Une promenade informelle dans le monde de l'optimisation mathématique

Namur Center for Complex Systems (naXys), University of Namur, Belgium

Collège Belgique, Namur, Octobre 2013

Mes objectifs :

- Montrer (par l'exemple) que l'optimisation est un langage dans lequel s'expriment et se résolvent naturellement un grand nombre de problèmes complexes et intéressants.
- Partager enthousiame et émerveillement devant la variété de son champ d'application.
- Brosser un portrait large du domaine et de sa structure
- Faire passer quelques moments intellectuellement stimulants. . .

Au programme

Concepts et exemples

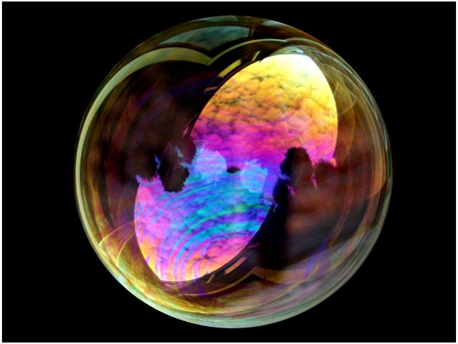
Qu'est-ce que l'optimisation mathématique ?

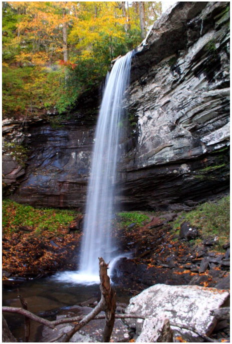
Utiliser les mathématiques pour

> effectuer le meilleur choix sous contraintes

meilleur	$\Rightarrow$	critère, fonction objective (maximiser/minimiser)
choix	$\Rightarrow$	variables dont on peut choisir la valeur
contraintes	$\Rightarrow$	restrictions sur les valeurs admissibles des variables

La nature optimise

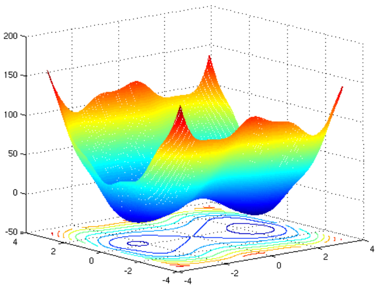


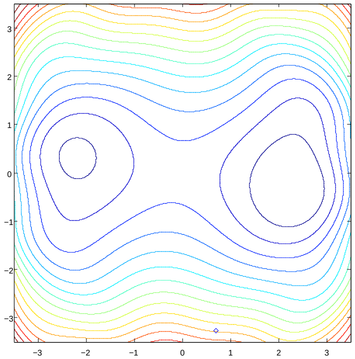


Les humains optimisent (tous les jours)

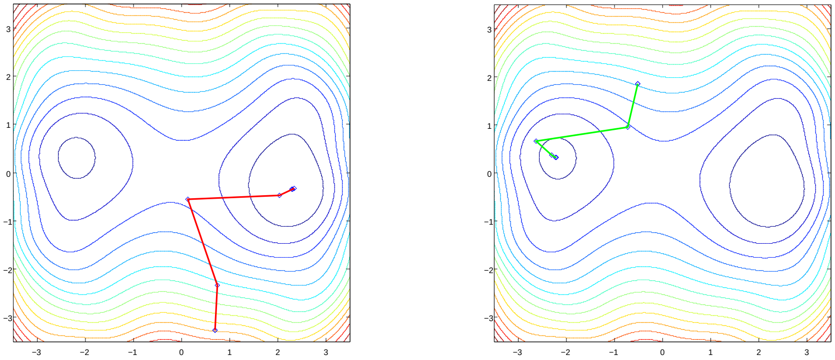
Un premier aperçu des méthodes et moyens

Les méthodes : une vision montagnarde . . .





Les méthodes : le chemin vers le lac



Les moyens

De puissants ordinateurs :

IBM Blue Gene

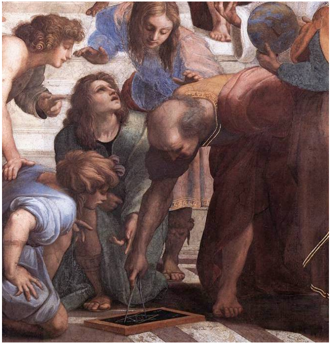
Une discipline ancrée dans l'histoire

Les pionniers de l'optimisation

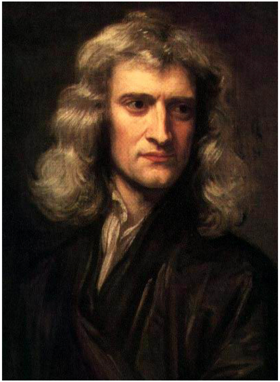
“Nous sommes comme des nains juchés sur des épaules de géants, de telle sorte que nous puissions voir plus de choses et de plus éloignées que n'en voyaient ces derniers. Et cela, non point parce que notre vue serait puissante ou notre taille avantageuse, mais parce que nous sommes portés et exhaussés par la haute stature des géants.”

Bernard de Chartres (1130-1160)

Euclide (300 BC) Al-Khwarizmi (783-850)

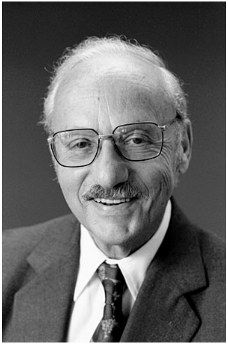


Isaac Newton (1642-1727) Leonhard Euler (1707-1783)



J. de Lagrange (1735-1813) Friedrich Gauss (1777-1855)

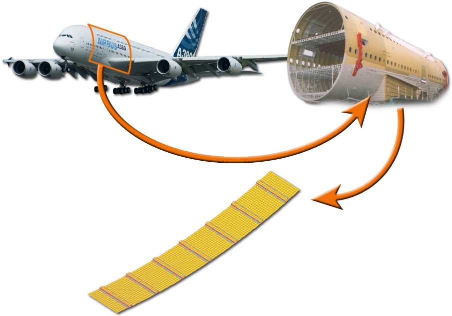
Augustin Cauchy (1789-1857) George Dantzig (1914-2005)



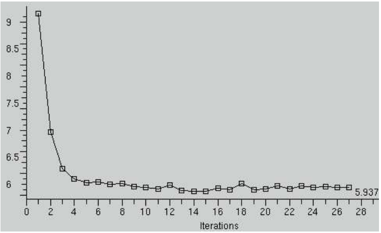
L'optimisation en action

La conception de structures pour l'aéronautique

minimiser le poids du fuselage en maintenant l'intégrité structurale



SAMTECH (2009)

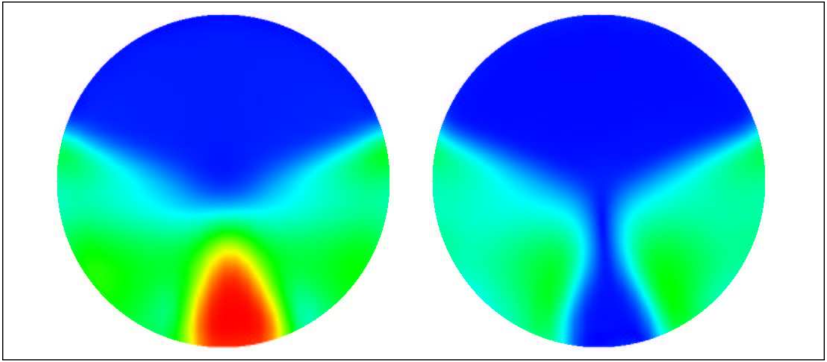


réduction de la masse
en cours d'optimisation

Le design des verres progressifs (1)

Mise au point des lentilles progressives modernes :

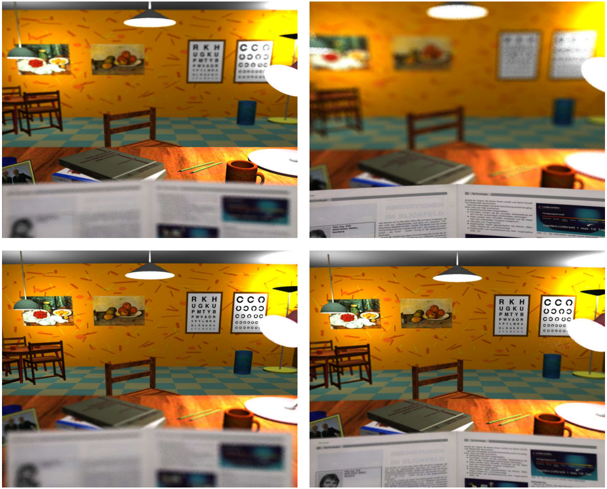
varier la puissance optique de la lentille en minimisant l'astigmatisme



Loos, Greiner, Seidel (1997)

Le design des verres progressifs (2)

Ce qu'on peut réaliser :



sans correction courte distance

longue distance verres progressifs

Le design des verres progressifs (3)

Ce problème est-il non linéaire ($\approx$ difficile) ?

Si la surface de la lentille est $z = z(x, y)$, la puissance optique est

$$p(x,y) = \frac{N^3}{2}\left[\left(1+\left[\frac{\partial z}{\partial x}\right]^2\right)\frac{\partial^2 z}{\partial y^2} + \left(1+\left[\frac{\partial z}{\partial y}\right]^2\right)\frac{\partial^2 z}{\partial x^2} - 2\frac{\partial z}{\partial x}\,\frac{\partial z}{\partial y}\,\frac{\partial^2 z}{\partial x \partial y}\right]$$

où

$$N = N(x,y) = \frac{1}{\sqrt{1+\left[\frac{\partial z}{\partial x}\right]^2+\left[\frac{\partial z}{\partial y}\right]^2}}.$$

L'astigmatisme de la surface est alors

$$a(x,y) = -2\sqrt{p(x,y) - N^4\left(\frac{\partial z}{\partial x}\,\frac{\partial z}{\partial y} - \left[\frac{\partial^2 z}{\partial x \partial y}\right]^2\right)}$$

D'un point de vue esthétique ...

$$p(x,y) = \frac{N^3}{2}\left[\left(1+\left[\frac{\partial z}{\partial x}\right]^2\right)\frac{\partial^2 z}{\partial y^2} + \left(1+\left[\frac{\partial z}{\partial y}\right]^2\right)\frac{\partial^2 z}{\partial x^2} - 2\frac{\partial z}{\partial x}\,\frac{\partial z}{\partial y}\,\frac{\partial^2 z}{\partial x \partial y}\right]$$

$$N = N(x,y) = \frac{1}{\sqrt{1+\left[\frac{\partial z}{\partial x}\right]^2+\left[\frac{\partial z}{\partial y}\right]^2}}.$$

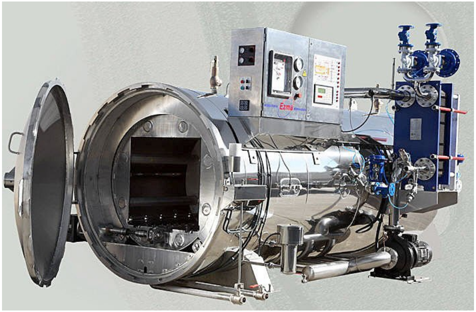
$$a(x,y) = -2\sqrt{p(x,y) - N^4\left(\frac{\partial z}{\partial x}\,\frac{\partial z}{\partial y} - \left[\frac{\partial^2 z}{\partial x \partial y}\right]^2\right)}$$

La stérilisation des aliments pour nourissons (1)

Un problème très courant dans l'industrie alimentaire :

garder un max de vitamines en éliminant une fraction donnée des bactéries

chauffage des aliments dans des autoclaves à vapeur ou eau chaude



Sachs (2003)

La stérilisation des aliments pour nourissons (2)

$\boxed{\text{Modèle :}}$ équations aux dérivées partielles (EDP) couplées

Concentration des micro-organismes et autres nutriments :

$$\frac{\partial C}{\partial t}(x,t) = -K[\theta(x,t)]C(x,t),$$

où $\theta(x,t)$ est la température, et où

$$K[\theta] = K_1 e^{-K_2\left(\frac{1}{\theta}-\frac{1}{\theta_r}\right)} \quad \text{(équation d'Arrhenius)}$$

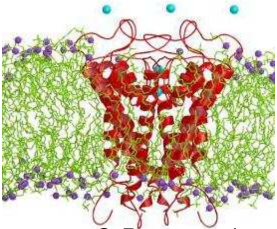
Evolution de la température :

$$\rho c(\theta)\frac{\partial \theta}{\partial t} = \nabla \cdot [k(\theta)\nabla\theta],$$

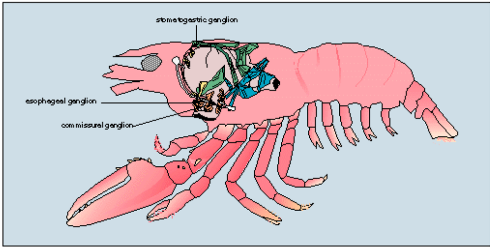
(avec conditions limites adéquates : refroidissant, température initiale, ...)

Une estimation de paramètres en biologie (1)

Un modèle de canal potassique dans une membrane neuronale :
Sansom (2001)



Mais où se trouvent ces neurones ? Dans un homard épineux du Pacifique (*Panulirus interruptus*) ! Simmers, Meyrand and Moulin (1995)



Une estimation de paramètres en biologie (2)

Après collection des données expérimentales
(application d'un courant à la cellule) :

estimer les paramètres du modèle pour reproduire les données au mieux

Le modèle (pour les biologistes) :

- Activation : p portes indépendantes
- Déactivation : n_h portes avec des dynamiques différentes
- BDF à 5-points sur $\approx$ 50000 pas de temps
- $\Rightarrow$ très non linéaire !

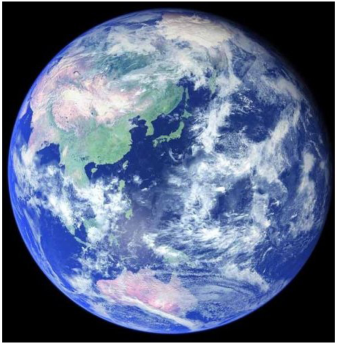
L'assimilation de données en météorologie (1)

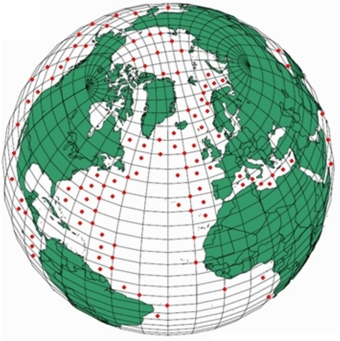
(Essayer de) prédire. . .

- le temps qu'il fera demain
- la température moyenne de l'océan le mois prochain
- le champ de gravité futur
- les prochains courants dans l'ionosphère
- . . .

L'assimilation de données en météorologie (2)

Données : température, vent, pression, ... partout et à tout moment !

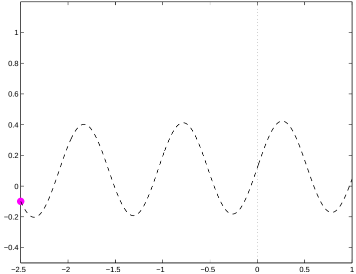




Peut contenir plus de 1.000.000.000 variables !

L'assimilation de données en météorologie (3)

Le principe :

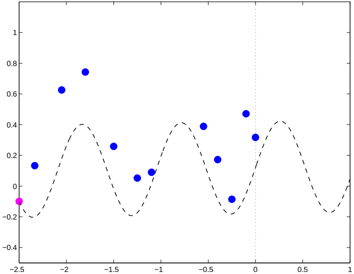


température/jours

- Situation connue il y a 2.5 jours et prévision "de fond"

L'assimilation de données en météorologie (3)

Le principe :



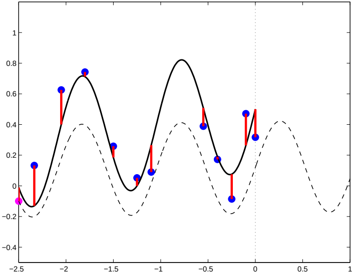
température/jours

- Situation connue il y a 2.5 jours et prévision "de fond"
- Température des 2.5 derniers jours

L'assimilation de données en météorologie (3)

Le principe :

Minimiser l'erreur entre modèle et observations passées



température/jours

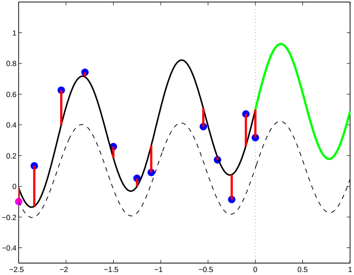
- Situation connue il y a 2.5 jours et prévision "de fond"
- Température des 2.5 derniers jours
- Faire tourner le modèle pour minimiser l'écart | entre modèle and observations

$$\min_{x_0} \frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{i=0}^{N} \|\mathcal{HM}(t_i, x_0) - b_i\|^2_{R_i^{-1}}.$$

L'assimilation de données en météorologie (3)

Le principe :

Minimiser l'erreur entre modèle et observations passées

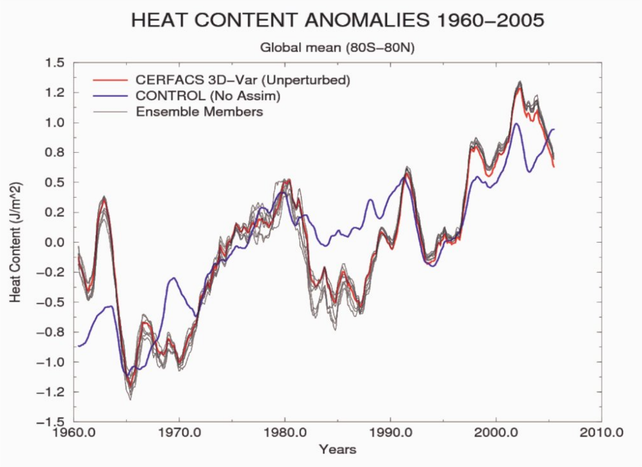


température/jours

- Situation connue il y a 2.5 jours et prévision "de fond"
- Température des 2.5 derniers jours
- Faire tourner le modèle pour minimiser l'écart | entre modèle and observations
- Predire la temperature pour demain

L'assimilation de données en météorologie (4)

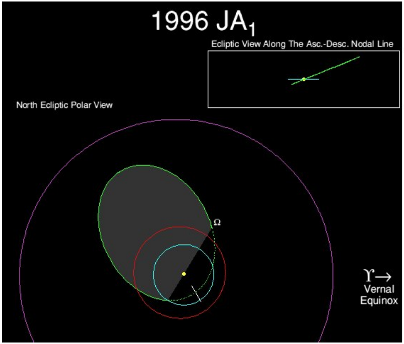
Analyse du contenu de chaleur de l'océan : CERFACS (2009)



Bien meilleur ajustement !

Une partie de chasse aux astéroides

trouver aujourd'hui un astéroide dont les paramètres orbitaux coincident le mieux avec des observations vieilles de 50 ans



Milani, Sansaturio et al. (2005)

Le débruitage d'images (1)

Considérons une image aléatoirement bruitée (noirceur des pixels incorrecte) proportionnellement à la magnitude du signal.

Comment retrouver les valeurs (noirceurs) originales ?

utiliser les valeurs des pixels le plus possible
en minimisant les transitions brutales (gradients)

Ceci donne le problème d'optimisation

$$\min_u \sum_{ij\in\Omega} (u_{ij} - z_{ij}\log(u_{ij})) + \alpha \int_\Omega \|\nabla u\|$$

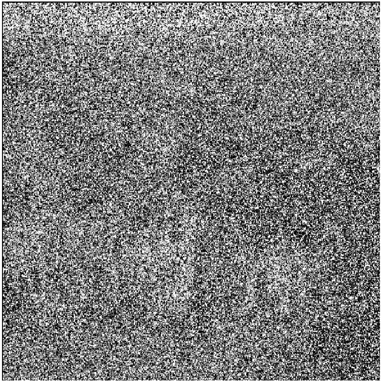
où

$$z_{ij} = u_{ij} + n f(u_{ij})$$

et n est un bruit Gaussien.

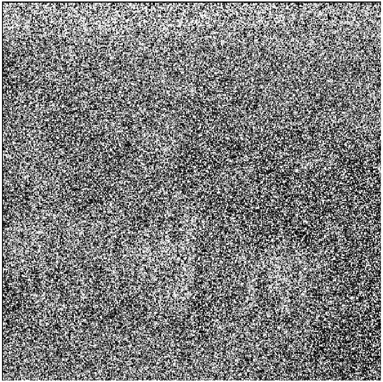
Le débruitage d'images (2)

Quelques résultats spectaculaires : une image 512×512 avec 95% de bruit



Le débruitage d'images (2)

Quelques résultats spectaculaires : une image 512×512 avec 95% de bruit



Chan and Chen (2007)

La reconstruction d'objets cachés

- un object $\mathcal{O}$ caché dans un milieu opaque (sol, eau, corps, ...)
- on "illumine" cet object par une onde (électromagnétique) plane dans "toutes" les directions

$$u^{inc}(x,d) = e^{ik\langle x,d\rangle} \quad d \in \mathcal{S}$$

- on mesure les ondes réfléchies $u^s(x,d)$ par l'objet dans "toutes" les directions

$$\Delta u^s + k^2 u^s = 0 \quad \text{(en dehors de } \mathcal{O}\text{)}$$

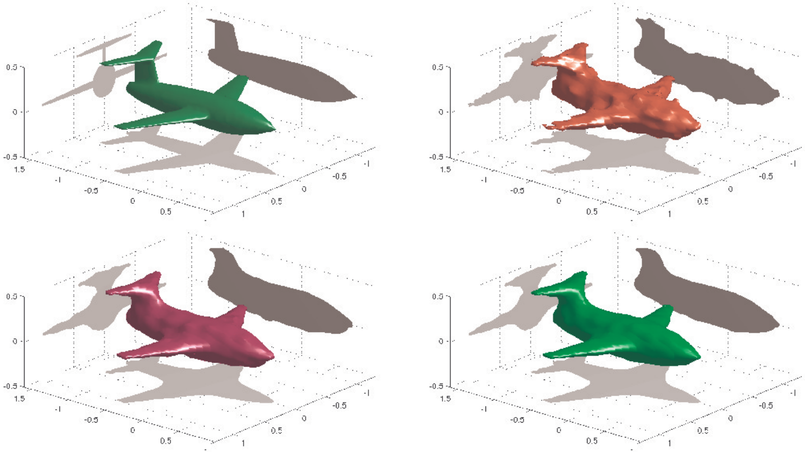
avec

$$u = u^{inc} + u^s, \quad u = 0 \text{ sur } \partial\mathcal{O} \quad \text{et} \quad \lim_{r\to\infty} \frac{\partial u^s}{\partial r} = iku^s.$$

Optimiser la forme de l'objet $\mathcal{O}$ à partir des mesures $u^s(x,d)$

Illustration : la reconstruction d'un avion

L'original et les reconstructions ($N = 1002, p = 5, 20, 50$)



Optimisation, transports et mobilité

Le contexte

Analyse de la mobilité quotidienne en Belgique

Modélisation des déplacements

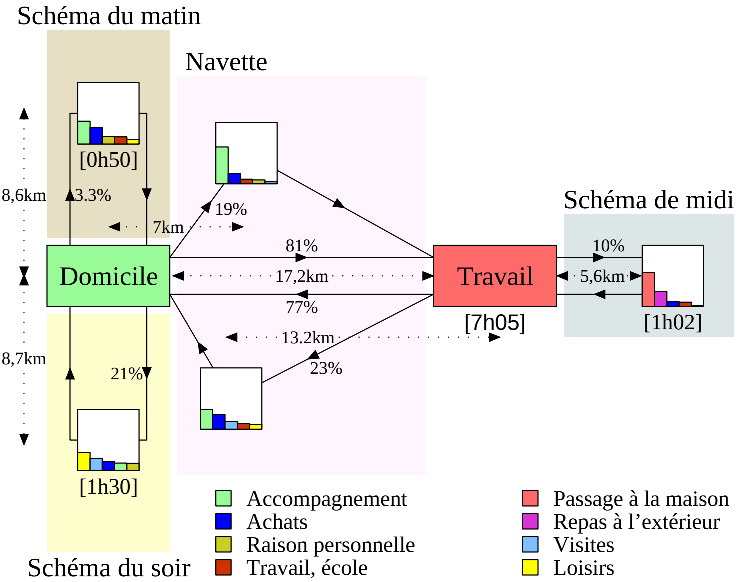
$\Downarrow$

Modélisation des activités

$\Downarrow$

Modélisation des individus (ou ménages)

La mobilité quotidienne : un système complexe



Comprendre les stratégies de choix en société

Contexte : simulation des choix individuels en Mobilité (mode, trajet, heure de départ, ...)

Théorie de l'utilité aléatoire

Un individu i assigne à l'alternative j l' "utilité"

$$U_{ij} = [\text{ paramètres } \times \text{ facteurs explicatifs}] + [\text{ erreur aléatoire }]$$

Illustration :

$$U_{bus} = \text{distance} - 1.2 \times \text{prix du billet} - 2.1 \times \text{retard par rap. au trajet en voiture} + \epsilon$$

Les modèles de choix discrets (1)

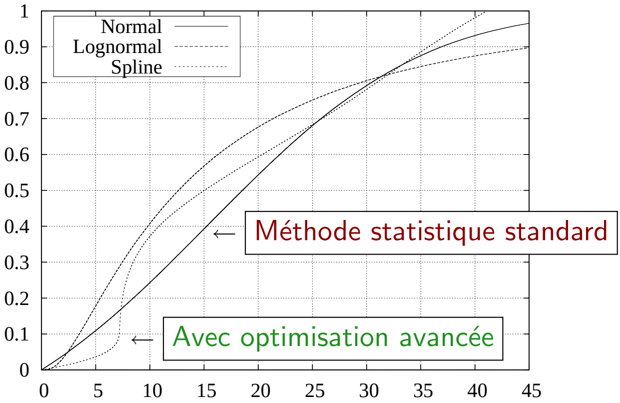
Probabilité qu'un individu i choisisse l'alternative j plutôt que l'alternative k donnée par

$$\text{prob}\left(U_{ij} \geq U_{ik} \quad \text{pour tout} \quad k\right)$$

Données : enquêtes de mobilité (MOBEL, BELDAM, etc.)

trouver les paramètres de l'utilité qui
maximisent la vraisemblance des comportements observés

Les modèles de choix discrets (2)



Estimation de la valeur du temps perdu dans les embouteillages (avec et sans optimisation avancée)

Modéliser les individus ?

Une difficulté majeure

- Données individuelles (souvent) indisponibles/incohérentes. . .
(coût, vie privée, sources variables)

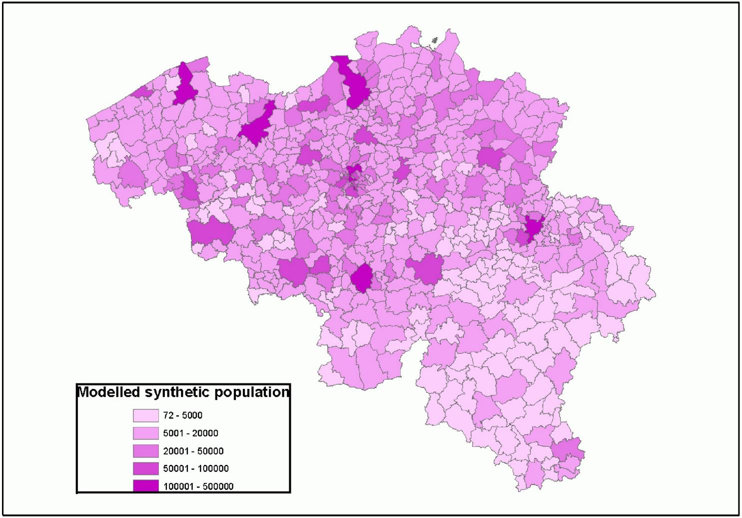
$\boxed{\text{Construire une population synthétique}}$

Méthode

- maximiser la vraisemblance des descriptions des individus/ménages
- en préservant les statistiques connues sur la population réelle

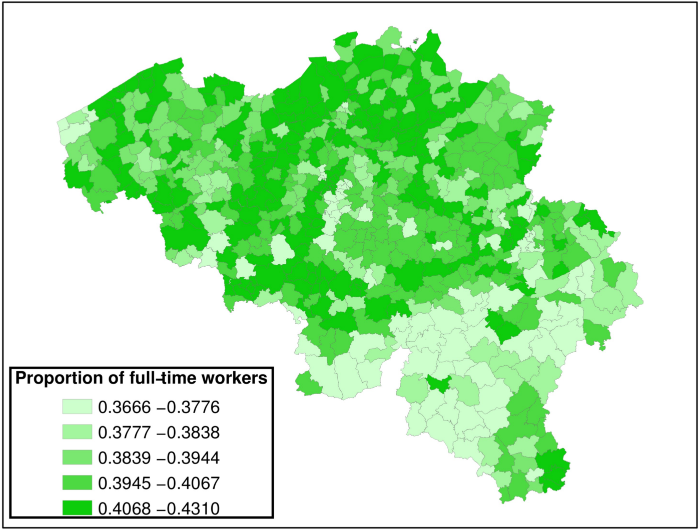
Une population synthétique pour la Belgique

Et l'on obtient $\approx$ 10.600.000 individus dans 4.400.000 ménages



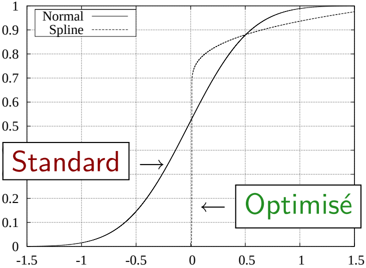
Une population synthétique pour la Belgique

D'où l'on peut déduire (en utilisant d'autres techniques d'optimisation)



Et en finance...

1. gestion du risque
2. analyse de portfolios
3. marchés des changes

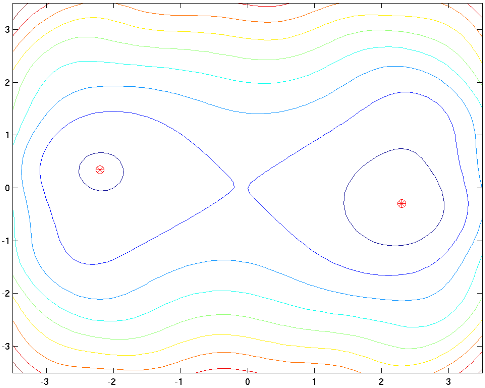


Distribution des investissements de la BoJ 1991-2004

4. ...

Les grands types de problèmes en variables continues

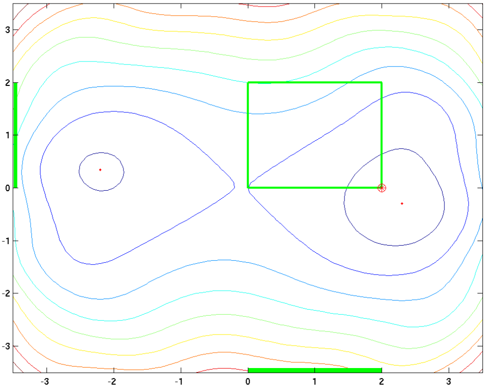
Les problèmes sans contrainte



Les variables peuvent prendre n'importe quelle valeur

$$\min_{x,y} f(x, y)$$

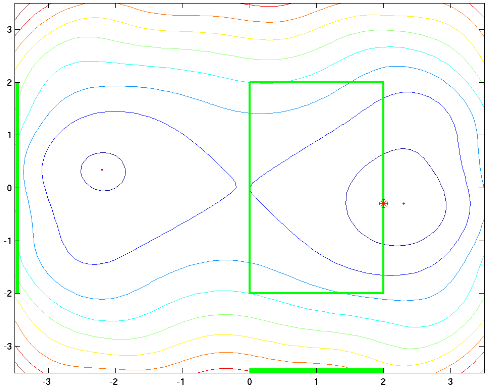
Les problèmes avec contraintes de bornes (1)



Les variables doivent rester dans certains intervalles

$$\min_{x,y} f(x,y) \quad \text{avec} \quad 0 \leq x \leq 2 \ \text{ et } \ 0 \leq y \leq 2$$

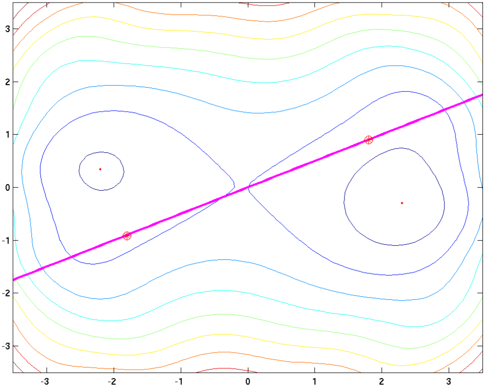
Les problèmes avec contraintes de bornes (2)



Les variables doivent rester dans certains intervalles

$$\min_{x,y} f(x, y) \quad \text{avec} \quad 0 \leq x \leq 2 \;\; \text{et} \;\; -2 \leq y \leq 2$$

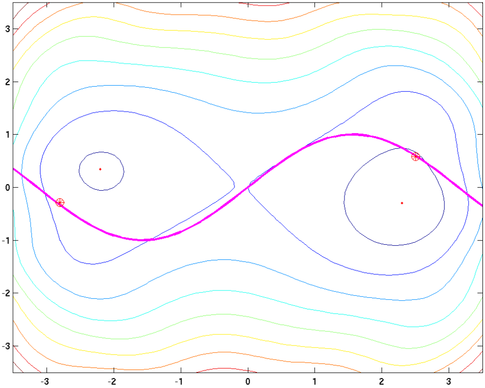
Les problèmes avec contrainte(s) d'égalité (1)



Les solutions doivent satisfaire certaine(s) égalité(s)

$$\min_{x,y} f(x,y) \quad \text{avec} \quad y = \frac{1}{2}x \quad \text{(linéaire)}$$

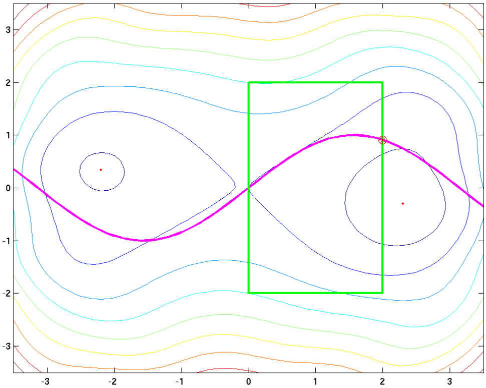
Les problèmes avec contrainte(s) d'égalité (2)



Les solutions doivent satisfaire certaine(s) égalité(s)

$$\min_{x,y} f(x, y) \quad \text{avec} \quad y = \sin x \quad \text{(non linéaire)}$$

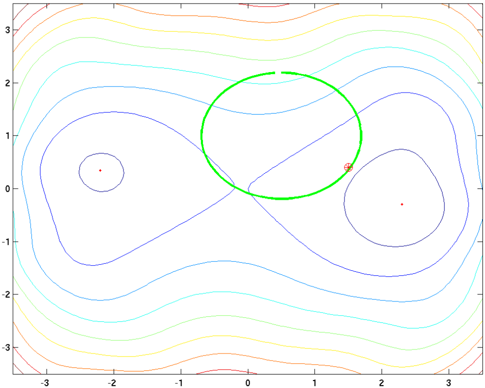
Les problèmes avec contraintes de bornes et d'égalité



Une combinaison des deux types précédents ...

$$\min_{x,y} f(x,y) \quad \text{avec} \quad 0 \leq x \leq 2, -2 \leq y \leq 2 \;\; \text{et} \;\; y = \sin x$$

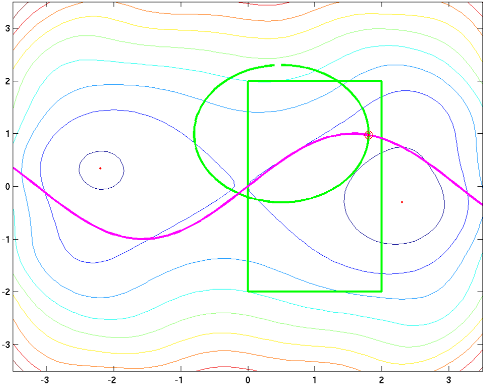
Les problèmes avec contrainte(s) d'inégalité générales



Les solutions doivent appartenir à un ensemble ...

$$\min_{x,y} f(x,y) \quad \text{avec} \quad (x-\tfrac{1}{2})^2 + (y-1)^2 \leq 1.19$$

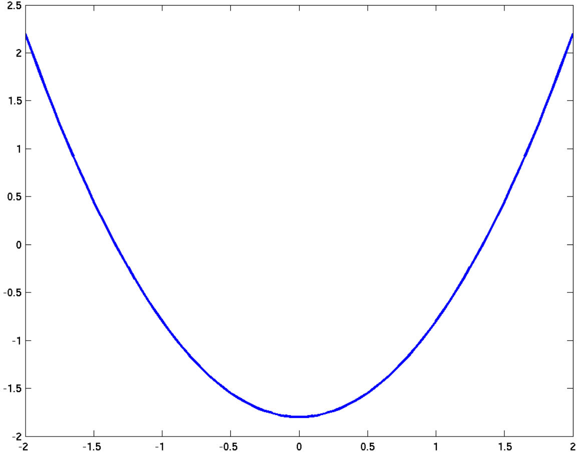
Le problème avec contraintes mixtes



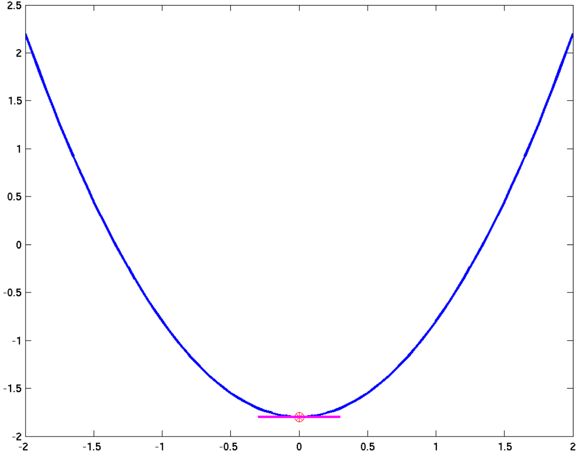
Ou bien tous ces types de contraintes ensemble ...

$$\min_{x,y} f(x,y) \quad \text{avec} \quad 0 \leq x \leq 2, -2 \leq y \leq 2, y = \sin x, (x-\tfrac{1}{2})^2+(y-1)^2 \leq 1.19$$

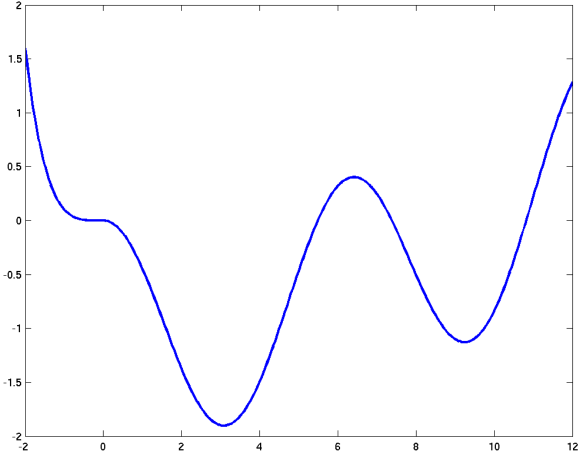
Qu'est-ce qu'une solution ? (simple)



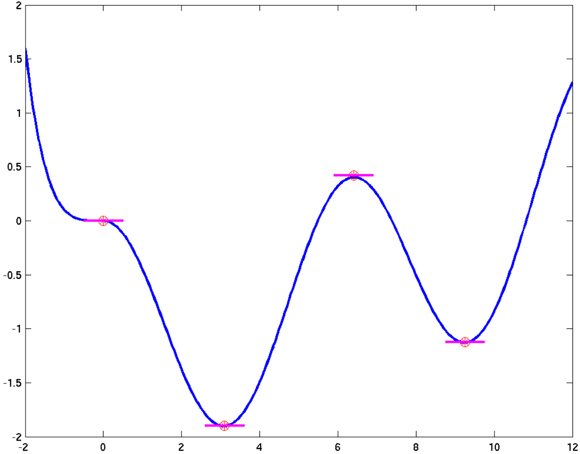
Qu'est-ce qu'une solution ? (simple)



Qu'est-ce qu'une solution ? (moins simple)



Qu'est-ce qu'une solution ? (moins simple)



Une condition nécessaire (sans contrainte)

Dans tous les cas

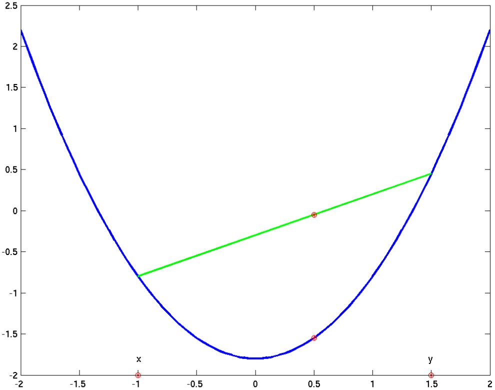
A une solution sans contrainte, la courbe (surface) de la fonction objective est plate

$$\nabla_x f(x_*) = 0$$

Le gradient $\nabla_x f(x) = g(x)$ indique la direction de la plus grande pente de la surface $f(x)$ en x.

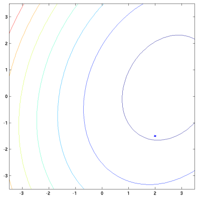
Plus compliqué s'il y a des contraintes ...

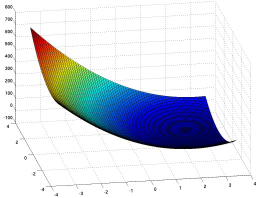
Une autre distinction fondamentale : la convexité



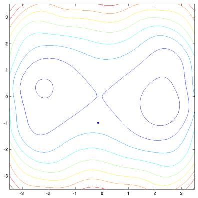
$$\boxed{\forall x, y\ \forall t \in [0,1], \quad f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)}$$

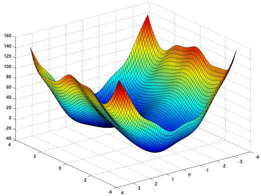
Convexe et non convexe





convexe





non convexe

Impact de la convexité

Si le problème est convexe, toute solution locale est globale

Si le problème est strictement convexe,
toute solution finie est unique

$\Rightarrow$ Les problèmes convexes sont beaucoup plus simples
(hélas, le monde n'est pas convexe. . .)

Quelques autres distinctions...

la taille (dimension)

- le nombre de variables : de 1 à ... plusieurs millions !
- le nombre de contraintes : de 0 à ... plusieurs millions !

le dégré de nonlinéarité

- fonction objective, contraintes

les coûts d'évaluation

- fonction objective, contraintes
- connues sous *forme analytique* ou non ?
- de "bon marché" à "tout-à-fait impossible"

la structure

- creux, partiellement séparable

$$\min_{x,y,z} f_1(x,y) + f_2(y,z) \text{ plus facile que } \min_{x,y,z} f(x,y,z)!$$

Les grands types d'algorithmes

Comment calculer une solution (approximative) ?

$\boxed{\text{Méthode de calcul} = \text{algorithme}}$

De quoi dépend le design d'un algorithme ?

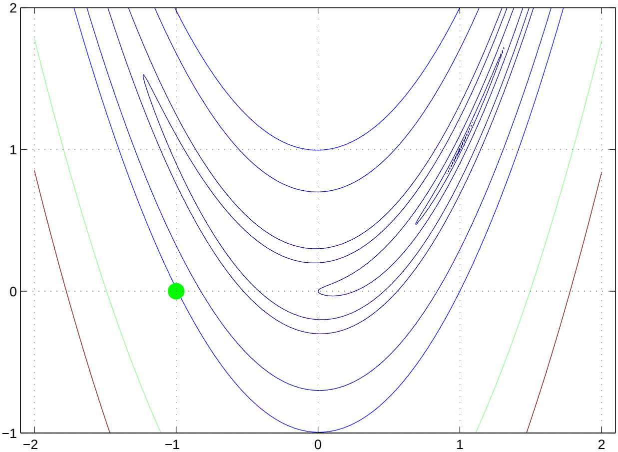
- le type de problème (contraintes ?)
- le type d'information disponible sur le problème
- un choix de modélisation
- des choix méthodologiques

Un exemple

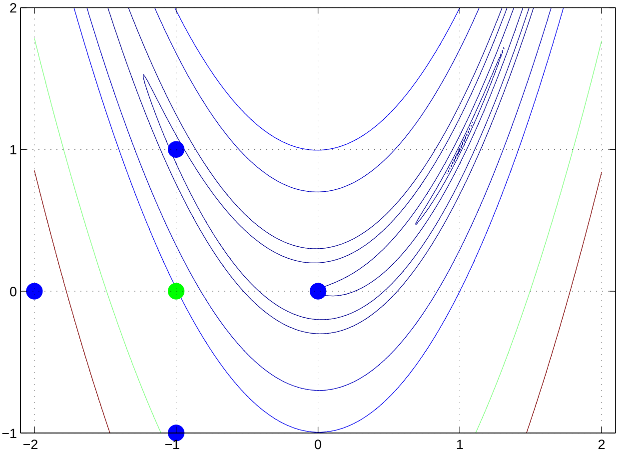
Un case très simple :

- pas de contraintes
- seulement les valeurs de la fonction objective sont disponibles

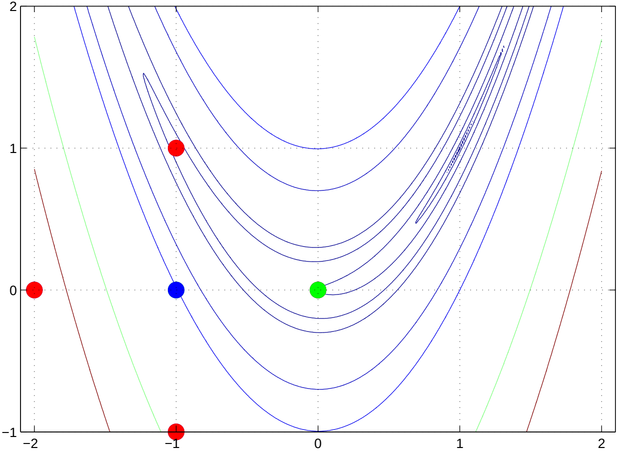
La méthode du compas sur la fonction "banane" de Rosenbrock



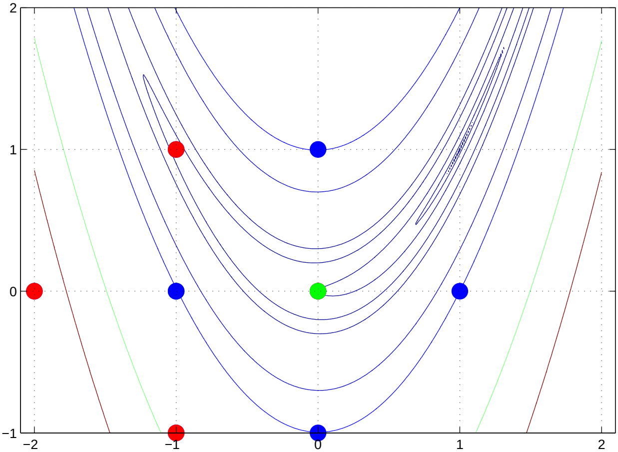
La méthode du compas sur la fonction "banane" de Rosenbrock



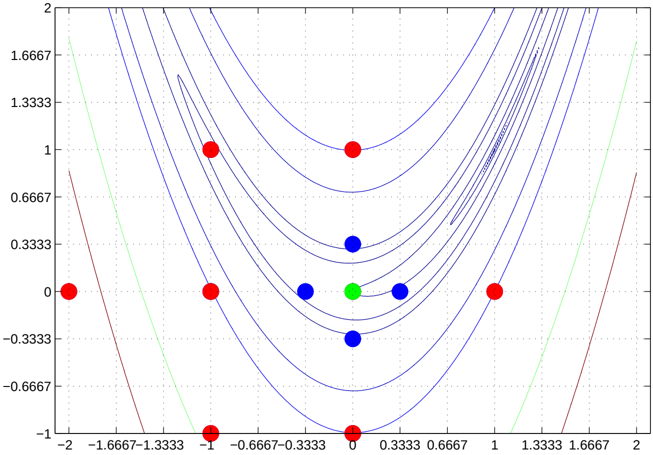
La méthode du compas sur la fonction "banane" de Rosenbrock



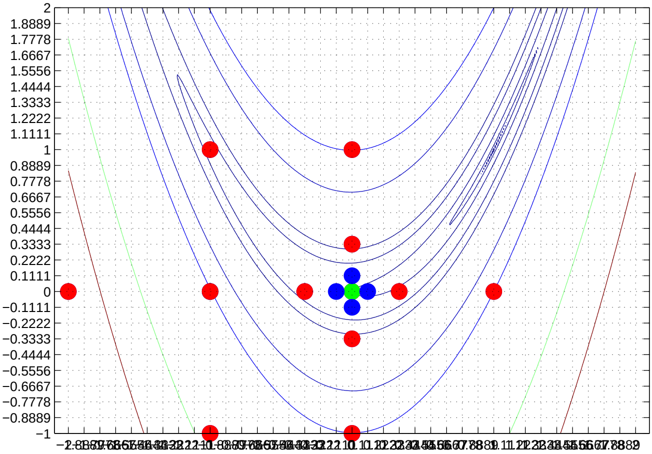
La méthode du compas sur la fonction "banane" de Rosenbrock



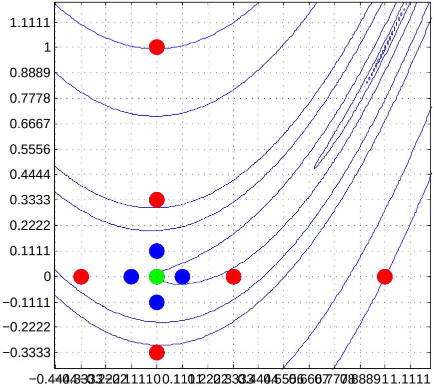
La méthode du compas sur la fonction "banane" de Rosenbrock



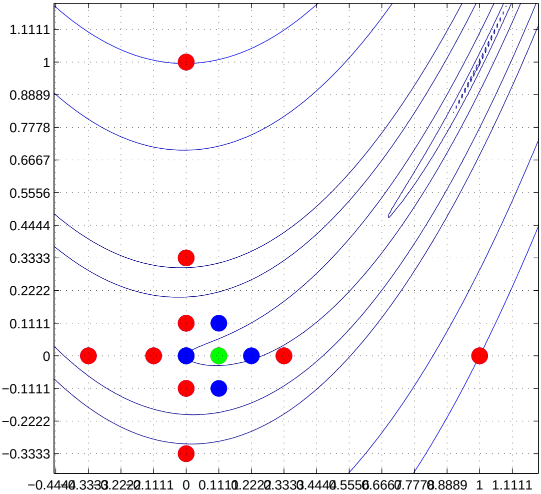
La méthode du compas sur la fonction "banane" de Rosenbrock



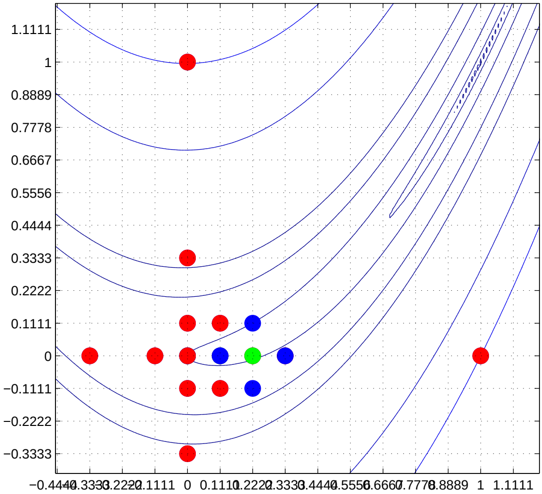
La méthode du compas sur la fonction "banane" de Rosenbrock



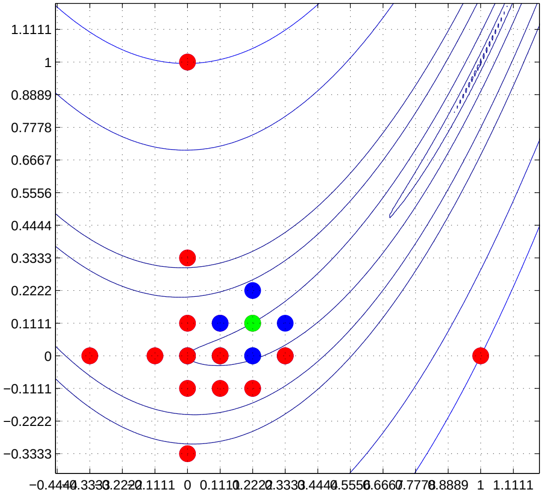
La méthode du compas sur la fonction "banane" de Rosenbrock



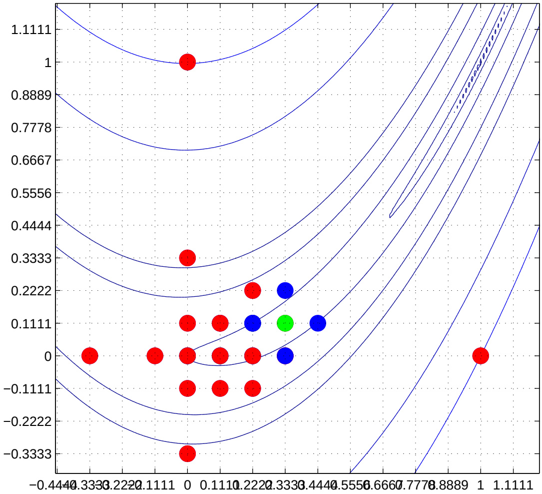
La méthode du compas sur la fonction "banane" de Rosenbrock



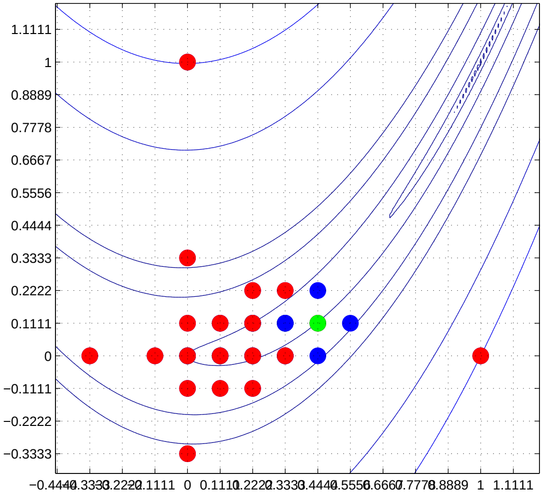
La méthode du compas sur la fonction "banane" de Rosenbrock



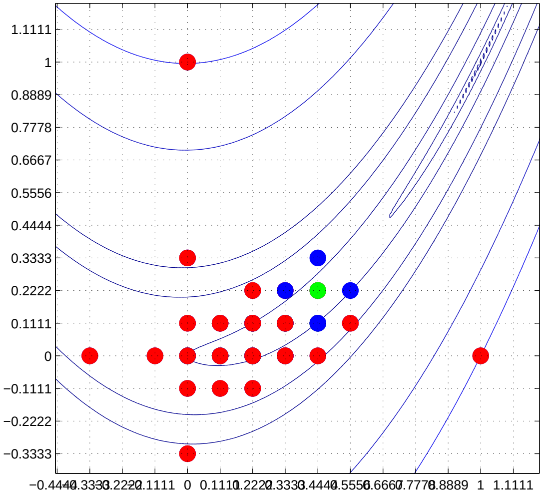
La méthode du compas sur la fonction "banane" de Rosenbrock



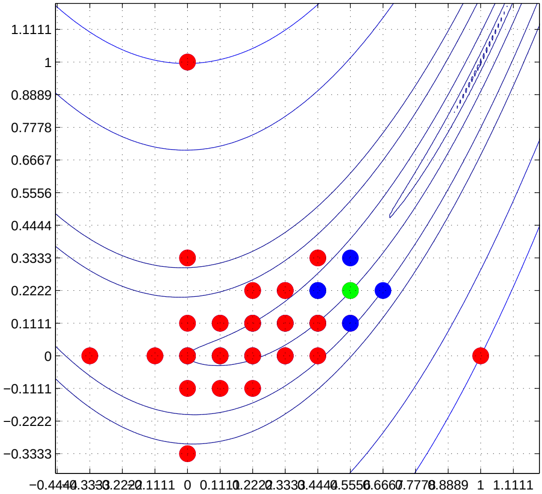
La méthode du compas sur la fonction "banane" de Rosenbrock



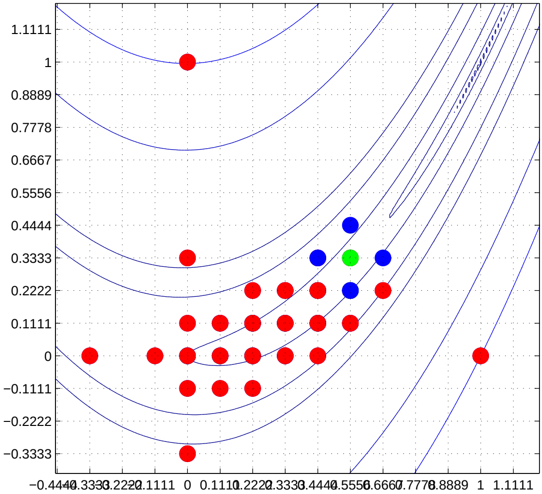
La méthode du compas sur la fonction "banane" de Rosenbrock



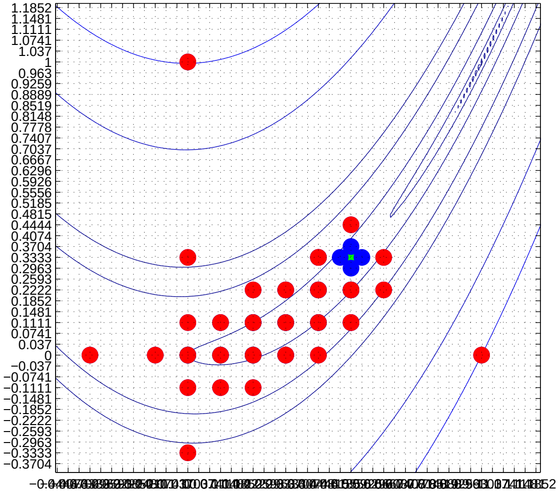
La méthode du compas sur la fonction "banane" de Rosenbrock



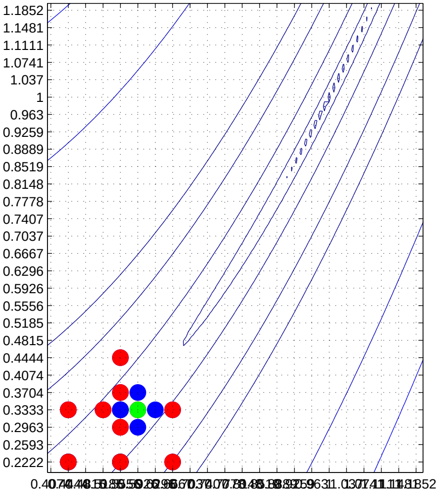
La méthode du compas sur la fonction "banane" de Rosenbrock



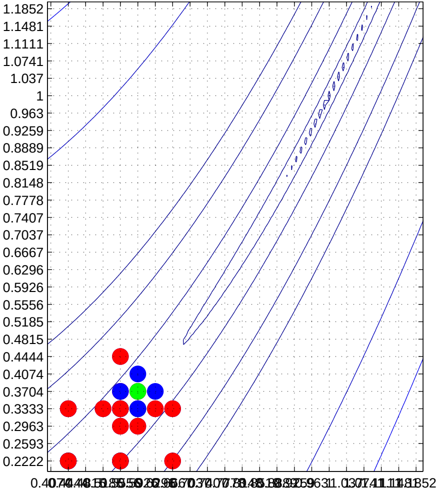
La méthode du compas sur la fonction "banane" de Rosenbrock



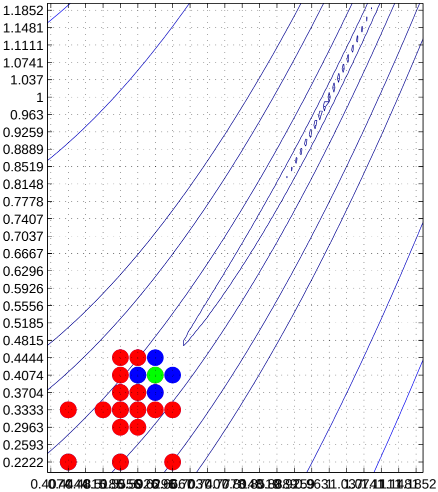
La méthode du compas sur la fonction "banane" de Rosenbrock



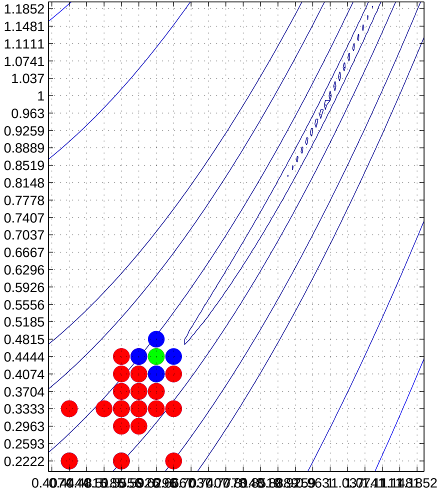
La méthode du compas sur la fonction "banane" de Rosenbrock



La méthode du compas sur la fonction "banane" de Rosenbrock



La méthode du compas sur la fonction "banane" de Rosenbrock



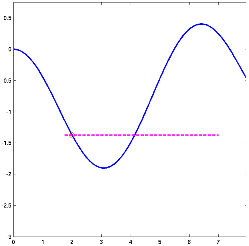
Les ingrédients d'une amélioration

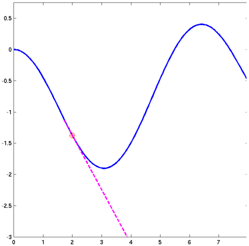
Malheureusement . . . très lent !

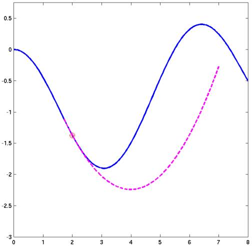
$\Rightarrow$ utiliser plus d'information (que simplement la valeur de f)

la pente locale $\longrightarrow$ le gradient (dérivées premières)

la courbure locale $\longrightarrow$ le Hessien (dérivées secondes)





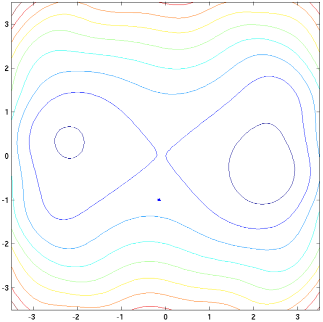


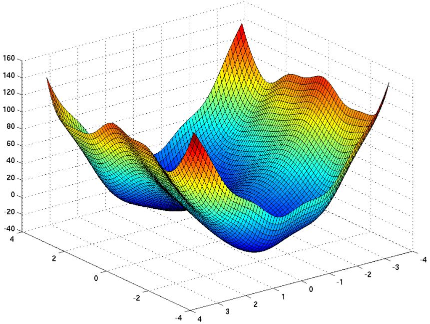
Comment construire des modèles locaux ?

La fonction objective f est compliquée dans le voisinage de x

$\Rightarrow$ construire un modèle local.

Revenons à notre exemple favori :

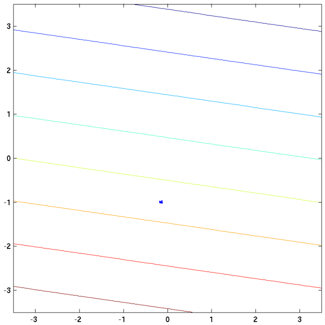


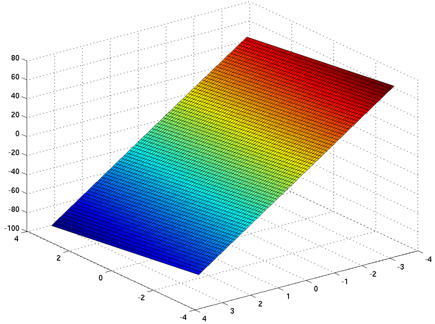


Assez facile : un modèle linéraire

Une interprétation simple :
la pente reste constante et il n'y a pas d'autre effet

$$\ell(u,v) = f(x,y) + g_1(x,y)(u-x) + g_2(x,y)(v-y)$$



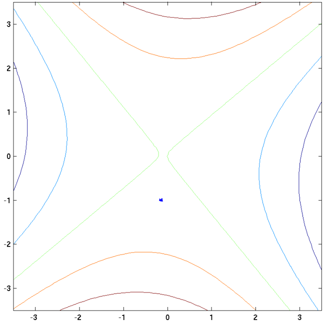


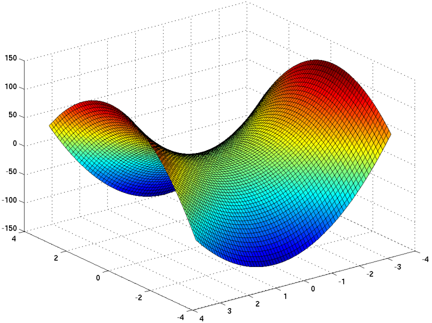
Un peu plus complexe : un modèle quadratique

Plus précis : la pente et la courbure restent constante et il n'y a pas d'autre effet

$$q(u,v) = f(x,y) + g_1(x,y)(u-x) + g_2(x,y)(v-y)$$

$$+\tfrac{1}{2}H_{11}(x,y)(u-x)^2 + H_{12}(x,y)(u-x)(v-y) + \tfrac{1}{2}H_{22}(x,y)(v-y)^2$$

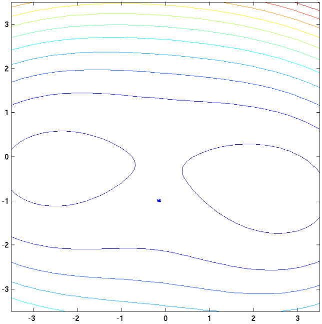


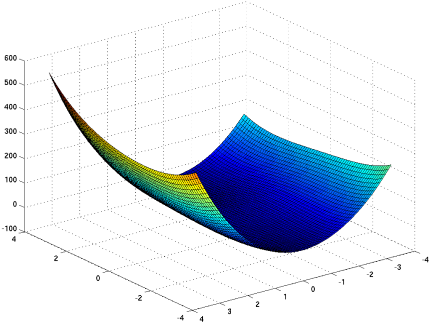


Sophistiqué : un modèle cubique

On ajoute au modèle quadratique une pénalisation de la distance à (x, y)

$$c(u, v) = q(u, v) + \frac{\sigma}{3}\left[(u-x)^2 + (v-y)^2\right]^{3/2}$$

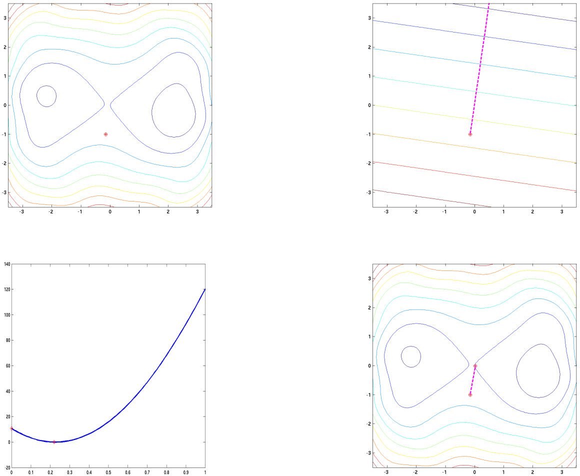




Avec le modèle linéaire : la méthode de Cauchy

Minimise la fonction objective f le long de la direction de la plus grande pente

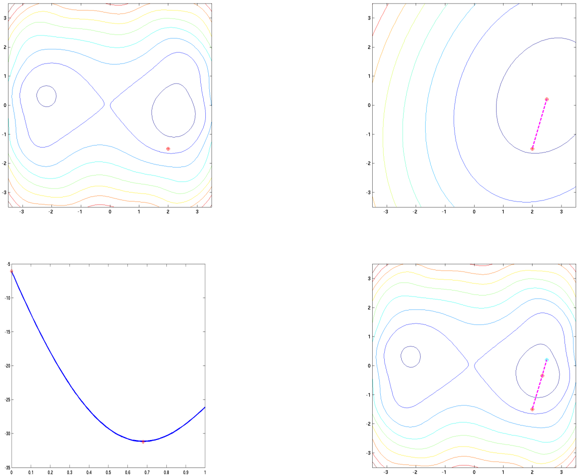
recherche linéaire



Avec le modèle quadratique : la méthode de Newton (1)

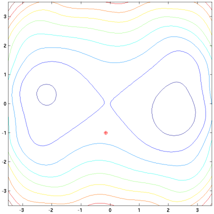
Minimise la fonction objective f le long de la direction du minimum du modèle

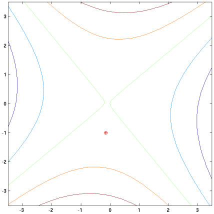
recherche linéaire



Avec le modèle quadratique : la méthode de Newton (2)

Minimise la fonction objective f le long de la direction du ? ? ?



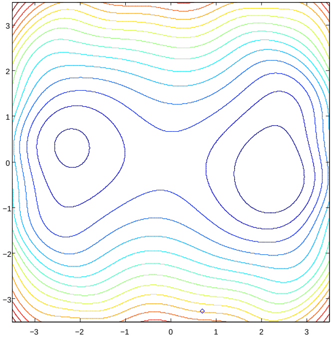


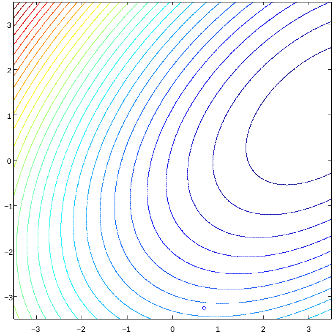
La méthode de Newton n'est pas toujours bien définie !

$\Rightarrow$ on ne peut pas faire confiance au modèle trop loin...

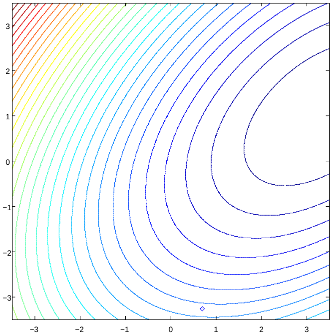
Avec le modèle quadratique : la méthode des régions de confiance

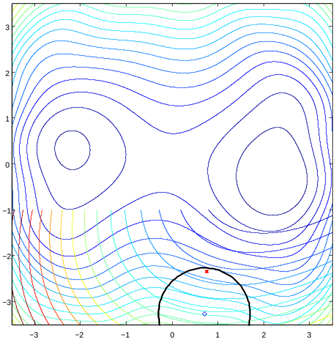
Idée : déterminer une adaptivement région (sphérique) dans laquelle on peut faire confiance au modèle



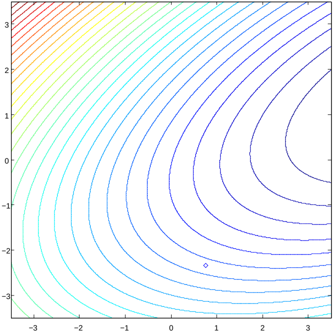


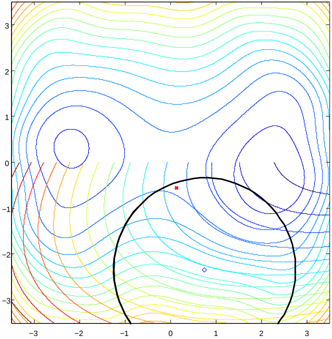
Les itérations de la méthode de région de confiance



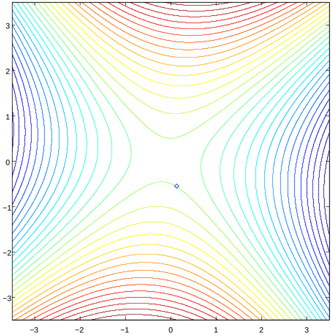


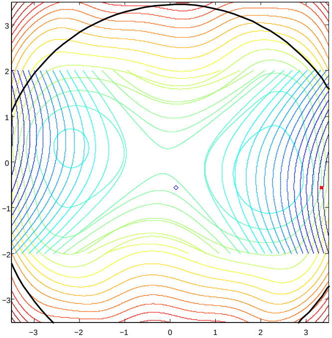
Les itérations de la méthode de région de confiance



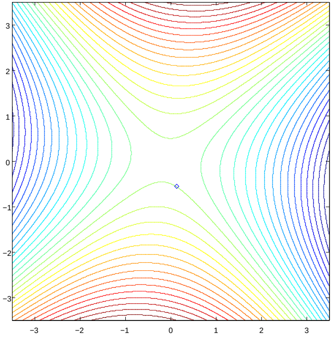


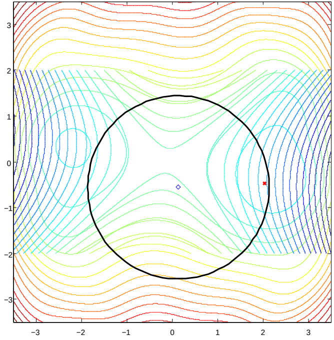
Les itérations de la méthode de région de confiance



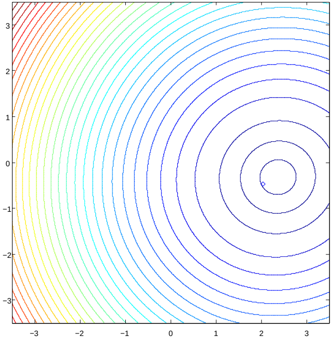


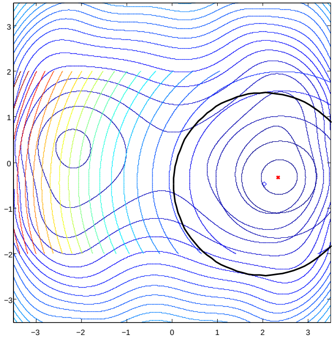
Les itérations de la méthode de région de confiance



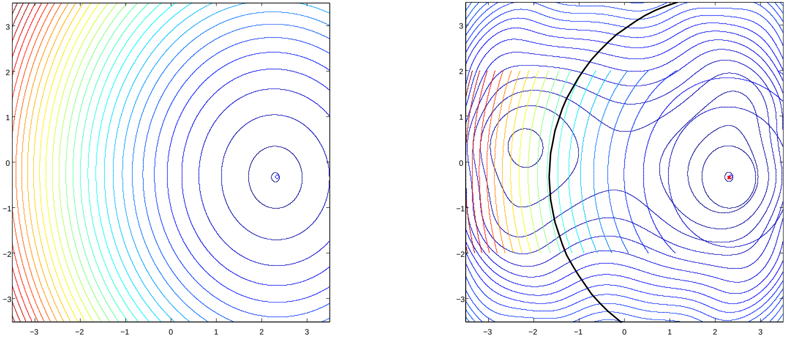


Les itérations de la méthode de région de confiance

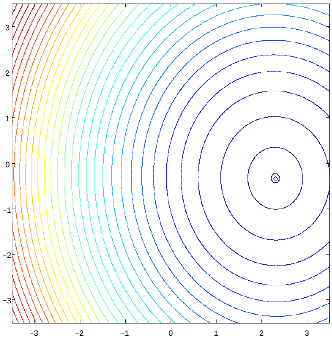


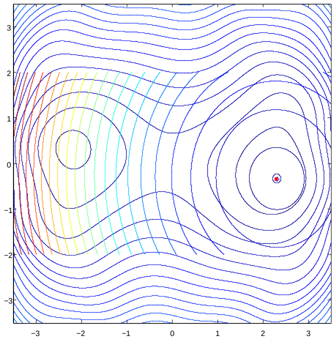


Les itérations de la méthode de région de confiance



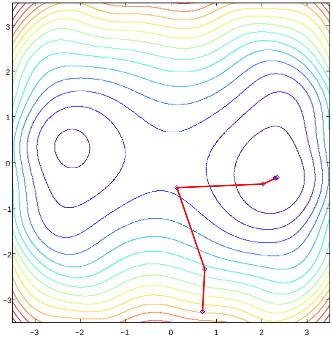
Les itérations de la méthode de région de confiance



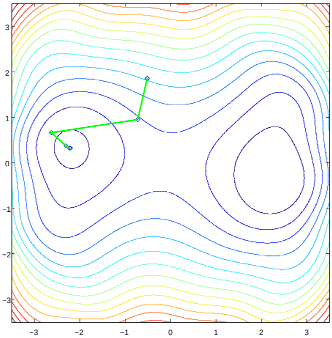


En résumé

Les itérations :



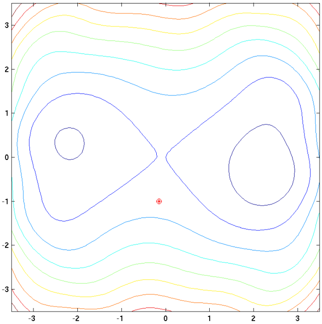
En partant d'ailleurs :

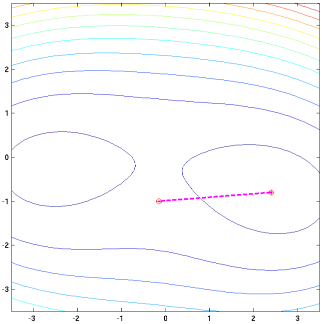


Avec le modèle cubique : la méthode ARC

Même idée, en utilisant le modèle cubique pour ne pas aller trop loin

Pour un pas :



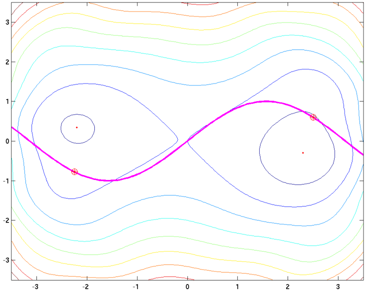


Ajuster le terme cubique plutôt que le rayon de la région

Problèmes avec contraintes : la méthode des pénalités

Idée : pénaliser la violation des contraintes $e(x) = 0$

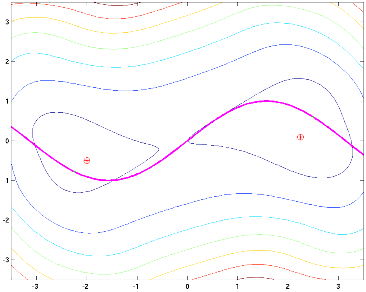
$$min_x f(x) + \rho e(x)^2 \quad \rho = 0$$



Problèmes avec contraintes : la méthode des pénalités

Idée : pénaliser la violation des contraintes $e(x) = 0$

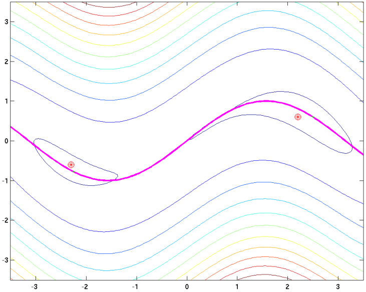
$$min_x f(x) + \rho e(x)^2 \quad \rho = 10$$



Problèmes avec contraintes : la méthode des pénalités

Idée : pénaliser la violation des contraintes $e(x) = 0$

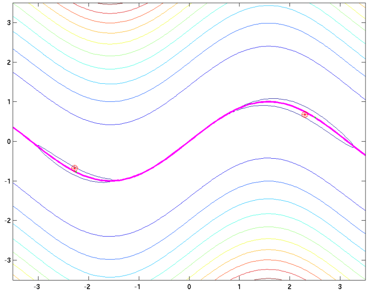
$$min_x f(x) + \rho e(x)^2 \quad \rho = 100$$



Problèmes avec contraintes : la méthode des pénalités

Idée : pénaliser la violation des contraintes $e(x) = 0$

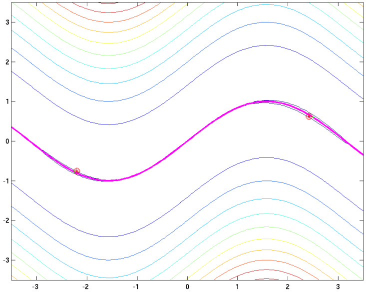
$$min_x f(x) + \rho e(x)^2 \qquad \rho = 1000$$



Problèmes avec contraintes : la méthode des pénalités

Idée : pénaliser la violation des contraintes $e(x) = 0$

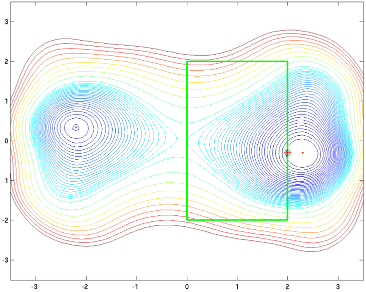
$$min_x f(x) + \rho e(x)^2 \qquad \rho = 5000$$



Problèmes avec contraintes d'inégalité : la méthode des points intérieurs

Idée : ériger une barrière le long des contraintes $i(x) \geq 0$

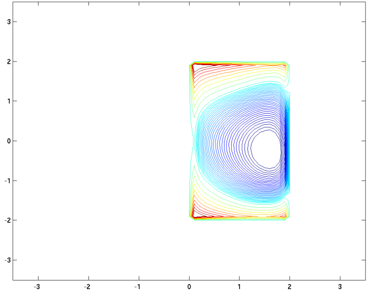
$$min_x f(x) - \mu \log x \quad \mu \text{ inactif}$$



Problèmes avec contraintes d'inégalité : la méthode des points intérieurs

Idée : ériger une barrière le long des contraintes $i(x) \geq 0$

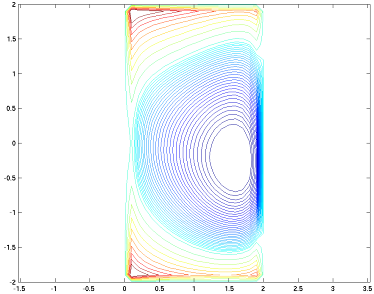
$$min_x f(x) - \mu \log x \quad \mu = 10$$



Problèmes avec contraintes d'inégalité : la méthode des points intérieurs

Idée : ériger une barrière le long des contraintes $i(x) \geq 0$

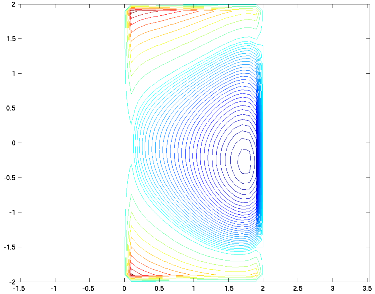
$$min_x f(x) - \mu \log x \qquad \mu = 10$$



Problèmes avec contraintes d'inégalité : la méthode des points intérieurs

Idée : ériger une barrière le long des contraintes $i(x) \geq 0$

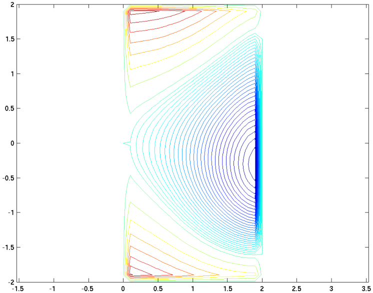
$$min_x f(x) - \mu \log x \qquad \mu = 5$$



Problèmes avec contraintes d'inégalité : la méthode des points intérieurs

Idée : ériger une barrière le long des contraintes $i(x) \geq 0$

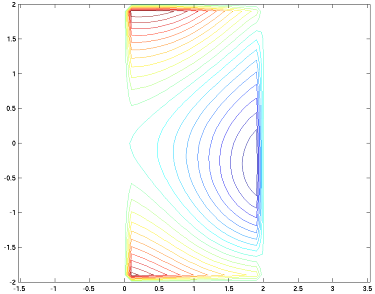
$$min_x f(x) - \mu \log x \qquad \mu = 1$$



Problèmes avec contraintes d'inégalité : la méthode des points intérieurs

Idée : ériger une barrière le long des contraintes $i(x) \geq 0$

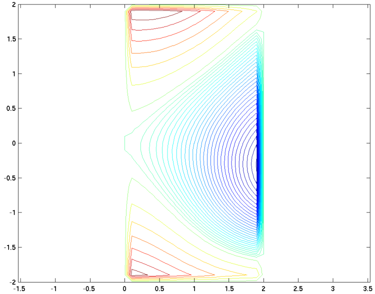
$$min_x f(x) - \mu \log x \qquad \mu = 0.1$$



Problèmes avec contraintes d'inégalité : la méthode des points intérieurs

Idée : ériger une barrière le long des contraintes $i(x) \geq 0$

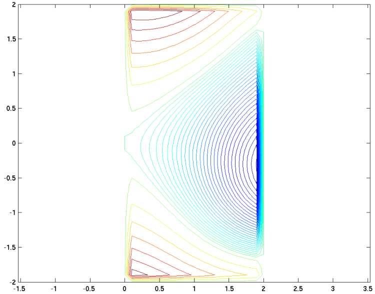
$$min_x f(x) - \mu \log x \quad \mu = 0.001$$



Problèmes avec contraintes d'inégalité : la méthode des points intérieurs

Idée : ériger une barrière le long des contraintes $i(x) \geq 0$

$$min_x f(x) - \mu \log x \quad \mu = 0.00001$$



Et puis bien sûr ...

... cela peut se compliquer : on mélange les types de contraintes

$\Rightarrow$ $\boxed{\text{un vrai zoo de méthodes}}$:

- méthodes du lagrangien augmenté
- méthodes d'optimisation quadratique séquentielle
- méthodes des filtres
- méthode de l'entonnoir
- ...

Une autre piste : abandonner la monotonicité, etc., etc.

Convergence et complexité

Comment appliquer d'un algorithme ?

Appliquer un algorithme, c'est (en principe) :

1. choisir un point de départ
2. le laisser calculer jusqu'à ce qu'une solution satisfaisante soit trouvée

$\Rightarrow$ qu'est-ce qu'une solution satisfaisante ? ? ?

Choisir un critère d'optimalité approximative

$$\|g(x)\| = 0 \text{ devient } \|g(x)\| \leq \epsilon \qquad \text{(sans contrainte)}$$

Note : ϵ est un nombre positif proche de zéro.

Qu'attendre d'un algorithme ?

Les questions qui se posent alors :

convergence :	l'algorithme va-t-il s'arreter pour n'importe quel choix de $\epsilon \in (0, 1)$?
vitesse de convergence :	une fois "proche" la solution, l'algorithme accélère-t-il ?
complexité :	si $\epsilon \in (0, 1)$ est choisi, combien d'évaluations de $f(x)$ l'algorithme va-t-il demander dans le pire des cas ?

Dans le cas sans contrainte

méthode	convergence	vitesse	complexité (évaluations de f)
Cauchy	tout juste	très lent	$O(1/\epsilon^2)$
Newton	oui (presque)	oui	$O(1/\epsilon^2)$
R. de confiance	oui	oui	$O(1/\epsilon^2)$
ARC	oui	oui	$O(1/\epsilon^{3/2})$

ϵ	0.1	0.01	0.001	0.0001
$1/\epsilon^2$	100	10000	1000000	100000000
$1/\epsilon^{3/2}$	32	1000	32622	1000000

Les surprises

Quelques résultats surprenants pour l'optimisation sans contrainte :

On peut donner une borne supérieure pour la complexité !

La méthode de Newton est aussi lente, dans le pire des cas, que la (très lente) methode de Cauchy dans le pire des cas !

La méthode ARC est la plus rapide possible (optimale) !

Encore plus surprenant !

Pour l'optimisation avec contraintes :

La complexité est, pour chaque "classe de méthodes", la même qu'il y ait des contraintes ou non !

La complexité ne dépend pas du type et du nombre de contrainte(s) !

Note : ces résultats sont *très* récents.

Conclusions et perspectives

Quelques conclusions

Le langage mathématique est extraordinairement *souple, adapté, élégant et efficace*.

Les mathématiques appliquées sont un domaine en plein développement.

L'optimisation est en expansion encore plus rapide.

Il reste *énormément* à comprendre et découvrir !

Merci de votre patiente attention.

(Merci aussi à N. Gould, A. Conn, C. Cartis, S. Gratton, A. Sartenaer, D. Orban, . . .)