<u>Linear least-squares problems.</u>

In some applications, one has observed data $b_1, \dots, b_m$ that are assumed to depend on a set of n variables. It is often appropriate or convenient to assume a <u>linear model</u>: For $1 \le i \le m$,

$$b_i \approx a_{i1} x_1 + \dots + a_{in} x_n, \qquad (*)$$

where $a_{i1}, \dots, a_{in}$ are the known i^{th} values of the variables and $x_1, \dots, x_n$ are parameters to be determined so that the data are "fit" as well as possible by the model.

Rewrite $(*)$ as $Ax \approx b$, where

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m.$$

What's a good fit? If $m \le n$, <u>may</u> hope to satisfy $Ax = b$ exactly, but this is unlikely to give satisfactory results in most applications. Typically, $m > n$ (often $m >> n$) and "fit" is determined by

<u>Method of least-squares</u>: Choose $x_1, \dots, x_n$ to minimize the <u>sum</u> <u>of</u> <u>squared residuals</u>

$$R(x) = \sum_{i=1}^{m} \Big(b_i - \sum_{j=1}^{n} a_{ij} x_j\Big)^2.$$

Here, $b_i - \sum_{j=1}^{m} a_{ij} x_j$ is the i^{th} residual.

The method of least-squares goes back to Gauss. It is widely used for several reasons: It is easy to implement. It often gives satisfactory results. The results sometimes have special meaning in particular applications, e.g., as maximum-likelihood estimates in some statistical applications.

The problem of minimizing $R(x)$ is called a <u>linear least-squares problem</u> because the parameters $x_1, \dots, x_n$ appear linearly in the model. We can rewrite the problem as

$$\min R(x) = \min \| b - Ax \|^2,$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^m$: $\|v\| = (v^T v)^{\frac{1}{2}}$, associated with the Euclidean inner product $\langle u, v \rangle = u^T v$.

From this it is apparent that a solution is an $x \in \mathbb{R}^n$ such that Ax is <u>the orthogonal projection of b onto $\mathcal{R}(A)$</u>.

For every $x \in \mathbb{R}^n$, we have $b = Ax + (b - Ax)$, and $Ax \in \mathcal{R}(A)$. It follows that Ax will be the orthogonal projection of b onto $\mathcal{R}(A)$ for x such that $(b - Ax)$ is orthogonal to $\mathcal{R}(A)$. Since $\mathcal{R}(A)$ is the span of the columns of A, we have that

$$(b - Ax) \in \mathcal{R}(A)^{\perp} \Leftrightarrow A^T(b - Ax) = 0 \Leftrightarrow A^T A x = A^T b$$

The last equation is the famous <u>normal equation of least squares</u>. ("Normal" since it requires the residual to be orthogonal - normal - to the columns of A.)

<u>Lemma.</u> $\mathcal{N}(A) = \mathcal{N}(A^TA)$.

Proof: clearly, $\mathcal{N}(A) \subseteq \mathcal{N}(A^TA)$. On the other hand, if $v \in \mathcal{N}(A^TA)$, then for all $u \in \mathbb{R}^n$,

$$\langle Au, Av \rangle = \langle u, A^TAv \rangle = 0.$$

In particular, $\langle Av, Av \rangle = 0$, and it follows that $Av = 0$, i.e., $v \in \mathcal{N}(A)$.

It follows that if A is full-rank, then $\mathcal{N}(A^TA) = \mathcal{N}(A) = \{0\}$, and A^TA is nonsingular.

So if A is full-rank, then the <u>least-squares problem</u>

$$\min_{x \in \mathbb{R}^n} \|b - Ax\|^2$$

<u>has a unique solution</u> $x = (A^TA)^{-1}A^Tb$.

In any case, the least-squares problem has a solution

$\Leftrightarrow$ $A^TAx = A^Tb$ has a solution $x \in \mathbb{R}^n$

$\Leftrightarrow$ $\langle A^Tb, z \rangle = 0$ whenever $z \in \mathcal{N}((A^TA)^T) = \mathcal{N}(A^TA) = \mathcal{N}(A)$

$\Leftrightarrow$ $\langle A^Tb, z \rangle = 0$ whenever $Az = 0$

$\Leftrightarrow$ $\langle b, Az \rangle = 0$ whenever $Az = 0$

Since the last condition clearly holds, we conclude that <u>the least-squares problem always has a solution.</u>

Suppose x and $\hat{x}$ are solutions. Then $Ax = A\hat{x}$, since the projection of b onto $\mathcal{R}(A)$ is unique. It follows that $A(x - \hat{x}) = 0$, i.e., $x - \hat{x} \in \mathcal{N}(A)$.

A traditional way to solve a linear least-squares problem is to form and solve the normal equation. This is frequently unsatisfactory, usually because the solution of the normal equation is very sensitive to errors, including roundoff errors in the solution process and errors in the data. We will explore an alternative solution process shortly.

<u>Unitary transformations</u>.

Suppose $\mathcal{V}$ and $\mathcal{W}$ are inner-product spaces and $U : \mathcal{V} \to \mathcal{W}$ is linear

<u>Def</u>: U is a <u>unitary transformation</u> if U is invertible and $U^{-1} = U^*$.

<u>See</u>: For $u, v \in \mathcal{V}$, $\langle T(u), T(v)\rangle = \langle u, T^*T(v)\rangle = \langle u, v\rangle$, and $\|T(v)\| = \langle T(v), T(v)\rangle^{1/2} = \langle v, v\rangle^{1/2} = \|v\|$

So unitary transformations preserve <u>lengths</u> of vectors and <u>angles</u> between vectors.

A particular consequence is that if $\{u_1, \ldots, u_n\}$ is an orthonormal basis for $\mathcal{V}$, then $\{T(u_1), \ldots, T(u_n)\}$ is an orthonormal basis for $\mathcal{W}$. In particular, if $\mathcal{V} = \mathcal{W}$, then a unitary transformation transforms one orthonormal basis into another.

When $\mathcal{V} = \mathcal{W}$, if $\lambda \in \sigma(U)$ and v is an eigenvector, then $\|v\| = \|U(v)\| = \|\lambda v\| = |\lambda|\,\|v\|$. Consequently, $|\lambda| = 1$.

Note: If either $\mathcal{V}$ or $\mathcal{W}$ is finite-dimensional, then both are finite-dimensional and $\dim \mathcal{V} = \dim \mathcal{W}$.

Consider the case when $\mathcal{V} = \mathcal{W} = \mathbb{R}^n$ or $\mathbb{C}^n$ and U is a matrix in $\mathbb{R}^{n\times n}$ or $\mathbb{C}^{n\times n}$.

If $U \in \mathbb{C}^{n\times n}$, it is a <u>unitary</u> matrix.
If $U \in \mathbb{R}^{n\times n}$, it is an <u>orthogonal</u> matrix.

<u>See:</u> $U^T U = I$ $(U^* U = I)$ $\Leftrightarrow$ the columns of U are orthonormal $\Leftrightarrow$ the rows of U are orthonormal.

<u>Ex.:</u> <u>Rotation matrix</u> in $\mathbb{R}^{2\times 2}$

$$U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \sim \text{rotation through angle } \theta$$

More generally, have

$$\begin{array}{cc} U = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \cos\theta & & -\sin\theta & & \\ & & & 1 & & & \\ & & \sin\theta & & \cos\theta & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} & \begin{array}{l} \\ \\ \leftarrow i^{th} \text{ row} \\ \\ \leftarrow j^{th} \text{ row} \\ \\ \\ \end{array} \\ \qquad\qquad \uparrow \qquad\quad \uparrow & \\ \qquad\quad i^{th} \text{ column} \quad j^{th} \text{ column} & \end{array}$$

Especially useful for matrix factorizations.
These are called <u>Givens rotations</u>, after Wallace Givens, who used them at Argonne Lab in the 1950s.

<u>See:</u> Uv rotates the component of v in the ij^{th} coordinate plane.

<u>Ex.:</u> Permutation matrix P = matrix obtained from I by permuting rows or columns.

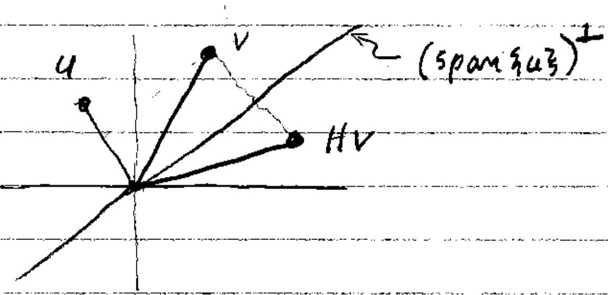
An example worth studying in depth is

<u>Householder transformations</u>.

Named after Alston Householder, who introduced them in 1958. Also known as Householder matrices or <u>elementary reflectors</u>. Like Givens rotations, they are widely used in computing matrix factorizations.

Form: $H = I - 2uu^T$, where $u \in \mathbb{R}^n$ satisfies $\|u\| = 1$.

Geometrically, $Hv = v - 2\langle v, u\rangle u$ is the reflection of v across the hyperplane of vectors orthogonal to u.



Algebraically,

(1) $H^T = H$ (symmetric)

(2) $H^2 = I$ (idempotent)

From these, see that H is <u>orthogonal</u>.

Suppose we have v, w and want $H = I - 2uu^T$ such that $Hv = w$.

Since $\|Hv\| = \|v\|$, must have $\|v\| = \|w\|$. Assuming this, set

$$u = \frac{v-w}{\|v-w\|}.$$

Then
$$Hv = v - 2\frac{(v-w)}{\|v-w\|}\frac{(\|v\|^2 - v^Tw)}{\|v-w\|} = v - 2\frac{(v-w)(\|v\|^2 - v^Tw)}{\|v\|^2 - 2v^Tw + \|w\|^2}$$
$$= v - 2\frac{(v-w)(\|v\|^2 - v^Tw)}{2(\|v\|^2 - v^Tw)} = v \quad \text{since } \|v\| = \|w\|.$$

Application to QR decomposition.

Suppose $A \in \mathbb{R}^{m\times n}$. For convenience, assume $m \geq n$. (Case $m < n$ is similar.)

Step 1: Find $H_1 = I - 2u_1u_1^T$

$$H_1 \times (1^{st} \text{ column of } A) = \begin{pmatrix} r_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where $|r_{11}|$ = norm of 1^{st} column of A.

Then $H_1 A = \begin{pmatrix} r_{11} & x & - \\ 0 & x & - \\ \vdots & \vdots & \\ 0 & x & - \end{pmatrix}$

Step 2: Find $H_2 = I - 2u_2u_2^T$ such that

$u_2 = \begin{pmatrix} 0 \\ x \\ \vdots \\ x \end{pmatrix}$ and $H_2H_1A = \begin{pmatrix} r_{11} & x & x & - \\ 0 & r_{22} & x & - \\ \vdots & 0 & x & - \\ 0 & 0 & x & - \end{pmatrix}$

Note: Since the 1^{st} component of u_2 is zero, H_2H_1A has the same 1^{st} row as H_1A.

Continue to get $H_{n-1} \cdots H_1 A = \begin{pmatrix} R \\ O \end{pmatrix} \begin{matrix} \}n \\ \}m-n \end{matrix}$

where $R = \begin{pmatrix} r_{11} & & = \\ & \ddots & \\ O & & r_{nn} \end{pmatrix} \in \mathbb{R}^{n\times n}$.

Multiplying on the left by $H_1 \cdots H_{n-1}$ gives

$$A = H_1 \cdots H_{n-1} \begin{pmatrix} R \\ O \end{pmatrix} = Q \begin{pmatrix} R \\ O \end{pmatrix},$$

where $Q = H_1 \cdots H_{n-1}$. This is a QR decomposition of A.

Note: A product of orthogonal (unitary) matrices is orthogonal (unitary).

So $Q = H_1 \cdots H_{n-1}$ is orthogonal.

Application to least-squares problems.

Suppose we have a least-squares problem $\min \|b - Ax\|$, where $A \in \mathbb{R}^{m \times n}$ with $m \leq n$.

Assume that A is full-rank, and suppose we have the QR decomposition $A = Q \binom{R}{0}$, where $Q \in \mathbb{R}^{m \times m}$ is orthogonal and $R \in \mathbb{R}^{n \times n}$ is upper-triangular. Note that R is nonsingular since A is full rank.

We have

$$A^T A = \binom{R}{0}^T Q^T Q \binom{R}{0} = (R^T \; 0^T) \binom{R}{0} = R^T R$$

and

$$A^T b = \binom{R}{0}^T Q^T b = R^T \hat{b},$$

where

$$\hat{b} = \text{vector of the } 1^{st} \text{ } n \text{ components of } Q^T b.$$

Then the normal equation $A^T A x = A^T b$ becomes $R^T R x = R^T \hat{b}$, or just $Rx = \hat{b}$ since R^T is nonsingular.

So the steps of the solution process are

(1) Determine $A = Q \binom{R}{0}$.

(2) Form $\hat{b} = 1^{st}$ n components of $Q^T b$

(3) Solve $Rx = \hat{b}$.

Note: Solving upper triangular systems is easy (use back substitution).

Remark. A popular alternative is the "thin" QR decomposition $A = QR$, where $Q \in \mathbb{R}^{m\times n}$ and $R \in \mathbb{R}^{n\times n}$. This can be obtained by performing Gram-Schmidt orthogonalization on the columns of A. The resulting orthonormal vectors are the columns of Q, and the scalar coefficients generated during the orthogonalization are the entries in R.

Self-adjoint transformations.

Suppose $T: \mathcal{V} \to \mathcal{V}$ is a linear transformation on an inner product space $\mathcal{V}$.

Def: T is self-adjoint if $T^* = T$.

Ex.: Symmetric $n\times n$ matrices determine self-adjoint transformations on $\mathbb{R}^n$.

Explore properties of eigenvalues and eigenvectors of a self-adjoint T. Assume $\mathcal{F} = \mathbb{C}$, since eigenvalues and eigenvectors may be complex in general.

Prop. $\sigma(T) \subseteq \mathbb{R}$.

Pf.: Suppose $T(v) = \lambda v$ for $v \neq 0$. Then

$$\lambda \langle v,v\rangle = \langle \lambda v, v\rangle = \langle T(v), v\rangle = \langle v, T(v)\rangle = \langle v, \lambda v\rangle = \bar{\lambda}\langle v, v\rangle.$$

It follows that $\lambda = \bar{\lambda}$, i.e., that λ is real.

Prop. Eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Pf.: Suppose $T(v) = \lambda v$ and $T(w) = \mu w$ for $\lambda \neq \mu$. Then

$$\lambda \langle v, w \rangle = \langle \lambda v, w \rangle = \langle T(v), w \rangle = \langle v, T(w) \rangle$$
$$= \langle v, \mu w \rangle = \mu \langle v, w \rangle \quad \text{since } \bar{\mu} = \mu.$$

Since $\lambda \neq \mu$, this implies $\langle v, w \rangle = 0$.

Say a subspace $S \subseteq \mathcal{V}$ is invariant under T if $T(S) \subseteq S$, i.e., $T(v) \in S$ whenever $v \in S$.

For $\lambda \in \sigma(T)$, set $S_\lambda = \{ v \in \mathcal{V} : T(v) = \lambda v \}$. Clearly, S_λ is invariant under T.

Lemma: If $\lambda \in \sigma(T)$, then $S_\lambda^\perp$ is invariant under T.

Pf.: Suppose $v \in S_\lambda$, i.e., $T(v) = \lambda v$, and $w \in S_\lambda^\perp$, i.e., $\langle v, w \rangle = 0$. Then

$$\langle v, T(w) \rangle = \langle T(v), w \rangle = \langle \lambda v, w \rangle = \lambda \langle v, w \rangle = 0.$$

Theorem: If T is a self-adjoint transformation on a finite-dimensional inner-product space $\mathcal{V}$, then $\mathcal{V}$ has an orthonormal basis of eigenvectors of T.

Proof: The proof is by induction on the dimension of $\mathcal{V}$. The result is trivial when $\dim \mathcal{V} = 1$. Suppose that, for $n > 1$, the result holds for spaces of dimension k, $1 \leq k \leq n-1$. Suppose $\dim \mathcal{V} = n$.

We know that T has an eigenvalue λ. We have that $\dim S_\lambda + \dim S_\lambda^\perp = n$, so $\dim S_\lambda^\perp = k$ for $1 \le k \le n-1$. Since $S_\lambda^\perp$ is invariant under T and the result holds for $S_\lambda^\perp$, there is an orthonormal basis of eigenvectors $\{u_1, -, u_k\}$ for $S_\lambda^\perp$. Letting $\{u_{k+1}, -, u_n\}$ be an orthonormal basis for S_λ, we have that $\{u_1, \dots, u_k, u_{k+1}, \dots, u_n\}$ is an orthonormal basis of eigenvectors for $\mathcal{V}$.

Remark: If $\{u_1, -, u_n\}$ is an orthonormal basis of eigenvectors of $T: \mathcal{V} \to \mathcal{V}$, then for $v \in \mathcal{V}$, have $v = \sum_{i=1}^{n} \langle v, u_i \rangle u_i$ and

$$T(v) = \sum_{i=1}^{n} \lambda_i \langle v, u_i \rangle u_i$$

This is called a spectral representation of T. (Non-unique since basis is non-unique.)

Can this be generalized? Is there a broader class of linear transformations that have orthonormal bases of eigenvectors and corresponding spectral representations?

Note that if $\mathcal{V} = \mathbb{R}^n$ and $T(v) = Av$ with $A = A^T$, then $A = U \Lambda U^T$ for orthogonal $U \in \mathbb{R}^{n \times n}$ and $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{pmatrix} \in \mathbb{R}^{n \times n}$. Similarly, if $A \in \mathbb{C}^{n \times n}$ with $A = A^*$, then $A = U \Lambda U^*$ for unitary $U \in \mathbb{C}^{n \times n}$ and $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{pmatrix} \in \mathbb{R}^{n \times n}$.