Analyse du problème

Michel Bierlaire

michel.bierlaire@epfl.ch

Laboratoire Transport et Mobilité

EPFL - ENAC - TRANSP-OR

Définition du problème

$$\min_{x \in \mathbb{R}^n} f(x)$$

sous contraintes

$$\begin{array}{rcl} h(x) & = & 0, \\ g(x) & \leq & 0, \\ x & \in & X. \end{array}$$

- $f : \mathbb{R}^n \to \mathbb{R}$, $n > 0$
- $h : \mathbb{R}^n \to \mathbb{R}^m$, $m \geq 0$
- $g : \mathbb{R}^n \to \mathbb{R}^p$, $p \geq 0$
- $X \subseteq \mathbb{R}^n$ ensemble convexe

Définition du problème

Minimum local

Appelons $Y = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0$ *et* $x \in X\}$ *l'ensemble des vecteurs vérifiant toutes les contraintes.*

Le vecteur $x^* \in Y$ *est appelé* minimum local *s'il existe* $\varepsilon > 0$ *tel que*

$$f(x^*) \leq f(x) \quad \forall x \in Y \text{ tel que } \|x - x^*\| < \varepsilon.$$

Définition du problème

Minimum local strict

Appelons $Y = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0$ *et* $x \in X\}$ *l'ensemble des vecteurs vérifiant toutes les contraintes.*

Le vecteur $x^* \in Y$ *est appelé* minimum local strict *s'il existe* $\varepsilon > 0$ *tel que*

$$f(x^*) < f(x) \quad \forall x \in Y, x \neq x^* \text{ tel que } \|x - x^*\| < \varepsilon.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Définition du problème

Minimum global

Appelons $Y = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0$ *et* $x \in X\}$ *l'ensemble des vecteurs vérifiant toutes les contraintes.*

Le vecteur $x^* \in Y$ *est appelé* minimum global *si*

$$f(x^*) \leq f(x) \quad \forall x \in Y.$$

Définition du problème

Minimum global strict

Appelons $Y = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0$ *et* $x \in X\}$ *l'ensemble des vecteurs vérifiant toutes les contraintes.*

Le vecteur $x^* \in Y$ *est appelé* minimum global strict *si*

$$f(x^*) < f(x) \quad \forall x \in Y, x \neq x^*.$$

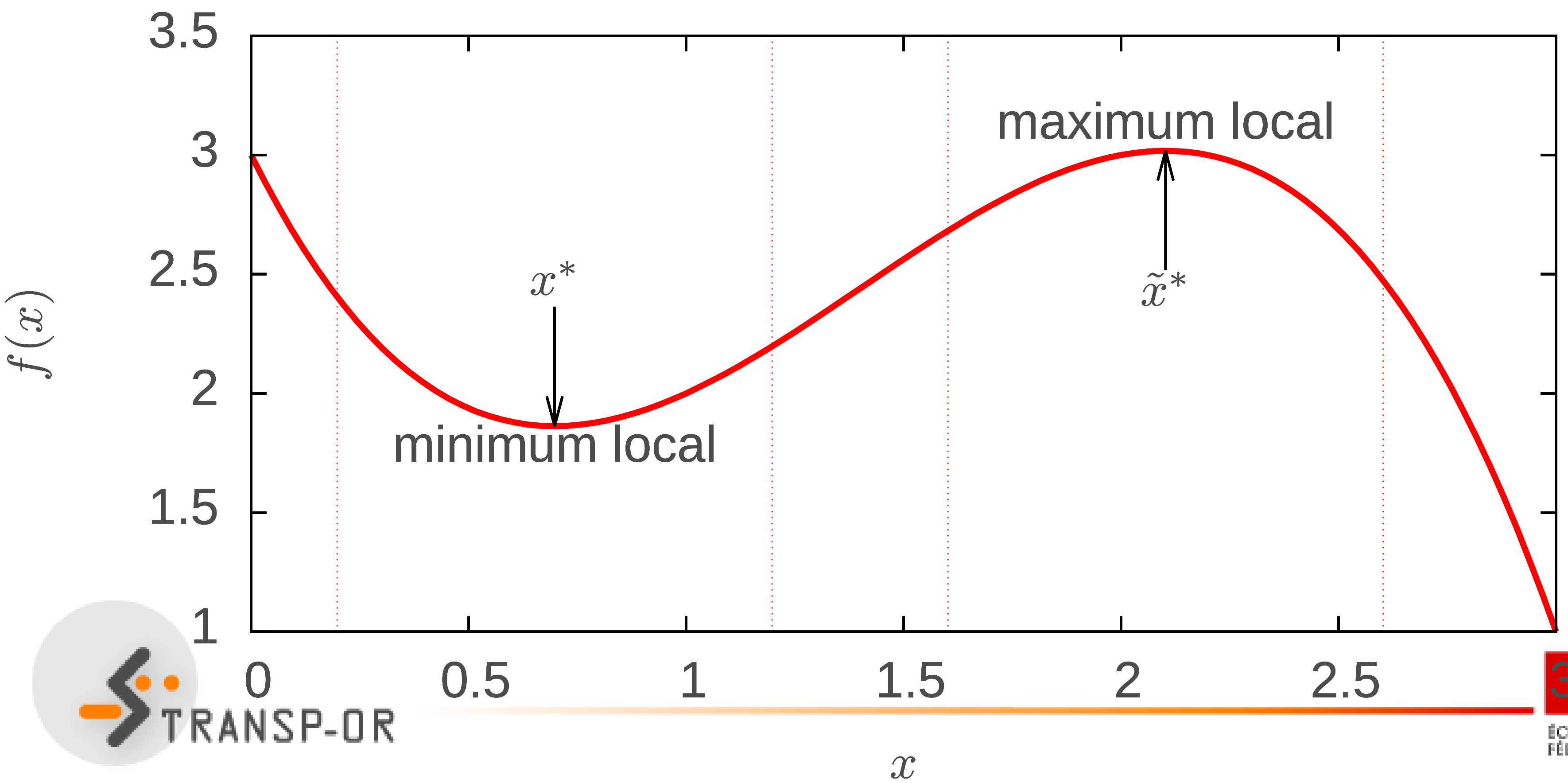
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Définition du problème

Exemple :

$$f(x) = -\frac{5}{6}x^3 + \frac{7}{2}x^2 - \frac{11}{3}x + 3.$$



Définition du problème

Attention : $\min_{x \in \mathbb{R}^n} f(x) = x$ n'a pas de solution

Fonction bornée inférieurement

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *et un ensemble* $Y \subseteq \mathbb{R}^n$*.*

La fonction f *est* bornée *inférieurement sur* Y *s'il existe un réel* M *tel que*

$$f(x) \geq M \quad \forall x \in Y.$$

Définition du problème

Infimum

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *bornée inférieurement sur un ensemble* $Y \subseteq \mathbb{R}^n$. *L'infimum de* f *sur* Y *est noté*

$$\inf_{y \in Y} f(y)$$

et est tel que

$$\inf_{y \in Y} f(y) \leq f(x) \quad \forall x \in Y$$

et

$$\forall M > \inf_{y \in Y} f(y), \exists x \in Y \text{ tel que } f(x) < M.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Infimum : exemple

Soit $f(x) = e^x$ et $Y = \mathbb{R}$. Nous avons

$$\inf_{y \in Y} f(y) = 0.$$

En effet,

$$0 = \inf_{y \in Y} f(y) \leq f(x) = e^x \quad \forall x \in \mathbb{R}.$$

Prenons $M > 0$ arbitraire, et $x = \ln M/2$. Alors,

$$f(x) = \frac{M}{2} < M,$$

et la définition est vérifiée.

Définition du problème

Théorème: **Théorème de Weierstrass** *Soit $Y \subset \mathbb{R}^n$ un sous-ensemble fermé et non-vide de $\mathbb{R}^n$, et soit $f : Y \longrightarrow \mathbb{R}$ une fonction semi-continue inférieurement en chaque point de Y.*
Alors, si Y est compact ou si f est coercitive, il existe $x^* \in Y$ tel que

$$f(x^*) = \inf_{y \in Y} f(y).$$

(p. 26)

Rappels

Semi-continuité

Soit $X \subseteq \mathbb{R}^n$*, et soit* $f : X \longrightarrow \mathbb{R}$ *une fonction à valeurs réelles.* f *est dite* semi-continue inférieurement *en* $x \in X$ *si pour toute suite* $(x_k)_k$ *d'éléments de* X *convergeant vers* x*, on a*

$$f(x) \leq \liminf_{k \to \infty} f(x_k)$$

Rappel

Fonction coercitive

Soit $X \subseteq \mathbb{R}^n$, *et soit* $f : X \longrightarrow \mathbb{R}$ *une fonction à valeurs réelles.* f *est dite* coercitive *si pour toute suite* $(x_k)_k$ *d'éléments de* X *telle que* $\|x_k\| \to \infty$ *pour une norme quelconque, on a*

$$\lim_{k\to\infty} f(x_k) = \infty$$

Définition du problème

Point intérieur

Soit $Y \subset \mathbb{R}^n$ et $y \in Y$. On dit que y est intérieur à Y s'il existe un voisinage de y contenu dans Y. D'une manière équivalente, y est intérieur à Y s'il existe $\varepsilon > 0$ tel que

$$z \in Y \quad \forall z \text{ tel que } \|z - y\| \leq \varepsilon.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Définition du problème

Théorème: **Solution optimale à l'intérieur des contraintes** *Soit x^* un minimum local.*
Si x^* est un point intérieur à l'ensemble des contraintes,
alors x^* est un minimum local du problème sans contrainte

$$\min_{x \in \mathbb{R}^n} f(x)$$

(p. 27)

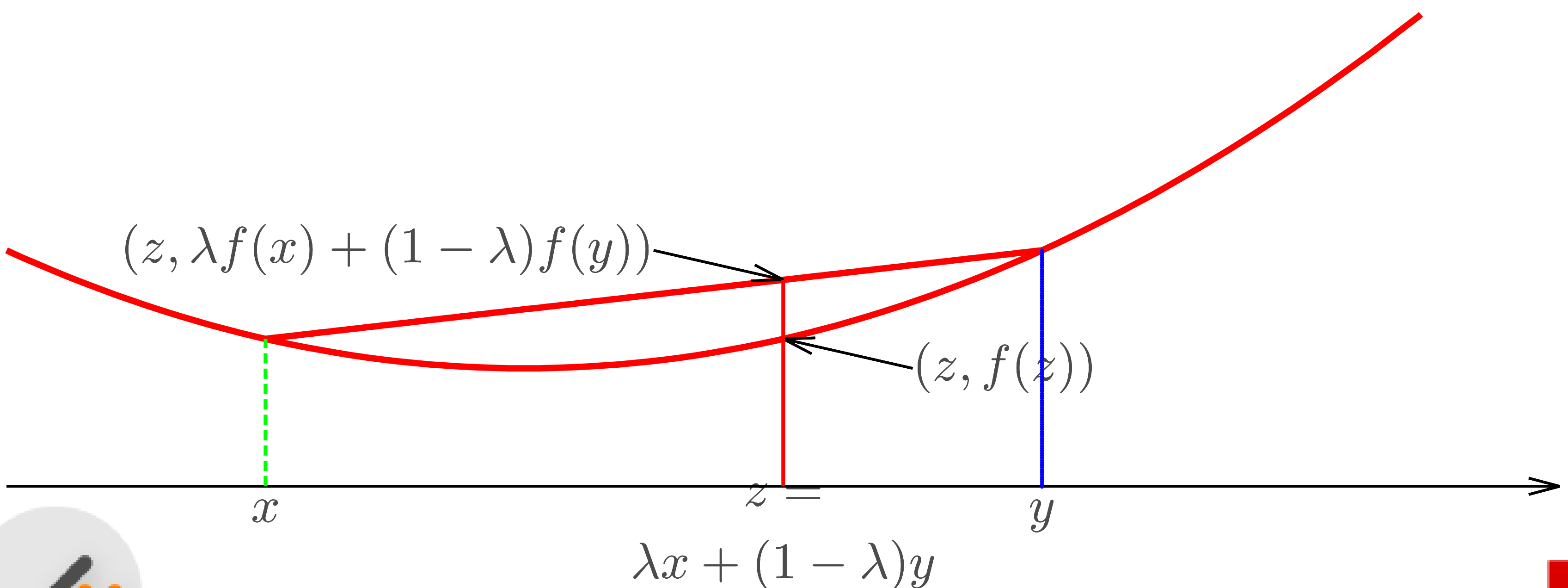
Fonction objectif

Fonction convexe

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *est dite convexe si,*

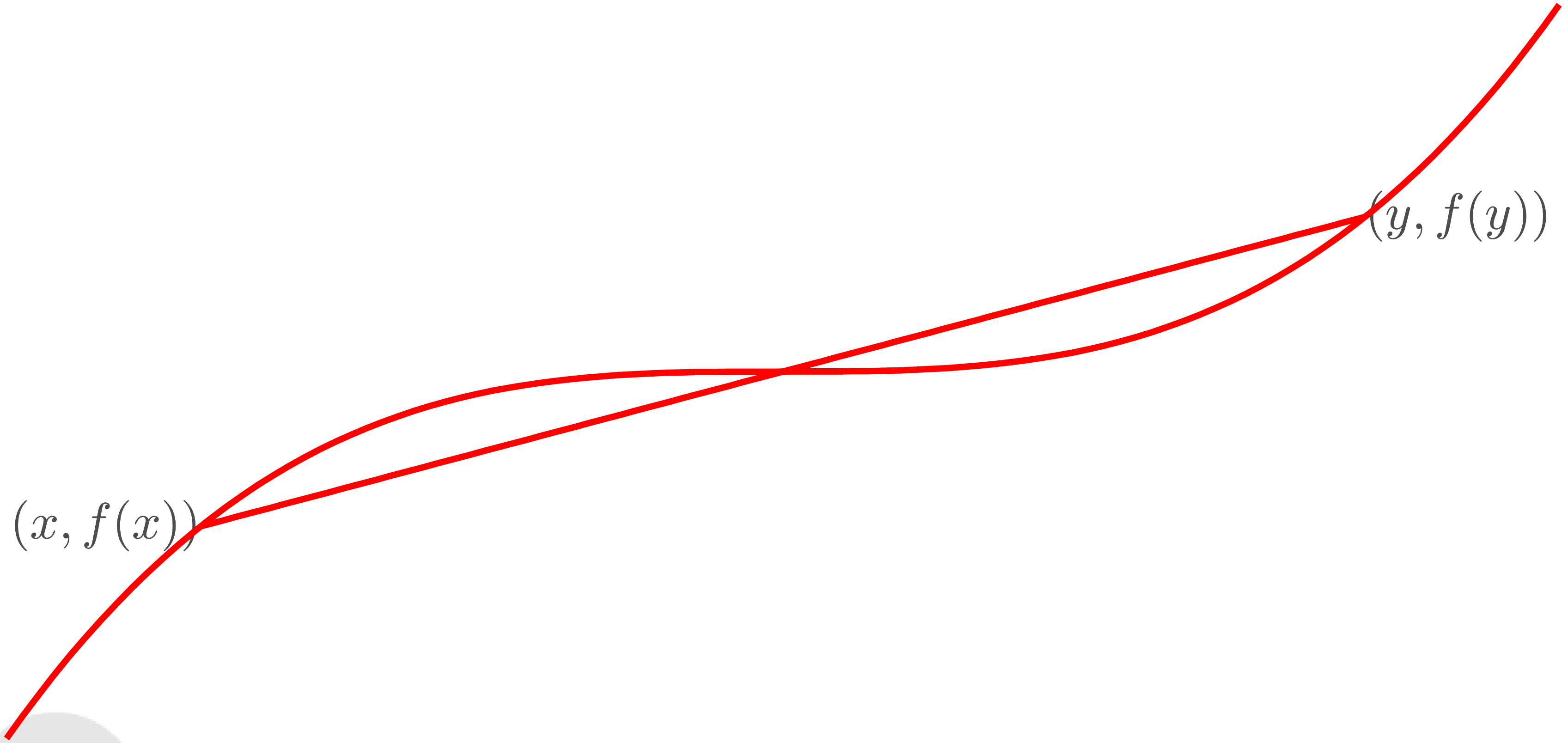
pour tout $x, y \in \mathbb{R}^n$, *on a*

$$f\left(\lambda x + (1-\lambda)y\right) \leq \lambda f(x) + (1-\lambda) f(y) \quad \forall \lambda \in [0,1]$$



Fonction objectif

Fonction non convexe :



Fonction objectif

Fonction strictement convexe

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *est dite strictement convexe si, pour tout* $x, y \in \mathbb{R}^n$*, on a*

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) \quad \forall x \neq y, \forall \lambda \in]0,1[$$

Fonction concave

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *est dite concave si* $-f$ *est une fonction convexe, c'est-à-dire si, pour tout* $x, y \in \mathbb{R}^n$*, on a*

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y) \quad \forall \lambda \in [0,1].$$

Différentiabilité : 1er ordre

Dérivée partielle

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction continue.*

La fonction notée $\nabla_i f(x) : \mathbb{R}^n \longrightarrow \mathbb{R}$, *également notée* $\frac{\partial f(x)}{\partial x_i}$ *est appelée* i*ème dérivée partielle de* f *et est définie par*

$$\lim_{\alpha \to 0} \frac{f(x_1, \ldots, x_i + \alpha, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{\alpha}.$$

Cette limite peut ne pas exister.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Différentiabilité : 1er ordre

Gradient

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction différentiable.*

La fonction notée $\nabla f(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *est appelée le gradient de* f *et est définie par*

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}.$$

Gradient : exemple

Soit $f(x_1, x_2, x_3) = e^{x_1} + x_1^2 x_3 - x_1 x_2 x_3$. Le gradient de f est donné par

$$\nabla f(x_1, x_2, x_3) = \left(\begin{array}{c} e^{x_1} + 2x_1 x_3 - x_2 x_3 \\ -x_1 x_3 \\ x_1^2 - x_1 x_2 \end{array} \right).$$

Dérivée directionnelle

Dérivée directionnelle

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction continue. Soient* $x \in \mathbb{R}^n$ *et* $d \in \mathbb{R}^n$.

La dérivée directionnelle de f *en* x *dans la direction* d *est donnée par*

$$\lim_{\alpha \xrightarrow{>} 0} \frac{f(x+\alpha d) - f(x)}{\alpha},$$

si la limite existe.

De plus, lorsque le gradient existe, la dérivée directionnelle est le produit scalaire entre le gradient de f *et la direction* d, *i.e.*

$$\nabla f(x)^T d.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Différentiabilité

Fonction différentiable

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction continue.*

Si, pour tout $d \in \mathbb{R}^n$*, la dérivée directionnelle de* f *dans la direction* d *existe,*

alors la fonction f *est dite* différentiable.

Dérivée directionnelle

Soit $f(x_1, x_2, x_3) = e^{x_1} + x_1^2 x_3 - x_1 x_2 x_3$, et soit

$$d = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}.$$

La dérivée directionnelle de f dans la direction d est

$$\begin{aligned}&(d_1\ d_2\ d_3)\nabla f(x_1, x_2, x_3) = \\ &\quad d_1(e^{x_1} + 2x_1x_3 - x_2x_3) - d_2x_1x_3 + d_3(x_1^2 - x_1x_2)\end{aligned}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Dérivée directionnelle

Dérivées partielles = dérivées directionnelles dans la direction des axes de coordonnées

$$\frac{\partial f(x)}{\partial x_i} = \nabla f(x)^T e_i$$

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \longleftarrow i\text{ème ligne}$$

Dérivée directionnelle

Direction de descente

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction différentiable.*

Soient $x, d \in \mathbb{R}^n$*.*

La direction d *est une* direction de descente *en* x *si*

$$d^T \nabla f(x) < 0.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Rappels

Théorème: **Taylor au premier ordre** *Soit* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction différentiable sur une sphère ouverte* S *centrée en* x*. Alors,*

- *pour tout* d *tel que* $x + d \in S$*, on a*

$$f(x+d) = f(x) + d^T \nabla f(x) + o(\|d\|),$$

- *pour tout* d *tel que* $x + d \in S$*, il existe* $\alpha \in [0,1]$ *tel que*

$$f(x+d) = f(x) + d^T \nabla f(x + \alpha d).$$

(p. 499)

Rappels

Notation de Landau $o(\cdot)$

Soient f et g deux fonctions de $\mathbb{R} \longrightarrow \mathbb{R}$, avec $f(x) \neq 0$, $\forall x$. La notation de Landau $g(x) = o(f(x))$ signifie que

$$\lim_{x \to 0} \frac{g(x)}{f(x)} = 0.$$

Par abus de language, on dit que $g(x)$ tend vers zéro plus vite que $f(x)$.

Direction de descente

Théorème: **Direction de descente** *Soit* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction différentiable.*
Soient $x \in \mathbb{R}^n$ *tel que* $\nabla f(x) \neq 0$ *et* $d \in \mathbb{R}^n$*.*
Si d *est une direction de descente,*
alors il existe $\eta > 0$ *tel que*

$$f(x + \alpha d) < f(x) \quad \forall 0 < \alpha \leq \eta.$$

De plus, pour tout $\beta < 1$*, il existe* $\widehat{\eta} > 0$ *tel que*

$$f(x + \alpha d) < f(x) + \alpha\beta\nabla f(x)^T d,$$

pour tout $0 < \alpha \leq \widehat{\eta}$*.(p. 36)*

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Idées de la preuve

Taylor:

$$f(x + \alpha d) = f(x) + \alpha d^T \nabla f(x) + o(\alpha \|d\|)$$

- $\alpha d^T \nabla f(x) < 0$ par hypothèse.
- $o(\alpha \|d\|)$ négligeable si α est suffisamment petit.
- Donc $f(x + \alpha d) < f(x)$ si α est suffisamment petit.

Intuition pour le second résultat : la diminution est proportionnelle à la pente de la fonction.

Plus forte pente

Théorème: **Plus forte pente** *Soit* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction différentiable.*
Soient $x \in \mathbb{R}^n$*, et* $d^* = \nabla f(x)$*.*
Alors, pour tout $d \in \mathbb{R}^n$ *tel que* $\|d\| = \|\nabla f(x)\|$*, on a*

$$d^T \nabla f(x) \leq {d^*}^T \nabla f(x) = \nabla f(x)^T \nabla f(x),$$

et la direction du gradient est celle dans laquelle la fonction a la plus forte pente.
(p. 37)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Preuve

Soit d une direction quelconque. Nous avons

$$\begin{aligned} d^T \nabla f(x) &\leq \|d\| \, \|\nabla f(x)\| && \text{Cauchy-Schwartz} \\ &= \|\nabla f(x)\|^2 && \text{par hypothèse } \|d\| = \|\nabla f(x)\| \\ &= \nabla f(x)^T \nabla f(x) && \text{par déf. du produit scalaire} \\ &= {d^*}^T \nabla f(x) && \text{par déf. de } d^* \end{aligned}$$

Comme ${d^*}^T \nabla f(x) = \|\nabla f(x)\|^2 \geq 0$, la fonction est non décroissante dans la direction d^*, qui correspond bien à la plus forte pente.

Plus forte descente

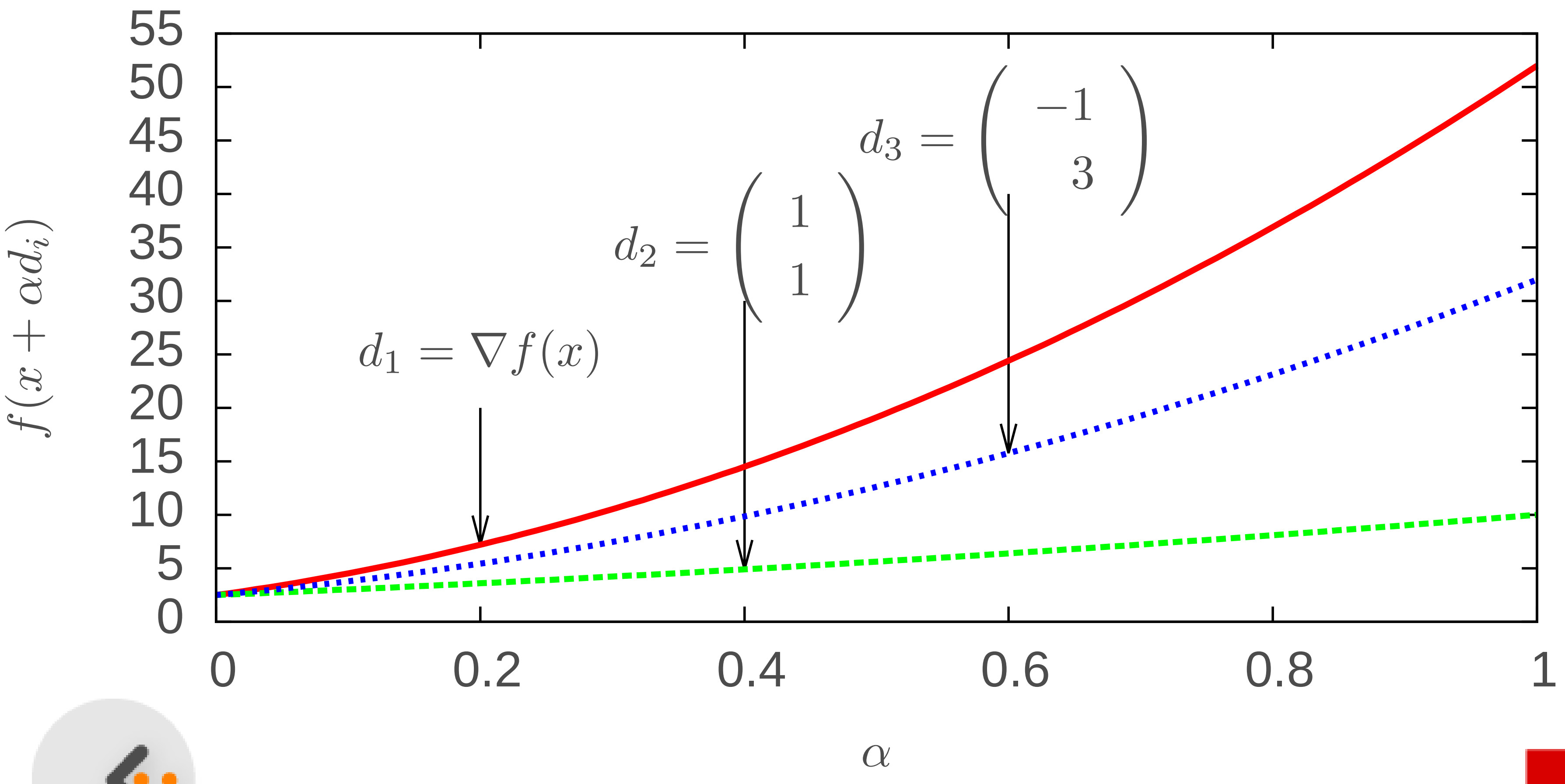
Théorème: **Plus forte descente** *Soit* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction différentiable.*
Soient $x \in \mathbb{R}^n$*, et* $d^* = -\nabla f(x)$*.*
Alors, pour tout $d \in \mathbb{R}^n$ *tel que* $\|d\| = \|\nabla f(x)\|$*, on a*

$$-\nabla f(x)^T \nabla f(x) = {d^*}^T \nabla f(x) \leq d^T \nabla f(x).$$

et la direction opposée au gradient est donc celle dans laquelle la fonction a la plus forte descente. (Immédiat - p. 38)

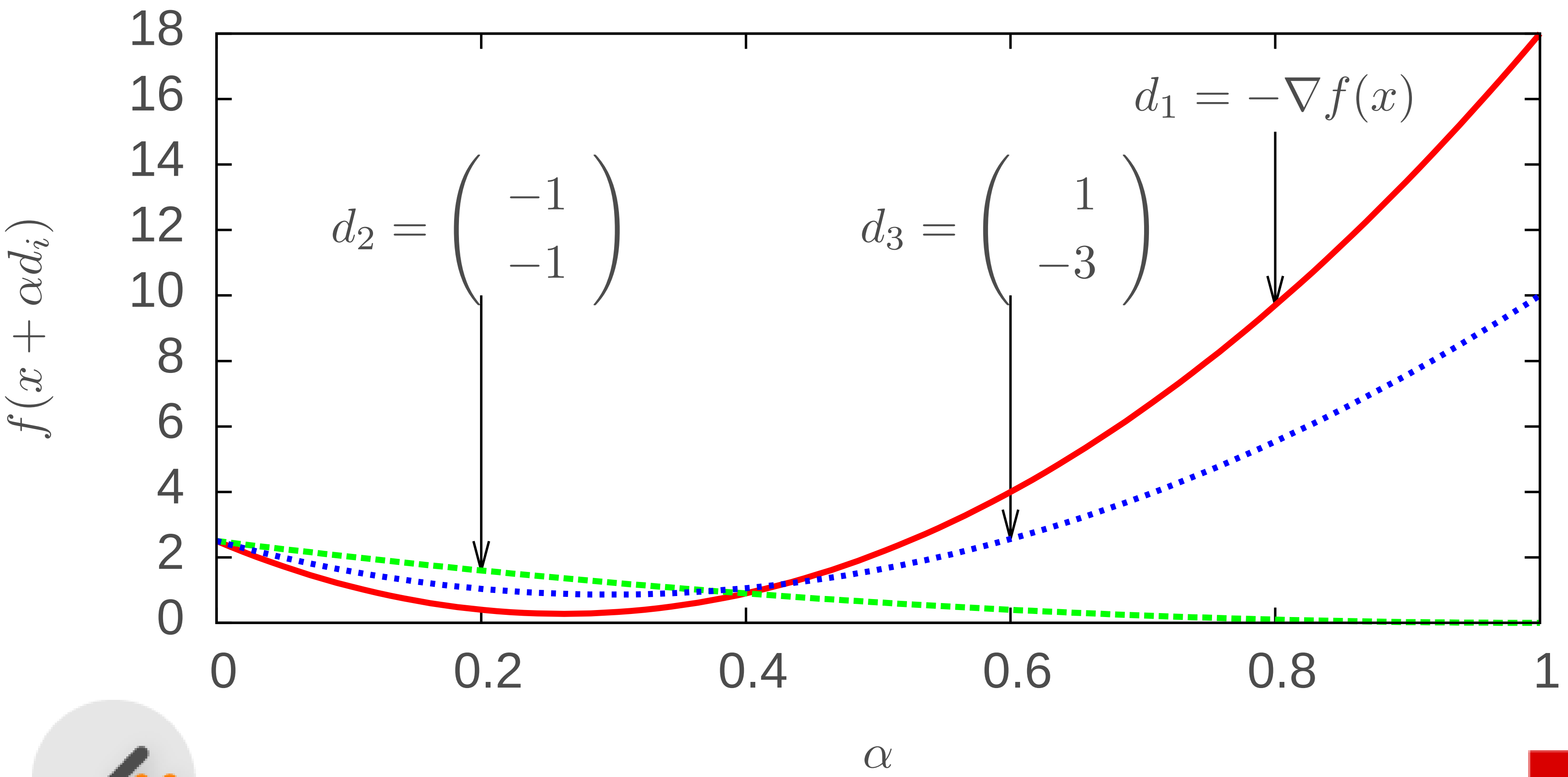
Plus forte pente

Pente de $\frac{1}{2}x_1^2 + 2x_2$ au point $(1,1)^T$



Plus forte descente

Pente de $\frac{1}{2}x_1^2 + 2x_2$ au point $(1,1)^T$



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Gradient

- Information sur la pente
- Information sur la convexité / concavité

Convexité

Théorème: **Convexité par le gradient** *Soit* $f : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction différentiable sur un ensemble convexe* X.
f *est convexe sur* X *si et seulement si*

$$f(y) - f(x) \geq (y-x)^T \nabla f(x), \quad \forall x, y \in X.$$

f *est strictement convexe sur* X *si et seulement si*

$$f(y) - f(x) > (y-x)^T \nabla f(x), \quad \forall x, y \in X.$$

(p. 40)

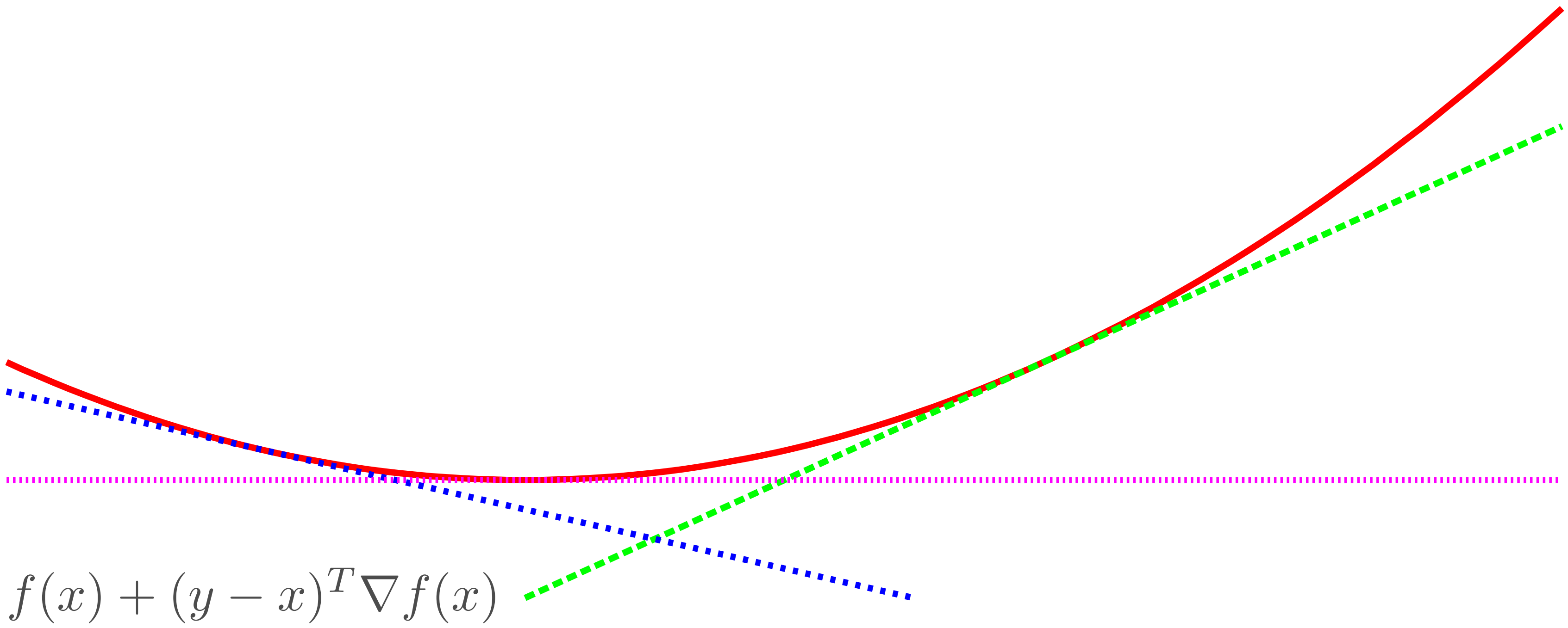
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Convexité par le gradient

Interprétation géométrique : le graphe de la fonction se trouve au dessus des hyperplans tangents

$$f(y) \geq f(x) + (y-x)^T \nabla f(x), \quad \forall x, y \in \mathbb{R}^n,$$



Fonctions de $\mathbb{R}^n \longrightarrow \mathbb{R}^m$

Matrice gradient

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *telle que* $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ *est différentiable, pour* $i = 1, \ldots, m$.

Dans ce cas, f *est différentiable, et la fonction* $\nabla f(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^{n \times m}$ *est appelée* matrice gradient, *et est définie par*

$$\nabla f(x) = \begin{pmatrix} | & & | \\ \nabla f_1(x) & \cdots & \nabla f_m(x) \\ | & & | \end{pmatrix}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Fonctions de $\mathbb{R}^n \longrightarrow \mathbb{R}^m$

$$\nabla f(x) \quad = \quad \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Fonctions de $\mathbb{R}^n \longrightarrow \mathbb{R}^m$

Matrice jacobienne

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$.

La fonction $J(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^{m \times n}$ *est appelée* matrice jacobienne, *et est définie par*

$$J(x) = \nabla f(x)^T = \begin{pmatrix} \text{———} & \nabla f_1(x)^T & \text{———} \\ & \vdots & \\ \text{———} & \nabla f_m(x)^T & \text{———} \end{pmatrix}.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Différentiabilité : le 2ième ordre

- Le gradient de f est une fonction de $\mathbb{R}^n \longrightarrow \mathbb{R}^n$
- On peut donc calculer sa matrice gradient dont les composantes sont

$$\frac{\partial \nabla_i f(x)}{\partial x_j} = \frac{\partial(\partial f(x)/\partial x_i)}{\partial x_j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}.$$

- Il s'agit de la matrice des dérivées secondes de f, ou *matrice hessienne*

Différentiabilité : le 2ième ordre

Matrice Hessienne

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction deux fois différentiable.*

La fonction notée $\nabla^2 f(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^{n\times n}$ *est appelée la* matrice hessienne *ou* Hessien *de* f *et est définie par*

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}.$$

La matrice hessienne est toujours symétrique.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Différentiabilité : le 2ième ordre

Soit $f(x_1,x_2,x_3) = e^{x_1} + x_1^2 x_3 - x_1 x_2 x_3$. Le gradient de f est donné par

$$\nabla f(x_1,x_2,x_3) = \begin{pmatrix} e^{x_1} + 2x_1x_3 - x_2x_3 \\ -x_1x_3 \\ x_1^2 - x_1x_2 \end{pmatrix}.$$

Le Hessien de f est donné par

$$\nabla^2 f(x_1,x_2,x_3) = \begin{pmatrix} e^{x_1} + 2x_3 & -x_3 & 2x_1 - x_2 \\ -x_3 & 0 & -x_1 \\ 2x_1 - x_2 & -x_1 & 0 \end{pmatrix}.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Convexité par le hessien

Théorème: **Convexité par le Hessien** *Soit $f : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction deux fois différentiable sur un ensemble convexe X.*
Si $\nabla^2 f(x)$ est semi définie positive (resp. définie positive) pour tout x dans X, alors f est convexe (resp. strictement convexe).(p. 44)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Rappels

Matrice semi définie positive

La matrice carrée $A \in \mathbb{R}^{n \times n}$ *est dite* semi définie positive *lorsque*

$$x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n.$$

Si de plus A *est symétrique, alors aucune de ses valeurs propres ne sont négatives.*

Rappels

Matrice définie positive

La matrice carrée $A \in \mathbb{R}^{n \times n}$ *est dite* définie positive *lorsque*

$$x^T A x > 0 \quad \forall x \in \mathbb{R}^n, x \neq 0.$$

Si de plus A *est symétrique, toutes ses valeurs propres sont strictement positives.*

Rappels

Théorème: **Taylor au second ordre** *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction deux fois différentiable sur une sphère ouverte S centrée en x. Alors,*

- *pour tout d tel que $x + d \in S$, on a*

$$f(x+d) = f(x) + d^T \nabla f(x) + \frac{1}{2} d^T \nabla^2 f(x) d + o(\|d\|^2),$$

- *pour tout d tel que $x + d \in S$, il existe $\alpha \in [0, 1]$ tel que*

$$f(x+d) = f(x) + d^T \nabla f(x) + \frac{1}{2} d^T \nabla^2 f(x + \alpha d) d$$

(p. 499)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Analyse directionnelle

Analysons la fonction f dans la direction d à partir de x

$$g : \mathbb{R} \to \mathbb{R} : \alpha \rightsquigarrow f(x + \alpha d).$$

Règle de différentiation en chaîne :

$$g'(\alpha) = d^T \nabla f(x + \alpha d).$$

Notons que $g'(0)$ est la dérivée directionnelle. Nous avons aussi

$$g''(\alpha) = d^T \nabla^2 f(x + \alpha d) d.$$

La dérivée seconde informe sur la courbure

Analyse directionnelle

Courbure

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction deux fois différentiable. Soient* $x, d \in \mathbb{R}^n$*. La quantité*

$$\frac{d^T \nabla^2 f(x) d}{d^T d}$$

représente la courbure *de la fonction* f *en* x *dans la direction* d*.*

Linéarité

Fonction linéaire

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *est dite linéaire si elle s'écrit*

$$f(x) = c^T x = \sum_{i=1}^{n} c_i x_i,$$

où $c \in \mathbb{R}^n$ *est un vecteur constant, i.e. indépendant de* x.

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *est linéaire si chacune de ses composantes* $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $i = 1, \ldots, m$, *est linéaire.*

Dans ce cas, elle peut s'écrire

$$f(x) = Ax$$

où $A \in \mathbb{R}^{m \times n}$ *est une matrice de constantes.*

Linéarité

En optimisation, on utilise souvent le terme *linéaire* pour *affine*

Linéarité

Fonction affine

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *est dite affine si elle s'écrit*

$$f(x) = c^T x + d,$$

où $c \in \mathbb{R}^n$ *est un vecteur de constantes et* $d \in \mathbb{R}$*.*

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *est affine si chacune de ses composantes* $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $i = 1, \ldots, m$, *est affine.*

Dans ce cas, elle peut s'écrire

$$f(x) = Ax + b$$

où $A \in \mathbb{R}^{m \times n}$ *est une matrice et* $b \in \mathbb{R}^m$ *un vecteur.*

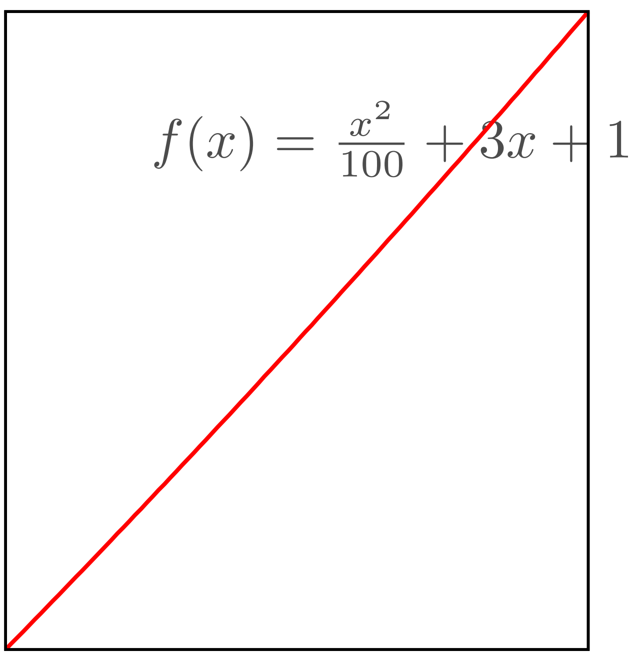
Linéarité

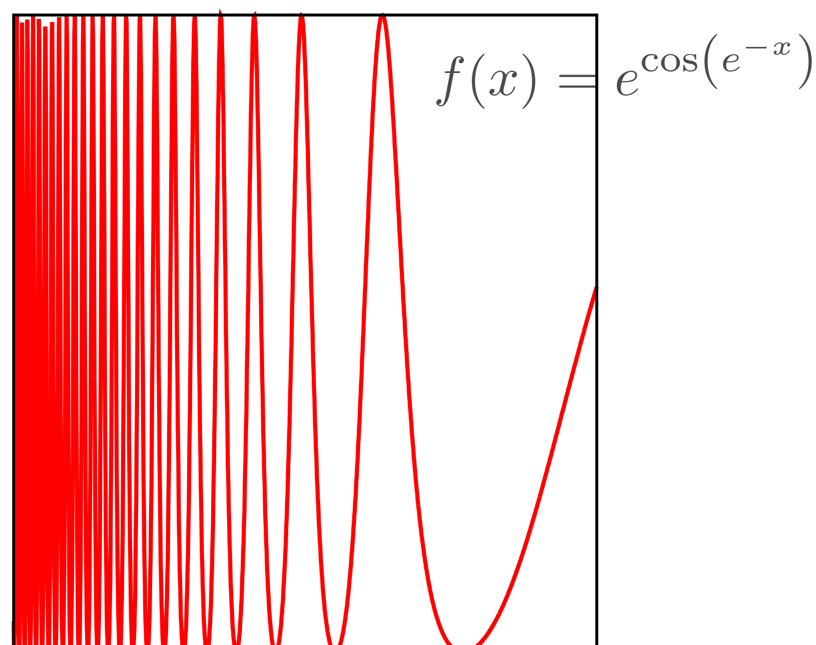
Minimiser $c^T x + d$ ou $c^T x$ est équivalent

Linéarité

Fonction non linéaire

Toute fonction qui n'est pas affine est dite non linéaire.





Il est important de caractériser la non linéarité

Linéarité

Continuité du gradient au sens de Lipschitz

Soit $f : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$.

La matrice gradient de la fonction est continue au sens de Lipschitz sur X s'il existe une constante $M > 0$ telle que, pour tout $x, y \in X$, on a

$$\|\nabla f(x) - \nabla f(y)\|_{n\times m} \leq M \|x - y\|_n,$$

où $\|\cdot\|_{n\times m}$ est une norme sur $\mathbb{R}^{n\times m}$ et $\|\cdot\|_n$ une norme sur $\mathbb{R}^n$.

La constante M est appelée constante de Lipschitz*.*

Linéarité

- M est une borne supérieure sur la courbure de la fonction.
- Si $M = 0$, la courbure est nulle et la fonction est linéaire.
- Cette constante est théorique, et il est très difficile d'en obtenir une valeur.

Fonction quadratique

Fonction quadratique

Une fonction $f : \mathbb{R}^n \to \mathbb{R}$ *sera dite quadratique si elle peut s'écrire*

$$f(x) = \frac{1}{2} x^T Q x + g^T x + c$$

où Q *est une matrice symétrique* $n \times n$, $g \in \mathbb{R}^n$ *et* $c \in \mathbb{R}$. *Nous avons*

$$\nabla f(x) = Qx + g \text{ et } \nabla^2 f(x) = Q.$$

Fonction quadratique

- L'hypothèse Q symétrique n'est pas restrictive. En effet,

$$x^TQx = \sum_{i=1}^{n}\sum_{j=1}^{n} Q_{ij}x_ix_j = \sum_{i=1}^{n}\sum_{j=i}^{n} \frac{1}{2}(Q_{ij}+Q_{ji})x_ix_j$$

- Soit Q' telle que $Q'_{ij} = Q'_{ji} = \frac{1}{2}(Q_{ij}+Q_{ji})$.
- Ainsi, $x^TQx = x^TQ'x$ avec Q' symétrique.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conditionnement et préconditionnement

Nombre de conditionnement

Soit $A \in \mathbb{R}^{n \times n}$ une matrice symétrique non singulière.

Le nombre de conditionnement de A est

$$\kappa(A) = \|A\| \|A^{-1}\|.$$

Conditionnement et préconditionnement

Nombre de conditionnement (suite)

Si la norme matricielle utilisée est la norme 2, nous avons

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n},$$

où σ_1 est la plus grande valeur singulière de A, et σ_n la plus petite. Par extension, nous considérerons que le nombre de conditionnement d'une matrice singulière est $+\infty$.

Si A est symétrique définie positive, les valeurs singulières de A sont ses valeurs propres.

Conditionnement et préconditionnement

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$
- Soit $x \in \mathbb{R}^n$
- Supposons $\nabla^2 f(x)$ définie positive
- Soient λ_1 sa plus grande valeur propre, et d_1 le vecteur propre associé.
- Comme $Ad_1 = \lambda_1 d_1$, nous avons

$$\lambda_1 = \frac{d_1^T A d_1}{d_1^T d_1}.$$

- λ_1 : courbure de f dans la direction d_1

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conditionnement et préconditionnement

Théorème: **Rayleigh-Ritz** *Soit* $A \in \mathbb{R}^{n\times n}$ *une matrice réelle symétrique. Soit* λ_1 *la plus grande valeur propre de* A *et* λ_n *la plus petite. Alors,*

$$\lambda_1 = \max_{x\neq 0} \frac{x^T A x}{x^T x},$$

et

$$\lambda_n = \min_{x\neq 0} \frac{x^T A x}{x^T x}.$$

(p. 500)

TRANSP-OR

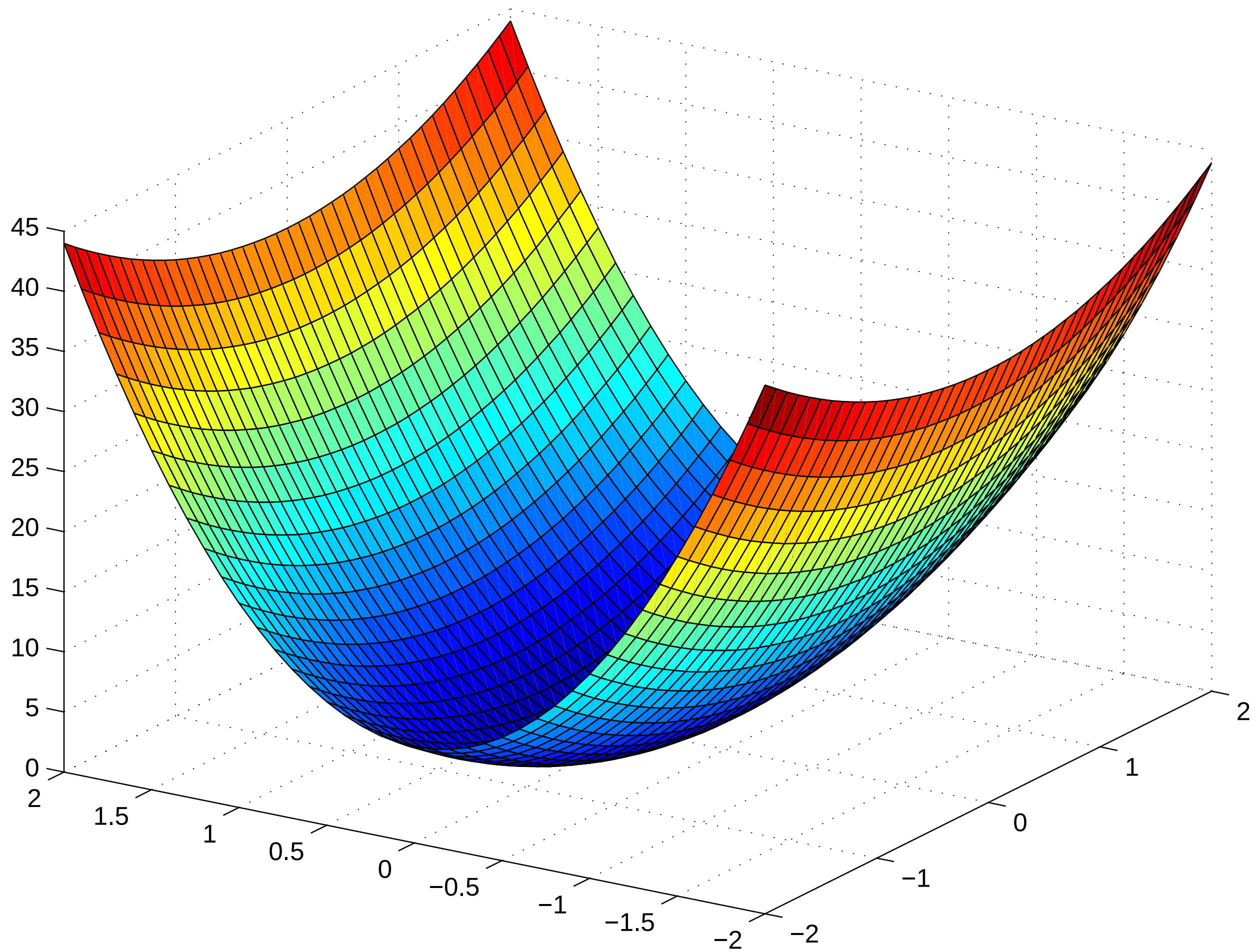
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conditionnement et préconditionnement

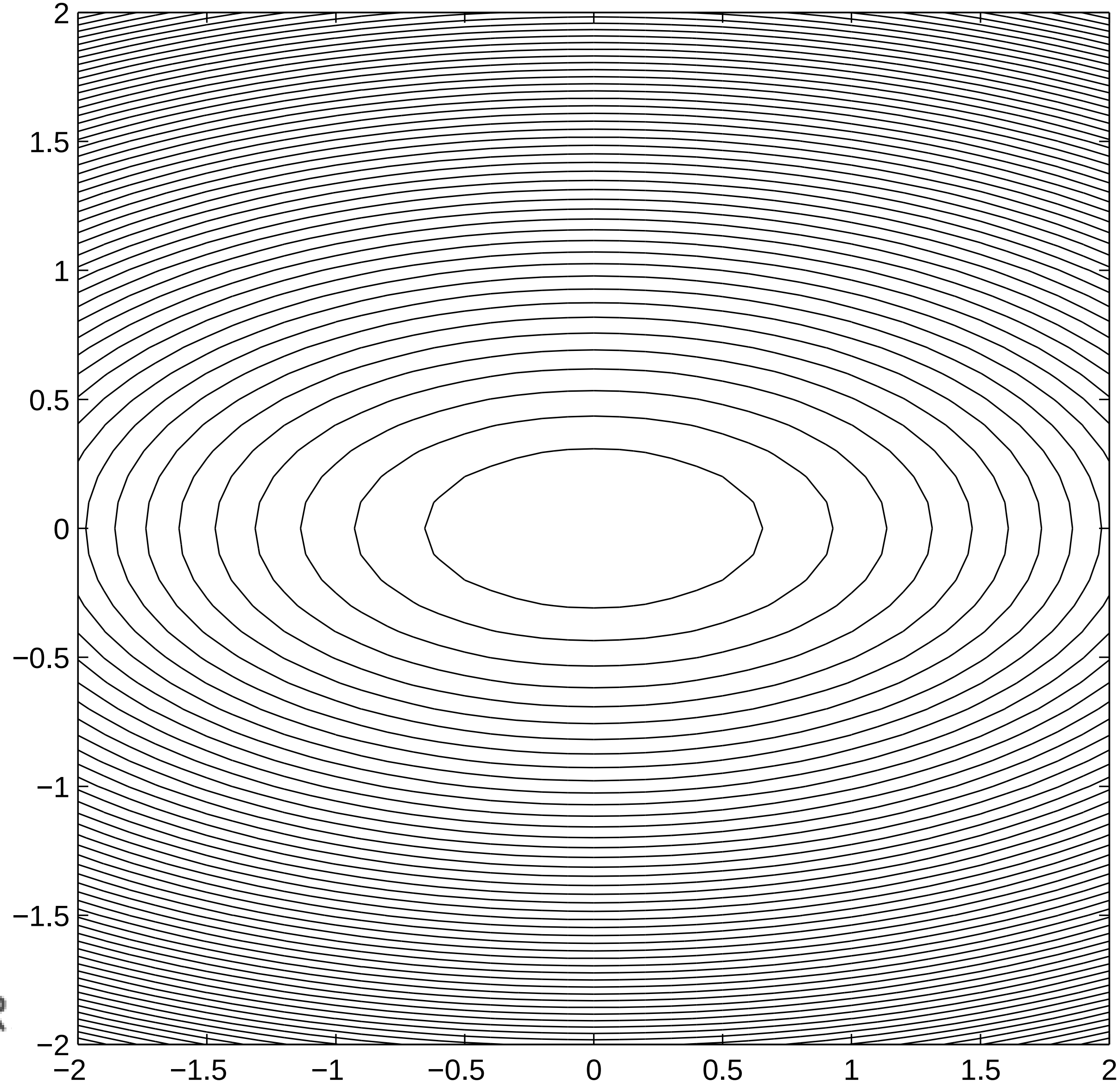
Interprétation géométrique

- λ_1 plus grande courbure de f en x
- λ_n plus petite courbure de f en x
- Nombre de conditionnement : rapport entre la plus grande et la plus petite des courbures parmi les directions partant de x.
- Si proche de 1, fonction bien conditionnée

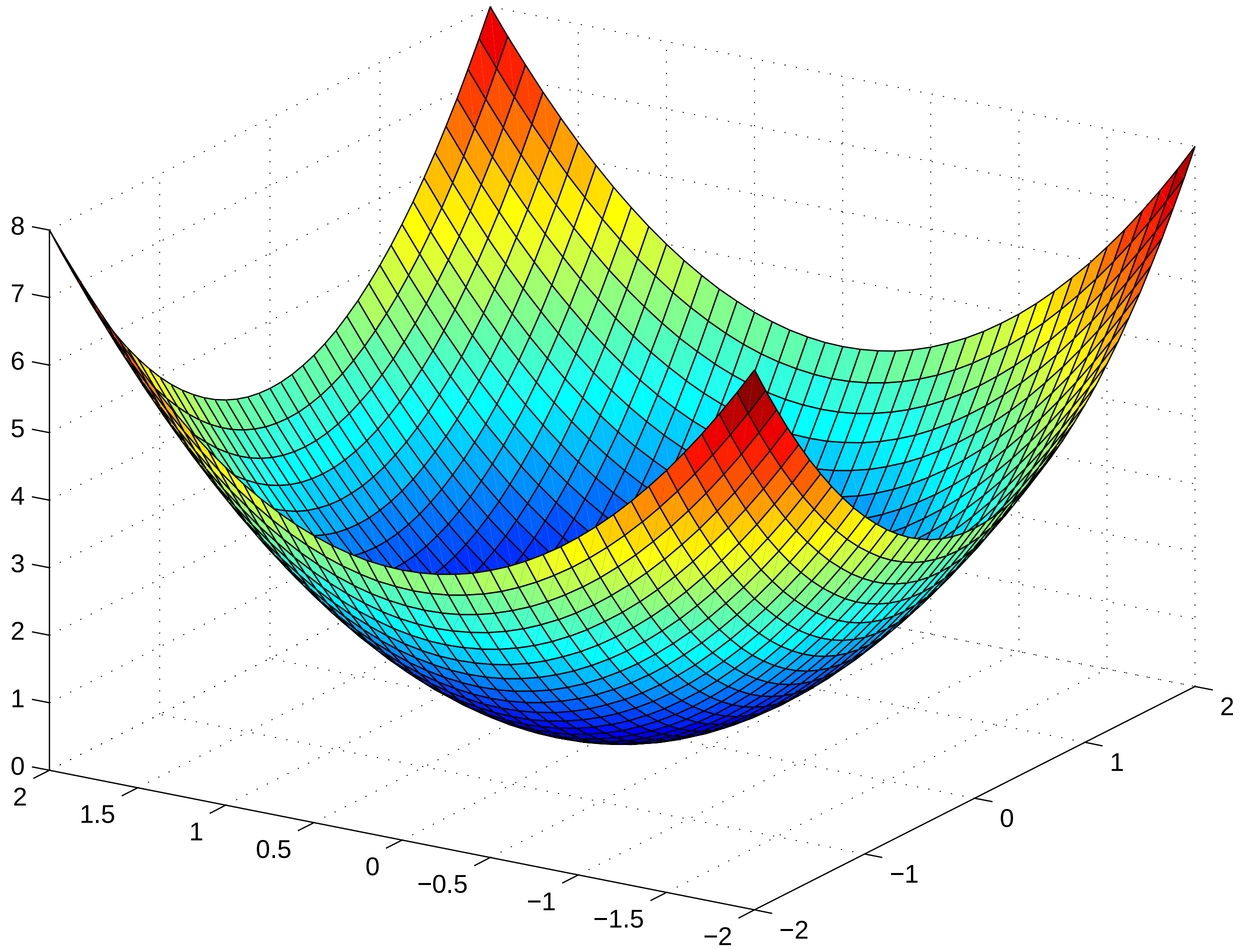
Conditionnement et préconditionnement



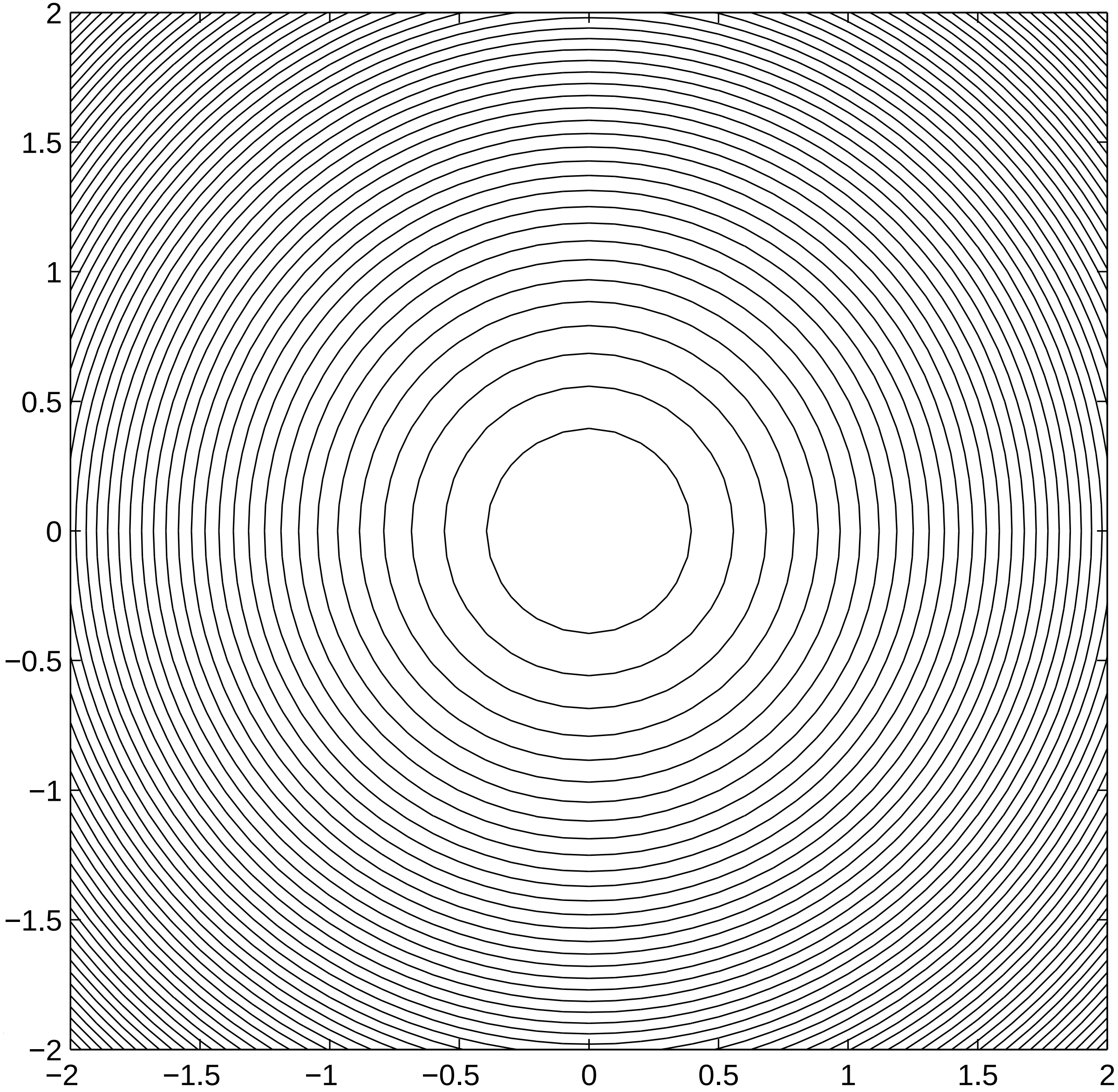
Conditionnement et préconditionnement



Conditionnement et préconditionnement



Conditionnement et préconditionnement



Conditionnement et préconditionnement

- Soit

$$f(x_1, x_2) = 2x_1^2 + 9x_2^2$$

- Nombre de conditionnement = $9/2$, $\forall x$, car

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 4 & 0 \\ 0 & 18 \end{pmatrix}.$$

- Appliquons maintenant le changement de variable

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3\sqrt{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Conditionnement et préconditionnement

- Changement de variable

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{\sqrt{2}}{6} \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}.$$

- Nous obtenons

$$f(x'_1, x'_2) = \frac{1}{2}{x'_1}^2 + \frac{1}{2}{x'_2}^2,$$

dont le Hessien est la matrice identité, et le nombre de conditionnement est 1, pour tout (x'_1, x'_2).

Conditionnement et préconditionnement

Changement de variable

Soit $x \in \mathbb{R}^n$. Soit $M \in \mathbb{R}^{n\times n}$ une matrice inversible.

Le changement de variable est l'application linéaire définie par M et transformant x en $x' = Mx$.

- Soit $f(x)$ et un changement de variables $x' = Mx$
- Nous obtenons

$$
\begin{aligned}
\tilde{f}(x') &= f(M^{-1}x') \\
\nabla \tilde{f}(x') &= M^{-T}\nabla f(M^{-1}x') \\
\nabla^2 \tilde{f}(x') &= M^{-T}\nabla^2 f(M^{-1}x')M^{-1}.
\end{aligned}
$$

Conditionnement et préconditionnement

- Conditionnement de $\tilde{f}$ en x'= conditionnement de

$$M^{-T}\nabla^2 f(M^{-1}x')M^{-1}$$

- Objectif : choisir M tel que ce conditionnement soit le plus proche possible de 1

Conditionnement et préconditionnement

Préconditionnement

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction deux fois différentiable, et soit un vecteur $x \in \mathbb{R}^n$.

Le préconditionnement de f en x revient à définir un changement de variable $x' = Mx$ et une fonction $\tilde{f}(x') = f(M^{-1}x')$, tels que le conditionnement de $\tilde{f}$ en Mx soit meilleur que le conditionnement de f en x.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Résumé

- Définition du problème
  - Minimum local – minimum global.
  - Existence d'une solution : Weierstrass.
- Gradient = dérivée du premier ordre
  - Informe sur la pente de la fonction.
  - Informe sur la convexité de la fonction.
  - Concept de direction de descente.
- Matrice hessienne = dérivée du second ordre
  - Informe sur la courbure de la fonction.
  - Informe sur la convexité de la fonction.
- Propriétés géométriques
  - Linéarité - non linéarité (Lipschitz).
  - Conditionnement.