Lecture12

Lecture 12:
Applications of Projections: Linear Least Squares problems

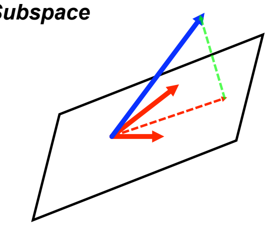
Outline:

1) Quick Review:
   A) projections and least-squares problems
   B) properties of projection matrices

2) Applications: Putting it to work
   A) fitting a straight line to noisy data
   B) fitting polynomials to data
   C) Using Matlab for least-squares problems
   D) General linear least Squares $f(\underline{x})=\sum c_i \Phi_i(\underline{x})$

3) Caveats and Cautions

Orthogonal Projection onto a Subspace



Properties of projection matrices

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

1) Projection matrices are always symmetric

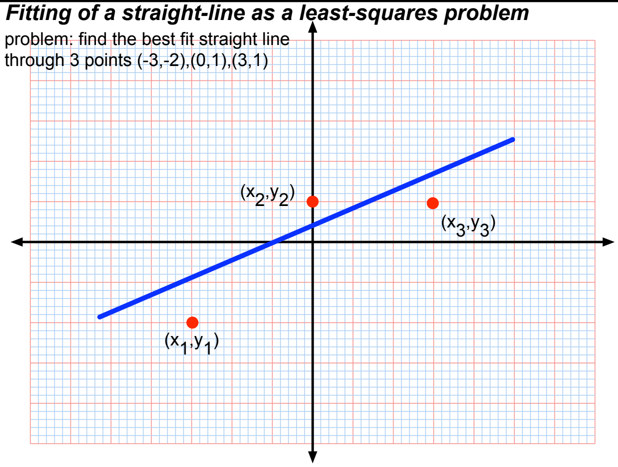
2) Projection matrices are usually singular (if $N(A^T)\neq Z$)

3) if A is invertible P=_____. Why?

4) if $\underline{p}=P\underline{b}$, then $P\underline{p}$=_____? therefore ___________

Fitting of a straight-line as a least-squares problem

problem: find the best fit straight line through 3 points (-3,-2),(0,1),(3,1)

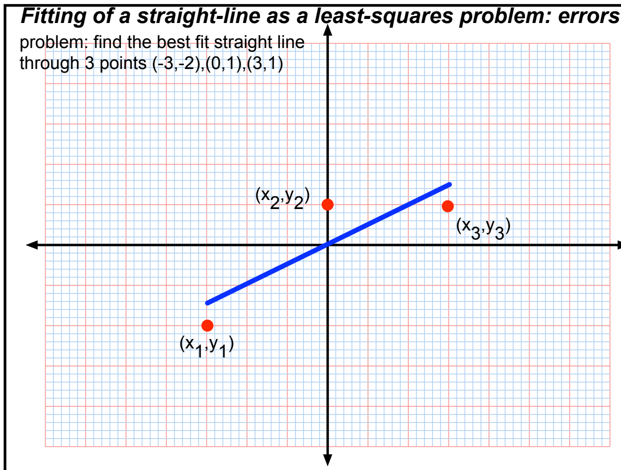


Fitting of a straight-line as a least-squares problem

problem: find the best fit straight line through 3 points (-3,-2),(0,1),(3,1)

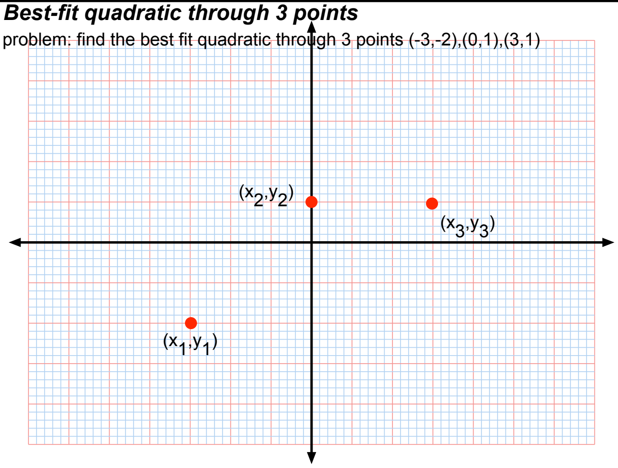
Fitting of a straight-line as a least-squares problem: errors

problem: find the best fit straight line
through 3 points (-3,-2),(0,1),(3,1)



Best-fit quadratic through 3 points

problem: find the best fit quadratic through 3 points (-3,-2),(0,1),(3,1)

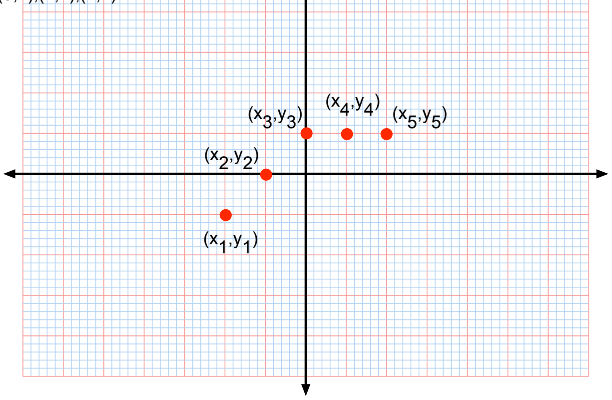


Best-fit quadratic through 3 points

problem: find the best fit quadratic through 3 points (-3,-2),(0,1),(3,1)

Best-fit quadratic through 5 points

problem: find the best fit quadratic through 3 points (-2,-1),(-1,0),(0,1),(1,1),(2,1)



Matlab Time:

Problem: fit a n-degree polynomial through m points

given y=[y_1 y_2 y_3 ... y_m]' at points x = [x_1 x_2 x_3 ... x_m]' find best fit coefficients to $f(x)=c_1x^n + c_2x^{n-1}+c_3x^{n-2}$ ...c_{n+1}

Method 1): Do it yourself (example y=[-1 0 1 1 1]', x=[-2 -1 0 1 2]' n=2)

A) form the Vandermonde matrix:

B) solve the normal equations

C) Better yet, use backslash

Matlab Time:

Problem: fit a n-degree polynomial through m points

given y=[y_1 y_2 y_3 ... y_m]' at points x = [x_1 x_2 x_3 ... x_m]' find best fit coefficients to $f(x)=c_1x^n + c_2x^{n-1}+c_3x^{n-2}$ ...$c_{n+1}f$

Method 2): use polyfit and polyval

(example y=[-1 0 1 1 1]', x=[-2 -1 0 1 2]' n=2)

A) find the best fit coefficients using polyfit (see help polyfit)

B) evaluate f(x) using polyval

C) plot and enjoy

General linear-least squares (linear regression)

***Problem:** find best-fit coefficients for*

$$f(\underline{x})= f(\underline{x})=\sum c_i\Phi_i(\underline{x})=c_1\Phi_1(\underline{x}) + c_2\Phi_2(\underline{x}) + \ldots\ c_n\Phi_n(\underline{x})$$

through m points

Examples:

1) best fit truncated cos series --

$f(x) = c_1 \cos(\pi x) + c_2 \cos(2\pi x) + c_3 \sin(3\pi x)$

2) best fit parabolic surface

$f(x,y) = c_1x^2 + c_2xy + c_3y^2$

General linear-least squares (linear regression)

Problem: ***find best-fit coefficients for***

$f(\underline{x}) = f(\underline{x}) = \sum c_i \Phi_i(\underline{x}) = c_1\Phi_1(\underline{x}) + c_2\Phi_2(\underline{x}) + \ldots c_n\Phi_n(\underline{x})$

through m points

Method:

1) ***form generalized Vandermonde matrix***

2) ***solve for coeffients***

3) ***plot to see if it makes any sense***

Caveats and Cautions:

1) ***The best-fit model isn't necessarily the best model***

   A) ***Choosing good models to fit is Science...***
   B) ***not all models are linear***
   C) ***beware over-fitting (too many parameters for the data)***

2) ***Uncertainty in the data propagates to uncertainty in the parameters:***

   You're not really done until you understand the quality and uncertainty in the fit: That's statistics