Final Exam 2016 - Derivative Free Optimization (Part I)

Anne Auger
anne.auger@inria.fr

February 2016

Exercice 1: Optimization of a noisy function (5 points)

We have considered during the class the optimization of numerical **noiseless** functions $f : \mathbf{x} \in \mathbb{R}^n \to \mathbb{R}$ where for a given input $\mathbf{x}$, the outcome $f(\mathbf{x})$ is **deterministic**. We now consider the optimization of **noisy** functions where two different calls to the function f will give two different **random** outputs or in other words where $f(\mathbf{x})$ is a random variable. The optimization goal generally formulated is then to optimize the unknown expected value $E[f(\mathbf{x})]$ while we only have access to the noisy function $f(\mathbf{x})$.

We consider first the sphere function with multiplicative noise defined as

$$f_{\text{ns1}}(\mathbf{x}) = f_{\text{sphere}}(\mathbf{x})(1 + \epsilon \mathcal{N}(0,1)) \tag{1}$$

where $f_{\text{sphere}}(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{x}_i^2$ and $\epsilon > 0$ and $\mathcal{N}(0,1)$ denotes a normal distribution with mean 0 and variance 1 that is sampled anew independently for each call of f. All the functions are to be minimized.

a) Compute for a given $\mathbf{x}$ the expected value (or mean value) of $f_{\text{ns1}}(\mathbf{x})$ that we denote $E[f_{\text{ns1}}(\mathbf{x})]$.

b) What is the minimum of $E[f_{\text{ns1}}(\mathbf{x})]$?

b) Compute for a given $\mathbf{x}$, the variance of $f_{\text{ns1}}(\mathbf{x})$ that we will denote $\text{Var}(f_{\text{ns1}}(\mathbf{x}))$. Show that $\text{Var}(f_{\text{ns1}}(\mathbf{x}))$ decreases to zero when $\mathbf{x}$ approaches the optimum of $E[f_{\text{ns1}}(\mathbf{x})]$.

We consider now the sphere function with additive noise defined as

$$f_{\text{ns2}}(\mathbf{x}) = f_{\text{sphere}}(\mathbf{x}) + \epsilon \mathcal{N}(0,1) \tag{2}$$

where $f_{\text{sphere}}(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{x}_i^2$ and $\epsilon > 0$ and $\mathcal{N}(0,1)$ denotes a normal distribution with mean 0 and variance 1 that is sampled anew independently for each call of f.

d) Compute for a given $\mathbf{x}$ the expected value of $f_{\text{ns2}}(\mathbf{x})$ that we denote $E[f_{\text{ns2}}(\mathbf{x})]$.

e) What is the minimum of $E[f_{\text{ns2}}(\mathbf{x})]$?

f) Compute, for a given $\mathbf{x}$, the variance of $f_{\text{ns2}}(\mathbf{x})$ that we will denote $\text{Var}(f_{\text{ns2}}(\mathbf{x}))$. Show that $\text{Var}(f_{\text{ns2}}(\mathbf{x}))$ does not decrease to zero when $\mathbf{x}$ approaches the optimum of $E[f_{\text{ns2}}(\mathbf{x})]$.

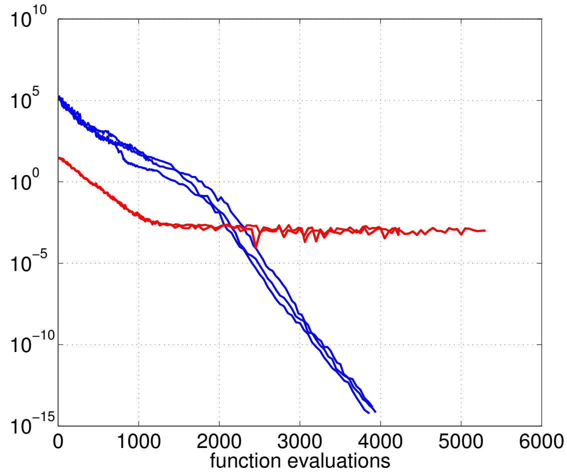

Figure 1: Evolution of the objective function value of the best solution per iteration for f_{ns1} and f_{ns2}.

g) We are using the CMA-ES algorithm to optimize the functions f_{ns1} and f_{ns2}. We set $\epsilon = 10^{-3}$. We display in Figure 1 the evolution of the objective function value of the best candidate solution at a given iteration versus the number of function evaluations for both the f_{ns1} and f_{ns2} functions. For each function, we display three independent runs. Identify to which function the red plots correspond and to which function the blue plots correspond. Explain your reasoning.

Exercice 2 (5 points)

We consider the following test functions to be minimized:

$$f_1(\mathbf{x}) = \mathbf{x}_1^2 + 10^6 \sum_{i=2}^n \mathbf{x}_i^2 \,; \quad f_2(\mathbf{x}) = 10^6 \mathbf{x}_1^2 + \sum_{i=2}^n \mathbf{x}_i^2 \,; \quad f_3(\mathbf{x}) = f_1(\mathbf{R}\mathbf{x}) \;; \quad f_4(\mathbf{x}) = f_2(\mathbf{R}\mathbf{x})$$

where for the functions f_3 and f_4 the matrix $\mathbf{R} \in \mathcal{M}_n(\mathbb{R})$ is a rotation matrix that is sampled uniformly among the set of rotation matrices.

a) What is the global minimum of each function (explain your reasoning)? Compute the Hessian matrix and its condition number for each function.

b) Fill in the table below by putting a cross whenever you think a property is true.

	separable	non-separable	unimodal	multi-modal	ill-conditioned
f_1					
f_2					
f_3					
f_4					

We are using the CMA-ES to minimize the functions f_1, f_2, f_3 et f_4 in dimension $n = 5$. The initial mean vector is taken equal to $(1, 1, 1, 1, 1)$ and the initial step-size is equal to 10. The

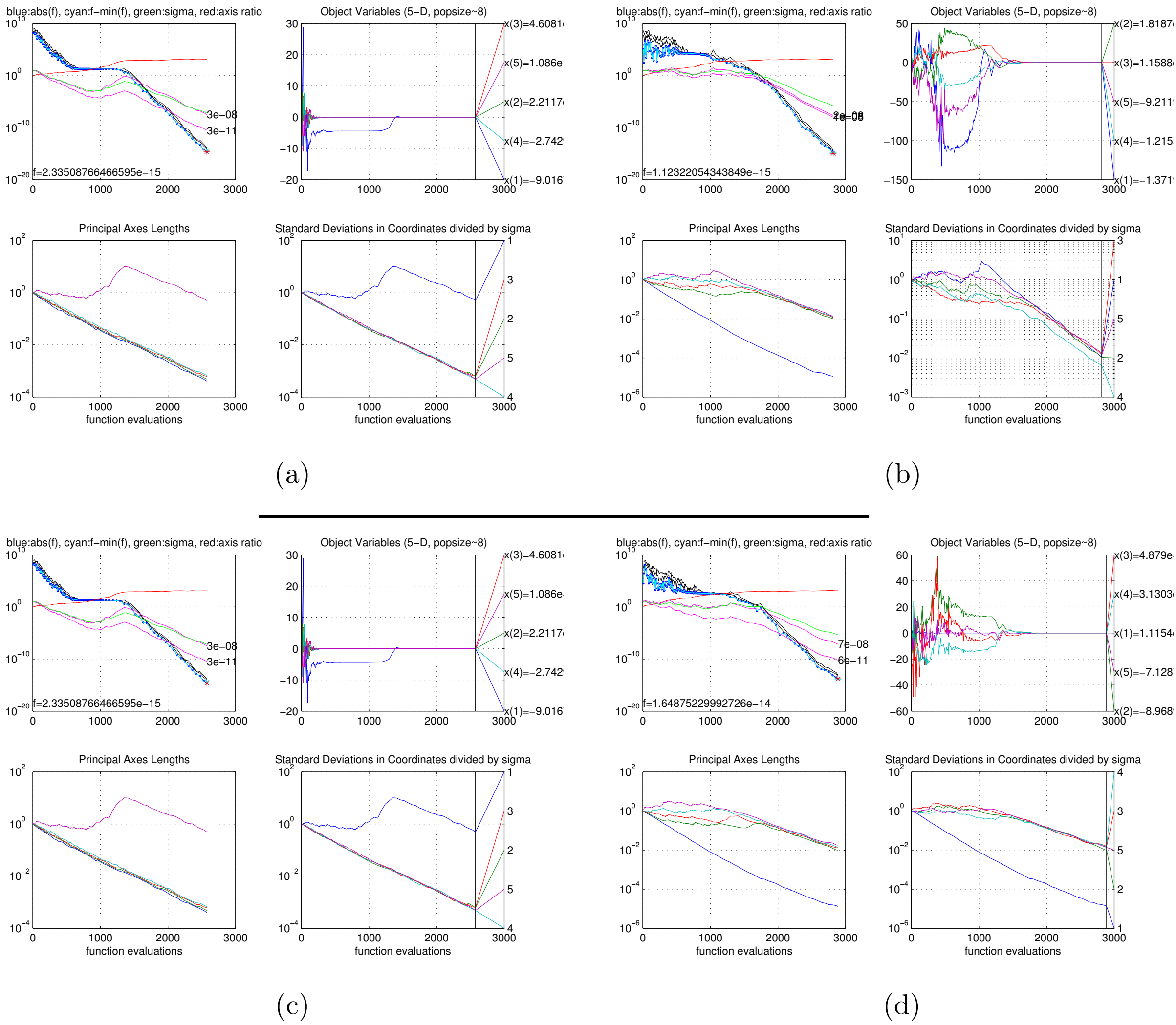

Figure 2: Single runs of the CMA-ES algorithm on the functions f_1, f_2, f_3 and f_4. Identify the function corresponding to each run.

graphical output of a single run of the CMA-ES algorithm minimizing each of the 4 functions is presented in Figure 2.

c) Explain what is presented on each of the 4 graphs of the graphical output (for the upper left plot we will only consider the blue and green curves).

d) Identify which function correspond to each run presented in Figure 2 (a), (b), (c) et (d). The reasoning should be carefully explained.

e) Explain in details what you observe on the plots in Figure 2 (c). Identify and explain in particular the different convergence phases.

DERVATIVE FREE OPTIMIZATION FINAL EXAM, PART 2

Exercice 1 *On the pattern search method*

Consider the classical pattern search method for the minimization of a function $f : \mathbb{R}^n \to \mathbb{R}$ with a fixed set of directions $\mathcal{D}$ such that

$$\forall d \in \mathcal{D}, \quad ||d|| = 1$$

and

$$\kappa = \min_{||v||=1} \max_{d \in \mathcal{D}} v^T d > 0$$

Denote $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ the sequence of points of the pattern search method and $(\alpha_k)_{k\in\mathbb{N}}$ the associated step size. We recall that the acceptation criterion for a new point is the following :

$$f(x_k + \alpha_k d) < f(x_k) - c\frac{\alpha_k^2}{2}$$

where $c > 0$ is fixed and that $\alpha_{k+1} = \theta\alpha_k$ (respectively $\alpha_{k+1} = \gamma\alpha_k$) in case of failure (respectively success) with $\theta \in]0, 1[$ and $\gamma \geq 1$.

The following lemma (admitted here) can be proven :

Lemma Assume that f is C^1, ∇f is ν-Lipschitz and that f is bounded from below by $m \in \mathbb{R}$. Then, the sequence of step size satisfies for all $N \in \mathbb{N}$:

$$\sum_{k=0}^{N} \alpha_k^2 \leq \frac{2\gamma^2}{c(1-\theta^2)} \left(\frac{c\alpha_0^2}{2\gamma^2} + f(x_0) - m \right)$$

1. Prove that

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \quad |f(y) - f(x) - \nabla f(x)^T (y - x)| \leq \frac{\nu}{2}||y - x||^2$$

2. Prove that $\lim_{k\to\infty} \alpha_k = 0$ and that the set of failure steps is infinite.
3. Prove that for a failure step

$$||\nabla f(x_k)|| \leq \frac{c + \nu}{2\kappa}\alpha_k$$

4. Prove that

$$\liminf_{k\to\infty} ||\nabla f(x_k)|| = 0$$

Exercice 2 *On the trust region method*

Consider the following function on $\mathbb{R}^2$:

$$f(x_1, x_2) = \begin{cases} x_1^2 + x_2^2 + (10 - x_1)x_2 & \text{if } x_1 < 10 \\ x_1^2 + \boldsymbol{x}_2^2 & \text{if } x_1 \geq 10 \end{cases}$$

and the following set of initial points

$$\mathcal{Y}_0 = \{(11, 1), (11, 0), (10, -1), (10, 1), (10, 0), (9, 0)\}$$

1. Prove that the first quadratic model around the initial point $x_0 = (10, 0)$ is equal to
$$m_0(x_1, x_2) = x_1^2 + x_2^2$$
2. Assume that is initial radius Δ_0 is equal to 2, what is the next possible iterate x_0^+ ?
3. Compute the ratio
$$\rho_0 = \frac{f(x_0) - f(x_0^+)}{m_0(x_0) - m_0(x_0^+)}$$
Is the point x_0^+ accepted and what is the set $\mathcal{Y}_1$?

Exercice 3 *On the RBF and the kriging method*

Consider the two following metamodels for a given function f defined on $\mathbb{R}^n$:

– A RBF metamodel with a radial basis function

$$h(r) = e^{-cr^2}$$

– A kriging model with a covariance function

$$c(x, y) = \theta_1 + \theta_2 \exp(-\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{2\sigma_i})$$

Prove that for a set of parameters for the kriging that will be given, the two metamodels are equal.