Derivative Free Optimization
Part 2 : Deterministic methods

1) Introduction
2) Local methods
- S1 (→ Pattern search
- → Nelder Mead
- S2 + S3 (→ Multi-Direction Search
- → DFO trust region

(Pattern search, Nelder Mead, Multi-Direction Search: direct methods; DFO trust region: trust region)

3) Global methods
- S4 + S5 (→ DIRECT
- → Response surface methods

For each method :
- → algorithmic presentation
- → theoretical study

★ Evaluation of this part :
- → Two small quizz (~20')
- → A final report (based on an article study)

Informations on the web page :

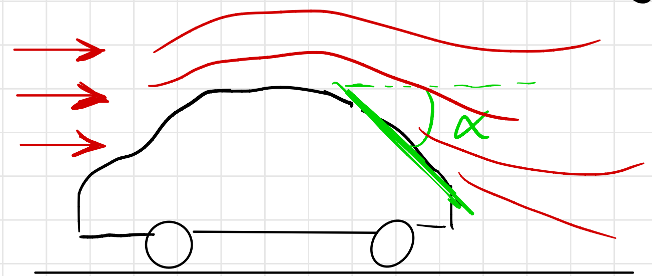
http://dumas.perso.math.cnrs.fr/V04.html

1) Introduction (same as part 1)

Pb to solve: find the / a local or global minimum of a given cost function $f : \mathbb{R}^n \to \mathbb{R}$ (or $\Omega \subset \mathbb{R}^n \to \mathbb{R}$) (continuous optimization)

We assume that f can be smooth eventually, but its gradient is unknown, or difficult to compute.

Example (among many): car shape optimization with respect to aerodynamic drag.



Minimise the drag coefficient:

C_d = (pressure + viscous) drag

(surface integration issued from a 3D Navier Stokes computation)

with respect to car shape parameters, such as α, ..., (angles, length, ...)

It is a very difficult optimization problem, in the sense :

→ The cost function is computationnally expensive to compute.

→ The gradient of the cost function is unknown or difficult and costly to compute.

→ The cost function is not smooth on its admissible set ($\alpha \mapsto C_d(\alpha)$ is not continous everywhere)

→ The contraints are complex to check (either geometrical or aerodynimical)

→ This problem need a Derivative Free Optimization approach!

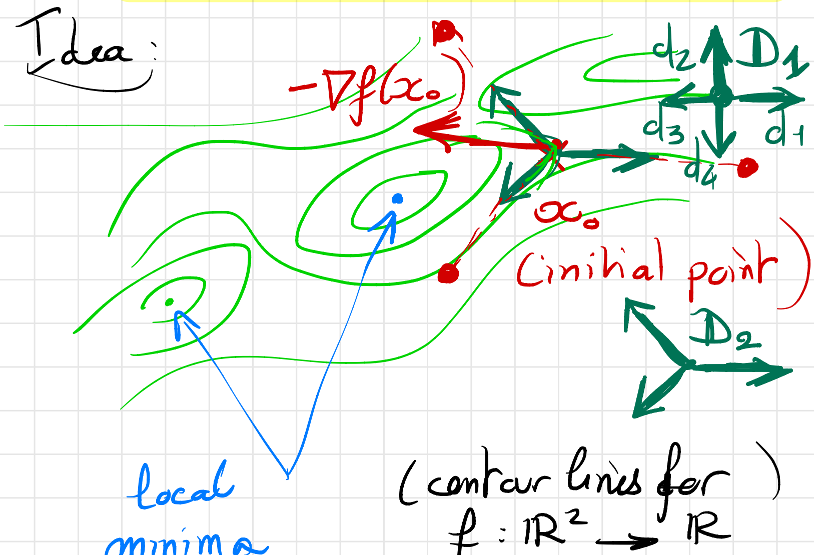
* The DFO method must use a reduced number of evaluations of the cost function.
* A local optimization method is sufficient here //

2) Local methods

→ chronological line of the Review article (JGO, 2013)

2.1) Pattern Search method

Idea:



Algorithm:

* Define a set of directions $D = \{d_1, \dots, d_N\}$ that positively spans $\mathbb{R}^n$ (or eventually a finite number of sets)
* Define an initial point x_0 and an initial step size α_0
* $k \to k+1$:
* Evaluate $\{f(x_k + \alpha_k d_i),\ 1 \leq i \leq N\}$
* If $f(x_k + \alpha_k d_{i_0}) < f(x_k)$

  then $\begin{cases} \to x_{k+1} = x_k + \alpha_k d_{i_0} \\ \to \text{increase the step size} \end{cases}$ success

  Else $\begin{cases} \to x_{k+1} = x_k \\ \to \text{decrease the step size} \end{cases}$ failure

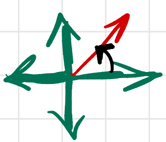
→ Implementation with Scilab on the sphere function: convergence in any dimension to the (global) minimum. //

Remarks:

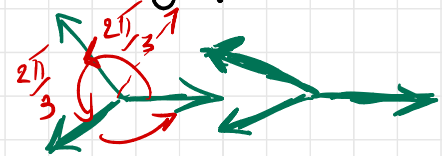
* A simple choice for the set of directions D can be:

$$D = \{e_i, -e_i, 1 \le i \le n\}$$

where $(e_i)_{1 \le i \le n}$ is the canonical basis of $\mathbb{R}^n$. ($N = 2n$)



* The minimal number of directions that positively spans $\mathbb{R}^n$ is $N = n+1$.



* A positive parameter is associated to any family:

$$K = \min_{\substack{\|v\|=1 \\ v \in \mathbb{R}^n}} \left(\max_{d \in D} \langle v, \frac{d}{\|d\|} \rangle \right)$$

→ For the "cross" family: $K = \frac{1}{\sqrt{2}}$

→ For the "$\{1, j, j^2\}$" family: $K = \cos\left(\frac{\pi}{3}\right) = \frac{1}{2}$

In other word, in the case of

a regular function (C^1 at least), there exists for any point x_k, a direction d_i which is a descent direction, that is:

$$\langle -\nabla f(x_k), d_i \rangle > 0$$

- There exists a convergence result of the pattern search method for smooth function towards a critical point.

Reference: Book "Introduction to derivative free Optimization" (Conn, Scheinberg, Vicente)

* Asumptions

→ f C^1

→ The level set $L(x_0) = \{x \in \mathbb{R}^n, f(x) \leq f(x_0)\}$ is compact (example: f is coercive that is $\lim_{\|x\| \to +\infty} f(x) = +\infty$)

→ ∇f is Lipschitz continuous on $L(x_0)$.

* Result:

There exists a subsequence of (x_k) such that

$$\lim_{k \to +\infty} \|\nabla f(x_{\sigma(k)})\| = 0$$

→ Exercice: a variant of this pattern search: replace

The original descent condition :

$$f(x_k + \alpha_k d_i) < f(x_k)$$

by a more strict condition :

$$f(x_k + \alpha_k d_i) \leq f(x_k) - c\,\frac{\alpha_k^2}{2}$$

(similar to an Armijo condition in a steepest descent method)

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 :

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

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

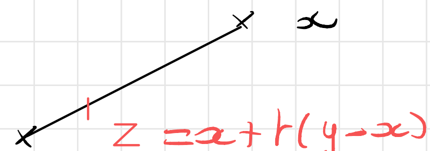
2. Prove that $\lim_{k\to\infty} \alpha_k = 0$ and that the set of failure steps is infinite.
3. Prove that for a failure step

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

4. Prove that

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

1)



Define: $g : \begin{pmatrix} [0,1] \longrightarrow \mathbb{R} \\ t \longmapsto f(x + t(y-x)) \end{pmatrix}$, g C^1 :

$$g'(t) = \langle y - x, \nabla f(x + t(y-x)) \rangle$$

$$g'(t) - g'(0) =$$

$$\dots = \langle y-x, \nabla f(x+t(y-x)) - \nabla f(x) \rangle$$

$\Rightarrow$ (integration between 0 and 1)

$$\left| \int_0^1 (g'(t) - g'(0)) dt \right| = |g(1) - g(0) - g'(0)|$$

$$\leq \|y-x\| \int_0^1 \|\nabla f(x+t(y-x)) - \nabla f(x)\| dt$$

(Cauchy-Schwarz)

$\Rightarrow$

$$|f(y) - f(x) - \langle y-x, \nabla f(x) \rangle|$$

$$\leq \|y-x\| L \int_0^1 t \|y-x\| dt$$

Lipsch. Condition

$$\frac{1}{2} \|y-x\|^2$$

2) $\sum_{k=0}^{N} \alpha_k^2 \leq M$

$\Rightarrow$ the positive serie is bounded and thus convergent. Its general term, α_k^2, goes to 0 when $k \to +\infty$.

Consequently, the number of unsuccessfull steps is infinite because such steps are the only steps where α_k is strictly decreasing.

3) A failure step means that

$$0 \leq f(x_{k+1}) - f(x_k) + c \frac{\alpha_k^2}{2} \quad \forall i \in \{1, \dots N\}$$

$x_{k+1} = x_k + \alpha_k d_i$

$$0 \le f(x_k + \alpha_k d_i) - f(x_k) + \frac{c}{2}\alpha_k^2$$

$(\forall i)$

$$\le \alpha_k \langle d_i, \nabla f(x_k) \rangle + \frac{c}{2}\alpha_k^2 + \frac{\nu}{2}\alpha_k^2$$

Q1

$\langle d_i, \nabla f(x_k)\rangle < 0$ (for some d_i)

Moreover

$$K = \underset{v}{\mathrm{Min}}(\quad) \le \max_{1 \le i \le N} \left\langle -\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}, d_i \right\rangle$$

achieved at i_0

$$\Longrightarrow \left\langle \frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}, d_{i_0} \right\rangle \le -K$$

9

$\Longrightarrow$

$$0 \le -K \|\nabla f(x_k)\| \alpha_k + \frac{c}{2}\alpha_k^2 + \frac{\nu}{2}\alpha_k^2$$

$(\alpha_k > 0) \Longrightarrow$

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

for unsuccessfull steps

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

Consider the subsequence of unsuccessfull steps (correctly defined Q2): $k \mapsto \varphi(k)$

For this subsequence

$$\|\nabla f(x_{\varphi(k)})\| \le \frac{c+\nu}{2K}\alpha_{\varphi(k)} \xrightarrow{Q2} 0$$

Thus, $\lim_{k \to +\infty} \|\nabla f(x_{\varphi(k)})\| = 0$

$\Longrightarrow$ More details for the proof of the lemma: PhD thesis B. Pauwels (2016)

or :

[61] Kolda, T. G., Lewis, R. M., and Torczon, V. Optimization by direct search : New perspectives on some classical and modern methods. *SIAM review 45*, 3 (2003), 385–482.

In the next session, we will present a new approach (simplex-type) for local optimization, in particular the Nelder-Mead algorithm (1965, still used in Matlab).