1 Exercise

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function and $\gamma, c > 0$. We consider the following algorithm:

$$X_{k+1} = X_k - \gamma \nabla f(X_k) + c\eta_{k+1}\,,$$

with $(\eta_k)_{k\geqslant 1}$ being i.i.d. $\mathbb{R}^d$-valued random variables such that $\mathbb{E}[\eta_1] = 0$ and $\mathbb{E}[\|\eta_1\|^2] < +\infty$. We assume moreover, that η has a density $g : \mathbb{R}^d \to \mathbb{R}$.

Question 1.1. Write down P the Markov kernel of (X_k)

In the following, we will assume that $\nabla f : \mathbb{R}^d \to \mathbb{R}^d$ is L-Lipschitz continuous and that this implies

$$\forall x, y \in \mathbb{R}^d \quad f(y) \leqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|^2\,.$$

Question 1.2. Show that this implies the existence of $K > 0$ such that for all $x \in \mathbb{R}^d$,

$$Pf(x) \leqslant f(x) - \gamma\|\nabla f(x)\|^2 + \frac{L}{2}\gamma^2\|\nabla f(x)\|^2 + K\,.$$

Assume moreover, that there is $\alpha > 0$ such that $\|\nabla f(x)\|^2 \geqslant \frac{1}{2\alpha}(f(x) - f^*)$, where $f^* = \inf_{x\in\mathbb{R}^d} f^*$ (this happens for instance if f is strongly-convex).

Question 1.3. Show that there are values of (c, γ) such that there is $V : \mathbb{R}^d \to [1, +\infty)$, $\lambda \in (0, 1)$ and $b \geqslant 0$ such that the drift inequality from the lecture notes (geometric ergodicity) is true:

$$PV(x) \leqslant \lambda V(x) + b$$

Question 1.4. What is a simple condition on V (related to f) and the law of η_1 to ensure that there is a unique invariant distribution π to which the law of X_k converges exponentially fast.

Answer

Question 1.1. Let $h : \mathbb{R}^d \to \mathbb{R}$ a bounded function and $A \in \mathcal{B}(\mathbb{R}^d)$ a borelian. It holds

$$\begin{aligned}
\mathbb{E}[h(X_1)\mathbb{1}_A(X_0)] &= \mathbb{E}\left[\mathbb{1}_A(X_0)\int_{y\in\mathbb{R}^d} h(X_0 - \gamma\nabla f(X_0) + cy)g(y)\,\mathrm{d}y\right] \\
&= \mathbb{E}\left[\mathbb{1}_A(X_0)\frac{1}{c^d}\int_{z\in\mathbb{R}^d} h(z)g((z - X_0 + \gamma\nabla f(X_0))/c)\,\mathrm{d}z\right] \\
&= \mathbb{E}[\mathbb{1}_A(X_0)\int h(z)P(X_0, \mathrm{d}z)]\,,
\end{aligned}$$

where we used the change of variable $y \mapsto (z - X_0 + \gamma\nabla f(X_0))/c$ in the penultimate equality and where $P(x, \mathrm{d}z) = \frac{1}{c^d}g((z - x + \gamma\nabla f(x))/c)\,\mathrm{d}z$ is the Markov kernel.

Question 1.2. It holds that

$$\begin{aligned}
Pf(x) = \mathbb{E}_x[f(X_1)] &\leqslant \mathbb{E}_x[f(x) + \langle\nabla f(x), X_1 - x\rangle + \frac{L}{2}\|X_1 - x\|^2] \\
&\leqslant f(x) - \gamma\mathbb{E}_x[\|\nabla f(x)\|^2] + \gamma c\mathbb{E}_x[\langle f(x), \eta_1\rangle] + \frac{L}{2}\mathbb{E}_x[\|\gamma\nabla f(x) + c\eta_1\|^2] \\
&\leqslant f(x) - \gamma\|\nabla f(x)\|^2 + \frac{L}{2}\gamma^2\mathbb{E}_x[\|\nabla f(x)\|^2] + \frac{L}{2}c^2\mathbb{E}_x[\|\eta_1\|^2]\,,
\end{aligned}$$

where in the first inequality we have used the L-smoothness of ∇f, and in the last the fact that η_1 is of zero-mean. Since, $\mathbb{E}[\|\eta_1\|^2] < K$ for some $K < +\infty$ the proof is finished.

Question 1.3. Using the inequality shown in the previous question we have:

$$P(f(x) - f^*) = Pf(x) - f^* \leqslant f(x) - f^* - \gamma(1 - \frac{L}{2}\gamma)\,\|\nabla f(x)\|^2 + K\,.$$

If $\gamma < \frac{2}{L}$, then $(1 - \frac{L}{2}\gamma) > 0$ and using the new "strongly-convex" inequality we obtain:

$$P(f(x) - f^*) \leqslant f(x) - f^* - \gamma(1 - \frac{L}{2}\gamma)(f(x) - f^*) + K = \lambda(f(x) - f^*) + K\,,$$

with $\lambda = \frac{L}{2}\gamma$. Finally, defining $V = f(x) - f^* + 1$, we obtain

$$PV(x) = P(f(x) - f^*) + 1 \leqslant \lambda(f(x) - f^*) + K + 1 \leqslant \lambda V(x) + K + 1 - \lambda = \lambda V(x) + b,\,,$$

and so obtain we obtain such an inequality, for any $c \geqslant 0$, $\gamma \leqslant \frac{L}{2}\gamma$ and $V = f(x) - f^* + 1$.

Question 1.4. To have geometric ergodicity we also need a minorizing condition: for any $d \geqslant 1$, there is a measure ν_d and $\varepsilon_d > 0$ such that for any $x \in C_d := \{x \in \mathbb{R}^d : V(x) \leqslant d\}$ it holds that $P(x, \cdot) \geqslant c_d \nu_d$.

Such a condition will be automatically satisfied if g, the density of η, is continuous and C_d is compact (for instance $\lim_{\|x\| \to +\infty} f(x) = +\infty$). Indeed, in that case:

$$\forall x \in \mathbb{R}^d\,, A \in \mathcal{B}(\mathbb{R}^d) \quad P(x, A) = \frac{1}{c^d}\int_{z \in A} g((z - x + \gamma\nabla f(x))/c)\,\mathrm{d}z \geqslant \frac{1}{c^d}\int_{z \in A} \inf_{x \in C_d} g((z - x + \gamma\nabla f(x))/c)\,\mathrm{d}z\,.$$

Denoting $l(z) = \inf_{x \in C_d} g((z - x + \gamma\nabla f(x))/c)$ we have that $l(z) > 0$ by compactness of C_d. Therefore, we can define ν_d as

$$\nu_d(\mathrm{d}z) = \frac{l(z)\,\mathrm{d}z}{\int_{\mathbb{R}^d} l(z)\,\mathrm{d}z}$$

and we obtain

$$\forall x \in \mathbb{R}^d\,, A \in \mathcal{B}(\mathbb{R}^d) \quad P(x, A) \geqslant \varepsilon_d \nu_d(A)\,,$$

with $\varepsilon_d = \frac{\int_{\mathbb{R}^d} l(z)\,\mathrm{d}z}{c^d}$.