Non-Square Matrices

There is a useful variation on the concept of eigenvalues and eigenvectors which is defined for both square and non-square matrices. Throughout this discussion, for $A \in \mathbb{C}^{m\times m}$, let $\|A\|$ denote the operator norm induced by the Euclidean norms on $\mathbb{C}^n$ and $\mathbb{C}^m$ (which we denote by $\|\cdot\|$), and let $\|A\|_F$ denote the Frobenius norm of A. Note that we still have

$$\langle Ax, y\rangle_{\mathbb{C}^m} = y^H Ax = \langle x, A^H y\rangle_{\mathbb{C}^n} \quad \text{for} \quad x \in \mathbb{C}^n,\ y \in \mathbb{C}^m.$$

From $A \in \mathbb{C}^{m\times n}$ one can construct the square matrices $A^H A \in \mathbb{C}^{n\times n}$ and $AA^H \in \mathbb{C}^{m\times m}$. Both of these are Hermitian positive semi-definite. In particular $A^H A$ and AA^H are diagonalizable with real non-negative eigenvalues. Except for the multiplicities of the zero eigenvalue, these matrices have the same eigenvalues; in fact, we have:

Proposition. Let $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$ with $m \leq n$. Then the eigenvalues of BA (counting multiplicity) are the eigenvalues of AB, together with $n-m$ zeroes. (Remark: For $n = m$, this was problem 4 on Problem Set 5.)

Proof. Consider the $(n+m)\times(n+m)$ matrices

$$C_1 = \begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix} \quad \text{and} \quad C_2 = \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix}.$$

These are similar since $S^{-1}C_1 S = C_2$ where

$$S = \begin{bmatrix} I & A \\ 0 & I \end{bmatrix} \quad \text{and} \quad S^{-1} = \begin{bmatrix} I & -A \\ 0 & I \end{bmatrix}.$$

But the eigenvalues of C_1 are those of AB along with n zeroes, and the eigenvalues of C_2 are those of BA along with m zeroes. The result follows. □

So for any m, n, the eigenvalues of $A^H A$ and AA^H differ by $|n-m|$ zeroes. Let $p = \min(m,n)$ and let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ (≥ 0) be the joint eigenvalues of $A^H A$ and AA^H.

Definition. The *singular values* of A are the numbers

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0,$$

where $\sigma_i = \sqrt{\lambda_i}$. (When $n > m$, one often also defines singular values $\sigma_{m+1} = \cdots = \sigma_n = 0$.)

It is a fundamental result that one can choose orthonormal bases for $\mathbb{C}^n$ and $\mathbb{C}^m$ so that A maps one into the other, scaled by the singular values. Let $\Sigma = \text{diag}\,(\sigma_1, \ldots, \sigma_p) \in \mathbb{C}^{m\times n}$ be the "diagonal" matrix whose ii entry is σ_i $(1 \leq i \leq p)$.

Singular Value Decomposition (SVD)

If $A \in \mathbb{C}^{m\times n}$, then there exists unitary matrices $U \in \mathbb{C}^{m\times m}$, $V \in \mathbb{C}^{n\times n}$ such that $A = U\Sigma V^H$, where $\Sigma \in \mathbb{C}^{m\times n}$ is the diagonal matrix of singular values.

Proof. As in the square case, $\|A\|^2 = \|A^H A\|$. But

$$\|A^H A\| = \lambda_1 = \sigma_1^2, \quad \text{so} \quad \|A\| = \sigma_1.$$

So we can choose $x \in \mathbb{C}^n$ with $\|x\| = 1$ and $\|Ax\| = \sigma_1$. Write $Ax = \sigma_1 y$ where $\|y\| = 1$. Complete x and y to unitary matrices

$$V_1 = [x, \widetilde{v}_2, \cdots, \widetilde{v}_n] \in \mathbb{C}^{n\times n} \quad \text{and} \quad U_1 = [y, \widetilde{u}_2, \cdots, \widetilde{u}_m] \in \mathbb{C}^{m\times m}.$$

Since $U_1^H A V_1 \equiv A_1$ is the matrix of A in these bases it follows that

$$A_1 = \begin{bmatrix} \sigma_1 & w^H \\ 0 & B \end{bmatrix}$$

for some $w \in \mathbb{C}^{n-1}$ and $B \in \mathbb{C}^{(m-1)\times(n-1)}$. Now observe that

$$\begin{aligned} \sigma_1^2 + w^* w &\leq \left\| \begin{bmatrix} \sigma_1^2 + w^* w \\ Bw \end{bmatrix} \right\| \\ &= \left\| A_1 \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\| \\ &\leq \|A_1\| \cdot \left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\| \\ &= \sigma_1 (s_1^2 + w^* w)^{\frac{1}{2}} \end{aligned}$$

since $\|A_1\| = \|A\| = \sigma_1$ by the invariance of $\|\cdot\|$ under unitary multiplication.

It follows that $(\sigma_1^2 + w^* w)^{\frac{1}{2}} \leq \sigma_1$, so $w = 0$, and thus

$$A_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix}.$$

Now apply the same argument to B and repeat to get the result. For this, observe that

$$\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & B^H B \end{bmatrix} = A_1^H A_1 = V_1^H A^H A V_1$$

is unitarily similar to $A^H A$, so the eigenvalues of $B^H B$ are $\lambda_2 \geq \cdots \geq \lambda_n$ (≥ 0). Observe also that the same argument shows that if $A \in \mathbb{R}^{m\times n}$, then U and V can be taken to be real orthogonal matrices. □

This proof given above is direct, but it masks some of the key ideas. We now sketch an alternative proof that reveals more of the underlying structure of the SVD decomposition.

Alternative Proof of SVD: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $A^H A$ associated with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ (≥ 0), respectively, and let $V = [v_1 \cdots v_n] \in \mathbb{C}^{n\times n}$. Then V is unitary, and

$$V^H A^H A V = \wedge \equiv \text{diag}\,(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n\times n}.$$

For $1 \leq i \leq n$,

$$\|Av_i\|^2 = e_i^H V^H A^H A V e_i = \lambda_i = \sigma_i^2 \ .$$

Choose the integer r such that

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$$

(r turns out to be the rank of A). Then for $1 \leq i \leq r$, $Av_i = \sigma_i u_i$ for a unique $u_i \in \mathbb{C}^m$ with $\|u_i\| = 1$. Moreover, for $1 \leq i, j \leq r$,

$$u_i^H u_j = \frac{1}{\sigma_i \sigma_j} v_i^H A^H A v_j = \frac{1}{\sigma_i \sigma_j} e_i^H \wedge e_j = \delta_{ij}.$$

So we can append vectors $u_{r+1}, \ldots, u_m \in \mathbb{C}^m$ (if necessary) so that $U = [u_1 \cdots u_m] \in \mathbb{C}^{m \times m}$ is unitary. It follows easily that $AV = U\Sigma$, so $A = U\Sigma V^H$. □

The ideas in this second proof are derivable from the SVD of A,

$$A = U\Sigma V^H$$

(no matter how it is constructed). The key insite in this derivation of the SVD is the relation

$$AV = U\Sigma \ .$$

Interpreting this equation columnwise gives

$$Av_i = \sigma_i u_i \quad (1 \leq i \leq p),$$

and

$$Av_i = 0 \quad \text{for } i > m \text{ if } n > m,$$

where $\{v_1, \ldots, v_n\}$ are the columns of V and $\{u_1, \ldots, u_m\}$ are the columns of U. So A maps the orthonormal vectors $\{v_1, \ldots, v_p\}$ into the orthogonal directions $\{u_1, \ldots, u_p\}$ with the singular values $\sigma_1 \geq \cdots \geq \sigma_p$ as scale factors. (Of course if $\sigma_i = 0$ for an $i \leq p$, then $Av = 0$, and the direction of u is not represented in the range of A.)

The vectors $v_1, \ldots, v_n$ are called the *right singular vectors* of A, and $u_1, \ldots, u_m$ are called the *left singular vectors* of A. Observe that

$$A^H A = V\Sigma^H \Sigma V^H \quad \text{and} \quad \Sigma^H \Sigma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2) \in \mathbb{R}^{n \times n}$$

even if $m < n$. So

$$V^H A^H A V = \wedge = \operatorname{diag}(\lambda_1, \ldots \lambda_n),$$

and thus the columns of V form an orthonormal basis consisting of eigenvectors of $A^H A \in \mathbb{C}^{n \times n}$. Similarly $AA^H = U\Sigma\Sigma^H U^H$, so

$$U^H A A^H U = \Sigma\Sigma^H = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_p^2, \quad \overbrace{0, \ldots, 0}^{(m-n\mathrm{zeroesif}m>n)} \quad) \in \mathbb{R}^{m \times m},$$

and thus the columns of U form an orthonormal basis of $\mathbb{C}^m$ consisting of eigenvectors of $AA^H \in \mathbb{C}^{m \times m}$.

Caution. We cannot choose the bases of eigenvectors $\{v_1, \ldots, v_n\}$ of $A^H A$ (corresponding to $\lambda_1, \ldots, \lambda_n$) and $\{u_1, \ldots, u_m\}$ of AA^H (corresponding to $\lambda_1, \ldots, \lambda_p, [0, \ldots, 0]$) independently: we must have $Av_i = \sigma_i u_i$ for $\sigma_i > 0$.

In general, the SVD is not unique. Σ is uniquely determined but if $A^H A$ has multiple eigenvalues, then one has freedom in the choice of bases in the corresponding eigenspace, so V (and thus U) is not uniquely determined. One has complete freedom of choice of orthonormal bases of $\mathcal{N}(A^H A)$ and $\mathcal{N}(AA^H)$: these form the right-most columns of V and U, respectively. For a nonzero multiple singular value, one can choose the basis of the eigenspace of $A^H A$ (choosing columns of V), but then the corresponding columns of U are determined; or, one can choose the basis of the eigenspace of AA^H (choosing columns of U), but then the corresponding columns of V are determined. If all the singular values $\sigma_1, \ldots, \sigma_n$ of A are distinct, then each column of V is uniquely determined up to a factor of modulus 1, i.e., V is determined up to right multiplication by a diagonal matrix

$$D = \operatorname{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n}).$$

Such a change in V must be compensated for by multiplying the first n columns of U by D (the first $n-1$ cols. of U by $\operatorname{diag}(e^{i\theta_1}, \ldots, e^{i\theta_{n-1}})$ if $\sigma_n = 0$); of course if $m > n$, then the last $m-n$ columns of U have further freedom (they are in $\mathcal{N}(AA^H)$).

There is an abbreviated form of SVD useful in computation. Since rank is presumed under unitary multiplication, $\operatorname{rank}(A) = r$ iff $\sigma_1 \geq \cdots \geq \sigma_r > 0 = \sigma_{n+1} = \cdots$. Let $U_r \in \mathbb{C}^{m \times r}$, $V_r \in \mathbb{C}^{n \times r}$ be the first r columns of U, V, respectively, and let $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r \times r}$. Then $A = U_r \Sigma_r V_r^H$ (exercise).

Applications of SVD

If $m = n$, then $A \in \mathbb{C}^{n \times n}$ has eigenvalues as well as singular values. These can differ significantly. For example, if A is nilpotent, then all of its eigenvalues are 0. But all of the singular values of A vanish iff $A = 0$. However, for A normal, we have:

Proposition. Let $A \in \mathbb{C}^{n \times n}$ be normal, and order the eigenvalues of A as

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Then the singular values of A are $\sigma_i = |\lambda_i|$, $1 \leq i \leq n$.

Proof. By the Spectral Theorem for normal operators, there is a unitary $V \in \mathbb{C}^{n \times n}$ for which $A = V \wedge V^H$, where $\wedge = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. For $1 \leq i \leq n$, choose $d_i \in \mathbb{C}$ for which $\bar{d}_i \lambda_i = |\lambda_i|$ and $|d_i| = 1$, and let $D = \operatorname{diag}(d_1, \ldots, d_n)$. Then D is unitary, and

$$A = (VD)(D^H \wedge)V^H \equiv U \Sigma V^H,$$

where $U = VD$ is unitary and $\Sigma = D^H \wedge = \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_n|)$ is diagonal with decreasing nonnegative diagonal entries. □

Note that both the right and left singular vectors (columns of V, U) are eigenvectors of A; the columns of U have been scaled by the complex numbers d_i of modulus 1.

The Frobenius and Euclidean operator norms of $A \in \mathbb{C}^{m\times n}$ are easily expressed in terms of the singular values of A:

$$\|A\|_F = \left(\sum_{i=1}^{n} \sigma_i^2\right)^{\frac{1}{2}} \quad \text{and} \quad \|A\| = \sigma_1 = \sqrt{p(A^H A)},$$

as follows from the unitary invariance of these norms. There are no such simple expressions (in general) for these norms in terms of the eigenvalues of A if A is square (but not normal). Also, one cannot use the spectral radius $\rho(A)$ as a norm on $\mathbb{C}^{n\times n}$ because it is possible for $\rho(A) = 0$ and $A \neq 0$; however, on the *subspace* of $\mathbb{C}^{n\times n}$ consisting of the normal matrices, $\rho(A)$ is a norm since it agrees with the Euclidean operator norm for normal matrices.

The SVD is useful computationally for questions involving rank. The rank of $A \in \mathbb{C}^{m\times n}$ is the number of nonzero singular values of A since rank is invariant under pre- and post-multiplication by invertible matrices. There are stable numerical algorithms for computing SVD (try on `matlab`). In the presence of round-off error, row-reduction to echelon form usually fails to find the rank of A when its rank is $< \min(m, n)$; for such a matrix, the computed SVD has the zero singular values computed to be on the order of machine ϵ, and these are often identifiable as "numerical zeroes." For example, if the computed singular values of A are $10^2, 10, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-15}, 10^{-15}, 10^{-16}$ with machine $\epsilon \approx 10^{-16}$, one can safely expect $\operatorname{rank}(A) = 7$.

Another application of the SVD is a way to prove the polar form of a matrix. This is the analogue of the polar form $z = re^{i\theta}$ in $\mathbb{C}$. (Note from problem 1 on Prob. Set 6, $U \in \mathbb{C}^{n\times n}$ is unitary iff $U = e^{iH}$ for some Hermitian $H \in \mathbb{C}^{n\times n}$).

Polar Form

Every $A \in \mathbb{C}^{n\times n}$ may be written as $A = PU$, where P is positive semi-definite Hermitian and U is unitary.

Proof. Let $A = U\Sigma V^H$ be a SVD for A, and write

$$A = (U\Sigma U^H)(UV^H).$$

Then $U\Sigma U^H$ is positive semi-definite Hermitian and UV^H is unitary. □

Observe in the proof that the eigenvalues of P are the singular values of A; this is true for any polar decomposition of A (exercise). We note that in the polar form $A = PU$, P is always uniquely determined and U is uniquely determined if A is invertible (as in $z = re^{i\theta}$). The uniqueness of P follows from the following two facts:

(i) $AA^H = PUU^H P^H = P^2$ and

(ii) every positive semi-definite Hermitian matrix has a unique positive semi-definite Hermitian square root (see H-J, Theorem 7.2.6).

If A is invertible, then so is P, so $U = P^{-1}A$ is also uniquely determined. There is also a version of the polar form for non-square matrices; see H-J for details.

Linear Least Squares Problems

If $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$, the linear system $Ax = b$ might not be solvable. Instead, we can solve the minimization problem. Find $x \in \mathbb{C}^n$ to attain $\inf_{x\in\mathbb{C}^n} \|Ax - b\|^2$ (Euclidean norm). This is called a least-squares problem since the square of the Euclidean norm is a sum of squares. At a minimum of $\varphi(x) = \|Ax - b\|^2$ we must have $\nabla\varphi(x) = 0$, or equivalently

$$\varphi'(x; v) = 0 \quad \forall\, v \in \mathbb{C}^n,$$

where

$$\varphi'(x; v) = \frac{d}{dt}\varphi(x + tv)\bigg|_{t=0}$$

is the directional derivative. If $y(t)$ is a differentiable curve in $\mathbb{C}^m$, then

$$\frac{d}{dt}\|y(t)\|^2 = \langle y'(t), y(t)\rangle + \langle y(t), y'(t)\rangle = 2\mathcal{R}e\langle y(t), y'(t)\rangle.$$

Taking $y(t) = A(x + tv) - b$, we obtain that

$$\nabla\varphi(x) = 0 \Leftrightarrow (\forall\, v \in \mathbb{C}^n)\ 2\mathcal{R}e\langle Ax - b, Av\rangle = 0 \Leftrightarrow A^H(Ax - b) = 0,$$

i.e.,

$$A^H Ax \ = \ A^H b\ .$$

These are called the *normal equations* (they say $(Ax - b) \perp \mathcal{R}(A)$).

Linear Least Squares, SVD, and Moore-Penrose Pseudoinverse

The Projection Theorem (for finite dimensional S)

Let V be an inner product space, and let S be a finite dimensional subspace. Then

(1) $V = S \oplus S^\perp$, i.e., given $v \in V$, $\exists$ unique $\bar{y} \in S$ and $\bar{z} \in S^\perp$ for which

$$v = \bar{y} + \bar{z}$$

(so $\bar{y} = Pv$, where P is the orthogonal projection of V onto S; also $\bar{z} = (I - P)v$ and $I - P$ is the orthogonal projection of V onto $S^\perp$).

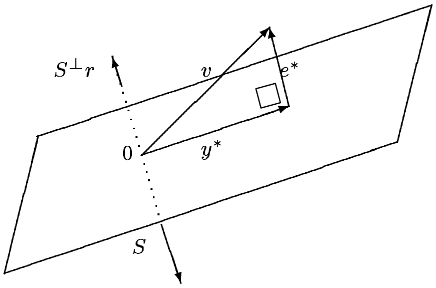
(2) Given $v \in V$, the $\bar{y}$ in (1) is the unique element of S which satisfies

$$(\forall\, y \in S)\ \langle v - \bar{y}, y\rangle = 0.$$

(3) Given $v \in V$, the $\bar{y}$ in (1) is the unique element of S solving the minimization problem

$$\underset{y\in S}{\text{minimize}}\ \|v - y\|^2\ .$$

Remark. The content of the Projection Theorem is contained in the following picture:



Proof. (1) Let $\{\psi_1, \ldots, \psi_r\}$ be an orthonormal basis of S. Given $v \in V$, let

$$\bar{y} = \sum_{j=1}^{r} \langle v, \psi_j \rangle \psi_j \quad \text{and} \quad \bar{z} = v - \bar{y}.$$

Then $v = \bar{y} + \bar{z}$ and $\bar{y} \in S$. For $1 \leq k \leq r$,

$$\langle \bar{z}, \psi_k \rangle = \langle v, \psi_k \rangle - \langle \bar{y}, \psi_k \rangle = \langle v, \psi_k \rangle - \langle v, \psi_k \rangle = 0,$$

so $\bar{z} \in S^\perp$. Uniqueness follows from the fact that $S \cap S^\perp = \{0\}$.
(2) Since $\bar{z} = v - \bar{y}$, this is just a restatement of $\bar{z} \in S^\perp$.
(3) For any $y \in S$,

$$v - y = \underbrace{\bar{y} - y}_{\in S} + \underbrace{\bar{z}}_{\in S^\perp},$$

so by the Pythagorean Theorem ($p \perp q \Rightarrow \|p \pm q\|^2 = \|p\|^2 + \|q\|^2$),

$$\|v - y\|^2 = \|\bar{y} - y\|^2 + \|\bar{z}\|^2.$$

Therefore, $\|v - y\|^2$ is minimized iff $y = \bar{y}$, and $\|v - \bar{y}\|^2 = \|\bar{z}\|^2$. □

The Projection Theorem as stated above is a special case of a much more general result that can be stated for closed convex sets on a Banach space. Below we give the Hilbert space version of this result.

The Projection Theorem for Convex Sets (on Hilbert Spaces)

Let X be a Hilbert space and let C be a closed convex subset of X. Then for each $x \in X$ there is a unique vector $y^0 \in C$ such that

$$\|x - y^0\| \leq \|x - y\| \qquad \forall\, y \in C.$$

Furthermore, a necessary and sufficient condition that y_0 be the unique minimizing vector is that

$$\mathrm{Re}(\langle x - y^0, y - y^0 \rangle) \leq 0 \qquad \forall\, y \subset C.$$

Proof. Let $\{y^i\} \subset C$ be such that

$$\|x - y^i\| \to \inf\{\|x - y\| : y \in C\} =: \delta.$$

By the parallelogram law

$$\|y^m - y^n\|^2 = 2\|x - y^m\|^2 + 2\|x - y^n\|^2 - 4\left\|x - \frac{y^n + y^m}{2}\right\|^2.$$

By convexity, $2^{-1}(y^n + y^m) \in C$; so

$$\|x - 2^{-1}(y^m + y^n)\| \geq \delta.$$

Therefore,

$$\|y^m - y^n\|^2 \leq 2\|y^m - x\|^2 + 2\|y^n - x\|^2 - 4\delta^2 \to 0.$$

Consequently, the sequence $\{y^n\}$ is Cauchy and so has a limit y^0 with $\|x - y^0\| = \delta$. The uniqueness follows by considering the sequence

$$y^{2n+1} = y^a \quad \text{and} \quad y^{2n} = y^b \quad n = 0, 1, \ldots,$$

where $y^a, y^b \in C$ with $\|h^a - x\| = \|y^b - x\| = \delta$. By applying the above argument, we find $y^a = y^b$.

We next show that y^0 is the unique vector satisfying

$$\text{Re}(\langle x - y^0, y - y^0\rangle) \leq 0 \text{ for all } y \in C.$$

Suppose to the contrary that there is a vector y^1 such that $\text{Re}(\langle x - y^0, y^1 - y^0\rangle) = \epsilon > 0$. Consider the vectors

$$y^\alpha = \alpha y^1 + (1 - \alpha)y^0 \in C \quad \text{for} \quad \alpha \in [0, 1].$$

Note that the function $\varphi : \mathbb{R} \to \mathbb{R}$ given by

$$\begin{aligned} \varphi(\alpha) &= \|x - y^\alpha\|^2 \\ &= (1 - \alpha)^2\|x - y^0\|^2 + 2\alpha(1 - \alpha)\text{Re}(\langle x - y^0, x - y^1\rangle) + \alpha^2\|x - y^1\|^2 \end{aligned}$$

is differentiable with

$$\begin{aligned} \varphi'(0) &= -2\|x - y^0\|^2 + 2\text{Re}(\langle x - y^0, x - y^1\rangle) \\ &= -2\text{Re}(\langle x - y^0, x - y^0\rangle + \langle x - y^0, y^1 - x\rangle) \\ &= -2\text{Re}\langle x - y^0, y^1 - y^0\rangle = -2\epsilon < 0. \end{aligned}$$

Hence, $\|x - y^\alpha\| < \|x - y^0\|$ for all $\alpha > 0$ sufficiently small. This contradiction implies that y^1 does not exist.

Conversely, suppose that $y^0 \in C$ is such that

$$\text{Re}(\langle x - y^0, y - y^0\rangle) \leq 0 \quad \forall\, y \in C.$$

Then for any $y \in C$ with $y \neq y^0$, we have

$$\begin{aligned} \|x - y\|^2 &= \|(x - y^0) + (y^0 - y)\|^2 \\ &= \|x - y^0\|^2 + 2\text{Re}(\langle x - y^0, y^0 - y\rangle) + \|y^0 - y\|^2 \\ &> \|x - y^0\|^2. \end{aligned}$$

□

Given $x \in X$, the unique solution to the problem

$$\min_{y \in C} \|x - y\|$$

is called the projection of x onto C and is often denoted $P_C(x)$. In the classical Projection Theorem the set C is a closed subspace S of X (recall that the subspace $S^\perp$ is always closed). In this case, the condition that

$$\mathrm{Re}(\langle x - y^0, y - y^0 \rangle) \leq 0 \qquad \forall\, y \subset S$$

is equivalent to the statement that $(x - P_C(x)) \in S^\perp$. Note that the Projection Theorem for Convex Sets extends the classical Projection Theorem to Hilbert spaces.

We leave application of the Projection Theorem for Convex Sets to later study. For the moment we focus on applications of the Projection Theorem in the Euclidean space setting (finite dimentional inner product spaces).

Theorem [Normal Equations for Linear Least Squares]
Let $A \in \mathbb{C}^{m \times n}$, $b \in \mathbb{C}^m$ and $\|\cdot\|$ be the Euclidean norm. Then $x \in \mathbb{C}^n$ solves

$$(*) \qquad \underset{x \in \mathbb{C}^n}{\text{minimize}} \ \|b - Ax\|^2$$

if and only if x is a solution to the normal questions $A^H Ax = A^H b$.

Proof. Let $\{a_1, \ldots, a_n\}$ be the columns of A, and let $S = \mathcal{R}(A) = \mathrm{span}\{a_1, \ldots, a_n\}$. Substituting $y \in S$ for Ax we arrive at the equivalent minimization problem

$$\underset{y \in S}{\text{minimize}} \ \|b - y\|^2 \ .$$

Apply the Projection Theorem to find that $y = \bar{y}$ iff $b - y \in S^\perp$, or equivalently, x solves (*) iff

$$b - Ax \in S^\perp = \mathcal{R}(A)^\perp = \mathcal{N}(A^T).$$

The condition that $b - Ax \in \mathcal{N}(A^T)$ is equivalent to the normal equations. □

Remarks.

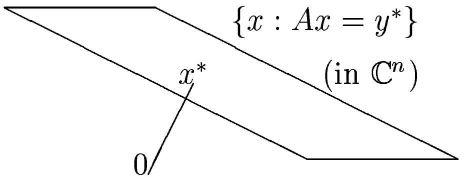
(i) The minimizing element $y = y^*$ of S is unique. Since $y^* \in \mathcal{R}(A)$, there exists $x \in \mathbb{C}^n$ for which $Ax = y^*$, or equivalently, there exists $x \in \mathbb{C}^n$ minimizing $\|b - Ax\|^2$. Consequently, there is an $x \in \mathbb{C}^n$ for which $A^H Ax = A^H b$, that is, the normal equations are consistent.

(ii) If rank$(A) = n$ (i.e. $\{a_1, \ldots, a_n\}$ are linearly independent in $\mathbb{C}^m$), then there is a unique $\bar{x} \in \mathbb{C}^n$ for which $A\bar{x} = \bar{y}$. This $\bar{x}$ is the unique minimizer of $\|b - Ax\|^2$ over $x \in \mathbb{C}^n$ as well as the unique solution of the normal equations $A^H Ax = A^H b$.

If $\operatorname{rank}(A) = r < n$, then the minimizing vector x is not unique; x can be modified by adding any element of $\mathcal{N}(A)$. (Exercise. Show $\mathcal{N}(A) = \mathcal{N}(A^H A)$.) However, there is a unique

$$\hat{x} \in \{\bar{x} \in \mathbb{C}^n : \|b - A\bar{x}\| \le \|b - Ax\| \;\forall\, x \in \mathbb{C}^n\} = \{x \in \mathbb{C}^n \;:\; A^H Ax = A^H b\}$$

of minimum norm.



Since $\{x \in \mathbb{C}^n \;:\; A^H Ax = A^H b\}$ is an affine translate of the subspace $\mathcal{N}(A^H A) = \mathcal{N}(A)$, a translated version of the Projection Theorem shows that there is a unique $\{x \in \mathbb{C}^n \;:\; A^H Ax = A^H b\}$ for which $\hat{x} \perp \mathcal{N}(A)$ and thus $\hat{x}$ is the unique element of $\{x \in \mathbb{C}^n \;:\; A^H Ax = A^H b\}$ of minimum norm.

In summary: given $b \in \mathbb{C}^m$, then $x \in \mathbb{C}^n$ minimizes $\|b - Ax\|^2$ over $x \in \mathbb{C}^n$ iff Ax is the orthogonal projection of b onto $\mathcal{R}(A)$, and among this set of solutions there is a unique such $\hat{x}$ of minimum norm; alternatively, $\hat{x}$ is the unique solution of the normal equations $A^H Ax = A^H b$ which also satisfies $\hat{x} \in \mathcal{N}(A)^\perp$.

The map $A^\dagger : \mathbb{C}^m \to \mathbb{C}^n$ ($A^\dagger$ is read "A dagger") which maps $b \in \mathbb{C}^m$ into the unique minimizer $\hat{x}$ of $\|b - Ax\|^2$ of minimum norm is called the Moore-Penrose pseudo-inverse of A. As we will see shortly, $A^\dagger$ is linear, so it is represented by an $n \times m$ matrix which we also denote by $A^\dagger$ (and we also call this matrix the Moore-Penrose pseudo-inverse of A). If $m = n$ and A is invertible, then every $b \in \mathbb{C}^n$ is in $\mathcal{R}(A)$, so $\bar{y} = b$, and the solution of $Ax = b$ is unique, given by $x = A^{-1}b$. In this case $A^\dagger = A^{-1}$. So the pseudo-inverse is a generalization of the inverse to possibly non-square, non-invertible matrices.

The pseudo-inverse of A can be expressed easily in terms of the abbreviated form of the SVD of A. Let $A = U\Sigma V^H$ be an SVD of A, let $r = \operatorname{rank}(A)$ (so

$$\sigma_1 \ge \cdots \ge r_r > 0 = \sigma_{r+1} = \cdots),$$

let U_r and V_r in $\mathbb{C}^{m\times r}$ be the first r columns of U, V, respectively, and let

$$\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{C}^{r\times r}.$$

Then as we have seen above, $A = U_r \Sigma_r V_r^H$. Let $\widetilde{U} \in \mathbb{C}^{m\times(m-r)}$, $\widetilde{V} = \mathbb{C}^{n\times(n-r)}$ be the remaining columns of U, V, respectively, so

$$U = [U_r \widetilde{U}] \quad \text{and} \quad V = [V_r, \widetilde{V}].$$

Note that

$$\operatorname{span}\{\text{cols. of } U_r\} = \mathcal{R}(U_r) = \mathcal{R}(A),\; \mathcal{R}(\widetilde{U}) = \mathcal{R}(A)^\perp,\; \mathcal{R}(\widetilde{V}) = \mathcal{N}(A),$$

and

$$\mathcal{R}(V_r) = \mathcal{N}(A)^{\perp}.$$

Now

$$\begin{aligned} \|b - Ax\|^2 &= \|b - U\Sigma V^H x\|^2 \\ &= \|U^H b - \Sigma V^H x\|^2 \\ &= \left\| \begin{bmatrix} U_r^H \\ \widetilde{U}^H \end{bmatrix} b - \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^H \\ \widetilde{V}^H \end{bmatrix} x \right\|^2 \\ &= \left\| \begin{bmatrix} U_r^H b - \Sigma_r V_r^H x \\ \widetilde{U}^H b \end{bmatrix} \right\|^2, \end{aligned}$$

so

$$\|b - Ax\|^2 = \|U_r^H b - \Sigma_r V_r^H x\|^2 + \|\widetilde{U}_b^H\|^2.$$

Thus

$$\left[x \text{ solves } \underset{x \in \mathbb{C}^n}{\text{minimize}} \ \|b - Ax\|^2 \right] \Leftrightarrow \Sigma_r V_r^H x = U_r^H b \ \Leftrightarrow \ V_r^H x = \Sigma_r^{-1} U_r^H b,$$

and, in addition,

$$x = \hat{x} = \text{ the unique minimizer of } \|b - Ax\|^2 \text{ of minimum norm}$$

if and only if

$$x \in \mathcal{N}(A)^{\perp} = \mathcal{R}(V_r),$$

i.e., $\widetilde{V}^H x = 0$. So $x = \hat{x}$ if and only if

$$V^H x = \begin{bmatrix} V_r^H x \\ \widetilde{V}^H x \end{bmatrix} = \begin{bmatrix} \Sigma_r^{-1} U_r^H b \\ 0 \end{bmatrix}$$

$$\Longleftrightarrow$$

$$x = V \begin{bmatrix} \Sigma_r^{-1} U_r^H b \\ 0 \end{bmatrix}$$

$$\Longleftrightarrow$$

$$x = [V_r \widetilde{V}] \begin{bmatrix} \Sigma_r^{-1} U_r^H b \\ 0 \end{bmatrix} = V_r \Sigma_r^{-1} U_r^H b \ .$$

So

$$\hat{x} = V_r \Sigma_r^{-1} U_r^H b \ .$$

We conclude that $\hat{x}$ is a linear function of b, so $A^{\dagger}$ is linear, and the matrix for $A^{\dagger}$ is

$$A^{\dagger} = V_r \Sigma_r^{-1} U_r^H = [V_r \widetilde{V}] \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^H \\ \widetilde{U}^H \end{bmatrix} = V \Sigma^{\dagger} U^H,$$

where

$$\Sigma^{\dagger} = \operatorname{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0) \in \mathbb{C}^{n \times m}.$$

It is appropriate to call this matrix $\Sigma^\dagger$ as it is easily shown that the pseudo-inverse of $\Sigma \in \mathbb{C}^{m\times n}$ is this matrix (exercise).

One rarely actually computes $A^\dagger$. Instead, to minimize $\|b - Ax\|^2$ using the SVD of A one computes

$$\hat{x} \;=\; V_r(\Sigma_r^{-1}(U_r^H b)) \;.$$

For $b \in \mathbb{C}^m$, we saw above that if $\hat{x} = A^\dagger b$, then $A\hat{x} = \bar{y}$ is the orthogonal projection of b onto $\mathcal{R}(A)$. Thus $AA^\dagger$ is the orthogonal projection of $\mathbb{C}^m$ onto $\mathcal{R}(A)$. This is also clear directly from the SVD:

$$AA^\dagger = U_r\Sigma_r V_r^H V_r \Sigma_r^{-1} U_r^H = U_r \Sigma_r \Sigma_r^{-1} U_r^H = U_r U_r^H = \Sigma_{j=1}^r u_j u_j^H$$

which is clearly the orthogonal projection onto $\mathcal{R}(A)$. Note that $V_r^H V_r = I_r$ since the columns of V are orthonormal. Similarly, since $w = A^\dagger(Ax)$ is the vector of least length satisfying $Aw = Ax$, $A^\dagger A$ is the orthogonal projection of $\mathbb{C}^n$ onto $\mathcal{N}(A)^\perp$. Again, this also is clear directly from the SVD:

$$A^\dagger A = V_r\Sigma_r^{-1} U_r^H U_r \Sigma_r V_r^H = V_r V_r^H = \Sigma_{j=1}^r v_j v_j^H,$$

is the orthogonal projection onto $\mathcal{R}(V_r) = \mathcal{N}(A)^\perp$. These relationships are substitutes for $AA^{-1} = A^{-1}A = I$ for invertible $A \in \mathbb{C}^{n\times n}$. Similarly, one sees that

(i) $AXA = A$,

(ii) $XAX = X$,

(iii) $(AX)^H = AX$,

(iv) $(XA)^H = XA$,

where $X = A^\dagger$. In fact, one can show that $X \in \mathbb{C}^{n\times m}$ is $A^\dagger$ if and only if X satisfies (i), (ii), (iii), (iv). (Exercise — see section 5.54 in Golub and Van Loan.)

The pseudo inverse can be used to extend the (Euclidean operator norm) condition number to general matrices: $\kappa(A) = \|A\| \cdot \|A^\dagger\| = \sigma_1/\sigma_r$ (where $r = \operatorname{rank} A$).

LU Factorization

All of the matrix factorizations we have studied so far are spectral factorizations in the sense that in obtaining these factorizations, one is obtaining the eigenvalues and eigenvectors of A (or matrices related to A, like A^HA and AA^H for SVD). We end our discussion of matrix factorizations by mentioning two non-spectral factorizations. These non-spectral factorizations can be determined directly from the entries of the matrix, and are computationally less expensive than spectral factorizations. Each of these factorizations amounts to a reformulation of a procedure you are already familiar with. The LU factorization is a reformulation of Gaussian Elimination, and the QR factorization is a reformulation of Gran-Schmitt orthogonalization.

Recall the method of Gaussian Elimination for solving a system $Ax = b$ of linear equations, where $A \in \mathbb{C}^{n\times n}$ is invertible and $b \in \mathbb{C}^n$. If the coefficient of x_1 in the first equation is nonzero, one eliminates all occurrences of x_1 from all the other equations by adding appropriate multiples of the first equation. This operations does not change the set of solutions to the equation. Now if the coefficient of x_2 in the new second equation is nonzero, it can be used to eliminate x_2 from the further equations, etc... In matrix terms, if

$$A = \begin{bmatrix} a & v^T \\ u & \widetilde{A} \end{bmatrix} \in \mathbb{C}^{n\times m},$$

with $a \neq 0$, $a \in \mathbb{C}$, $u, v \in \mathbb{C}^{n-1}$, and $\widetilde{A} \in \mathbb{C}^{(n-1)\times(n-1)}$, then using the first row to zero out u amounts to left multiplication of the matrix A by the matrix

$$\begin{bmatrix} 1 & 0 \\ -\frac{u}{a} & I \end{bmatrix}$$

to get

$$(*) \qquad \begin{bmatrix} 1 & 0 \\ -\frac{u}{a} & I \end{bmatrix} \begin{bmatrix} a & v^T \\ u & \widetilde{A} \end{bmatrix} \in \mathbb{C}^{n\times m} = \begin{bmatrix} a & v^T \\ 0 & A_1 \end{bmatrix} .$$

Define

$$L_1 = \begin{bmatrix} 1 & 0 \\ \frac{u}{a} & I \end{bmatrix} \in \mathbb{C}^{n\times n} \quad \text{and} \quad U_1 = \begin{bmatrix} a & v^T \\ 0 & A_1 \end{bmatrix} .$$

and observe that

$$L_1^{-1} = \begin{bmatrix} 1 & 0 \\ -\frac{u}{a} & I \end{bmatrix} .$$

Hence (*) becomes

$$L_1^{-1}A = U_1, \text{ or equivalently,} \quad A = L_1 U_1 .$$

Note that L_1 is lower triangular and U_1 is block upper-triangular with one 1×1 block and one $(n-1) \times (n-1)$ block on the block diagonal. The elements of $\frac{u}{a} \in \mathbb{C}^{n-1}$ are called *multipliers*, they are the multiples of the first row subtracted from subsequent rows, and they are computed in the Gaussian Elimination algorithm. The multipliers are usually denoted

$$u/a = \begin{bmatrix} m_{21} \\ m_{31} \\ \vdots \\ m_{n1} \end{bmatrix} .$$

Now, if the $(1,1)$ entry of A_1 is not 0, we can apply the same procedure to A_1: if $A_1 = \begin{bmatrix} a_1 & v_1^T \\ u_1 & \widetilde{A}-1 \end{bmatrix} \in \mathbb{C}^{(n-1)\times(n-1)}$ with $a_1 \neq 0$, letting $\widetilde{L}_2 = \begin{bmatrix} I & 0 \\ \frac{u_1}{a_1} & I \end{bmatrix} \in \mathbb{C}^{(n-1)\times(n-1)}$, and forming $\widetilde{L}_2^{-1}A_1 = \begin{bmatrix} 1 & 0 \\ -\frac{u_1}{a_1} & I \end{bmatrix} \begin{bmatrix} a_1 & v_1^T \\ u_1 & \widetilde{A}-1 \end{bmatrix} = \begin{bmatrix} a_1 & v_1^T \\ 0 & A_2 \end{bmatrix} \equiv \widetilde{U}_2 \in \mathbb{C}^{(n-1)\times(n-1)}$ (where $A - 2 \in \mathbb{C}^{(n-2)\times(n-2)}$) amounts to using the second row to zero out elements of the second column below the diagonal: setting $L_2 = \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{L}_2 \end{bmatrix}$ and $u_2 = \begin{bmatrix} a & v^T \\ 0 & \widetilde{u}_2 \end{bmatrix}$, we have $L_2^{-1}L_1^{-1}A =$

$\begin{bmatrix} 1 & 0 \\ 0 & \widetilde{L}_2^{-1} \end{bmatrix} \begin{bmatrix} a & v^T \\ 0 & A_1 \end{bmatrix} = U_2$, which is block upper triangular with two $|x|$ blocks and one $(n-2) \times (n-2)$ block on the block diagonal. The elements of $\frac{u_1}{a_1}$ are multipliers, usually denoted $\frac{u_1}{a_1} = \begin{bmatrix} m_{32} \\ m_{52} \\ \vdots \\ m_{n2} \end{bmatrix}$. Notice that these multipliers appear in L_2 in the *second* column, below the diagonal. Continuing in a similar fashion, $L_{n-1}^{-1} \cdots L_2^{-1} L_1^{-1} A = U_{n-1} \equiv U$ is upper triangular (provided along the way that the $(1,1)$ entries of $A, A_1, A_2, \ldots, A_{n-2}$ are nonzero so the process can continue). Define $L = (L_{n-1}^{-1} \cdots L_1^{-1})^{-1} = L_1 L_2 \cdots L_{n-1}$. Then $A = LU$. (Remark: A lower triangular matrix with 1's on the diagonal is called a *unit* lower triangular matrix, so $L_j, L_j^{-1}, L_{j-1}^{-1} \cdots L_1^{-1}, L_1 \cdots L_j, L^{-1}, L$ are all unit lower triangular.) For an invertible $A \in \mathbb{C}^{n \times n}$, writing $A = LU$ as a product of a unit lower triangular matrix L and a (necessarily invertible) upper triangular matrix U (both in $\mathbb{C}^{n \times n}$) is called the *LU factorization* of A.

Remarks.

(1) If $A \in \mathbb{C}^{n \times n}$ is invertible and has an LU factorization, it is unique (exercise).

(2) One can show that $A \in \mathbb{C}^{n \times n}$ has an LU factorization iff for $1 \leq j \leq n$, the upper left $j \times j$ principal submatrix $\begin{bmatrix} a_{11} & \cdots & a_{ij} \\ \vdots & & \\ a_{j1} & \cdots & a_{jj} \end{bmatrix}$ is invertible.

(3) Not every invertible $A \in \mathbb{C}^{n \times n}$ has an LU-factorization. (Example: $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ doesn't.) Typically, one must permute the rows of A to move nonzero entries to the appropriate spot for the elimination to proceed. Recall that a permutation matrix $P \in \mathbb{C}^{n \times n}$ is the identity I with its rows (or columns) permuted: so $P \in \mathbb{R}^{n \times n}$ is orthogonal, and $P^{-1} = P^T$. Permuting the rows of A amounts to left multiplication by a permutation matrix P^T; then $P^T A$ has an LU factorization, so $A = PLU$ (called the PLU factorization of A).

(4) Fact: Every invertible $A \in \mathbb{C}^{n \times n}$ has a (not necessarily unique) PLU factorization.

(5) It turns out that $L = L_1 \cdots L_{n-1} = \begin{bmatrix} 1 & \ddots & \\ m_{21} & \ddots & \\ \vdots & & \\ m_{n-1} & \cdots & 1 \end{bmatrix}$ has the multipliers m_{ij} below the diagonal.

(6) The LU factorization can be used to solve linear systems $Ax = b$ (where $A = LU \in \mathbb{C}^{n \times n}$ is invertible). The system $Ly = b$ can be solved by forward substitution (1^{st} eqn. gives x_1, etc.), and $Ux = y$ can be solved by back-substitution (n^{th} eqn. gives x_n, etc.), giving solu. of $Ax = LUx = b$. See section 3.5 of H-J.

QR Factorization

Recall first the Gram-Schmidt orthogonalization process. Let V be an inner product space, and suppose $a_1, \ldots, a_n \in V$ are linearly independent. Define $q_1, \ldots, q_n$ inductively, as follows: let $p_1 = a_1$ and $q_1 = p_1/\|p_1\|$: for $2 \leq j \leq n$, let $p_j = a_j - \sum_{i=1}^{j-1}\langle a_j, q_j\rangle q_i$ and $q_j = p_j/\|p_j\|$. Since clearly for $1 \leq k \leq n$ $q_k \in \text{span}\{a_1, \ldots, a_k\}$, each p_j is nonzero by the linear independence of $\{a_1, \ldots, a_n\}$, so each q_j is well-defined. It is easily seen that $\{q_1, \ldots, q_n\}$ is an orthonormal basis for span$\{a_1, \ldots, a_n\}$. Note also that for $1 \leq k \leq n$ $a_k \in \text{span}\{q_1, \ldots, q_k\}$ (and thus $\{q_1, \ldots, q_k\}$ is an orthonormal basis of span$\{a_1, \ldots, a_k\}$: defining $r_{jj} = \|p\|_j$ (so $p_j = r_{jj}q_j$) and $r_{ij} = \langle a_j, q_i\rangle$ for $1 \leq i < j \leq n$, we have: $a_1 = r_{11}q_1$, $a_2 = r_{12}q_1 + r_{22}q_2$, in general $a_j = \sum_{i=1}^{j} r_{ij}q_i$.

Remarks.

(1) If $a_1, a_2, \cdots$ is a linearly independent sequence in V, we can apply the Gram-Schmidt process to obtain an orthonormal sequence $q_1, q_2, \ldots$ with the property that for $k \geq 1$, $\{q_1, \ldots, q_k\}$ is an orthonormal basis for span$\{a_1, \ldots, a_k\}$.

(2) If the a_j's are linearly dependent, then for some value(s) of k, $a_k \in \text{span}\{a_1, \ldots, a_{k-1}\}$, and then $p_k = 0$. The process can be modified by setting $q_k = 0$ and proceeding. We end up with orthogonal q_j's, some of which have $\|q_j\| = 1$ and some have $\|q_j\| = 0$. Then for $k \geq 1$, the nonzero vectors in the set $\{q_1, \ldots, q_k\}$ form an orthonormal basis for span$\{a_1, \ldots, a_k\}$.

(3) The classical Gram-Schmidt algorithm described above applied to n linearly independent vectors $a_1, \ldots, a_n \in \mathbb{C}^m$ (where of course $m \geq n$) does no behave well computationally. Due to the accumulation of round-off error, the computed q_j's are not as orthogonal as one would want (or need in applications): $\langle q_j, q_k\rangle$ is small for $j \neq k$ with j near k, but not so small for $j \ll k$ or $j \gg k$. An alternate version, "Modified Gram-Schmidt," is equivalent in exact arithmetic, but behaves better numerically. In the following "pseudo-codes," p denotes a temporary storage vector used to accumulate the sums defining the p_j's.

Classic Gram-Schmidt

```
For  $j = 1, \cdots, n$ do
    $p := a_j$
    For $i = 1, \ldots, j-1$ do
        $r_{ij} = \langle a_j, q_i\rangle$
        $p := p - r_{ij}q_i$
    $r_{jj} := \|p\|$
    $q_j := p/r_{jj}$
```

Modified Gram-Schmidt

```
For  $j = 1, \ldots, n$ do
    $p := a_j$
    For $i = 1, \ldots, j-1$ do
        $r_{ij} = \langle p, q_i\rangle$
        $p := p - r_{ij}q_i$
    $r_{jj} = \|p\|$
    $q_j := p/r_{jj}$
```

The only difference is in the computation of r_{ij}: in Modified Gram-Schmidt, we orthogonalize the accumulated partial sum for p_j against each q_i successively.

Proposition. Suppose $A \in \mathbb{C}^{m\times n}$ with $m \geq n$. Then $\exists\, Q \in \mathbb{C}^{m\times m}$ which is unitary and an upper triangular $R \in \mathbb{C}^{m\times n}$ (i.e. $r_{ij} = 0$ for $i > j$) for which $A = QR$. If $\widetilde{Q} \in \mathbb{C}^{m\times n}$ denotes the first n columns of Q and $\widetilde{R} \in \mathbb{C}^{n\times n}$ denotes the first n rows of R, then clearly also $A = QR = [\widetilde{Q} *] \begin{bmatrix} \widetilde{R} \\ 0 \end{bmatrix} = \widetilde{Q}\widetilde{R}$. Moreover

(a) We may choose an R 2with nonnegative diagonal entries.

(b) If A is of full rank (i.e. rank$\,(A) = n$, or the cols. of A are linearly independent), then we may choose an R with positive diagonal entries, in which case the condensed factorization $A = \widetilde{Q}\widetilde{R}$ is unique (and thus in this case if $m = n$, the factorization $A = QR$ is unique since then $Q = \widetilde{Q}$ and $R = \widetilde{R}$).

(c) If A is of full rank, the condensed factorization $A = \widetilde{Q}\widetilde{R}$ is essentially unique: if $A = \widetilde{Q}_1\widetilde{R}_1 = \widetilde{Q}_2\widetilde{R}_2$, then $\exists$ a unitary diagonal matrix $D \in \mathbb{C}^{n\times n}$ for which $\widetilde{Q}_2 = \widetilde{Q}_1 D^H$ (rescaling cols. of $\widetilde{Q}_1$) and $\widetilde{R}_2 = D\widetilde{R}_1$ (rescaling rows of $\widetilde{R}_1$).

Proof. If the columns of A are linearly independent, we can apply the Gram-Schmidt process described above. Let $\widetilde{Q} = [q_1, \ldots, q_n] \in \mathbb{C}^{m\times n}$, and define $\widetilde{R} \in \mathbb{C}^{n\times n}$ by setting $r_{ij} - 0$ for $i > j$, and r_{ij} to be the value computed in J-S for $i \leq j$. Then $A = \widetilde{Q}\widetilde{R}$. Extending $\{q_1, \ldots, q_n\}$ to an orthonormal basis $\{q_1, \ldots, q_m\}$ of $\mathbb{C}^m$, and setting $Q = [q_1, \ldots, q_m]$ and $R = \begin{bmatrix} \widetilde{R} \\ 0 \end{bmatrix} \in \mathbb{C}^{m\times n}$, we have $A = QR$. As $r_{jj} > 0$ in J-S, we have (b): uniqueness follows by induction passing through the J-S process again, noting that at each step we have no choice. (c) follows easily from (b) since if rank$\,(A) = n$, then rank$\,(\widetilde{R}) = n$ in any $\widetilde{Q}\widetilde{R}$ factorization of A. If the columns of A are linearly dependent, we alter the Gram-Schmidt algorithm as in Remark (2) above. Notice that $q_k = 0$ iff $r_{kj} = 0\,\forall j$, so if $\{q_{k_1}, \ldots, q_{k_r}\}$ are the nonzero vectors in $\{q_1, \ldots, q_n\}$ (where of course $r =$ rank$\,(A)$), then the nonzero rows in R are precisely rows $k_1, \ldots, k_r$. So if we define $\widehat{Q} = [q_{k_1} - q_{k_r}] \in \mathbb{C}^{m\times r}$ and $\widehat{R} \in \mathbb{C}^{r\times n}$ to be these nonzero rows, then $\widehat{Q}\widehat{R} = A$ where $\widehat{Q}$ has orthonormal columns and $\widehat{R}$ is upper triangular. Let Q be a unitary matrix whose first r columns are $\widehat{Q}$, and let $R = \begin{bmatrix} \widehat{R} \\ 0 \end{bmatrix} \in \mathbb{C}^{m\times m}$. Then $A = QR$. (Notice that in addition to (a), we actually have constructed an R for which, in each nonzero row, the first nonzero element is positive.) □

Remarks.

(4) If $A \in \mathbb{R}^{m\times n}$, everything can be done in real arithmetic, so, e.g., $Q \in \mathbb{R}^{m\times m}$ is orthogonal and $R \in \mathbb{R}^{m\times n}$ is real, upper triangular.

(5) In practice, there are more efficient and better computationally behaved ways of calculating the Q and R factors. The idea is to create zeros below the diagonal (successively in columns $1, 2, \ldots$) as in Gaussian Elimination, except we now use Householder transformations (which are unitary) instead of the unit lower triangular matrices L_j. Details will be described in an upcoming problem set.

Using QR Factorization to Solve Least Squares Problems

Suppose $A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$, and $m \geq n$. Assume A has full rank ($\operatorname{rank}(A) = n$). To solve the least squares problem $\min \|b - Ax\|^2$ (which has a unique solution in this case). Let $A = QR$ be a QR factorization of A, with condensed form $\widetilde{Q}\widetilde{R}$, and write $Q = [\widetilde{Q}\widetilde{Q}]$ where $\widetilde{\widetilde{Q}} \in \mathbb{C}^{m\times(n-n)}$. Then $\|b - Ax\|^2 = \|b - QRx\|^2 = \|Q^H b - Rx\|^2 = \left\| \begin{bmatrix} \widetilde{Q}^H \\ \widetilde{\widetilde{Q}}^H \end{bmatrix} b - \begin{bmatrix} \widetilde{R} \\ 0 \end{bmatrix} x \right\|^2 =$ $\left\| \begin{bmatrix} \widetilde{Q}^H b - \widetilde{R}x \\ \widetilde{\widetilde{Q}}^H b \end{bmatrix} \right\|^2 = \|\widetilde{Q}^H b - \widetilde{R}x\|^2 + \|\widetilde{\widetilde{Q}}^H b\|^2$. Here $\widetilde{R} \in \mathbb{C}^{n\times n}$ is an invertible upper triangle matrix, so x minimizes $\|b - Ax\|^2$ iff $\widetilde{R}x = \widetilde{Q}^H b$. This invertible upper triangular $n \times n$ system for x can be solved by back-substitution. Note that we only need $\widetilde{Q}$ and $\widetilde{R}$ to solve for x.

The QR Algorithm

The QR algorithm is used to compute a specific Schur unitary triangularization of a matrix $A \in \mathbb{C}^{n\times n}$. The algorithm is *iterative*: We generate a sequence $A = A_0, A_1, A_2, \ldots$ of matrices which are unitarily similar to A; the goal is to get the subdiagonal elements to converge to zero, as then the eigenvalues will appear on the diagonal. If A is Hermitian, then so also are $A_1, A_2, \ldots$, so if the subdiagonal elements $\to 0$, also the superdiagonal elements $\to 0$, and (in the limit) we have diagonalized A. The QR algorithm is the most commonly used method for computing all the eigenvalues (and eigenvectors if wanted) of a matrix. It behaves well numerically since all the similarity transformations are unitary.

When used in practice, a matrix is first reduced to *upper-Hessenberg form* $\begin{bmatrix} x & \cdots & x \\ & \ddots & \vdots \\ x & \ddots & \\ 0 & x & x \end{bmatrix}$ ($h_{ij} = 0$ for $i > j+1$) using unitary similarity transformations built from Householder reflections (or Givens rotations), quite analogous to computing a QR factorization. Here, however, similarity transformations are being performed, so they require left and right multiplication by the Householder transformations — leading to an inability to zero out the first subdiagonal ($i = j+1$) in the process. If A is Hermitian ad upper-Hessenberg, A is tridiagonal. This initial reduction is to decrease the computational cost of the iterations in the QR algorithm. It is successful because upper-Hessenberg form is preserved by the iterations: if A_k is upper Hessenberg, so is A_{k+1}.

There are many sophisticated variants of the QR algorithm (shifts to speed up convergence, implicit shifts to allow computing a real quasi-upper triangular matrix similar to a real matrix using only real arithmetic, etc.). We consider the basic algorithm over $\mathbb{C}$.

The (Basic) QR Algorithm

Given $A \in \mathbb{C}^{n\times n}$, let $A_0 = A$. For $k = 0, 1, 2, \ldots$, starting with A_k, do a QR factorization of A_k. $A_k = Q_k R_k$, and define $A_{k+1} = R_k Q_k$.

Remark. $R_k = Q_k^H A_k$, so $A_{k+1} = Q_k^H A_k Q_k$ is unitarily similar to A_k. The algorithm uses the Q of the QR factorization of A_k to perform the next unitary similarity transformation.

Convergence of the QR Algorithm

We will show under mild hypotheses that all of the subdiagonal elements of A_k converge to 0 as $k \to \infty$. See section 2.6 in H-J for examples where the QR algorithm does not converge. See also sections 7.5, 7.6, 8.2 in Golub and Van Koen for more discussion.

Lemma. Let Q_j $(j = 1, 2, \ldots)$ be a sequence of unitary matrices in $\mathbb{C}^{n\times n}$ and R_j $(j = 1, 2, \ldots)$ be a sequence of upper triangular matrices in $\mathbb{C}^{n\times n}$ with positive diagonal entries. Suppose $Q_j R_j \to I$ as $j \to \infty$. Then $Q_j \to I$ and $R_j \to I$.

Proof Sketch. Let Q_{j_k} be any subsequence of Q_j. Since the set of unitary matrices in $\mathbb{C}^{n\times n}$ is compact, $\exists$ a sub-subsequence $Q_{j_{k_l}}$ and a unitary $Q \ni Q_{j_{k_l}} \to Q$. So $R_{j_{k_l}} = Q_{j_{k_l}}^H Q_{j_{k_l}} R_{j_{k_l}} \to Q^H \cdot I = Q^H$. So Q^H is unitary, upper triangular, with nonnegative diagonal elements, which implies easily that $Q^H = I$. Thus every subsequence of Q_j has in turn a sub-subsequence converging to I. By standard metric space theory, $Q_j \to I$, and thus $R_j = Q_j^H Q_j R_j \to I \cdot I = I$. □

Theorem. *Suppose $A \in \mathbb{C}^{n\times n}$ has eigenvalues $\lambda_1, \ldots, \lambda_n$ with $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0$. Choose $X \in \mathbb{C}^{n\times n} \ni X^{-1}AX = \wedge \equiv \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$, and suppose X^{-1} has an LU decomposition. Generate the sequence $A_0 = A, A_1, A_2, \ldots$ using the QR algorithm. Then the subdiagonal entries of $A_k \to 0$ as $k \to \infty$, and for $1 \le j \le n$, the j^{th} diagonal entry $\to \lambda_j$.*

Proof. Define $\widetilde{Q}_k = Q_0 Q_1 \cdots Q_k$ and $\widetilde{R}_k = R_k \cdots R_0$. Then $A_{k+1} = \widetilde{Q}_k^H A \widetilde{Q}_k$.

Claim: $\widetilde{Q}_k \widetilde{R}_k = A^{k+1}$ [Proof: Clear for $k = 0$¿ Suppose $\widetilde{Q}_{k-1}\widetilde{R}_{k-1} = A^k$. Then $R_k = A_{k+1} Q_k^H = \widetilde{Q}_k^H A \widetilde{Q}_k Q_k^H = \widetilde{Q}_k^H A \widetilde{Q}_{k-1}$, so $\widetilde{R}_k = R_k \widetilde{R}_{k-1} = \widetilde{Q}_k^H A \widetilde{Q}_{k-1} \widetilde{R}_{k-1} = \widetilde{Q}_k^H A^{k+1}$, so $\widetilde{Q}_k \widetilde{R}_k = A^{k+1}$.] Now, choose a QR factorization of X and an LU factorization of X^{-1}: $X = QR$, $X^{-1} = LU$ (Q unitary, L unit lower triang., R and U upper triangular with nonzero diagonal entries). Then $A^{k+1} = X \wedge^{k+1} X^{-1} = QR \wedge^{k+1} LU = QR(\wedge^{k+1} L \wedge^{-(k+1)}) \wedge^{k+1} U$. Let $E_{k+1} = \wedge^{k+1} L \wedge^{-(k+1)} - I$ and $F_{k+1} = RE_{k+1}R^{-1}$. *Claim:* $E_{k+1} \to 0$ (and this $F_{k+1} \to 0$) as $k \to \infty$. [Proof: Let ℓ_{ij} denote the elements of L. E_{k+1} is strictly lower triangular, and for $i > j$ its ij element is $\left(\frac{\lambda_i}{\lambda_j}\right)^{k+1} \ell_{ij} \to 0$ as $k \to \infty$ since $|\lambda_i| < |\lambda_j|$.] Now $A^{k+1} = QR(I + E_{k+1}) \wedge^{k+1} U$, so $A^{k+1} = Q(I + F_{k+1})R \wedge^{k+1} U$. Choose a QR factorization of $I + F_{k+1}$ (which is invertible) $I + F_{k+1} = \widehat{Q}_{k+1}\widehat{R}_{k+1}$ where $\widehat{R}_{k+1}$ has positive diagonal entries. By the Lemma, $\widehat{Q}_{k+1} \to I$ and $\widehat{R}_{k+1} \to I$. Since $A^{k+1} = (Q\widehat{Q}_{k+1})(\widehat{R}_{k+1} R \wedge^{k+1} U)$ and $A^{k+1} = \widetilde{Q}_k \widetilde{R}_k$, the essential uniqueness of QR factorizations of invertible matrices implies $\exists$ a unitary diagonal matrix D_k for which $Q\widehat{Q}_{k+1} D_k^H = \widetilde{Q}_k$ and$D_k \widehat{R}_{k+1} \wedge^{k+1} U = \widetilde{R}_k$. So $\widetilde{Q}_k D_k = Q\widehat{Q}_{k+1} \to Q$, and thus $D_k^H A_{k+1} D_k = D_k^H \widetilde{Q}_k^H A \widetilde{Q}_k D_k \to Q^H A Q$ as $k \to \infty$¿ But $Q^H A Q = Q^H (QR \wedge X^{-1}) QRR^{-1} = R \wedge R^{-1}$ is upper triangular with diagonal entries $\lambda_1, \ldots, \lambda_n$ in that order. Since D_k is unitary and diagonal, the lower triangular part of $R \wedge R^{-1}$ and of $D_k R \wedge R^{-1} D_k^H$ are the same, namely $\begin{bmatrix} \lambda_1 & \\ & \ddots \\ 0 & \lambda_n \end{bmatrix}$, and $\|A_{k+1} - D_k R \wedge R^{-1} D_k^H\| =$

$\|D_k^H A_{k+1} D_k - R \wedge R^{-1}\| \to 0$. The Theorem follows. □

(Note that the proof shows that $\exists$ a sequence $\{D_k\}$ of unitary diagonal matrices for which $D_k^H A_{k+1} D_k \to R \wedge R^{-1}$. So although the superdiagonal $(i < j)$ elements of A_{k+1} may not converge, the magnitude of each superdiagonal element converges.)

As a partial explanation for why the QR algorithm works, we show how the convergence of the first column of A_k to $\begin{bmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ follows from the power method. (See problem 6 on Prob. Set 5.) Suppose $A \in \mathbb{C}^{n\times n}$ is diagonalizable and has a unique eigenvalue λ_1 of maximum modules, and suppose for simplicity that $\lambda_1 > 0$. Then if $x \in \mathbb{C}^n$ has nonzero component in the direction of the eigenvector corresponding to λ_1 when expanded in terms of the eigenvectors of A, it follows that the sequence $A^k x/\|A^k x\|$ converges to a unit eigenvector corresponding to λ_1. The condition in the Theorem above that X^{-1} has an LU factorization implies that the $(1,1)$ entry of X^{-1} is nonzero, so when e_1 is expanded in terms of the eigenvectors $x_1, \ldots, x_n$ (cols. of X), the x_1-coefficient is nonzero. So $A^{k+1}e_1/\|A^{k+1}e_1\|$ converges to αx_1 for some $\alpha \in \mathbb{C}$ with $|\alpha| = 1$. Let $(\widetilde{q}_k)_1$ denote the first column of $\widetilde{Q}_k$ and $(\widetilde{r}_k)_{11}$ denote the $(1,1)$-entry of $\widetilde{R}_k$; then $A^{k+1}e_1 = \widetilde{Q}_k\widetilde{R}_k e_1 = (\widetilde{r}_k)_{11}\widetilde{Q}_k e_1 = (\widetilde{r}_k)_{11}(\widetilde{q}_k)_1$, so $(\widetilde{q}_k)_1 \to \alpha x_1$. Since $A_{k+1} = \widetilde{Q}_k^H A \widetilde{Q}_k$, the first column of $A_{k+1} \to \begin{bmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

Further insight into the relationship between the QR algorithm and the power method, inverse power method, and subspace iteration, can be found in this delightful paper "Understanding the QR Algorithm" by D. S. Watkins (SIAM Review, vol. 24, 1982, pp. 427–440).