LEAST SQUARES and NORMAL EQUATIONS

Background

- Overdetermined Linear systems: consider $A\mathbf{x} = \mathbf{b}$
  if A is $m \times n$, $\mathbf{x}$ is $n \times 1$, $\mathbf{b}$ is $m \times 1$ with $m > n$.
  The linear system is **inconsistent** if
  no $\mathbf{x}$ satisfies all equations.
  Note: too many equations, not enough unknowns.
  Example:

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

- The **least squares solution**: the $\mathbf{x}$ that minimizes
  $||\mathbf{r}||_2 = ||A\mathbf{x} - \mathbf{b}||_2 = \left(\sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}x_j - b_i \right)^2 \right)^{1/2}$

LS and NORMAL EQUATIONS

- Geometric least squares solution $\bar{\mathbf{x}}$:
  $A\bar{\mathbf{x}} - \mathbf{b}$ should be *orthogonal* to all $A\mathbf{x}$.

- Algebraic least squares solution: consider
  $||A\mathbf{x} - \mathbf{b}||_2^2 = ||A(\bar{\mathbf{x}} + \mathbf{e}) - \mathbf{b}||_2^2$.

LEAST SQUARES, NORMAL EQUATIONS

The Normal Equations

- The **normal equations** are $A^T A\mathbf{x} = A^T\mathbf{b}$.
- If $rank(A) = n$ the normal equations have a unique solution $\bar{\mathbf{x}}$.
- Example

- SE and $RMSE$: with $\mathbf{r} = A\bar{\mathbf{x}} - \mathbf{b}$
  squared error
  $SE = ||\mathbf{r}||_2^2 = r_1^2 + r_2^2 + \cdots + r_m^2$;
  root mean squared error
  $RMSE = \sqrt{\sum_{i=1}^{m} r_i^2/m} = \sqrt{SE/m}$.

LEAST SQUARES CONTINUED

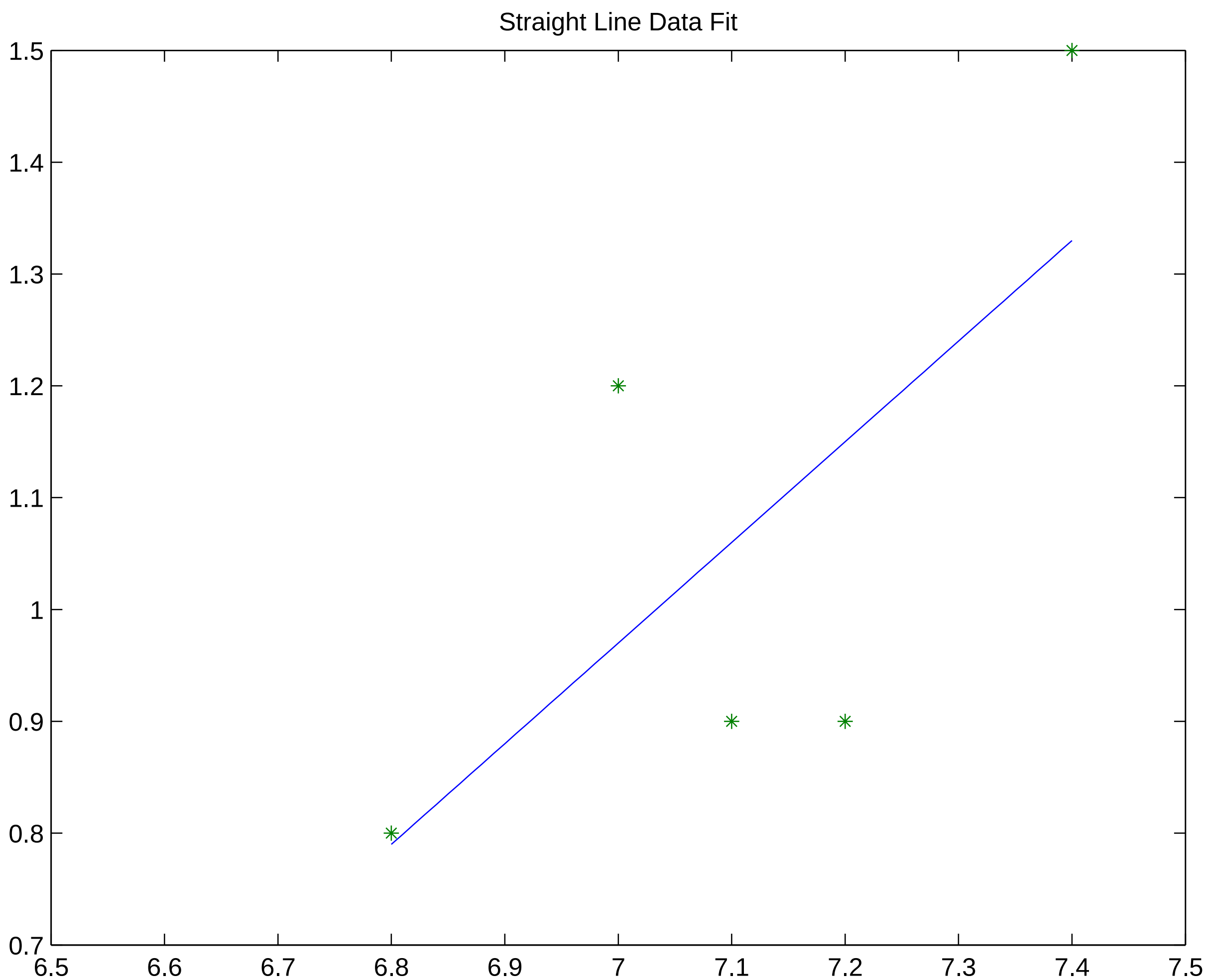
Data Fitting and Linear Models

- Fitting data to straight line: given data $\{(t_i, y_i)\}_{i=1}^{m}$, find the line $y(t) = a+bt$ "closest" to the data points.

"Least Squares" line minimizes sum of squared errors.

LEAST SQUARES CONTINUED

Example: $t = [6.8\ 7\ 7.1\ 7.2\ 7.4]$, $y = [.8\ 1.2\ .9\ .9\ 1.5]$



Matlab

```
t = [6.8 7 7.1 7.2 7.4]; y = [.8 1.2 .9 .9 1.5];
A = [ones(5,1) t']; p = (A'*A)\(A'*y');
tp = [6.8:.01:7.4];
plot(tp,p(1)+p(2)*tp,t,y,'*')
```

LEAST SQUARES CONTINUED

- General linear models have $y(t) = \sum_{j=1}^{n} c_j f_j(t)$,
  for some model functions $f_i(t)$ (e.g. $f_j(t) = t^{j-1}$):
  given data $\{(t_i, y_i)\}_{i=1}^{m}$, find best c_j's.

$$A = \begin{bmatrix} f_1(t_1) & f_2(t_1) & f_3(t_1) & \dots & f_n(t_1) \\ f_1(t_2) & f_2(t_2) & f_3(t_2) & \dots & f_n(t_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_1(t_m) & f_2(t_m) & f_3(t_m) & \dots & f_n(t_m) \end{bmatrix}$$

  Normal equations solution minimizes SE.
  If n is small and m is large with small SE, you have **data compression**.
  Examples:

LEAST SQUARES CONTINUED

Conditioning for Normal Equations

- Vandermonde Example:
  find best $y(t) = \sum_{j=1} c_j t^{j-1}$ for data.
  Linear system is $V\mathbf{c} = \mathbf{y}$, with "Vandermonde" V

$$\begin{bmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \ldots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 & \ldots & t_m^{n-1} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix},$$

  Normal equations solution: solve $V^T V\mathbf{c} = V^T\mathbf{y}$.
  Example: $t_i = 1, 1.1, \ldots, 2$, with $n = 8$. Matlab

```
t = [10:20]'/10; m = 11; n = 8;
V = (t*ones(1,n)).^(ones(m,1)*[0:n-1]);
disp([cond(V) cond(V'*V)])
   2.2303e+08   4.8086e+16
```

  Fit to data from $y(t) = 1 + t + t^2 + \cdots + t^7$.

```
P = @(t)sum(t.^[0:n-1]);
for i = 1:m, y(i) = P(t(i)); end
c = (V'*V)\(V'*y'); disp(c')
.98829 1.058 .87824 1.1407 .90343 1.0394 ...
  .99116 1.0008
disp(norm(V*c-y')/sqrt(m)); % RMSE
  1.3149e-07
```

  True solution should have all $c_i = 1$.

- Theory shows $K_2(A^T A) = K_2(A)^2$:
  normal equations are often **illconditioned**.