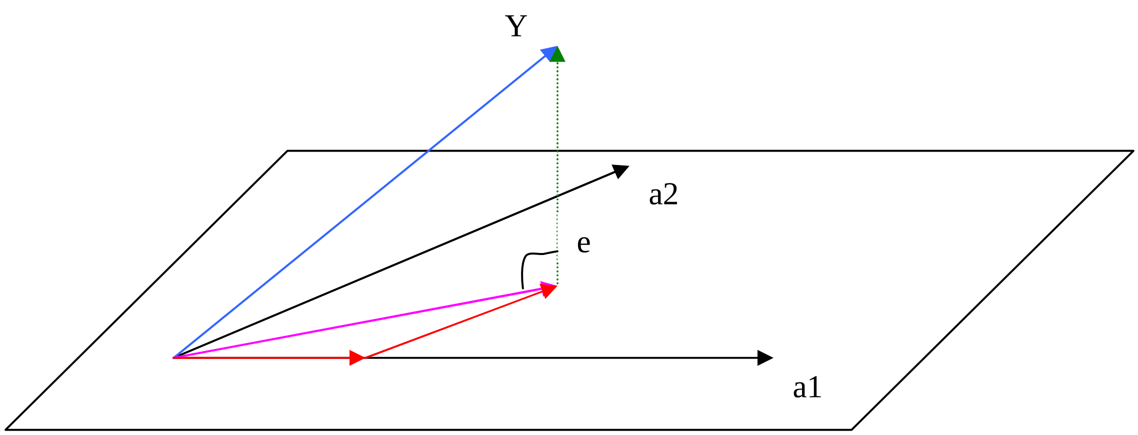
Lecture Date: $16^{th}$ & $17^{th}$ Apr '04

# Solving Least Squares

The problem

Given a vector Y and a set of basis vectors A = $\{a_1, a_2, \ldots a_n\}$, we want to find the measure of Y in terms of A i.e. find M such that
MA = Y
The vector space spanned by Y and A differs, giving us a set of over-determined equations. So, we find an M such that
$|| Y - MA ||$ is minimum.

So, we are basically trying to - find the shadow of Y in the subspace spanned by A, and also find the linear combinants, which add up to that shadow. This is illustrated by the figure below. The vectors in red are the linear combinants in the direction of a1 & a2.




The vector in pink is the shadow of Y in the subspace of A. 'e' is the error, which is **orthogonal to the subspace spanned by A.**

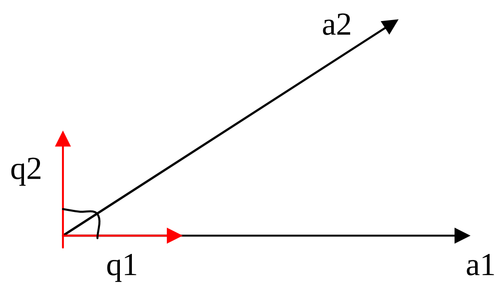
We notice that if the A is an orthogonal basis, the solution to the problem is simplified. If the basis is orthogonal, then the linear combinants are just the coordinate values of Y (the analysis and synthesis matrix is same).

Mathematically, the same can be justified by the fact that the inverse of an orthogonal matrix is its conjugate transpose.
To find the linear combinants we need the inverse of the basis matrix as shown below.

AC = Y, C is [ $\lambda_1$, $\lambda_2$….., $\lambda_n$], the scales by which each direction in A needs to be scaled so that they all add up to Y.
C = $A^{-1}$ Y.

Thus the least square problem is reduced to finding a set of orthogonal set of basis Q such that
Span { Q } = Span { A }
i.e. a set of Orthogonal vector space spanning the same subspace as that of the given basis.

For example in a 2-D subspace, it can be shown as below. q1 and q2 are the normalized orthogonal vectors for the subspace of a1 & a2



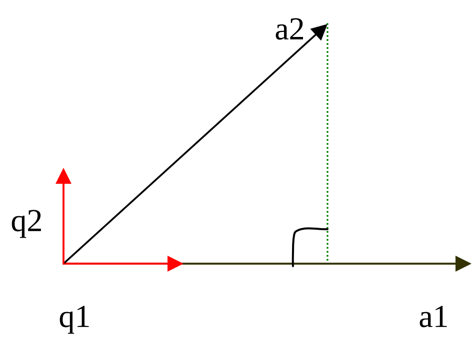

Therefore we have different methods of orthogonalizing the basis vector A.

I. <u>Graham Schmidt Method</u>

This is the successive orthogonalization (the QR method).
Let A contain n linearly independent vectors. We want to find the orthonormal basis for A

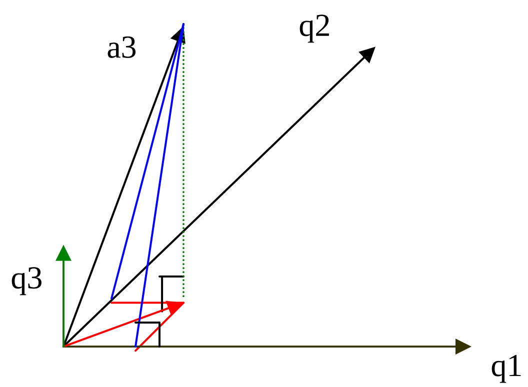
q1 is a1/|a1|


For a2, we take the projection of a2 on a1, to find the orthogonal vector q2



∴ q2 = a2 – proj of a2   on a1 / |a2 – proj of a2   on a1|
q3 is a vector orthogonal to subspace of a1& a2. Its simple to see this a least squares problem. (given vector a3, and need to find its shadow in a1 & a2, so that we can find the direction of the error which is orthogonal to a1 & a2)

Hence we use the orthogonal vectors q1 and a2 to find q3.
First we drop perpendicular to q1 and q2 (refer to the figure below) shown in blue. Assume a3 is coming out of the monitor.




Now, if we look at a3 to be made up of 2 parts, 1 which lies in the subspace of q1, q2 and the other remaining is the component orthogonal to q1 & q2.
By, dropping perpendiculars on q1 and q2 we find the projection of a3 in q1 & q2. Since q1,q2 are ortho to each other, these projections when added up will give the shadow of a3 in q1 & q2. Subtracting this from a3 gives us the component that not in q1 & q2 i.e orthogonal.

In the figure, the shadow of a3 is in red.

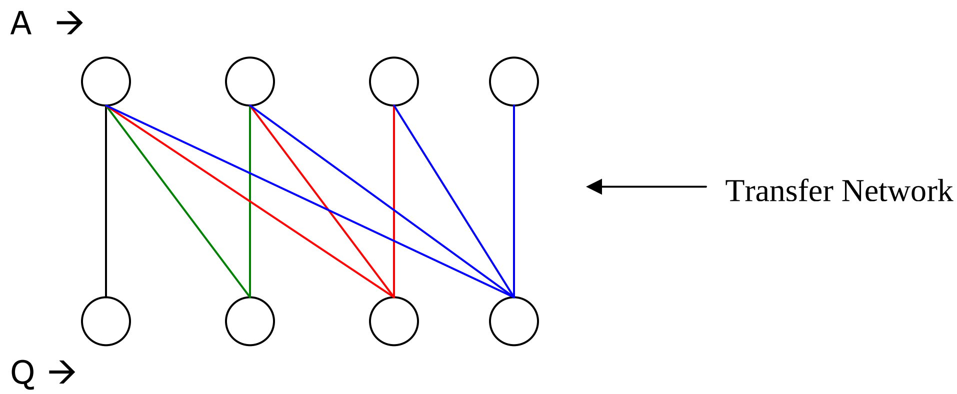
∴ q1 = a1
∴ q2 = a2 – $\text{proj}_{q1}$ a2
∴ q3 = a3 - $\text{proj}_{q1}$ a3 - $\text{proj}_{q2}$ a3

The same can be shown by the following transfer network

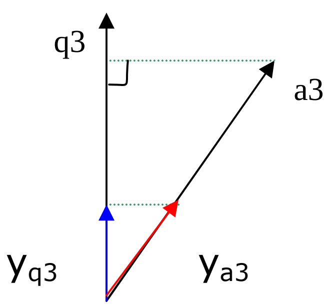
A can be considered to be made of 2 vectors Q and R, such that

A = QR, Q is the orthogonal basis and R is an upper triangular matrix, which is the transfer n/w shown in the figure.

The reason it is easy to invert A is that triangular matrices are simple to invert.




We have the linear combinants of Y in terms of Q, namely Y's coordinate values. But, what we actually want are the linear combinants in terms of A. This can be done in a simple way shown by the figure below.




We find the projection vector of a3 on q3. (Note, for a 3-D space, only a3 can have any energy along q3). We use the same proj vector to find a vector along a3 such that $y_{q3}$ is its projection. By similarity we see that this all the energy Y can have along a3. We subtract this from Y, so the only components left in Y would be of a1 and a2. They can too be found in the same way.

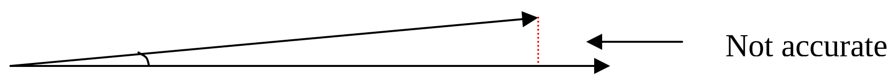
<u>Why this method is not that good</u>
It is easy to see in this method that the error seeps through. If we have to find the Q for n vector space, the final error is n times as large as the initial error.

<u>Solution</u>: Modified Graham Schimdt's Method
Similar to GSM, but after finding an ortho vector qn, we find the energy of each remaining vector in A along qn and subtract this energy. It is easy to see that the two methods are equivalent, but the

energy is reduced stepwise from each vector. This gives better numerical results.

Why both methods are bad

For ill-conditioned vectors, we get numerically bad results. Consider two vectors very close to each other, the result after subtraction has low resolution, giving bad results.




II Givens Rotation

Instead of finding the orthogonal basis, we start with an orhto basis and then rotate it into to mach the given subspace. The error here will be because of the fact that the subspaces spanned by the 2 basis may not be identical, but the error doesn't get amplified proportionally to the no. of vectors in the basis.

The important thing to note here is that are doing planar rotations, so at a time only 2 co-ordinates changes, the ones which are in the plane of rotations

We need to match q1 with a1, q2 with a2 and so on. We do this in steps.

1. Take a1 and rotate it into q1 (the inverse rotations are the ones required to rotate q1 into a1).
2. Then we rotate a2 into the plane of q1 & q2 such that not component along q1 does not change.

We successively do this till A is rotated into Q.

Consider a 3-D space. If we want to rotate a1 into X axis, we could first rotate in the X-Z plane and then in the X-Y plane or first in the X-Y plane and then in the X-Z plane. We choose the following method

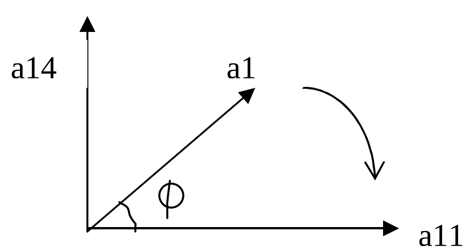
> Choose the plane with one axis as the one you want to zero out, and the other the axis which we want to move into.

For a 4-D space

a1 = $\begin{pmatrix} a11 \\ a12 \\ a13 \\ a14 \end{pmatrix}$   a2 = $\begin{pmatrix} a21 \\ a22 \\ a23 \\ a24 \end{pmatrix}$   and so on.

For moving a1 in q1, we want to zero out a14 component, so choose the a11-a14 plane of rotation.

$$\therefore \text{ Rotation matrix } = \begin{pmatrix} \cos\Phi & 0 & 0 & \sin\Phi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sin\Phi & 0 & 0 & \cos\Phi \end{pmatrix}$$




Then we use a11-a13 plane of rotation followed by a11 – a12.

For rotating a2 into X-Y plane, we have to ensure that X component is not in the plane of rotation.
So, we choose a22 – a24 as plane of rotation and then a22 – a23 as plane of rotation.

We want to solve for 'x' using least squares

$\therefore$ Ax = B

$\because$ A = QR

QR x = B
Rx = Q' B   $(Q' = Q^{-1})$
x = $R^{-1}$ Q'B

Since, we need the inverse of Q anyway, we do not solve for the inverse rotations. The above rotation matrices are $Q^{-1}$, and the upper triangular matrix formed by applying these transformations to A, is R.

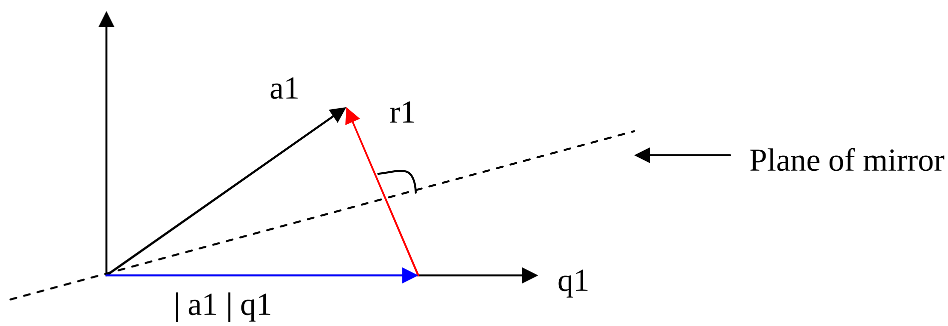
Householder Reflections

"No our reflections are upside down, left-right flipped, back-fwd flipped. The man in the mirror is actually an alien, with no symmetry or is it up down symmetric. It's all about how we turn!"

"Dabum dubum thubum…!"

Householder Reflections

The reason we use reflections to match the orthogonal basis to A, is that reflections are less expensive than rotations.




Therefore, r1 = a1 - |a1 | q1

We want to reflect A in the direction of r1.
This is given by $r_1^T$ A. If we think of A to consists of two parts, one component along in the mirror, and the other ortho to it. So, the component in the mirror does not change, and the other gets flipped.

Therefore, after flipping we get $- r_1^T A$.
Then we reconstruct the vector i.e $- r_1 r_1^T A$.
The component that doesn't change is $A - r_1 r_1^T A$.

∴ the final vector is the addition of these two vectors

$A - r_1 r_1^T A - r_1 r_1^T A$
$= (I - 2 r_1 r_1^T) A$

For, reflecting a2 into q2, we must ensure that component along q1 doesn't change. So, we flatten this dimension, so that it is just point and lies in the mirror.

$$a2 = \begin{pmatrix} a21 \\ a22 \\ a23 \\ a24 \end{pmatrix} \begin{matrix} \leftarrow \text{Don't consider this component} \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \end{matrix}$$

Thus Q is the matrix formed by $(I - r\, r^T)$ and R is the triangular matrix formed transforming A.