# Exercises 4: Better stochastic methods

**Exercise 1** (SVRG)**.**

*The goal of this exercise is to show the benefit of variance reduction techniques in stochastic optimization. We consider empirical risk minimization*

$$\min_{\theta \in \mathbb{R}^d} F(\theta) = \frac{1}{n} \sum_{i=1}^{n} f_i(\theta).$$

*through the use of Stochastic Variance Reduced Gradient (SVRG).*

---

**Algorithm 1:** SVRG

---

Input : $\tilde{\theta}$
**for** e $= 1, \ldots, \#$ of epochs **do**
  Compute all the gradients $\nabla f_i(\tilde{\theta})$ and store $\nabla F(\tilde{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(\tilde{\theta})$
  Initialize $\theta_0 = \tilde{\theta}$
  **for** $t = 1, \ldots,$ length of epoch **do**
    $\theta_t = \theta_{t-1} - \gamma \left( \nabla F(\tilde{\theta}) + \nabla f_{i(t)}(\theta_{t-1}) - \nabla f_{i(t)}(\tilde{\theta}) \right)$
  **end for**
  Update $\tilde{\theta} = \theta_t$
**end for**
Output $\tilde{\theta}$

---

*Assume that each $f_i$ is $L$-smooth, and that $F$ is $\mu$-strongly convex, and call $\theta^\star$ the minimizer of $F$.*

1. *Show that for $i(t) \sim \mathcal{U}(\{1, \ldots, n\})$, and for all $\theta$,*

$$\mathbb{E} \left\| \nabla f_{i(t)}(\theta) - \nabla f_{i(t)}(\theta^\star) \right\|^2 \leq 2L(F(\theta) - F^\star(\theta^\star)).$$

2. *Fix an epoch $e$, the goal of what follows is to understand the benefit at the end of one epoch.*

   (a) *Set, at iteration $t$ of the epoch $e$, $g_t = \nabla F(\tilde{\theta}) + \nabla f_{i(t)}(\theta_{t-1}) - \nabla f_{i(t)}(\tilde{\theta})$. Check that*

$$\|\theta_{t+1} - \theta^\star\|^2 = \|\theta_t - \theta^\star\|^2 - 2\gamma\langle \theta_{t-1} - \theta^\star, g_t \rangle + \gamma^2 \|g_t\|^2.$$

   (b) *Show that*

$$\mathbb{E}\left[\|\theta_{t+1} - \theta^\star\|^2\right] \leq \mathbb{E}\left[\|\theta_t - \theta^\star\|^2\right] - 2\gamma(1-2\gamma L)\mathbb{E}\left[F(\theta_t) - F(\theta^\star)\right] + 4\gamma^2 L \mathbb{E}\left[F(\tilde{\theta}) - F(\theta^\star)\right]$$

   *and therefore*

$$\mathbb{E}\left[\|\theta_t - \theta^\star\|^2\right] \leq \|\theta_0 - \theta^\star\|^2 - 2\gamma(1-2\gamma L)\sum_{s=1}^{t} \mathbb{E}\left[F(\theta_{s-1}) - F(\theta^\star)\right] + 4\gamma^2 L t \mathbb{E}\left[F(\tilde{\theta}) - F(\theta^\star)\right].$$

(c) *Deduce that*

$$\mathbb{E}\left[F\left(\frac{1}{t}\sum_{s=1}^{t}\theta_t\right)-F^\star\right]\leq\left(\frac{1}{t\gamma(1-2\gamma L)\mu}+\frac{2\gamma L}{1-2\gamma L}\right)\left(F(\theta_0)-F^\star\right).$$

(d) *Choose* $t=20L/\mu$ *steps in epoch* $e$ *with* $\gamma=1/10L$ *and show that in a fixed epoch* $e$

$$\mathbb{E}\left[F\left(\frac{1}{t}\sum_{s=1}^{t}\theta_t\right)-F^\star\right]\leq 0.9\left(F(\theta_0)-F^\star\right),$$

*with* $\theta_0$ *the initial point in epoch* $e$*. Overall after* $n_e$ *epochs,*

$$\mathbb{E}\left[F\left(\frac{1}{t}\sum_{s=1}^{t}\theta_t\right)-F^\star\right]\leq 0.9^{n_e}\left(F(\theta_0)-F^\star\right),$$

*with* $\theta_0$ *the initial point of SVRG.*