$$C = \{x : \|x\| \le t\}$$

often FW is simpler / more efficient than

- Proj onto C
- prox operator of $\|\cdot\|$

$$f(x) = \frac{1}{n} \sum f_i(x)$$

$$E[\nabla f_{i_k}(x)] = \nabla f(x)$$

$$E\left[\frac{1}{b} \sum_{i \in I_k} \nabla f_{i}(x)\right] = \nabla f(x)$$

Expectation over random choices of indices

SG : after k iterations $\quad O(1/\sqrt{k})$

mini-batch SG : after k iters $\quad O(1/\sqrt{bk})$
(b small)

BUT. each iter of mini-batch SG is b times more expensive

so compare k SG iters with k/b mini-batch iters

$$\underline{O(1/\sqrt{k})} \quad \text{versus} \quad O\left(1/\sqrt{b \cdot \tfrac{k}{b}}\right) = \underline{O\left(\tfrac{1}{\sqrt{k}}\right)}$$

SAG :
initialize $\quad g_i^{(0)} = \nabla f_i(x^{(0)}), \quad i = 1 \ldots n$

OR $\quad g_i^{(0)} = 0$

$$\underbrace{g_{ik}^{(k)} - g_{ik}^{(k-1)} + \underbrace{\sum g_i^{(k-1)}}_{\text{old sum}}}_{\substack{\sum g_i^{(k)} \\ \text{new sum}}}$$