The QR factorization

1) Used as an alternative to Gaussian Elimination for solving $Ax=b$. work estimate is twice that of G.E. but is more stable.

2) Used in solving the Least-Squares problem

3) Main ingredient in the QR method for finding all eigenvalues of a matrix.

Defn Outer product of two vectors let $v \in \mathbb{R}^m$, $w \in \mathbb{R}^n$

The outer product vw^T of v and w is the $m \times n$ matrix

$$(vw^T)_{ij} = v_i w_j, \quad i=1,\dots,m,\ j=1,\dots,n.$$

Ex. $v=(1,-2,3)^T$, $w=(2,4)^T$

$$vw^T = \begin{bmatrix} 2 & 4 \\ -4 & -8 \\ 6 & 12 \end{bmatrix}$$

Note vw^T is a rank-one matrix since all rows (or columns) are multiples of each other

Defn. let $v \in \mathbb{R}^n$ $v \neq 0$. The matrix

$$H(v) = I - \frac{2\, v v^T}{\|v\|_2^2} \qquad (\text{note } \|v\|_2^2 = v^T v)$$

is called a Householder matrix.

If $v=0$, we set $H(0)=I$.

we immediately see that $H(v)$ is <u>symmetric</u>

Indeed

$$H(v)^T = \left(I - 2\frac{vv^T}{\|v\|_2^2}\right) = I - \frac{2}{\|v\|_2^2}(vv^T)^T$$

$$= I - \frac{2}{\|v\|_2^2} vv^T = H(v).$$

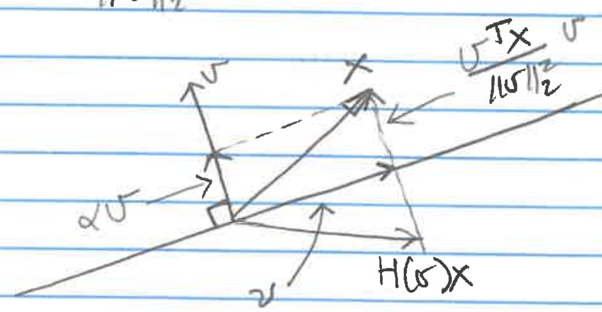
Also, $H(v)$ is <u>orthogonal</u>, i.e. $H(v)H(v)^T = I$

$$H(v)H(v)^T = H^2 = \left(I - 2\frac{vv^T}{\|v\|_2^2}\right)\left(I - 2\frac{vv^T}{\|v\|_2^2}\right)$$

$$= I - \frac{2vv^T}{\|v\|_2^2} - \frac{2vv^T}{\|v\|_2^2} + 4\frac{v\overset{\|v\|_2^2}{v^Tv}v^T}{\|v\|_2^4}$$

$$= I - 4\frac{vv^T}{\|v\|_2^2} + 4\frac{vv^T}{\|v\|_2^2} = I. \checkmark$$

$H(v)$ is a <u>reflector</u>



i.e. Given x, $H(v)x$ is the <u>reflection</u> of x in the hyperplane orthogonal to v.

Indeed, we can decompose x into:

$$x = \frac{v^Tx}{\|v\|_2^2}v + \nu \quad \text{where } \frac{v^Tx}{\|v\|_2^2}v \text{ is the}$$

projection of x along v and ν is orthogonal to v, i.e. $\nu^Tv = 0$.

$$H(v)x = \left(I - 2\frac{vv^T}{\|v\|_2^2}\right)x = \left(I - \frac{2vv^T}{\|v\|_2^2}\right)\left(\overset{\text{scalar}}{\frac{v^Tx}{\|v\|_2^2}}v + \nu\right)$$

$$= \frac{v^Tx}{\|v\|_2^2}v + \nu - \frac{2vv^T}{\|v\|_2^2}\frac{v^Tx}{\|v\|_2^2}v - \frac{2v\cancelto{0}{v^T\nu}}{\|v\|_2^2}$$

$$= \frac{v^Tx}{\|v\|_2^2}v + \nu - \frac{2v^Tx}{\|v\|_2^4}vv^Tv$$

$$= -\frac{v^Tx}{\|v\|_2^2}v + \nu$$

and this the reflection of $x = +\frac{v^Tx}{\|v\|_2^2}v + \nu$.

Main Idea we would like to use Householder type matrices to introduce zeros in a given vector just as was done using elimination matrices. In other words, Given x, we want to find v such that

$$H(v)x = \begin{bmatrix}\alpha\\0\\|\\|\\0\end{bmatrix} = \alpha e_1 \quad , \quad e_1 = \begin{bmatrix}1\\0\\|\\|\\0\end{bmatrix}.$$

We find v and α as follows: Assume $x \neq 0$.

$$\alpha e_1 = H(v)x = \left(I - 2\frac{vv^T}{\|v\|_2^2}\right)x = x - 2\frac{v^Tx}{\|x\|_2^2}v$$

$$\Rightarrow 2\frac{v^Tx}{\|v\|_2^2}v = x - \alpha e_1 \Rightarrow v = \frac{\|v\|_2^2}{2v^Tx}x - \frac{\alpha\|v\|_2^2}{2v^Tx}e_1$$

i.e. v is a linear combination of x and e_1.

We also make the following observation: For any $\beta \neq 0$

$$H(\beta v) = I - 2\frac{(\beta v)(\beta v)^T}{\|\beta v\|_2^2} = I - 2\frac{\cancel{\beta^2}\, vv^T}{\cancel{\beta^2}\,\|v\|_2^2}$$

$$= I - 2\frac{vv^T}{\|v\|_2^2} = H(v).$$

i.e. we can discard the constant $\frac{\|v\|_2^2}{2v^Tx}$.

So we are looking for a vector v of the form $\boxed{x - \alpha e_1}$

Now

① $v^Tx = (x - \alpha e_1)^Tx = \|x\|_2^2 - \alpha x_1$ ← first component

and ② $v^Tv = (x - \alpha e_1)^T(x - \alpha e_1) = \|x\|_2^2 - 2\alpha x_1 + \alpha^2$

$$\Rightarrow H(v)x = \left(I - 2\frac{vv^T}{\|v\|_2^2}\right)x = x - 2\frac{v^Tx}{\|v\|_2^2}v$$

using (1) & (2)

$$= x - \frac{2(\|x\|_2^2 - \alpha x_1)}{\|x\|_2^2 - 2\alpha x_1 + \alpha^2} \cdot (x - \alpha e_1)$$

(3)
$$= \left(1 - \frac{2(\|x\|_2^2 - \alpha x_1)}{\|x\|_2^2 - 2\alpha x_1 + \alpha^2}\right)x + \underline{\frac{2\alpha(\|x\|_2^2 - \alpha x_1)}{\|x\|_2^2 - 2\alpha x_1 + \alpha^2}}\, e_1$$

<u>Note</u> we want $H(v)x$ to be a scalar multiple of e_1 hence we need to make the coefficient of x above equal to zero, i.e. we want

$$1 - \frac{2(\|x\|_2^2 - \alpha x_1)}{\|x\|_2^2 - 2\alpha x_1 + \alpha^2} = 0$$

$$\Leftrightarrow \quad \|x\|_2^2 - \cancel{2\alpha x_1} + \alpha^2 = 2\|x\|_2^2 - \cancel{2\alpha x_1}$$

$$\Leftrightarrow \quad \alpha^2 = \|x\|_2^2 \Leftrightarrow \boxed{\alpha = \pm \|x\|_2}$$

with one of these two choices of α, we have

$$H(v)x = \frac{2\alpha(\|x\|_2^2 - \alpha x_1)}{\|x\|_2^2 - 2\alpha x_1 + \alpha^2}\, e_1 = \alpha e_1 .$$

We have completely solved the pb. and found <u>two</u> solutions for v!

$$\boxed{\begin{cases} \alpha = +\|x\|_2^2 \Rightarrow v = x - \|x\|_2 e_1 \Rightarrow H(v)x = \|x\|_2 e_1 \\ \alpha = -\|x\|_2^2 \Rightarrow v = x + \|x\|_2 e_1 \Rightarrow H(v)x = -\|x\|_2 e_1 . \end{cases}}$$

Ex. $x = (3,1,5,1)^T$, $\quad \|x\|_2 = \sqrt{36} = 6$

$$v = x - 6e_1 = \begin{pmatrix}3\\1\\5\\1\end{pmatrix} - 6\begin{pmatrix}1\\0\\0\\0\end{pmatrix} = \begin{pmatrix}-3\\1\\5\\1\end{pmatrix}, \quad \|v\|_2^2 = 36$$

$$H(v)x = \left(I - 2\,\frac{\begin{pmatrix}-3\\1\\5\\1\end{pmatrix}(-3,1,5,1)}{36}\right)\begin{pmatrix}3\\1\\5\\1\end{pmatrix}$$

$$= \begin{pmatrix}3\\1\\5\\1\end{pmatrix} - \frac{1}{\cancel{18}}\begin{pmatrix}-3\\1\\5\\1\end{pmatrix}\overbrace{(-9+1+25+1)}^{\cancel{18}} = \begin{pmatrix}6\\0\\0\\0\end{pmatrix}$$

$$= +\|x\|_2\, e_1 \quad \text{as expected.}$$

We can <u>explicitly</u> form $H(v)$

$$H(v) = I - 2\,\frac{\begin{pmatrix}3\\1\\5\\1\end{pmatrix}(3\ 1\ 5\ 1)}{36} = \frac{1}{54}\begin{bmatrix}27 & 9 & 45 & 9\\ 9 & 51 & -15 & -3\\ 45 & -15 & -21 & -15\\ 9 & -3 & -15 & 51\end{bmatrix}$$

In actual computations $H(v)$ is <u>never</u> explicitly formed since all the information about $H(v)$ is contained in the vector v.

Furthermore, since any nonzero multiple of v will also work, we can make the first component v_1 of v equal to 1, so it does not have to be stored.

Another important issue is the choice of sign in v. Recall: $v = x \pm \|x\|_2 e_1$. We have to be careful if $x \approx e_1$. To avoid loss of significance due to subtraction of nearly equal numbers, we take

$$\boxed{v = x + \operatorname{sign}\{x_1\}\, \|x\|_2\, e_1}$$

Algorithm (Householder vector) Given $x \in \mathbb{R}^n$, this program computes $v \in \mathbb{R}^n$ with $v_1 = 1$ and $\beta \in \mathbb{R}$ such that $H = I_n - \beta v v^T$ is the Householder matrix with

$H x = \|x\|_2 e_1$

$$\sigma = x(2:n)^T x(2:n) = x_2^2 + \cdots + x_n^2$$

$$v = \begin{bmatrix} 1 \\ x(2:n) \end{bmatrix}$$

if $\sigma = 0$

  $\beta = 0$

else

  $\mu = \sqrt{x_1^2 + \sigma} \quad (= \|x\|_2^2)$

  if $x_1 \le 0$

    $v_1 = x_1 - \mu$

  else

    $v_1 = \dfrac{-\sigma}{x_1 + \mu}$

  end

  $\beta = 2 v_1^2 / (\sigma + v_1^2) \quad \left(= \dfrac{2}{\|v\|_2^2} \right)$

  $v = v / v_1$

end

This algorithm requires $3n$ flops.

Again, a Householder matrix is not explicitly computed. Rather, its action on a vector x or a matrix is computed as follows

$$Hx = \left(I - 2\frac{vv^T}{\|v\|_2^2}\right)x = x - \beta(v^T x) v$$

Form $v^T x \longrightarrow 2n-1$ flops (recall $v_1 = 1$)

Form $\beta(v^T x) \longrightarrow 1$

Form $(\beta(v^T x))v \longrightarrow n$ flops

Form $x - (\beta(v^T x)v) \longrightarrow n$ Flops

$$\boxed{\text{Total } 4n \text{ Flops}}$$

To perform HB, where B is an $n \times m$ matrix apply H to each column of $B \Rightarrow 4nm$ Flops.

"Partial" Householder matrices

Suppose $v \in R^n$, first j components are zero $v = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \hline \tilde{v} \end{pmatrix}$, $\tilde{v} \in R^{n-j}$

$$H(v) = I - \frac{2vv^T}{\|v\|_2^2} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} - \frac{2}{\|\tilde{v}\|_2^2} \begin{array}{c|cc|c} & j & n-j & \\ & \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \tilde{v}\tilde{v}^T \end{array}\right. & \left.\right] & \begin{array}{c} j \\ n-j \end{array} \end{array}$$

$$= \left[\begin{array}{c|c} I_j & 0 \\ \hline 0 & I_{n-j} - \frac{2\tilde{v}\tilde{v}^T}{\|\tilde{v}\|_2^2} \end{array}\right] = \begin{array}{cc} \left[\begin{array}{c|c} I_j & 0 \\ \hline 0 & H(\tilde{v}) \end{array}\right] & \begin{array}{c} j \\ n-j \end{array} \end{array}$$

where $H(\tilde{v})$ is $(n-j) \times (n-j)$ Householder matrix.

Theorem Let A be an $m \times n$ real matrix with $m \geq n$ and assume $\operatorname{rank}(A) = n$. We say A has full column rank. Then there exist an $m \times m$ orthogonal matrix Q and an upper triangular matrix R such that $A = QR$.

proof.

The proof, which is an algorithm, proceeds by constructing Householder matrices for columns of A.

To begin, let a_1 be the first column of A.
Let v_1 be the vector such that $H(v)a_1 = \pm \|a_1\|_2 e_1$
Here the sign depends on the choice of sign in v_1.

$$\Rightarrow \underbrace{H(v_1)}_{H_1} A = \left[\begin{array}{c|c|c} \pm\|a_1\|_2 & \times & \times \;\text{—}\; \times \\ \hline 0 & & \\ | & \tilde{a}_2 & \\ 0 & & \end{array}\right] \qquad \begin{array}{l} \tilde{a}_2 \in R^{m-1} \\ H_1 \text{ is } m \times m \end{array}$$

Note unlike in Gauss Elimination, the first row of A has changed.

In the 2nd step, we choose $v_2 = \begin{bmatrix} 0 \\ \hline \tilde{v}_2 \end{bmatrix}$

Such that $H(\tilde{v}_2)\tilde{a}_2 = \pm \|\tilde{a}_2\|_2 e_1$

$$\text{let } H_2 = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & H(\tilde{v}_2) \end{array}\right]$$

Then

$$H_2 H_1 A = \begin{bmatrix} \pm\|a_1\|_2 & \times & \text{—} & \times & \times \\ 0 & \pm\|\tilde{a}_2\|_2 & \times & \text{—} & \times \\ | & 0 & \boxed{\begin{matrix} \\ \tilde{a}_3 \\ \\ \end{matrix}} & \text{—} & \times \\ | & | & & & | \\ 0 & 0 & & & \times \end{bmatrix}$$

Note that given the structure of H_2, the first row of $H_1 A$ is not affected.

Proceeding this way, we eventually arrive at

Case 1 $\quad m > n$

$$H_n \cdots H_1 A = \begin{array}{c} n \\ \begin{bmatrix} \tilde{R} \\ 0 \end{bmatrix} \end{array} \begin{array}{l} n \\ m-n \end{array} = R$$

$$\Rightarrow \quad A = H_1^{-1} H_2^{-1} \cdots H_n^{-1} R = \underbrace{H_1 H_2 \cdots H_n}_{} R$$

$$\qquad \parallel \qquad = Q R$$

Q is orthogonal since $QQ^T = H_1 H_2 \cdots H_n (H_1 \cdots H_n)^T$

$$= H_1 H_2 \cdots H_n H_n^T \cdots H_1^T$$
$$= H_1 H_2 \cdots H_n H_n \cdots H_1$$
$$= I$$

Case 2 $\quad m = n$. Note we don't need H_n.

$$H_{n-1} \cdots H_1 A = R \Rightarrow A = \underbrace{H_1 \cdots H_{n-1}}_{Q} R = QR$$

work estimate.

First step: construction of $v_1 \longrightarrow 3m$ Flops

mult. $H_1 A \longrightarrow 4m \times \#$ of columns $= 4mn$

2nd step construction of $\tilde{v}_2 \longrightarrow 3(m-1)$ Flops

mult. $H_2 H_1 A \longrightarrow 4(m-1)(n-1)$

So total work for $m > n$ is

$$\underbrace{3m + 4m(n-1)}_{H_1 A} + \underbrace{3(m-1) + 4(m-1)(n-2)}_{H_2 H_1 A} + \cdots$$

$$+ \underbrace{3(m-n+2) + 4(m-n+2)(1)}_{H_{n-1} \cdots H_1 A} + \underbrace{3(m-n+1) + 0}_{H_n \cdots H_1 A}$$

Adds to $2mn^2 - \frac{2}{3}n^3$ + lower order terms

In case $\underline{m = n}$

work $\cong \frac{4n^3}{3}$ about twice that of LU factorization.

<u>Application to solving the linear system $Ax = b$</u>

Suppose $A \in \mathbb{R}^{n \times n}$ and is invertible. Then

$$A = QR \quad \text{with } u_{ii} \neq 0, \ i = 1, \ldots, n$$

$$Ax = b \Rightarrow QRx = b, \quad Q = H_1 \cdots H_{n-1}$$

1) $Rx = Q^T b = H_{n-1} \cdots H_1 b$

Each multiplication by H_i requires $4n$ Flops.

So total to compute $Q^T b$ is $4n(n) = 4n^2$

2) Solve for x from $Rx = Q^T b$ using Back substitution; requires n^2 flops.

Properties of the QR factorization

Theorem 5.2.1 If $A = QR$ is a QR factorization of a full column rank matrix $A \in \mathbb{R}^{m \times n}$ and $A = [a_1, a_2, \ldots, a_n]$ and $Q = [q_1, \ldots, q_m]$ are column partitionings of A and Q, then

$$\operatorname{span}\{a_1, \ldots, a_k\} = \operatorname{span}\{q_1, \ldots, q_k\}, \quad k = 1:n.$$

In particular, if $Q_1 = Q(1:m, 1:n)$ and $Q_2 = Q(1:m, n+1:m)$, then

$$\operatorname{ran}(A) = \operatorname{ran}(Q_1)$$
$$\operatorname{ran}(A)^\perp = \operatorname{ran}(Q_2),$$

$$Q = \begin{array}{c} \left[\begin{array}{c|c} Q_1 & Q_2 \end{array}\right] m \\ \quad n \quad\;\; m-n \end{array}$$

and

$$A = Q_1 R_1 \quad \text{with} \quad R_1 = R(1:n, 1:n).$$

Proof.

$$a_1 = r_{11} q_1$$
$$a_2 = r_{12} q_1 + r_{22} q_2$$
$$\vdots$$
$$a_k = r_{1k} q_1 + \cdots + r_{kk} q_k = \sum_{i=1}^{k} r_{ik} q_i.$$

$$\left[\begin{array}{c|c|c|c} a_1 & a_2 & \cdots & a_n \end{array}\right] = \left[\begin{array}{c|c|c|c} q_1 & q_2 & \cdots & q_m \end{array}\right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ 0 & & \ddots & r_{nn} \\ & & 0 & \end{bmatrix}$$

$\Rightarrow \operatorname{span}\{a_1, \ldots, a_k\} \subseteq \operatorname{span}\{q_1, \ldots, q_k\}$. However, since

$\operatorname{rank}(A) = n$, $\dim \operatorname{span}\{a_1, \ldots, a_k\} = k$ (columns are lin. indep.)
$= \dim \operatorname{span}\{q_1, \ldots, q_k\}$

$\Rightarrow \operatorname{span}\{a_1, \ldots, a_k\} = \operatorname{span}\{q_1, \ldots, q_k\}$. ∎

Note This also shows that $r_{ii} \neq 0$ and if we choose $v = x \stackrel{+}{-} \|x\| e_1$, then $r_{ii} > 0$.

Algorithm 5.2.1 (Householder QR) Given $A \in \mathbb{R}^{m \times n}$ with $\boxed{m \geq n}$, the following algorithm finds Householder matrices $H_1, \ldots, H_n$ such that if $Q = H_1 \cdots H_n$, then $Q^T A = R$ is upper triangular. The upper triangular part of A is overwritten by the upper triangular part of R and components $j+1:m$ of the j-th Householder vector $v^{(j)}$ are stored in $A(j+1:m, j)$, $j < m$.

For $j = 1:n$

  $[v, \beta] = \text{house}(A(j:m, j))$

  $A(j:m, j:n) = (I_{m-j+1} - \beta v v^T) A(j:m, j:n)$

  if $j < m$

    $A(j+1:m, j) = v(2:m-j+1)$

  end

end

This algorithm requires $2n^2(m - \frac{n}{3})$ flops.

$\frac{4n^3}{3}$ when $m = n$ = twice LU factorization.

$$A \rightarrow \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ v_2^{(1)} & r_{22} & \cdots & r_{2n} \\ v_3^{(1)} & & \ddots & \\ & & & r_{nn} \\ & & & v_{n+1}^{(n)} \\ \vdots & & & \vdots \\ v_m^{(1)} & \cdots & \cdots & v_m^{(n)} \end{bmatrix}$$

Note If Q is required then it can be accumulated using (5.1.5).
Cost: $4(m^2 n - m n^2 + \frac{n^3}{3})$ flops.

Note we can store $\beta_1, \ldots, \beta_n$ separately as a vector.

The fact that $\text{ran}(A)^{\perp} = \text{ran}(Q_2)$ follows from the fact that the columns of Q are orthogonal. $\square$

<u>Defn.</u> The factorization $A = Q_1 R_1$ is called the "Thin" QR factorization of A.

<u>Theorem 5.2.2</u> Suppose $A \in \mathbb{R}^{m \times n}$ has full column rank. The Thin QR factorization $A = Q_1 R_1$ is unique where $Q_1 \in \mathbb{R}^{m \times n}$ has orthonormal columns and R_1 is upper triangular with positive diagonal entries. Moreover, $R_1 = G^T$ where G is the lower triangular choleski factor of $A^T A$.

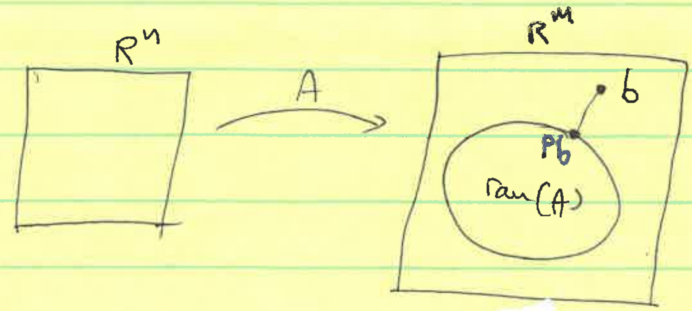
<u>proof.</u> Since A has full column rank, $A^T A$ is s.p.d. $\Rightarrow A^T A = G G^T$. Also, since $A^T A = R_1^T Q_1^T Q_1 R_1 = R_1^T I_{n \times n} R_1 = R_1^T R_1$, we see that $G = R_1^T$ is a cholesky factor of $A^T A$. This factor is unique by theorem 4.2.5. Since $Q_1 = A R_1^{-1}$, it follows that Q_1 is also unique. $\square$

The full-rank Least Squares (LS) problem.

Consider the system $Ax=b$ where $m \geq n$. → overdetermined system

If $b \notin \text{ran}(A)$, then system cannot be solved in the classical sense.

Alternatively, we may try to find $x \in \mathbb{R}^n$ s.t. $\|Ax-b\|_p$ is minimized for some $p \geq 1$.



It can be shown that this problem has always a ~~unique~~ solution. But depending on p, we may get a different solution.

Ex. $A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$, $b_1 \geq b_2 \geq b_3 \geq 0$. Then

for $p=1$, $x_{opt} = b_2$

$p=2$, $x_{opt} = (b_1+b_2+b_3)/3$

$p=\infty$, $x_{opt} = \dfrac{b_1+b_3}{2}$.

For $p=2$, x_{opt} is called the least-squares solution and the problem:

Find $x \in \mathbb{R}^n$ such that $\|Ax-b\|_2$ is minimized

is called the least-squares problem (LS).

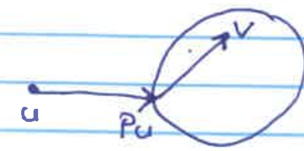
Existence

Theorem (The projection Theorem) Let $V, \langle \cdot, \cdot \rangle$ be a Hilbert space and S a closed, convex subset of V.

(i) There exists a well-defined operator $P : V \to S$, the "projection" operator such that for any $u \in V$

$$\|u - Pu\| = \inf_{v \in S} \|u - v\|.$$

(ii) The projection Pu of u satisfies



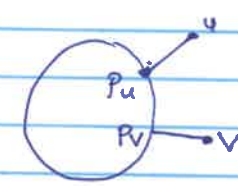
$$\langle Pu - u, v - Pu \rangle \geq 0 \quad \forall v \in S.$$

Conversely, if there is an element $z \in S$ such that

$$\langle z - u, v - z \rangle \geq 0, \quad \text{then} \quad z = Pu.$$

(iii) The projection operator P is <u>non expansive</u> in the sense that

$$\|Pu - Pv\| \leq \|u - v\|, \quad \forall u, v \in V$$

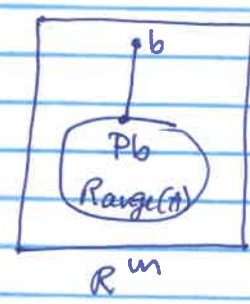


(iv) If S is a linear subspace of V, then the following orthogonality condition holds:

(*) $$\langle Pu - u, v \rangle = 0 \quad \forall v \in S.$$

Conversely, if (*) holds, then S is a linear subspace of V. $\square$

R^m equipped with the Euclidean dot product $\langle u,v\rangle = u^T v$ is a Hilbert space. Range(A) is a subspace and is both closed (R^m is finite dimensional) and convex. By the projection theorem, there exists a unique $Pb \in \text{Range}(A)$ such that



$$\|Pb - b\|_2 = \inf_{v \in \text{Range}(A)} \|b - v\|_2 .$$

Since $Pb \in \text{Range}(A)$, clearly there are solutions x_{LS} to $Ax_{LS} = Pb$. Such x_{LS} are also solutions to the Least-Squares problem, since

$$\|b - Ax_{LS}\|_2 = \|b - Pb\|_2 = \inf_{v \in \text{Range}(A)} \|b - v\|_2 = \inf_{v \in \text{Range}(A)} \|b - Ax\|_2$$

Furthermore, if $\text{rank}(A) = n$, then the solution is unique.

<u>Existence via minimization</u>

Consider the function $\phi : R^n \to R$ given by

$$\phi(x) = \frac{1}{2}\|Ax - b\|_2^2 .$$

It is true that minimizing ϕ is equivalent to minimizing $\|Ax - \|_2$. Now, some algebra shows that

$$\phi(x) = \frac{1}{2} x^T A^T A x - x^T A^T b + \frac{1}{2} b^T b .$$

ϕ is a quadratic form and is clearly differenciable with gradient

$$\nabla\phi(x) = A^T A x - A^T b .$$

If x is a minimizer of ϕ, then $\nabla\phi(x) = 0$. These are the Cauchy–Euler equations.

For Existence, we need the converse of this.

If A has full rank, n, then it is an easy exercise to show that A^TA is symmetric and positive definite. Using some results from the theory of optimization, we can infer the existence of a unique minimizer. However, we can do this directly exploiting the fact that ϕ is a quadratic.

$$\text{let} \quad x = z + (A^TA)^{-1}A^Tb.$$

(we know from $\nabla\phi(x)=0$ that $(A^TA)^{-1}A^Tb$ is the solution.)

we have

$$\phi(x) = \underbrace{\frac{1}{2} z^T(A^TA)z}_{=\frac{1}{2}\|Az\|_2^2} - \frac{1}{2} b^TA(A^TA)^{-1}A^Tb + \frac{1}{2}\|b\|_2^2.$$

The last two terms are fixed whereas $\frac{1}{2}z^T(A^TA)z = \frac{1}{2}\|Az\|_2^2$ nonnegative. Hence ϕ is minimized by $z=0$. If A has full rank then this is the only minimizer.

<u>Solution via normal Equations.</u>

The ~~system~~ $\nabla\phi(x_{LS}) = (A^TA)x_{LS} - A^Tb = 0$

are called the normal equations. One way of calculating x_{LS} is to use the Choleski factorization

$$A^TA = LL^T.$$

Note that L is an $n\times n$ matrix. So the factorization is not costly. However this method is not recommended if the condition number of A is large. The problem will be compounded upon taking A^TA.

Solution of the Full-rank Least-Squares problem via QR

Since Q is orthogonal, we have

$$\|Ax-b\|_2^2 = \|QRx-b\|_2^2 = \|Q(Rx-Q^Tb)\|_2^2$$
$$= \|Rx-Q^Tb\|_2^2 .$$

let $\tilde{R}$ be the $n\times n$ upper triangular matrix such that

$$R = \begin{array}{c} \left[\begin{array}{c} \tilde{R} \\ \hline 0 \end{array}\right] \end{array}\begin{array}{l} n \\ m-n \end{array}$$ and partition as $$\left[\begin{array}{c} \tilde{b}_1 \\ \hline \tilde{b}_2 \end{array}\right]$$

we have

$$Rx - Q^Tb = \left[\begin{array}{c} \tilde{R}x-\tilde{b}_1 \\ \hline -\tilde{b}_2 \end{array}\right] \Rightarrow \underbrace{\|Rx-Q^Tb\|_2^2}_{\|Ax-b\|_2^2} = \|\tilde{R}x-\tilde{b}_1\|_2^2 + \|\tilde{b}_2\|_2^2 .$$

It is clear that the quantity $\|Ax-b\|_2$ is minimized when $\tilde{R}x-\tilde{b}_1=0$. Now $\tilde{R}$ is nonsingular since the columns of R are linearly independent in the full-rank case.

QR least-Squares

1) Factor A into QR
2) Form $\tilde{b} = Q^Tb$
3) Extract the first n components of $\tilde{b}$
4) Solve for x_{LS} from $\tilde{R}x_{LS} = \tilde{b}_1$ by back substitution.

The Singular value Decomposition

Theorem 2.5.1 If $V_1 \in \mathbb{R}^{n \times r}$ has orthonormal columns, then there exists $V_2 \in \mathbb{R}^{n \times (n-r)}$ such that the matrix $V = [V_1 \ V_2]$ is orthogonal. Note that $\text{Range}(V_1)^\perp = \text{ran}(V_2)$.

proof.

Note that $V_1 \in \mathbb{R}^{n \times r}$ has orth. cols. $\Rightarrow$ $r \le n$. If $r = n$, then $V = V_1$. If $r < n$, then the rows of V_1 must be linearly dependant $\Rightarrow \exists\, b \in \mathbb{R}^n$, $b \ne 0$, such that $b^T V_1 = 0$. But this is equivalent to $b^T c_i = 0$ for every column c_i of V_1, $i = 1, \dots, n$. So we can make $\frac{b}{\|b\|_2}$ the first column of V_2. proceeding this way, it follows that $V = [V_1 \ V_2]$ with $V^T V = I$.

It is clear that if $v_1 \in \text{ran}\{V_1\}$ and $v_2 \in \text{ran}(V_2)$, then $v_1 \perp v_2$. Thus $v_1 \in \text{ran}(V_2)^\perp$ and $v_2 \in \text{ran}(V_1)^\perp \Rightarrow \text{ran}(V_1)^\perp = \text{ran}(V_2)$ and $\text{ran}(V_2)^\perp = \text{ran}(V_1)$.

Lemma (i) If Q is orthogonal, then $\|Qx\|_2 = \|x\|_2 \quad \forall x$.

(ii) If $Q \in \mathbb{R}^{m \times m}$ and $Z \in \mathbb{R}^{n \times n}$ are orthogonal, then

$$\|QAZ\|_2 = \|A\|_2 \qquad \forall A \in \mathbb{R}^{m \times n} \text{ and}$$

$$\|QAZ\|_F = \|A\|_F \qquad \forall A \in \mathbb{R}^{m \times n}.$$

proof. (i) $\|x\|_2^2 = x^T x = x^T \underbrace{Q^T Q}_{I} x = \|Qx\|_2^2$.

(ii) $$\|QAZ\|_2^2 = \sup_{x \ne 0} \frac{\|QAZx\|_2^2}{\|x\|_2^2} = \sup_{x \ne 0} \frac{\|AZx\|_2^2}{\|x\|_2^2} = \sup_{x \ne 0} \frac{\|Ay\|_2^2}{\|Z^T y\|_2^2}$$

$$= \sup_{y \ne 0} \frac{\|Ay\|_2^2}{\|y\|_2^2} = \|A\|_2^2.$$

$$\|QA\|_F^2 = \|[Qa_1 | \cdots | Qa_n]\|_F^2 = \|Qa_1\|_2^2 + \cdots + \|Qa_n\|_2^2 = \|a_1\|_2^2 + \cdots + \|a_n\|_2^2$$
$$= \|A\|_F^2. \quad \text{Similarly, } \|AZ\|_F = \|A\|_F.$$

Theorem 2.5.2 (Singular value Decomposition (SVD)). If A is a real $m\times n$ matrix, then there exist orthogonal matrices $U=[u_1,\dots,u_m]\in R^{m\times m}$ and $V=[v_1,\dots,v_n]\in R^{n\times n}$ such that

$$U^T A V = \operatorname{diag}(\sigma_1,\dots,\sigma_p)\in R^{m\times n},\quad p=\min\{m,n\}$$

where $\sigma_1 \geq \dots \geq \sigma_p \geq 0$.

proof.

let $x\in R^n$ and $y\in R^m$ be such that $\|y\|_2=\|x\|_2=1$ and $Ax=\sigma_1 y$ with $\sigma_1=\|A\|_2$. To show this, we use the fact that in the definition $\|A\|_2=\sup_{\|x\|_2=1}\|Ax\|_2$ the supremum is attained for some x with $\|x\|_2=1$.

let $y=\dfrac{Ax}{\|A\|_2} \Rightarrow Ax=\|A\|_2 y$

Also

$$\|y\|_2=\frac{\|Ax\|_2}{\|A\|_2}=\frac{\|A\|_2}{\|A\|_2}=1.$$

By Theorem 2.5.1, there exist $V_2\in R^{n\times(n-1)}$ such that $V=[x\,|\,V_2]\in R^{n\times n}$ is orthogonal and $U_2\in R^{m\times(m-1)}$ such that $U=[y\,|\,U_2]\in R^{m\times m}$ is orthogonal. Now

$$U^TAV=\begin{bmatrix} y^T \\ \hline U_2^T\end{bmatrix}\begin{bmatrix} A \end{bmatrix}\begin{bmatrix} x & V_2\end{bmatrix}=\begin{bmatrix} y^T \\ \hline U_2^T\end{bmatrix}\begin{bmatrix} Ax & AV_2\end{bmatrix}$$

$$=\left[\begin{array}{c|c} y^TAx & y^TAV_2 \\ \hline U_2^TAx & U_2^TAV_2\end{array}\right].$$

Now $Ax=\|A\|_2 y$ and since y is orthogonal to every col. of U_2, $U_2^TAx=0$.

$$\Rightarrow U^TAV=\left[\begin{array}{c|c} \sigma_1 & w^T \\ \hline 0 & B\end{array}\right]\equiv A_1$$

where $\sigma_1=y^TAx=y^T\|A\|_2 y=\|A\|_2\|y\|_2^2=\|A\|_2$.

We next show that $w = 0$. Now

$$\left\| A_1 \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2^2 = \left\| \begin{bmatrix} \sigma_1^2 + w^T w \\ Bw \end{bmatrix} \right\|_2^2 = (\sigma_1^2 + w^T w)^2 + \|Bw\|_2^2$$

$$\geq (\sigma_1^2 + \|w\|_2^2)^2.$$

Thus

$$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \geq \frac{\left\| A \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2} \geq (\sigma_1^2 + \|w\|_2^2)^{1/2}.$$

But since U and V are orthogonal,

$$\|A_1\|_2 = \|U^T A V\|_2 = \|A\|_2 = \sigma_1 .$$ This means $w = 0$.

$$\Rightarrow \quad U^T A V = \left[\begin{array}{c|c} \sigma & 0 \\ \hline 0 & \mathbf{B} \end{array} \right] \quad \text{with } B \in R^{m-1 \times n-1}.$$

The argument can now be continued to get the desired result.

Remark (i) we use the fact that the product of any number of orthogonal matrices is also orthogonal.

(ii) $\|A\|_2^2 = \sigma^2 + \|B\|_2^2 \Rightarrow \|B\|_2 \leq \|A\|_2 \Rightarrow$ ordering of $\sigma_1 \geq \cdots \geq \sigma_p$.

$\sigma_1, \ldots, \sigma_p$ are the singular values of A

$u_1, \ldots, u_m$ are the left singular vectors of A

$v_1, \ldots, v_n$ are the right singular vectors of

Some properties of the SVD

Lemma Let $U^T A V = \Sigma$ be the SVD of A. Suppose for some $r \in \{1, \dots, p\}$ we have $\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0$ then $r = \text{rank}(A)$.

proof We use the following facts:

(i) $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$

(ii) If A is invertible, then $\text{rank}(AB) = \text{rank}(BA) = \text{rank}(B)$.
Here it is assumed that AB or BA is defined.

Clearly $\text{rank}(\Sigma) = r$. The result now follows upon application of (ii). $\boxtimes$

The next result shows that any matrix A can be expressed as the sum of rank-one matrices.

Lemma Let $U^T A V = \Sigma$ be the SVD of A. Then

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^T$$

where u_k and v_k are the k-th columns of U and V respectively.

proof. We have $A = U \Sigma V^T$, which can be written in partitioned form

$$A = \begin{array}{cc} \begin{array}{cc} r & m-r \end{array} & \begin{array}{cc} r & n-r \end{array} \\ \left[\begin{array}{c|c} U_1 & U_2 \end{array}\right] & \left[\begin{array}{c|c} \tilde{\Sigma} & 0 \\ \hline 0 & 0 \end{array}\right] \end{array} \begin{array}{c} \left[\begin{array}{c} V_1^T \\ \hline V_2^T \end{array}\right] \end{array} \begin{array}{l} r \\ n-r \end{array}$$

$$= \left[\begin{array}{c|c} U_1 & U_2 \end{array}\right] \left[\begin{array}{c} \tilde{\Sigma} V_1^T \\ \hline 0 \end{array}\right] = U_1 \tilde{\Sigma} V_1^T.$$

$$= \left[\begin{array}{c|c|c} u_1 & \cdots & u_r \end{array}\right] \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix} \left[\begin{array}{c} v_1^T \\ \hline \vdots \\ \hline v_r^T \end{array}\right] = \left[\begin{array}{c|c|c} u_1 & \cdots & u_r \end{array}\right] \left[\begin{array}{c} \sigma_1 v_1^T \\ \hline \vdots \\ \hline \sigma_r v_r^T \end{array}\right].$$

Now for $i=1,\dots,m$, $j=1,\dots,n$

$$a_{ij} = \sum_{k=1}^{r} u_{ik}\,\sigma_k (v_k^T)_j = \sum_{k=1}^{r} \sigma_k\, u_{ik} v_{jk} = \sum_{k=1}^{r} \sigma_k (u_k v_k^T)_{ij},$$

which means

$$A = \sum_{k=1}^{r} \sigma_k\, u_k v_k^T. \quad \blacksquare$$

Remark The decomposition $U_1^T A V_1 = \tilde{\Sigma}$ is called the "Thin" SVD of A. Note that U_1 is $m\times r$ and V_1 is $r \times n$ with orthonormal columns.

Lemma Let $U^T A V = \Sigma$ be the SVD of A. Then

(i) $\text{Range}(A) = \text{span}\{u_1,\dots,u_r\}$

(ii) $\text{Null}(A) = \text{span}\{v_{r+1},\dots,v_n\}$

proof. (i) Suppose $y \in \text{Range}(A)$, i.e. $y = Ax$, $x \in \mathbb{R}^n$. Then,

$$y = Ax = \sum_{k=1}^{r} \sigma_k\, u_k v_k^T x = \sum_{k=1}^{r} (\sigma_k v_k^T x)\, u_k,$$

i.e. y is a linear combination of $u_1,\dots,u_r$.

(ii) In the language of orthogonal complements, we have

$$\text{span}\{v_1,\dots,v_r\}^{\perp} = \text{span}\{v_{r+1},\dots,v_n\}.$$

Using the Thin SVD $A = U_1 \tilde{\Sigma} V_1^T$,

if $Ax=0$, then $U_1 \tilde{\Sigma} V_1^T x = 0$ which means that $V_1^T x = 0$ since $\tilde{\Sigma}$ is invertible and U_1 has orthonormal columns. But $V_1^T x = 0$ implies that x is orthogonal to $\text{span}\{v_1,\dots,v_r\}$, i.e. it belongs to $\text{span}\{v_{r+1},\dots,v_n\}$.

Conversely, if $x \in \text{span}\{v_{r+1},\dots,v_n\}$. Then clearly $V_1 x = 0$. Hence $Ax = U_1 \tilde{\Sigma} V_1^T x = 0$. $\blacksquare$

An interesting result concerning invertible matrices is the following: Let $R^{n\times n}$ denote the vector space of all $n\times n$ matrices. Equip with the topology of some norm $\|\cdot\|$. Since $R^{n\times n}$ is finite dimensional, it does not matter which norm is chosen. Then the set of all invertible matrices is open. One equivalent fact is that if A is invertible, then there exists $\epsilon>0$ such that the open ball $\{B\in R^{n\times n} \mid \|B-A\|<\epsilon\}$ contains only invertible matrices. As a consequence, $\inf_{B \text{ singular}} \|B-A\| \geq \epsilon > 0$. Indeed, it can be shown (Exercise)

that $\inf_{B \text{ singular}} \|B-A\| = \sigma_n$, the smallest singular value of A. Note that σ_n is positive since A is invertible.

It turns out that this is a special case of the following more general result.

Theorem. Let $U^TAV=\Sigma$ be the SVD of A. For $k=1,\dots,r-1$, $r=\text{rank}(A)$, let

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T .$$

Then

$$\min_{\text{rank}(B)=k} \|A-B\|_2 = \|A-A_k\|_2 = \sigma_{k+1} .$$

proof. We can write

$$U^T = \sum_{j=1}^{m} e_j u_j^T , \quad \{e_1,\dots,e_m\} \text{ canonical basis of } R^m$$

$$V = \sum_{l=1}^{n} v_l e_l^T \quad \{e_1,\dots,e_n\} \text{ " " " } R^n.$$

Hence

$$U^TA_kV = \left(\sum_{j=1}^{m} e_j u_j^T\right)\left(\sum_{i=1}^{k}\sigma_i u_i v_i^T\right)\left(\sum_{l=1}^{n} v_l e_l^T\right)$$

$$= \sum_{i=1}^{k}\sigma_i \left(\sum_{j=1}^{m} e_j \underbrace{u_j^T u_i}_{\delta_{ij}}\right)\left(\sum_{l=1}^{n} \underbrace{v_i^T v_l}_{\delta_{il}} e_l^T\right)$$

$$= \sum_{i=1}^{k} \sigma_i e_i e_i^T$$

$$= \mathrm{diag}\{\sigma_1, \ldots, \sigma_k, 0 \text{---} 0\} \in \mathbb{R}^{m \times n}.$$

Thus $\mathrm{rank}(A_k) = k$. Also,

$$U^T (A - A_k) V = \mathrm{diag}\{0, \ldots, 0, \sigma_{k+1}, \ldots, \sigma_p\},$$

and hence

$$\|A - A_k\|_2 = \sigma_{k+1}.$$

Now suppose $\mathrm{rank}(B) = k$. We will show that $\|A - B\|_2 \geq \sigma_{k+1}$. We can find orthonormal vectors $x_1, \ldots, x_{n-k}$ in $\mathbb{R}^n$ so that $\mathrm{Null}(B) = \mathrm{span}\{x_1, \ldots, x_{n-k}\}$ in view of

$$\mathbb{R}^n = \mathrm{Null}(B) \oplus \mathrm{range}(B^T).$$

$\mathrm{span}\{x_1, \ldots, x_{n-k}\}$ and $\mathrm{span}\{v_1, \ldots, v_{k+1}\}$ are both subspaces of $\mathbb{R}^n$ and since $n - k + k + 1 = n + 1 > n$,

$$\mathrm{span}\{x_1, \ldots, x_{n-k}\} \cap \mathrm{span}\{v_1, \ldots, v_{k+1}\} \neq \{0\}.$$

Let z be in this intersection with $\|z\|_2 = 1$. Now z belongs to $\mathrm{Null}(B)$, hence $Bz = 0$. Also

$$Az = \sum_{i=1}^{k+1} \sigma_i u_i v_i^T z = \sum_{i=1}^{k+1} \sigma_i (v_i^T z) u_i,$$

hence

$$\|A - B\|_2^2 \geq \|(A - B) z\|_2^2 = \|Az\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 (v_i^T z)^2$$

$$\geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} (v_i^T z)^2.$$

It turns out that $\sum_{i=1}^{k+1} (v_i^T z)^2 = 1$. Indeed, since $\{v_1, \ldots, v_{k+1}\}$ is an orthonormal set,

$$z = \sum_{i=1}^{k+1} (v_i^T z) v_i \quad \text{and} \quad 1 = \|z\|_2^2 = \sum_{i=1}^{k+1} (v_i^T z)^2.$$ Using this in the above, we get $\|A - B\|_2 \geq \sigma_{k+1}$ and this concludes the proof. $\blacksquare$

Relationship of singular values to eigenvalues

From $U^T A V = \Sigma$, we have

$$A^T A = V \Sigma^T U^T U \Sigma V^T = V(\Sigma^T \Sigma) V^T.$$

$$= V \left[\begin{array}{ccc|c} \sigma_1^2 & & & \\ & \ddots & & 0 \\ & & \sigma_r^2 & \\ \hline & 0 & & 0 \end{array}\right] V.$$

This shows that $A^T A$ and $\Sigma^T \Sigma$ are similar. Hence the eigenvalues of $\Sigma^T \Sigma$ are $\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0$ and these are equal to the eigenvalues of $A^T A$. Hence the singular values of A squared are the eigenvalues of $A^T A$.

Application of SVD to image compression

The original image (top left) consists of a 200×320 matrix of pixels requiring 64,000 bytes of memory.

$$A = \sum_{i=1}^{200} \sigma_i u_i v_i^T \quad \text{is the SVD of the matrix.}$$

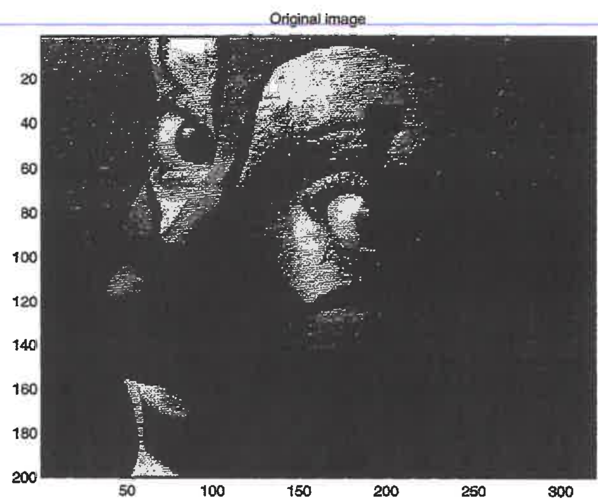
The next 3 pictures show the result of keeping only the largest k singular values for $k = 3, 10, 20$.

we measure the relative error by σ_{k+1}/σ_k

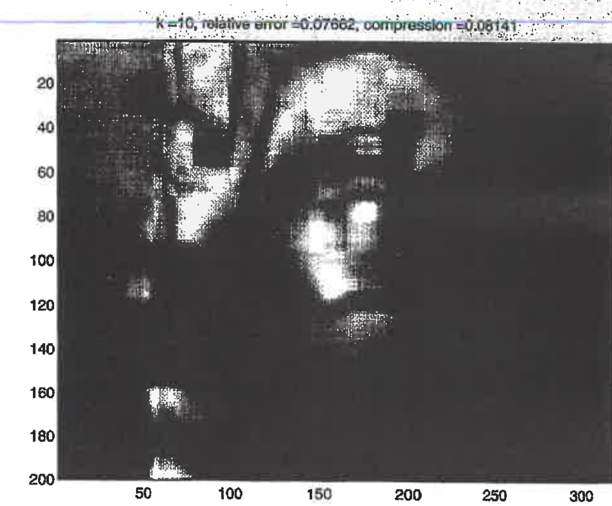
" " " compression ratio by $520K/64{,}000$.

$k=3 \Rightarrow$ error $= .155$, $k=10 \Rightarrow$ error $= .077$, $k=20 \Rightarrow$ error $= .04$

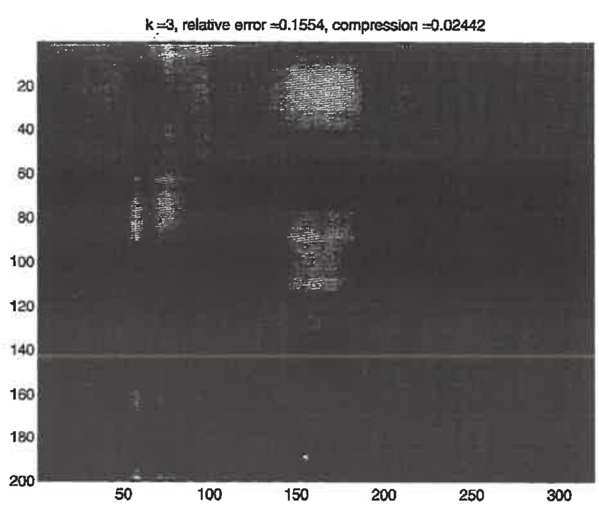
$k=3$ compression $= .024$, $k=10 \Rightarrow$ compression $= .081$, $k=2 \Rightarrow$ compr. $= .16$



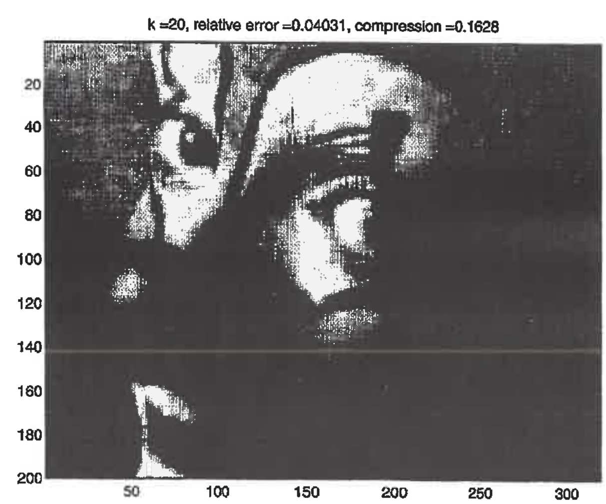
(a)



(c)



(b)



(d)

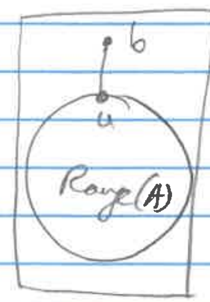
rank(A) < n ≤ m

Application to the least-squares pb.: the rank deficient case

(LS) Find $u \in R^n$ such that $\|Au - b\|_2 = \inf_{x \in R^n} \|Ax - b\|_2$

The (LS) problem always has a solution. If rank(A) = n then the solution is unique. If rank(A) < n, then there are infinitely many solutions.

proof



write $b = b_1 + b_2$

$b_1 \in Null(A)$, $b_2 \in Range(A)$

$$\|Ax - b\|_2^2 = \|Ax - b_1 - b_2\|_2^2$$

$$= \|b_1 + (Ax - b_2)\|_2^2 = \|b_1\|_2^2 + \|Ax - b_2\|_2^2.$$

$\|b_1\|_2$ is independent of x. Hence $\|Ax - b\|_2^2$ is minimized when $\|Au - b_2\|_2^2 = 0$. $Au = b_2$.

This is possible since $b_2 \in Range(A)$.

Now if rank(A) = n, then the solution is unique. If rank(A) < n, then there are infinitely many u such that $Au = b_2$.

It can be shown that among these infinitely many solutions, there is one with smallest norm.

$$\|Ax-b\|_2^2 = \|U\Sigma V^T x - b\|_2^2$$

$$= \|U(\Sigma V^T x - \underbrace{U^T b}_{\tilde{b}})\|_2^2 = \|\Sigma V^T x - \tilde{b}\|_2^2$$

Recall

$$\Sigma V^T = \begin{bmatrix} \tilde{\Sigma} V_1^T \\ \hline 0 \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix} \quad , \text{ let } \tilde{b} = \begin{bmatrix} w \\ \hline z \end{bmatrix}$$

$$\Rightarrow \|Ax-b\|_2^2 = \left\| \begin{bmatrix} \tilde{\Sigma} V_1^T \\ \hline 0 \end{bmatrix} x - \begin{bmatrix} w \\ \hline z \end{bmatrix} \right\|_2^2$$

$$= \|z\|_2^2 + \|\tilde{\Sigma} V_1^T x - w\|_2^2$$

$\|Ax-b\|_2^2$ is minimized by setting $\tilde{\Sigma} V_1^T x - w = 0$

i.e. $\begin{bmatrix} \sigma_1 v_1^T \\ \vdots \\ \sigma_r v_r^T \end{bmatrix} x = w \Rightarrow$ (1) $\boxed{\sigma_i (v_i^T x) = w_i \quad , i = 1, \dots, r}$

Now note that $w_i = (U^T b)_i = u_i^T b \ , \ i = 1, \dots, r$ (2)

Also, the LS solution x belongs to R^n and $v_1, \dots, v_n$ form an orthonormal basis for R^n. Hence

(3) $x = \sum_{i=1}^{n} \alpha_i v_i$ where the coefficients α_i are given by $\alpha_i = v_i^T x$ (This is because $v_1, \dots, v_n$ are orthonormal)

Hence from (1) and (2)

$$\alpha_i = \frac{1}{\sigma_i} u_i^T b \ , \ i = 1, \dots, r$$

From (3) $\boxed{x = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i}$ is the LS solution.