Global sensitivity analysis for stochastic optimization

October 2020

1 Introduction

Let us consider the following optimization under uncertainty problem:

$$\min_{x \in \mathbb{R}^n} F(x) := \int_\Omega f(x, \xi) \mu(d\xi) \tag{1}$$

Here ξ denotes realizations of a d-dimensional random vector that has distribution law μ.

We consider the mapping

$$\mu \mapsto F(x^*(\mu)),$$

and ask the "question"

> "How sensitive is $F(x^*(\mu))$ to μ"

Here x^* denotes the solution of the optimization problem.

The problem can be made more concrete by (i) focusing on a specific set of distribution laws for ξ; and, more importantly, (ii) specifying what is meant by "sensitivity" in the "question" above.

Let's consider the simpler first task: we may assume μ is a multivariate Gaussian: $\mu = N(m, C)$, where m and C are the mean and covariance matrix, respectively. Then, we can consider the mapping

$$\Phi(m, C) = F(x^*(m, C)),$$

and its sensitivity to (some) entries of m and C.

1.1 A more general problem statement

Let the law of ξ have density $\pi = \pi(\xi;\zeta)$, where ζ is a vector of parameters characterizing π. For a given ζ, we consider the objective function

$$F(x;\zeta) = \int f(x,\xi)\pi(\xi;\zeta).$$

A minimizer x^* of F will be a function of ζ, $x^* = x^*(\zeta)$. We can then consider the sensitivity of x^* to the components of ζ.

2 Stochastic Rosenbrock 2-D Function

The Rosenbrock function is commonly used as a test for optimization algorithms due to the fact that its local minima are easy to find while the global minimum is more difficult to find.

2.1 Problem Statement

Consider the Rosenbrock 2-D function $R : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$R(x) = (a - x_1)^2 + b(x_2 - x_1^2)^2$$

where a and b are constants, typically set to $a = 1$ and $b = 100$. The global minimum occurs at $x = (a, a^2)^T$. The Rosenbrock 2-D function can be turned into a stochastic function by replacing the constants a and b with random variables θ_1 and θ_2 (together denoted $\Theta = (\theta_1\, \theta_2)^T$) which are determined by a bivariate normal distribution. This bivariate normal distribution is, in turn, determined by means μ_1, μ_2, standard deviations σ_1, σ_2, and covariance ρ, all of which are uniform random variables. Denoting the covariance matrix by $C = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}$ and mean vector by $m = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, we have that

$$\Theta \sim N(m, C) \text{ and } \pi(\Theta) \sim e^{-\frac{1}{2}(\theta - m)^T C^{-1}(\theta - m)}$$

The problem of optimizing the stochastic Rosenbrock function is then expressed as solving

$$\min_{x \in \mathbb{R}^2} F(x) := \int_{\mathbb{R}^2} R(x,\Theta) e^{-\frac{1}{2}(\theta - m)^T C^{-1}(\theta - m)} d\Theta \tag{2}$$

where the objective function is defined as $F(x) = \int_{\mathbb{R}^2} R(x, \Theta)\pi(\Theta)d\Theta$ for a fixed Θ. Let x^* denote the minimizer such that $F(x^*)$ is the solution to the problem (2). If Θ is varied then the minimizer x^* may also vary. It is then natural to consider a mapping

$$(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \mapsto F(x^*) = \int_{\mathbb{R}^2} R(x^*, \Theta))e^{-\frac{1}{2}(\theta-m)^T C^{-1}(\theta-m)}d\Theta$$

Recalling that $\Theta = (\theta_1, \theta_2)^T$, we can then ask what is the sensitivity of the minimizer $x^*(\Theta)$ to θ_1 and θ_2. Since Θ is determined by a bivariate normal distribution, we can instead ask what is the sensitivity of x^* to $\mu_1, \mu_2, \sigma_1, \sigma_2$, and ρ by computing their respective Sobol' indices.

2.2 Methods

To attempt to find a solution to this problem, sample average optimization can be employed. For some N, an approximation for $F(x)$ can be made by

$$F(x) \approx \frac{1}{N}\sum_{i=1}^{N} R(x, \Theta_i)$$

where $\{\Theta_1, ..., \Theta_N\}$ is a sample taken from the bivariate normal distribution. More explicitly, the approximation is

$$F(x, y) \approx \frac{1}{N}\sum_{i=1}^{N} \left((\theta_{1,i} - x_1)^2 + \theta_{2,i}(x_2 - x_1^2)^2\right)$$

So, an approximated solution can be found by optimizing this approximated objective function. Sensitivity of x^* to the random variables $\mu_1, \mu_2, \theta_1, \theta_2, \rho$ can then be measured by computing the respective Sobol' indices.

It is expected that the random variables associated to θ_2 will not display any sensitivity since, in the deterministic Rosenbrock function, the global minimum depends only on the constant a.

2.3 Analytic Solution

We try to find an analytic solution by getting an explicit formula for

$$F(x) = \int_{\mathbb{R}^2} R(x, \Theta)e^{-\frac{1}{2}(\theta-m)^T C^{-1}(\theta-m)}d\Theta$$

First, we write $\pi(\Theta)$ as

$$\pi(\theta) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - \rho^2}} e^{\frac{2\rho(\mu_2-\theta_2)(\mu_1-\theta_1)-\sigma_1^2(\mu_2-\theta_2)^2-\sigma_2^2(\mu_1-\theta_1)^2}{2(\rho^2-\sigma_1^2\sigma_2^2)}}$$

Then first integrate with respect to θ_1

$$\int_{-\infty}^{\infty} \frac{(\theta_1 - x_1)^2 + \theta_2(x_2 - x_1^2)^2}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - \rho^2}} e^{\frac{2\rho(\mu_2-\theta_2)(\mu_1-\theta_1)-\sigma_1^2(\mu_2-\theta_2)^2-\sigma_2^2(\mu_1-\theta_1)^2}{2(\rho^2-\sigma_1^2\sigma_2^2)}} d\theta_1 =$$

$$\left(\frac{(x_2 - x_1^2)^2\theta_2 + (\mu_1 - x_1)^2 + \sigma_1^2}{\sqrt{2\pi}\sigma_2} + \frac{2\rho(\mu_2 - \theta_2)(x_1 - \mu_1)}{\sqrt{2\pi}\sigma_2^2} - \frac{\rho^2}{\sqrt{2\pi}\sigma_2^3} + \frac{\rho^2}{\sqrt{2\pi}\sigma_2^5}(\mu_2 - \sigma_2)^2 \right) e^{-\frac{1}{2}\left(\frac{\mu_2-\theta_2}{\sigma_2}\right)^2}$$

Then, we integrate this with respect to θ_2

$$\int_{-\infty}^{\infty} \left(\frac{(x_2 - x_1^2)^2\theta_2 + (\mu_1 - x_1)^2 + \sigma_1^2}{\sqrt{2\pi}\sigma_2} + \frac{2\rho(\mu_2 - \theta_2)(x_1 - \mu_1)}{\sqrt{2\pi}\sigma_2^2} - \frac{\rho^2}{\sqrt{2\pi}\sigma_2^3} + \frac{\rho^2}{\sqrt{2\pi}\sigma_2^5}(\mu_2 - \sigma_2)^2 \right) e^{-\frac{1}{2}\left(\frac{\mu_2-\theta_2}{\sigma_2}\right)^2} d\theta_2$$
$$= \mu_1(x_2 - x_1^2)^2 + (\mu_1 - x_1)^2 + \sigma_1^2$$

The objective function is then given by

$$F(x) = \mu_1(x_2 - x_1^2)^2 + (\mu_1 - x_1)^2 + \sigma_1^2$$

Then, the partial derivatives of F are given by

$$F_{x_1}(x) = -4\mu_1(x_2 - x_1^2) - 2(\mu_1 - x_1)$$
$$F_{x_2}(x) = 2\mu_1(x_2 - x_1^2)$$

Therefore, $x^* = (\mu_1, \mu_1^2)$ and $F(x^*) = \sigma_1^2$.

Alternatively:

$$f(x, \theta) = (\theta_1\theta_2 - x_1)^2 + \theta_2(x_2 - x_1^2)^2$$

Then for $C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

This gives objective function

$$F(x) = 4\mu_1\mu_2 + (\mu_1^2 + \sigma_1^2)(\mu_2^2 + \sigma_2^2) - 2(\rho\sigma_1\sigma_2 + \mu_1\mu_2)x_1 + x_1^2 + \mu_2(x_2 - x_1^2)^2$$

So $x_1^* = \sqrt{2(\rho\sigma_1\sigma_2 + \mu_1\mu_2)}$ and $x_2^* = 2(\rho\sigma_1\sigma_2 + \mu_1\mu_2)$

Beale Function

$$f(x,y) = (a - x + xy)^2 + (b - x + xy^2)^2 + (c - x + xy^3)^2$$
$$c = \frac{(a^3-(b-a)^3)}{(a(2a-b))}$$

For $a = 1.5$, $b = 2.25$, $c = 2.625$, we have that $f(x^*, y^*) = 0$ for $(x^*, y^*) = (3, 0.5)$.

Let $a = \theta_1$ with mean $\mu_1 = 1.5$, $b = \frac{\theta_1\theta_2}{1.75}$ and $c = \theta_2$ with mean $\mu_2 = 2.625$.

Then

$$F(x,y) = \sigma_1^2 + \mu_1^2 + \frac{2\rho^2 + 4\mu_1\mu_2\rho + (\mu_1^2+\sigma_1^2)(\mu_2^2+\sigma_2^2)}{1.75^2} + \sigma_2^2 + \mu_2^2 + 2\mu_1(-x+y) + \frac{2}{1.75}(\rho + \mu_1\mu_2)(-x+xy^2) + 2\mu_2(-x+xy^3) + (-x+xy)^2 + (-x+xy^2)^2 + (-x+xy^3)^2$$

Branin Function

$$f(x,y) = a(y - bx^2 + cx - r)^2 + s(1-t)\cos(x) + s$$

2.4 Modified Rosenbrock Function

Consider the modified Rosenbrock function

$$F(x) = (\theta_1 - \theta_2 x_1)^2 + \theta_3(x_2 - x_1^2)^2$$

where $x \in \mathbb{R}^3$ and $\theta \in \mathbb{R}^3$. For fixed θ, the minimum is attained at $x^* = (\xi\ \xi^2)^T$ where $\xi = \frac{\theta_1}{\theta_2}$.

Assuming θ is distributed according to a multivariate normal, $\mu = N(m, C)$ where

$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & 0 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

then the minimizer of the OUU problem is $x^* = (\xi\ \xi^2)^T$ where $\xi = \frac{\rho\sigma_1\sigma_2 + m_1 m_2}{\sigma_2^2 + m_2^2}$ and $F(x^*) = \frac{(\sigma_1^2+m_1^2)(\sigma_2^2+m_2^2) - (\rho\sigma_1\sigma_2+m_1m_2)^2}{\sigma_2^2 m_2^2}$.