A Review of Linear Algebra

(see Gilbert Strang: Linear Algebra and its Applications, 3rd edition for all the material that follows).

Linearly-independent (page 80).

Given a collection of vectors:

$$\mathbf{V}_1 = \begin{bmatrix} v_{1,1} \\ \vdots \\ v_{1,N} \end{bmatrix}, \quad \ldots, \quad \mathbf{V}_k = \begin{bmatrix} v_{k,1} \\ \vdots \\ v_{k,N} \end{bmatrix},$$

We say that $V_1, V_2, \ldots, V_k$ are linearly-independent if

$$c_1 V_1 + c_2 V_2 + \cdots + c_k V_k = 0$$

implies that: $c_1 = c_2 = \cdots = c_k = 0$.

If a set of vectors is not linearly-independent, then it is linearly-dependent.

Column Space (page 90)

If $V_1, V_2, \ldots, V_k$ represent the columns of a matrix, then

$$A x = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ V_1 & V_2 & \cdots & V_k \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = x_1 V_1 + x_2 V_2 + \cdots + x_k V_k$$

for arbitrary x, will span the column space of A.

Null Space

Given a matrix A, the set of vectors $\{x_i\}$ satisfying $Ax_i = 0$ define the <u>nullspace</u> of A. If A is invertible, $Ax_i = 0 \Rightarrow x_i = A^{-1}0 = \underline{0}$ (only one vector in the null-space).

Left Null Space

Given a matrix A, the left nullspace is the set of vectors y_i:

such that $A^T y_i = 0$.

Row Space

The row-space of A is the column-space of A^T.

Projections and Least Squares Approximations

(from Strang: "Linear Algebra and its Applications", 3rd edition, section 3.3 (review chapter 3))

Theorem The least-squares solution to an inconsistent system $Ax=b$ of m equations in n unknowns satisfies (n ≠ m, else P = I, page 159 for A square, invertible)

$$A^T A \bar{x} = A^T b$$

If the columns are linearly independent, then $A^T A$ is invertible and:

$$\bar{x} = (A^T A)^{-1} A^T b$$

The projection of b onto the column space is therefore:

$$p = A\bar{x} = A(A^T A)^{-1} A^T b$$

(inconsistent means b is not in the column-space of A.)

Notes:

1. $P = A(A^T A)^{-1} A^T$ is called a projection matrix.
2. $P^2 = P$
3. $P^T = P$
4. Any symmetric matrix with $P^2 = P$, represents a projection.
5. $(I-P)$ will project b onto the nullspace of the columns of A. NOTE $(I-P)b + Pb = b$ (Perfect reconstruction

Interpretation 1

Suppose that we would like to express every 2×2 subimage in terms of 2 subimages

■ □
□ ■

and

■ ■
■ ■

. What would be the least-squares decomposition and how can it be computed?

The problem is:

Find α, β such that:

$$\begin{bmatrix} c & d \\ e & f \end{bmatrix} \cong \alpha \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \beta \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

is optimal in that:

$$\left\| \begin{bmatrix} c & d \\ e & f \end{bmatrix} - \alpha \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \beta \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\|^2$$

is minimal.

First, we note that $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ are linearly independent since:

$$\alpha \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \beta \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \Longrightarrow \alpha = \beta = 0.$$

To reformulate the problem in terms of vectors, we scan each matrix rowise:

Replace $\begin{bmatrix} c & d \\ e & f \end{bmatrix}$ by $\begin{bmatrix} c \\ d \\ e \\ f \end{bmatrix}$

Replace $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ by $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

Replace $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ by $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

The approximation problem is one of determining α, β so that:

$$\begin{bmatrix} c \\ d \\ e \\ f \end{bmatrix} \cong \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Rewrite RHS as:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} c \\ d \\ e \\ f \end{bmatrix}$$

Define $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$, $x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, $b = \begin{bmatrix} c \\ d \\ e \\ f \end{bmatrix}$

Using the least-squares theorem, define the Projection:

$$P = A(A^T A)^{-1} A^T$$

The least-squares $\bar{x}$ is given by:

$$A\bar{x} = Pb. \quad \blacksquare$$

Interpretation 2

We modify the first problem:

We recognize that 2×2 subimages have $2 * 2 = 4$ degrees of freedom, while the specified subimages only have 2. Can we find another 2 subimages that capture the other 2 degrees of freedom?

If we had these two subimages, we can express any 2×2 subimage in-terms of 4.

In order to answer this question, we need to introduce the Singular Value Decomposition SVD.

Interpretation 3

Not every subimage satisfies:

$$A\bar{x} = b$$

Least-square guarantees that among all possible choices for $\bar{x}$, $\bar{x}$ is such that:

$$\|b - A\bar{x}\|^2 \text{ is } \underline{\text{minimized}}.$$

Here $\left\| \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \right\|^2 = \sum_{i=1}^{N} y_i^2$.

When $A\bar{x} = b$ is solvable, $\|b - A\bar{x}\|^2 = 0$, otherwise it is <u>not</u>.

Weighted-Least Squares (p 204 of Strang)

For an incosistent system $Ax=b$, we can weigh the system using W by:

$$(WA)x = (Wb)$$

The least-squares solution (asuming that the (WA) columns are linearly-independent), satisfies:

$$(WA)^T (WA)\bar{x} = (WA)^T (Wb)$$

or:

$$(A^T W^T W A)\bar{x} = A^T W^T W b$$

As before, the solution is now given by:

$$\bar{x} = (A^T W^T W A)^{-1} (WA)^T (Wb)$$ ∎

The new solution is minimizing $\|\bar{x}\|_W = \|W\bar{x}\|$ for <u>W invertible</u>.

In this case, the inner-product is:

$$(x,y)_W = (Wy)^T(Wx) = y^T(W^TW)x$$

For random-vectors, $D = W^TW$ is the inverse of the covariance matrix, and $D = C^{-1}$, $C(i,j) = E\{(b_i - \bar{b}_i)(b_j - \bar{b}_j)\}$

How succesful is our Transform?

We will assume that a transform is succesful, if it captures most of the energy in the signal:

$\left\| b - A\bar{x} \right\|^2$ is very small compared to $\|b\|^2$.

For example: $\dfrac{\|b - A\bar{x}\|^2}{\|b\|^2} < 5\,\%$.

Computing the energy in the transform domain, from the components of $\bar{x}$ is <u>not</u> always possible for Least-squares projections, because lengths are not generally preserved.

However, if the columns of A are <u>orthonormal</u>:

$$A^T A = \begin{bmatrix} \longleftarrow v_1 \longrightarrow \\ \longleftarrow v_2 \longrightarrow \\ \vdots \\ \longleftarrow v_k \longrightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & \cdots & v_k \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

Then $\|Ax\|^2 = (Ax)^T (Ax) = x^T A^T A x = \|x\|^2$, and we may thus use $\|x\|^2$ instead of $\|Ax\|^2$.

In this case $\dfrac{\|x\|^2}{\|b\|^2} > 95\%$ indicates a good transform

Least-Squares Examples

(Examples from Strang, 3rd edition, page 179).

Compute a least-squares approximation to $y = x^5$ using a straight line $C + Dx$ for $x \in [0, 1]$.

Solution 1 Minimize $E^2 = \int_0^1 [x^5 - (C + Dx)]^2 \, dx$ w.r.t. C, D:

$$E^2 = \tfrac{1}{11} - \tfrac{2}{6} C - \tfrac{2}{7} D + C^2 + CD + \tfrac{1}{3} D^2$$

$$\frac{\partial E^2}{\partial C} = -\tfrac{2}{6} + 2C + D = 0$$

$$\frac{\partial E^2}{\partial D} = -\tfrac{2}{7} + C + \tfrac{2}{3} D = 0$$

which leads to:

$$\begin{bmatrix} 2 & 1 \\ 1 & 2/3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1/3 \\ 2/7 \end{bmatrix}$$

which can be solved for C, D.

<u>solution 2</u>: Use an orthonormal basis ex-

Apply Gram-Schmidt orthogonalization to obtain an orthonormal expansion, given $1, x$ as basis functions.

The inner-product is:

$$< \phi(x), \psi(x) > = \int_0^1 \phi(x)\,\psi(x)\,dx$$

for ϕ, ψ being real-valued functions, and integration over the given range of $[0, 1]$.

The two basis function given are:

$$\phi_1(x) = 1, \quad x \in [0,1]$$

$$\phi_2(x) = x, \quad x \in [0,1]$$

Using Gram-Schmidt, an orthonormal-basis is obtained using:

$$\psi_1(x) = \frac{\phi_1(x)}{\sqrt{<\phi_1(x), \phi_1(x)>}} = \frac{1}{\sqrt{\int_0^1 1\,dx}} = \underline{\underline{1}}$$

a) $\psi_2'(x) = x - <\psi_2(x), 1> \cdot 1 = x - \int_0^1 x\,dx = \underline{\underline{x - 1/2}}$

(b) Normalize: $\psi_2(x) = \dfrac{\psi_2'(x)}{\| \psi_2'(x) \|}$

$$\psi_2(x) = \frac{x - 1/2}{\sqrt{\int_0^1 (x-1/2)^2\,dx}} = \frac{x - 1/2}{\sqrt{1/12}}$$

Now, to expand x^5 as $c + Dx$, we have:

$$C + Dx = \langle x^5, \psi_1(x) \rangle \psi_1(x) + \langle x^5, \psi_2(x) \rangle \psi_2(x)$$

$$= \int_0^1 x^5\,dx \cdot 1 + 12 \int_0^1 x^5 (x - 1/2)\,dx \quad (x - 1/2)$$

$$= 1/6 + 12\left(1/7 - 1/12\right)(x - 1/2)$$

$$= 1/6 + 5/7\,(x - 1/2)$$

$$\Rightarrow \quad C = 1/6 - 5/14 ,$$

$$D = 5/7$$

Solution 3: Use a least-squares Projection.

Place the basis functions along the columns of A:

$$\underbrace{\begin{bmatrix} \uparrow & \uparrow \\ 1 & x \\ \downarrow & \downarrow \end{bmatrix}}_{\text{Note that these two are functions}} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{\text{This is regular scalar multiplication}} = \underbrace{\begin{bmatrix} \uparrow \\ x^5 \\ \downarrow \end{bmatrix}}_{\text{This is also a function}}$$

It is now in the form "$Ax = b$"

For the least-squares projected solution, multiply by A^T on both sides:

$$\begin{bmatrix} \longleftarrow 1 \longrightarrow \\ \longleftarrow x \longrightarrow \end{bmatrix} \underbrace{\begin{bmatrix} \uparrow & \uparrow \\ 1 & x \\ \downarrow & \downarrow \end{bmatrix}}_{\text{Multiplication of function by function is done using the inner-product } \langle \cdot, \cdot \rangle} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{\text{scalar multiplication}} = \begin{bmatrix} \leftarrow 1 \rightarrow \\ \leftarrow x \rightarrow \end{bmatrix} \underbrace{\begin{bmatrix} \uparrow \\ x^5 \\ \downarrow \end{bmatrix}}_{\text{Use } \langle \cdot, \cdot \rangle}$$

$$\begin{bmatrix} \leftarrow 1 \rightarrow \\ \leftarrow x \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow \\ 1 & x \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} \langle 1, 1 \rangle & \langle 1, x \rangle \\ \langle x, 1 \rangle & \langle x, x \rangle \end{bmatrix}$$

$$= \begin{bmatrix} \int_0^1 1\,dx & \int_0^1 x\,dx \\ \int_0^1 x\,dx & \int_0^1 x^2\,dx \end{bmatrix} = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} x^5 \end{bmatrix} = \begin{bmatrix} \int_0^1 x^5\,dx \\ \int_0^1 x^6\,dx \end{bmatrix} = \begin{bmatrix} 1/6 \\ 1/7 \end{bmatrix}$$

Hence, the basic equation becomes:

$$\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1/6 \\ 1/7 \end{bmatrix}$$

which can be solved for C, D.

5.5 Complex Matrices:

Symmetric vs Hermitian

Orthogonal vs Unitary

(Gilbert Strang, Section 5.5).

A matrix A is <u>symmetric</u> if: $A = A^T$

(or $a_{ij} = a_{ji}$)

A matrix A is <u>hermitian</u> if: $A = A^H$

(or $a_{ij} = \bar{a}_{ji}$).

<u>Properties</u>

1. $A = A^H \Longrightarrow x^H A x$ is real.

2. A can be expressed as:

$$A = Q\Lambda Q^T = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ x_1 & x_2 & \cdots & x_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} \leftarrow & x_1^T & \rightarrow \\ \leftarrow & x_2^T & \rightarrow \\ & \vdots & \\ \leftarrow & x_n^T & \rightarrow \end{bmatrix}$$

$$= \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ x_1 & x_2 & \cdots & x_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \leftarrow & \lambda_1 x_1^T & \rightarrow \\ \leftarrow & \lambda_2 x_2^T & \rightarrow \\ & \vdots & \\ \leftarrow & \lambda_n x_n^T & \rightarrow \end{bmatrix}$$

$$= \lambda_1 x_1 x_1^T + \lambda_2 x_2 x_2^T + \cdots + \lambda_n x_n x_n^T$$

where: $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues,

$x_1, x_2, \ldots, x_n$ are the eigenvectors, and they satify: (i) the eigenvalues are real,

(ii) the eigenvectors are orthonormal.

Recall that:

$$A x_i = \lambda_i x_i$$

implies that x_i is an eigenvector, and λ_i is the corresponding eigenvalue (Section 5.1 of Strang).

Unitary Matrices

Suppose that $U^H U = I$, where U is a square matrix. Then:

1. $\|Ux\|^2 = \|x\|^2$

Proof: $\|Ux\|^2 = (Ux)^H Ux = x^H U^H U x = x^H x = \|x\|^2$

2. All the eigenvalues of U are $|\lambda| = 1$, and all the eigenvectors are orthonormal.

Examples of Unitary Matrices:

(i) The 1-D FFT Matrix

(ii) The 1-D DCT Matrix

(iii) The 1-D orthonormal Discrete Wavelet Transform Matrix

Chapter 6 Positive Definite Matrices (from Strang)

Given a function F of $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$. A Taylor series approximation around y is given by:

$$F(x) = F(y) + (x^T - y^T)\nabla F(y)$$
$$+ \tfrac{1}{2}(x-y)^T A (x-y) + \text{3rd order term}$$

where:

$$\nabla F = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \vdots \\ \frac{\partial F}{\partial x_k} \end{bmatrix},$$

A is the Hessian with:

$$a_{ij} = \frac{\partial^2 F}{\partial x_i \partial x_j}$$

If y is a local minimum, we require

(i) $\nabla F(y) = 0$,

(ii) A is positive definite

Condition (ii) guarantees that $F(x)$ increases away from y since: $\tfrac{1}{2}(x-y)^T A (x-y) > 0$

Tests that are necessary and sufficient for the real-symmetric matrix A to be positive definite:

(I) $x^T A x > 0$ for all $x \neq 0$.

(II) All eigenvalues satisfy: $\lambda_i > 0$

Hence, the general form for A is:

$$A = Q \Lambda Q^T = \begin{bmatrix} \uparrow & & \uparrow \\ x_1 & \cdots & x_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \leftarrow x_1^T \rightarrow \\ \vdots \\ \leftarrow x_n^T \rightarrow \end{bmatrix}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n > 0$.

Singular Value Decomposition (SVD)

(Source: Appendix A of Strang)

Suppose that the columns of A are not linearly independent. For example, let V_1 be expressed as a function of the rest of the vectors:

$$V_1 = c_2 V_2 + c_3 V_3 + \cdots + c_K V_K$$

In this case, we cannot compute the least-square solution to "$Ax = b$" using just the least-squares solution (linearly independence was required).

Furthermore, it is clear that if $\bar{x}$ is such that:

$\| A\bar{x} - b \|^2$ is minimized, then:

$A\bar{y}$ where:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{bmatrix} = \begin{bmatrix} 0 \\ x_2 \\ x_3 \\ \vdots \\ x_K \end{bmatrix} + x_1 \begin{bmatrix} 0 \\ c_2 \\ c_3 \\ \vdots \\ c_K \end{bmatrix}$$

from the elements of $\bar{x}$.

satisfies $A\bar{x} = A\bar{y}$ and hence:

$\| A\bar{y} - b \|^2$ is also minimized.

Hence, we need to define the optimal solution by imposing more constrains:

Find x^+ such that (p 446-447):

(i) $\|Ax^+ - b\|^2$ is minimized, and
(ii) $\|x^+\|$ is minimized.

The solution is computed using the Pseudoinverse A^+. We then have: $\boxed{x^+ = A^+ b}$

As we shall see, the pseudoinverse is easilly computed from the SVD of A.

Singular Value Decomposition

(source: Appendix A of "Linear Algebra and its Applications", 3rd edition, by Gilbert Strang).

SVD Theorem

Any $m \times n$ matrix A can be factored into:

$$A = Q_1 \Sigma Q_2^T$$

where:

Q_1 is m by m, and its columns are the eigenvectors of AA^T (guaranteed to exist since $(AA^T)^T = AA^T$).

Q_2 is n by n, and its columns are the eigenvectors of A^TA (guaranteed to exist since $(A^TA)^T = A^TA$).

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$ satisfy $\sigma_1, \sigma_2, \ldots, \sigma_r > 0$

and $\sigma_i = \sqrt{\underbrace{|\lambda_i(AA^T)|}} = \sqrt{|\lambda_i(A^TA)|}$

meaning the eigenvalue of AA^T

Note: σ_i come from the common eigenvalues.

Interpretations:

1. If A is positive-definite, then the SVD is the regular eigen decomposition $Q \Lambda Q^T$.

2. We can use the SVD to construct basis for every subspace associated with A:

   (i) first r columns of Q_1: column space of A
   (ii) last $m-r$ columns of Q_1: left nullspace of A
   (iii) first r columns of Q_2: row space of A
   (iv) last $n-r$ columns of Q_2: nullspace of A

3. Image Compression

   Idea: transmit eigenvectors for significant σ_i.

$$Q_1 \Sigma Q_2^T = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \cdots & u_m \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} \leftarrow v_1^T \rightarrow \\ \leftarrow v_2^T \rightarrow \\ \vdots \\ \leftarrow v_n^T \rightarrow \end{bmatrix}$$

$$= u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + \cdots + u_r \sigma_r v_r^T + 0 + \cdots + 0$$

note this (and every other) term contain matrices of rank one.

We can hence drop the last terms for small σ_i:

$$\sigma_{n+1} u_{n+1} v_{n+1}^T + \cdots + \sigma_r u_r v_r^T \quad \text{for } \sigma_{n+1}, \ldots, \sigma_r \text{ small}$$

For any image A, transmitting the first p terms requires transmitting $(n+m)p + p$ numbers. 3

The compression ratio is:

$$\frac{n \cdot m}{(n+m)p + p}$$

For a $1000 \cdot 1000$ image, keeping $p=60$ coefficients yields:

$$\frac{1000^2}{2000 \cdot 60 + 60} \simeq \frac{1000}{120} \simeq 8 \text{ times}$$

④ Effective Rank

1. Compute the singular values:

$$\sigma_1, \sigma_2, \ldots, \sigma_r$$

2. Find the $\sigma_i > \varepsilon$ for ε small:

$$\sigma_1, \sigma_2, \ldots, \sigma_p > \varepsilon.$$

3. Take p, the number of $\sigma_i > \varepsilon$ as the effective rank.

⑤ Computing the Pseudo-inverse:

Let $A = Q_1 \Sigma Q_2^T$ (its SVD decomposition).

The pseudo-inverse A^+ is simply:

$$A^+ = Q_2 \Sigma^+ Q_1^T \qquad \underline{\underline{n \times m \text{ size}}}$$

where:

$\sigma_1, \sigma_2, \ldots, \sigma_r$ are the singular values of Σ,

$1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_r$ are the singular values of Σ^+

clearly:

$$\begin{aligned} A^+ A &= (Q_2 \Sigma^+ Q_1^T)(Q_1 \Sigma Q_2^T) \\ &= Q_2 \Sigma^+ (Q_1^T Q_1) \Sigma Q_2^T \\ &= Q_2 (\Sigma^+ \Sigma) Q_2^T \\ &= Q_2 \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} Q_2^T \end{aligned}$$

r entries

n×n matrix

An example

$$A = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

To solve "$Ax = b$", note:

$$A\bar{x} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \end{bmatrix} = \begin{bmatrix} \sigma_1 \bar{x}_1 \\ \sigma_2 \bar{x}_2 \\ 0 \end{bmatrix}$$

To have:

"$A\bar{x} = b$", we get:

$$\sigma_1 \bar{x}_1 = b_1 \Rightarrow \bar{x}_1 = b_1 / \sigma_1,$$
$$\sigma_2 \bar{x}_2 = b_2 \Rightarrow \bar{x}_2 = b_2 / \sigma_2,$$

but $\bar{x}_3$ and $\bar{x}_4$ can be anything.

The <u>minimum-length</u> solution is:

$$x^+ = A^+ b = \begin{bmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1/\sigma_1 \\ b_2/\sigma_2 \\ 0 \\ 0 \end{bmatrix},$$

that is: $\bar{x}_3 = \bar{x}_4 = 0$ for

$$\|\bar{x}\| = \sqrt{\bar{x}_1^2 + \bar{x}_2^2 + \bar{x}_3^2 + \bar{x}_4^2} = \sqrt{\bar{x}_1^2 + \bar{x}_2^2} \quad \text{minimum}$$

Chapter 7: Computation with Matrices (Strang, 3rd edition).

Definition The norm of A is defined by:

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

Using $\|A\|$, we see that:

$$\|Ax\| \leq \|A\| \, \|x\|,$$

in other words, a change in the data $\|Ax\|$ is bounded above by $\|A\| \, \|x\|$. In general, $\|A\| \neq 1$, except for Unitary matrices, or orthonormal matrices (problem 1, page 369).

Definition The condition number of A is:

$$c = \|A\| \, \|A^{-1}\|.$$

If: $Ax = b$, then (p 367, 369)

$$\frac{1}{c} \frac{\|\delta b\|}{\|b\|} \leq \frac{\|\delta x\|}{\|x\|} \leq c \frac{\|\delta b\|}{\|b\|}$$

Also, a change in the matrix such that:
$Ax = b$ and $(A + \delta A)(x + \delta x) = b$ satisfy:

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq c \frac{\|\delta A\|}{\|A\|}$$

Thm (p368) The norm of A is the square root of the largest eigenvalue of $A^T A$:

$$\|A\|^2 = \lambda_{max}(A^T A)$$

If x_{max} corresponds to λ_{max} as the eigenvector of $A^T A$, then

$$\frac{\|Ax\|^2}{\|x\|^2} = \frac{x^T A^T A x}{x^T x} \quad \text{is maximum when } x = x_{max}:$$

$$\frac{\|A x_{max}\|^2}{\|x_{max}\|^2} = \frac{x_{max}^T (A^T A) x_{max}}{1} = x_{max}^T \lambda_{max} x_{max} = \underline{\lambda_{max}}$$

Notes

1) If $A = A^T$, then $\|A\| = \max_i |\lambda_i|$

2) Consider the SVD for A, A^T

$$A = Q_1 \Sigma Q_2^T, \quad A^T = Q_2 \Sigma Q_1^T$$

Then: $A^T A = Q_2 \Sigma^2 Q_2^T$, and by definition: $\|A\| = \sigma_{max}$, the largest σ along Σ.

SVD & Condition Number Examples

Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

Generally, $\det(B - \lambda I) = 0$ is used to find the eigenvalues of B

Then:

$$AA^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix} \Rightarrow \lambda_1 = 1, \lambda_2 = 4, \lambda_3 = 9$$

The eigen decoposition is:

$\lambda = 1$ $\quad AA^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 1 \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$\Rightarrow \begin{cases} x_1 = x_1 \\ x_2 = 4x_2 \Rightarrow x_2 = 0 \\ x_3 = 9x_3 \Rightarrow x_3 = 0 \end{cases}$$

$\Rightarrow$ Set $x_1 = \pm 1$ so that $\sqrt{x_1^2 + x_2^2 + x_3^2} = 1$

$\lambda = 4$ $\quad AA^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 4 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$\Rightarrow \begin{cases} x_1 = -2x_1 \Rightarrow x_1 = 0 \\ 4x_2 = 4x_2 \\ 9x_3 = -2x_3 \Rightarrow x_3 = 0 \end{cases}$$

$\Rightarrow$ Set $x_2 = \pm 1$ so that $\sqrt{x_1^2 + x_2^2 + x_3^2} = 1$

$\lambda_3 = 9$

$$AA^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 9 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$\Rightarrow x_3 = \pm 1 \quad x_1 = x_2 = 0$

$$AA^T = \begin{bmatrix} \uparrow & \uparrow & \uparrow \\ x_1 & x_2 & x_3 \\ \downarrow & \downarrow & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \leftarrow & x_1^T & \rightarrow \\ \leftarrow & x_2^T & \rightarrow \\ \leftarrow & x_3^T & \rightarrow \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is the eigen decomposition.

$A^T A = AA^T$ in this <u>square matrix</u> example

The σ_i are the square-roots of the common eigenvalues:

$$\sigma_1 = \sqrt{9} = 3$$
$$\sigma_2 = \sqrt{4} = 2$$
$$\sigma_3 = \sqrt{1} = 1$$

The <u>SVD</u> is:

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Note that we picked these eigenvectors so that we get -2 for $A(2,2)$ when all is multiplied.

The norm $\|A\| = \sigma_{max} = \underline{\underline{3}}$

For $\|A^{-1}\|$, $\|A^{-1}\| = 1/\sigma_{min} = \underline{\underline{1}}$

Condition-number $c = \|A\| \|A^{-1}\| = \underline{\underline{3}}$

<u>Note</u>: If $\sigma_{min} = 0$, the condition number is <u>infinite</u>.

Interpretation of the Condition Number

Assume that A is quantized to $A+\delta A$.
Assume that the input b is changed to $b+\delta b$.
Also, assume that A is invertible, and A^{-1} exists (Otherwise, the condition number is infinite).

How is the "spectrum" x affected?

Since A^{-1} exists: $b = A^{-1}x$.

Also, let x be the "correct" spectrum, and δx be the error in the spectrum.

Case 1. When b changes to $b+\delta b$:

$$Ax=b, \quad A(x+\delta x) = b+\delta b \Rightarrow \boxed{\delta x = A^{-1}\delta b}$$

Case 2. When A is quantized to $A+\delta A$:

$(A+\delta A)(x+\delta x) = b$ is assumed.

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

$$\cancel{Ax} + A\delta x + \delta A x + \delta A \delta x = \cancel{b}$$

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

In Case 1: $b \to b+\delta b$:

$$\frac{1}{c}\frac{\|\delta b\|}{\|b\|} \leqslant \frac{\|\delta x\|}{\|x\|} \leqslant c\,\frac{\|\delta b\|}{\|b\|}$$

Large condition number c, implies that a relatively small change δb (relatively) can cause a significantly large relative change to either

$\frac{\|\delta x\|}{\|x\|} = \frac{1}{c}\frac{\|\delta b\|}{\|b\|}$, the lower bound, or

$\frac{\|\delta x\|}{\|x\|} = c\,\frac{\|\delta b\|}{\|b\|}$, the upper bound.

In case 2: $A \to A+\delta A$:

$$\frac{\|\delta x\|}{\|x+\delta x\|} \leqslant c\,\frac{\|\delta A\|}{\|A\|}$$

Large condition number c implies that

$\frac{\|\delta x\|}{\|x+\delta x\|} = c\,\frac{\|\delta A\|}{\|A\|}$ can occur for a suitable b.

Ideally: $c=1$ for Unitary matrices.

Building Transforms

Let A contain the basis functions of a desired transformation in its columns:

$$A = \begin{bmatrix} \uparrow & & \uparrow \\ v_1 & \cdots & v_n \\ \downarrow & & \downarrow \end{bmatrix},$$ and we wish to

expand b as: $b \simeq \bar{x}_1 v_1 + \bar{x}_2 v_2 + \cdots + \bar{x}_n v_n$

① To compute $\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_n \end{bmatrix}$, we can use a <u>least-squares projection</u>, when $v_1, v_2, \ldots, v_n$ are linearly independent.

② When $v_1, v_2, \ldots, v_n$ are <u>linearly-dependent</u>, we can use the pseudo-inverse A^+ and:

$$x^+ = A^+ b,$$

which also gives a <u>least-squares</u>, <u>minimum-length</u> solution.

(3) Given a number of basis functions in 1-D, 2-D, or 3-D, we can place them in the columns of A, and form a projection so that:

$$\|A\bar{x} - b\|^2 \text{ is minimized.}$$

For minimum error, we require that:

$$\frac{\partial}{\partial \bar{x}}\left[\|A\bar{x} - b\|^2\right] = 0$$

This means that:
$$\begin{bmatrix} \frac{\partial}{\partial \bar{x}_1}\left[\|A\bar{x} - b\|^2\right] \\ \frac{\partial}{\partial \bar{x}_2}\left[\|A\bar{x} - b\|^2\right] \\ \vdots \\ \frac{\partial}{\partial \bar{x}_n}\left[\|A\bar{x} - b\|^2\right] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

<u>We have:</u>

$$\|A\bar{x} - b\|^2 = (A\bar{x} - b)^H (A\bar{x} - b)$$

$$\frac{\partial}{\partial \bar{x}}\left[\|A\bar{x} - b\|^2\right] = \frac{\partial}{\partial \bar{x}}\left[(A\bar{x} - b)^H\right](A\bar{x} - b) + (A\bar{x} - b)^H \frac{\partial}{\partial \bar{x}}\left[A\bar{x} - b\right]$$

Hence:

$$0 = A^H (A\bar{x} - b) + (A\bar{x} - b)^H A$$
$$= A^H (A\bar{x} - b) + \left(A^H (A\bar{x} - b)\right)^H$$

Since $\bar{x}$ can be anything, we see that we get zero if and only if

$$\boxed{A^H (A\bar{x} - b) = 0}$$

If $e = A\bar{x} - b$ is the error in our transform, then $A^H e = 0$ implies that the error in our approximation is orthogonal to the columns of A^H. By definition, this implies that e belongs to the left nullspace of A (which is the nullspace of A^H of-course)

Let V, W be:

V is the column-space of A.
W is the left-nullspace of A.

Then:

$$\mathbb{R}^m = V + W \quad \text{(eg4, p199 and 3D, page 139)}$$

and

$$V = W^\perp, \quad W = V^\perp$$

(for an $m \times n$ matrix)

If $\underset{m\times m}{A = Q_1} \Sigma \underset{n\times n}{Q_2^T}$ (its SVD decomposition),

then: the last $m-r$ columns of Q_1 will form an <u>orthonormal basis</u> for the left nullspace.

$$Q_1 = \begin{bmatrix} \uparrow & & \uparrow & & \uparrow \\ u_1 & \cdots & u_{r+1} & \cdots & u_m \\ \downarrow & & \downarrow & & \downarrow \end{bmatrix}$$

$m-r$ columns

Since they are orthonormal, we can form a projection by transposing:

"
$$\underset{m\times(m-r)}{\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_{r+1} & u_{r+2} & \cdots & u_m \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}} \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_{m-r} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$
"

has a least-squares projection:

$$P_{ln(A)} = \begin{bmatrix} \leftarrow u_{r+1}^T \rightarrow \\ \leftarrow u_{r+2}^T \rightarrow \\ \vdots \\ \leftarrow u_m^T \rightarrow \end{bmatrix}$$

so that:

$$\begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_{m-r} \end{bmatrix} = P_{ln(A)} b$$

meaning left-nullspace of A.

Given an arbitrary input b, for a transform represented by A, compute: 5

$v = P_A b \in V$, least-squares projection on columns of A.

$w = P_{\mathcal{N}(A)} b \in W$, least-squares projection on left-nullspace of A.

Then: $v \perp w$, and:

$$\boxed{Av + Bw = b}$$ <u>Perfect Reconstruction</u>

where:

$$B = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_{r+1} & u_{r+2} & \cdots & u_m \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

The columns are an orthonormal basis for the left-nullspace of A.

The condition number for the transformation associated with A, B is the condition number of $[A \; B]$.