The Linear Least Squares Problems

Many real-world problems could be formulated as solving $A\mathbf{x} = \mathbf{b}$, where $A \in R^{m \times n}$, $\mathbf{x} \in R^n$, $\mathbf{b} \in R^m$, with $m > n$. The linear least squares problem is to find a solution which minimizes $\|\mathbf{b} - A\mathbf{x}\|_2^2$.

- Examples of Polynomial Regression
- Existence and Uniqueness
- Normal Equations $A^t A\mathbf{x} = A^t\mathbf{b}$
- Gram-Schmidt Orthogonalization and QR Factorization
- Orthogonal Transforms

  Householder Transforms

  Jacobi Transforms (Givens Rotations)

The Linear Least Squares Problems

Consider the problem of determining an $\mathbf{x} \in R^n$ such that the *residual sum of squares* $\rho^2(\mathbf{x}) = \|\mathbf{b} - A\mathbf{x}\|_2^2$ is minimized for given $\mathbf{b} \in R^n$, $A \in R^{n\times n}$.

◇ *Best Line Fit:*

Given $[x_i, y_i]^t \in R^2$ for $1 \le i \le n$, find a line which best fits these points. The problem is equivalent to finding m and b to minimize

$$f(m,b) = \sum_{i=1}^{n} (y_i - m x_i - b)^2$$

or to solve

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix}$$

◇ *Best Parabola Fit:*

Given $[x_i, y_i]^t \in R^2$ for $1 \le i \le n$, find a parabola which best fits these points. The problem is equivalent to finding a, b, c to minimize

$$f(a,b,c) = \sum_{i=1}^{n} (y_i - a x_i^2 - b x_i - c)^2$$

or to solve

$$\begin{bmatrix} \sum_{i=1}^n x_i^4 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i & \sum_{i=1}^n 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i^2 y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix}$$

An Example of Best Line Fit

1. The respective high school and college GPAs for 20 college seniors as ordered pairs (x, y) are

(3.75, 3.19)	(3.45, 3.34)	(2.87, 2.23)	(3.60, 3.46)	(3.42, 2.97)
(4.00, 3.79)	(2.65, 2.55)	(3.10, 2.50)	(3.47, 3.15)	(2.60, 2.26)
(4.00, 3.76)	(2.30, 2.11)	(2.47, 2.11)	(3.36, 3.01)	(3.60, 2.92)
(3.65, 3.09)	(3.30, 3.05)	(2.58, 2.63)	(3.80, 3.22)	(3.79, 3.27)

(a) Verify that $\bar{x}$= 3.2880, $\bar{y}$ =2.9305, $s_x^2 = 0.283$, $s_y^2 = 0.260$, and $r = 0.92$.

(b) The equation of best fitting line is $y = 0.8822x + 0.0298$.

(c) Plot the 20 points and the best fitting line on the same graph.

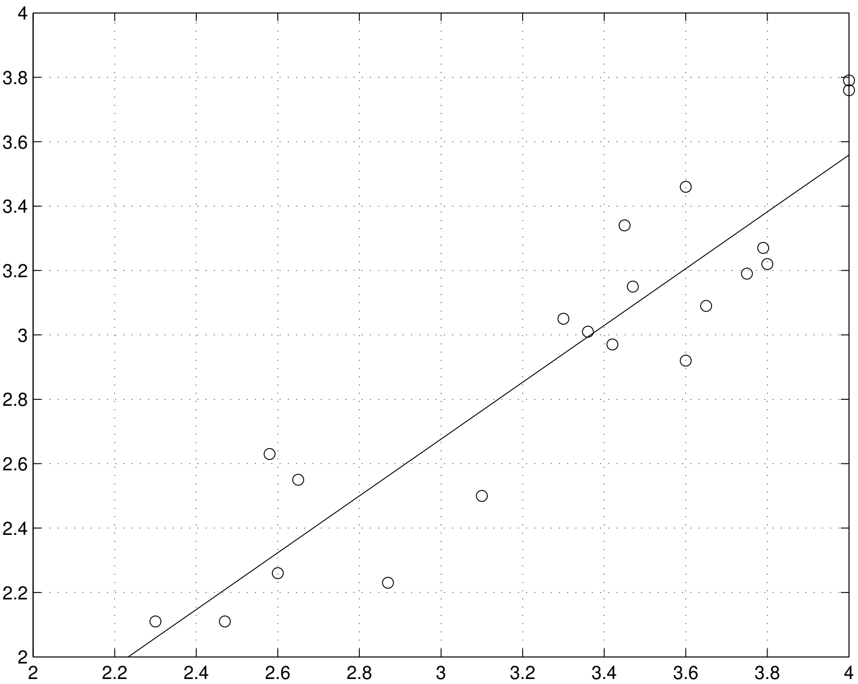

Figure 1: Plot of A Best Fitted Line

An Example of Best Parabola Fit

2. A time series of 21 data points (t, y) are given below.

t	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y	2.9	2.7	4.8	5.3	7.1	7.6	7.7	7.6	9.4	9.0	9.6
t	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	
y	10.0	10.2	9.7	8.3	8.4	9.0	8.3	6.6	6.7	4.1	

(a) The best fitted parabola is $y = -0.238t^2 + 2.67t + 2.18$

(b) Plot the 21 points and the best fitting parabola on the same graph.

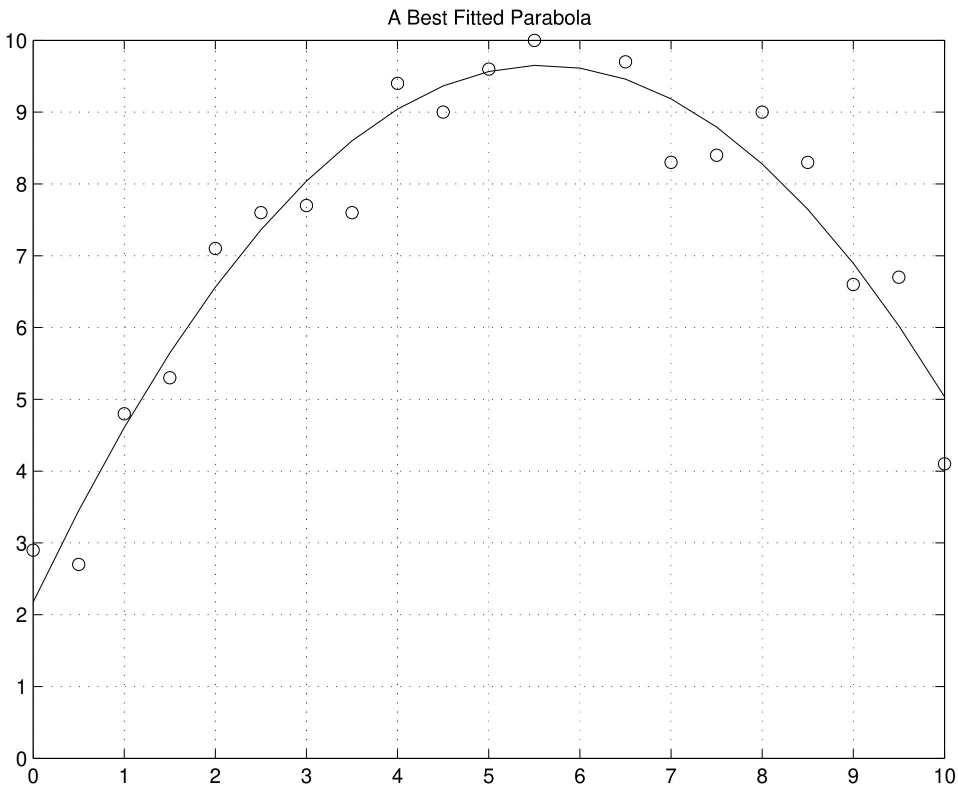

Figure 2: Plot of A Best Fitted Parabola

Existence and Uniqueness

Theorem: The linear least squares problem of minimizing $\|\mathbf{b}-A\mathbf{x}\|_2$ always has a solution. The solution is unique iff $Null(A) = \{\mathbf{0}\}$.

Corollary: Let $\mathbf{x}$ be a linear least squares solution of minimizing $\|\mathbf{b} - A\mathbf{x}\|_2$, then the residual vector $\mathbf{r} = \mathbf{b} - A\mathbf{x}$ satisfies the following *normal equations*.

$$A^t\mathbf{r} = A^t(\mathbf{b} - A\mathbf{x}) = \mathbf{0} \quad \textit{or} \quad A^tA\mathbf{x} = A^t\mathbf{b}$$

Theorem: $A\mathbf{x} = \mathbf{b}$ has a solution iff $\mathbf{b} \in R(A)$.

If the columns of A are linearly independent, then A^tA is invertible and $\mathbf{x} = (A^tA)^{-1}A^t\mathbf{b}$. The projection of $\mathbf{b}$ onto the column space of matrix A is $\mathbf{p} = A(A^tA)^{-1}A^t\mathbf{b}$.

Example:

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \quad A^tA = \begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}$$

Theorem: If $A \in R^{m\times n}$ has rank n $(n \le m)$, the normal equations $A^tA\mathbf{x} = A^t\mathbf{b}$ has a unique solution $\hat{\mathbf{x}} = (A^tA)^{-1}A^t\mathbf{b}$ and $\hat{\mathbf{x}}$ is the unique LLS solution to $A\mathbf{x} = \mathbf{b}$.

Orthonormal Basis and Orthogonal Matrices

Definition: The vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ are orthonormal if $\|\mathbf{u}_k\|_2 = 1,\ 1 \le k \le n$, and $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0\ \forall\ i \ne j$.

Definition: An *orthogonal matrix* is simply a matrix with *orthonormal columns*. That is, $Q \in R^{m\times k}$ is orthogonal if $Q^tQ = I_k$. In particular, if $m = k$, then $Q^{-1} = Q^t$.

♣ *Some Properties of Orthogonal Matrices*

(a) The columns of Q form an orthonormal basis for R^n

(b) $Q^tQ = I$ and $Q^{-1} = Q^t$

(c) $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2,\ \ \forall\ \mathbf{x} \in R^n$

(d) $\langle Q\mathbf{x}, Q\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle,\ \forall\ \mathbf{x}, \mathbf{y} \in R^n$

(e) $\|QA\|_2 = \|A\|_2,\ \ \forall\ A \in R^{n\times k}$

(f) $det(Q) = |Q| = 1\ or\ -1$

♣ *Least Squares and Orthonormal Sets*

Theorem: *If the column vectors of $A \in R^{m\times n}$ form an orthonormal set of vectors in R^m, then $A^tA = I$ and the LLS solution to $A\mathbf{x} = \mathbf{b}$ is $\hat{\mathbf{x}} = (A^tA)^{-1}A^t\mathbf{b} = A^t\mathbf{b}$.*

Theorem: *Let S be a subspace of an inner product vector space V and $\mathbf{x} \in V$. Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ be an orthonromal basis for S. If*

$$\mathbf{p} = \sum_{i=1}^{n} c_i\mathbf{u}_i, \quad where\ c_i = \langle \mathbf{x}, \mathbf{u}_i \rangle$$

Then, $(\mathbf{x} - \mathbf{p}) \in S^{\perp}$

Proof: $\langle \mathbf{x} - \mathbf{p},\ \mathbf{u}_i \rangle = \langle \mathbf{x},\ \mathbf{u}_i \rangle - \langle \mathbf{p},\ \mathbf{u}_i \rangle = c_i - c_i = 0$

Gram-Schmidt Orthogonalization Process

Let $V = \{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$ *be a set of independent vectors. The Gram-Schmidt process transforms the set* V *to an orthonormal set of* $U = \{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n\}$ *such that*

$$span(\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n) \;=\; span(\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)$$

(a) $\mathbf{q}_1 \;\leftarrow\; \mathbf{a}_1/\|\mathbf{a}_1\|_2$

(b) $\mathbf{t}_2 = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1; \quad \mathbf{q}_2 \;\leftarrow\; \mathbf{t}_2/\|\mathbf{t}_2\|_2$

(c) $\mathbf{t}_k = \mathbf{a}_k - \sum_{i=1}^{k-1} \langle \mathbf{a}_k, \mathbf{q}_i \rangle \mathbf{q}_i; \quad \mathbf{q}_k \;\leftarrow\; \mathbf{t}_k/\|\mathbf{t}_k\|_2$ *for* $3 \le k \le n$.

Example:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \ \mathbf{a}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{a}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}; \ \ \mathbf{q}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \ \mathbf{q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \ \mathbf{q}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Example (QR Factorization):

$$A = [\mathbf{a}_1, \ \mathbf{a}_2, \ \mathbf{a}_3] = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$$

Example:

$$A = [\mathbf{a}_1, \ \mathbf{a}_2, \ \mathbf{a}_3] = [\mathbf{q}_1, \ \mathbf{q}_2, \ \mathbf{q}_3] \begin{bmatrix} \mathbf{q}_1^t \mathbf{a}_1 & \mathbf{q}_1^t \mathbf{a}_2 & \mathbf{q}_1^t \mathbf{a}_3 \\ 0 & \mathbf{q}_2^t \mathbf{a}_2 & \mathbf{q}_2^t \mathbf{a}_3 \\ 0 & 0 & \mathbf{q}_3^t \mathbf{a}_3 \end{bmatrix} = QR$$

QR Factorization

Theorem: *Every* $A \in R^{m\times n}$ *with linearly independent columns can be factored into* $A = QR$*, where* Q *is orthogonal,* R *is upper-*Δ *and invertible.*

Proof: *Successively applied Householder matrices* $\{H_j's\}$ *on* A*, we can get* $H_1H_2\cdots H_mA = R$*, where ,* R *is upper-*Δ*. If* R *is not invertible, then* $\exists$ $\mathbf{x} \in R^n$ *such that* $R\mathbf{x} = \mathbf{0}$*, then* $QR\mathbf{x} = \mathbf{0}$ *and hence* $A\mathbf{x} = \mathbf{0}$ *which contradicts that* A *has linearly independent column vectors.*

Note: *Suppose* $A = QR$*, the LLS solution of* $A\mathbf{x} = \mathbf{b}$ *is reduced to solving a triangular system of equations* $R\mathbf{x} = Q^t\mathbf{b}$*.*

Example:

$$A = \begin{bmatrix} 1 & -2 & -1 \\ 2 & 0 & 1 \\ 2 & -4 & 2 \\ 4 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.2 & -0.4 & -0.8 \\ 0.4 & 0.2 & 0.4 \\ 0.4 & -0.8 & 0.4 \\ 0.8 & 0.4 & -0.2 \end{bmatrix} \begin{bmatrix} 5 & -2 & 1 \\ 0 & 4 & -1 \\ 0 & 0 & 2 \end{bmatrix} = QR$$

Let $\mathbf{b} = [-1.4, 0.2, 1.2, -1.6]^t$*. By solving* $R\mathbf{x} = Q^t\mathbf{b}$*, we have* $\mathbf{x} = [-0.4, 0, 1]^t$ *for the LLS solution of* $A\mathbf{x} = \mathbf{b}$*.*

Householder Transforms

♣ *A Householder transform (matrix) can be defined as*

$$H = I - 2\mathbf{u}\mathbf{u}^t, \quad \textit{where } \mathbf{u} \in R^n \textit{ with } \|\mathbf{u}\|_2 = 1$$

♣ *A Householder matrix H is symmetric, orthogonal, and det$(H) = -1$*

Theorem: *Let $\mathbf{x} = [x_1, x_2, \cdots, x_n]^t \in R^n$ and $\|\mathbf{x}\|_2 = \alpha = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$, define $\mathbf{v} = \mathbf{x} + \alpha\mathbf{e}_1$ with $\|\mathbf{v}\|_2 = r$ and $\mathbf{u} = \frac{\mathbf{v}}{r}$. If $H = I - 2\mathbf{u}\mathbf{u}^t$, then $H\mathbf{x} = -\|\mathbf{x}\|_2\mathbf{e}_1 = -\alpha\mathbf{e}_1$.*

Proof: $r^2 = \mathbf{v}^t\mathbf{v} = \sum_{i=1}^n v_i^2 = (\alpha + x_1)^2 + \sum_{j=2}^n x_j^2$, *then* $r^2 = 2(\alpha^2 + \alpha x_1)$. *On the other hand,* $\mathbf{v}^t\mathbf{x} = \sum_{i=1}^n v_i x_i = (\alpha + x_1)x_1 + \sum_{j=2}^n v_j x_j$, *then* $\mathbf{v}^t\mathbf{x} = \|\mathbf{x}\|_2^2 + \alpha x_1 = \alpha^2 + \alpha x_1$.
Thus

$$\begin{aligned} H\mathbf{x} &= (I - 2\mathbf{u}\mathbf{u}^t)\mathbf{x} \\ &= \left(I - 2 \cdot \frac{\mathbf{v}\mathbf{v}^t}{\mathbf{v}^t\mathbf{v}}\right)\mathbf{x} \\ &= \mathbf{x} - 2 \cdot \frac{(\alpha^2 + \alpha x_1)}{2(\alpha^2 + \alpha x_1)} \cdot \mathbf{v} \\ &= \mathbf{x} - \mathbf{v} \\ &= -\alpha\mathbf{e}_1 \;=\; -\|\mathbf{x}\|_2\mathbf{e}_1 \end{aligned}$$

□ Example*: Let $\mathbf{x} = [3, 1, 5, 1]^t$, then $\|\mathbf{x}\|_2 = \sqrt{3^2 + 1^2 + 5^2 + 1^2} = 6$. Define $\mathbf{v} = \mathbf{x} + \|\mathbf{x}\|_2\mathbf{e}_1$, and let $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|_2$, then*

$$H = I - 2\mathbf{u}\mathbf{u}^t = \frac{1}{54}\begin{bmatrix} -27 & -9 & -45 & -9 \\ -9 & 53 & -5 & -1 \\ -45 & -5 & 29 & -5 \\ -9 & -1 & -5 & 53 \end{bmatrix}, \quad \textit{and} \quad H\mathbf{x} = \begin{bmatrix} -6 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Givens Rotations (Jacobi Transforms)

$$
J(i,k;\theta) = \begin{bmatrix}
1 & \cdot & \cdot & \cdots & \cdot & \cdot & 0 \\
0 & \ddots & \cdot & \cdots & \cdot & \vdots & 0 \\
0 & \cdot & c & \cdots & s & \cdot & 0 \\
\cdot & \vdots & \cdot & \ddots & \cdot & \vdots & \cdot \\
0 & \cdot & -s & \cdots & c & \cdot & 0 \\
0 & \vdots & \cdot & \cdots & \cdot & \ddots & 0 \\
\cdot & \cdot & 0 & \cdots & 0 & \cdot & 1
\end{bmatrix}
$$

$J_{hh} = 1 \ \textit{if}\ h \neq i \ \textit{or}\ h \neq k, \ \textit{where}\ i < k$

$J_{ii} = J_{kk} = c = \cos\theta$

$J_{ki} = -s = -\sin\theta, \ J_{ik} = s = \sin\theta$

Let $\mathbf{x}, \mathbf{y} \in R^n$*, then* $\mathbf{y} = J(i,k;\theta)\mathbf{x}$ *implies that*

$y_i = cx_i + sx_k$

$y_k = -sx_i + cx_k$

$c = \frac{x_i}{\sqrt{x_i^2 + x_k^2}}, \ s = \frac{x_k}{\sqrt{x_i^2 + x_k^2}},$

$$
\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}, \quad \textit{then} \quad J(2,4;\theta)\mathbf{x} = \begin{bmatrix} 1 \\ \sqrt{20} \\ 3 \\ 0 \end{bmatrix}
$$

Sensitivity and Conditioning

Define $Cond(A) = \|A\|_2 \cdot \|A^+\|_2$, *where for* $A \in R^{m\times n}$ *with* $rank(A) = n < m$ *and* $A^+ = (A^tA)^{-1}A^t$. *Let* $span(A)$ *be the column space of matrix* A, *an appropriate measure of the closeness of* $\mathbf{b}$ *to* $span(A)$ *is the ratio*

$$\frac{\|A\mathbf{x}\|_2}{\|\mathbf{b}\|_2} = \frac{\|\mathbf{y}\|_2}{\|\mathbf{b}\|_2} = \cos(\theta),$$

where θ *is the angle between* $\mathbf{y}$ *and* $\mathbf{b}$.

♣ Analysis of the sensitivity of LLS solution to $A\mathbf{x} = \mathbf{b}$ with $rank(A) = n < m$

$$A^tA(\mathbf{x} + \Delta\mathbf{x} = A^t(\mathbf{b} + \Delta\mathbf{b}) \quad and \quad A^tA\Delta\mathbf{x} = A^t\Delta\mathbf{b}$$

Then

$$\Delta\mathbf{x} = (A^tA)^{-1}A^t\Delta\mathbf{b} = A^+\Delta\mathbf{b} \quad and \quad \|\Delta\mathbf{x}\|_2 \le \|A^+\|_2 \cdot \|\Delta\mathbf{b}\|_2$$

Thus

$$\frac{\|\Delta\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \le \|A^+\|_2 \cdot \frac{\|\Delta\mathbf{b}\|_2}{\|\mathbf{x}\|_2} \le Cond(A) \cdot \frac{1}{\cos(\theta)} \cdot \frac{\|\Delta\mathbf{b}\|_2}{\|\mathbf{b}\|_2}$$

On the other hand,

$$(A+E)^t(A+E)(\mathbf{x} + \Delta\mathbf{x}) = (A+E)^t\mathbf{b} \quad and \quad A^tA\Delta\mathbf{x} = A^t\Delta\mathbf{b}$$

By dropping the 2nd-order terms, we have

$$A^tA\Delta\mathbf{x} \approx E^t\mathbf{b} - E^tA\mathbf{x} - A^tE\mathbf{x} = E^t(\mathbf{b} - A\mathbf{x}) - A^tE\mathbf{x} = E^t\mathbf{r} - A^tE\mathbf{x}$$

$$\Delta\mathbf{x} \approx (A^tA)^{-1}E^t\mathbf{r} - (A^tA)^{-1}A^tE\mathbf{x} = (A^tA)^{-1}E^t\mathbf{r} - A^+E\mathbf{x}$$

Taking norms to obtain

$$\|\Delta\mathbf{x}\|_2 \le \|(A^tA)^{-1}\|_2 \cdot \|E\|_2 \cdot \|\mathbf{r}\|_2 + \|A^+\|_2 \cdot \|E\|_2 \cdot \|\mathbf{x}\|_2$$

Thus

$$\frac{\|\Delta\mathbf{x}\|_2}{\|\mathbf{x}\|_{\mathbf{2}}} \le \Big([Cond(A)]^2 tan(\theta) + Cond(A)\Big) \cdot \frac{\|E\|_2}{\|A\|_2}$$

♣ Examples 3.5 and 3.6 on *Page 116*

Spectrum Decomposition for Symmetric Matrices

Schur's Theorem: $\forall\, A \in R^{n\times n}$, $\exists$ an orthogonal matrix U such that $U^tAU = T$ is upper-Δ. The eigenvlues must be shared by the similarity matrix T and appear along its main diagonal.

Hint: By induction, suppose that the theorem has been proved for all matrices of order $n-1$, and consider an matrix $A \in R^{n\times n}$ with $A\mathbf{x} = \lambda\mathbf{x}$ and $\|\mathbf{x}\|_2 = 1$, then $\exists$ a Householder matrix H_1 such that $H_1\mathbf{x} = \beta\mathbf{e}_1$, *e.g.*, $\beta = -\|\mathbf{x}\|_2$, hence

$$H_1AH_1^t\mathbf{e}_1 = H_1A(H_1^{-1}\mathbf{e}_1) = H_1A(\beta^{-1}\mathbf{x}) = H_1\beta^{-1}A\mathbf{x} = \beta^{-1}\lambda(H_1\mathbf{x}) = \beta^{-1}\lambda(\beta\mathbf{e}_1) = \lambda\mathbf{e}_1$$

Thus,

$$H_1AH_1^t = \begin{bmatrix} \lambda & | & * \\ --- & | & --- \\ O & | & A^{(1)} \end{bmatrix}$$

Spectrum Decomposition Theorem: Every real symmetric matrix can be diagonalized by an orthogonal matrix.

$\diamond$ $Q^tAQ = \Lambda$ or $A = Q\Lambda Q^t = \sum_{i=1}^n \lambda_i \mathbf{q}_i\mathbf{q}_i^t$

Definition: A symmetric matrix $A \in R^{n\times n}$ is nonnegative definite if $\mathbf{x}^tA\mathbf{x} \geq 0\ \forall\ \mathbf{x} \in R^n$.

Definition: A symmetric matrix $A \in R^{n\times n}$ is positive definite if $\mathbf{x}^tA\mathbf{x} > 0\ \forall\ \mathbf{x} \in R^n$, $\mathbf{x} \neq \mathbf{0}$.

Singular Value Decomposition

Singular Value Decomposition Theorem: Each matrix $A \in R^{m\times n}$ can be decomposed as $A = U\Sigma V^t$, where both $U \in R^{m\times m}$ and $V \in R^{n\times n}$ are orthogonal. Moreover, $\Sigma \in R^{m\times n} = diag[\sigma_1, \sigma_2, \ldots, \sigma_k, 0, \ldots, 0]$ is essentially diagonal with the singular values satisfying $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k > 0$.

◇ $A = U\Sigma V^t = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^t$

Examples:

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$$