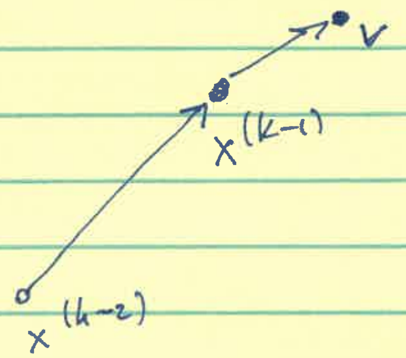
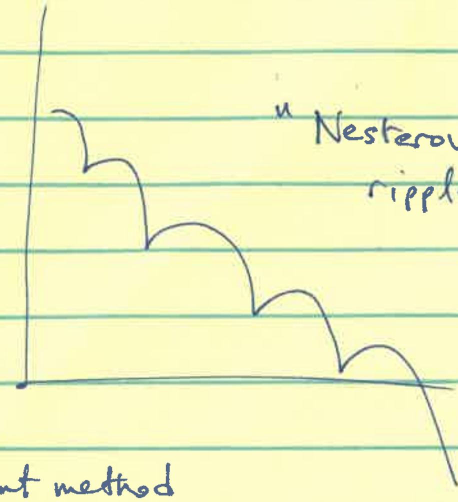
$$\operatorname{prox}_t(x) = \underset{z}{\operatorname{argmin}} \frac{1}{2t} \|x - z\|_2^2 + h(z)$$





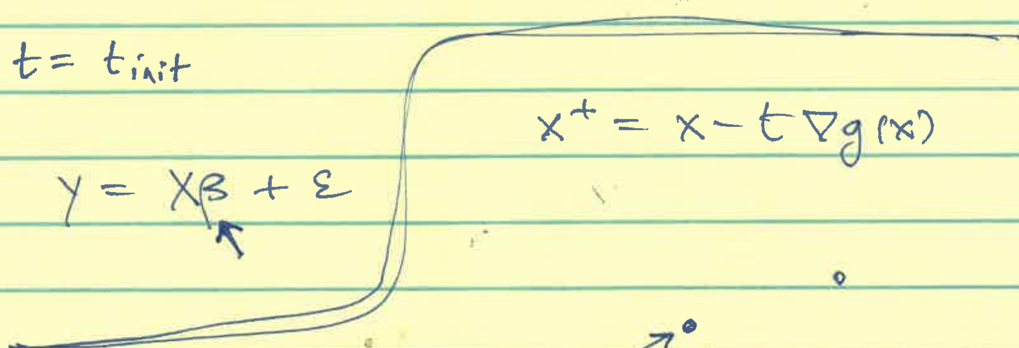
"Nesterov ripples"

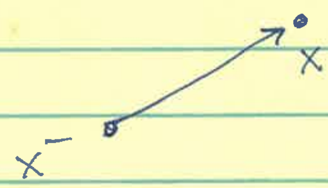
Not a descent method

$t = t_{init}$

$x^+ = x - t \nabla g(x)$

$Y = X\beta + \varepsilon$





$$\underset{B}{\min} \ \frac{1}{2} \sum_{(i,j) \in \Omega} (Y_{ij} - B_{ij})^2 + \lambda \|B\|_*$$

$\|B\|_{tr}$.

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\frac{1}{n}\|y - X\beta\|_2^2 = \frac{1}{n}\sum_{i=1}^{n} \underbrace{(y_i - x_i^T\beta)^2}_{f_i(\beta)}$$

X $n \times p$.

x_i rows of x

$i = 1, \dots, n$

$$\mathbb{E}[\nabla f_{i_k}(x)] = \frac{1}{m}\sum_{i=1}^{m} \nabla f_i(x)$$

$$p_i(\beta) = \frac{1}{1 + \exp(-x_i^T\beta)}$$ logistic.

$$\begin{aligned} x^{(k+2)} &= x^{(k+1)} - t\nabla f_2(x^{(k+1)}) \\ &= x^{(k)} - t\left[\nabla f_1(x^{(k)}) + \nabla f_2(x^{(k+1)})\right] \end{aligned}$$

$$F(x, \xi) = f(x) + \xi^2$$

$$f(x - \xi)$$

SGE

Experience (Duchi)