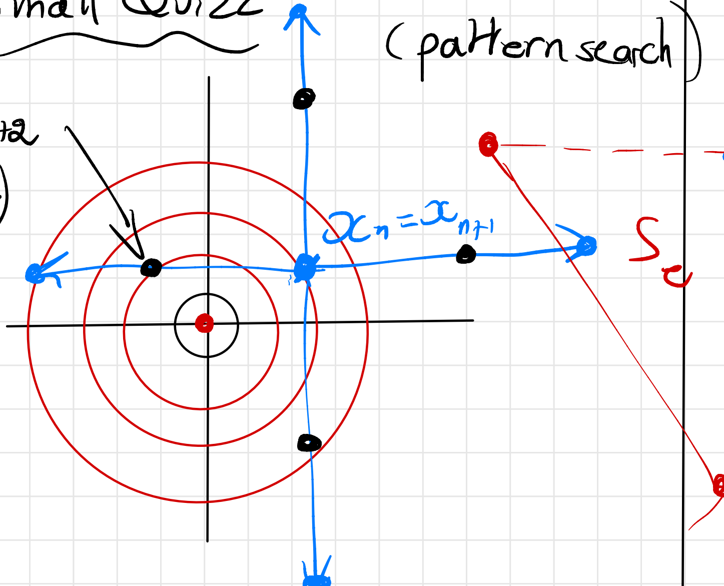
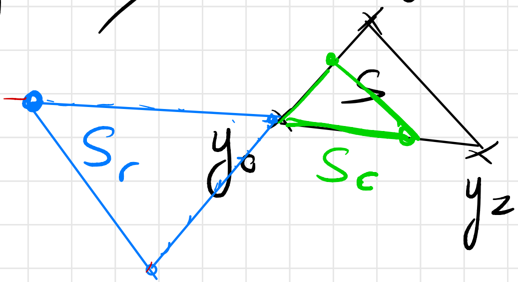
Small Quizz

20

Ex1) (pattern search)



Ex3) (MDS)

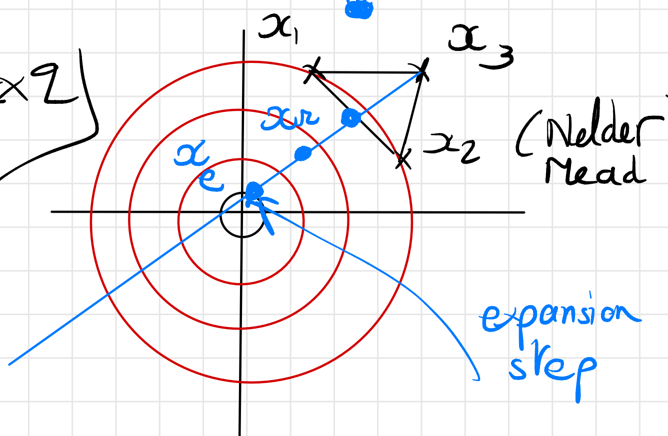


$$Vol(S_r) = Vol(S)$$

$$Vol(S_e) = 2^n \, Vol(S)$$

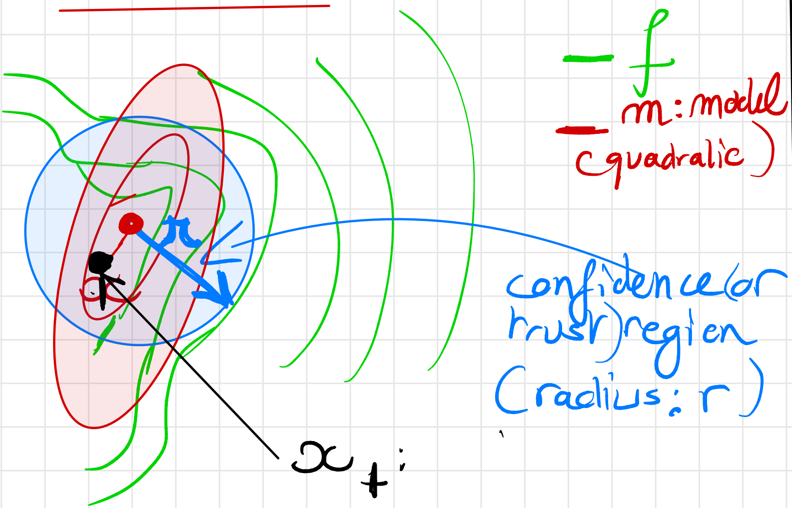
$$Vol(S_c) = \left(\frac{1}{2}\right)^n Vol(S)$$

Ex2) (Nelder Mead)



2.4) Trust region methods

The idea is to locally replace the function to minimize (f) by a quadratic (or polynomial) model and locally minimize this model.



At each step, the model is minimized on the trust region (result is called x^+).
The algorithm iterates in the following way :

* If $f(x_+)$ is sufficiently decreasing compared to $f(x)$, replace x by x_+ and increase the trust region.
* Else, x is kept and the confidence region is reduced.

Q1 → Which model to chose ?

Q2 → Which criteria for a succesful step ?

Answer to question 1 :

* In a gradient / Hessian context for f, the best choice for a quadratic model is given by a 2nd order Taylor expansion around x :

$$\boxed{m(x+h) = f(x) + \langle h, \nabla f(x) \rangle + \frac{1}{2} \langle h, Hf(x) h \rangle}$$

When no gradient information is available for f, a Lagrange interpolation model is built (see below).

Answer to question 2 :

* We say that a step is successful: if :

$$\frac{\overbrace{f(x) - f(x_+)}^{\text{function decrease}}}{\underbrace{m(x) - m(x_+)}_{\text{model decrease}}} \geq \eta$$

where $\eta < 1$ $(\eta \approx 0.1)$.

In this case, the trust region radius is increased by a factor $\gamma_1 > 1$ (else it is reduced by $\gamma_2 < 1$).

Example of implementation of this method with Scilab

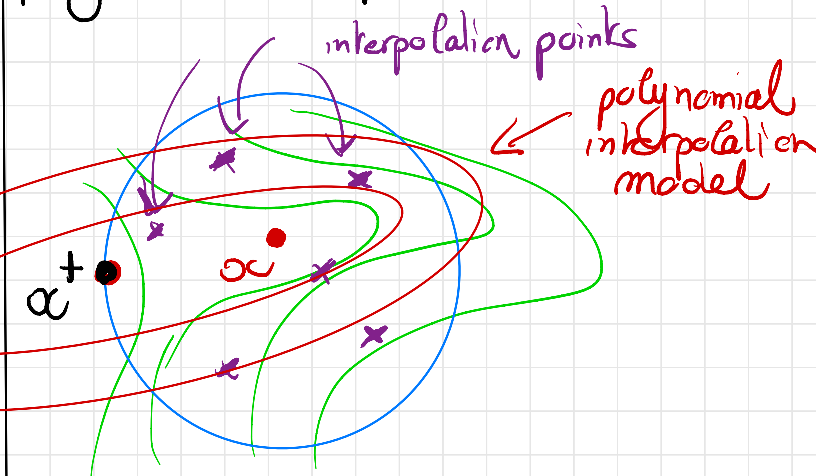
$f(x,y) = (x-1)^4 + (y-2)^4$

(and a Taylor expansion).

In a gradient / Hessian context, there exists a convergence result of $(x_k)_{k \in \mathbb{N}}$ to a critical point of f.

(see lecture notes Gould & Leyffer)

In a DFO context, the Taylor expansion is replaced by a Lagrange polynomial interpolation :



After initialization, a new point x_+ enters in the set of interpolation points.

($\longrightarrow$ NEWUOA, Powell, 2004)

Ex 5)

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

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

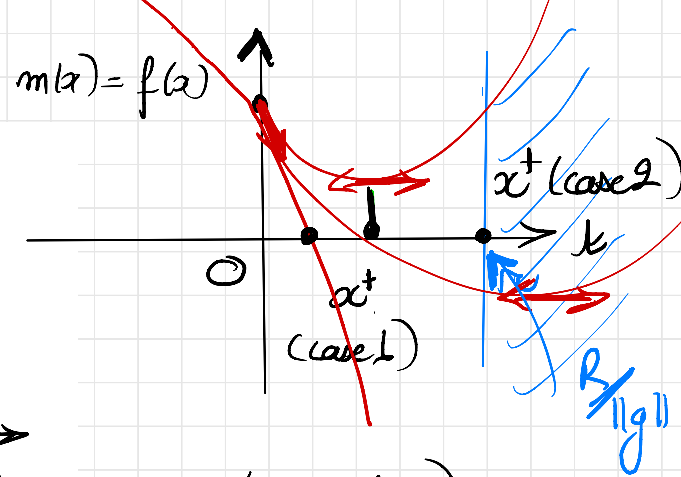
where $\frac{||g||}{|||H|||} = +\infty$ if $H = 0$.

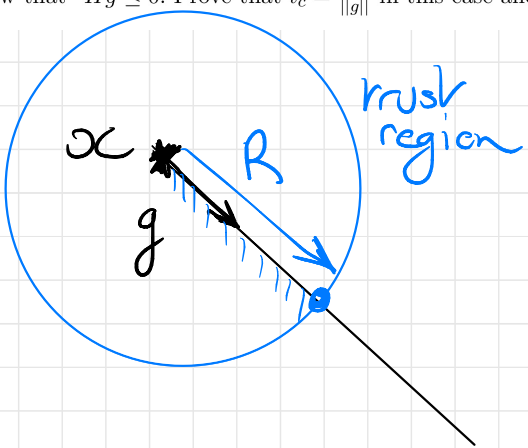
1. First, we assume that ${}^t gHg > 0$. Prove that in this case

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

   and conclude

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





$\rightarrow$

$$\varphi(t) = m(x_0 - tg) =$$

$$\varphi'(t) = -\langle g, \nabla m(x_0 - tg)\rangle$$

where $\varphi(t) = -t||g||^2 + \frac{t^2}{2}\,{}^t gHg$

1) If ${}^t gHg > 0$ (case 1 or 2)

φ is minimal at

$$t^* = \frac{||g||^2}{{}^t gHg}$$

$*$ Two cases can occur:

→ if $t^* \leq \frac{R}{\|g\|}$

$$t_c = t^* = \frac{\|g\|^2}{{}^t g H g}$$

→ if $t^* > \frac{R}{\|g\|}$

$$t_c = \frac{R}{\|g\|}$$

In any cases

$$t_c = \mathrm{Min}\left(\frac{R}{\|g\|}, \frac{\|g\|^2}{{}^t g H g}\right)$$

2) If ${}^t g H g \leq 0$, φ goes to $-\infty$ and is minimal at $t_c = \frac{R}{\|g\|}$

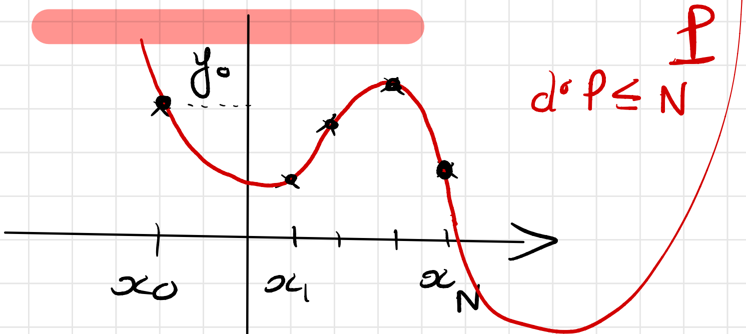
The conclusion follows easily:

$$\underbrace{m(x_0) - m(x_0 - t_c g)}_{\text{decrease of the model}} \geq \frac{1}{2}\|g\| \,\mathrm{Min}\left(\frac{\|g\|}{\|H\|}, R\right)$$

Remark: this condition gives a sufficient decrease to prove the convergence of the trust region method in a gradient / Hessian context.

Lagrange interpolation for a function $f : \mathbb{R}^n \to \mathbb{R}$

- Recall in 1D :



$\exists !\, P \in \mathbb{R}_N[x]$ such that $\forall i \in \{0, \dots, N\},\ P(x_i) = y_i$

* In dimension $n \geq 1$, we seek for interpolation with:

$P \in \mathbb{R}_2[x_1, \dots x_n]$: set of polynomials with n-variables of total degree ≤ 2

(it contains only monomial $x_1^{\alpha_1} \dots x_n^{\alpha_n}$ with $\alpha_1 + \dots + \alpha_n \leq 2$)

In this case,

$$\dim \left(\mathbb{R}_2[x_1, \dots x_n] \right) = p_n \text{ with}$$

$$p_n = \underbrace{1}_{d^\circ \leq 0} + \underbrace{n}_{d^\circ \leq 1} + \underbrace{\binom{n}{2}}_{x_1 x_2 \dots} + \underbrace{n}_{x_1^2, \dots x_n^2}$$

$$= \frac{(n+1)(n+2)}{2}$$

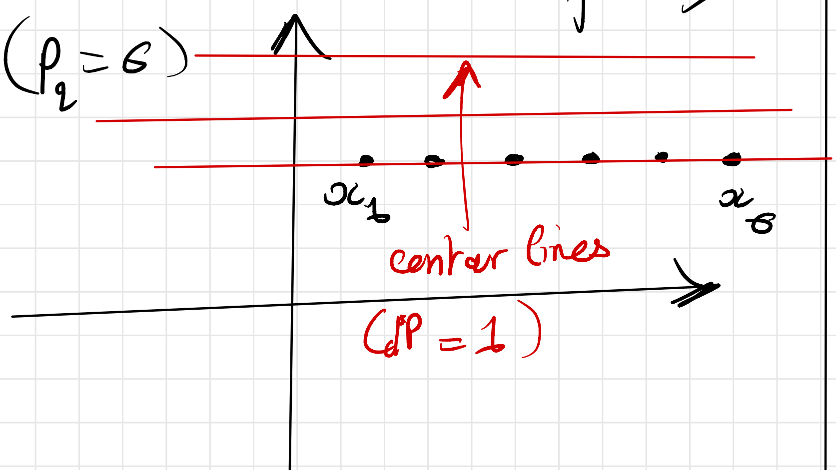
(if $n = 1$, $p_1 = 3$)

A number of points

$$p_n = \frac{(n+1)(n+2)}{2}$$

is needed to built a quadratic interpolation//

To ensure a unique interpolation quadratic polynomial, there is an additional condition if $n \geq 2$:



Moreover, the number of interpolation points needed is quadratically increasing:

$n = 1$, $p_1 = 3$

$n = 2$, $p_2 = 6$

$n = 3$, $p_3 = 10$

$n = 4$, $p_4 = 15$ ......

Def: $Y = \{x_1, \dots, x_N\}$ (with $N = p_n$) is poised if there exists a unique polynomial interpolation

$\forall (y_i)_{1 \leq i \leq N}$, $\exists! \, P \in \mathbb{R}_2[X_1, \dots, X_n]$ such that $P(x_i) = y_i$.

In this case,

$$P(x_1, \ldots x_n) = \sum_{j=1}^{N} y_j \, l_j(x_1, \ldots x_n)$$

where $\{l_1, \ldots l_N\}$ is the basis of elementary Lagrange polynomials.

→ Implementation with Scilab of Lagrange interpolation

→ solve a Vandermonde system in the canonical basis:

$\{1, x_1, \ldots x_n, x_1^2, x_1x_2, \ldots x_1x_n, \ldots x_n^2\}$

Ex 6) (computation of an elementary Lagrange polynomial basis (≡ similar to Gauss method or Gram Schmidt)

Ex 7)

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

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

and the following set of initial points

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

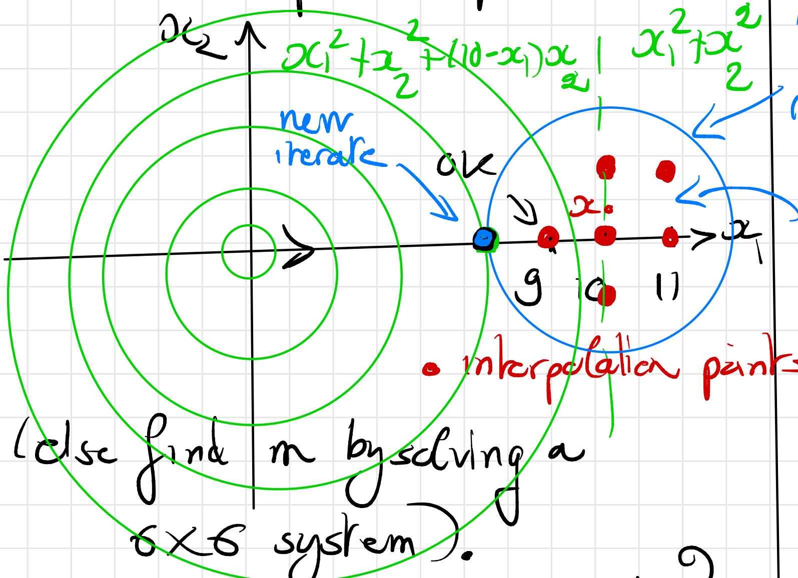
1. Prove that the first quadratic model around the initial point $x_0 = (10, 0)$ is equal to
$$m_0(x_1, x_2) = x_1^2 + x_2^2$$
2. Assume that is initial radius Δ_0 is equal to 2, what is the next possible iterate x_0^+ ?
3. Compute the ratio
$$\rho_0 = \frac{f(x_0) - f(x_0^+)}{m_0(x_0) - m_0(x_0^+)}$$

Is the point x_0^+ accepted and what is the set $\mathcal{Y}_1$?

1) It is enough to check that

$$m(x_1, x_2) = x_1^2 + x_2^2 = f(x_1, x_2)$$

at the interpolation points



(else find m by solving a 6x6 system).

2) What is the value of x_0^+?

$x_0^+ = (8, 0)$ (minimum of m on $B(x_0, 2)$)

$$\rho = \frac{f(x_0) - f(x_0^+)}{m(x_0) - m(x_0^+)} = 1 \geq \eta$$

$\rightarrow$ x_0^+ is accepted:

$$\begin{cases} x_1 = x_0^+ \\ r_1 = \gamma r_0 \geq 2 \end{cases}$$

At this new iteration, there are 7 possible interpolation points

We take the new set by

- including x_1
- excluding the farest point: $(11, 1)$

and we iterate.....

→ Two problems remain :

* The cost of the first interpolation model

* The possible non poised system that can be obtained