Singular Value Decomposition (SVD)

Considered the most stable method for calculating solutions to system of linear equations and least squares estimators.

X is an $n \times p$ matrix with $n \geq p$. The SVD of X is

$$X = UDV^T$$

where

- U is an $n \times p$ matrix with orthonormal columns
- D is an $p \times p$ diagonal matrix with $d_{ii} \geq 0$
- U is an $p \times p$ orthogonal matrix

We will assume that $d_{11} \geq \ldots \geq d_{pp}$.

Note there is an alternate form of the decomposition where U is an $n \times n$ orthogonal matrix and D is extended to an $n \times p$ matrix by adding $n - p$ rows with all elements set to 0.

The S-Plus/R function `svd()` uses the first form, whereas the Matlab function `svd()` uses the second form (use `svd(X,0)` to get the first form).

I'll use the first form, since it's a bit easier to talk about.

Knowing SVD of X immediately gives us what we need to know about $X^T X$ since

$$X^T X = VDU^T UDV^T = VD^2 V^T$$

So V are the eigenvectors of $X^T X$ and the d_{ii}^2 are the corresponding eigenvalues.

If X is a square, non-singular matrix, then both U and V are orthogonal matrices and

$$X^{-1} = VD^{-1}U^T$$

As this is also the SVD of X^{-1}, the singular values of X^{-1} are just the reciprocals of the singular values of X.

If X is square, symmetric and positive definite, $U = V$ as the SVD reduces to the usual eigenvalue/eigenvector decomposition of X.

This won't hold as X moves from symmetry and positive definiteness.

Generalized Inverses

A matrix X^- is a generalized inverse of a matrix X if

$$XX^-X = X$$

Note that X^- is usually not unique, except if X is invertible. See Rao, 1973 section 1b for the forms of all generalized inverses of X.

Sometimes it's useful to pick a generalized inverse that satisfies

$$X^-XX^- = X^-$$

This condition will not hold for some generalized inverses.

The most popular generalized inverse is the Moore-Penrose generalized inverse which satisfies

$$X^-X = (X^-X)^T \text{ and } XX^- = (XX^-)^T$$

The Moore-Penrose generalized inverse can easily gotten from the SVD

Let D^+ be the diagonal matrix with entries

$$d_{ii}^{+} = \begin{cases} 1/d_{ii} & d_{ii} > 0 \\ 0 & d_{ii} = 0 \end{cases}$$

Then the Moore-Penrose generalized inverse is

$$X^{+} = VD^{+}U^{T}$$

Solving $Ax = b$ via the SVD

If A is invertible, then $VD^{-1}U^{T}b$ is the route to go since $D^{-1}U^{T}b$ involves only p divisions with $U^{T}b$ since D is diagonal.

The more interesting case occurs if A is singular.

If b is not in the range space of A, then there is no solution.

Otherwise there are an infinite number of solutions, with $A^{+}b$ being one.

The other solutions will be of the form

$$A^{+}b + \sum_{j=k+1}^{p} \alpha_j V_j$$

where $d_{11}, \ldots, d_{kk} > 0$ and $d_{k+1,k+1} = \ldots = d_{pp} = 0$

SVD and Least Squares

Want to look at two cases

1) rank(X) = p

2) rank(X) < p

1) rank(X) = p

This is the usual regression situation.

$$X^T X = VD^2V^T$$

so

$$\hat{\beta} = VD^{-1}U^T y$$

In addition, other useful quantities can be calculated in terms of V and D, though sometimes, its also useful to have X and y around as well.

For example, $\text{Var}(\hat{\beta}) = \sigma^2 VD^{-2}V^T$.

For the “hat” matrix, there are a couple of ways to do it. One is

$$H = XVD^{-2}V^T X^T$$

This form is more useful if you only need the leverages h_{ii}, the diagonal entries of H since

$$h_{ii} = x_i^T V D^{-2} V^T X^T x_i$$

H can also be gotten directly from the SVD since

$$H = UU^T$$

2) rank(*X*) < *p*

This situation occurs when some of the predictors are linear combinations of the others.

One situation where this can occur is with the ANOVA model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

when no constraints are put on the τ_i's (which corresponds to dropping columns from *X*).

The least squares solution, based on the generalized inverse

$$\hat{\beta} = VD^{+}U^T y = X^{+} y$$

still satisfies the normal equations

$$X^T X \beta = X^T y$$

For the ANOVA problem rank(*X*) = *p* – 1.

In S-Plus/R, this isn't a problem since they put constraints based on the setting of `options()$contrasts`. (See Statistical Models in S or Modern Applied Statistics with S for a discussion)

However in SAS, no constraints are used directly if this model is fit with `PROC GLM`.

In the output, they note the problem and mention a generalized inverse is used.

However underlying the SAS output is the constraint to force the last group to have $\hat{\tau}_k = 0$ (equivalent to `'contr.treatment'` in S-Plus/R)

```
Margarine Experiment                   18:55 Tuesday, February 17, 2004   3

The GLM Procedure

                                              Standard
Parameter            Estimate                    Error    t Value    Pr > |t|

Intercept        0.0047940406 B      0.00005339      89.79      <.0001
brand     1      0.0008799749 B      0.00007551      11.65      <.0001
brand     2      -.0005979730 B      0.00007551      -7.92      <.0001
brand     3      0.0010435329 B      0.00007551      13.82      <.0001
brand     4      0.0000000000 B       .                 .          .

NOTE: The X'X matrix has been found to be singular, and a generalized
      inverse was used to solve the normal equations.  Terms whose
      estimates are followed by the letter 'B' are not uniquely estimable.
```

Exactly what generalized inverse they are using isn't clear. It's not the Moore-Penrose as that doesn't force one of the coefficients to 0.

However, given that any valid solution can be gotten from

$$(X^T X)^+ X^T y + \sum_{j=k+1}^{p} \alpha_j V_j = VD^+ U^T y + \sum_{j=k+1}^{p} \alpha_j V_j$$

they may start with it and then figure out the α_j needed to set certain components to 0.

For the one way ANOVA problem, it appears that the Moore-Penrose solution is equivalent to the constraint $\mu = \sum \tau_i$.

Since the V_j corresponding to the 0 singular values are a basis for the null space of X, checking these will give you the constraints underlying the Moore-Penrose solution. For one-way ANOVA, it appears that

$$V_p^T \propto [1 \quad -1 \quad -1 \quad \dots \quad -1]$$

which corresponds to the constraint $\mu = \sum \tau_i$.

(Actually SAS uses a modified sweep routine instead of SVD or QR in `PROC GLM`, but the previous describes how they would do things if they used SVD instead.)

Estimable functions ($l^T\beta$)

A function $l^T\beta$ is estimable from data if l is in the range space of X^T. Since the range space of X^T is the same as the range space of VD, it follows that the columns of V corresponding to d_{ii} > 0, for a basis for the space of estimable functions.

Iterative methods for solving $Ax = b$

If the matrix A is sparse (most entries are 0), it can be preferable to solve the system by the iterative scheme

$$Mx^{(k+1)} = Nx^{(k)} + b$$

where $A = M - N$.

If $x^{(k)}$ converges to $x^{(\infty)}$, then $Ax^{(\infty)} = b$

This scheme will converge if all the eigenvalues λ_j of $M^{-1}N$ satisfy $\left|\lambda_j\right|$ < 1.

Note that this is equivalent to the conditions of proposition 6.4.1 of Lange, where he only considers the situation $M = I$ and $N = I - A$.

This sort of scheme will be reasonable if $x^{(k)}$ can be easily computed.

One scheme (Gauss-Seidel) that allows this is as follows

- Let D be the diagonal matrix with entries a_{ii}
- Let U be the matrix with entries $u_{ij} = a_{ij}$ for $i < j$ and 0 otherwise (elements above the main diagonal)
- Let L be the matrix with entries $u_{ij} = a_{ij}$ for $i > j$ and 0 otherwise (elements below the main diagonal)

Then $A = L + D + U$

Let $M = D + L$ and $N = -U$.

The iterates $x_j^{(k+1)}$ satisfy

$$x_j^{(k+1)} = \frac{1}{a_{jj}}\left(b_j - \sum_{l<j} a_{jl} x_l^{(k+1)} - \sum_{l>j} a_{jl} x_l^{(k)} \right)$$

Vector Norms

Euclidean norm $\|x\|_2 = \sqrt{\sum x_i^2}$

There are other norms that can be useful

A norm on $\mathbb{R}^m$ must satisfy the following 4 conditions

1) $\|x\| \geq 0$

2) $\|x\| = 0$ iff $x = 0$

3) $\|cx\| = |c|\|x\|$ for every real number c

4) $\|x+y\| \leq \|x\| + \|y\|$ (triangle inequality)

Other common norms are

L_1: $\|x\|_1 = \sum_{i=1}^{m} |x_i|$

L_∞: $\|x\|_\infty = \max_{1 \leq i \leq m} |x_i|$

These are all special cases of

L_p: $\|x\|_p = \sqrt[p]{\sum |x_i|^p}$

It ends up that all of these norms induce the same topology on the space $\mathbb{R}^m$ due to the following proposition (6.2.1 of Lange)

Let $\|x\|$ be any norm of $\mathbb{R}^m$. The there exist positive constants such that

$$k_l \|x\|_1 \leq \|x\| \leq k_l \|x\|_u$$

Two consequences of this proposition are

- $\|x\|_q \leq \|x\|_p$
- $\|x\|_p \leq m^{\frac{1}{p}-\frac{1}{q}} \|x\|_q$

when $p < q$ and p and q are taken from {1, 2, ∞}. (Actually I believe the result holds when p and q are both ≥ 1)

So a consequence of these, is that you can pick the norm which makes your problem the easiest to deal with.

Matrix Norms (for square matrices)

Want them to have properties 1 through 4 of vector norms, plus

5) $\|AB\| \le \|A\|\|B\|$ for any product of $m \times m$ matrices

One possible matrix norm is the Euclidean norm

$$\|A\|_E = \sqrt{\sum\sum a_{ij}^2} = \sqrt{\mathrm{tr}(AA^T)} = \sqrt{\mathrm{tr}(A^T A)}$$

Another way of getting matrix norms is by inducing them from vector norms via

$$\begin{aligned}\|A\| &= \sup_{x \ne 0} \frac{\|Ax\|}{\|x\|} \\ &= \sup_{\|x\|=1} \|Ax\|\end{aligned}$$

Note that the second form of the definition implies that $\|I\|$ = 1 for any induced matrix norm

Since $\|I\|_E = \sqrt{m}$, $\|A\|_E$ and $\|A\|_2$ are different norms

For the popular choices of *p*, the matrix norms are

- $\|A\|_1 = \max_j \sum_i |a_{ij}|$ (maximum column sum)
- $\|A\|_2 = \sqrt{\rho(A^T A)}$ which reduces to $\rho(A)$ when *A* is symmetric
- $\|A\|_\infty = \max_i \sum_j |a_{ij}|$ (maximum row sum)

where $\rho(A)$ is the absolute value of the dominant eigenvalue of *A*. Its sometimes referred to as the spectral radius of *A*.

The spectral radius can be tied to any induced matrix norm through proposition 6.3.2 as

$$\rho(A) \le \|A\|$$

In addition, for any *A* and $\varepsilon > 0$, there exists some induced matrix norm such that

$$\|A\| \le \rho(A) + \varepsilon$$

These results can be used to justify the Gauss-Seidel scheme talked about earlier.

Lets focus on the form $x^{(k+1)} = Bx^{(k)} + b$ which can be used to solve $(I - B)x = b$.

Underlying this scheme is the map $f(x) = Bx + b$, which satisfies

$$\begin{aligned} \|f(y) - f(x)\| &= \|B(y-x)\| \\ &\leq \|B\|\|y-x\| \end{aligned}$$

which is contractive if $\|B\| < 1$.

This is analogous to $|f\,'(x)| < 1$ in some region for the fixed point methods discussed earlier.

In fact, if we replace absolute values with norms, the earlier proof of the functional iteration results generalize to this vector setting.

Thus the iterates $x^{(k)}$ converge to the unique solution of $(I - B)x = b$.

It also follows that $I - B$ must be invertible as well.

The results are a consequence of proposition 6.4.1

Let B be an arbitrary matrix with special radius $\rho(B)$. Then $\rho(B) < 1$ if and only if $\|B\| < 1$ for some induced matrix norm. $\|B\| < 1$ implies

a) $\lim_{n\to\infty} \|B^n\| = 0$

b) $(I - B)^{-1} = \sum_{n=0}^{\infty} B^n$

c) $\frac{1}{1+\|B\|} \le \|(I-B)^{-1}\| \le \frac{1}{1-\|B\|}$

Part a) follows from $\|B^n\| \le \|B\|^n$

Part b) come from starting the iteration scheme at $x^{(0)}$ which gives $x^{(n)} = \sum_{i=0}^{n-1} B^i b$. If we pass to the limit, we get $(I-B)^{-1}b = \sum_{i=0}^{\infty} B_i b$. Since b is arbitrary, the result holds.

Part c) is useful is that is allows us to put bounds on the norm of the solution of $(I - B)x = b$.

Condition numbers

While at its simplest level, a square matrix is either singular or not, the situation is a bit more complicated.

Lets look at two systems of equations

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2.0001 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 2 \\ 2.0002 \end{bmatrix}$$

As can be easily seen the solutions are

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

Underlying the problem here is that the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix}$$

is very close to the singular matrix

$$A' = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

So let $Ax = b$ and $A(x+\Delta x) = b + \Delta b$ be the two systems of interest.

Then by the definition of the induced matrix norm

$$\|b\| \le \|A\|\|x\|$$

$$\|\Delta x\| \le \|A^{-1}\|\|\Delta b\|$$

These imply

$$\frac{\|\Delta x\|}{\|x\|} \le \|A\|\|A^{-1}\|\frac{\|\Delta b\|}{\|b\|} = \mathrm{cond}(A)\frac{\|\Delta b\|}{\|b\|}$$

where $\mathrm{cond}(A) = \|A\|\|A^{-1}\|$ is known as the condition number.

Instead of changing b, lets change A giving the systems $Ax = b$ and $(A+\Delta A)(x+\Delta x) = b$.

Then it can be shown that

$$\Delta x = -A^{-1}\Delta A(x+\Delta x)$$

so that

$$\|\Delta x\| = \|A^{-1}\|\|\Delta A\|\|x+\Delta x\|$$

which implies

$$\frac{\|\Delta x\|}{\|x+\Delta x\|} \leq \text{cond}(A)\frac{\|\Delta A\|}{\|A\|}$$

Going through a bunch of algebra gives

$$\frac{\|\Delta x\|}{\|x\|} \leq \text{cond}(A)\frac{\|\Delta A\|}{\left(\|A\| - \text{cond}(A)\|\Delta A\|\right)}$$

Note that these results hold for whatever matrix norm is being used. However usually condition numbers based on the $\|A\|_2$ norm are used.

For this norm, $\text{cond}_2(A)$ is the ratio of the largest and smallest eigenvalues of A.

For the example above λ_1 = 2.00005 and λ_2 = 0.00005 giving $\text{cond}_2(A)$ = 40002.

Why care about condition numbers?

The numerical stability of many matrix routines can be described through the condition number.

For example, the LU decomposition can become unstable with matrices with large condition numbers, particularly when a poor choice of pivots are used.