Convexity, Optimization and Stochastic Control
Exam 1st Session

3h, documents are not allowed

January 6, 2025

Optimal step size in gradient descent

In the sequel, we denote by $|\cdot|$ the euclidean norm on $\mathbb{R}^n$ given by $|x| = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $n \geq 1$.

We consider a function $f : \mathbb{R}^n \to \mathbb{R}$ that we suppose to be continuous and to satisfy the following condition

$$S_x := \{y \in \mathbb{R}^n : f(y) \leq f(x)\} \text{ is compact for all } x \in \mathbb{R}^n . \tag{1}$$

1. Show that under condition (1), the function f admits a global minimizer over $\mathbb{R}^n$.

2. Suppose in addition that f is strictly convex. Show that the global minimizer is unique.

We shall denote the unique global minimizer of f by x^* in the sequence. We now propose an algorithm to approximate the minimal value $f(x^*)$ and the global minimizer. For that, we additionally suppose that $f \in C^1(\mathbb{R}^n, \mathbb{R})$. We next define the sequence $(t_k, d_k, x_k)_{k\geq 0}$ of $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n$ by

$$\left\{ \begin{array}{rcl} d_k & = & -\nabla f(x_k) \\ t_k & \in & \underset{t\in\mathbb{R}}{\operatorname{argmin}}\, f(x_k + t d_k) \\ x_{k+1} & = & x_k + t_k d_k \end{array} \right. \tag{2}$$

for $k \geq 0$, where $x_0 \in \mathbb{R}^n$ is given.

3. Show that the function $t \in \mathbb{R} \mapsto f(x + td)$ admits a minimum over $\mathbb{R}$ for any $x, d \in \mathbb{R}^n$. Deduce that the sequence $(t_k, d_k, x_k)_{k\geq 0}$ defined by x_0 and (2) is well defined.

We now prove that the sequence $(x_k)_{k\geq 0}$ converges to x^*. For that we make the additional assumptions

$$f \in C^2(\mathbb{R}^n, \mathbb{R}) , \tag{3}$$

and

$$\exists M > 0 : \nabla^2 f(x) \leq M I_n , \quad \forall x \in \mathbb{R}^n . \tag{4}$$

4. Let $g: t \in \mathbb{R} \mapsto f(x_k + td_k)$. Using exact Taylor's formula and (4), show that

$$g(t) \leq f(x_k) + \Big(\frac{M}{2}t^2 - t\Big)|\nabla f(x_k)|^2 ,$$

for all $t \in \mathbb{R}$.

5. Deduce that the series $\sum_k |\nabla f(x_k)|^2$ converges and that $\lim_{k\to\infty} |\nabla f(x_k)|^2 = 0$.

We now show that the sequence (x_k) converges to x^*.

6. Show that the sequence $(f(x_k))_{k\geq 0}$ is nonincreasing. Deduce that it is bounded.
7. Show that the sequence $(x_k)_{k\geq 0}$ admits x^* as the unique adherence value.
8. Deduce that the sequence $(x_k)_{k\geq 0}$ converges to x^* as k goes to ∞.

We now consider a specific example of application. Suppose that f is given by

$$f(x) = \frac{1}{2}x^\top Qx + b^\top x , \quad x \in \mathbb{R}^n$$

where Q is a symmetric positive $n \times n$ matrix and $b \in \mathbb{R}^n$.

9. Show that f is strictly convex and satisfies the conditions (1), (4) and (3).
10. Give the explicit expressions of d_k, t_k and x_{k+1} depending on x_k for $k \geq 0$.

Optimal importance sampling and stochastic control

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space supporting a one-dimensional Brownian motion W. We denote by $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ the completed natural filtration induced by W, where $T > 0$.

For $(t, x) \in [0, T] \times \mathbb{R}$, we define $X^{t,x} = (X^{t,x}_s)_{s\in[t,T}$ as the solution to

$$X^{t,x}_s = x + \int_t^s \sigma(X^{t,x}_r)dW_r , \quad s \in [t, T] ,$$

where σ is Lipschitz and satisfies $\inf_{x\in\mathbb{R}} \sigma(x) =: \underline{\sigma} > 0$.

We consider the optimal importance sampling problem for the Monte-Carlo computation of

$$m(t, x) := \mathbb{E}[g(X^{t,x}_T)] , \quad (t, x) \in [0, T] \times \mathbb{R} ,$$

where g is continuous and bounded.

The idea consists in introduction a family of density processes $\{H^{t,x,\nu} , \nu \in \mathcal{A}\}$ defined by

$$H^{t,x,\nu}_s := 1 + \int_t^s \frac{\nu_r}{\sigma(X^{t,x}_r)} H^{t,x,\nu}_r dW_r , \quad s \in [t, T]$$

where $\mathcal{A}$ stands for the set of of progressively measurable processes ν with values in $[-M, M]$, where $M > 0$ is a fixed constant. We then observe that

$$m(t,x) = \mathbb{E}^{\nu}\big[g(X_T^{t,x})/H_T^{t,x,\nu}\big] \tag{5}$$

where $\mathbb{E}^{\nu}$ is the expectation under $\mathbb{P}^{\nu}$ defined as

$$d\mathbb{P}^{\nu}/d\mathbb{P} = H_T^{t,x,\nu}.$$

We then look for $\hat{\nu} \in \mathcal{A}$ such that

$$\mathrm{Var}^{\hat{\nu}}\big[g(X_T^{t,x})/H_T^{t,x,\hat{\nu}}\big] = \inf_{\nu\in\mathcal{A}} \mathrm{Var}^{\nu}\big[g(X_T^{t,x})/H_T^{t,x,\nu}\big] =: w(t,x) \tag{6}$$

for $(t,x) \in [0,T] \times \mathbb{R}$, where Var^{ν} denotes the variance under $\mathbb{P}^{\nu}$. We finally replace the standard Monte-Carlo estimator of $\mathbb{E}[g(X_T^{t,x})]$ by the one associated to $\mathbb{E}^{\hat{\nu}}\big[g(X_T^{t,x})/H_T^{t,x,\hat{\nu}}\big]$, obtained by sampling $g(X_T^{t,x})/H_T^{t,x,\hat{\nu}}$ under $\mathbb{P}^{\hat{\nu}}$. The aim is to treat problem (6) by stochastic control technics.

1. Show that $\mathbb{P}^{\nu}$ is a well defined probability measure equivalent to $\mathbb{P}$ and that (5) holds true, for $\nu \in \mathcal{A}$.

2. Show that the problem defined in the right-hand side of (6) is equivalent to

$$v(t,x) := \inf_{\nu\in\mathcal{A}} \mathbb{E}^{\nu}\big[\big(g(X_T^{t,x})/H_T^{t,x,\hat{\nu}}\big)^2\big], \quad (t,x) \in [0,T] \times \mathbb{R};,$$

   and that

$$v(t,x) = w(t,x) + m(t,x)^2, \quad (t,x) \in [0,T] \times \mathbb{R}. \tag{7}$$

In the following, we set

$$J(t,x;\nu) := \mathbb{E}^{\nu}\big[\big(g(X_T^{t,x})/H_T^{t,x,\nu}\big)^2\big], \quad (t,x) \in [0,T] \times \mathbb{R};,$$

and we admit that v is continuous on $[0,T) \times \mathbb{R}$.

3. Show that

$$J(t,x;\nu) = \mathbb{E}\big[g(X_T^{t,x})^2/H_T^{t,x,\nu}\big], \quad (t,x) \in [0,T] \times \mathbb{R};,$$

4. Let $\nu^1, \nu^2 \in \mathcal{A}$, $(t,x) \in [0,T] \times \mathbb{R}$ and $\tau \in \mathcal{T}_{[t,T]}$ where $\mathcal{T}_{[t,T]}$ denotes the set of stopping times with values in $[t,T]$.

   (a) Show that $\nu := \nu^1 \mathbb{1}_{[0,\tau)} + \nu^2 \mathbb{1}_{[\tau,T]} \in \mathcal{A}$.

   (b) Show that

$$J(t,x;\nu) = \mathbb{E}\big[\mathbb{E}\big[g(X_T^{t,x})^2/H_T^{\tau,X_\tau^{t,x},\nu^2}\big|\mathcal{F}_\tau\big]/H_\tau^{t,x,\nu^1}\big]$$

   (c) Deduce by a formal argument that

$$v(t,x) \geq \inf_{\nu\in\mathcal{A}} \mathbb{E}\big[v(\tau, X_\tau^{t,x})/H_\tau^{t,x,\hat{\nu}}\big].$$

(d) We assume that there exists a measurable map $\phi : [0,T] \times \mathbb{R} \to \mathcal{A}$ such that

$$v(\theta,\xi) \;=\; \mathbb{E}\big[g(X_T^{\theta,\xi})^2 / H_T^{\theta,\xi,\phi(\theta,\xi)} | \mathcal{F}_\theta\big]\,, \quad \mathbb{P}-a.s.$$

for all $\theta \in \mathcal{T}_{[0,T]}$ and all real valued $\mathcal{F}_\theta$ -mesurable random variable ξ. Deduce that

$$v(t,x) \;\leq\; \inf_{\nu\in\mathcal{A}} \mathbb{E}\big[v(\tau, X_\tau^{t,x})/H_\tau^{t,x,\nu}\big]\,.$$

From now on, we admit that v is continuous and satisfies

$$v(t,x) \;=\; \inf_{\nu\in\mathcal{A}} \mathbb{E}\big[v(\tau, X_\tau^{t,x})/H_\tau^{t,x,\nu}\big]\,. \tag{8}$$

We denote

$$\mathcal{L}^a\varphi(t,x) \;=\; \partial_t\varphi(t,x) - u\partial_x\varphi(t,x) + \frac{1}{2}\sigma^2(x)\partial^2_{xx}\varphi(t,x) + (u/\sigma(x))^2\varphi(t,x)$$

for $\varphi \in C^{1,2}([0,T]\times\mathbb{R})$, $a \in \mathbb{R}$ and $(t,x) \in [0,T]\times\mathbb{R}$.

5. By considering constant controls show that v is a viscosity subsolution on $[0,T)\times\mathbb{R}$ to

$$-\inf_{a\in[-M,M]} \mathcal{L}^a\varphi \;=\; 0\,. \tag{9}$$

6. Let $(t_0,x_0) \in [0,T)\times\mathbb{R}$ and $\varphi \in C^{1,2}([0,T]\times\mathbb{R})$ be such that (t_0,x_0) achieves a strict local minimum of $v-\varphi$ and $(v-\varphi)(t_0,x_0) = 0$.

   (a) Show that the assertion $-\inf_{a\in[-M,M]}\mathcal{L}^a\varphi(t_0,x_0) < 0$ leads to a contradiction to (8).

   (b) Deduce that v is a viscosity supersolution on $[0,T)\times\mathbb{R}$ to (9).

7. We now study the terminal condition at T.

   (a) Deduce from (7) that

   $$\liminf_{(t',x')\to(T-,x)} v(t',x') \;\geq\; g(x)^2$$

   for all $x \in \mathbb{R}$.

   (b) Show that $v(t,x) \leq \mathbb{E}[g(X_T^{t,x})^2]$ for all $(t,x) \in [0,T]\times\mathbb{R}$.

   (c) Deduce that

   $$\lim_{(t',x')\to(T-,x)} v(t',x') \;=\; v(T,x) \;=\; g(x)^2$$

   for all $x \in \mathbb{R}$.