Exercise List: Variance Reduced Gradient Methods

Francis Bach and Robert M. Gower

October 12, 2019

1 The SAG algorithm

Ex. 1 — The SAG Algorithm. Consider the optimization problem

$$\min_{x\in\mathbb{R}^d} \frac{1}{n}\sum_{j=1}^{n} f_j(x), \tag{1}$$

and the following implementation of the SAG algorithm given in Algorithm 1.

Algorithm 1 SAG: Stochastic Average Gradient descent

Initialize $x^0, g_i = 0 \in \mathbb{R}^d$ for $i = 1, \ldots, n$. Choose $\eta > 0$ the stepsize.
for $k = 1, \ldots, T-1$ **do**
  Sample $i_k \in \{1, \ldots, n\}$
  $g_{i_k} = \nabla f_{i_k}(x^k)$
  $G^k = \frac{1}{n}\sum_{j=1}^{n} g_j$
  $x^{k+1} = x^k - \eta G^k$
Output: x^T

Part I

Assume that calculating $\nabla f_{i_k}(x^k)$ costs $O(d)$ operations and that sampling i_k costs $O(1)$. What is the computational cost of a single iteration of Algorithm 1?

Part II

Re-write this implementation of SAG in such a way that the computational cost of a single iteration is $O(d)$.