OPT : an introduction to Numerical and Combinatorial Optimization

E. Fabre[1], R. Andonov[2], S. Pinchinat[3]

[1]CR INRIA, DistribCom team, IRISA
[2]Pr, Symbiose team, IRISA
[3]MC, S4 team, IRISA

Master 2, research in Computer Science, '07

Download the lecture notes from my web-page :
http://www.irisa.fr/distribcom/Personal_Pages/fabre/fabre.html

Outline

Outline

What is it all about?

$$x^* = \arg\min_{x \in \mathcal{D}} f(x)$$

- domain $\mathcal{D}$
- cost function (min) or objective function (max) $f : \mathcal{D} \to \mathbb{R}$
- vector of parameters x^*

In general, the value $f(x^*)$ of secondary interest compared to x^*.
We distinguish

- numerical optimization where $\mathcal{D} \subseteq \mathbb{R}^d$, d possibly large, from
- combinatorial optimization where $\mathcal{D}$ is a discrete domain.

Numerical optimization

Corresponds to the case $f : \mathcal{D} \subseteq \mathbb{R}^d \to \mathbb{R}$, with d possibly large.

Notations :

- $x = [x_1, ..., x_d]^t$, a column-vector
- we often use $f(x) = f(x_1, ..., x_d)$

A few classes of problems, sorted by increasing difficulty.

Problems without constraints

Quadratic problem

$$f(x) = x^t A x + b^t x + c$$

where A symmetric matrix in $\mathbb{R}^{d\times d}$, $b \in \mathbb{R}^d$ and $c \in \mathbb{R}$.

Recall :

- $b^t x = \sum_{i=1}^{d} b_i x_i$ is the scalar product in $\mathbb{R}^d$
- Ax is the matrix product, and so $x^t(Ax)$ yields a scalar

Problems without constraints

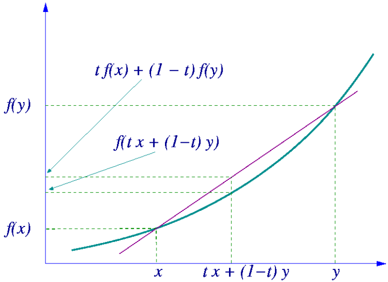
Convex problem

$$\forall\, 0 \leq \alpha \leq 1, \qquad f[\alpha x + (1-\alpha) y] \;\leq\; \alpha f(x) + (1-\alpha) f(y)$$

x^0 is a local minimum of f iff :

$$\exists \epsilon > 0, \quad \|x - x^0\| \leq \epsilon \;\;\Rightarrow\;\; f(x) \geq f(x^0)$$

Interest of convex functions : any local minimum x^* of f is also a global minimum.



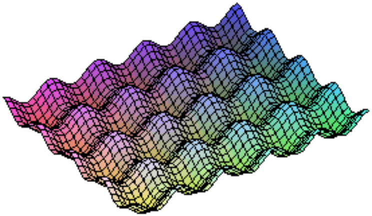
Problems without constraints

Non-linear problem

All other cases.

The most regular (=differentiable) f is, the easier the problem.

In these lectures, we will assume f $\mathcal{C}^1$, or $\mathcal{C}^2$ when necessary.



Problems with constraints

$$\min_x f(x) \qquad \text{subject to} \qquad \theta_j(x) \leq 0, \quad 1 \leq j \leq m$$

$$\theta_j : \mathbb{R}^d \longrightarrow \mathbb{R}$$

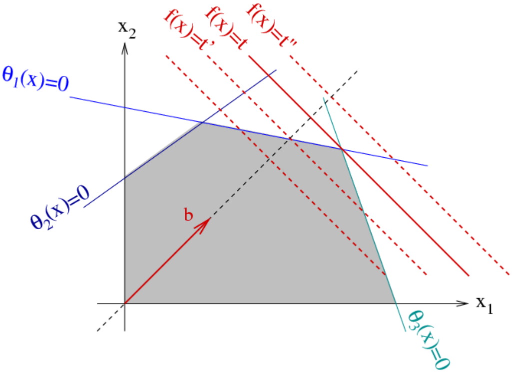
Difficulties :

- In general, equality constraints are simpler than inequalities.
- The simpler and most regular are the constraints, the simpler the problem.

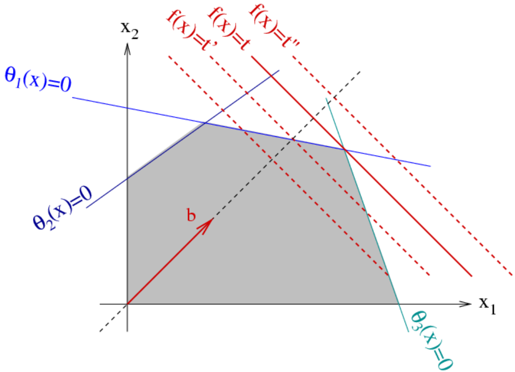
Problems with constraints

Linear program

- f is an linear function $f(x) = b^t x$, to maximize
- constraints are affine : $\theta_j(x) = b_j^t x + c_j \leq 0$
- one often has $x_i \geq 0$ in the constraints



Problems with constraints



- $\mathcal{D}$ is a convex volume, limited by hyperplanes,
- its boundary is called a simplex
- the maximum of f is found at a corner of the simplex
- the simplex algorithm explores these corners
- also addressed by interior points methods

Problems with constraints

Quadratic program

f is a quadratic function, constraints are affine

Addressed by quadratic programming

Non-linear program

All the rest. By increasing complexity:

- f convex, affine constraints,
- f regular (C^n), affine constraints,
- f convex, convex constraints : convex programming
- f regular, convex constraints,
- f regular, regular constraints,
- etc.

Combinatorial optimization

Corresponds to the case where x varies in a discrete domain $\mathcal{D}$.

- $\mathcal{D} = \mathbb{N}^d$,
- $\mathcal{D} =$ set of paths on a graph,
- etc.

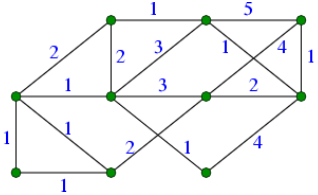
Problems are sorted by complexity.
In front of a combinatorial optimization problem, try to express it as one of the standard examples.

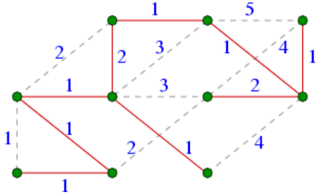
Easy problems (polynomial complexity)

Optimal covering tree

Graph $G = (V, E, c)$ with $c : E \to \mathbb{R}$ assigning costs to edges.

Problem : find a min cost tree in G that reaches all vertices





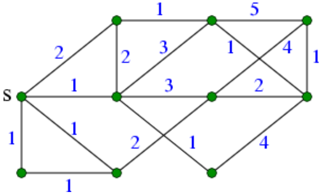
Solution : a greedy procedure (Prim/Kruskal)

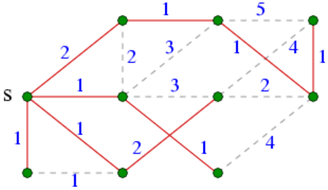
- explore edges by increasing cost in order to build a tree
- if an edge closes a cycle, reject it and take the next
- stop when all nodes reached

Easy problems (polynomial complexity)

Dynamic programming

Problem : In G, find the shortest path from s to s', where $c(e)$ is the “length” of edge e.





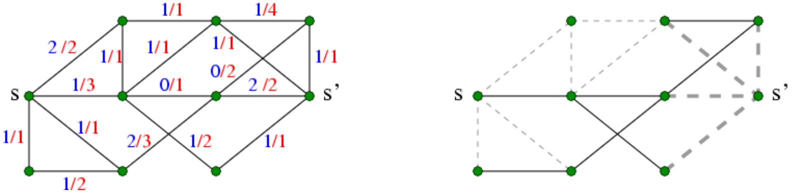
Solution : A recursion

- if the shortest path from s to s' goes through s'', then the section $s \longrightarrow s''$ is also optimal
- start from s, and find the shortest path to its neighbours, and so on recursively.

Easy problems (polynomial complexity)

Simple flow problems

Problem : Maximize the flow from s to s', where $c(e)$ is the maximal capacity of edge e.



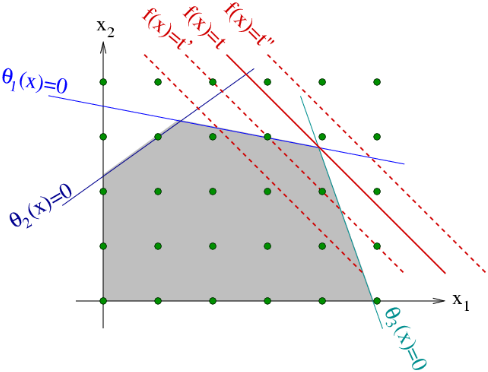
Solution : Ford-Fulkerson algorithm

- find a path from s to s', with minimal capacity $c_1 > 0$
- reduce capacities on edges of this path by c_1
- repeat until s and s' are disconnected

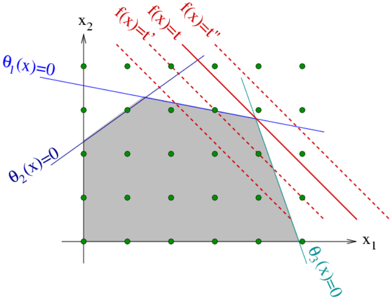
Complex problems (NP)

Integer linear program (NP)

$$\max_{x \in \mathbb{N}^d} b^t x \qquad \text{s.t.} \qquad \theta_j(x) = b_j^t x + c_j \le 0, \quad 1 \le j \le m$$



Complex problems (NP)



- the extra constraint $x \in \mathbb{N}^d$ instead of $x \in \mathbb{R}^d$ changes dramatically the complexity
- rounding the solution of a linear program may give a good approximation, or initial guess
- when the simplex is too "flat", this doesn't work
- complexity is NP in general...

Complex problems (NP)

Example: the knapsack problem, a typical integer linear programming problem

- d objects O_i of volume v_i and utility u_i,
- knapsack of limited volume v,
- select objects in order to maximize the utility of the knapsack, under the volume constraint.

Complex problems (NP)

Traveling salesman : find a hamiltonian cycle in a graph, *i.e.* visit all nodes exactly once with the shortest cycle.
Maximal coupling : set of edges in G that have no neighbour in common.
Maximal clique : find the largest clique, *i.e.* set of points that are pairwise neighbours.

Solutions by branch and bound techniques.

Some problems are even more complex : it is already hard to find an element of the domain $\mathcal{D}$.
Examples : planning problems, puzzle resolution, etc.
Addressed by constraint solving methods.

Variations calculus (not covered)

The parameter x has infinite dimension, and becomes a function.
Examples :

- Optimal shape of a rope hanged by its two extremities (minimizes potential energy).
- Shape of a toboggan that maximizes the exit speed : the brachistochrone.

Others (not covered)

Optimization multi-criterion :
Minimize *at the same time* $f_1(x)$ and $f_2(x)$ in the same value of x.

Games :

- We have K players,
- Divide $x \in \mathbb{R}^d$ into $x = (x_1, ..., x_K)$ where $x_i \in \mathbb{R}^{d_i}$, $\sum_i d_i = d$
- Player i is in charge of adjusting the (sub)vector x_i.

We distinguish

- cooperative games : players jointly minimize $f(x)$
- competitive games : player k minimizes its own $f_k(x)$.

Outline

Scalar function

$$\min_x f(x) = ax^2 + bx + c$$

- Solve $f'(x) = 0$, *i.e.* try $x^* = -\frac{b}{2a}$,
- equivalent to expressing f as $f(x) = a(x - x^*)^2 + c'$
- min or max ? Depends on the sign of a...

Vector function

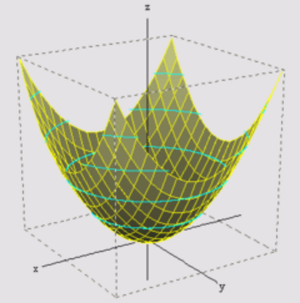
$$\min_x f(x) = x^t A x + b^t x + c$$

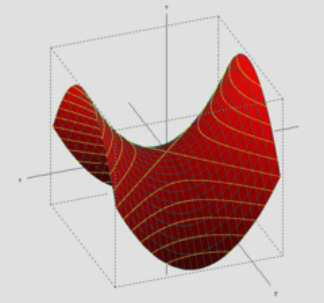
where A is a diagonal matrix $A = Diag(a_1, ..., a_d)$.

- equivalent to d independent problems $f_i = a_i x_i^2 + b_i x_i + c_i$
- joint resolution of all $f_i'(x_i) = 0$:
  equivalent to canceling the gradient

$$\nabla f(x) = \begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i} \\ \vdots \end{bmatrix} = 2Ax + b$$

- try $x^* = -\frac{1}{2} A^{-1} b$, when all $a_i \neq 0$
- equivalent to expressing f as $f(x) = (x - x^*)^t A (x - x^*) + c'$
- min or max ? depends on the signs of the a_i...





- all a_i positive : elliptic paraboloid (left), and x^* is a min
- there exist a_i of opposite signs : hyperbolic paraboloid (right), and x^* is not a min
- what if some of the a_i vanish ?

General case

$$f(x) = x^t A x + b^t x + c$$

One always has $A = P \Delta P^t$ where

- Δ is *diagonal*, formed by the eigenvalues λ_i of A
- P is *unitary*, *i.e.* $P^t P = \mathbf{1} = P P^t$,
  and formed by the eigenvectors p_i of A
- $A p_i = \lambda_i p_i$

The change of variables $y = P^t x$ takes us back to the previous case.

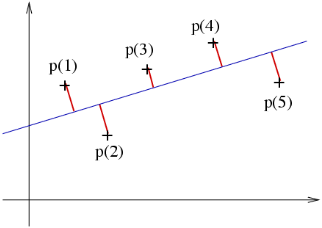
Direct resolution

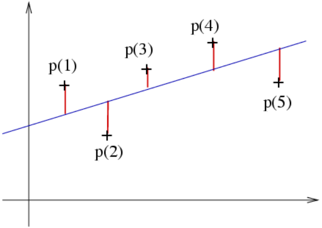
- gradient $\nabla f(x) = 2Ax + b$,
- $x^* = -\frac{1}{2}A^{-1}b$ yields a stationary point (obtained by solving a linear system)
- verify that this is a min, by checking the positivity of the eigenvalues of A

Example 1: curve fitting

Problem : adjust the line $\mathcal{L} : x_1 p_1 + x_2 p_2 + x_3 = 0$ to a cloud of N points $p(1), \ldots, p(N)$ in the plane $\mathbb{R}^2$

Distances to the line





$$D(p, \mathcal{L}) = \frac{x_1\, p_1 + x_2\, p_2 + x_3}{\sqrt{{x_1}^2 + {x_2}^2}} \qquad D_v(p, \mathcal{L}) = \left| \frac{x_1\, p_1 + x_2\, p_2 + x_3}{x_2} \right|$$

$x = (x_1, x_2, x_3)$ is defined up to a constant.
Assume $x_2 = 1$: $D_v(p, \mathcal{L}) = |x_1\, p_1 + p_2 + x_3|$

$$\min_x f(x) \quad \text{where} \quad f(x) = f(x_1, x_3) = \sum_i D_v(p(i), \mathcal{L})^2$$

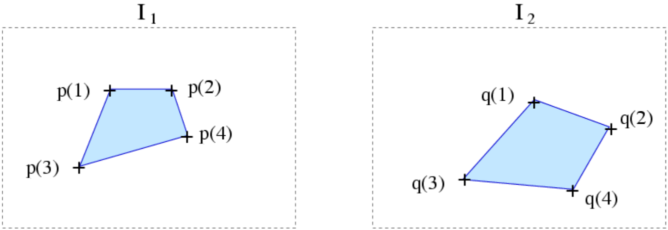
This method can be used to adjust

- a parabola $\mathcal{P}$: $p_2 = x_1 p_1 + x_2 p_1^2 + x_3$ to a cloud of points,
- any polynomial curve in p_1,
- a hyperplane $\mathcal{H}$: $x_1 p_1 + x_2 p_2 + \ldots + x_k p_k + x_{k+1} = 0$ in $\mathbb{R}^k$

Caution : very sensitive to outliers

Example 2: best image transform

Problem : Estimate the apparent movement of an object between two consecutive images. Characterized by the best affine transform that sends points $p(n)$ to points $q(n)$.



Model :

$$q = Hp + T \quad \text{where} \quad H = \begin{bmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}$$

Resolution by $\min_x \sum_n \|HP(n) + T - q(n)\|^2$

Example 3 : Best solution of a linear system

Consider the overconstrained linear system $Mx = y$
where $M \in \mathbb{R}^{n \times d}$ and $n \geq d$: more equations than unknowns.

Approximate resolution :

- minimize the norm of the error $e = y - Mx$

$$f(x) = \|Mx - y\|^2 = (Mx - y)^t (Mx - y) = x^t M^t Mx - 2y^t Mx + y^t y$$

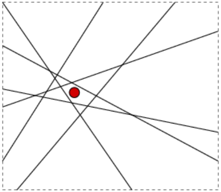
- $A = M^t M$ is a positive symmetric matrix: there is a min
- stationary point(s) : $\nabla f(x) = 2M^t Mx - 2M^t y = 0$
- if M has full rank, *i.e.* $Rank(M) = Rank(M^t M) = d$, then

$$x^* = (M^t M)^{-1} M^t y = M^\dagger y$$

Example 4 : intersection of N lines

Find a point $x = (x_1, x_2)$ such that

$$\mathcal{L}(n): \quad a_{n,1}\, x_1 + a_{n,2}\, x_2 = a_{n,0} \qquad 1 \leq n \leq N$$



The approximate resolution of this linear system gives the point that is the closest to all lines :

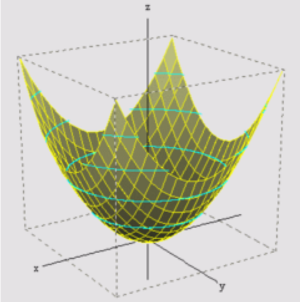
$$\min_{x_1, x_2} \sum_{n=1}^{N} (a_{n,1}\, x_1 + a_{n,2}\, x_2 - a_{n,0})^2$$

Q: when does the linear system violate $rank(M) = d = 2$?

Outline

Representing a function $f : \mathbb{R}^n \to \mathbb{R}$

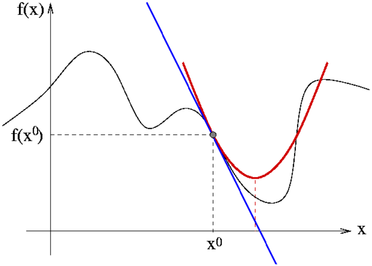
1. As a “surface” in $\mathbb{R}^{d+1}$:
   - the $d+1^{st}$ dimension z is for the value of f,
   - $z = f(x)$ defines a manifold of degree d in $\mathbb{R}^{d+1}$.
2. By its level lines in $\mathbb{R}^d$:
   - level line of “height” h : $\mathcal{L}(f,h) = \{x \in \mathbb{R}^d : f(x) = d\}$,
   - $\mathcal{L}(f,h)$ is a manifold of degree $d-1$ in $\mathbb{R}^d$,
   - similar to geographical or meteorological maps.



Taylor expansion in $\mathbb{R}$

For $f : \mathbb{R} \to \mathbb{R}$ that is C^2 (*i.e.* has a cont. 2nd order derivative), one can approximate f around any point x^0 by

$$f(x) = f(x^0) + f'(x^0)(x - x^0) + \frac{1}{2} f''(x^0)(x - x^0)^2 + o(x - x^0)^2$$



- $y = f(x^0) + f'(x^0)(x - x^0)$: tangent line
- $y = f(x^0) + f'(x^0)(x - x^0) + \frac{1}{2} f''(x^0)(x - x^0)^2$: tg parabola
- Can be used to find a local minimum of f.

Taylor expansion in $\mathbb{R}^d$: 1st order

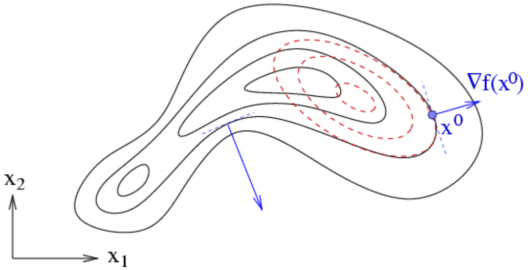
1st order : let $f : \mathbb{R}^d \to \mathbb{R}$ be a C^2 function,

$$f(x) = f(x^0) + \nabla f(x^0)^t (x - x^0) + o(\|x - x^0\|)$$

- if the move $x - x^0$ is perpendicular to the gradient $\nabla f(x^0)$, f doesn't change
- if the move $x - x^0$ is parallel to the gradient, one gets the maximal change for f

Level lines

Level line for f at $x^0 = \{x \in \mathbb{R}^d : f(x) = f(x^0)\}$



- at any point, the gradient is orthogonal to the level line
- tangent hyperplane to the “surface” $z = f(x)$ in $\mathbb{R}^{d+1}$:
  $z = f(x^0) + \nabla f(x^0)^t(x - x^0)$
- tangent hyperplane to the level line at x^0 in $\mathbb{R}^d$:
  intersection with $z = f(x^0)$, *i.e.* $\nabla f(x^0)^t(x - x^0) = 0$

Taylor expansion in $\mathbb{R}^d$: 2nd order

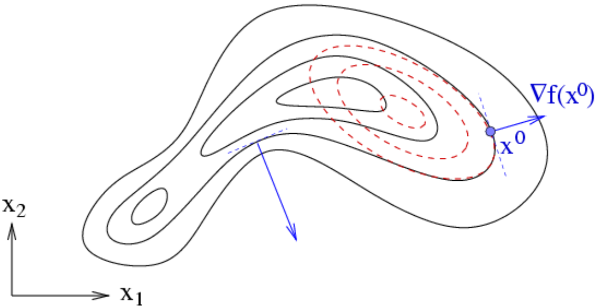
2nd order :

$$\begin{aligned} f(x) &= f(x^0) + \nabla f(x^0)^t\,(x - x^0) \\ &\quad + \frac{1}{2}(x - x^0)^t\,\nabla^2 f(x^0)\,(x - x^0) + o(\|x - x^0\|^2) \end{aligned}$$

where $\nabla^2 f(x^0)$ is the Hessian of f at x^0

$$\nabla^2 f(x^0) = \begin{bmatrix} & \vdots & \\ \cdots & \frac{\partial^2 f(x^0)}{\partial x_i \partial x_j} & \cdots \\ & \vdots & \end{bmatrix}$$

Level lines of f and of its best approximation by a quadratic form at x^0.



Characterization of a local minimum x^* of f :

- NC: it's a stationary point, $\nabla f(x^*) = 0$
- NC: the Hessian is semi-definite positive $\nabla^2 f(x^*) \geq 0$
- SC: the Hessian is non-negative $\nabla^2 f(x^*) > 0$

Directional derivatives

Consider the parametric curve/line :

$$x(t) = x^0 + tu, \qquad t \in \mathbb{R}, \quad u = \text{direction in } \mathbb{R}^d$$

1st order : For $f \circ x : \mathbb{R} \to \mathbb{R}$, one has

$$\begin{aligned} (f \circ x)'(t) &= \sum_i \frac{\partial f[x(t)]}{\partial x_i} \frac{dx_i(t)}{dt} \\ &= \nabla f[x(t)]^t \, x'(t) \\ &= \nabla f[x(t)]^t \, u \end{aligned}$$

Whence the 1st order Taylor expansion of $f \circ x$

$$(f \circ x)(t) = x^0 + t \cdot \nabla f(x^0)^t \, u + o(t)$$

2nd order : One has

$$\begin{aligned}(f \circ x)''(t) &= x'(t)^t \, \nabla^2 f[x(t)] \, x'(t) + \nabla f[x(t)] \, x''(t) \\ &= u^t \, \nabla^2 f[x(t)] \, u \; + \; 0\end{aligned}$$

Whence the 2nd order Taylor expansion of $f \circ x$

$$(f \circ x)(t) \;=\; x^0 + t \cdot \nabla f(x^0)^t \, u + \frac{1}{2} \, t^2 \cdot u^t \, \nabla^2 f(x^0) \, u + o(t^2)$$