Linear Least Squares

- Recall: general quadratic norm of vector $\{x\}$

$$Q = \{x\}^T [W] \{x\} = \sum_i \sum_j x_i W_{ij} x_j$$

• Derivative wrt x_k becomes

$$\frac{\partial Q}{\partial x_k} = \sum_j W_{kj} x_j + \sum_i x_i W_{ik}$$

$$= \underbrace{\left[\longleftarrow W_{kj} \longrightarrow \right] \begin{Bmatrix} \uparrow \\ x \\ \downarrow \end{Bmatrix}}_{\substack{\text{dot product of } x \\ \text{w/ row } k \text{ of } [W]}} + \underbrace{\{ \longleftarrow x \longrightarrow \} \begin{bmatrix} \uparrow \\ W_{ik} \\ \downarrow \end{bmatrix}}_{\substack{\text{dot product of} \\ x \text{ w/ column } k \text{ of } [W]}}$$

$$= \left[\longleftarrow W_{kj} \longrightarrow \right] \{ x \} + \left[\longleftarrow W^T_{kj} \longrightarrow \right] \begin{Bmatrix} \uparrow \\ x \\ \downarrow \end{Bmatrix}$$

• then for all x_k's

$$\nabla Q = [W]\{x\} + \{x\}^T [W]$$

$$= \left([W] + [W]^T \right) \{x\}$$

IF $[W]$ symmetric, $\nabla Q = 2 [W] \{x\}$

- Scalar product $\{X\}$ with any vector $\{V\}$

$$S = \{X\}^T\{V\} = \sum_i X_i V_i = \{V\}^T\{X\}$$

$$\frac{\partial S}{\partial X_k} = V_k \Rightarrow \nabla S = \{V\}$$

- Assume we have

$$[A]\{X\} = \{b\} \quad \text{with } [A] \ m \times n, \ m > n$$

nonsymmetric

$\{b\}$ is known

$\{X\}$ is unknown

Since $m>n$, $[A]$ has more than n independent rows
System is over-determined, no solution exists which satisfies all equations

$$\therefore \ \{r\} \equiv [A]\{x\} - \{b\} \neq 0$$

seek the minimum residual sol'n.

$$\Omega = Var(r) = \{r\}^T\{r\}$$

$$= \left([A]\{x\} - \{b\}\right)^T\left([A]\{x\} - \{b\}\right)$$

$$= \{X\}^T[A]^T[A]\{X\} - \{X\}^T[A]^T\{b\}$$
$$- \{b\}^T[A]\{X\} + \{b\}^T\{b\}$$

$$= \{X\}^T[A]^T[A]\{X\} - 2\{X\}^T[A]^T\{b\}$$
$$+ \{b\}^T\{b\}$$

The first order conditions for minimizing Ω are the vanishing of all components of its gradient

$$\{\nabla\Omega\} = \left([A]^T[A] + \left[[A]^T[A]\right]^T \right)\{x\} - 2[A]^T\{b\}$$

$$= 2\left[[A]^T[A]\right]\{x\} - 2[A]^T\{b\}$$

since $[A]^T[A]$ is symmetric

We want $\nabla\Omega = 0 \Rightarrow \underbrace{[A]^T[A]\{x\} = [A]^T\{b\}}$

"Normal Equations" for ordinary least squares defines solution with minimum residual variance

$$\{x\} = \left[[A]^T[A]\right]^{-1}\{b\}$$

essentially have premultiplied non-square system $[A]\{x\} = \{b\}$ with $[A]^T$ to produce $n \times n$ system... key is what is the conditioning of $[A]^T[A]$ and does an inverse exist.

- often normal equations are poorly conditioned and produce noisy results... one strategy, solve with SVD, to control the rank to filter out modes of sol'n with small singular values.

- Weighted Least Squares (WLS)

- Suppose "all residuals not equal"... can insert a weighting matrix into Ω

$$\Omega = \{r\}^T [W] \{r\}$$

e.g. diagonals of $[W]$ adjust for the expected size of each residual but could also weight cross-products of residuals or could make differences among residuals or other linear combinations

$$\{r'\} = [V]\{r\}$$

then $$Var(r') = \Omega' = \{r\}^T \underbrace{[V]^T [V]}_{[W]} \{r\}$$

So Normal Equations for WLS become

$$\Omega' = \left([A]\{x\} - \{b\}\right)^T [W] \left([A]\{x\} - \{b\}\right)$$

$$= \{x\}^T [A]^T [W] [A] \{x\} - 2\{x\}^T [A]^T [W] \{b\} + \{b\}^T [W] \{b\}$$

$$\{\nabla \Omega'\} = 2\left[[A]^T [W] [A]\right]\{x\} - 2\left[[A]^T [W]\right]\{b\}$$

so $$\left[[A]^T [W] [A]\right]\{x\} = \left[[A]^T [W]\right]\{b\}$$

Generalized Least Squares (GLS)

- most general case, weight both residual and sol'n

$$\Omega'' = \{r\}^T [W_r] \{r\} + \{x\}^T [W_x] \{x\}$$

$$\{\nabla \Omega''\} = \{\nabla \Omega'\} + 2 [W_x] \{x\}$$

Setting $\nabla \Omega'' = 0$ gives GLS Normal Equations

$$\Big[[A]^T [W_r] [A] + [W_x] \Big] \{x\} = \Big[[A]^T [W_r] \Big] \{b\}$$

or

$$\{x\} = \Big[[A]^T [W_r] [A] + [W_x] \Big]^{-1} \Big[[A]^T [W_r] \Big] \{b\}$$

- $[W_x]$ can add desirable conditioning, e.g. $[W_x] = \frac{1}{\sigma^2} I$ would prefer answers which are not big compared to σ

Then $[W_x]$ just adds to diagonal of GLS system "regularization" effect to make $[A]^T[A]$ more invertable (i.e. better conditioned) to avoid big and noisy $\{x\}$ solutions

- Preferred choice for weight matrices
  - Inverse of the covariance of vector being estimated

$$\text{i.e.} \quad [W_x] = [\mathrm{Cov}(x)]^{-1}$$
$$[W_r] = [\mathrm{Cov}(r)]^{-1}$$

- In under-determined case, $m<n$, GLS provides a unique answer

  $[A]^T[A]$ has no inverse... there are multiple solutions which produce $\{r\}=0$. OLS/WLS provides no choice among possibilities.

  But GLS adds a second term to Ω: $\{x\}^T[W]_x\{x\}$ which favors certain size/shape of $\{x\}$... tries to achieve a balance between small $\{r\}$ and "credible" $\{x\}$ depending on details of $[W_r]$, $[W_x]$.

Overall Strategy: formulate GLS normal equations, concentrating on weight matrices $[W_r]$, $[W_x]$ in the design of the inversion, then use SVD to solve resulting square system using condition number control to avoid near singularities and the resulting noise amplification in the normal equations.