Batch method

fixed step. Recall

$$\lim_{k\to\infty} f(x^{(k)}_{best}) \leq f^* + \frac{G^2_{batch}}{2} t$$

Stochatistic method: each f_i is G-Lipschitz

$$\Rightarrow f = \sum_{i=1}^{m} f_i \text{ is } mG\text{-Lipschitz}$$

$$\text{''} \qquad \leq \text{''} + \frac{5(mG)^2 t}{2}$$

Batch rate $O\left(\frac{G^2_{batch}}{\varepsilon^2}\right)$

Cyclic rate $O\left(m \cdot \frac{(mG)^2}{\varepsilon^2}\right)$ iteration complexity

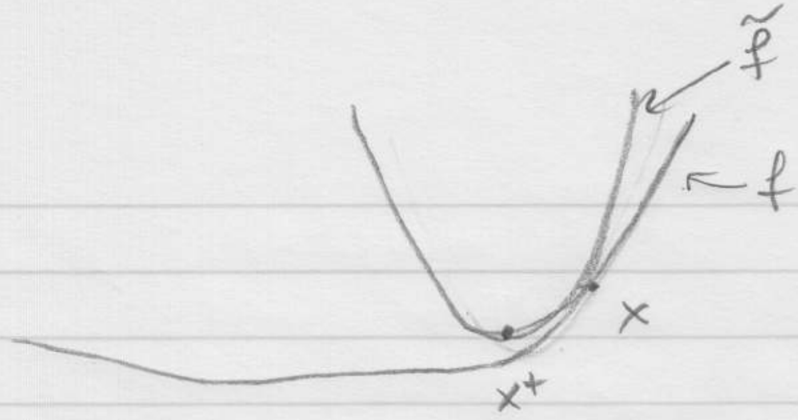
$O\left(\frac{(mG)^2}{\varepsilon^2}\right)$ cycle complexity

Randomized rule $O\left(\frac{(mG)^2}{\varepsilon^2}\right)$ iteration complexity

$O\left(\frac{mG^2}{\varepsilon^2}\right)$ cykle complexity

reduced by factor of m

Lower bound is $O(1/\varepsilon^2)$ for any n.s. first-order method



$$\frac{1}{2t}\|z-(x-t\nabla g(x))\|_2^2$$

$$=\frac{1}{2t}\|z-x+t\nabla g(x)\|_2^2$$

$$=\frac{1}{2t}\|z-x\|_2^2+(z-x)^T\nabla g(x)+(x \text{ terms})$$

$$\operatorname{prox}_{h,t}(x)=\operatorname*{argmin}_z \frac{1}{2t}\|x-z\|_2^2+h(z)$$

$$x^+=\operatorname*{argmin}_z \frac{1}{2t}\|z-(x-t\nabla g(x))\|_2^2+h(z)$$

$$=\operatorname{prox}_{h,t}(x-t\nabla g(x))$$

$$f = g + h$$

loss regularizer/penalty

$$\text{prox}_t(\beta) = \underset{z}{\text{argmin}}\ \frac{1}{2}\|z-\beta\|_2^2 + \lambda t \|z\|_1$$

$$= \text{soft-thresholding } \beta \text{ by amt } \lambda t$$

$$B = U \Sigma V^T$$ singular-value decomposition

$$\text{prox}_\lambda(B) = U \cdot \text{diag}((\Sigma_{11}-\lambda)_+, \ldots (\Sigma_{rr}-\lambda)_+) V^T$$

$$\min_{Z}\ \frac{1}{2}\|B-Z\|_F^2 + \lambda t \|Z\|_{tr}$$

$$0 = Z - B + \lambda t \cdot \Gamma, \quad \Gamma \in \partial \|Z\|_{tr}$$ homework

$$\partial \|Z\|_{tr} = \{ UV^T + W : \|W\|_{op} \leq 1, U^T W = 0, WV = 0 \}$$

$$Z = U \Sigma V^T$$

---

plug-in $Z = U \Sigma_{\lambda t} V^T \quad (B = U \Sigma V^T)$

$$U(\Sigma_{\lambda t} - \Sigma)V^T + \lambda t \Gamma = 0$$

$$\Gamma = \frac{1}{\lambda t} U(\Sigma - \Sigma_{\lambda t})V^T$$

$$= UV^T + \underbrace{U\left(\boxed{\frac{\Sigma - \Sigma_{\lambda t}}{\lambda t} - I}\right)V^T}_{W}$$

\# of diag elements $> \lambda t$

$\|W\|_{op} = ?$

< 1 ✓

$$\Sigma - \Sigma_{\lambda t} = (\underbrace{\lambda t, \ldots \lambda t}_{k}, \underbrace{\Sigma_{(k+1),k+1} \ldots}_{})$$

<1

$$\frac{\Sigma - \Sigma_{\lambda t}}{\lambda t} = (\underbrace{1 \ldots 1}_{k}, a \ldots)$$