigma_{TL}^{-1} U_L^H y == V_L \Sigma_{TL}^{-1} U_L^H \widehat{y}.$$

The sequence of pictures, and their explanations, suggest a much simply path towards the formula for solving the LLS problem.

- We know that we are looking for the solution x to the equation

$$Ax = U_L U_L^H y.$$

- We know that there must be a solution x_r in the row space of A. It suffices to find w_T such that $x_r = V_L w_T$.
- Hence we search for w_T that satisfies

$$A V_L w_T = U_L U_L^H y.$$

- Since there is a one-to-one mapping by A from the row space of A to the column space of A, we know that w_T is unique. Thus, if we find *a* solution to the above, we have found *the* solution.

- Multiplying both sides of the equation by U_L^H yields

$$U_L^H A V_L w_T = U_L^H y.$$

- Since $A = U_L \Sigma_{TL}^{-1} V_L^H$, we can rewrite the above equation as

$$\Sigma_{TL} w_T = U_L^H y$$

so that $w_T = \Sigma_{TL}^{-1} U_L^H y$.

- Hence

$$x_r = V_L \Sigma_{TL}^{-1} U_L^H y.$$

- Adding any vector in the null space of A to x_r also yields a solution. Hence all solutions to the LLS problem can be characterized by

$$x = V_L \Sigma_{TL}^{-1} U_L^H y + V_R w_R.$$

Here is yet another important way of looking at the problem:

- **We start by considering the LLS problem: Find $x \in \mathbb{C}^n$ such that**

$$\|Ax - y\|_2^2 = \max_{z \in \mathbb{C}^n} \|Az - y\|_2^2.$$

- **We changed this into the problem of finding w_L that satisfied**

$$\Sigma_{TL} w_L = v_T$$

where $x = V_L w_L$ and $\hat{y} = U_L U_L^H y = U_L v_T$.

- **Thus, by expressing x in the right basis (the columns of V_L) and the projection of y in the right basis (the columns of U_L), the problem became trivial, since the matrix that related the solution to the right-hand side became diagonal.**

8 Exercise: Using the the LQ factorization to solve underdetermined systems

We next discuss another special case of the LLS problem: Let $A \in \mathbb{C}^{m \times n}$ where $m < n$ and A has linearly independent rows. A series of exercises will lead you to a practical algorithm for solving the problem of describing all solutions to the LLS problem

$$\|Ax - y\|_2 = \min_z \|Az - y\|_2.$$

You may want to review "Notes on the QR Factorization" as you do this exercise.

Exercise 2. *Let $A \in \mathbb{C}^{m \times n}$ with $m < n$ have linearly independent rows. Show that there exist a lower triangular matrix $L_L \in \mathbb{C}^{m \times m}$ and a matrix $Q_T \in \mathbb{C}^{m \times n}$ with orthonormal rows such that $A = L_L Q_T$, noting that L_L does not have any zeroes on the diagonal. Letting $L = \left(\begin{array}{c|c} L_L & 0 \end{array} \right)$ be $\mathbb{C}^{m \times n}$ and unitary $Q = \left(\begin{array}{c} Q_T \\ \hline Q_B \end{array} \right)$, reason that $A = LQ$.*

Don't overthink the problem: use results you have seen before.

Algorithm: $[L, Q] := \text{LQ\_CGS\_UNB}(A, L, Q)$

Partition $A \rightarrow \left(\begin{array}{c} A_T \\ \hline A_B \end{array} \right)$, $L \rightarrow \left(\begin{array}{c|c} L_{TL} & L_{TR} \\ \hline L_{BL} & L_{BR} \end{array} \right)$, $Q \rightarrow \left(\begin{array}{c} Q_T \\ \hline Q_B \end{array} \right)$

where A_T has 0 rows, L_{TL} is 0×0, Q_T has 0 rows

while $m(A_T) < m(A)$ **do**

Repartition

$$\left(\begin{array}{c} A_T \\ \hline A_B \end{array} \right) \rightarrow \left(\begin{array}{c} A_0 \\ \hline a_1^T \\ \hline A_2 \end{array} \right), \left(\begin{array}{c|c} L_{TL} & L_{TR} \\ \hline L_{BL} & L_{BR} \end{array} \right) \rightarrow \left(\begin{array}{c|c|c} L_{00} & l_{01} & L_{02} \\ \hline l_{10}^T & \lambda_{11} & l_{12}^T \\ \hline L_{20} & l_{21} & L_{22} \end{array} \right), \left(\begin{array}{c} Q_T \\ \hline Q_B \end{array} \right) \rightarrow \left(\begin{array}{c} Q_0 \\ \hline q_1^T \\ \hline Q_2 \end{array} \right)$$

where a_1 has 1 row, λ_{11} is 1×1, q_1 has 1 row

Continue with

$$\left(\begin{array}{c} A_T \\ \hline A_B \end{array} \right) \leftarrow \left(\begin{array}{c} A_0 \\ \hline a_1^T \\ \hline A_2 \end{array} \right), \left(\begin{array}{c|c} L_{TL} & L_{TR} \\ \hline L_{BL} & L_{BR} \end{array} \right) \leftarrow \left(\begin{array}{c|c|c} L_{00} & l_{01} & L_{02} \\ \hline l_{10}^T & \lambda_{11} & l_{12}^T \\ \hline L_{20} & l_{21} & L_{22} \end{array} \right), \left(\begin{array}{c} Q_T \\ \hline Q_B \end{array} \right) \leftarrow \left(\begin{array}{c} Q_0 \\ \hline q_1^T \\ \hline Q_2 \end{array} \right)$$

endwhile

Figure 1: Algorithm skeleton for CGS-like LQ factorization.

Exercise 3. *Let $A \in \mathbb{C}^{m \times n}$ with $m < n$ have linearly independent rows. Consider*

$$\|Ax - y\|_2 = \min_z \|Az - y\|_2.$$

Use the fact that $A = L_L Q_T$, where $L_L \in \mathbb{C}^{m \times m}$ is lower triangular and Q_T has orthonormal rows, to argue that any vector of the form $Q_T^H L_L^{-1} y + Q_B^H w_B$ (where w_B is any vector in $\mathbb{C}^{n-m}$) is a solution to the LLS problem. Here $Q = \left(\begin{array}{c} Q_T \\ \hline Q_B \end{array} \right)$.

Exercise 4. *Continuing Exercise 2, use Figure 1 to give a Classical Gram-Schmidt inspired algorithm for computing L_L and Q_T. (The best way to check you got the algorithm right is to implement it!)*

Exercise 5. *(Optional) Continuing Exercise 2, use Figure 2 to give a Householder QR factorization inspired algorithm for computing L and Q, leaving L in the lower triangular part of A and Q stored as Householder vectors above the diagonal of A. (The best way to check you got the algorithm right is to implement it!)*

Algorithm: $[A, t] := \text{HLQ\_UNB}(A, t)$

Partition $A \rightarrow \left(\begin{array}{c|c} A_{TL} & A_{TR} \\ \hline A_{BL} & A_{BR} \end{array}\right)$, $t \rightarrow \left(\begin{array}{c} t_T \\ \hline t_B \end{array}\right)$

where A_{TL} is 0×0, t_T has 0 rows

while $m(A_{TL}) < m(A)$ **do**

Repartition

$$\left(\begin{array}{c|c} A_{TL} & A_{TR} \\ \hline A_{BL} & A_{BR} \end{array}\right) \rightarrow \left(\begin{array}{c|c|c} A_{00} & a_{01} & A_{02} \\ \hline a_{10}^T & \alpha_{11} & a_{12}^T \\ \hline A_{20} & a_{21} & A_{22} \end{array}\right), \left(\begin{array}{c} t_T \\ \hline t_B \end{array}\right) \rightarrow \left(\begin{array}{c} t_0 \\ \hline \tau_1 \\ \hline t_2 \end{array}\right)$$

where α_{11} is 1×1, τ_1 has 1 row

Continue with

$$\left(\begin{array}{c|c} A_{TL} & A_{TR} \\ \hline A_{BL} & A_{BR} \end{array}\right) \leftarrow \left(\begin{array}{c|c|c} A_{00} & a_{01} & A_{02} \\ \hline a_{10}^T & \alpha_{11} & a_{12}^T \\ \hline A_{20} & a_{21} & A_{22} \end{array}\right), \left(\begin{array}{c} t_T \\ \hline t_B \end{array}\right) \leftarrow \left(\begin{array}{c} t_0 \\ \hline \tau_1 \\ \hline t_2 \end{array}\right)$$

endwhile

Figure 2: Algorithm skeleton for Householder QR factorization inspired LQ factorization.