Chapter 2: Introduction to stochastic approximation algorithms and their applications in mathematical finance

Noufel Frikha

Université Paris 1 Panthéon-Sorbonne

23th February 2026

Contents

Introduction

Introduction

In mathematical finance, one often has to solve optimization or inverse problems (zero search problems) such as

- Implicit volatility: find $\sigma \in \mathbb{R}_+$ such that
$$\mathbb{E}[(S_T(\sigma) - K)_+] = P_{market}$$
- Implicit correlation: find $\rho \in [-1, 1]$ such that
$$\mathbb{E}[(\max(S_T^1(\rho), S_T^2(\rho)) - K)_+] = P_{market}$$
where
$$S_T^1(\rho) = S_0^1 e^{(r-\sigma_1^2/2)T+\sigma_1 W_T^1}, \quad S_T^2(\rho) = S_0^2 e^{(r-\sigma_2^2/2)T+\sigma_2(\rho W_T^1 + \sqrt{1-\rho^2} W_T^2)},$$
- Computation of the (VaR_α, CVaR_α): find $(\xi^*, C^*) \in \mathbb{R}^2$ such that
$$\mathbb{P}(L \leq \xi^*) = \alpha, \quad C^* = \xi^* + \frac{1}{1-\alpha}\mathbb{E}[(L - \xi^*)_+].$$
- Static Risk minimization portfolio:
$$\inf_{\theta \in \Delta_d^+} \rho(-\langle \theta . X \rangle)$$
where ρ is a convex risk measure and X stands for assets return.

In all these situations one has to compute the zero of a function h which writes

$$h(y) = \mathbb{E}[H(y, Z)], \quad H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$$

where Z is a random variable easy to simulate.

In general, to estimate $\{y^*\} = \{h = 0\}$, one may use a Newton-Raphson like procedure with dynamics

$$y_{n+1} = y_n - \gamma_{n+1} h(y_n), \quad n \geq 0, \quad y_0 \in \mathbb{R}^d \tag{1}$$

$\rightsquigarrow$ Local convergence of Newton-Raphson algorithm: $\exists \delta > 0$ s.t.

$$\gamma_{n+1} = Dh(y_n)^{-1} \text{ and } |y_0 - y^*| \leq \delta \Rightarrow y_n \to y^*, n \to \infty.$$

To derive the convergence of $(y_n)_{n\geq 0}$ given by (1) one usually supposes that y^* is attractive in the following sense

$$\forall y \neq y^*, \quad \langle y - y^*, h(y) \rangle > 0 \qquad \text{(Mean reverting assumption)}$$

or, that there exists a Lyapunov function $L : \mathbb{R}^d \to \mathbb{R}_+$, s.t. $\lim_{|y|\to\infty} L(y) = +\infty$ and

$$\forall y \neq y^*, \quad \langle \nabla L(y), h(y) \rangle > 0 \quad \text{and} \quad \{\nabla L = 0\} = \{y^*\}$$

Convergence of the deterministic algorithm

S

uppose that $h : \mathbb{R}^d \to \mathbb{R}^d$ is continuous, $\{y^*\} = \{h = 0\}$ and satisfies:

- Mean reverting:

$$\forall y \neq y^*, \quad \langle y - y^*, h(y) \rangle > 0,$$

- Sub-linear growth:

$$\forall y \in \mathbb{R}^d, \quad |h(y)|^2 \leq C(1 + |y - y^*|^2)$$

- Step of the algorithm: $(\gamma_n)_{n\geq 1}$ is a positive step sequence satisfying

$$\sum_{n\geq 1} \gamma_n = \infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < \infty.$$

Then, one has

$$y_n \to y^*, \quad n \to \infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n \langle y_n - y^*, h(y_n) \rangle < \infty$$

Proof: Use the dynamics (1) to write

$$
\begin{aligned}
|y_{n+1}-y^*|^2 &= |y_n-y^*|^2 - 2\gamma_{n+1}\langle y_n-y^*, h(y_n)\rangle + \gamma_{n+1}^2|h(y_n)|^2 \\
&\leq (1+C\gamma_{n+1}^2)|y_n-y^*|^2 - 2\gamma_{n+1}\langle y_n-y^*, h(y_n)\rangle + C\gamma_{n+1}^2
\end{aligned}
$$

Let $(s_n)_{n\geq 1}$ be the positive sequence defined by

$$
s_n := \frac{|y_n-y^*|^2 + 2\sum_{k=0}^{n-1}\gamma_{k+1}\langle y_k-y^*, h(y_k)\rangle + C\sum_{k\geq n}\gamma_{k+1}^2}{\prod_{k=0}^{n-1}(1+C\gamma_{k+1}^2)}
$$

then, using the preceding inequality, one easily checks that

$$
\begin{aligned}
s_{n+1} &\leq \frac{(1+C\gamma_{n+1}^2))|y_n-y^*|^2 + 2\sum_{k=0}^{n-1}\gamma_{k+1}\langle y_k-y^*, h(y_k)\rangle + C\sum_{k\geq n}\gamma_{k+1}^2}{\prod_{k=0}^{n}(1+C\gamma_{k+1}^2)} \\
&\leq s_n
\end{aligned}
$$

so that, $(s_n)_{n\geq 1}$ is a positive non-increasing sequence so that it converges to some $s_\infty \geq 0$.
Since $\sum_{n\geq 1}\gamma_n^2 < \infty$, we deduce the two following facts

1. $(|y_n-y^*|^2)_{n\geq 1}$ is bounded,
2. $\sum_{n\geq 0}\gamma_{n+1}\langle y_n-y^*, h(y_n)\rangle < \infty$

so that, $|y_n-y^*|^2 \to L_\infty$ as $n\to\infty$.

From the above partial conclusion, we deduce that $\liminf_n \langle y_n - y^*, h(y_n)\rangle = 0$ so that there exists an increasing ϕ s.t.

$$y_{\phi(n)} \to y_\infty \quad \text{and} \quad \langle y_{\phi(n)} - y^*, h(y_{\phi(n)})\rangle \to 0.$$

Since h is continuous, one gets $\langle y_\infty - y^*, h(y_\infty)\rangle = 0$ so that $y_\infty = y^*$ and $L_\infty = 0$. We thus conclude that $(y_n)_{n\geq 1}$ converges to y^*.

Remarks:

- When $h(y) = \mathbb{E}[H(y,Z)]$, this procedure is not satisfactory because one does not have a direct access to the values of h.
- One may replace h by its Monte Carlo approximation using M i.i.d. samples $(Z^i)_{1\leq i\leq M}$:

$$h_M(y) := \frac{1}{M}\sum_{i=1}^{M} H(y, Z^i)$$

and then run the algorithm...

Robbins-Monro algorithm

Robbins-Monro theorem

Let $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ and set $h(y) = \mathbb{E}[H(y, Z)]$. Assume that h is continuous with $\{h = 0\} = \{y^*\}$ and satisfies the two following conditions:

(H1) Mean-reverting: $\forall y \neq y^*$, $\langle y - y^*, h(y)\rangle > 0$.

(H2) Linear growth: $\forall y \in \mathbb{R}^d$, $\mathbb{E}[|H(y, Z)|^2]^{\frac{1}{2}} \leq C(1 + |y - y^*|^2)$

Let $(\gamma_n)_{n\geq 1}$ be a sequence of positive numbers such that

$$\sum_{n\geq 1} \gamma_n = \infty, \quad \sum_{n\geq 1} \gamma_n^2 < \infty.$$

Finally, assume that Y_0 is a r.v. independent of $(Z^{(n)})_{n\geq 1}$ such that $\mathbb{E}[|Y_0|^2] < \infty$ then the recursive procedure

$$Y_{n+1} = Y_n - \gamma_{n+1} H(Y_n, Z^{(n+1)}), \quad n \geq 0$$

satisfies

$$Y_n \xrightarrow{a.s.} y^*, \text{as } n \to \infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n \langle Y_n - y^*, h(Y_n)\rangle < \infty, \ a.s.$$

The key of the proof is the convergence theorem for non-negative supermartingales:

◦ Let $(S_n)_{\geq 0}$ be a non-negative super-martingale w.r.t to $(\mathcal{F}_n)_{n\geq 0}$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, i.e. $S_n \geq 0$, $S_n \in L^1(\mathbb{P})$ and $\mathbb{E}[S_{n+1}|\mathcal{F}_n] \leq S_n$ a.s. then $S_n \stackrel{a.s.}{\to} S_\infty \geq 0$ as $n \to \infty$.

◦ Set $\mathcal{F}_n = \sigma(Y_0, Z^{(1)}, \cdots, Z^{(n)})$, $n \geq 1$ and $\Delta M_{n+1} := H(Y_n, Z^{(n+1)}) - h(Y_n)$. Note that

$$\mathbb{E}[\Delta M_{n+1}|\mathcal{F}_n] = \mathbb{E}[H(Y_n, Z^{(n+1)})|\mathcal{F}_n] - h(Y_n) = h(Y_n) - h(Y_n) = 0.$$

Hence,

$$\begin{aligned}
|Y_{n+1} - y^*|^2 &= |Y_n - y^*|^2 - 2\gamma_{n+1}\langle Y_n - y^*, H(Y_n, Z^{(n+1)})\rangle + \gamma_{n+1}^2 |H(Y_n, Z^{(n+1)})\\
&= |Y_n - y^*|^2 - 2\gamma_{n+1}\langle Y_n - y^*, h(Y_n)\rangle + \gamma_{n+1}^2 |H(Y_n, Z^{(n+1)})|^2\\
&\quad - 2\gamma_{n+1}\langle Y_n - y^*, \Delta M_{n+1})\rangle
\end{aligned}$$

which proves (by induction on n) that $|Y_n - y^*|^2 \in L^1(\mathbb{P})$. Then, one defines

$$\forall n \geq 1,\ S_n := \frac{|Y_n - y^*|^2 + 2\sum_{k=0}^{n-1} \gamma_{k+1}\langle Y_k - y^*, h(Y_k)\rangle + C\sum_{k\geq n}\gamma_{k+1}^2}{\prod_{k=0}^{n-1}(1 + C\gamma_{k+1}^2)}$$

$\circ$ $(S_n)_{n\geq 1}$ is a positive super-martingale (prove it!) so that it $a.s.$ converges to $S_\infty \in L^1(\mathbb{P})$ and

$$|Y_n - y^*|^2 + 2\sum_{k=0}^{n-1} \gamma_{k+1}\langle Y_k - y^*, h(Y_k)\rangle \overset{a.s.}{\to} S_\infty \prod_{k\geq 0}(1 + C\gamma_{k+1}^2) \in L^1(\mathbb{P})$$

so that, one has

- $\sum_{n\geq 0} \gamma_{n+1}\langle Y_n - y^*, h(Y_n)\rangle < \infty$, $a.s.$
- $|Y_n - y^*|^2 \overset{a.s.}{\to} L_\infty$, as $n \to \infty$

and $\exists$ ϕ such that

$$Y_{\phi(n)} \to y_\infty, \quad \text{and} \quad \lim_n \langle Y_{\phi(n)} - y^*, h(Y_{\phi(n)})\rangle = \langle y_\infty - y^*, h(y_\infty)\rangle = 0$$

so that

$$\langle y_\infty - y^*, h(y^*)\rangle = 0$$

$(Y_n)_{n\geq 1}$ admits only one “valeur d'adhérence” which is y^*.

∘ Possible extension of this theorem: If $L : \mathbb{R}^d \to \mathbb{R}_+$ is a differentiable function such that ∇L is Lipschitz continuous and $|\nabla L|^2 \leq C(1+L)$, $\lim_{|y|\to\infty} L(y) = \infty$, $\{\nabla L = 0\} = \{y^*\}$ with

$$\forall y \neq y^*, \quad \langle \nabla L(y), y - y^* \rangle > 0.$$

and

$$\mathbb{E}[|H(y,Z)|^2]^{\frac{1}{2}} \leq C(1 + L(y))$$

Then, if $L(Y_0) \in L^1(\mathbb{P})$, one has $Y_n \overset{a.s.}{\to} y^*$, as $n \to \infty$.

Central limit theorem

We want to investigate the convergence rate of the sequence $(Y_n)_{n\geq 0}$ toward y^*. We will assume that

$$Y_{n+1} = Y_n - \gamma_{n+1} h(Y_n) + \gamma_{n+1}\Delta M_{n+1}, \quad \Delta M_{n+1} = h(Y_n) - H(Y_n, Z^{(n+1)})$$

satisfies the following assumptions:

- h is twice continuously differentiable
- $Dh(y^*)$ is an Hurwitz matrix, that is, the smallest real part λ_m of the eigenvalues of $Dh(y^*)$ is positive.
- $y \mapsto \mathbb{E}[H(y,Z)H(y,Z)^T]$ is continuous and $y \mapsto \mathbb{E}[|H(y,Z)|^{2+\delta}]$ is locally bounded for some $\delta > 0$.
- $\Sigma^* = \mathbb{E}[H(y^*,Z)H(y^*,Z)^T]$ is definite positive.
- The step sequence $(\gamma_n)_{n\geq 1}$ satisfies one of the following conditions:
  - $\gamma_n = \frac{\gamma_0}{n+1}$, γ_0 s.t. $2\gamma_0\lambda_m > 1$.
  - $\gamma_n = \frac{\gamma_0}{(n+1)^\rho}$, $\rho \in (\frac{1}{2}, 1)$.

Then, one has

$$\sqrt{\gamma_n^{-1}}(Y_n - y^*) \Rightarrow \mathcal{N}(0, \Gamma^*), \text{ with } \Gamma^* := \int_0^\infty (e^{-(Dh(y^*)-\xi I_d)s})^T \Sigma^* e^{-(Dh(y^*)-\xi I_d)s} ds$$

SGD and Projected SGD

SGD and Projected SGD

◦ The classical SA algorithm solves problem

$$\min_{x \in X} g(x) = \mathbb{E}[G(x, Z)] \tag{2}$$

where $X \subset \mathbb{R}^d$ is a nonempty bounded closed convex set and Z a random variable from which we can easily sample, by mimicking the simplest subgradient descent method.
◦ For chosen $x_1 \in X$ and a sequence $\gamma_j > 0$, $j = 1, \ldots$, of stepsizes, it generates the iterates by the formula

$$x_{j+1} := \Pi_X \left(x_j - \gamma_j \nabla_x G(x_j, Z^j) \right), \tag{3}$$

where $\Pi_X(x) = \arg\min_{x' \in X} \|x - x'\|_2$ and $(Z^j)_{j \geq 1}$ is an i.i.d. sequence with the same law as Z.

◦ Of course, the crucial question of that approach is how to choose the stepsizes γ_j.

◦ Let $\bar{x}$ be an optimal solution of problem (2). Note that since the set X is compact and g is continuous, problem (2) has an optimal solution. Note also that the iterate $x_j = x_j(Z_{[j-1]})$ is a function of the history $Z_{[j-1]} := (Z_1, \ldots, Z_{j-1})$ of the generated random process and hence is random.

◦ Denote $A_j := \frac{1}{2}\|x_j - \bar{x}\|_2^2$ and $a_j := \mathbb{E}[A_j] = \frac{1}{2}\mathbb{E}[\|x_j - \bar{x}\|_2^2]$. By using the fact that Π_X is a contraction operator and since $\bar{x} \in X$ and hence $\Pi_X(\bar{x}) = \bar{x}$, we can write

$$\begin{aligned} A_{j+1} &= \frac{1}{2}\left\|\Pi_X\left(x_j - \gamma_j \nabla_x G(x_j, Z^j)\right) - \bar{x}\right\|_2^2 \\ &\leq A_j + \frac{\gamma_j^2}{2}\|\nabla_x G(x_j, Z^j)\|_2^2 - \gamma_j (x_j - \bar{x})^\top \nabla_x G(x_j, Z^j). \end{aligned}$$

We also have

$$\begin{aligned} \mathbb{E}[(x_j - \bar{x})^\top \nabla_x G(x_j, Z^j)] &= \mathbb{E}\left\{\mathbb{E}\left[(x_j - \bar{x})^\top \nabla_x G(x_j, Z^j) | \mathcal{F}_{j-1}\right]\right\} \\ &= \mathbb{E}[(x_j - \bar{x})^\top \nabla g(x_j)] \end{aligned}$$

using the fact that x_j is $\mathcal{F}_{j-1}$ measurable and Z^j is independent of $\mathcal{F}_j$.

◦ Therefore, by taking expectation on both sides of the previous inequality, we obtain

$$a_{j+1} \le a_j - \gamma_j \mathbb{E}[(x_j - \bar{x})^\top \nabla g(x_j)] + \frac{\gamma_j^2}{2} M^2, \tag{4}$$

where

$$M^2 := \sup_{x \in X} \mathbb{E}[\|\nabla_x G(x, \xi)\|_2^2]. \tag{5}$$

$\rightsquigarrow$ We assume that the above constant M is finite.

◦ Suppose, further, that the expectation function g is strongly convex on X, i.e., there is a constant $c > 0$ such that

$$g(x') \ge g(x) + (x' - x)^\top \nabla g(x) + \frac{c}{2}\|x' - x\|_2^2, \quad \forall x', x \in X, \tag{6}$$

or equivalently

$$(x' - x)^\top (\nabla g(x') - \nabla g(x)) \ge c\|x' - x\|_2^2, \quad \forall x', x \in X. \tag{7}$$

◦ Strong convexity of g implies that the minimizer $\bar{x}$ is unique. By optimality of $\bar{x}$, we have that

$$(x-\bar{x})^\top \nabla g(x) \geq 0, \quad \forall x \in X. \tag{8}$$

Combining (8) with (7), we obtain

$$\begin{aligned} \mathbb{E}[(x_j-\bar{x})^\top \nabla g(x_j)] &\geq \mathbb{E}[(x_j-\bar{x})^\top (\nabla g(x_j) - \nabla g(\bar{x}))] \\ &\geq c\mathbb{E}[\|x_j-\bar{x}\|_2^2] = 2ca_j. \end{aligned} \tag{9}$$

Substituting (9) into (4), we get

$$a_{j+1} \leq (1-2c\gamma_j)a_j + \frac{\gamma_j^2}{2}M^2. \tag{10}$$

◦ Let us take stepsizes $\gamma_j = \frac{\gamma}{j}$ for some constant $\gamma > \frac{1}{2c}$. Then by (10), we have

$$a_{j+1} \leq \left(1 - \frac{2c\gamma}{j}\right) a_j + \frac{\gamma^2 M^2}{2j^2}. \tag{11}$$

By induction, it follows that

$$a_j \leq \frac{\kappa}{j}, \tag{12}$$

where

$$\kappa := \max\left\{\frac{\gamma^2 M^2}{2(2c\gamma - 1)}, a_1\right\}. \tag{13}$$

◦ Suppose further that $\bar{x}$ is an interior point of X and $\nabla g(x)$ is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|\nabla g(x') - \nabla g(x)\|_2 \leq L\|x' - x\|_2, \quad \forall x', x \in X. \tag{14}$$

Then,

$$g(x) \leq g(\bar{x}) + \frac{L}{2}\|x - \bar{x}\|_2^2, \quad \forall x \in X, \tag{15}$$

and hence

$$\mathbb{E}[g(x_j) - g(\bar{x})] \leq L a_j \leq \frac{L\kappa}{j}. \tag{16}$$

◦ Conclusion: Under the specified assumptions, it follows that after n iterations, the expected error of the current solution is of order $O(n^{-1/2})$, and the expected error of the corresponding objective value is of order $O(n^{-1})$, provided that $\gamma > \frac{1}{2c}$.

◦ Caution: However, this result is highly sensitive to the choice of c. Overestimating c can lead to suboptimal convergence,

Some difficulties of GD and SGD

- Implementing a GD or SGD scheme requires fine tuning of learning sequence $(\gamma_k)_{k\geq 1}$.

Some difficulties of GD and SGD

- Implementing a GD or SGD scheme requires fine tuning of learning sequence $(\gamma_k)_{k\geq 1}$.

- **Basic example:** Consider the optimization problem:

$$\text{Minimize } \left\{ f(x) = \frac{x^2}{10} \right\}, \quad \text{over} \quad X = [-1,1] \subset \mathbb{R}.$$

Some difficulties of GD and SGD

- Implementing a GD or SGD scheme requires fine tuning of learning sequence $(\gamma_k)_{k\geq 1}$.
- **Basic example:** Consider the optimization problem:

$$\text{Minimize } \left\{ f(x) = \frac{x^2}{10} \right\}, \quad \text{over} \quad X = [-1, 1] \subset \mathbb{R}.$$

- **Iteration Rule:** of the standard GD scheme with learning rate $\gamma_k = \frac{\gamma}{k}$, $\gamma > 0$.

$$x_{k+1} = x_k - \frac{f'(x_k)}{k} = \left(1 - \frac{\gamma}{5k}\right) x_k.$$

Some difficulties of GD and SGD

- Implementing a GD or SGD scheme requires fine tuning of learning sequence $(\gamma_k)_{k\geq 1}$.

- **Basic example:** Consider the optimization problem:

$$\text{Minimize } \left\{ f(x) = \frac{x^2}{10} \right\}, \quad \text{over} \quad X = [-1, 1] \subset \mathbb{R}.$$

- **Iteration Rule:** of the standard GD scheme with learning rate $\gamma_k = \frac{\gamma}{k}$, $\gamma > 0$.

$$x_{k+1} = x_k - \frac{f'(x_k)}{k} = \left(1 - \frac{\gamma}{5k}\right) x_k.$$

- **Parameters:** $x_1 = 1$ and $\gamma = 1$. It holds

$$x_n = \prod_{k=1}^{n-1} (1 - \frac{1}{5k}) = e^{-\sum_{k=1}^{n-1} \ln(1 + \frac{1}{5k-1})} > e^{-(0.25 + \int_1^{n-1} \frac{1}{5t-1}\, dt)} > 0.8 n^{-\frac{1}{5}}$$

Some difficulties of GD and SGD

- Implementing a GD or SGD scheme requires fine tuning of learning sequence $(\gamma_k)_{k\geq 1}$.

- **Basic example:** Consider the optimization problem:

$$\text{Minimize } \left\{ f(x) = \frac{x^2}{10} \right\}, \quad \text{over} \quad X = [-1, 1] \subset \mathbb{R}.$$

- **Iteration Rule:** of the standard GD scheme with learning rate $\gamma_k = \frac{\gamma}{k}$, $\gamma > 0$.

$$x_{k+1} = x_k - \frac{f'(x_k)}{k} = \left(1 - \frac{\gamma}{5k}\right) x_k.$$

- **Parameters:** $x_1 = 1$ and $\gamma = 1$. It holds

$$x_n = \prod_{k=1}^{n-1} (1 - \frac{1}{5k}) = e^{-\sum_{k=1}^{n-1} \ln(1 + \frac{1}{5k-1})} > e^{-(0.25 + \int_1^{n-1} \frac{1}{5t-1}\, dt)} > 0.8 n^{-\frac{1}{5}}$$

$\rightsquigarrow$ Slow convergence: For $n = 10^9$, error remains > 0.013.

Some difficulties of GD and SGD

- Implementing a GD or SGD scheme requires fine tuning of learning sequence $(\gamma_k)_{k\geq 1}$.
- **Basic example:** Consider the optimization problem:

$$\text{Minimize } \left\{ f(x) = \frac{x^2}{10} \right\}, \quad \text{over} \quad X = [-1, 1] \subset \mathbb{R}.$$

- **Iteration Rule:** of the standard GD scheme with learning rate $\gamma_k = \frac{\gamma}{k}$, $\gamma > 0$.

$$x_{k+1} = x_k - \frac{f'(x_k)}{k} = \left(1 - \frac{\gamma}{5k}\right) x_k.$$

- **Parameters:** $x_1 = 1$ and $\gamma = 1$. It holds

$$x_n = \prod_{k=1}^{n-1} (1 - \frac{1}{5k}) = e^{-\sum_{k=1}^{n-1} \ln(1+\frac{1}{5k-1})} > e^{-(0.25+\int_1^{n-1} \frac{1}{5t-1}\, dt)} > 0.8 n^{-\frac{1}{5}}$$

⇝ Slow convergence: For $n = 10^9$, error remains > 0.013.
⇝ Optimal choice: $\gamma = 1/c = 5$, yields $x_k = 0$ in a single iteration.

Mirror Descent Algorithm

Notations and tools

∘ Our goal here is to develop a generalization of the stochastic approximation scheme introduced so far.

We say that a function $w : X \to \mathbb{R}$ is a distance generating function modulus $\alpha > 0$, if w is convex and continuous on X, the set

$$X^o := \left\{ x \in X : \text{there exists } p \in \mathbb{R}^d \quad \text{such that } x \in \arg\min_{u \in X} (p^t u + w(u)) \right\}$$

is convex, and restricted to X^o, w is continuously differentiable and strongly convex with parameter α, namely

$$(x' - x)^t (\nabla w(x') - \nabla w(x)) \geq \alpha \|x - x'\|^2, \quad \forall x, x' \in X^o.$$

• Example: $w(x) = \|x\|^2/2$, $\rightsquigarrow X^o = X$,

$$\Pi_X(x - y) = \arg\min_{z \in X} \left\{ y^T (z - x) + \frac{1}{2} \|z - x\|_2^2 \right\}$$

and

$$\frac{1}{2} \|z - x\|_2^2 = w(z) - [w(x) + \nabla w(x)^T (z - x)]$$

◦ Let us define the *proxy-function* $V : X^o \times X \to \mathbb{R}_+$ as follows

$$V(x, z) := w(z) - [w(x) + \nabla w(x)^T (z - x)]$$

associated to the distance generating function w.
◦ We now define the *proxy mapping* $P_x : \mathbb{R}^d \to X^o$ associated with w and a point $x \in X^o$:

$$P_x(y) = \arg\min_{z \in X} \left\{ y^T (z - x) + V(x, z) \right\}.$$

• Example: For $w(x) = \frac{1}{2}\|x\|_2^2$, we have $P_x(y) = \Pi_X(x - y)$.

Note that the projected SGD writes

$$x_{j+1} = \Pi_X(x_j - \gamma_j \nabla_x G(x_j, Z^j)) = P_{x_j}(\gamma_j \nabla_x G(x_j, Z^j)), \quad j \geq 1$$

$\rightsquigarrow$ Study the above recursion in the general case of distance generating function w.

Lemma

For any $u \in X$, any $x \in X^o$ and y, the following inequality holds

$$V(P_x(y), u) \leq V(x, u) + y^T(u - x) + \frac{1}{2\alpha}\|y\|_\star^2$$

where $\|x\|_\star = \sup_{\|y\| \leq 1} y^T x$.

Using the above inequality with $x = x_j$, $y = \gamma_j \nabla_x G(x_j, Z^j)$ and $u = \bar{x}$, we get

$$\gamma_j \langle \nabla_x G(x_j, Z^j), x_j - \bar{x} \rangle \leq V(x_j, \bar{x}) - V(x_{j+1}, \bar{x}) + \frac{\gamma_j^2}{2\alpha}\|\nabla_x G(x_j, Z^j)\|_\star^2$$

which equivalently writes

$$\gamma_j \langle \nabla g(x_j), x_j - \bar{x} \rangle \leq V(x_j, \bar{x}) - V(x_{j+1}, \bar{x}) - \gamma_j \langle \Delta M_j, x_j - \bar{x} \rangle + \frac{\gamma_j^2}{2\alpha}\|\nabla_x G(x_j, Z^j)\|_\star^2$$

where $\Delta M_j := \nabla_x G(x_j, Z^j) - \nabla g(x_j)$.

Summing up from $j = 1, \cdots, n$, and taking into account that $V(x_j, u) \geq 0$ for any $u \in X$, we get

$$\sum_{j=1}^{n} \gamma_j \langle \nabla g(x_j), x_j - \bar{x} \rangle \leq V(x_1, \bar{x}) - \sum_{j=1}^{t} \gamma_j \langle \Delta M_j, x_j - \bar{x} \rangle + \sum_{j=1}^{n} \frac{\gamma_j^2}{2\alpha} \|\nabla_x G(x_j, Z^j)\|_\star^2$$

Introduce the sequence

$$\bar{x}_n = \frac{1}{\sum_{j=1}^n \gamma_j} \sum_{j=1}^{n} \gamma_j x_j$$

By convexity of g,

$$\begin{aligned} g(\bar{x}_n) - g(\bar{x}) &\leq \frac{1}{\sum_{j=1}^n \gamma_j} \sum_{j=1}^{n} \gamma_j (g(x_j) - g(\bar{x})) \\ &\leq \frac{1}{\sum_{j=1}^n \gamma_j} \sum_{j=1}^{n} \gamma_j \langle \nabla g(x_j), x_j - \bar{x} \rangle \\ &\leq \frac{V(x_1, \bar{x}) - \sum_{j=1}^t \gamma_j \langle \Delta M_j, x_j - \bar{x} \rangle + \sum_{j=1}^n \frac{\gamma_j^2}{2\alpha} \|\nabla_x G(x_j, Z^j)\|_\star^2}{\sum_{j=1}^n \gamma_j} \end{aligned}$$

As previously done, let us assume that there exists $M_\star < \infty$ such that $\|\nabla_x G(x,Z)\|_\star \leq M_\star$.
We also remark that if $\mathcal{F}_j = \sigma(x_0, Z^1, \cdots, Z^j)$, then x_{j+1} is $\mathcal{F}_j$-mesurable and Z^{j+1} is independent of $\mathcal{F}_j$.
Hence,

$$\mathbb{E}[\Delta M_{j+1} \,|\, \mathcal{F}_j] = \mathbb{E}[\nabla_x G(x_{j+1}, Z^{j+1})|\, \mathcal{F}_j] - \nabla g(x_{j+1}) = \nabla g(x_{j+1}) - \nabla g(x_{j+1}) = 0$$

We eventually get

$$g(\bar{x}_n) - \min_{x \in X} g(x) \leq \frac{V(x_1, \bar{x}) + (2\alpha)^{-1} M_\star^2 \sum_{j=1}^n \gamma_j^2}{\sum_{j=1}^n \gamma_j}.$$

◦ Question : what about the stepsize strategy?

Robust Mirror Descent SA algorithm with constant stepsize policy:
If the number of iterations N of the mirror descent scheme is fixed in advance then the constant stepsize strategy $\gamma_j = \gamma$, $j = 1, \cdots, N$ can be implemented. Indeed, it holds

$$\mathbb{E}[g(\bar{x}_N)] - \min_{x \in X} g(x) \leq \frac{V(x_1, \bar{x})}{\gamma N} + \frac{M_\star^2 \gamma}{2\alpha}$$

If we set $\gamma = \frac{V(x_1,\bar{x})}{M_\star}\sqrt{\frac{2\alpha}{N}}$, then we get

$$\mathbb{E}[g(\bar{x}_N)] - \min_{x \in X} g(x) \leq V(x_1, \bar{x}) M_\star \sqrt{\frac{2}{\alpha N}}$$

By the Markov inequality, we get

$$\mathbb{P}(g(\bar{x}_N) - \min_{x \in X} g(x) > \varepsilon) \leq \frac{V(x_1, \bar{x}) M_\star}{\varepsilon} \sqrt{\frac{2}{\alpha N}}$$

To obtain much finer bounds, let us assume that

$$\mathbb{E}[\exp(\|G(x,Z)\|_\star^2 / M_\star^2)] \leq \exp(1), \quad \forall x \in X.$$

Then, for the above constant stepsize choice, for some constant $C > 0$, for any $\varepsilon > 0$

$$\mathbb{P}(g(\bar{x}_N) - \min_{x \in X} g(x) > \varepsilon) \leq C \exp(-\kappa \varepsilon \sqrt{N})$$

where $\kappa = \sqrt{\alpha}/(\sqrt{8} M_\star V(x_1, \bar{x}))$. It follows that for chosen accuracy ε and confidence level $\delta \in (0,1)$, the sample size

$$N \geq \frac{C M_\star^2 V(x_1, \bar{x})^2 \log^2(\delta^{-1})}{\alpha \varepsilon^2}$$

guarantees that $\bar{x}_N$ is an ε-optimal solution of the true problem with probability at least $1 - \delta$.

Applications

Computation of VaR and ES

Aim: Compute the VaR and ES of the loss L of a financial portfolio, where L is easily simulable. Recall that

$$\mathsf{VaR}_\alpha(L) = \inf\{y : \mathbb{P}(L \le y) \ge \alpha\}, \quad \alpha \in (0,1).$$

We will assume that the c.d.f. of L is continuous and strictly increasing so that $\mathsf{VaR}_\alpha(L)$ is the unique solution ξ^* of

$$\mathbb{P}(L \le \xi) = \alpha.$$

If $L \in L^1(\mathbb{P})$, the Expected Shortfall of L is defined by

$$\mathsf{ES}_\alpha(L) = \mathbb{E}[L | L \ge \mathsf{VaR}_\alpha(L)] = \xi^* + \frac{1}{1-\alpha}\mathbb{E}[(L-\xi^*)_+].$$

Actually, $(\mathsf{VaR}_\alpha(L), \mathsf{ES}_\alpha(L))$ solves a convex optimization problem:

Proposition

Set $V(\xi) = \mathbb{E}[v(\xi, L)]$ with $v(\xi,\ell) := \xi + \frac{1}{1-\alpha}(\ell-\xi)_+$. Then, V is convex, coercice, differentiable and $\mathsf{VaR}_\alpha(L)$ is the solution of

$$\mathsf{ES}_\alpha(L) = \min_{\xi\in\mathbb{R}} V(\xi), \quad \arg\min V = \{\xi : V'(\xi) = \mathbb{E}[\partial_\xi v(\xi, L)] = 0\}.$$

To estimate $\xi^* = \mathsf{VaR}_\alpha(\varphi(X))$ satisfying $V'(\xi^*) = 0$, we devise the stochastic approximation algorithm

$$\xi_{n+1} = \xi_n - \gamma_{n+1} H_1(\xi_n, L^{(n+1)}), \quad H_1(\xi, \ell) = \partial_\xi v(\xi, \ell) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\ell \geq \xi\}}$$

where $\xi_0 \in L^1(\mathbb{P})$ is independent of $(L^{(n)})_{n\geq 1}$, $\sum_{n\geq 1}\gamma_n = \infty$ and $\sum_{n\geq 1}\gamma_n^2 < \infty$. Then, according to Robbins-Monro theorem

$$\xi_n \xrightarrow{a.s.} \xi^*, \quad n \to \infty.$$

To compute the $\mathsf{ES}_\alpha(L)$, one uses the following approximation

$$\mathsf{ES}_\alpha(L) = \xi^* + \frac{1}{1-\alpha}\mathbb{E}[(L-\xi^*)_+] \approx C_n := \frac{1}{n}\sum_{k=0}^{n-1} \xi_k + \frac{1}{1-\alpha}(L^{(k+1)} - \xi_k)_+$$

using the same r.v. $(L^{(k)})_{k\geq 1}$ as the one employed in the first algorithm. Note that it may be written in recursive form

$$C_{n+1} = C_n - \frac{1}{n+1}H_2(C_n, \xi_n, L^{(n+1)}), \quad H_2(c, \xi, \ell) := c - \left(\xi + \frac{1}{1-\alpha}(\ell - \xi)_+\right).$$

Hence, we see that the couple VaR-CVaR may be approximated by the stochastic algorithm

$$\begin{cases} \xi_{n+1} & = \xi_n - \gamma_{n+1} H_1(\xi_n, L^{(n+1)}) \\ C_{n+1} & = C_n - \frac{1}{n+1} H_2(\xi_n, C_n, L^{(n+1)}) \end{cases}$$

Once, one has proved that $\xi_n \xrightarrow{a.s.} \xi^*$, one can prove that

$$C_n \xrightarrow{a.s.} C^* = \mathsf{ES}_\alpha(L), \; n \to \infty.$$

Then, under the additional assumption $\mathbb{E}[|L|^{2+\delta}] < \infty$ for some $\delta > 0$, taking $\gamma_n = \gamma_1/n$, $\gamma_1 > 0$, one may prove the CLT

$$\sqrt{n}\begin{pmatrix} \xi_n - \xi^* \\ C_n - C^* \end{pmatrix} \Rightarrow \mathcal{N}(0, \Sigma^*), \quad \Sigma^* = \begin{pmatrix} \frac{\alpha(1-\alpha)}{f_L(\xi^*)^2} & \frac{\alpha}{1-\alpha}\mathbb{E}[(L-\xi^*)_+] \\ \frac{\alpha}{1-\alpha}\mathbb{E}[(L-\xi^*)_+] & \frac{Var((L-\xi^*)_+)}{(1-\alpha)^2} \end{pmatrix}$$

if $2\gamma_1 f_L(\xi^*) > 1$ where $y \mapsto f_L(y)$ is the density function of the loss L.

Variance reduction by SGD

∘ The price of an option often writes $\mathbb{E}[\varphi(Z)]$, where $Z \sim \mathcal{N}(0, I_d)$.
We use the transformation

$$\forall \theta \in \mathbb{R}^d, \quad \mathbb{E}[\varphi(Z)] = \mathbb{E}[\varphi(Z+\theta) e^{-\langle Z,\theta\rangle - \frac{|\theta|^2}{2}}]$$

so that

$$\mathrm{Var}(\varphi(Z+\theta) e^{-\langle Z,\theta\rangle - \frac{|\theta|^2}{2}}) = L(\theta) - \mathbb{E}[\varphi(Z)]^2$$

with

$$L(\theta) = e^{-|\theta|^2}\mathbb{E}[\varphi(Z+\theta)^2 e^{-2\langle Z,\theta\rangle}] = \mathbb{E}[\varphi(Z)^2 e^{-\langle Z,\theta\rangle + \frac{|\theta|^2}{2}}].$$

We have seen that if $\forall a$, $\mathbb{E}[\varphi(Z)^2 |Z| e^{a|Z|}] < +\infty$ then $L \in \mathcal{C}^1(\mathbb{R}^d)$, strictly convex with $\lim_{|\theta|\to\infty} L(\theta) = +\infty$ if $\mathbb{P}(\varphi(Z) \neq 0) > 0$ and

$$\nabla L(\theta) = \mathbb{E}[\varphi(Z)^2(\theta - Z) e^{-\langle Z,\theta\rangle + \frac{|\theta|^2}{2}}] = \mathbb{E}[\varphi(Z-\theta)^2(2\theta - Z)] e^{|\theta|^2}$$

Hence, L has a unique global minimum θ^* satisfying $\nabla L(\theta^*) = 0$ and since L is strictly convex

(Mean-reverting) $\quad \forall \theta \neq \theta', \quad \langle \nabla L(\theta) - \nabla L(\theta'), \theta - \theta'\rangle > 0$

◦ Assume moreover the following control of the growth of the payoff:

$$0 \leq \varphi(z) \leq C e^{\frac{a}{2}|z|}, \quad \forall z \in \mathbb{R}^d.$$

Define

$$H(\theta, z) = e^{-a(1+\theta|^2)^{\frac{1}{2}}} \varphi(z-\theta)^2 (2\theta - z)$$

then one has

$$\{\theta^*\} = \left\{\theta \in \mathbb{R}^d : h(\theta) = \mathbb{E}[H(\theta, Z)] = 0\right\} = \left\{\theta \in \mathbb{R}^d : \nabla L(\theta) = 0\right\}$$

and

$$\forall \theta \in \mathbb{R}^d, \quad \mathbb{E}[|H(\theta, Z)|^2] \leq C(1 + |\theta|^2)$$

so that, by the Robbins-Monro theorem, the procedure

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, Z^{(n+1)}),$$

$(Z^{(n)})_{n\geq 1}$ being an i.i.d. sequence of r.v. with common law $\mathcal{N}(0, I_d)$ satisfies

$$\theta_n \xrightarrow{a.s.} \theta^* = \arg\min L.$$

One finally implements the Monte Carlo estimator which satisfies

$$\sqrt{M}\left(\frac{1}{M}\sum_{k=1}^{M} \varphi(Z^{(k+1)} + \theta_k) e^{-\langle Z^{(k+1)}, \theta_k\rangle - \frac{|\theta_k|^2}{2}} - \mathbb{E}[\varphi(Z)]\right) \Rightarrow \mathcal{N}(0, L(\theta^*) - \mathbb{E}[\varphi(Z)]^2)$$

Risk minimization portfolio by SMD

We consider a financial market composed of d assets whose returns are given by the $\mathbb{R}^d$-valued random variable X.

A financial portfolio is identified with its corresponding vector of weights $u = (u_1, \cdots, u_d)$ belonging to the simplex $\Delta_d = \left\{u \in \mathbb{R}_+^d : u_1 + \cdots + u_d = 1\right\}$.

If the portfolio is given by the vector $u \in \Delta_d$, then $-\langle u, X\rangle = -\sum_{i=1}^d u_i X_i$ corresponds to its loss.

We want to find the following risk minimization portfolio

$$\bar{u} \in \arg\min_{u \in \Delta_d} \rho(-\langle u, X\rangle)$$

where ρ is a convex risk measure.

◦ Examples:

- The variance $\rho(Z) = \mathrm{Var}(Z)$
- The Expected Shortfall $\rho(Z) = \mathrm{ES}_\alpha(Z) := \min_\xi \left\{\mathbb{E}[v(\xi, Z)] := \xi + \frac{1}{1-\alpha}\mathbb{E}[(Z-\xi)_+]\right\}$.

To handle the optimization domain $\Delta_d = \left\{u \in \mathbb{R}^d_+ : u_1 + \cdots + u_d = 1\right\}$, we can use the Mirror Descent algorithm with the *negative entropy* as a *distance generating function*

$$w(u) = \sum_{i=1}^{d} u_i \log(u_i), \quad u \in \Delta_d,$$

so that the *proxy function*

$$V(u, u') = \sum_{i=1}^{d} u'_i \log(u'_i / u_i) \rightsquigarrow V(u, u') \geq \frac{1}{2}\|u - u'\|^2$$

The deterministic Mirror Descent algorithm writes

$$\begin{aligned} u_{k+1} &= P_{u_k}(\gamma_k \nabla_u [\rho(-\langle u_k, X \rangle)]) \\ &= \arg\min_{u \in \Delta_d} \{\gamma_k \langle \nabla_u [\rho(-\langle u_k, X \rangle)], u - u_k\rangle + V(u_k, u)\} \end{aligned}$$

which can be solved explicitly

$$u^i_{k+1} = \frac{u^i_k \exp(-\gamma_k \partial_{u_i}[\rho(-\langle u_k, X \rangle)])}{\sum_{j=1}^{d} u^j_k \exp(-\gamma_k \partial_{u_j}[\rho(-\langle u_k, X \rangle)])}, \quad i = 1, \cdots, d.$$

To handle the case of the Expected Shortfall, one has to rely on the Stochastic Mirror Descent Scheme. The portfolio optimization problem writes

$$(\bar{u}, \bar{\xi}) \in \arg \min_{(u,\xi)\in\Delta_d\times\mathbb{R}} \mathbb{E}[v(\xi, -\langle u, X\rangle)]$$

Define

$$w(u,\xi) = \sum_{i=1}^{d} u_i \log(u_i) + \frac{1}{2}\xi^2, \quad (u,\xi) \in \Delta_d \times \mathbb{R},$$

so that the *proxy function*

$$V((u,\xi),(u',\xi')) = \sum_{i=1}^{d} u'_i \log(u'_i/u_i) + \frac{1}{2}(\xi - \xi')^2$$

which satisfies

$$V((u,\xi),(u',\xi')) \geq \frac{1}{2}\|u - u'\|^2 + \frac{1}{2}(\xi - \xi')^2.$$

The SMD now writes

$$(u_{k+1}, \xi_{k+1})$$
$$= \arg \min_{(u,\xi)\in\Delta_d\times\mathbb{R}} \left\{\gamma_k \langle \nabla_{(\xi,u)} v(\xi, -\langle u_k, X^k\rangle), (u,\xi) - (u_k,\xi_k)\rangle + V((u,\xi),(u_k,\xi_k))\right\}$$

The solution to the previous minimization problem is explicit

$$u_{k+1}^i = \frac{u_k^i \exp(-\gamma_k \partial_{u_i} v(\xi_k, -\langle u_k, X^k\rangle))}{\sum_{j=1}^d u_k^j \exp(-\gamma_k \partial_{u_j} v(\xi_k, -\langle u_k, X^k\rangle))}, \quad i = 1, \cdots, d$$
$$\xi_{k+1} = \xi_k - \gamma_k \partial_\xi v(\xi_k, -\langle u_k, X^k\rangle)$$

where $(X^k)_{k\geq 1}$ is an i.i.d. sequence or random variables with the same law as X.

If we set $x_k = (u_k, \xi_k)$, $k \geq 1$, and

$$g(x) = \mathbb{E}[v(\xi, -\langle u, X\rangle)], \quad x = (u, \xi).$$

Take a constant step size $\gamma = \frac{V(x_1,\bar{x})}{M_\star}\sqrt{\frac{2}{N}}$, then we get

$$\mathbb{E}[g(\bar{x}_N)] - \min_{x\in\Delta_d\times\mathbb{R}} \mathbb{E}[v(\xi, -\langle u, X\rangle)] \leq V(x_1, \bar{x}) M_\star \sqrt{\frac{2}{N}}.$$