The QR factorization

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

2) Used in solving the <u>least-squares</u> problem

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

<u>Defn</u> <u>outer product of two vectors</u> 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

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

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

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

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

<u>Defn.</u> Let $v \in \mathbb{R}^n$ $v \neq 0$. The matrix

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

is called a <u>Householder matrix</u>.

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

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

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

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

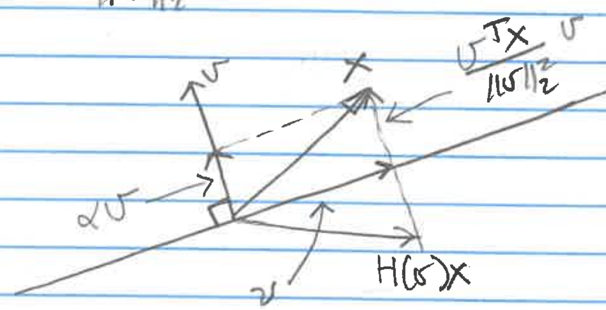
Also, $H(v)$ is <u>orthogonal</u>, i.e. $H(v)H(v)^T = I$

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

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

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

$H(v)$ 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. $\nu^Tv = 0$.

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

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

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

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

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

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 $x \neq 0$.

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

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

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 $\frac{\|v\|_2^2}{2v^T x}$.

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

Now

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

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

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

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

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

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

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

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

$= +\|x\|_2 e_1$ as expected.

We can explicitly 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 never 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

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

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

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

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

Form $(\beta(v^Tx))v \longrightarrow n$ flops

Form $x - (\beta(v^Tx)v) \longrightarrow n$ 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 $v \in R^n$, $v = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \hline \tilde{v} \end{pmatrix}$, $\tilde{v} \in R^{n-j}$
first j components ~~are zero~~

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

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

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.

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

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

In the 2nd step, we choose $v_2 = \left[\begin{array}{c} 0 \\ \hline \tilde{v}_2 \end{array}\right]$

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

$$H_2 H_1 A = \left[\begin{array}{ccccc} \pm\|a_1\|_2 & \times & \cdots & \times & \times \\ 0 & \pm\|\tilde{a}_2\|_2 & \times & \cdots & \times \\ \vdots & 0 & \boxed{\begin{array}{c} \\ \tilde{a}_3 \\ \\ \end{array}} & \cdots & \times \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & & \times \end{array}\right]$$

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 $m > n$

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

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

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

Q is orthogonal since
$$\begin{aligned} QQ^T &= H_1 H_2 \cdots H_n (H_1 \cdots H_n)^T \\ &= H_1 H_2 \cdots H_n H_n^T \cdots H_1^T \\ &= H_1 H_2 \cdots H_n H_n \cdots H_1 \\ &= I \end{aligned}$$

Case 2 $m = n$. 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 $v_1 \rightarrow 3m$ flops

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

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

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

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.

Application to solving the linear system $Ax = b$

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

$$A = QR \quad \text{with } r_{ii} \neq 0, \; i = 1, \dots, n$$

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

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.

The least squares problem: The full rank case.

Let A be an $m \times n$ matrix with $m > n$.
Typically $m >> n$.

Given an m-vector b, there is little chance that there will be a classical solution to $Ax = b$.
This motivates the following least-squares problem

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

Thm. Suppose $\operatorname{rank}(A) = n$. Then there is a unique solution u to the (LS) problem. It is given by

$$A^T A u = A^T b \quad \leftarrow \text{normal Equations} \quad \square$$

Finding the LS solution u using the normal equations is not recommended due to ill-conditioning of $A^T A$. Indeed, overdetermined systems contain many redundant equations $\Rightarrow$ ill-conditioning.
The problem is compounded upon forming $A^T A$.

Solution of Least-Squares pb. via QR. Full rank case

Lemma Let $A = QR$ be the QR factorization of A where $m \geq n$ and $\operatorname{rank}(A) = n$, i.e. the columns of A are linearly independent. Then the columns of R are linearly independent.

proof. Let $x \in \mathbb{R}^n$. Suppose $Rx = 0$. Note Rx is $\sum_{i=1}^{n} x_i r_i$ where $r_1, \dots, r_n$ are the cols. of R

Then from $A = QR$, we have $Rx = Q^T A x$

$0 = Rx = Q^T A x \Rightarrow Ax = Q0 = 0$. Now

$$Ax = \sum_{i=1}^{n} x_i a_i, \quad a_1, \dots, a_n \text{ being the columns of } A.$$

Since these are linearly independent, $x = 0$. Hence the columns of R are linearly independent.

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

$$= \|Rx - Q^T b\|_2^2 \quad , \text{ let } \tilde{b} = Q^T b$$

$\tilde R$ is $n \times n$ invertible.

$$\begin{array}{c} n \\ \begin{bmatrix} \tilde{R} \\ 0 \end{bmatrix} \end{array} \begin{bmatrix} x \end{bmatrix} - \begin{array}{cc} \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{bmatrix} & \begin{array}{c} n \\ m-n \end{array} \end{array} = \begin{bmatrix} \tilde{R}x - \tilde{b}_1 \\ 0 - \tilde{b}_2 \end{bmatrix}$$

$$\Rightarrow \|Ax - b\|_2^2 = \|\tilde{R}x - \tilde{b}_1\|_2^2 + \|\tilde{b}_2\|_2^2 .$$

We want to find x that minimizes $\|Ax - b\|_2^2$.

This is equivalent to minimizing $\|\tilde{R}x - \tilde{b}_1\|_2^2$

which can be zero upon choosing $\boxed{x = \tilde{R}^{-1}\tilde{b}_1}$.

QR Least-Squares

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

The singular value decomposition

Theorem (SVD) Let A be a real $m\times n$ matrix. Then there exists an $m\times m$ orthogonal matrix U and an $n\times n$ orthogonal matrix V and an $m\times n$ "diagonal" matrix Σ such that

$$U^TAV=\Sigma. \qquad \Sigma=\operatorname{diag}\{\sigma_1,\dots,\sigma_p\},\quad p=\min\{m,n\}$$

$\sigma_1\geq\sigma_2\geq\dots\geq\sigma_p\geq 0$. The sigma's are called the <u>singular values</u> of A.

<u>proof</u>

<u>Claim</u> There exists vectors $x\in\mathbb{R}^n$, $y\in\mathbb{R}^m$ with $\|x\|_2=1$, $\|y\|_2=1$ such that $\boxed{Ax=\|A\|_2\,y}$

<u>proof of claim</u>: let x with $\|x\|_2=1$ and $\|A\|_2=\|Ax\|_2$.

$$\text{let } y=\frac{Ax}{\|A\|_2}\implies \|y\|_2=\frac{\|Ax\|_2}{\|A\|_2}=\frac{\|A\|_2}{\|A\|_2}=1$$

$$\Updownarrow$$

$$Ax=\|A\|_2\,y \quad \checkmark$$

Let $V_1=\left[\,x\;\middle|\;\tilde{V}_1\,\right]$ $n\times n$ orthogonal

$U_1=\left[\,y\;\middle|\;\tilde{U}_1\,\right]$ $m\times m$ orthogonal.

Now

$$U_1^TAV_1=\begin{bmatrix} y^T \\ \hline \tilde{U}_1^T\end{bmatrix}\begin{bmatrix} A \end{bmatrix}\left[\,x\;\middle|\;\tilde{V}_1\,\right]=\begin{bmatrix} y^T \\ \hline \tilde{U}_1^T\end{bmatrix}\left[\,Ax\;\middle|\;\tilde{V}_1\,\right]$$

$$=\left[\begin{array}{c|c} y^TAx & y^T\tilde{V}_1 \\ \hline \tilde{U}_1^TAx & \tilde{U}_1^T\tilde{V}_1\end{array}\right]$$

Now $Ax = \|A\|_2 y$ and since y is orthogonal to every column of $\tilde{U}_1$, $\tilde{U}_1^T A x = \tilde{U}_1^T \|A\|_2 y = \|A\|_2 \tilde{U}_1^T y = 0$.

$$\Rightarrow U_1^T A V_1 = \left[\begin{array}{c|c} \sigma_1 & w^T \\ \hline 0 & B \end{array}\right] \equiv A_2 \qquad \begin{array}{l} \sigma_1 = y^T A x = y^T(\|A\|_2 y) = \|A\|_2 \\ B = \tilde{U}_1^T \tilde{V}_1 \end{array}$$

we next show that $w = 0$. Indeed,

$$A_2 \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} = \left[\begin{array}{c|c} \sigma_1 & w^T \\ \hline 0 & B \end{array}\right] \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} = \begin{bmatrix} \sigma_1^2 + w^T w \\ Bw \end{bmatrix}$$

$$\Rightarrow \left\| A_2 \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2^2 = (\sigma_1^2 + w^T w)^2 + \|Bw\|_2^2 \geq (\sigma_1^2 + \|w\|_2^2)^2.$$

Thus

$$\|A_2\|_2 \geq \frac{\left\| A_2 \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2} \geq \frac{\sigma_1^2 + \|w\|_2^2}{\sqrt{\sigma_1^2 + \|w\|_2^2}} = \sqrt{\sigma_1^2 + \|w\|_2^2}. \quad (*)$$

Since U_1 and V_1 are orthogonal (isometries)

$\|A_2\|_2 = \|U_1^T A V_1\|_2 = \|A\|_2 = \sigma_1$. From this and $(*)$

$$\sigma_1 \geq \sqrt{\sigma_1^2 + \|w\|_2^2} \Rightarrow w = 0.$$

Thus

$$U_1^T A V_1 = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & B \end{array}\right] = A_2$$

This argument will be continued until the desired result is obtained. Note also that

$$\|B\|_2 = \sup_{\substack{\|x\|_2 = 1 \\ x \in R^{n-1}}} \|Bx\|_2 = \sup_{\substack{x \in R^n \\ x_1 = 0 \\ \|x\|_2 = 1}} \|A_2 x\|_2 \leq \sup_{\substack{x \in R^n \\ \|x\|_2 = 1}} \|A_2 x\|_2$$

we now treat B the same way we treated A:

$\exists$ $(m-1)\times(m-1)$ orthogonal $\tilde{U}_2$
$(n-1)\times(n-1)$ orthogonal $\tilde{V}_2$

such that
$$\tilde{U}_2^T B \tilde{V}_2 = \left[\begin{array}{c|c} \sigma_2 & 0 \\ \hline 0 & C \end{array}\right] \qquad \sigma_2 = \|B\|_2 \leq \sigma_1$$

$$U_2 = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \tilde{U}_2 \end{array}\right] \quad m \times m \quad \text{orthogonal}$$

$$V_2 = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \tilde{V}_2 \end{array}\right] \quad \text{orthogonal}$$

$$\Rightarrow U_2^T U_1^T A V_1 V_2 = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \tilde{U}_2^T \end{array}\right]\left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & B \end{array}\right]\left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \tilde{V}_2 \end{array}\right]$$

$$= \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & \tilde{U}_2^T B \tilde{V}_2 \end{array}\right] = \left[\begin{array}{c|c|c} \sigma_1 & & \\ & \sigma_2 & \\ \hline & & C \end{array}\right]$$

$\underline{m \leq n}$
$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ 0 & & \ddots & \\ & & & \sigma_p \end{bmatrix}$$

$m > n$
$$\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ & & \sigma_p \\ & 0 & \end{bmatrix}$$

$m = n$
$$\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{bmatrix}$$

Some properties and applications of the SVD

$U^T A V = \Sigma$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0 \quad p = \min\{m, n\}$ 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 A

Suppose $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0$

Then $\boxed{r = \operatorname{rank}(A)}$

proof We use the fact that

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

(ii) If A is invertible, then $\operatorname{rank}(AB) = \operatorname{rank}(B)$
$\operatorname{rank}(BA) = \operatorname{rank}(B)$.

Applying (ii) since U, V are invertible

$r = \operatorname{rank}(\Sigma) = \operatorname{rank}(U^T A V) = \operatorname{rank}(AV) = \operatorname{rank}(A)$ ✓

lemma A can be expressed as a sum of rank-one matrices, i.e.

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

where u_k is the k-th column of U and v_k is the k-th column of V

proof.

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

U_1 is $m\times r$ $\qquad$ V_1^T is $r \times n$

U_2 is $m\times(m-r)$ $\qquad$ V_2^T is $(n-r)\times n$

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

$\Rightarrow$ $A = U_1 \tilde{\Sigma} V_1^T$ is called the "Thin" SVD.

$$\Rightarrow \quad A = \left[\begin{array}{c|c|c|c} u_1 & u_2 & \cdots & u_r \end{array}\right]\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}\begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$$

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

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

Now

$$u_{ik} v_{jk} = (u_k v_k^T)_{ij} \quad \Rightarrow \boxed{A = \sum_{k=1}^{r} \sigma_k u_k v_k^T}$$

From this it follows that $\text{Range}(A) = \text{span}\{u_1, \ldots, u_r\}$

proof.

Suppose $y \in R^m$ is such that $y \in \text{Range}(A)$, i.e.

$y = Ax$ for some $x \in R^n$.

$$y = Ax = \sum_{k=1}^{n} \sigma_k u_k \underbrace{(v_k^T x)}_{\alpha_k} = \sum_{k=1}^{n} \sigma_k \alpha_k u_k$$

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

It is also the case that $Null(A) = span\{v_{r+1}, \ldots, v_n\}$

<u>proof.</u>
$$U^T A = \Sigma V^T$$

$Null(U^T A) = Null(A)$ since U is invertible.

also as above
$$\Sigma V^T = \begin{array}{c} \begin{bmatrix} \tilde{\Sigma} V_1^T \\ \hline 0 \end{bmatrix} \begin{array}{l} r \\ n-r \end{array} \end{array} \Rightarrow Null(\Sigma V^T) = span\{v_{r+1}, \ldots, v_n\}$$

<u>Relationship of singular values to eigenvalues.</u>

$$U^T A V = \Sigma \Rightarrow A = U \Sigma V^T$$

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

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

This shows that $A^T A$ and $\Sigma^T \Sigma$ are similar. Hence they have the same eigenvalues.

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.

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

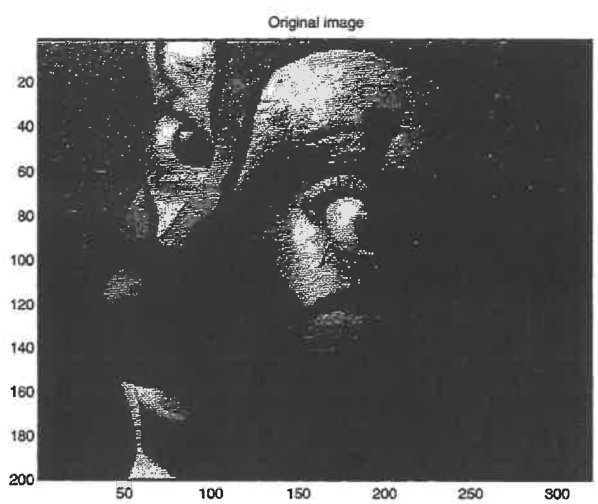
The next 3 pictures show the result of keeping only the largest k singular values for $k = 3, 10, 20$.

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

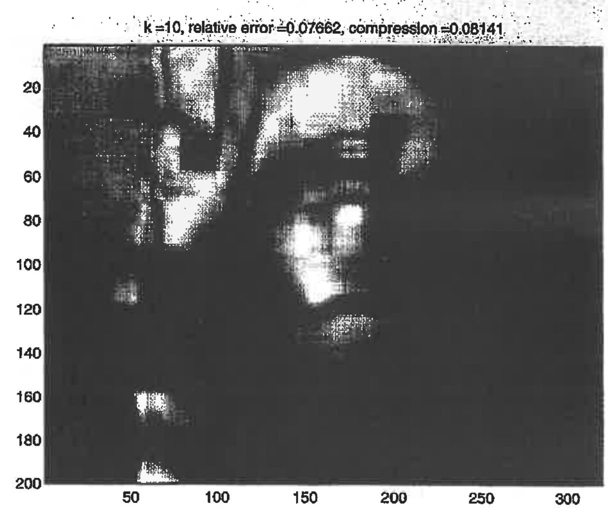
" " " compression ratio by $520k/64{,}000$.

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

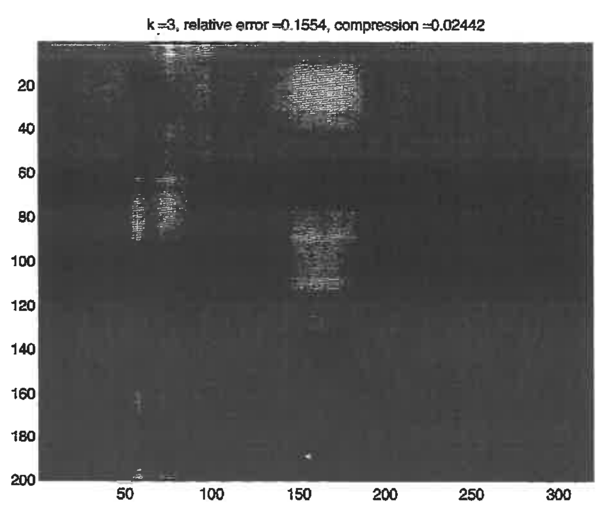
$k=3$ compression $= .024$, $k=10 \Rightarrow$ compression $= .081$, $k=2 \Rightarrow$ compr. $= .16$



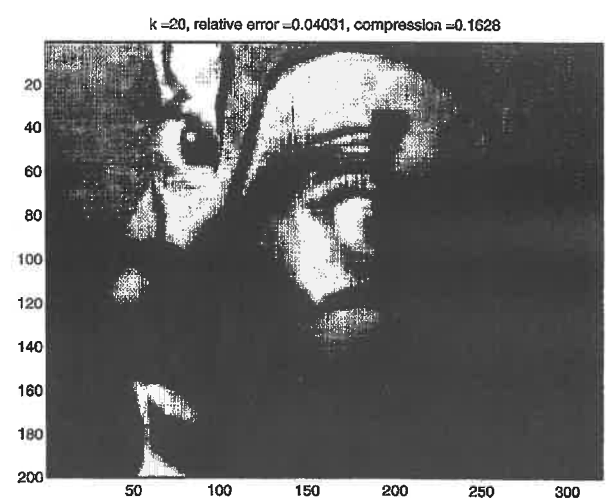
(a)



(c)



(b)



(d)

rank(A) < n ≤ m

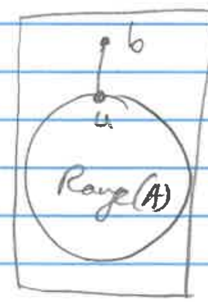
Application to the least-squares pb.: The rank deficient case

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

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

$$\Sigma V^T = \begin{array}{c} \left[\begin{array}{c} \tilde{\Sigma} V_1^T \\ \hline 0 \end{array}\right] \begin{array}{l} r \\ n-r \end{array} \end{array}, \quad \text{let } \tilde{b} = \left[\begin{array}{c} w \\ \hline z \end{array}\right]$$

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

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

$\|Ax-b\|_2^2$ 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) $\boxed{\sigma_i (v_i^T x) = w_i \quad , i = 1, \ldots, r}$

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

Also, the LS solution x belongs to R^n and $v_1, \ldots, v_n$ 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 $v_1, \ldots, v_n$ are orthonormal)

Hence from (1) and (2)

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

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