Math 416 - Abstract Linear Algebra
Fall 2011, section E1
Least squares solution

1. Curve fitting

The least squares solution can be used to fit certain functions through data points.

Example: Find the best fit line through the points $(1,0),(2,1),(3,1)$.

Solution: We are looking for a line with equation $y = a + bx$ that would ideally go through all the data points, i.e. satisfy all the equations

$$\begin{cases} a + b(1) = 0 \\ a + b(2) = 1 \\ a + b(3) = 1. \end{cases}$$

In matrix form, we want the unknown coefficients $\begin{bmatrix} a \\ b \end{bmatrix}$ to satisfy the system

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

but the system has no solution. Instead, we find the least squares fit, i.e. minimize the sum of the squares of the errors

$$\sum_{i=1}^{3} |(a + bx_i) - y_i|^2$$

which is precisely finding the least squares solution of the system above. Writing the system as $A\vec{c} = \vec{y}$, the normal equation is

$$A^T A\vec{c} = A^T\vec{y}$$

and we compute

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

$$A^T\vec{y} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}.$$

The normal equation has the unique solution

$$\vec{c} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ \frac{1}{2} \end{bmatrix}$$

so that the best fit line through the data points is $y = -\frac{1}{3} + \frac{1}{2}x$.

Remark: If we hate the formula for the inverse of a 2×2 matrix, or if we need to solve a bigger system, we can always use Gauss-Jordan:

$$\left[\begin{array}{cc|c} 3 & 6 & 2 \\ 6 & 14 & 5 \end{array}\right] \sim \left[\begin{array}{cc|c} 3 & 6 & 2 \\ 0 & 2 & 1 \end{array}\right] \sim \left[\begin{array}{cc|c} 3 & 0 & -1 \\ 0 & 2 & 1 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & -\frac{1}{3} \\ 0 & 1 & \frac{1}{2} \end{array}\right].$$

The unique solution to the system is indeed $\begin{bmatrix} -\frac{1}{3} \\ \frac{1}{2} \end{bmatrix}$.

2. Arbitrary inner product spaces

Just like Gram-Schmidt, the least squares method works in any inner product space V, not just $\mathbb{R}^n$ (or $\mathbb{C}^n$). Assume that the subspace $E \subseteq V$ onto which we are projecting is finite-dimensional.

Example: Consider the real inner product space $C[0,1] := \{f : [0,1] \to \mathbb{R} \mid f \text{ is continuous }\}$ with its usual inner product

$$(f, g) = \int_0^1 f(t)g(t)\,\mathrm{d}t.$$

Find the best approximation of the function t^2 by a polynomial of degree at most one.

Solution using least squares: We are looking for a polynomial of degree at most one $a + bt$ that would ideally satisfy

$$a + bt = t^2$$

which is clearly impossible, i.e. $t^2 \notin \operatorname{Span}\{1, t\}$. The best approximation is the vector in $\operatorname{Span}\{1, t\}$ minimizing the error vector

$$a + bt - t^2$$

which is achieved exactly when the error vector is orthogonal to $\operatorname{Span}\{1, t\}$. This imposes two conditions:

$$\begin{cases} (a + bt - t^2, 1) = 0 \\ (a + bt - t^2, t) = 0 \end{cases}$$

which we can rewrite as

$$\begin{cases} a(1,1) + b(t,1) = (t^2, 1) \\ a(1,t) + b(t,t) = (t^2, t) \end{cases}$$

or in matrix form:

$$\begin{bmatrix} (1,1) & (t,1) \\ (t,1) & (t,t) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} (t^2,1) \\ (t^2,t) \end{bmatrix}.$$

(This is the normal equation. The coefficient matrix here plays the role of $A^T A$ in the previous example, i.e. the square matrix of all possible inner products between vectors in the basis of E, in this case $\{1, t\}$. Likewise, the right-hand side plays the role of $A^T \vec{y}$ in the previous example, i.e. the list of all possible inner products between the basis vectors $\{1, t\}$ of E and the vector t^2 not in E which we want to project down to E.)

Computing the inner products involved, the system can be written as

$$\left[\begin{array}{cc|c} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \end{array}\right]$$

which we now solve:

$$\left[\begin{array}{cc|c} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & 1 & 1 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & -\frac{1}{6} \\ 0 & 1 & 1 \end{array}\right].$$

The least squares solution is $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -\frac{1}{6} \\ 1 \end{bmatrix}$ so that the best approximation of t^2 by a polynomial of degree at most one is $-\frac{1}{6} + t$.

Remark: If we hate Gauss-Jordan, we can always use the formula for the inverse of a 2×2 matrix, so that the unique solution to the system is

$$\begin{aligned}
\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{4} \end{bmatrix} &= \frac{1}{1/12} \begin{bmatrix} \frac{1}{3} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{4} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{3} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} \\
&= \frac{1}{6} \begin{bmatrix} 2 & -3 \\ -3 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} \\
&= \frac{1}{6} \begin{bmatrix} -1 \\ 6 \end{bmatrix} \\
&= \begin{bmatrix} -\frac{1}{6} \\ 1 \end{bmatrix}.
\end{aligned}$$

http://www.youtube.com/watch?v=lBdASZNPIv8

Solution using Gram-Schmidt: In a previous exercise, we obtained the orthonormal basis $\{u_1 = 1, u_2 = \sqrt{3}(2t-1)\}$ of Span$\{1, t\}$. Using this, we compute the projection

$$\begin{aligned}
\mathrm{Proj}_{\{1,t\}}(t^2) &= \mathrm{Proj}_{\{u_1,u_2\}}(t^2) \\
&= (t^2, u_1)u_1 + (t^2, u_2)u_2 \\
&= (t^2, 1)\,1 + (t^2, \sqrt{3}(2t-1))\,\sqrt{3}(2t-1) \\
&= \frac{1}{3} + 3\left(2(t^2, t) - (t^2, 1)\right)(2t-1) \\
&= \frac{1}{3} + 3\left(\frac{2}{4} - \frac{1}{3}\right)(2t-1) \\
&= \frac{1}{3} + \frac{1}{2}(2t-1) \\
&= -\frac{1}{6} + t.
\end{aligned}$$