Exemple Kepler's problem

Find the parallelepiped of maximum volume inscribed in the ellipsoid

$$\mathcal{E} = \{x \in \mathbb{R}^3, {x_1}^2/a_1^2 + x_2^2/a_2^2 + x_3^2/a_3^2 = 1\}$$

Write the problem as a canonical optimisation problem

$$\inf_{\substack{{x_1}^2/a_1^2 + x_2^2/a_2^2 + x_3^2/a_3^2 = 1 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ x_3 \geq 0}} f(x), \quad f(x) = -\prod_{i=1}^{3} x_i$$

these constraints are necessarily inactive at x^* because otherwise $V(x^*) = 0$

- Can we apply KKT theorem?
- Which constraints are active ?
- Lagrangian ?

Exemple Kepler's problem

$x_1 = 0$ or $x_2 = 0$ or $x_3 = 0 \Rightarrow f(x) = 0$! inequality constraints are inactive

$$\ell(x, y) = -\prod_{i=1}^{3} x_i + y({x_1}^2/a_1^2 + x_2^2/a_2^2 + x_3^2/a_3^2 - 1)$$

Gradient

Exemple Kepler's problem

$$
\begin{aligned}
\left(-x_2 x_3 + 2yx_1/a_1^2 \right. &= \left. 0 \right) \times x_1 \\
\left(-x_1 x_3 + 2yx_2/a_2^2 \right. &= \left. 0 \right) \times x_2 \\
\left(-x_2 x_1 + 2yx_3/a_3^2 \right. &= \left. 0 \right) \times x_3
\end{aligned}
$$

then sum the 3 equations;

then find x^*

Exemple Kepler's problem

$$-3x_1x_2x_3 + 2y\left(\underbrace{x_1^2/a_1^2 + x_2^2/a_2^2 + x_3^2/a_3^2}_{=1}\right) = 0$$

$$\begin{aligned} -x_2x_3 + 2yx_1/a_1^2 &= 0 \\ -x_1x_3 + 2yx_2/a_2^2 &= 0 \\ -x_2x_1 + 2yx_3/a_3^2 &= 0 \end{aligned}$$

We multiply each equation by the corresponding x_i component

$$\begin{aligned} -x_2x_3x_1 + 2yx_1^2/a_1^2 &= 0 \\ -x_1x_3x_2 + 2yx_2^2/a_2^2 &= 0 \\ -x_2x_1x_3 + 2yx_3^2/a_3^2 &= 0 \end{aligned}$$

Then sum $-3\prod x_i + 2y = 0$ from which $x_2x_3 = 2y/(3x_1)$ Then inserting in the first equation leads to

$$-2y/(3x_1) + 2yx_1/a_1^2 = 2y(x_1/a_1^2 - 1/(3x_1)) = 0$$

Exemple Kepler's problem

$x_1 = 0$ is impossible (or $\prod x_i = 0$ therefore $3x_1^2 = a_1^2$. We obtain similarly

$$x_i = \frac{a_i}{\sqrt{3}}, \quad i = 1, \dots 3$$

Finally we get

$$Vol = 8\prod x_i = 8\frac{\prod x_i}{3\sqrt{3}}, \quad y = \frac{3\prod x_i}{2} = \frac{\prod x_i}{2\sqrt{3}}$$

$$y = \frac{3}{2}\frac{\prod a_i}{(\sqrt{3})^3} = \frac{\prod a_i}{2\sqrt{3}}$$

Exemple

$f(x) = f(0) + \langle \nabla f(0), x \rangle + \frac{1}{2} \langle Hf(0)x, x \rangle$

degree 2 $\Rightarrow$ $f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + c$

Let the function

$$f(x) = 2x_1^2 + 3x_2^2 + 2x_1 x_2.$$

and domain $K = \{x_1^2 + 4x_2^2 \leq 1 \text{ and } x_1 + x_2 \geq 1\}$ $= D_a$

Seek $\inf_{x \in K} f(x)$

1. Show that f has a global minimum over $\mathbb{R}^2$ and calculate it
2. Show that K is nonempty convex.
3. Draw K in solid lines and isovalues of $f(x)$ in dotted lines
4. Write the Lagrangian for the problem (P).
5. Write the 1st order optimality conditions (or KKT conditions).
6. Use the drawing and the result of question 1 to decide which of the two constraints will be active at least.
7. Calculate $x^\star$ and the associated Lagrange multipliers.

Example

1. Compute $\nabla f(x) = (4x_1 + 2x_2, 2x_1 + 6x_2)^T$ and $Hf(x) = \begin{pmatrix} 4 & 2 \\ 2 & 6 \end{pmatrix}$. The Hessian eigenvalues are $(5 \pm \sqrt{5})/2$. Donc

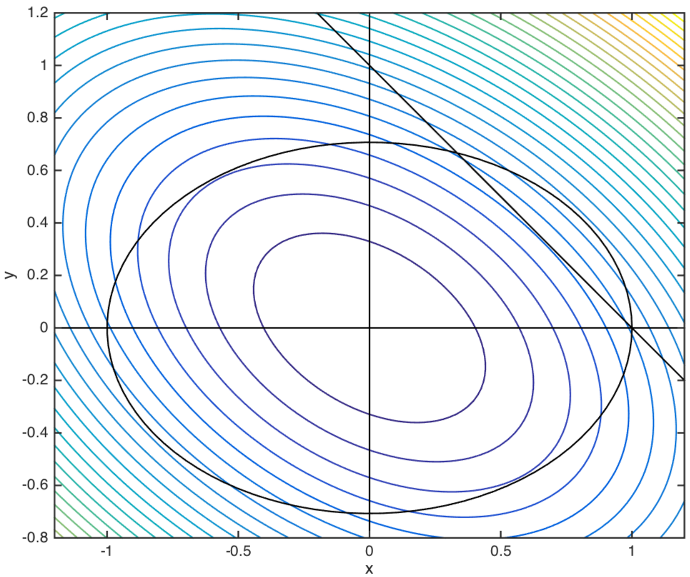
$$f(x) = f(0) + \langle \nabla f(0), x \rangle + \frac{1}{2} \langle Hf(0)x, x \rangle$$

is a quadratic form with a symmetric positive definite matrix, so the minimum 0 is reached in $x = (0, 0)$ and unique.

Example

1. On considère $K = \{x_1^2 + 4x_2^2 \leq 1 \text{ et } x_1 + x_2 \geq 1\}$. Le point $(1, 0) \in K \neq \emptyset$ L'ensemble $\{x_1^2 + 4x_2^2 \leq 1\}$ est l'ellipse de demi grand axe $x_1 \in [-1, 1]$, $x_2 = 0$ et de demi petit axe $x_1 = 0$, $x_2 \in [-1/2, 1/2]$. Il est donc convexe. Le demi plan $\{x_1 + x_2 \geq 1\}$ est lui aussi convexe. L'intersection non vide de 2 convexes est convexe.

Exemple



Exemple

1. On applique la méthode des multiplicateurs de Lagrange:

$$\ell(x, y) = f(x) + y_1(x_1^2 + 4x_2^2 - 1) + y_2(1 - x_1 - x_2)$$

Exemple

1. si $(x_1^\star, x_2^\star)$ minimise f on K il existe $y_1 \geq 0$ et $y_2 \geq 0$ tels que

$$\begin{aligned}
\nabla_x \ell(x^\star, y^\star) &= \begin{pmatrix} 4x_1^\star + 2x_2^\star + 2y_1^\star x_1^\star - y_2^\star \\ 2x_1^\star + 6x_2^\star + 8y_1^\star x_2^\star - y_2^\star \end{pmatrix} = 0_{\mathbb{R}^2}, \\
y_1^\star((x_1^\star)^2 + 4(x_2^\star)^2 - 1) &= 0, \\
y_2^\star(1 - x_1^\star - x_2^\star) &= 0
\end{aligned}$$

Exemple

Comme le minimum global de $f(x)$ n'appartient pas à K, l'une des deux contraintes au moins est forcément active. Vu la convexité de f on prévoit que le minimum sera atteint sur le segment $\{x_1 + x_2 = 1\} \cap K$ et le maximum sur l'arc $\{x_1^2 + 4x_2^2 = 1\} \cap K$.
Pour le minimum on cherche donc x, $y_2 > 0$ tels que

$$\begin{aligned} 4x_1^\star + 2x_2^\star &= y_1^\star \\ 2x_1^\star + 6x_2^\star &= y_2^\star \\ {x_1^\star}^2 + 4{x_2^\star}^2 &< 1 \\ x_1^\star + x_2^\star &= 1 \end{aligned}$$

ce qui conduit à $x_1^\star = 2/3$, $x_2^\star = 1/3$, $y_2^\star = 10/3$, $y_1^\star = 0$ et $f(x^\star) = 5/3$.

Algorithms for constrained optimization

$$\begin{cases} \inf & f(x) \\ s.c. & c^E x) = 0 \\ s.c. & c^I(x) \leq 0 \\ & x \in \mathbb{R}^n \end{cases}$$

with

$$\begin{aligned} f &: \mathbb{R}^n \longrightarrow \mathbb{R}, \\ c^E &: \mathbb{R}^n \longrightarrow \mathbb{R}^m, \\ c^I &: \mathbb{R}^n \longrightarrow \mathbb{R}^p, \\ f \quad c & \quad \text{smooth.} \end{aligned}$$

- Change of unknowns

Algorithms for constrained optimization

best ever method: use the python scipy.optimize package

$$\begin{cases} \inf & f(x) \\ s.c. & c^E x) = 0 \\ s.c. & c^I(x) \leq 0 \\ & x \in \mathbb{R}^n \end{cases}$$

with

$$\begin{aligned} f &: \mathbb{R}^n \longrightarrow \mathbb{R}, \\ c^E &: \mathbb{R}^n \longrightarrow \mathbb{R}^m, \\ c^I &: \mathbb{R}^n \longrightarrow \mathbb{R}^p, \\ f \quad c & \quad \text{smooth.} \end{aligned}$$

- Change of unknowns
- Projection

Algorithms for constrained optimization

$$\begin{cases} \inf & f(x) \\ s.c. & c^E x) = 0 \\ s.c. & c^I(x) \leq 0 \\ & x \in \mathbb{R}^n \end{cases}$$

with

$$\begin{aligned} f &: \mathbb{R}^n \longrightarrow \mathbb{R}, \\ c^E &: \mathbb{R}^n \longrightarrow \mathbb{R}^m, \\ c^I &: \mathbb{R}^n \longrightarrow \mathbb{R}^p, \\ f \quad c & \quad \text{smooth.} \end{aligned}$$

- Change of unknowns
- Projection
- Penalisation

Algorithms for constrained optimization

$$\begin{cases} \inf & f(x) \\ s.c. & c^E x) = 0 \\ s.c. & c^I(x) \leq 0 \\ & x \in \mathbb{R}^n \end{cases}$$

with

$$\begin{aligned} f &: \mathbb{R}^n \longrightarrow \mathbb{R}, \\ c^E &: \mathbb{R}^n \longrightarrow \mathbb{R}^m, \\ c^I &: \mathbb{R}^n \longrightarrow \mathbb{R}^p, \end{aligned}$$

f c smooth.

▶ SQP algorithm $\Leftrightarrow$ $\left.\begin{aligned} \nabla_x \ell(x,y) &= 0_{\mathbb{R}^n} \\ \nabla_y \ell(x,y) &= 0_{\mathbb{R}^m} \end{aligned}\right\}$ m+n system solved by Newton

- Change of unknowns
- Projection
- Penalisation
- Methods using Lagrange multipliers

Change of unknowns

$$\inf_{x_1^2+x_2^2=1} x_1^4 + x_2^4 \Rightarrow \begin{cases} x_1 = \cos\theta \\ x_2 = \sin\theta \end{cases}$$

diffeomorphism $\varphi : \mathbb{R}^n \to K \subset \mathbb{R}^n$

Examples:

- $K = (\mathbb{R}+)^n \longrightarrow$ set $x = y^2$ and optimize without constraints with respect to y.
  $\varphi : (\mathbb{R}^*)^n \to (\mathbb{R}^{+*})^n$, $x_i = \varphi_i(y) = y_i^2$

$$\inf_{\substack{x \geq 0 \\ x \in (\mathbb{R}^+)^n}} f(x) \qquad \tilde{f}(y) = f\left(\begin{pmatrix} y_1^2 \\ \vdots \\ y_n^2 \end{pmatrix}\right) \qquad \inf_{\mathbb{R}^n} \tilde{f}(y)$$

$$x_i^* = {y_i^*}^2$$

Change of unknowns

$\tilde{f}(\theta^*) = \inf_{\mathbb{R}^n} \tilde{f}(\theta)$

$x_i^* = \frac{a_i+b_i}{2} + \frac{b_i-a_i}{2} \cos\theta_i^*$

diffeomorphism $\varphi : \mathbb{R}^n \to K \subset \mathbb{R}^n$

Examples:

- $K = (\mathbb{R}+)^n \longrightarrow$ set $x = y^2$ and optimize without constraints with respect to y.
  $\varphi : (\mathbb{R}^*)^n \to (\mathbb{R}^{+*})^n$, $x_i = \varphi_i(y) = y_i^2$
- $K = \prod_{i=1,\dots,n}[a_i, b_i] \longrightarrow$ set $x_i = \frac{a_i+b_i}{2} + \frac{b_i-a_i}{2} cos\theta_i$ and optimize without constraints with respect to θ_i
  $\varphi : \mathbb{R}^n \to K$, $x_i = \varphi_i(\theta_i)$

$a_i \leq x_i \leq b_i$

$f(x)$ where

$$\tilde{f}(\theta) = f\begin{pmatrix} \frac{a_1+b_1}{2} + \frac{(b_1-a_1)}{2}\cos\theta_1 \\ \vdots \\ \frac{a_n+b_n}{2} + \frac{(b_n-a_n)}{2}\cos\theta_n \end{pmatrix}$$

Consequence of the change of unknowns on the calculation of the gradient

if you know $\nabla f(x)$ $\qquad \nabla \tilde{f}(y)$

$x = \varphi(y)$

$$\tilde{f}(y) = f(\varphi(y))$$

$$\nabla \tilde{f}(y) = J\varphi(y) \nabla f(\varphi(y))$$

φ diffeomorphism $\Rightarrow J\varphi$ defined on $\varphi^{-1}(\mathring{K})$

Consequence of the change of unknowns on the calculation of the gradient

φ diffeomorphism $\Rightarrow J\varphi$ defined on $\varphi^{-1}(\mathring{K})$
$\tilde{f}(y) = fo\varphi(y) = f(x)$

Consequence of the change of unknowns on the calculation of the gradient

φ diffeomorphism $\Rightarrow J\varphi$ defined on $\varphi^{-1}(\overset{\circ}{K})$
$\tilde{f}(y) = fo\varphi(y) = f(x)$ $\nabla_y \tilde{f}(y) = J(y)\nabla_x f(x)$ avec $J(x)$ jacobian matrix of φ

Example of change of unknowns

Example $\varphi : (\mathbb{R}^*)^n \to (\mathbb{R}^{+*})^n$, $x_i = \varphi_i(y) = y_i^2$

Example of change of unknowns

Example $\varphi : (\mathbb{R}^*)^n \to (\mathbb{R}^{+*})^n$, $x_i = \varphi_i(y) = y_i^2$

$$J(y) = \begin{pmatrix} 2y_1 & 0 & \cdots & 0 \\ 0 & 2y_2 & \cdots & 0 \\ \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 2y_n \end{pmatrix} = \begin{pmatrix} 2\sqrt{x_1} & 0 & \cdots & 0 \\ 0 & 2\sqrt{x_2} & \cdots & 0 \\ \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 2\sqrt{x_n} \end{pmatrix},$$

Example of change of unknowns

Example $\varphi : (\mathbb{R}^*)^n \to (\mathbb{R}^{+*})^n$, $x_i = \varphi_i(y) = y_i^2$

$$J(y) = \begin{pmatrix} 2y_1 & 0 & \cdots & 0 \\ 0 & 2y_2 & \cdots & 0 \\ \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 2y_n \end{pmatrix} = \begin{pmatrix} 2\sqrt{x_1} & 0 & \cdots & 0 \\ 0 & 2\sqrt{x_2} & \cdots & 0 \\ \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 2\sqrt{x_n} \end{pmatrix},$$

$$\tilde{f}(y) = f(y_1^2, \ldots, y_n^2), \quad \nabla_y \tilde{f}(y) = J(y)\nabla f(x)$$

$$\frac{\partial \tilde{f}(y)}{\partial y_i} = 2y_i \frac{\partial f}{\partial x_i}(y_1^2, \ldots, y_n^2)$$

Projection and projected gradient methods

loop of gradient descent

$x^*_{k+1} = x_k - \alpha_k \nabla f(x_k)$

$x_{k+1} = \mathbb{P}_K(x^*_{k+1})$ projection of x^*_{k+1} on K

Inequality constraints ⇔ Belong to a non empty closed convex set $K \subset \mathbb{R}^n$. (not necessarily bounded)

$$\inf_{x \in K} f(x)$$

Projection and projected gradient methods

Inequality constraints $\Leftrightarrow$ Belong to a non empty closed convex set $K \subset \mathbb{R}^n$.

$$\inf_{x \in K} f(x)$$

Theorem

Necessary local optimality condition Let $f : \mathbb{R}^n \to \mathbb{R}$ a differentiable function and K a convex non empty subset of $\mathbb{R}^n$. Let $x^\star$ a local minimizer of f in K then

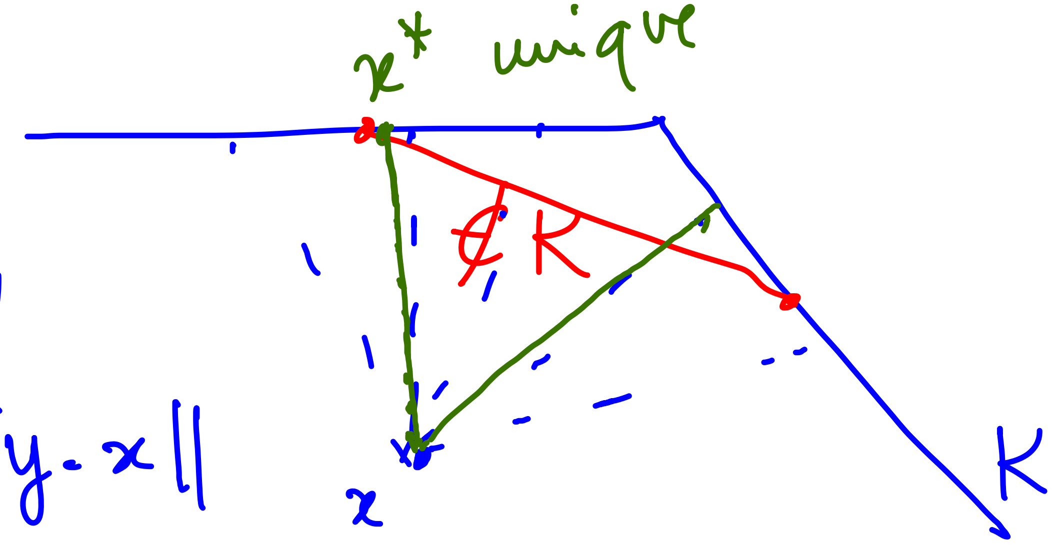
$$\langle \nabla f(x^\star), x - x^\star \rangle \geq 0, \forall x \in K$$

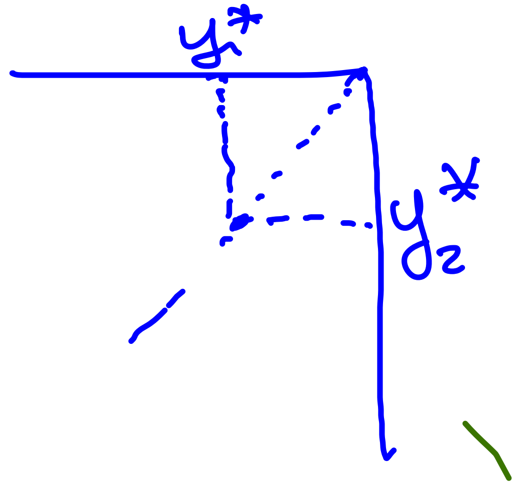
necessary

sufficient

If moreover f is convex the condition becomes ~~necessary~~, ie any point $x^\star$ satisfying it is a global minimum of f on K.

Counter exemple   K not convex set





$$\|P_K(x) - x\| = \inf_{y \in K} \|y - x\|$$
$$= \inf f_x(y)$$

f_x coercive and strictly convex still holds

$\exists\, y^*$ solution

Definition/proposition of the projection on a convex set

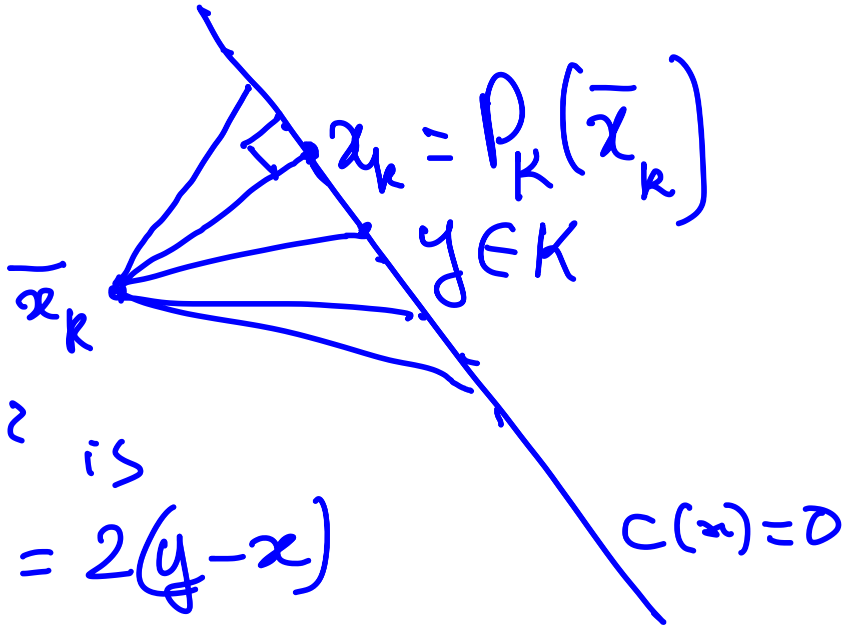
orthogonal projection:

$$f_x(y^*) = \inf_{y \in K} f_x(x) = p^*$$

p^* is global and unique
if K is convex then y^* is unique.

Let K a non empty closed convex set in $\mathbb{R}^n$. The projection of a point $x \in \mathbb{R}^n$ on K, denoted $P_K(x)$, is defined as the unique solution

$$\|P_K(x) - x\|^2 = \inf_{y \in K} \|x - y\|_2^2 = f_x(y).$$



Why does $P_K(x)$ $\exists!$?

for fixed x, $f_x(y) = \|y - x\|^2$ is
- strictly convex, $\nabla f_x(y) = 2(y - x)$
- $f_x(y)$ concave $\Rightarrow \exists$ minimum.

Definition/proposition of the projection on a convex set

Let K a non empty closed convex set in $\mathbb{R}^n$. The projection of a point $x \in \mathbb{R}^n$ on K, denoted $P_K(x)$, is defined as the unique solution

$f_x(y) = \|y - x\|^2$

$$\inf_{y \in K} \|x - y\|_2^2. = \inf_{y \in K} f_x(y)$$

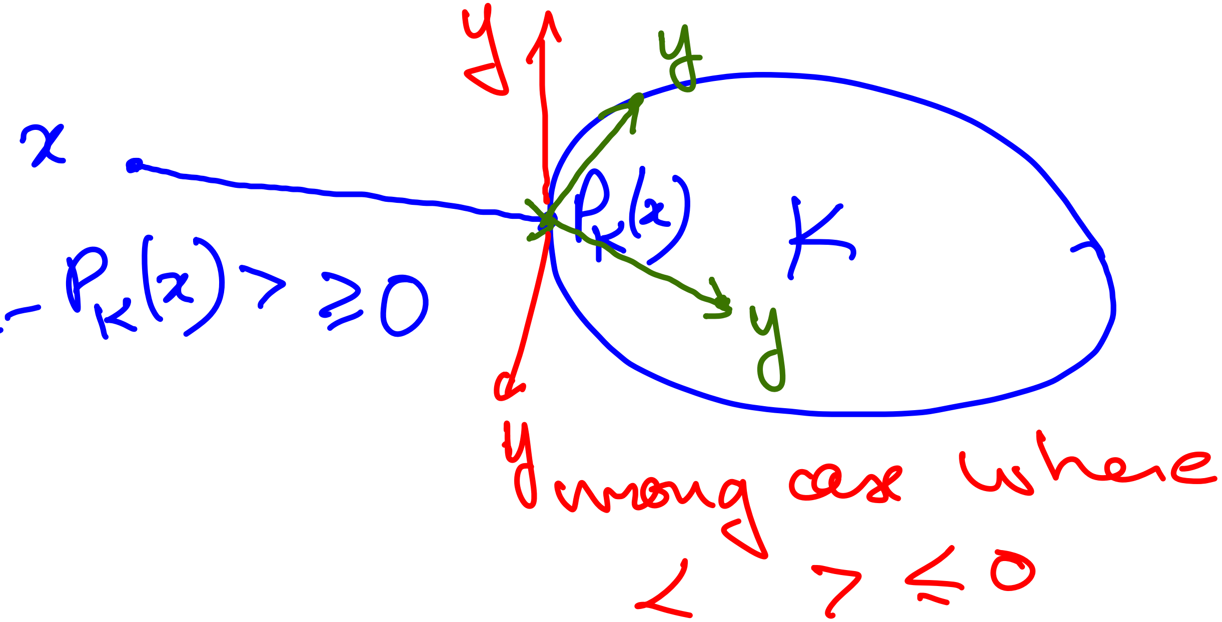
Furthermore $P_K(x)$ is the only point in K satisfying

$$\langle P_K(x) - x, y - P_K(x) \rangle \geq 0, \quad \forall y \in K$$

Proof

- Existence of $P_K(x)$ as minimum on a closed set of $f(y) = \|x - y\|^2$ coercive (Theorem 1.1)
-
-

$$\langle P_K(x) - x,\ y - P_K(x) \rangle \geq 0$$



Proof

- Existence of $P_K(x)$ as minimum on a closed set of $f(y) = \|x - y\|^2$ coercive (Theorem 1.1)
- Unicity : $f(y)$ is convex (Hessien $=2I_{\mathbb{R}^n}$) therefore unique minimum
-

Proof

- Existence of $P_K(x)$ as minimum on a closed set of $f(y) = \|x - y\|^2$ coercive (Theorem 1.1)
- Unicity : $f(y)$ is convex (Hessien $=2I_{\mathbb{R}^n}$) therefore unique minimum
- Equivalence between the two properties:
  $\Rightarrow$ for $y \in K$ and $\theta \in [0, 1]$, $\theta y + (1 - \theta)P_K(x) \in K$
  - Develop $\|x - P_K(x)\|^2 \leq \|\theta y + (1 - \theta)P_K(x) - x\|^2$
  - Simplify θ
  - Do $\theta = 0$

Proof

- Existence of $P_K(x)$ as minimum on a closed set of $f(y) = \|x - y\|^2$ coercive (Theorem 1.1)
- Unicity : $f(y)$ is convex (Hessien $=2I_{\mathbb{R}^n}$) therefore unique minimum
- Equivalence between the two properties:
  $\Rightarrow$ for $y \in K$ and $\theta \in [0,1]$, $\theta y + (1-\theta)P_K(x) \in K$
    - Develop $\|x - P_K(x)\|^2 \leq \|\theta y + (1-\theta)P_K(x) - x\|^2$
    - Simplify θ
    - Do $\theta = 0$

  $\Leftarrow$ for $y \in K$ develop

$$
\begin{aligned}
\|x - y\|^2 &= \|x - P_K(x) + P_K(x) - y\|^2 \\
&= \|x - P_K(x)\|^2 + \|P_K(x) - y\|^2 + 2\langle x - P_K(x), P_K(x) - y\rangle \\
&\geq \|x - P_K(x)\|^2
\end{aligned}
$$

Particular cases of projection on a convex.

Orthogonal projection of x on a subset K

$$P_K(x) := \operatorname*{argmin}_{y \in K} \|x - y\|.$$

Specific cases that can be solved explicitly

Projection on an intersection of half spaces

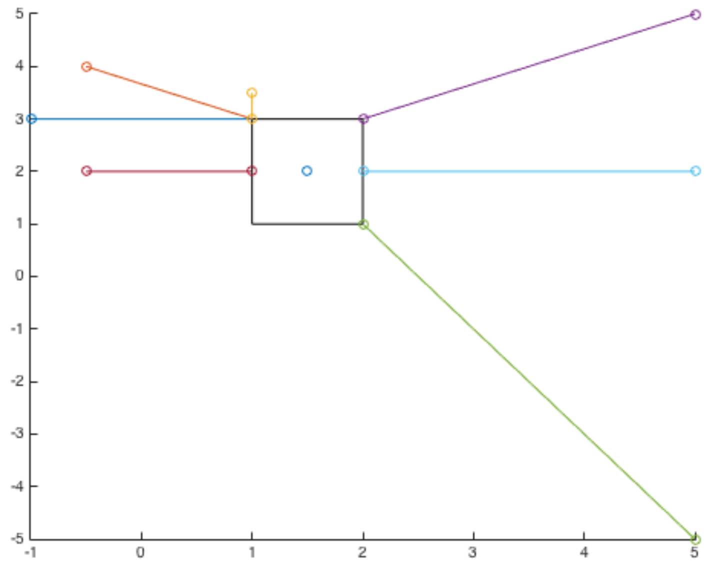
Supposons que $K = \{x \in \mathbb{R}^n, x_i \geq a_i, i \in I, x_j \leq b_j, j \in J\}$ avec $I, J \subset \{1, \ldots, n\}$. On a alors

$$P_K(x)_i = \begin{cases} \max(a_i, x_i), \text{ pour } i \in I \backslash J \\ \min(b_i, x_i), \text{ pour } i \in J \backslash I \\ \min(b_i, \max(a_i, x_i)), \text{ pour } i \in I \cap J \end{cases}$$

Projection on an intersection of half spaces

Supposons que $K = \{x \in \mathbb{R}^n, x_i \geq a_i, i \in I, x_j \leq b_j, j \in J\}$ avec $I, J \subset \{1, \ldots, n\}$. On a alors

$$P_K(x)_i = \begin{cases} \max(a_i, x_i), \text{ pour } i \in I \backslash J \\ \min(b_i, x_i), \text{ pour } i \in J \backslash I \\ \min(b_i, \max(a_i, x_i)), \text{ pour } i \in I \cap J \end{cases}$$



Projection on a line $(A, \vec{v})$

Projection on a line $(A, \vec{v})$

Euler inequation

Let $f : K \subset E \to \mathbb{R}$, where K is a convex included in E, a Hilbert space. We suppose that f is differentiable in $x^\star \in K$. If $x^\star$ is a local minimum of f over K, then $x^\star$ satisfies the Euler inequality:

$$Df(x^\star)(y - x^\star) \geq 0, \forall y \in K.$$

Algorithm of the projected gradient

Data: Function f, convex K, step $(\alpha_k)_{k\geq 0}$, tolerance τ, max number of iterations $k_{\max}$

Result: $\min_{x\in K} f(x)$

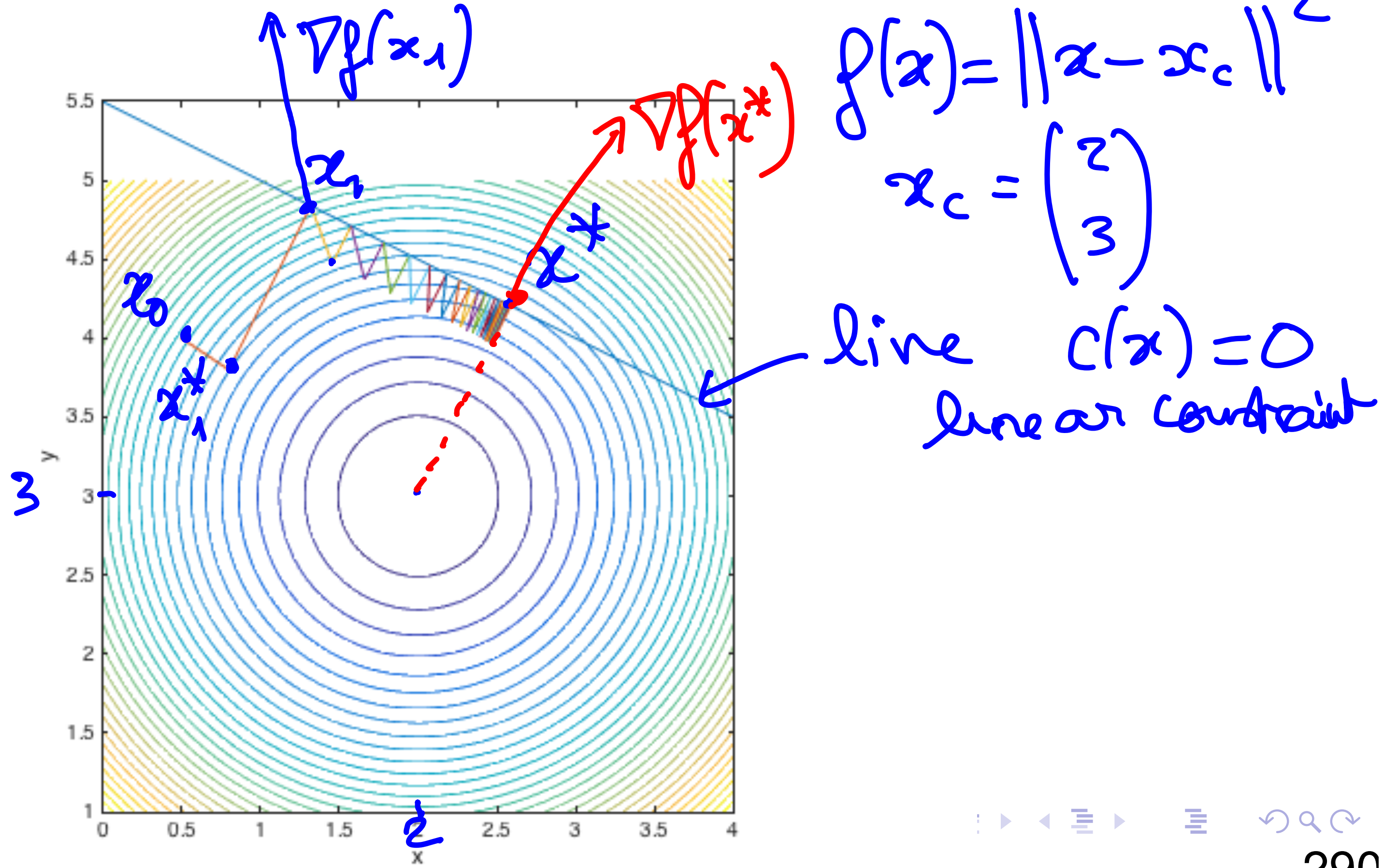
Initialisation : choice of $x_0 \in \mathbb{R}^n$

while $\|x^{k+1} - x^k\| \geq \tau$ *or* $k < k_{\max}$ **do**

  Solve $x^{k+1} = P_K(x^k - \alpha_k \nabla f(x^k))$, $k \leftarrow k+1$

end

$x^\star = x_k$



Convergence of projected gradient algorithm

Theorem

Let f differentiable sur $\mathbb{R}^n$ and $K \in \mathbb{R}^n$ closed non empty convex subset. Denote by x_k the current solution of the projected gradient algorithm and

$$d(\alpha) = P_K(x_k - \alpha \nabla f(x_k)) - x_k$$

If $d(\alpha) \neq 0$ then $d(\alpha)$ is a descent direction $\forall \alpha > 0$

Proof

Let $\alpha > 0$ fixed. Suppose : $d(\alpha) = p_K(x_k - \alpha \nabla f(x_k)) - x_k \neq 0$.
$d(\alpha)$ descent direction f in $x_k \Leftrightarrow \langle \nabla f(x_k), d(\alpha) \rangle < 0$

Proof

Let $\alpha > 0$ fixed. Suppose : $d(\alpha) = p_K(x_k - \alpha \nabla f(x_k)) - x_k \neq 0$.
$d(\alpha)$ descent direction f in $x_k \Leftrightarrow \langle \nabla f(x_k), d(\alpha) \rangle < 0$
Indeed $\forall y \in K$,
$\Leftrightarrow \langle P_k(x_k - \alpha \nabla f(x_k)) - (x_k - \alpha \nabla f(x_k)), y - P_K(x_k - \alpha \nabla f(x_k)) \rangle \geq 0$
Therefore for all $y \in K$ $\langle d(\alpha) + \alpha \nabla f(x_k)), y - x_k - d(\alpha) \rangle \geq 0$

Proof

Let $\alpha > 0$ fixed. Suppose : $d(\alpha) = p_K(x_k - \alpha \nabla f(x_k)) - x_k \neq 0$.
$d(\alpha)$ descent direction f in $x_k \Leftrightarrow \langle \nabla f(x_k), d(\alpha) \rangle < 0$
Indeed $\forall y \in K$,
$\Leftrightarrow \langle P_k(x_k - \alpha \nabla f(x_k)) - (x_k - \alpha \nabla f(x_k)), y - P_K(x_k - \alpha \nabla f(x_k)) \rangle \geq 0$
Therefore for all $y \in K$ $\langle d(\alpha) + \alpha \nabla f(x_k)), y - x_k - d(\alpha) \rangle \geq 0$
Since $x_k \in K$ choose $y = x_k$ on a
$\langle d(\alpha) + \alpha \nabla f(x_k), d(\alpha) \rangle \leq 0$ or $\alpha \langle \nabla f(x_k), d(\alpha) \rangle \leq -\|d(\alpha)\|^2 < 0$

Convergence of projected gradient algorithm

Theorem

If f is differentiable, α-elliptic and its gradient is C-Lipschitzien, the projected gradient algorithm converges towards $x^\star$ quand $k \to \infty$ for $(\alpha_k)_{k\geq 0}$ sufficiently small: $\alpha_k \leq \dfrac{2\alpha}{C^2}$ où α and C are two constants such that

$$
\begin{aligned}
(\nabla f(x) - \nabla f(y), x - y) &\geq \alpha \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n, \\
\|\nabla f(x) - \nabla f(y)\| &\leq C \|x - y\|, \quad \forall x, y \in \mathbb{R}^n.
\end{aligned}
$$

Proof

Compose two contracting applications

- $x \to x - \alpha \nabla f(x)$ with conditions on α_k (See unconstrained gradient convergence theorem)

Proof

Compose two contracting applications

- $x \to x - \alpha \nabla f(x)$ with conditions on α_k (See unconstrained gradient convergence theorem)
- $x \to P_K(x)$. indeed

$$
\begin{aligned}
\|x-y\|^2 &= \|x-y-(P_K(x)-P_K(y))+(P_K(x)-P_K(y))\|^2 \\
&= \|x-y-(P_K(x)-P_K(y))\|^2 + \\
&\quad \langle x-y-(P_K(x)-P_K(y)), P_K(x)-P_K(y)\rangle + \|P_K(x)-P_K(y)\|^2 \\
&= \|x-y-(P_K(x)-P_K(y))\|^2 \\
&\quad + \langle x-P_K(x), P_K(x)-P_K(y)\rangle \\
&\quad + \langle -y+P_K(y)), P_K(x)-P_K(y)\rangle + \|P_K(x)-P_K(y)\|^2
\end{aligned}
$$

but $\langle x-P_K(x), P_K(x)-P_K(y)\rangle \geq 0$ car $P_K(y) \in K$
Therefore

$$
\begin{aligned}
\|x-y\|^2 &\geq \|x-y-(P_K(x)-P_K(y))\|^2 + \|P_K(x)-P_K(y)\|^2 \\
&\geq \|P_K(x)-P_K(y)\|^2
\end{aligned}
$$