Lecture 12

Plane Transformations

& Least Squares

Plane Transformations

Consider a general 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Clearly, A can only multiply vectors in $I\!R^2$ and the result of such a multiplication is another vector in $I\!R^2$. Note that we can define a *function* $T(\boldsymbol{x}) = A\boldsymbol{x}$ in this manner, *i.e.*

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix}.$$

We say that T maps from $I\!R^2$ to $I\!R^2$ and it is actually an example of a *linear transformation*.

Notice that $A = I$ is not particularly interesting, as $A\boldsymbol{x} = I\boldsymbol{x} = \boldsymbol{x}$.
Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. Then

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = 2\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

i.e. multiplication by A results in a *stretch* of the vector $\boldsymbol{x}$ by a factor of 2.

Similarly, a matrix of the form $A = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$ would result in stretch by a factor k, assuming $k > 1$. (The effect in the case of $0 < k < 1$ is called a *compression*.)

Reflections can also be easily simulated with matrix multiplication. Consider $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Then,

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix},$$

i.e. the resulting vector (or point, if you like) is a *reflection* of the original vector in the x_1 axis. Similarly, $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ yields

$$A\boldsymbol{x} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix},$$

i.e. a reflection in the line $x_2 = x_1$!

Finally, rotations can be induced by the general *rotation matrix*,

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Let $\theta = \frac{\pi}{4}$ and $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then

$$A\boldsymbol{x} = \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\frac{\pi}{4} \\ \sin\frac{\pi}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Note that the resulting vector still has length 1, but it is has been rotated by $\theta = 45^\circ$ in the *anti-clockwise* direction. Similarly, for a general θ, this matrix will induce an anti-clockwise rotation of angle θ.

The other type of operation we can achieve with a matrix multiplication is a *shear* in the x_1 and x_2 directions. Anyhow, it be can be easily shown that *multiplication by an invertible* 2×2 *matrix is equivalent to a sequence of shears, compressions, stretches and reflections.*

Overdetermined & Inconsistent Equations

Systems of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ can either be consistent (*i.e.* a solution exists) or inconsistent (*i.e.* no solution exists). A system will be inconsistent if $r(A) \neq r([A|\boldsymbol{b}])$. In situations where no solution exists but a solution is still required, we then search for the best possible approximate solution which is called a least squares solution.

Consider the system of m linear equations in n unknown variables $A\boldsymbol{x} = \boldsymbol{b}$ where A is an $m \times n$ matrix, where $m > n$ (*i.e.* we have more equations than unknown variables). Such systems are called overdetermined systems of linear equations and for these we typically expect to find the system to be inconsistent.

Ex: Show that the following overdetermined system of linear equations is inconsistent.

$$\begin{aligned} x_1 - 2x_2 &= 3 \\ x_1 + 2x_2 &= -1 \\ -2x_1 + 3x_2 &= -4 \end{aligned}$$

The pseudoinverse of A (also called the generalized inverse or Moore-Penrose inverse) is denoted by pinv(A), and defined as

$$\text{pinv}(A) = (A^T A)^{-1} A^T$$

Multiplying each side of $A\boldsymbol{x} = \boldsymbol{b}$ by A^T leads to the normal system of equations given by

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

The solution to these normal equations is what is referred to as the least squares solutions of the system $A\boldsymbol{x} = \boldsymbol{b}$. If the system $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent, then these least squares solutions are taken as the approximate solution to $A\boldsymbol{x} = \boldsymbol{b}$.

The normal system of equations is itself a system of linear equations with an $n \times n$ coefficient matrix $(A^T A)$. If $r(A) = n$ then the square matrix $(A^T A)$ is invertible, which leads to a unique least squares solution, denoted by $\widehat{\boldsymbol{x}}$. Multiplying both sides of the normal equations by the inverse $(A^T A)^{-1}$,

$$\begin{aligned}
(A^T A)^{-1}(A^T A)\widehat{\boldsymbol{x}} &= (A^T A)^{-1} A^T \boldsymbol{b} \\
I\widehat{\boldsymbol{x}} &= (A^T A)^{-1} A^T \boldsymbol{b} \\
\widehat{\boldsymbol{x}} &= (A^T A)^{-1} A^T \boldsymbol{b}
\end{aligned}$$

i.e. $$\widehat{\boldsymbol{x}} = \text{pinv}(A)\boldsymbol{b}$$

This unique least squares solution $\widehat{\boldsymbol{x}} = \text{pinv}(A)\boldsymbol{b}$ is generally taken to be the best approximate solution to the inconsistent system.

Ex: Determine the least squares solution for the inconsistent system of linear equations

$$\begin{aligned}
x_1 - 2x_2 &= 3 \\
x_1 + 2x_2 &= -1 \\
-2x_1 + 3x_2 &= -4
\end{aligned}$$

If the system of equations $A\boldsymbol{x} = \boldsymbol{b}$ actually has a unique solution, then the least squares solution will be that unique solution (*i.e.* $\widehat{\boldsymbol{x}} = \boldsymbol{x}$).

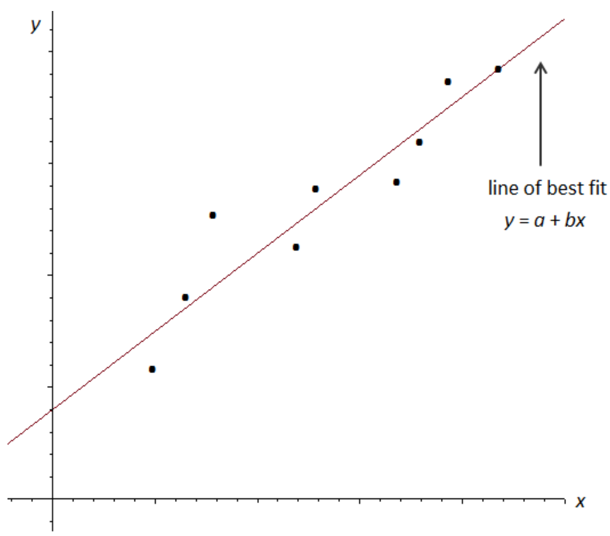
When solving the normal equations $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$ it is computationally more efficient to use Gaussian Elimination rather than to determine $(A^T A)^{-1}$ and then take the product $(A^T A)^{-1} A^T\boldsymbol{b}$.

<u>**Ex:**</u> For the following linear system, use Gaussian Elimination to find the least squares solution to the normal equations.

$$\begin{aligned} x_1 - 2x_2 &= 3 \\ x_1 + 2x_2 &= -1 \\ -2x_1 + 3x_2 &= -4 \end{aligned}$$

Least Squares Lines & Curves

A common task in engineering is the gathering of experimental data, in the form of data points (x_i, y_i), which is then used to develop a relationship between the variables x and y allowing us to predict new values of y for given values of x. Often when these points (x_i, y_i) are plotted they seem to lie close to a line and hence we wish to find an equation of a line $y = a + bx$ that best fits these points.



When carrying out an engineering investigation typically more than two data points will be collected. However having multiple data points leads to an overdetermined system when trying to fit a linear equation $y = a + bx$, and being overdetermined we would expect the system to be inconsistent, having no actual solution for a and b.

Given m data points (x_1, y_1), (x_2, y_2),...,(x_m, y_m), where at least two of the x_i are distinct, our objective is to find a linear function $y = a_0 + a_1 x$ that best fits the data points (x_i, y_i). Substituting the points (x_i, y_i) into the linear function gives the linear system of equations,

$$
\begin{aligned}
a_0 + a_1 x_1 &= y_1 \\
a_0 + a_1 x_2 &= y_2 \\
a_0 + a_1 x_3 &= y_3 \\
&\vdots \\
a_0 + a_1 x_m &= y_m
\end{aligned}
$$

which can be written in matrix form $A\boldsymbol{x} = \boldsymbol{b}$,

where $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$.

This linear system $A\boldsymbol{x} = \boldsymbol{b}$ is typically inconsistent, due to being overdetermined. Therefore we wish to find a least squares solution (*i.e.* best approximate solution) to $A\boldsymbol{x} = \boldsymbol{b}$.

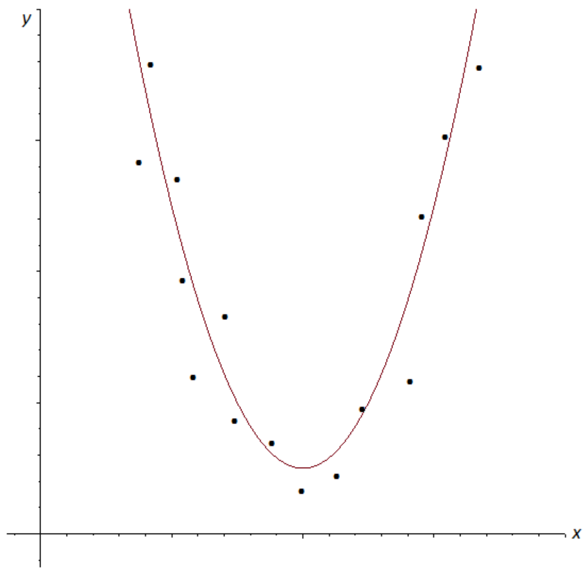
Since at least two of the x_i are distinct it means $r(A) = 2$, and as a result $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least square solution,

$$\widehat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b} = \text{pinv}(A)\boldsymbol{b}$$

The least squares solution $\widehat{\boldsymbol{x}}$ to the system $A\boldsymbol{x} = \boldsymbol{b}$ gives the coefficients a_0 and a_1 of the linear function $y = a_0 + a_1 x$, we refer to this line as the least squares line.

<u>**Ex:**</u> Find an equation of the least squares line for the data points $(-1, 0)$, $(0, 1)$, $(1, 2)$, $(2, 4)$.

In some situations a straight line does a poor job in representing the data points. In cases such as this it would make more sense to find a parabola.



Suppose we are given m data points (x_1, y_1), (x_2, y_2),...,(x_m, y_m), where at least $n + 1$ of the x_i are distinct (in particular $m \geq n + 1$) and we are interested in finding the polynomial of degree n,

$$y = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$$

that best fits the given data.

By analogy with the preceding least squares line analysis, we define

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ 1 & x_3 & x_3^2 & \cdots & x_3^n \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix}, \; \boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \; \boldsymbol{b} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

As the linear system of equations $A\boldsymbol{x} = \boldsymbol{b}$ is overdetermined, we expect the system to be inconsistent. Given that we can't determine a polynomial satisfying all the data points we instead find the best fitting approximating polynomial.

The condition that at least $n + 1$ of the x_i are distinct ensures that $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least square solution,

$$\widehat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b} = \text{pinv}(A)\boldsymbol{b}$$

This least squares solution $\widehat{\boldsymbol{x}}$ gives the coefficients a_0, a_1, a_2,..., a_n of the polynomial function $y = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$, this polynomial is called the least squares polnomial of degree n.

Ex: Find a quadratic least squares polynomial $y = a_0 + a_1 x + a_2 x^2$ for the data points $(-3, 5)$, $(-1, 3)$, $(0, 2)$, $(1, 2)$, $(3, 4)$.