$$X^T X \to \Sigma = \begin{pmatrix} 1+\lambda & & & & \\ & 1+\lambda & & & \\ & & 1+\lambda & & \\ & & & 1+\lambda & \\ & & & & 0.001+\lambda \end{pmatrix} \qquad \Sigma^{-1} = \begin{pmatrix} a & & & & 0 \\ & b & & & \\ & & c & & \\ & & & d & \\ 0 & & & & 1000 \end{pmatrix} \begin{pmatrix} \frac{1}{1+\lambda} & & & \\ & \ddots & & \\ & & & \frac{1}{0.001+\lambda} \end{pmatrix}$$

12

5×5

$$\Sigma \Sigma^{-1} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ 0 & & & & 0.01 \end{pmatrix} \begin{pmatrix} a & & & & 0 \\ & b & & & \\ & & c & & \\ & & & d & \\ 0 & & & & e \end{pmatrix} = \begin{pmatrix} a & & & & \\ & b & & & \\ & & c & & \\ & & & d & \\ & & & & 0.01e \end{pmatrix} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots \end{pmatrix}$$

Near singular matrix leads an instability!

$$w = (X^T X)^{-1} X^T y \Rightarrow (X^T X + \lambda I)^{-1} X^T y = w$$

Tikhonov regularization

$$(X^T X + \lambda I) w = X^T y$$

$$\chi^2 = \sum_i \left(y_i - \sum w_i x_{ij}\right)^2 \Longrightarrow$$

$$X^T X w - \lambda \vec{w} - X^T y = \emptyset$$

after λ $\chi^2 = \sum_i \left(y_i - \sum x_{ij} w_i\right)^2 + \lambda \sum_j w_j^2$

$2\lambda \vec{w}_j$

$$\chi^2 = \sum\left(y_i - \sum_j x_{ij} w_j\right)^2 + \lambda \sum w_j^2$$

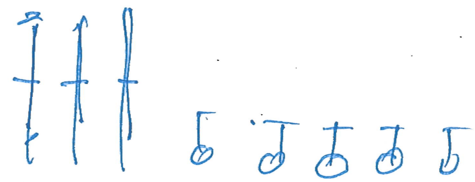
1 1 1 1 1000 $10^6 \lambda$

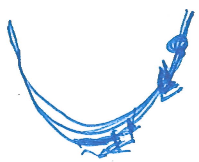
Compromise solution $\mathcal{O}(1/\sqrt{\lambda})$ RIDGE REGULARIZATI[ON]

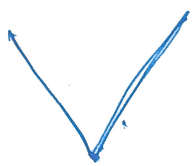
L1 regularization.

$\chi^2 = \|Y - Xw\|_2^2 \Rightarrow -\lambda \|w\|_2^2$

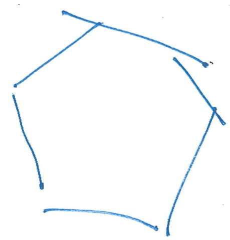
$-\lambda \|w\|_1$



$\|w\|_2^2$  gradient descent

$\|w\|_1$ 

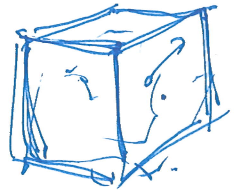
concave



convex region

$\lambda \|w\|_1$ $\qquad \|w\|_1 < t$ $\qquad t \sim 1/\lambda$

$|w_i| < t$



convex

L1 cheat

LASSO Least Absolute Selection and Shrinkage Operator

$w_1 \quad w_2 \quad w_3$ $\qquad \|\vec{w}\|_1 < t \;=\; |w_1| + |w_2| + |w_3| < t$

$$
\begin{array}{l}
|w_1| + |w_2 + w_3 \le t \\
-w_1 + w_2 + w_3 \le t \\
\quad \vdots \\
-w_1 - w_2 - w_3 \le t
\end{array}
$$

$8 = 2 \cdot 2 \cdot 2$

2^{100}

Shrinkage!

Variant: mix LASSO + RIDGE

$$\hat{w} = \underset{w}{\operatorname{argmin}} \left[\|Y - Xw\|_2^2 + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2 \right]$$

Elastic Net

NON-LINEAR LEAST SQUARES

$$r_i = y_i - f(x_i, \vec{\beta}_j) \qquad i = 1, 2, \dots m \qquad j = 1, \dots, n \qquad m \geq n$$

$$S = \sum_i r_i^2$$

"JACOBIAN"

Minimum of S: $\dfrac{\partial S}{\partial \beta_j} = 2 \sum_i r_i \dfrac{\partial r_i}{\partial \beta_j} = \emptyset$ $\qquad \dfrac{\partial r_i}{\partial \beta_j} = - J_{ij}$

$$f(x_i, \beta) = f(x_i, \beta^k) + \sum \frac{\partial f(x_i, \beta^k)}{\partial \beta_j} (\beta_j - \beta_j^k) = f(x_i, \beta^k) + \sum_j J_{ij} \Delta \beta_j$$

$$\Delta y_i = y_i - f(x_i, \beta^k)$$

$$r_i = y_i - f(x_i, \beta) = \left(y_i - f(x_i, \beta^k)\right) + \left(f(x_i, \beta^k) - f(x_i, \beta)\right) =$$

$$r_i = \Delta y_i - \sum_j J_{ij} \Delta \beta_j ,$$

$$\beta_j = \beta_j^k + \Delta \beta_j$$

Substituting into the $\frac{\partial S}{\partial \beta_j} = \emptyset$

$$-2\sum_{i=1}^{m} J_{ij}\left(\Delta y_i - \sum_{s=1}^{n} J_{is}\,\Delta\beta_s\right) = \emptyset$$

$$\sum_i \sum_s J_{ij} J_{is}\,\Delta\beta_s = \sum_i J_{ij}\,\Delta y_i$$

$$(J^T \cdot J)\,\Delta\vec{\beta} = J^T \cdot \Delta\vec{y}$$

$$\boxed{\Delta\vec{\beta} = (J^T J)^{-1} J^T \Delta\vec{y}}$$

Similar to linear models, just incrementally.

Using SVD: $J = U\Sigma V^T$

$$J^T J = V\Sigma^T \underbrace{U^T U}_{1} \Sigma V^T = V\Sigma^T\Sigma V^T$$

$$(J^T J)^{-1} = V\Sigma^{-1}(\Sigma^T)^{-1} V^T$$

$$\Delta\beta = (V\Sigma^{-1}U^T)\cdot\Delta\vec{y}$$

$$(J^T J)^{-1} J^T = V\Sigma^{-1}\overbrace{(\Sigma^T)^{-1}\underbrace{V^T \cdot V}_{1}\Sigma^T}^{1} U^T$$

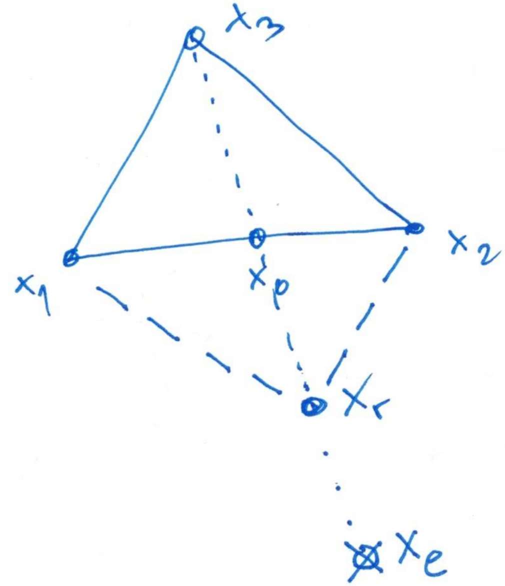
$$= V\Sigma^{-1}U^T$$

NL3

<u>Direct search</u>

Nelder-Mead (simplex / amoeba)

Illustrate in 2D $n=2$.



$n+1$ points form an $n+1$ simplex.

$$f(x_1) \le f(x_2) \le f(x_3)$$

$$x_0 = \text{CM of the smallest } n.$$
$$= \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Reflection

$$\vec{x}_r = \vec{x}_0 + \alpha(\vec{x}_0 - \vec{x}_{n+1}) \qquad \alpha \approx 1$$

If $\quad f(x_1) \le f(x_r) \le f(x_n) \Rightarrow x_{n+1} \leftarrow x_r$

Expansion

If $\quad f(x_r) < f(x_1)$, reflected is the best

$\vec{x}_e = x_0 + \gamma(x_r - x_0)$, expand $\quad \gamma \approx 2$

If $f(x_e) < f(x_r)$. $\quad x_{n+1} \leftarrow x_e$

else $\quad x_{n+1} \leftarrow x_r$

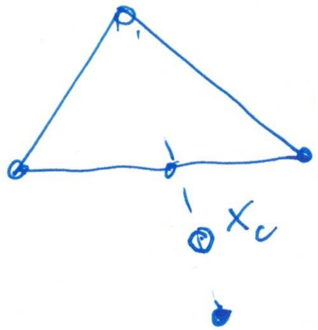
Contraction

If $\quad f(x_r) > f(x_n)$

If $f(x_r) \le f(x_{n+1})$

$$x_c = x_0 + \rho(x_r - x_0) \qquad 0 < \rho \approx \frac{1}{2}$$

NL4

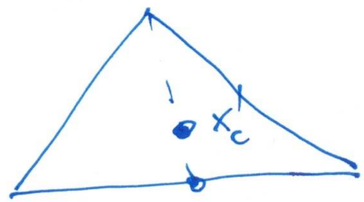


If $f(x_c) < f(x_r)$ $\quad x_{n+1} \leftarrow x_c$

Else $\rightarrow$ Shrink

$\leftarrow$ If $f(x_r) \geq f(x_{n+1})$

Compute contracted on the inside



$x_c' = x_0 + \rho(x_{n+1} - x_0) \quad 0 \leq \rho \leq 1/2$

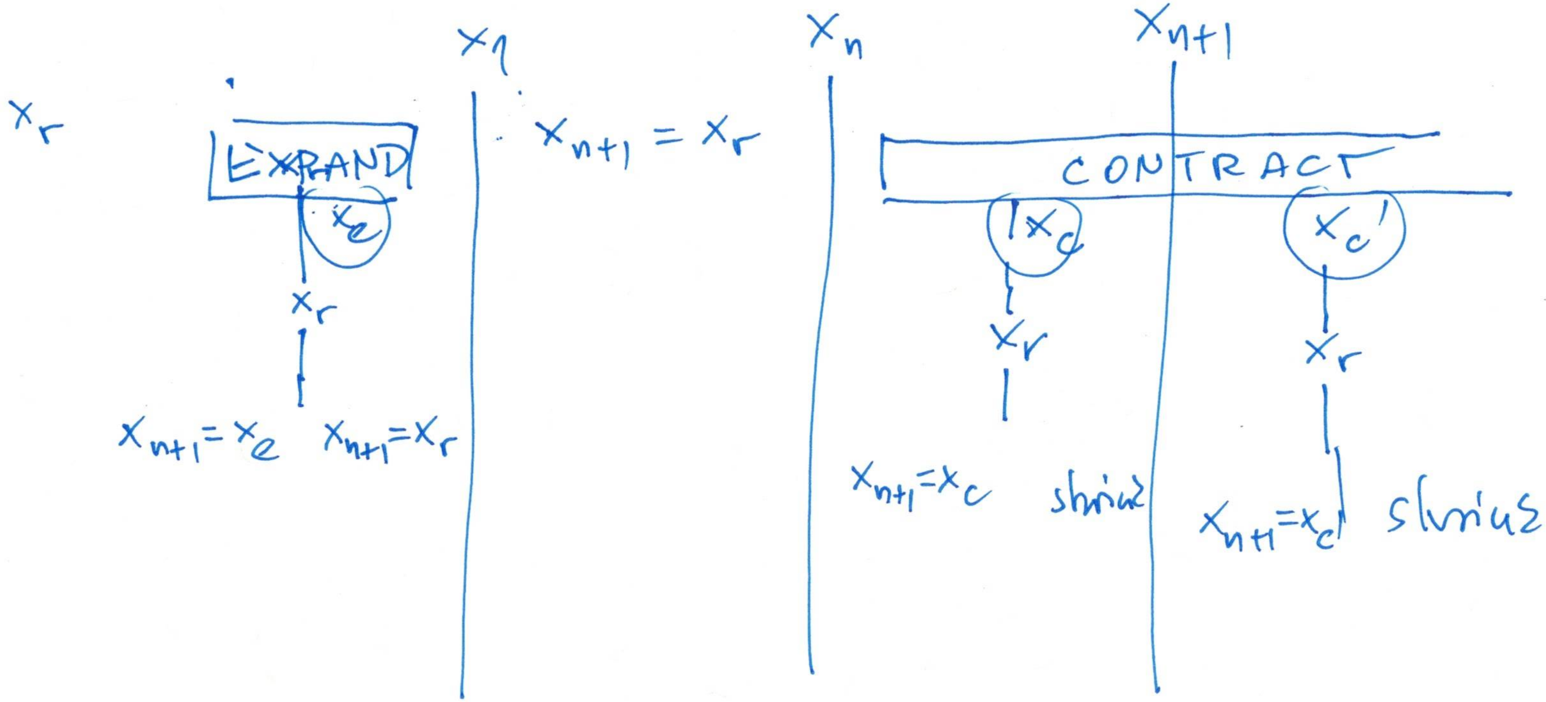
If $f(x_c') < f(x_{n+1})$ $\quad x_{n+1} \leftarrow x_c'$

Else Shrink

Shrink: Replace all points ~~with~~ except $\vec{x}_1$

$x_i = x_1 + \sigma(x_i - x_1) \quad \sigma \sim 1/2$

NLS


x_r
EXPAND
x_e
x_r
$x_{n+1} = x_e$
$x_{n+1} = x_r$
x_1
$x_{n+1} = x_r$
x_n
x_{n+1}
CONTRACT
x_c
x_c'
x_r
x_r
$x_{n+1} = x_c$ shrink
$x_{n+1} = x_c'$ shrink