

# Antithetic variates in higher dimensions

Third Conference on Numerical Methods in Finance

**Joan Andreu Lázaro Camí**    Sebastian del Baño Rollin



Centre de Recerca Matemàtica

Marne-la-Vallée. 15-17 April 2009.

# Introduction

Suppose we want to estimate

$$m = \mathbb{E}[f(\xi)]$$

where

- $\xi \sim \mathcal{N}(0, \text{Id}_N)$  is an  $N$ -dimensional standard Gaussian vector and
- $f : \mathbb{R}^N \rightarrow \mathbb{R}$  is a measurable function (i.e., the payoff of a derivative contract),

by means of a Monte Carlo. That is,

$$m \approx \frac{1}{n} \sum_{i=1}^n f(\xi_i), \quad n \text{ large enough,}$$

where  $\xi_1, \dots, \xi_n$  are independent simulations of the random variable  $\xi$ .

# Introduction

Suppose we want to estimate

$$m = \mathbb{E}[f(\xi)]$$

where

- $\xi \sim \mathcal{N}(0, \text{Id}_N)$  is an  $N$ -dimensional standard Gaussian vector and
- $f : \mathbb{R}^N \rightarrow \mathbb{R}$  is a measurable function (i.e., the payoff of a derivative contract),

by means of a Monte Carlo. That is,

$$m \approx \frac{1}{n} \sum_{i=1}^n f(\xi_i), \quad n \text{ large enough,}$$

where  $\xi_1, \dots, \xi_n$  are independent simulations of the random variable  $\xi$ .

- The accuracy of the method is proportional to

$$\sigma / \sqrt{n}, \quad \sigma^2 = \text{Var}[f(\xi)].$$

- Antithetic variates appear for the first time in the seminal work of Hammersley, Morton, and Mauldon (1956):

- Antithetic variates appear for the first time in the seminal work of Hammersley, Morton, and Mauldon (1956):
  - The idea is to recycle the simulations of  $\zeta$  as samples of  $-\zeta$ .

- Antithetic variates appear for the first time in the seminal work of Hammersley, Morton, and Mauldon (1956):
  - The idea is to recycle the simulations of  $\xi$  as samples of  $-\xi$ .
  - Then, we approximate  $m$  by

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{f(\xi_i) + f(-\xi_i)}{2} \right).$$

- Antithetic variates appear for the first time in the seminal work of Hammersley, Morton, and Mauldon (1956):
  - The idea is to recycle the simulations of  $\xi$  as samples of  $-\xi$ .
  - Then, we approximate  $m$  by

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{f(\xi_i) + f(-\xi_i)}{2} \right).$$

- The trivial equality

$$\text{Var} \left[ \frac{f(\xi) + f(-\xi)}{2} \right] = \frac{1}{2} (\text{Var} [f(\xi)] + \text{Cov} (f(\xi), f(-\xi)))$$

shows that if the dependence between  $f(\xi)$  and  $f(-\xi)$  is such that  $\text{Cov} (f(\xi), f(-\xi)) < 0$ , then *antithetics* improve the accuracy of the crude Monte Carlo.

- **Antithetic variates in higher dimensions:** If  $A \in O(N)$  is an orthogonal matrix, then  $A\xi \sim \mathcal{N}(0, \text{Id}_N)$ . Therefore, we can estimate  $m$  as

$$m \approx \frac{1}{n} \sum_{i=1}^n \frac{f(\xi_i) + f(A\xi_i)}{2}, \quad n \text{ large enough.}$$

- **Antithetic variates in higher dimensions:** If  $A \in O(N)$  is an orthogonal matrix, then  $A\xi \sim \mathcal{N}(0, \text{Id}_N)$ . Therefore, we can estimate  $m$  as

$$m \approx \frac{1}{n} \sum_{i=1}^n \frac{f(\xi_i) + f(A\xi_i)}{2}, \quad n \text{ large enough.}$$

- Now, the error or the Monte Carlo estimation is given by

$$\text{Var} \left[ \frac{f(\xi) + f(A\xi)}{2} \right] = \frac{1}{2} (\text{Var} [f(\xi)] + \text{Cov} (f(\xi), f(A\xi))), \quad \text{so}$$

- **Antithetic variates in higher dimensions:** If  $A \in O(N)$  is an orthogonal matrix, then  $A\xi \sim \mathcal{N}(0, \text{Id}_N)$ . Therefore, we can estimate  $m$  as

$$m \approx \frac{1}{n} \sum_{i=1}^n \frac{f(\xi_i) + f(A\xi_i)}{2}, \quad n \text{ large enough.}$$

- Now, the error or the Monte Carlo estimation is given by

$$\text{Var} \left[ \frac{f(\xi) + f(A\xi)}{2} \right] = \frac{1}{2} (\text{Var} [f(\xi)] + \text{Cov} (f(\xi), f(A\xi))), \quad \text{so}$$

- **Problem:** Find  $A^* \in O(N)$  which (globally) minimises

$$\min_{A \in O(N)} \text{Cov} (f(\xi), f(A\xi)).$$

$A^*$  will be called the **optimal antithetic** or simply the **antithetic** for  $f$ .

- Given a function  $f : \mathbb{R}^N \rightarrow \mathbb{R}$ , we present an iterative annealing algorithm that finds the optimal antithetic for  $f$ .
- Our work aims at shedding some light in the use of antithetics in the industry, where practitioners frequently apply antithetics in the original Hammersley sense.
- We will discuss a couple of elementary examples where the antithetic matrix optimises the calculation of  $\mathbb{E}[f(\xi)]$  in a Monte Carlo simulation.
- Preliminary version available at



del Baño, S. and Lázaro-Camí, J.-A. Antithetic variates in higher dimensions, [math NA/0902.4211](#)

# Splitting the problem

- First step (simplified problem): we review some of the existing methods to find the absolute minima of a function of the form

$$\begin{aligned}\bar{U} : \mathbb{R}^r &\longrightarrow \mathbb{R} \\ x &\longmapsto \mathbb{E}[U(x, \xi)],\end{aligned}$$

where  $\xi \sim \mathcal{N}(0, \text{Id}_N)$  is a Gaussian random vector and  $U : \mathbb{R}^r \times \mathbb{R}^N \rightarrow \mathbb{R}$  is a measurable function continuously differentiable with respect to the first  $r$  entries.

# Splitting the problem

- First step (simplified problem): we review some of the existing methods to find the absolute minima of a function of the form

$$\begin{aligned}\bar{U} : \mathbb{R}^r &\longrightarrow \mathbb{R} \\ x &\longmapsto \mathbb{E}[U(x, \xi)],\end{aligned}$$

where  $\xi \sim \mathcal{N}(0, \text{Id}_N)$  is a Gaussian random vector and  $U : \mathbb{R}^r \times \mathbb{R}^N \rightarrow \mathbb{R}$  is a measurable function continuously differentiable with respect to the first  $r$  entries.

- Second step: we take care of the fact that  $A \in O(N)$ .

# First step: Gradient method

## Problem

We want to (globally) minimise  $\bar{U}(x) = \mathbb{E}[U(x, \xi)]$ ,  $x \in \mathbb{R}^r$ .

Assuming differentiability, the minimum  $x^* \in \mathbb{R}^r$  satisfies

$$\nabla \bar{U}(x^*) = 0.$$

We might propose to find  $x^*$  as the limit of

$$x_n = x_{n-1} - h \nabla \bar{U}(x_{n-1}), \quad h > 0 \text{ small increment.}$$

# First step: Gradient method

## Problem

We want to (globally) minimise  $\bar{U}(x) = \mathbb{E}[U(x, \xi)]$ ,  $x \in \mathbb{R}^r$ .

Assuming differentiability, the minimum  $x^* \in \mathbb{R}^r$  satisfies

$$\nabla \bar{U}(x^*) = 0.$$

We might propose to find  $x^*$  as the limit of

$$x_n = x_{n-1} - h \nabla \bar{U}(x_{n-1}), \quad h > 0 \text{ small increment.}$$

Drawbacks:

- 1 The algorithm will probably converge to a local minimum.

# First step: Gradient method

## Problem

We want to (globally) minimise  $\bar{U}(x) = \mathbb{E}[U(x, \xi)]$ ,  $x \in \mathbb{R}^r$ .

Assuming differentiability, the minimum  $x^* \in \mathbb{R}^r$  satisfies

$$\nabla \bar{U}(x^*) = 0.$$

We might propose to find  $x^*$  as the limit of

$$x_n = x_{n-1} - h \nabla \bar{U}(x_{n-1}), \quad h > 0 \text{ small increment.}$$

Drawbacks:

- 1 The algorithm will probably converge to a local minimum.
- 2 At each step, we need to calculate  $\nabla \bar{U}(x_{n-1}) = \mathbb{E}[\nabla U(x_{n-1}, \xi)]$  through a Monte Carlo. Too costly. This can be solved by using Robbins-Monro algorithms.

# First step: Robbins-Monro algorithms

- RM algorithms can be generally written as

$$x_n = x_{n-1} - \gamma_n \nabla U(x_{n-1}, \xi_n),$$

where  $\{\gamma_n\}_{n \in \mathbb{N}}$  is a non-negative sequence of real numbers and  $\{\xi_n\}_{n \in \mathbb{N}}$  are independent Gaussian vectors.

# First step: Robbins-Monro algorithms

- RM algorithms can be generally written as

$$x_n = x_{n-1} - \gamma_n \nabla U(x_{n-1}, \xi_n),$$

where  $\{\gamma_n\}_{n \in \mathbb{N}}$  is a non-negative sequence of real numbers and  $\{\xi_n\}_{n \in \mathbb{N}}$  are independent Gaussian vectors.

- Let  $F(x, \xi) = -\nabla U(x, \xi)$  and  $\bar{F}(x) = -\nabla \bar{U}(x)$ .

# First step: Robbins-Monro algorithms

- RM algorithms can be generally written as

$$x_n = x_{n-1} - \gamma_n \nabla U(x_{n-1}, \xi_n),$$

where  $\{\gamma_n\}_{n \in \mathbb{N}}$  is a non-negative sequence of real numbers and  $\{\xi_n\}_{n \in \mathbb{N}}$  are independent Gaussian vectors.

- Let  $F(x, \xi) = -\nabla U(x, \xi)$  and  $\bar{F}(x) = -\nabla \bar{U}(x)$ .

# First step: Robbins-Monro algorithms

- RM algorithms can be generally written as

$$x_n = x_{n-1} - \gamma_n \nabla U(x_{n-1}, \xi_n),$$

where  $\{\gamma_n\}_{n \in \mathbb{N}}$  is a non-negative sequence of real numbers and  $\{\xi_n\}_{n \in \mathbb{N}}$  are independent Gaussian vectors.

- Let  $F(x, \xi) = -\nabla U(x, \xi)$  and  $\bar{F}(x) = -\nabla \bar{U}(x)$ . RM algorithms converge to  $x^*$  **almost surely** under the following hypotheses:
  - $\langle x - x^*, \bar{F}(x) \rangle < 0$  for any  $x \in \mathbb{R}^r$ ,

# First step: Robbins-Monro algorithms

- RM algorithms can be generally written as

$$x_n = x_{n-1} - \gamma_n \nabla U(x_{n-1}, \xi_n),$$

where  $\{\gamma_n\}_{n \in \mathbb{N}}$  is a non-negative sequence of real numbers and  $\{\xi_n\}_{n \in \mathbb{N}}$  are independent Gaussian vectors.

- Let  $F(x, \xi) = -\nabla U(x, \xi)$  and  $\bar{F}(x) = -\nabla \bar{U}(x)$ . RM algorithms converge to  $x^*$  **almost surely** under the following hypotheses:
  - 1  $\langle x - x^*, \bar{F}(x) \rangle < 0$  for any  $x \in \mathbb{R}^r$ ,
  - 2  $\sum_{n \in \mathbb{N}} \gamma_n = \infty$  and  $\sum_{n \in \mathbb{N}} \gamma_n^2 < \infty$  (for example,  $\gamma_n = \frac{b}{n}$ ,  $b > 0$ ),

# First step: Robbins-Monro algorithms

- RM algorithms can be generally written as

$$x_n = x_{n-1} - \gamma_n \nabla U(x_{n-1}, \xi_n),$$

where  $\{\gamma_n\}_{n \in \mathbb{N}}$  is a non-negative sequence of real numbers and  $\{\xi_n\}_{n \in \mathbb{N}}$  are independent Gaussian vectors.

- Let  $F(x, \xi) = -\nabla U(x, \xi)$  and  $\bar{F}(x) = -\nabla \bar{U}(x)$ . RM algorithms converge to  $x^*$  **almost surely** under the following hypotheses:
  - 1  $\langle x - x^*, \bar{F}(x) \rangle < 0$  for any  $x \in \mathbb{R}^r$ ,
  - 2  $\sum_{n \in \mathbb{N}} \gamma_n = \infty$  and  $\sum_{n \in \mathbb{N}} \gamma_n^2 < \infty$  (for example,  $\gamma_n = \frac{b}{n}$ ,  $b > 0$ ),
  - 3  $\mathbb{E} \left[ \|F(x_{n-1}, \xi_n)\|^2 \mid \mathcal{F}_{n-1} \right] \leq K \left( 1 + \|\xi_n\|^2 \right)$  a.s. for some constant  $K > 0$  where  $\mathcal{F}_{n-1}$  is the  $\sigma$ -algebra generated by  $\{\xi_k \mid k \leq n-1\}$ .

# First step: Robbins-Monro algorithms

- RM algorithms can be generally written as

$$x_n = x_{n-1} - \gamma_n \nabla U(x_{n-1}, \xi_n),$$

where  $\{\gamma_n\}_{n \in \mathbb{N}}$  is a non-negative sequence of real numbers and  $\{\xi_n\}_{n \in \mathbb{N}}$  are independent Gaussian vectors.

- Let  $F(x, \xi) = -\nabla U(x, \xi)$  and  $\bar{F}(x) = -\nabla \bar{U}(x)$ . RM algorithms converge to  $x^*$  **almost surely** under the following hypotheses:
  - 1  $\langle x - x^*, \bar{F}(x) \rangle < 0$  for any  $x \in \mathbb{R}^r$ ,
  - 2  $\sum_{n \in \mathbb{N}} \gamma_n = \infty$  and  $\sum_{n \in \mathbb{N}} \gamma_n^2 < \infty$  (for example,  $\gamma_n = \frac{b}{n}$ ,  $b > 0$ ),
  - 3  $\mathbb{E} \left[ \|F(x_{n-1}, \xi_n)\|^2 \mid \mathcal{F}_{n-1} \right] \leq K \left( 1 + \|\xi_n\|^2 \right)$  a.s. for some constant  $K > 0$  where  $\mathcal{F}_{n-1}$  is the  $\sigma$ -algebra generated by  $\{\xi_k \mid k \leq n-1\}$ .
- Drawbacks:

# First step: Robbins-Monro algorithms

- RM algorithms can be generally written as

$$x_n = x_{n-1} - \gamma_n \nabla U(x_{n-1}, \xi_n),$$

where  $\{\gamma_n\}_{n \in \mathbb{N}}$  is a non-negative sequence of real numbers and  $\{\xi_n\}_{n \in \mathbb{N}}$  are independent Gaussian vectors.

- Let  $F(x, \xi) = -\nabla U(x, \xi)$  and  $\bar{F}(x) = -\nabla \bar{U}(x)$ . RM algorithms converge to  $x^*$  **almost surely** under the following hypotheses:

- 1  $\langle x - x^*, \bar{F}(x) \rangle < 0$  for any  $x \in \mathbb{R}^r$ ,
- 2  $\sum_{n \in \mathbb{N}} \gamma_n = \infty$  and  $\sum_{n \in \mathbb{N}} \gamma_n^2 < \infty$  (for example,  $\gamma_n = \frac{b}{n}$ ,  $b > 0$ ),
- 3  $\mathbb{E} \left[ \|F(x_{n-1}, \xi_n)\|^2 \mid \mathcal{F}_{n-1} \right] \leq K \left( 1 + \|\xi_n\|^2 \right)$  a.s. for some constant  $K > 0$  where  $\mathcal{F}_{n-1}$  is the  $\sigma$ -algebra generated by  $\{\xi_k \mid k \leq n-1\}$ .

- Drawbacks:

- 1  $x^*$  has to be the unique minimum of  $\bar{U}(x) = \mathbb{E}[U(x, \xi)]$ .

# First step: Robbins-Monro algorithms

- RM algorithms can be generally written as

$$x_n = x_{n-1} - \gamma_n \nabla U(x_{n-1}, \xi_n),$$

where  $\{\gamma_n\}_{n \in \mathbb{N}}$  is a non-negative sequence of real numbers and  $\{\xi_n\}_{n \in \mathbb{N}}$  are independent Gaussian vectors.

- Let  $F(x, \xi) = -\nabla U(x, \xi)$  and  $\bar{F}(x) = -\nabla \bar{U}(x)$ . RM algorithms converge to  $x^*$  **almost surely** under the following hypotheses:

- 1  $\langle x - x^*, \bar{F}(x) \rangle < 0$  for any  $x \in \mathbb{R}^r$ ,
- 2  $\sum_{n \in \mathbb{N}} \gamma_n = \infty$  and  $\sum_{n \in \mathbb{N}} \gamma_n^2 < \infty$  (for example,  $\gamma_n = \frac{b}{n}$ ,  $b > 0$ ),
- 3  $\mathbb{E} \left[ \|F(x_{n-1}, \xi_n)\|^2 \mid \mathcal{F}_{n-1} \right] \leq K \left( 1 + \|\xi_n\|^2 \right)$  a.s. for some constant  $K > 0$  where  $\mathcal{F}_{n-1}$  is the  $\sigma$ -algebra generated by  $\{\xi_k \mid k \leq n-1\}$ .

- Drawbacks:

- 1  $x^*$  has to be the unique minimum of  $\bar{U}(x) = \mathbb{E}[U(x, \xi)]$ .
- 2 If  $\bar{U}(x)$  has several minima, Robbins-Monro algorithms may converge to a local minimum.

# First step: simulated annealing algorithms

General form:

$$x_n = x_{n-1} - \gamma_n \nabla_x U(x_{n-1}, \tilde{\xi}_n) + \sqrt{\gamma_n h_n} \zeta_n,$$

where

- 1  $\{\tilde{\xi}_n\}_{n \in \mathbb{N}}$ ,  $\{\zeta_n\}_{n \in \mathbb{N}}$  are two independent sequences of i.i.d Gaussian random vectors.
- 2  $\gamma_n = 1/n^\alpha$  and  $h_n = d/(1-\alpha) \ln n$ , with  $0 < \alpha < 1$  and  $d$  large enough. Other possibilities for  $\{\gamma_n\}_{n \in \mathbb{N}}$  and  $\{h_n\}_{n \in \mathbb{N}}$  exist.

# First step: simulated annealing algorithms

General form:

$$x_n = x_{n-1} - \gamma_n \nabla_x U(x_{n-1}, \tilde{\xi}_n) + \sqrt{\gamma_n h_n} \zeta_n,$$

where

- 1  $\{\tilde{\xi}_n\}_{n \in \mathbb{N}}$ ,  $\{\zeta_n\}_{n \in \mathbb{N}}$  are two independent sequences of i.i.d Gaussian random vectors.
- 2  $\gamma_n = 1/n^\alpha$  and  $h_n = d/(1-\alpha) \ln n$ , with  $0 < \alpha < 1$  and  $d$  large enough. Other possibilities for  $\{\gamma_n\}_{n \in \mathbb{N}}$  and  $\{h_n\}_{n \in \mathbb{N}}$  exist.



Fang, H., Gong, G., and Quian, M. Annealing of iterative stochastic schemes. *SIAM J. Control Optim.* **35** (6), pp. 1886–1907, 1997.

## Theorem

Let  $S = \{x \in \mathbb{R}^r \mid \bar{U}(x) = \min_{y \in \mathbb{R}^r} \bar{U}(y)\}$  and

$$x_n = x_{n-1} - \gamma_n \nabla_x U(x_{n-1}, \xi_n) + \sqrt{\gamma_n h_n} \zeta_n.$$

Then, for any  $l > 0$ ,

$$P(\{x_n \in B_l(S)\}) \rightarrow 1 \text{ as } n \rightarrow \infty$$

uniformly for any initial condition  $x_0$  in an arbitrary compact set provided that

# First step: simulated annealing algorithms

## Theorem

Let  $S = \{x \in \mathbb{R}^r \mid \bar{U}(x) = \min_{y \in \mathbb{R}^r} \bar{U}(y)\}$  and

$$x_n = x_{n-1} - \gamma_n \nabla_x U(x_{n-1}, \xi_n) + \sqrt{\gamma_n h_n} \zeta_n.$$

Then, for any  $l > 0$ ,

$$P(\{x_n \in B_l(S)\}) \rightarrow 1 \text{ as } n \rightarrow \infty$$

uniformly for any initial condition  $x_0$  in an arbitrary compact set provided that

- 1  $\limsup_{\|x\| \rightarrow \infty} \frac{\|\bar{F}(x)\|}{\|x\|} \leq M_1 < \infty, M_1 > 0,$
- 2  $\limsup_{\|x\| \rightarrow \infty} \frac{\langle \bar{F}(x), x \rangle}{\|x\|^2} \leq -c < 0, c > 0, \text{ and}$
- 3  $\limsup_{\|x\| \rightarrow \infty} \frac{E[(F(x, \xi) - \bar{F}(x))^2]}{\|x\|^2} \leq M_2 < \infty, M_2 > 0.$

## Second step: general facts about $O(N)$

Let  $O(N) = \{A \in GL(N, \mathbb{R}) \mid AA^T = \text{Id}\}$  denote the orthogonal Lie group and  $\mathfrak{so}(N) = \{Y \in \mathfrak{gl}(N, \mathbb{R}) \mid Y = -Y^T\}$  its Lie algebra.

### Facts

1  $O(N) = SO(N) \cup SO(N)^-$ .

## Second step: general facts about $O(N)$

Let  $O(N) = \{A \in GL(N, \mathbb{R}) \mid AA^T = \text{Id}\}$  denote the orthogonal Lie group and  $\mathfrak{so}(N) = \{Y \in \mathfrak{gl}(N, \mathbb{R}) \mid Y = -Y^T\}$  its Lie algebra.

### Facts

- 1  $O(N) = SO(N) \cup SO(N)^-$ .
- 2  $\mathfrak{so}(N) \cong \mathbb{R}^{N(N-1)/2}$  as *Euclidean spaces with the product*

$$\langle X, Y \rangle = \frac{1}{2} \text{trace} (XY^T).$$

## Second step: general facts about $O(N)$

Let  $O(N) = \{A \in GL(N, \mathbb{R}) \mid AA^T = \text{Id}\}$  denote the orthogonal Lie group and  $\mathfrak{so}(N) = \{Y \in \mathfrak{gl}(N, \mathbb{R}) \mid Y = -Y^T\}$  its Lie algebra.

### Facts

- 1  $O(N) = SO(N) \cup SO(N)^-$ .
- 2  $\mathfrak{so}(N) \cong \mathbb{R}^{N(N-1)/2}$  as Euclidean spaces with the product

$$\langle X, Y \rangle = \frac{1}{2} \text{trace} \left( XY^T \right).$$

- 3 Let  $E_{ij} \in \mathfrak{so}(N)$  be the matrix with  $+1$  in the  $(i, j)$ th position,  $-1$  in the  $(j, i)$ th position and zero elsewhere. Then  $\{E_{ij}\}_{0 < i < j \leq N}$  is an orthonormal basis of  $\mathfrak{so}(N)$ .

## Second step: general facts about $O(N)$

Let  $O(N) = \{A \in GL(N, \mathbb{R}) \mid AA^T = \text{Id}\}$  denote the orthogonal Lie group and  $\mathfrak{so}(N) = \{Y \in \mathfrak{gl}(N, \mathbb{R}) \mid Y = -Y^T\}$  its Lie algebra.

### Facts

- 1  $O(N) = SO(N) \cup SO(N)^-$ .
- 2  $\mathfrak{so}(N) \cong \mathbb{R}^{N(N-1)/2}$  as Euclidean spaces with the product

$$\langle X, Y \rangle = \frac{1}{2} \text{trace} (XY^T).$$

- 3 Let  $E_{ij} \in \mathfrak{so}(N)$  be the matrix with  $+1$  in the  $(i, j)$ th position,  $-1$  in the  $(j, i)$ th position and zero elsewhere. Then  $\{E_{ij}\}_{0 < i < j \leq N}$  is an orthonormal basis of  $\mathfrak{so}(N)$ .
- 4 The compact ball  $B(0, \pi\sqrt{\lfloor N/2 \rfloor}) \subset \mathfrak{so}(N)$  surjects onto  $SO(N)$  by the exponential map  $\exp : \mathfrak{so}(N) \rightarrow SO(N)$ .

## Second step: computing gradients on $O(N)$

- Given a smooth function  $F : G \rightarrow \mathbb{R}$ ,  $TF : TG \rightarrow T\mathbb{R}$  denotes its tangent map,

$$\begin{aligned} TF : T_g G &\longrightarrow T_{F(g)} \mathbb{R} = \mathbb{R} \\ Y &\longmapsto T_g F(Y). \end{aligned}$$

## Second step: computing gradients on $O(N)$

- Given a smooth function  $F : G \rightarrow \mathbb{R}$ ,  $TF : TG \rightarrow T\mathbb{R}$  denotes its tangent map,

$$\begin{aligned} TF : T_g G &\longrightarrow T_{F(g)} \mathbb{R} = \mathbb{R} \\ Y &\longmapsto T_g F(Y). \end{aligned}$$

- We use  $\mathfrak{g}$  to coordinate  $G$  and consider a (local) diffeomorphism  $\varphi : \mathfrak{g} \rightarrow G$  around  $0 \in \mathfrak{g}$ . Then,

$$\nabla (F \circ \varphi)(Y) = \left( \sum_{i < j} T_{\varphi(Y)} F \circ T_Y \varphi(E_{ij}) \right) E_{ij} \in \mathfrak{so}(N).$$

## Second step: computing gradients on $O(N)$

- Given a smooth function  $F : G \rightarrow \mathbb{R}$ ,  $TF : TG \rightarrow T\mathbb{R}$  denotes its tangent map,

$$\begin{aligned} TF : T_g G &\longrightarrow T_{F(g)} \mathbb{R} = \mathbb{R} \\ Y &\longmapsto T_g F(Y). \end{aligned}$$

- We use  $\mathfrak{g}$  to coordinate  $G$  and consider a (local) diffeomorphism  $\varphi : \mathfrak{g} \rightarrow G$  around  $0 \in \mathfrak{g}$ . Then,

$$\nabla (F \circ \varphi)(Y) = \left( \sum_{i < j} T_{\varphi(Y)} F \circ T_Y \varphi(E_{ij}) \right) E_{ij} \in \mathfrak{so}(N).$$

- If  $\varphi : \mathfrak{g} \rightarrow G$  equals  $\varphi(Y) := \exp(Y) g_0$  for some  $g_0 \in G$ , then

$$\begin{aligned} T_Y \varphi(X) &= \frac{\exp(\operatorname{ad}_Y) - \operatorname{Id}}{\operatorname{ad}_Y}(X) \varphi(Y) \\ &= \left( \sum_{j \geq 0} \frac{1}{(j+1)!} \operatorname{ad}_Y \circ \dots \circ \operatorname{ad}_Y(X) \right) \varphi(Y). \end{aligned}$$

- Recall that we want to minimise

$$\min_{A \in O(N)} \text{Cov}(f(\xi), f(A\xi)) = \min_{A \in O(N)} \mathbb{E}[f(\xi)f(A\xi)].$$

- We will apply *Step 1 + Step 2* to

$$\begin{aligned} U : \mathfrak{so}(N) \times \mathbb{R}^N &\longrightarrow \mathbb{R} \\ (Y, x) &\longmapsto f(x)f(\exp(Y)A_0x) \end{aligned}$$

for some  $A_0 \in O(N)$ .

- **Problem:** If  $\bar{F} = -\nabla \mathbb{E}[f(\xi) f(\exp(\cdot)A_0\xi)]$  denotes (minus) the gradient of  $\bar{U} \circ \exp$  then, since  $\bar{U}$  is bounded,

$$\lim_{\|Y\| \rightarrow \infty} \frac{\langle \bar{F}(Y), Y \rangle}{\|Y\|^2} = 0,$$

which is not strictly negative as simulated annealing algorithms require.

- **Problem:** If  $\bar{F} = -\nabla \mathbb{E}[f(\xi) f(\exp(\cdot)A_0\xi)]$  denotes (minus) the gradient of  $\bar{U} \circ \exp$  then, since  $\bar{U}$  is bounded,

$$\lim_{\|Y\| \rightarrow \infty} \frac{\langle \bar{F}(Y), Y \rangle}{\|Y\|^2} = 0,$$

which is not strictly negative as simulated annealing algorithms require.

- **Solution:** We add to  $U(Y, x)$  the penalty function

$$P(Y) = \|Y\|^2 \cdot \mathbf{1}_{B(0, \pi\sqrt{\lfloor N/2 \rfloor})^c}(Y), \quad Y \in \mathfrak{so}(N)$$

which satisfies all the hypothesis **at infinity**.

# Antithetic variates revisited

The algorithm

$$Z_n = \nabla (U_{\xi_n} \circ \varphi) (Y_{n-1}) = \sum_{i < j} (f(\xi_n) \nabla f^T |_{A_{n-1} \xi_n} T_{Y_{n-1}} \exp^{sc}(E_{ij}) A_{n-1} \xi_n) E_{ij},$$

$$Y_n = Y_{n-1} - \frac{1}{\sqrt{n}} \left( Z_n + 2Y_{n-1} \cdot \mathbf{1}_{B(0, \pi \sqrt{\lfloor N/2 \rfloor})^c} (Y_{n-1}) \right) + \sqrt{\frac{2d}{n \ln n}} \zeta_n,$$

$$A_n = \exp(Y_n) A_0,$$

converges in probability to the optimal antithetic  $A^*$  (or the antithetic locus if there are multiple minima) in the connected component of  $SO(N)$  containing  $A_0$ . Here

- 1  $\{\xi_n\}_{n \in \mathbb{N}}$  and  $\{\zeta_n\}_{n \in \mathbb{N}}$  are  $N$ -dimensional and  $N(N-1)/2$ -dimensional sequences of i.i.d standard Gaussian vectors respectively.
- 2  $d > 4 \text{Var}[f(\xi)]$ .

## Example: Asian option

- Let  $S$  be an asset in the Black-Scholes model and consider an Asian option with payoff

$$\text{Max} \left( 0, \frac{1}{N} \sum_{i=1}^N S_{T_i} - K \right).$$

## Example: Asian option

- Let  $S$  be an asset in the Black-Scholes model and consider an Asian option with payoff

$$\text{Max} \left( 0, \frac{1}{N} \sum_{i=1}^N S_{T_i} - K \right).$$

- Data

$$S_0 = 100 \quad r = 2.83\% \text{ (risk free rate)}$$

$$K = 100 \quad \sigma = 10.36\% \text{ (volatility)}$$

$$N = 12 \quad T_i = i \frac{T}{12}, \quad T = 1 \text{ year (expiry time)}$$

# Example: Asian option

- Let  $S$  be an asset in the Black-Scholes model and consider an Asian option with payoff

$$\text{Max} \left( 0, \frac{1}{N} \sum_{i=1}^N S_{T_i} - K \right).$$

- Data

$$S_0 = 100 \quad r = 2.83\% \text{ (risk free rate)}$$

$$K = 100 \quad \sigma = 10.36\% \text{ (volatility)}$$

$$N = 12 \quad T_i = i \frac{T}{12}, \quad T = 1 \text{ year (expiry time)}$$

- Results

	Price	Variance	Time (s)	MC Error ( $\sigma/\sqrt{n}$ )
<b>Crude MC</b>	3.15	17.25	8.75	0.01
<b>Optimal Antithetic</b>	3.15	3.60	3.54	0.01
<b>Antithetic <math>A = -\text{Id}</math></b>	3.16	3.50	3.41	0.01

## Example: Option on a Covariance Swap

- Let  $S_1, S_2$  be two independent assets in the Black-Scholes model and consider an option with payoff

$$100 \text{Max} \left( 0, \frac{1}{N} \sum_{i=1}^N \left( \frac{(S_1)_{T_i}}{(S_1)_{T_{i-1}}} - 1 \right) \left( \frac{(S_2)_{T_i}}{(S_2)_{T_{i-1}}} - 1 \right) \right).$$

## Example: Option on a Covariance Swap

- Let  $S_1, S_2$  be two independent assets in the Black-Scholes model and consider an option with payoff

$$100 \text{Max} \left( 0, \frac{1}{N} \sum_{i=1}^N \left( \frac{(S_1)_{T_i}}{(S_1)_{T_{i-1}}} - 1 \right) \left( \frac{(S_2)_{T_i}}{(S_2)_{T_{i-1}}} - 1 \right) \right).$$

- Data

$$(S_1)_{T=0} = 100 \quad \sigma_1 = 17.36\%$$

$$(S_2)_{T=0} = 100 \quad \sigma_2 = 10.12\%$$

$$r = 2.83\% \text{ (risk free rate)} \quad d_2 = 2.03\% \text{ (dividends asset } S_2)$$

$$N = 6 \quad T_i = i \frac{T}{6}, \quad T = 1 \text{ year (expiry time)}$$

# Example: Option on a Covariance Swap

- Let  $S_1, S_2$  be two independent assets in the Black-Scholes model and consider an option with payoff

$$100 \text{Max} \left( 0, \frac{1}{N} \sum_{i=1}^N \left( \frac{(S_1)_{T_i}}{(S_1)_{T_{i-1}}} - 1 \right) \left( \frac{(S_2)_{T_i}}{(S_2)_{T_{i-1}}} - 1 \right) \right).$$

- Data

$$(S_1)_{T=0} = 100 \quad \sigma_1 = 17.36\%$$

$$(S_2)_{T=0} = 100 \quad \sigma_2 = 10.12\%$$

$$r = 2.83\% \text{ (risk free rate)} \quad d_2 = 2.03\% \text{ (dividends asset } S_2)$$

$$N = 6 \quad T_i = i \frac{T}{6}, \quad T = 1 \text{ year (expiry time)}$$

- Results

	Price	Variance	Time (s)	MC Error ( $\sigma/\sqrt{n}$ )
<b>Crude MC</b>	0.276	$17.77 \cdot 10^{-2}$	9.97	0.001
<b>Optimal Antithetic</b>	0.276	$6.27 \cdot 10^{-2}$	6.72	0.001
<b>Antithetic A = - Id</b>	0.276	$17.77 \cdot 10^{-2}$	18.76	0.001

- Improve the efficiency (speed) of the annealing algorithm.

# Future work

- Improve the efficiency (speed) of the annealing algorithm.
- Study other products under more realistic models: local volatility, etc.

- Improve the efficiency (speed) of the annealing algorithm.
- Study other products under more realistic models: local volatility, etc.
- Study the viability of using antithetics dynamically,








$$\mathbb{E}[f(\xi)] \approx \frac{1}{n} \sum_{k=1}^n \frac{f(\xi_k) + f(A_{k-1}\xi_k)}{2}$$

instead of

$$\mathbb{E}[f(\xi)] \approx \frac{1}{n} \sum_{k=1}^n \frac{f(\xi_k) + f(A^*\xi_k)}{2},$$

where  $A_k \rightarrow A^*$  in probability.

# Some references

-  Arouna, B. Robbins-Monro algorithms and variance reduction in Finance. *The Journal of Computational Finance* **7** (2), pp. 1245-1255, 2003.
-  Arouna, B. Adaptive Monte Carlo method, a variance reduction technique. *Monte Carlo Methods and Appl.* **10** (1), pp. 1-24, 2004.
-  Duflo, M. *Random Iterative Models*. Springer, 1997.
-  Engø, K. On the construction of geometric integrators in the RKMK class. Technical Report N<sup>o</sup> **158**, Dept Comp. Sc., University of Bergen, 1998.
-  Fang, H., Gong, G., and Quian, M. Annealing of iterative stochastic schemes. *SIAM J. Control Optim.* **35** (6), pp. 1886–1907, 1997.
-  Hammersley, J. M. and Mauldon, J. G. General principles of antithetic variates. *Proc. Cambridge Philosophical Society* **52**, pp. 476-481, 1956.
-  Hammersley, J. M. and Morton, K. W. A new Monte Carlo technique: antithetic variates. *Proc. Cambridge Philosophical Society* **52**, pp. 449-475, 1956.