Université de Versailles — Year 2020-2021
http://dumas.perso.math.cnrs.fr/V04.html — M2 'Optimization' and AMS

DERVATIVE FREE OPTIMIZATION

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 Nelder Mead algorithm*

Consider the Nelder Mead algorithm for the sphere function in $\mathbb{R}^2$:

$$f(x, y) = x^2 + y^2$$

and the initial simplex made of $A = (4, 5)$, $B = (5, 3)$ and $C = (5, 6)$.

Compute the best obtained value by this algorithm after one iteration.

Exercice 3 *On the Nelder Mead algorithm*

1. Recall briefly the main principles of the Nelder Mead algorithm. A 2D illustration of the possible steps can be used.
2. Prove that no shrinkage steps are performed when the Nelder Mead algorithm is applied to a strictly convex function. We recall that $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if and only if :

   $$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n,\ \forall \lambda \in]0, 1[,\ f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) \text{ if } x \neq y$$

Exercice 4 *On the Nelder Mead algorithm*

Consider the Nelder Mead algorithm with its nominal parameters. Denote $S_k = \{y_k^0, ..., y_k^n\}$ the simplex obtained at iteration k with the corresponding values by the function f : $f_k^0 \leq f_k^1 \leq ... \leq f_k^n$.

Assume that the function f is bounded from below

1. Prove that the sequence $(f_k^0)_{k \in \mathbb{N}}$ is convergent.
2. If only a finite number of shrink steps occurs, prove that all the sequences $(f_k^i)_{k \in \mathbb{N}}$ are convergent $(0 \leq i \leq n)$.
3. Define the volume of the simplex :

   $$V(S_k) = \frac{1}{n!} \det[y_k^0 - y_k^n, ...y_k^{n-1} - y_k^n]$$

   Check that for $n = 2$ the definition of $V(S_k)$ corresponds to the area of the triangle (S_k).
4. Give the value of $V(S_{k+1})$ in terms of $V(S_k)$ when an expansion occurs (for sake of simplicity, we can assume that $y_k^n = 0$).
5. Give the value of $V(S_{k+1})$ in terms of $V(S_k)$ when a shrink step occurs.

Exercice 5 *On the Cauchy Step in a trust region method*

Consider the following quadratic model in the closed ball $B(x_0, R)$ of $\mathbb{R}^n$:

$$m(x_0 + h) = c + < g, h > + \frac{1}{2}{}^t hHh$$

where $c \in \mathbb{R}$, $g \in (\mathbb{R}^n)^*$ and $H \in \mathcal{M}_n(\mathbb{R})$.

We denote by t_c the Cauchy step in the steepest descent direction, that is :

$$t_c = \text{argmin}\{m(x_0 - tg), t > 0, x_0 - tg \in B(x_0, R)\}$$

The aim is to prove that

$$m(x_0) - m(x_0 - t_c g) \geq \frac{1}{2}||g|| \min(\frac{||g||}{|||H|||}, R)$$

where $\frac{||g||}{|||H|||} = +\infty$ if $H = 0$.

1. First, we assume that ${}^t gHg > 0$. Prove that in this case

$$t_c = \min(\frac{R}{||g||}, \frac{||g||^2}{{}^t gHg})$$

   and conclude

2. Assume now that ${}^g Hg \leq 0$. Prove that $t_c = \frac{R}{||g||}$ in this case and conclude.

Exercice 6 *On the Lagrange interpolation*

Consider a set $\mathcal{Y} = \{X_1, ..., X_p\}$ of p points in $\mathbb{R}^n$ where p is the cardinality of the polynomial space $\mathbb{R}_d[x_1, ..., x_n]$ $(d \geq 1)$. Assume that the set is poised. Denote $\mathcal{B} = \{\Phi_1, ..., \Phi_p\}$ the monomial basis of $\mathbb{R}_d[x_1, ..., x_n]$.

The following algorithm is proposed to define a new polynomial basis :

Initialisation : set $l_j = \Phi_j$ for all $j = 1, ..., p$.

For $i = 1, 2, ..., p$:

— *Point selection* : find $j_0 = argmax_{i \leq j \leq p}|l_i(X_j)|$. If $l_i(X_{j_0}) = 0$ then stop (the set is not poised). Otherwise, swap points X_i and X_{j_0} in $\mathcal{Y}$.

— *Normalisation* : change $l_i(x) \leftarrow \dfrac{l_i(x)}{l_i(X_i)}$

— *Orthogonalization* : for $j = 1, ..., p$, $j \neq i$, change $l_j(x) \leftarrow l_j(x) - l_j(X_i)l_i(x)$

1. If $d \in \{1, 2\}$, what is the value of p for a given n ?
2. Give a condition on a matrix, built with $\mathcal{B}$ and $\mathcal{Y}$, so that the set is poised.
3. Prove that the previous algorithm transforms the basis $\mathcal{B}$ into the Lagrange basis (which definition will be recalled).

Exercice 7 *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 8 *On a DFO trust region method*

Consider a DFO trust region method : denote $f : \mathbb{R}^n \to \mathbb{R}$ the function to minimize and m_k the quadratic model at iteration k around the current point x_k (in a ball of radius Δ_k). Assume that the hessian of m_k at x_k, H_k, has at least one negative eigenvalue and let τ_k its minimal eigenvalue. Denote g_k the gradient of m_k at x_k.

1. Recall the general principles of a DFO trust region method
2. Give the expression of $s \mapsto m_k(x_k + s)$ with respect to g_k and H_k.
3. Prove that there exists a vector $s_k \in \mathbb{R}^n$ such that
$$\begin{cases} < s_k, g_k > \leq 0 \\ ||s_k|| = \Delta_k \\ < s_k, H_k s_k >= \tau_k \Delta_k^2 \end{cases}$$
4. Give a positive lower bound of $m_k(x_k) - m_k(x_k + s_k)$ in terms of τ_k and Δ_k.

Exercice 9 *On a first order DFO trust region method*

The following algorithm in Matlab gives an example of a first order DFO trust region method. The objective is here to use a classical trust region method in dimension n, based on a linear interpolation of the function to minimize f made with a Lagrange interpolation from a set of p points :

```
n=3;  % dimension
p=n+1;
gamma=1.1;
theta=0.9;
eta=0.01;
Nstep=100;
X=rand(n,1); delta=0.1;           % initialization
Xla=[X,X*ones(1,p-1)+delta*(ones(n,p-1)-2*rand(n,p-1))];
Xlatot=Xla;  % total set of possible interpolation points
```

```
Xtot=[X];
for i=1:Nstep
    k=size(Xlatot,2);
    u=zeros(k,1);
    for j=1:k
        u(j)=norm(Xlatot(:,j)-X);
    end
    [a,b]=sort(u);
    Xla=Xlatot(:,b(1:p));      % choice of the nearest p points from X
    w=linlagrange(X,Xla);g=w(2:p);A=zeros(n,n);b=zeros(n,1);
    hplus=linprog(g,A,b,A,b,-delta*ones(n,1),delta*ones(n,1));
    Xplus=X+hplus;
    Xlatot=[Xlatot,Xplus];
    rhok=(f(X)-f(Xplus))/(f(X)-linmodel(g,f(X),hplus)+1E-16);
    if (rhok>eta)
       X=Xplus;delta=gamma*delta;
    else
       delta=theta*delta;
    end
    Xtot=[Xtot,X];
end
disp('best value:');disp(X)
```

In particular, the Matlab instruction `linprog` is used to minimize the function $m(x) = g' * x$ for $-\delta \leq x_i \leq \delta$. $(1 \leq i \leq n)$. The functions `linlagrange`, `linmodel` and `f` need to be defined to complete the code.

1. Give a global description of the script above.
2. Write a possible function `linmodel.m`
3. Write a possible function `linlagrange.m`, either in the particular case where $n = 2$ or in the general case.

Exercice 10 *On the DIRECT method*

Consider the 1D function $f(x) = x^4 - 2x$ to be minimized on the interval $[-1, 3]$.

Make three iterations of the DIRECT method on this case.

Exercice 11 *On the Kriging model*

The Kriging model is a surrogate model of a given function $J : \mathbb{R}^n \to \mathbb{R}$ that can be written as :

$$\hat{J}(X) = \sum_{i=1}^{N} \omega(X_i) J(X_i)$$

where the N points $X_i \in \mathbb{R}^n$ have been previously evaluated by J.

In this expression, J and $\hat{J}$ are viewed as a random fields where J is assumed to have a zero mean value everywhere. Moreover, the covariance between the evaluation at two points X and Y has the following form :

$$cov(J(X), J(Y)) = c(X, Y)$$

where the function $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is supposed to be known.

At each point $X \in \mathbb{R}^n$, the random field $\hat{J}$ is defined so as to minimize the standard deviation between $J(X)$ and $\hat{J}(X)$, while ensuring $E(\hat{J}(X)) = E(J(X))$.

1. Prove that
$$\hat{J}(X) =^t KC^{-1}z$$
where $K =^t (c(X_1, X), ..., c(X_N, X))$, $z =^t (J(X_1), ...J(X_N))$ and C is a $N \times N$ matrix such that $C_{i,j} = c(X_i, X_j)$.
2. Verify that the Kriging function is an interpolation model, that is $\hat{J}(X_i) = J(X_i)$ for all $i \in \{1, ..., N\}$.
3. Give an expression of $Var(J(X) - \hat{J}(X))$ using $c(X, X)$, K and C.

Exercice 12 *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.