Thus

$$A \equiv Q R$$

$$\text{s.t. } Q^T Q = I$$

Each matrix P can be easily formed.

Suppose we want to reduce a column of A given by a :

$$P a = \alpha (1, 0, 0 \cdots 0)^T \equiv \alpha \hat{e}_1$$

Remember from orthogonality that

$$\| P a \|_2 = \| a \|_2 , \text{ thus } \alpha = \pm \| a \|_2$$

It can be shown that :

$$P = I - 2 \frac{v v^T}{v^T v}$$

where $v = a - \alpha \hat{e}_1$

accomplishes this reduction.

Let's examine this matrix.

First, $\underset{\approx}{P}$ is symmetric such that

$$\underset{\approx}{P}^T = \underset{\approx}{P} \quad \text{or} \quad P_{ij} = P_{ji}$$

Second, it is orthogonal:

$$\underset{\approx}{P}^T \underset{\approx}{P} = \left(\underset{\approx}{I} - 2\frac{\underset{\sim}{V}\underset{\sim}{V}^T}{(\underset{\sim}{V}\cdot\underset{\sim}{V})}\right)\left(\underset{\approx}{I} - 2\frac{\underset{\sim}{V}\underset{\sim}{V}^T}{(\underset{\sim}{V}\cdot\underset{\sim}{V})}\right)$$

$$= \underset{\approx}{I} - 4\frac{\underset{\sim}{V}\underset{\sim}{V}^T}{(\underset{\sim}{V}\cdot\underset{\sim}{V})} + 4\frac{\underset{\sim}{V}\cancel{(\underset{\sim}{V}^T\underset{\sim}{V})}\underset{\sim}{V}^T}{(\underset{\sim}{V}\cdot\underset{\sim}{V})^{\cancel{2}}}$$

where the $\underset{\sim}{V}\cdot\underset{\sim}{V}$'s cancel out! Thus:

$$\underset{\approx}{P}^T \underset{\approx}{P} = \underset{\approx}{I}$$

Finally, we want to prove that our choice for $\underset{\approx}{P}$ accomplishes the desired reduction. It's actually easier to determine the choice

63B

for $\underset{\sim}{v}$ which does the job!

We seek a vector $\underset{\sim}{v}$ such that

$\underset{\approx}{P}\,\underset{\sim}{a} = \alpha \underset{\sim}{\hat{e}}_1$ where $\alpha \equiv \pm \|\underset{\sim}{a}\|_2$

Note that this must be true since $\underset{\approx}{P}$ is orthogonal and thus preserves the 2-norm!

So: $\underset{\approx}{P}\,\underset{\sim}{a} = \left(\underset{\approx}{I} - 2\dfrac{\underset{\sim}{v}\,\underset{\sim}{v}^T}{(\underset{\sim}{v}\cdot\underset{\sim}{v})}\right)\underset{\sim}{a}$

$$= \underset{\sim}{a} - 2\frac{\underset{\sim}{v}\,\underset{\sim}{v}^T\underset{\sim}{a}}{(\underset{\sim}{v}\cdot\underset{\sim}{v})} = \alpha\,\underset{\sim}{\hat{e}}_1$$

Rearranging:

$$\underbrace{2\frac{(\underset{\sim}{v}\cdot\underset{\sim}{a})}{(\underset{\sim}{v}\cdot\underset{\sim}{v})}}_{\text{scalar}}\underset{\sim}{v} = \underset{\sim}{a} - \alpha\,\underset{\sim}{\hat{e}}_1$$

Now we note that $\underset{\approx}{P}$ is independent of the magnitude of $\underset{\sim}{v}$ – you can multiply it by any scalar, and the resulting $\underset{\approx}{P}$ will be unchanged! Thus any vector proportional to:

$$\underset{\sim}{v} = \underset{\sim}{a} - \alpha \underset{\sim}{\hat{e}}_1$$

will accomplish the desired reduction.

OK, so how do we use this?

We have the problem:

$$\min_{\underset{\sim}{x}} \| \underset{\approx}{A} \underset{\sim}{x} - \underset{\sim}{b} \|_2$$

$$= \min_{\underset{\sim}{x}} \| \underset{\approx}{Q} \underset{\approx}{R} \underset{\sim}{x} - \underset{\sim}{b} \|_2$$

$$= \min_{\underset{\sim}{x}} \| \underset{\approx}{R} \underset{\sim}{x} - \underset{\approx}{Q}^T \underset{\sim}{b} \|_2$$

by orthogonality

$$= \min_{\underset{\sim}{x}} \| \underset{\approx}{R} \underset{\sim}{x} - \underset{\sim}{c} \|_2$$

where $\underset{\sim}{c} \equiv \underset{\approx}{Q}^T \underset{\sim}{b}$

we solve the first m rows exactly, while the last $n-m$ elements of $\underset{\sim}{c}$ is the residual!

accomplishes this reduction

~~Work out matlab demonstration of QR factorization.~~

Degenerate Problems

What happens if the problem is under determined??

Corresponds to matrix $\underset{\sim}{R}$ not being of full rank

rank ≡ largest submatrix w/ non-zero determinant.

In linear regression this arises if the modelling functions are linearly dependent

Example:

$$b(t) = X_1 \cdot (1) + X_2 \cdot (t) + X_3 \cdot (2t+1)$$

For this problem there will be an

infinite number of ways to approximate data w/ the same residual (minimum)

In this case:

$$\alpha\left[1\cdot(1)+2(t)-1\cdot(2t+1)\right]=0$$

for all α, thus

$$\underset{\sim}{X}=(x_1,x_2,x_3)$$

can be replaced with

$$\underset{\sim}{X}^*=\underset{\sim}{X}+\alpha(1,2,-1)$$

without changing the residual

A linear regression problem will be degenerate if the columns of $\underset{=}{A}$ (an $m\times n$ matrix) are linearly dependent

skip The line between degeneracy & non-degeneracy may be blurred by round off error!

Suppose we had:

$$\begin{pmatrix} 1 & 1 \\ 0 & 10^{-k} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

This has the solution:

$$x_2 = 10^k, \quad x_1 = \frac{1-10^k}{1} \approx -10^k$$

Now suppose $k >> 1$ and A_{22} is non-zero <u>only</u> due to roundoff error

We would prefer to fit the data with:

$$x_2 = 0, \quad x_1 = 1$$

The <u>size</u> of $\|\underset{\sim}{x}\|_2 = 1$

vs. $\sqrt{2}\ 10^k$!

You usually want to keep your

modelling parameters <u>small</u>, thus we seek a vector with the shortest length for degenerate problems!

Sometimes it's hard to tell how close a matrix is to degeneracy by looking at it. Take:

$$\underset{\approx}{A} = \begin{pmatrix} 1 & -1 & -1 & -1 & \cdots & -1 \\ & 1 & -1 & -1 & \cdots & -1 \\ & & 1 & -1 & \cdots & -1 \\ & 0 & & 1 & \cdots & -1 \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix}$$

The det. of this matrix is $\det(A) = 1$

If $A_{n1} = 2^{-(n-2)}$ then the matrix is singular!

Although $\underset{\approx}{A}$ is in upper triangular form already, we can do a QR on it w/ column pivoting!

Pick the column w/ lgest norm, move to column 1 & reduce it. Then pick next largest column of $(n-1)\times(n-1)$ submatrix & move to column 2, etc.

This yields:

$$\begin{pmatrix} 1 & -1 & -1 & -1 & -1 \\ & 1 & -1 & -1 & -1 \\ & & 1 & -1 & -1 \\ & 0 & & 1 & -1 \\ & & & & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 2.24 & 0.89 & .45 & 0 & -.45 \\ & 1.79 & .34 & 0 & -.34 \\ & & -1.64 & 0 & .42 \\ & 0 & & -1.4 & .71 \\ & & & & 0.108 \end{pmatrix}$$

Note that the last diagonal element is an order of magnitude smaller than the other diag elements! This is because A is close to singularity!

This is equivalent to the factorization:

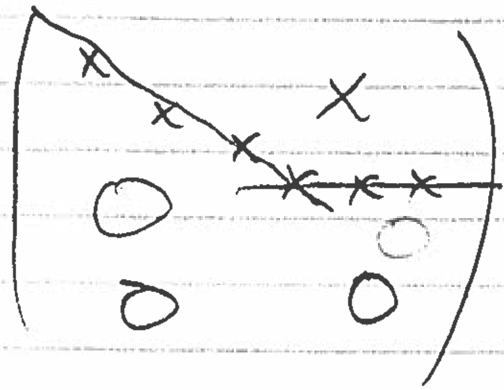
$$\underset{\sim}{A} = \underset{\sim}{Q}\,\underset{\sim}{R}\,\underset{\sim}{P} \leftarrow \text{rt. mult interchanging columns of } \underset{\sim}{A}$$

skip

69

If $\underset{\approx}{A}$ is singular, R is of form:

(w/ QRP)



The number of non-zero rows in $\underset{\approx}{R}$ is the <u>rank</u> of $\underset{\approx}{A}$

We thus reduce the regression problem to:

$$\| \underset{\approx}{A} \underset{\sim}{x} - \underset{\sim}{b} \|_2 = \| \underset{\approx}{R} \underset{\approx}{P} \underset{\sim}{x} - \underset{\approx}{Q}^T \underset{\sim}{b} \|_2$$

To solve this, let:

$$\underset{\sim}{y} = \underset{\approx}{P} \underset{\sim}{x} \quad \text{and} \quad \underset{\sim}{c} = \underset{\approx}{Q}^T \underset{\sim}{b}$$

Then

$$\min_{\underset{\sim}{y}} \| \underset{\approx}{R} \underset{\sim}{y} - \underset{\sim}{c} \|_2 \text{ may be}$$

solved via back substitution

We can go even further than this, and do Singular Value Decomposition!

We operated on $\underset{\approx}{A}$ to factor it into an orthogonal matrix and an upper triangular matrix. We can operate on $\underset{\approx}{R}$ to factor it into a diagonal matrix and an orthogonal matrix!

Thus:

$$\underset{m\times n}{\underset{\approx}{A}} = \underset{\substack{\text{orthogonal}\\ m\times m}}{\underset{\approx}{U}} \; \overset{\substack{\text{diagonal}\\ m\times n}}{\underset{\approx}{\Sigma}} \; \underset{\substack{\text{orthogonal}\\ n\times n}}{\underset{\approx}{V}}^{T}$$

Using pivoting, we order the elements of Σ s.t.

$$|\sigma_1| \geq |\sigma_2| \geq |\sigma_3| \geq \cdots \geq |\sigma_n|$$

If $\sigma_n = 0$ then $\underset{\approx}{A}$ is not of full rank! $(m > n)$

If A is square, then $|\sigma_1/\sigma_n|$ is the condition number of $\underset{\approx}{A}$

SVD is useful for many purposes. First, we have the data fitting/regression problem:

$$\min_{\underset{\sim}{x}} \|\underset{\approx}{A}\,\underset{\sim}{x} - \underset{\sim}{b}\|_2 = \min_{\underset{\sim}{x}} \|\underset{\approx}{U}\,\underset{\approx}{\Sigma}\,\underset{\approx}{V}^T\underset{\sim}{x} - \underset{\sim}{b}\|_2$$

$$= \min_{\underset{\sim}{x}} \|\underset{\approx}{\Sigma}\,\underset{\approx}{V}^T\underset{\sim}{x} - \underset{\approx}{U}^T\underset{\sim}{b}\|_2$$

let $\underset{\approx}{V}^T\underset{\sim}{x} = \underset{\sim}{z}$, $\underset{\approx}{U}^T\underset{\sim}{b} = \underset{\sim}{d}$

$\therefore \underset{\approx}{V}\,\underset{\approx}{V}^T\underset{\sim}{x} \equiv \underset{\sim}{x} = \underset{\approx}{V}\,\underset{\sim}{z}$

and $z_i = d_i/\sigma_i$

and $r^2 = \sum_{n+1}^{m} d_i^2$ <u>if</u> $\sigma_n \neq 0$

If $\sigma_n = 0$ then the problem is <u>singular</u>

If $\sigma_n \approx 0$ (or of order ε_{mach}) we may want to <u>regard</u> it as zero so it doesn't mess up the calculation of $\underset{\sim}{x}$!

Thus we let $z_i = \begin{cases} d_i/\sigma_i & \sigma_i > \varepsilon \\ 0 & \sigma_i < \varepsilon \end{cases}$

This has the property that among the set of vectors $\underset{\sim}{x}^*$ which minimize the residual:

$$\| \underset{\approx}{A} \underset{\sim}{x} - \underset{\sim}{b} \|_2$$

That computed via SVD will have the minimum norm:

$$\| \underset{\sim}{x}_{SVD} \|_2 \le \| \underset{\sim}{x} \|_2$$

---

skip

Another use for SVD is <u>data compression</u>

Recall $\underset{\approx}{A} = \underset{\approx}{U} \underset{\approx}{\Sigma} \underset{\approx}{V}^T$

SVD: Application to Control

Control problems often reduce to a set of linear equations where a limited number of control actions (parameters) are chosen to achieve some desired outcomes. Even if the underlying problem is <u>non-linear</u>, use of the Taylor Series can lead to linearization

These problems are often addressed via SVD. Suppose we want to <u>minimize</u>: model

target

$$\underset{\sim}{r} = \underset{\approx}{A}\,\underset{\sim}{x} - \underset{\sim}{b}$$

control parameters

We can "solve" this in a least-squares sense via SVD as we just showed!

In control, however, the problem is often close to singular — thus since the smallest singular values are close to zero, the solution vector $\underset{\sim}{x}$ blows up! We wind up with large errors in this case, and can have <u>unstable</u> <u>control</u> due to problems w/ the solution vector.

Example: plastic bag fabrication
Addition of more actuators for control of molding / bag thickness actually led to <u>worse</u> performance! (Richard Braatz, MIT)

We can fix this by doing SVD and setting the <u>smallest</u> singular values to zero! This is equiv. to controlling actuators in a reduced pattern (not independently)

SVD: Sensitivity Analysis ①

We look at a vector function of a vector ($\underset{\sim}{x}$)

$$\underset{\sim}{z} = \underset{\sim}{f}(\underset{\sim}{x})$$

Expand about $\underset{\sim}{x}_0$:

$$\underset{\sim}{z}_0 = \underset{\sim}{f}(\underset{\sim}{x}_0), \quad \underset{\sim}{z}' = \underset{\sim}{z} - \underset{\sim}{z}_0, \quad \underset{\sim}{x}' = \underset{\sim}{x} - \underset{\sim}{x}_0$$

$$\underset{\sim}{z} = \underset{\sim}{f}(\underset{\sim}{x}_0) + \underset{\sim}{\nabla}\underset{\sim}{f}\,(\underset{\sim}{x} - \underset{\sim}{x}_0) + O(\underset{\sim}{x}'^2)$$

so

$$\underset{\sim}{z}' \simeq \underset{\sim}{\nabla}\underset{\sim}{f}\,\underset{\sim}{x}'$$

$\underset{\sim}{\nabla}\underset{\sim}{f}$ details the sensitivity of $\underset{\sim}{z}$ to changes in $\underset{\sim}{x}$!

What is $\underset{\sim}{\nabla}\underset{\sim}{f}$?

$$\underset{\sim}{\nabla}\underset{\sim}{f} \equiv \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & & & \vdots \\ \frac{\partial f_n}{\partial x_1} & & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}$$

We can study this using SVD!

Let $\underset{\approx}{A} = \underset{\sim}{\nabla} \underset{\sim}{f}$

So $\underset{\sim}{z}' = \underset{\approx}{A}\,\underset{\sim}{x}'$

Let $\underset{\approx}{U}\,\underset{\approx}{\Sigma}\,\underset{\approx}{V}^T = \underset{\approx}{A}$

where $\underset{\approx}{U}^T \underset{\approx}{U} = \underset{\approx}{I}$

$\underset{\approx}{V}^T \underset{\approx}{V} = \underset{\approx}{I}$

and

$$\underset{\approx}{\Sigma} = \begin{bmatrix} \sigma_{11} & & 0 \\ & \ddots & \\ 0 & & \sigma_{nn} \end{bmatrix}$$

min of n, m if matrix isn't square

Now because $\underset{\approx}{U}$ and $\underset{\approx}{V}$ are orthogonal, they preserve the 2-norm of any vector.

Thus: $\|\underset{\approx}{U}\,\underset{\sim}{y}\|_2 = \|\underset{\sim}{y}\|_2$