Optimization and Simulation

Constrained optimization

Michel Bierlaire

michel.bierlaire@epfl.ch

Transport and Mobility Laboratory

Optimization: the problem

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

subject to

$$\begin{array}{rcl} h(x) & = & 0 \\ g(x) & \leq & 0 \\ x & \in & X \subseteq \mathbb{R}^n \end{array}$$

Modeling elements:

1. Decision variables: x
2. Objective function: $f : \mathbb{R}^n \to \mathbb{R}$ $(n > 0)$
3. Constraints:
   - equality: $h : \mathbb{R} \to \mathbb{R}^m$ $(m \geq 0)$
   - inequality: $g : \mathbb{R}^n \to \mathbb{R}^p$ $(p \geq 0)$
   - X is a convex set

The problem

- x_i, $i = 1, \ldots, n$, are continuous variables
- f, g and h are sufficiently differentiable
- $Y = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0$ and $x \in X\}$ is non empty

Local minimum $x^* \in Y$ is a local minimum of the above problem if there exists $\varepsilon > 0$ such that

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

Global minimum $x^* \in Y$ is a global minimum of the above problem if

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

Lagrangian

- Assume $X = \mathbb{R}^n$ in the above problem
- Consider $\lambda \in \mathbb{R}^m$
- Consider $\mu \in \mathbb{R}^p$

The function $L : \mathbb{R}^{n+m+p} \to \mathbb{R}$ defined as

$$\begin{aligned} L(x, \lambda, \mu) &= f(x) + \lambda^T h(x) + \mu^T g(x) \\ &= f(x) + \textstyle\sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x) \end{aligned}$$

is called the lagrangian function.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Dual function

- The function $q : \mathbb{R}^{m+p} \to \mathbb{R}$ defined as

$$q(\lambda, \mu) = \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$$

  is called the dual function of the optimization problem.
- Parameters λ and μ are called dual variables. x are called primal variables.
- If x^* is a global minimum of the optimization problem, then, for any $\lambda \in \mathbb{R}^m$ and any $\mu \in \mathbb{R}$, $\mu \geq 0$, we have

$$q(\lambda, \mu) \leq f(x^*).$$

Dual problem

Let $X_q \subseteq \mathbb{R}^{m+p}$ be the domain of q, that is

$$X_q = \{\lambda, \mu | q(\lambda, \mu) > -\infty\}$$

The optimization problem

$$\max_{\lambda,\mu} q(\lambda, \mu)$$

subject to

$$\mu \geq 0$$

and

$$(\lambda, \mu) \in X_q$$

is called the dual problem of the original problem, which is called the primal problem in this context.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Duality results

Weak duality theorem Let x^* be a global minimum of the primal problem, and (λ^*, μ^*) a global maximum of the dual problem. Then,

$$q(\lambda^*, \mu^*) \leq f(x^*).$$

Convexity-concavity of the dual problem

- The objective function of the dual problem is concave.
- The feasible set of the dual problem is convex.

Outline

- Feasible directions, constraint qualification
- Optimality conditions
  - Convex constraints
  - Lagrange multipliers: necessary conditions
  - Lagrange multipliers: sufficient conditions
- Algorithms
  - Constrained Newton
  - Interior point
  - Augmented lagrangian
  - Sequential quadratic programming

Feasible directions

Definitions:

- $x \in \mathbb{R}^n$ is a feasible point if it verifies the constraints
- Given x feasible, d is a feasible direction in x if there is $\eta > 0$ such that
$$x + \alpha d$$
is feasible for any $0 \leq \alpha \leq \eta$.

Convex constraints:

- Let $X \subseteq \mathbb{R}^n$ be a convex set, and $x, y \in X$, $x \neq y$.
- The direction
$$d = y - x$$
is feasible in x.
- Moreover, for each $0 \leq \alpha \leq 1$, $\alpha x + (1 - \alpha) y$ is feasible.

Feasible directions

Corollary:

- Let $X \subseteq \mathbb{R}^n$
- Let x be an interior point, that is there exists $\varepsilon > 0$ such that

$$\|x - z\| \leq \varepsilon \Longrightarrow z \in X.$$

- Then, any direction d is feasible in x.

Feasible sequences

- Consider the generic optimization problem
- Let $x^+ \in \mathbb{R}^n$ be a feasible point
- The sequence $(x_k)_k$ is said to be feasible in x^+ if
  - $\lim_{k\to\infty} x_k = x^+$,
  - $\exists k_0$ such that x_k is feasible if $k \geq k_0$,
  - $x_k \neq x^+$ for all k.

Feasible sequence: example

- One equality constraint

$$h(x) = x_1^2 - x_2 = 0,$$

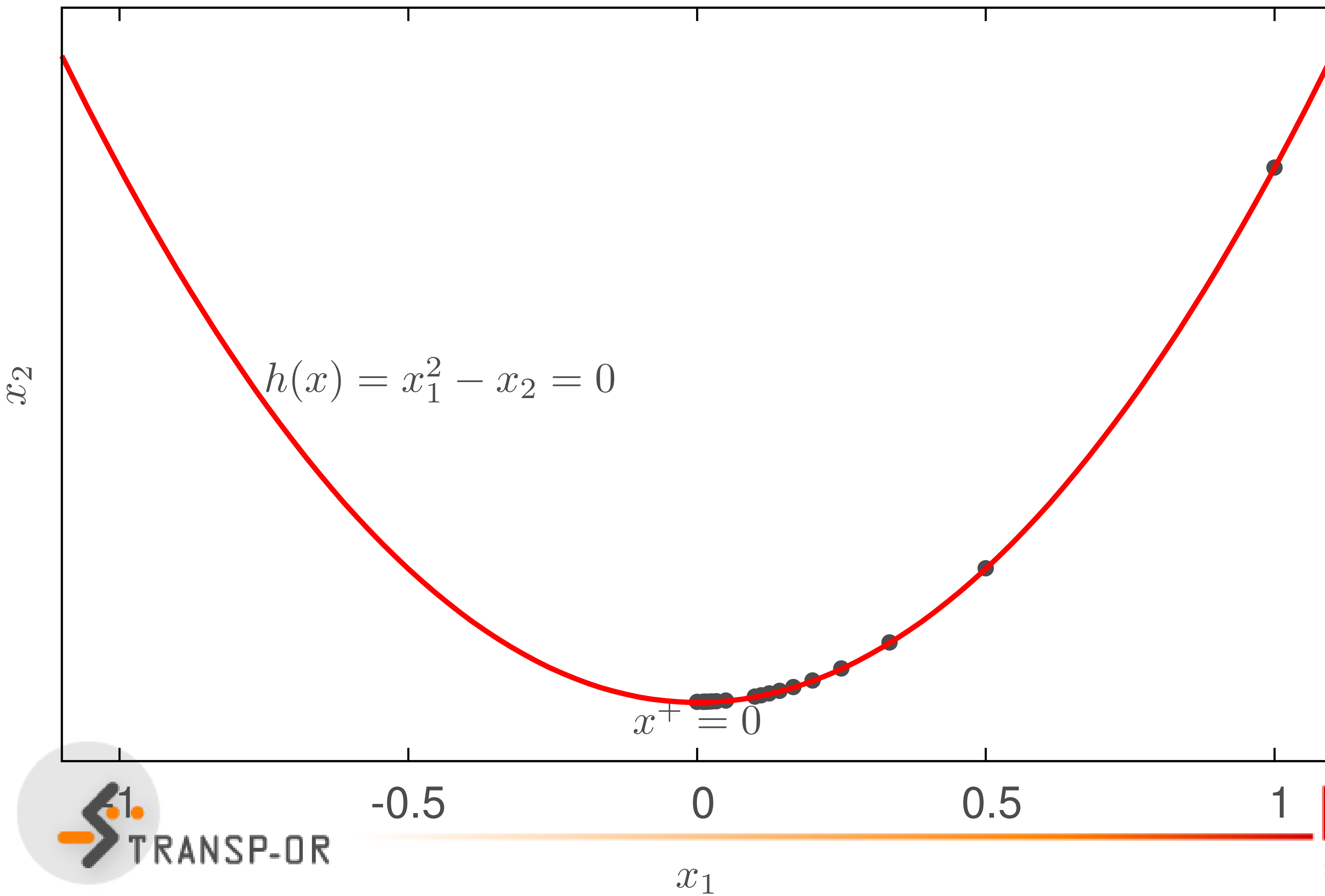
- Feasible point: $x^+ = (0,0)^T$
- Feasible sequence:

$$x_k = \begin{pmatrix} \frac{1}{k} \\ \frac{1}{k^2} \end{pmatrix}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Feasible sequence: example



Feasible limiting direction

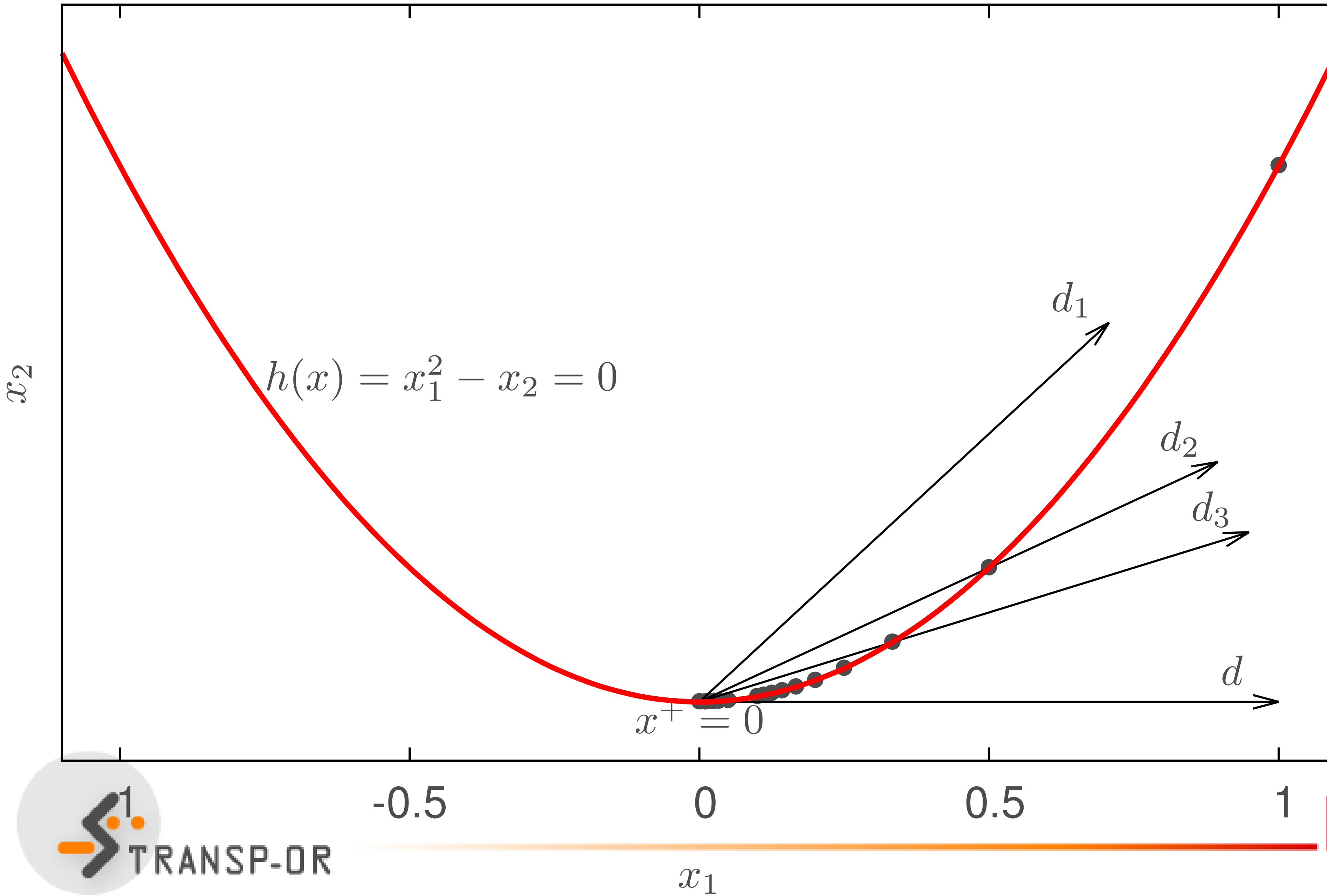
Idea: consider the sequence of directions

$$d_k = \frac{x_k - x^+}{\|x_k - x^+\|},$$

and take the limit.

- Directions d_k are not necessarily feasible
- The sequence may not always converge
- Subsequences must then be considered

Feasible limiting direction: example



Feasible limiting direction: example

- Constraint: $h(x) = x_1^2 - x_2 = 0$
- Feasible point: $x^+ = (0,0)^T$
- Feasible sequence:

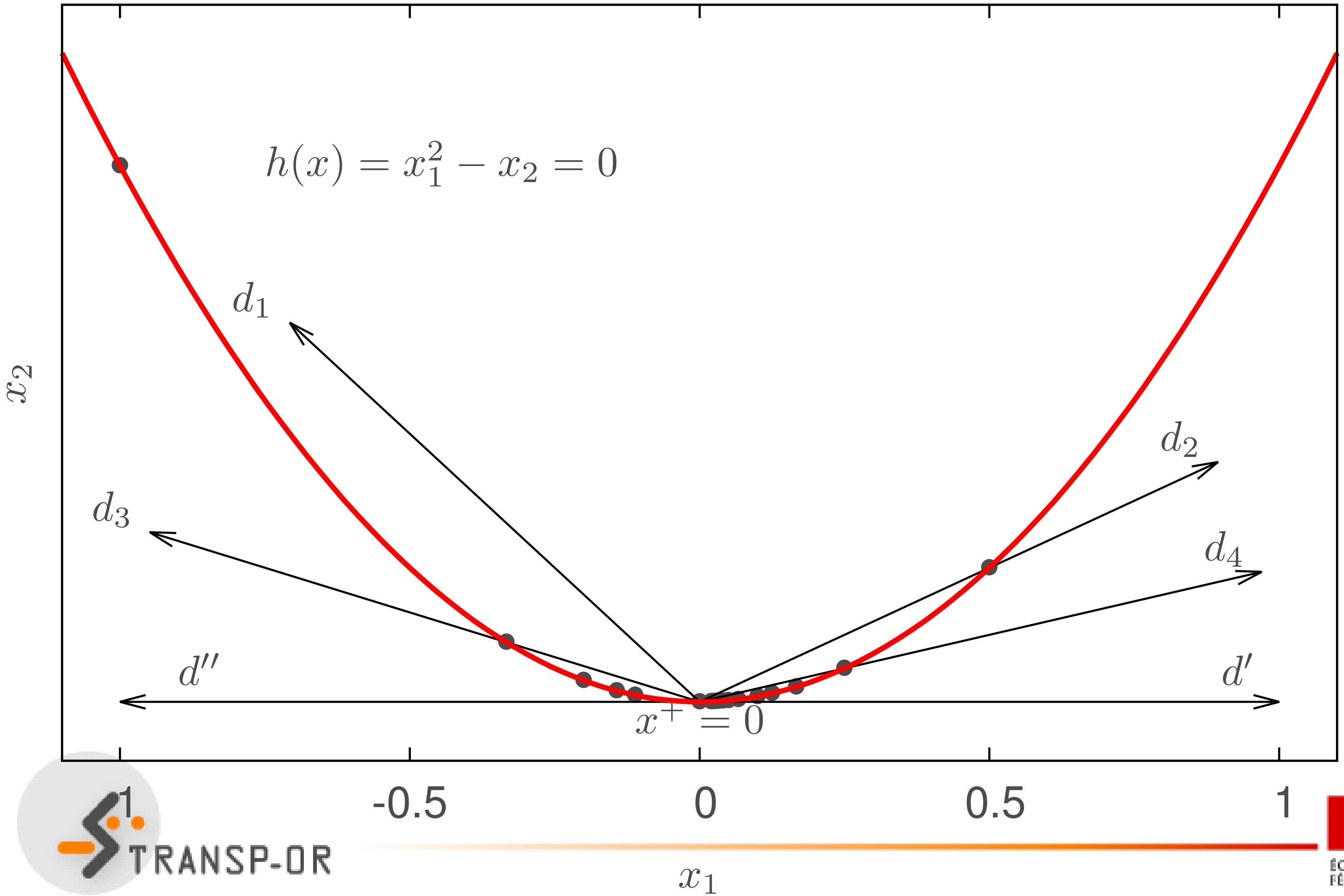
$$x_k = \begin{pmatrix} \frac{(-1)^k}{k} \\ \frac{1}{k^2} \end{pmatrix}$$

- Sequence of directions:

$$d_k = \begin{pmatrix} \frac{(-1)^k k}{\sqrt{k^2+1}} \\ \frac{1}{\sqrt{k^2+1}}, \end{pmatrix}$$

- Two limiting directions

Feasible limiting direction: example



Feasible limiting direction

- Consider the generic optimization problem
- Let $x^+ \in \mathbb{R}^n$ be feasible
- Let $(x_k)_k$ be a feasible sequence in x^+
- Then, $d \neq 0$ is a *feasible limiting direction* in x^+ for the sequence $(x_k)_k$ if there exists a subsequence $(x_{k_i})_i$ such that

$$\frac{d}{\|d\|} = \lim_{i \to \infty} \frac{x_{k_i} - x^+}{\|x_{k_i} - x^+\|}.$$

Notes:

- It is sometimes called a *tangent* direction.
- Any feasible direction d is also a limiting feasible direction, for the sequence

$$x_k = x^+ + \frac{1}{k} d$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Cone of directions

- Consider the generic optimization problem
- Let $x^+ \in \mathbb{R}^n$ be feasible
- The set of directions d such that

$$d^T \nabla g_i(x^+) \leq 0, \quad \forall i = 1, \ldots, p \text{ such that } g_i(x^+) = 0,$$

  and

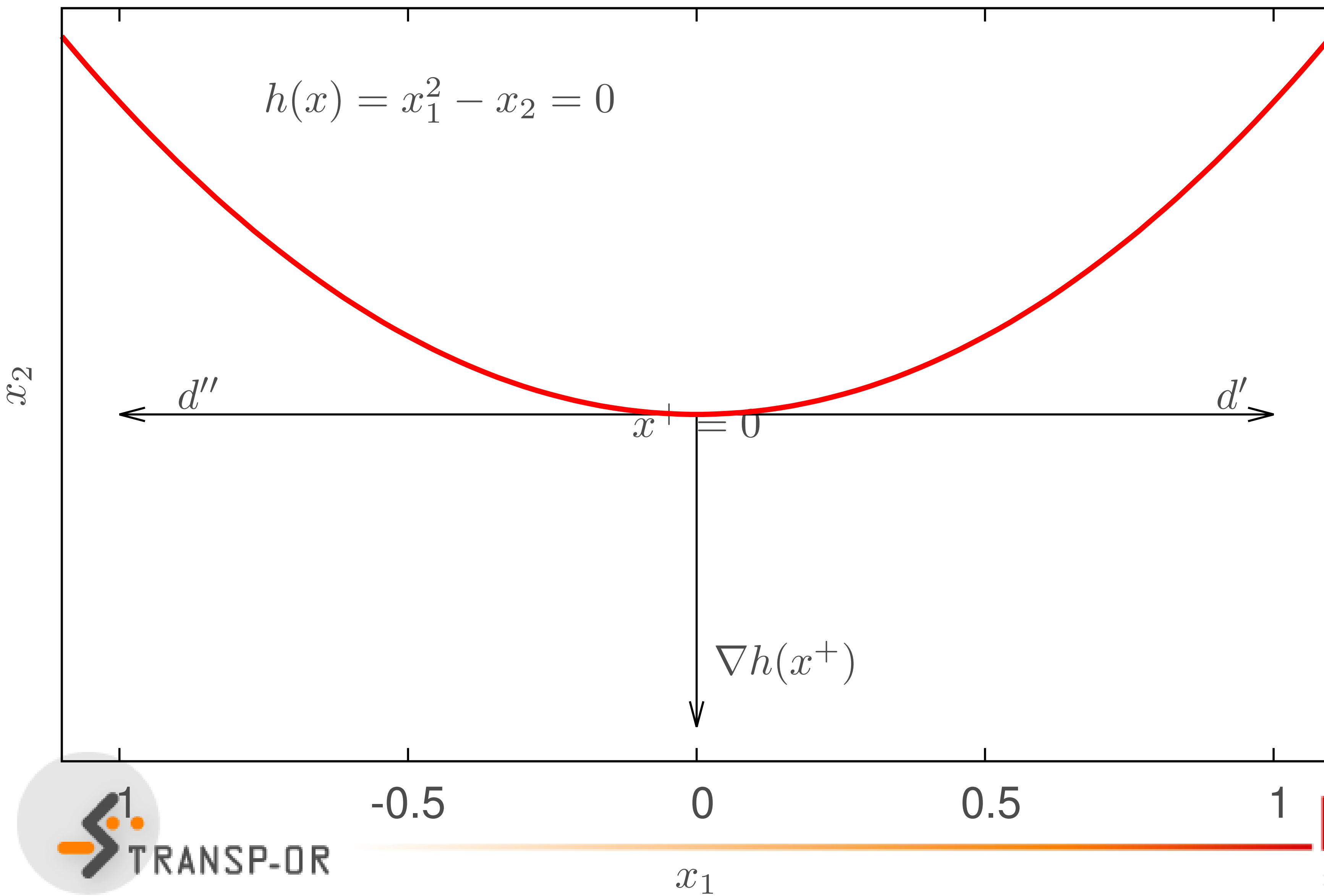
$$d^T \nabla h_i(x^+) = 0, \quad i = 1, \ldots, m,$$

  as well as their multiples αd, $\alpha > 0$, is the *cone of directions* at x^+.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Cone of directions



Cone of directions

Theorem:

- Consider the generic optimization problem
- Let $x^+ \in \mathbb{R}^n$ be feasible
- If d is a limiting feasible direction at x^+
- Then d belongs to the cone of directions at x^+

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Constraint qualification

Definition:

- Consider the generic optimization problem
- Let $x^+ \in \mathbb{R}^n$ be feasible
- The *constraint qualification* condition is verified if every direction in the cone of directions at x^+ is a feasible limiting direction at x^+.

This is verified in particular

- if the constraints are linear, or
- if the gradients of the constraints active at x^+ are linearly independent.

Optimality conditions

Necessary condition for the generic problem:

- Let x^* be a local minimum of the generic problem
- Then

$$\nabla f(x^*)^T d \geq 0$$

for each direction d which is feasible limiting at x^*.

Intuition: no "feasible" direction is a descent direction

Optimality conditions: convex problem (I)

Consider the problem

$$\min_x f(x)$$

subject to

$$x \in X \subseteq \mathbb{R}^n$$

where X is convex and not empty.

- If x^* is a local minimum of this problem
- Then, for any $x \in X$,

$$\nabla f(x^*)^T (x - x^*) \geq 0.$$

Optimality conditions: convex problem (II)

- Assume now that X is convex and closed.
- For any $y \in \mathbb{R}^n$, we note by $[y]^P$ the projection of y on X.
- If x^* is a local minimum, then

$$x^* = [x^* - \alpha \nabla f(x^*)]^P \quad \forall \alpha > 0.$$

- Moreover, if f is convex, the condition is sufficient.

Note: useful when the projection is easy to compute (e.g. bound constraints)

Optimality conditions: Karush-Kuhn-Tucker

The problem:

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

subject to

$$\begin{array}{rclr} h(x) & = & 0 & [h : \mathbb{R}^n \to \mathbb{R}^m] \\ g(x) & \leq & 0 & [g : \mathbb{R}^n \to \mathbb{R}^p] \\ x & \in & X & = \mathbb{R}^n \end{array}$$

- Let x^* be a local minimum
- Let L be the Lagrangian

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x).$$

- Assume that the constraint qualification condition is verified.
- Then...

Optimality conditions: Karush-Kuhn-Tucker

... there exists a unique $\lambda^* \in \mathbb{R}^m$ and a unique $\mu^* \in \mathbb{R}^p$ such that

$$\nabla_x L(x^*, \lambda^*, \mu^*) = \nabla f(x^*) + (\lambda^*)^T \nabla h(x^*) + (\mu^*)^T \nabla g(x^*) = 0,$$

$$\mu_j^* \geq 0 \quad j = 1, \ldots, p,$$

and

$$\mu_j^* g_j(x^*) = 0 \quad j = 1, \ldots, p.$$

If f, g and h are twice differentiable, we also have

$$\begin{array}{l} y^T \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) y \geq 0 \quad \forall y \neq 0 \text{ such that} \\ \quad y^T \nabla h_i(x^*) = 0 \quad i = 1, \ldots, m \\ \quad y^T \nabla g_i(x^*) = 0 \quad i = 1, \ldots, p \text{ such that } g_i(x^*) = 0. \end{array}$$

KKT: sufficient conditions

Let $x^* \in \mathbb{R}^n$, $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ be such that

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0$$

$$h(x^*) = 0, \quad g(x^*) \leq 0$$

$$\mu^* \geq 0,$$

$$\mu_j^* g_j(x^*) = 0 \quad \forall j, \quad \mu_j^* > 0 \quad \forall j \text{ such that } g_i(x^*) = 0.$$

$$y^T \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) y > 0 \quad \forall y \neq 0 \text{ such that}$$
$$y^T \nabla h_i(x^*) = 0 \quad i = 1, \ldots, m$$
$$y^T \nabla g_i(x^*) = 0 \quad i = 1, \ldots, p \text{ such that } g_i(x^*) = 0.$$

Then x^* is a strict local minimum of the problem.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Algorithms

- Constrained Newton
- Interior point
- Augmented lagrangian
- Sequential quadratic programming

Here: we give the main ideas.

Constrained Newton

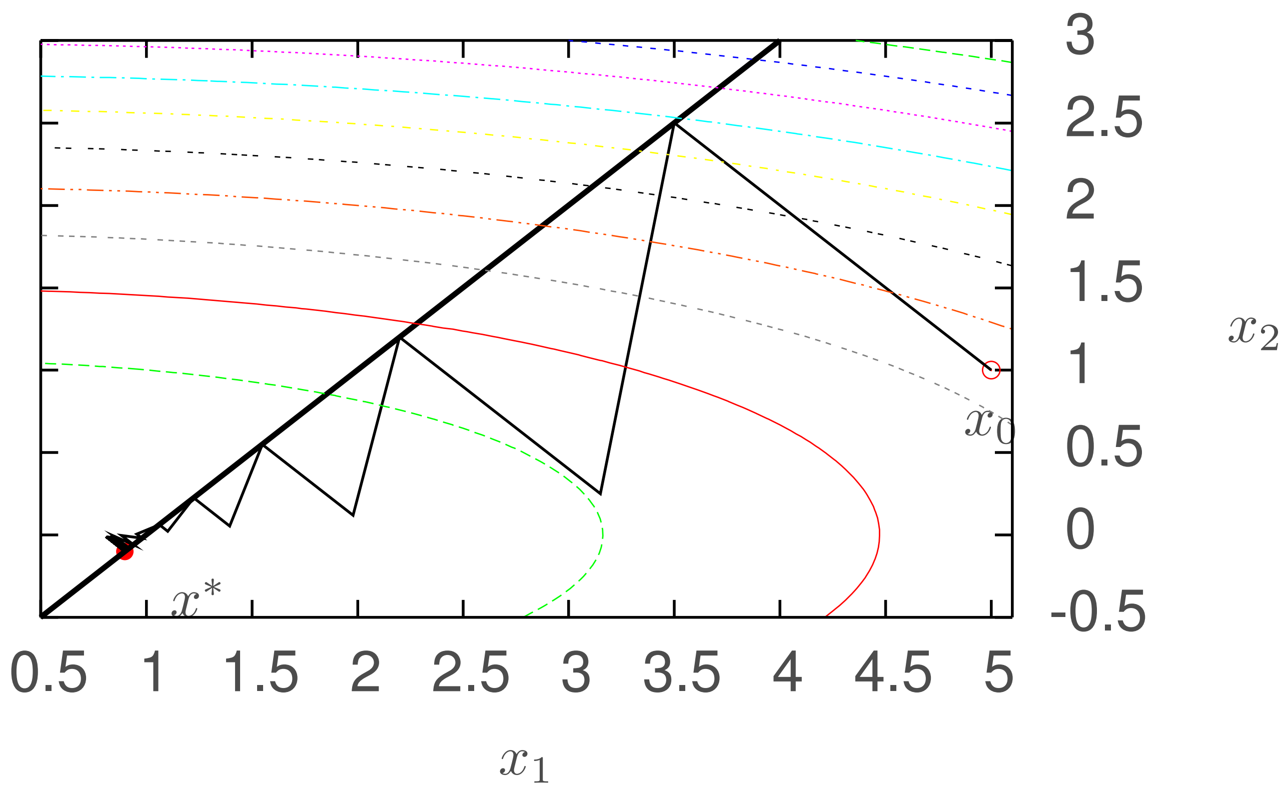
Context:

- Problem with a convex constraint set.
- Assumption: it is easy to project on the set.
- Examples: bound constraints, linear constraints.

Main idea:

- In the unconstrained case, Newton = preconditioned steepest descent
- Consider first the projected gradient method
- Precondition it.

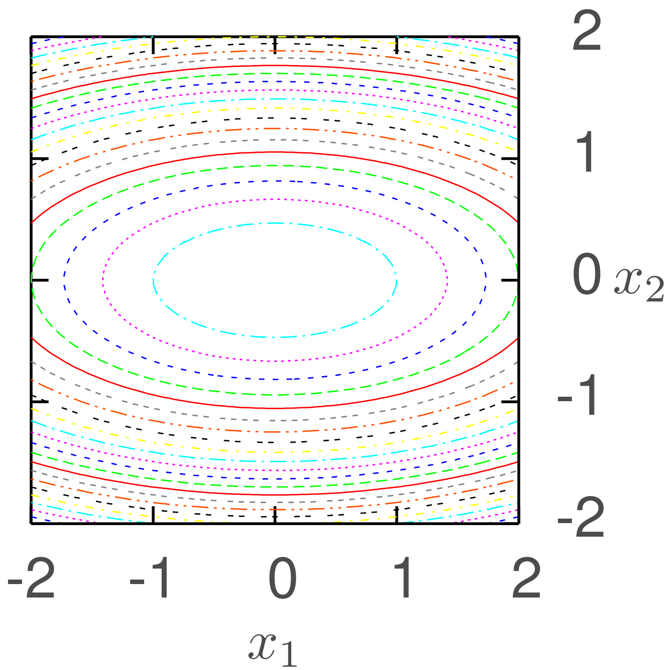
Projected gradient method

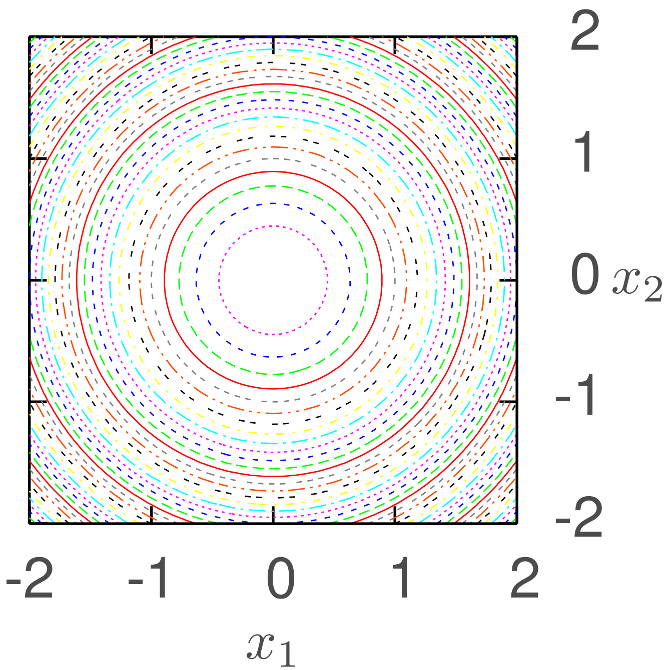


Condition number

- Consider $\nabla^2 f(x)$ positive definite.
- Let λ_1 be the largest eigenvalue, and λ_n the smallest.
- The condition number is equal to λ_1/λ_n.
- Geometrically, it is the ratio between the largest and the smallest curvature.
- The closest it is to one, the better.

Condition number





Cond = 9/2

Cond = 1

Preconditioning

Preconditioning = appropriate change of variables.

- Let $M \in \mathbb{R}^{n\times n}$ be invertible.
- Change of variables = linear application $x' = Mx$.

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

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

Now, consider $\nabla^2 f(x) = LL^T$, and $x' = L^T x$. Then,

$$\begin{array}{rcl}\nabla^2 \tilde{f}(x') & = & L^{-1}\nabla^2 f(x)L^{-T} \\ & = & L^{-1}LL^T L^{-T} \\ & = & I.\end{array}$$

Readings

- Bierlaire (2006) Chapter 18.
- Bertsekas (1999) Section 2.3.

Algorithms

- Constrained Newton
- Interior point
- Augmented lagrangian
- Sequential quadratic programming

Interior point methods

Motivation:

- At an interior point, every direction is feasible.
- It gives more freedom to the algorithm.

Main ideas:

- Focus first on being feasible.
- Then try to become optimal.

Barrier functions

- Let $X \subset \mathbb{R}^n$ be a closed set.
- Let $g : \mathbb{R}^n \to \mathbb{R}^m$ a convex function.
- Let $\mathcal{S}$ be the set of interior points for g:

$$\mathcal{S} = \{x \in \mathbb{R}^n | x \in X, g(x) < 0\}.$$

- A function barrier $B : \mathcal{S} \to \mathbb{R}$ is continuous and such that

$$\lim_{x \in S, g(x) \to 0} B(x) = +\infty.$$

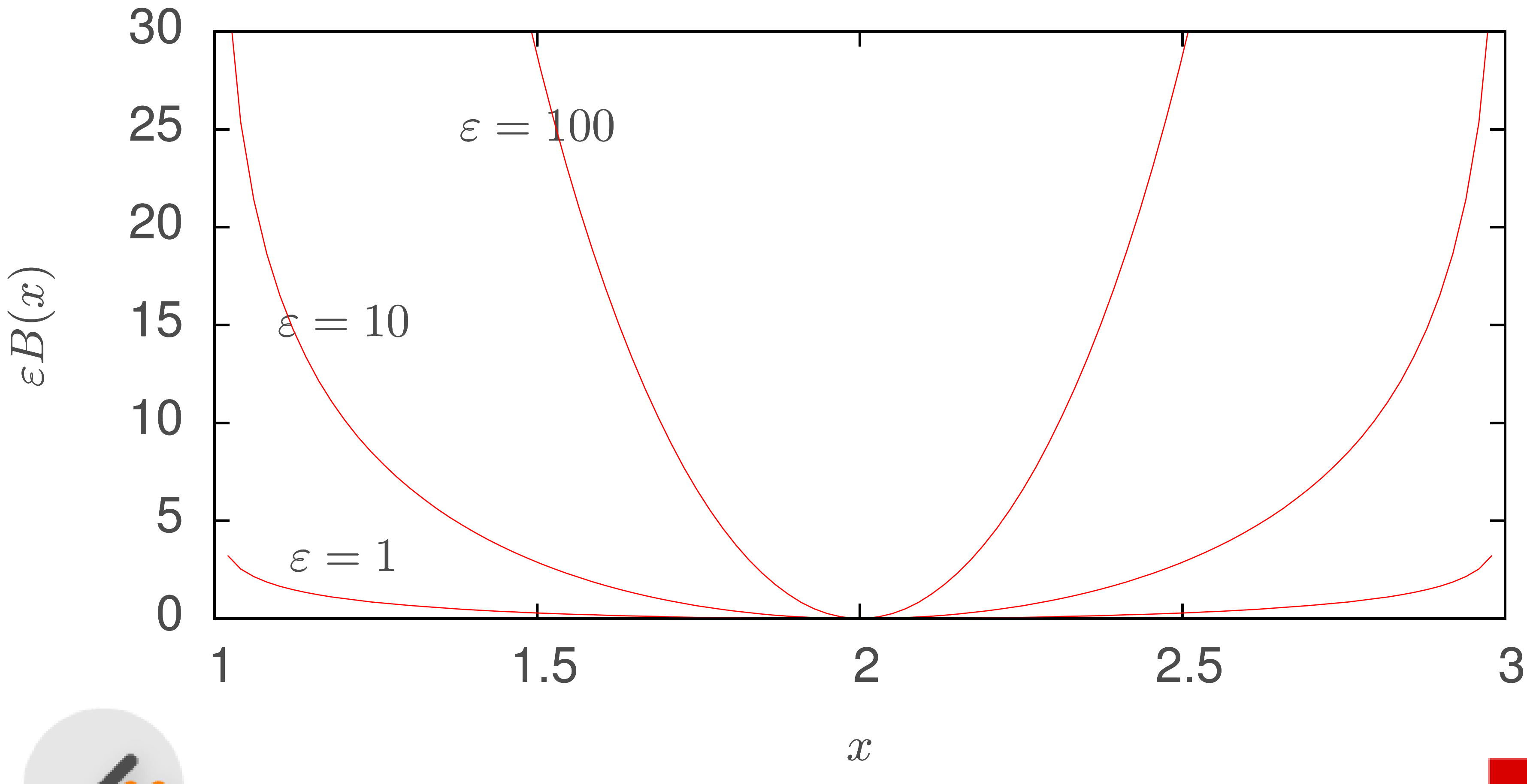
- Examples:

$$B(x) = -\sum_{j=1}^{m} \ln(-g_j(x))$$

$$B(x) = -\sum_{j=1}^{m} \frac{1}{g_j(x)}.$$

Barrier functions: example (logarithmic)

$$1 \leq x \leq 3 \implies B(x) = -\ln(x-1) - \ln(3-x).$$



Barrier methods

- Define a sequence of parameters $(\varepsilon_k)_k$ such that
  - $0 < \varepsilon_{k+1} < \varepsilon_k$, $k = 0, 1, \ldots$
  - $\lim_k \varepsilon_k = 0$.
- At each iteration, solve

$$x_k = \operatorname{argmin}_{x \in \mathcal{S}} f(x) + \varepsilon_k B(x).$$

Issues:

- The subproblem should be easy to solve.
- In particular, we should rely on unconstrained optimization. A descent method should not go outside the constraints, thanks to the barrier.
- The speed of convergence of $(\varepsilon_k)_k$ is critical.

Typical applications: linear programming, convex programming

Readings

- Bierlaire (2006) Chapter 19.
- Bertsekas (1999) Section 4.1.

See also: Wright, S. J. (1997) *Primal-Dual Interior-Point Methods*, SIAM

Algorithms

- Constrained Newton
- Interior point
- Augmented lagrangian
- Sequential quadratic programming

Augmented Lagrangian

Main ideas:

- Focus first on reducing the objective function, even if constraints are violated.
- Then recover feasibility.
- Inspired by the optimality conditions.

We assume that the problem has only equality constraints

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

subject to

$$h(x) = 0 \quad [h : \mathbb{R}^n \to \mathbb{R}^m]$$

Augmented Lagrangian

- Solve a sequence of unconstrained optimization problems.
- Penalize the constraint violation using
  - a lagrangian relaxation, and
  - a quadratic penalty function.

  Augmented lagrangian

$$L_c(x,\lambda) = f(x) + \lambda^T h(x) + \frac{c}{2}\|h(x)\|^2.$$

Augmented Lagrangian: lagrangian relaxation

- If λ^* is known (see optimality conditions).
- Then the solution is given by solving the unconstrained problem

$$\min_{x \in \mathbb{R}^n} L_c(x, \lambda^*) = f(x) + (\lambda^*)^T h(x) + \frac{c}{2} \|h(x)\|^2.$$

  with c sufficiently large.
- Unfortunately, λ^* is not known by default.
- But we will be able to approximate it.

Augmented Lagrangian: quadratic penalty

- If c becomes large enough, any non feasible point will be non optimal for

$$\min_{x \in \mathbb{R}^n} L_c(x,\lambda) = f(x) + \lambda^T h(x) + \frac{c}{2}\|h(x)\|^2,$$

  for any λ.
- Consider a sequence $(c_k)_k$ such that

$$\lim_{c_k \to \infty} = +\infty.$$

- Then, for a given λ, the sequence

$$x_k = \operatorname{argmin}_{x \in \mathbb{R}^n} L_{c_k}(x,\lambda)$$

  converges to a solution of the constrained problem.

Augmented Lagrangian: quadratic penalty

Main issue:

- If c_k is large, $L_{c_k}(x,\lambda)$ is ill-conditioned.
- Methods for unconstrained optimization become slow, or may even fail to converge.
- But... if λ is close to λ^*, no need for large values of c_k.

Theoretical result:

- Under relatively general conditions, the sequence

$$\lim_k \lambda_k + c_k h(x_k)$$

converges to λ^*.

Augmented Lagrangian: algorithm

1. Use an unconstrained optimization algorithm to solve

$$x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} L_{c_k}(x, \lambda_k)$$

   to a given precision ε_k.
2. If x_{k+1} is close to feasibility:
   - update the estimate of the multipliers: $\lambda_{k+1} = \lambda_k + c_k h(x_k)$
   - keep $c_k = c_{k+1}$,
   - require more precision: $\varepsilon_{k+1} = \varepsilon_k / c_k$.
3. If x_{k+1} is far from feasibility:
   - keep $\lambda_{k+1} = \lambda_k$
   - increase c_k,
   - relax the precision: $\varepsilon_{k+1} = \varepsilon_0 / c_{k+1}$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Readings

- Bierlaire (2006) Chapter 20.
- Bertsekas (1999) Section 4.2.

Sequential quadratic programming

Main ideas:

- Apply Newton's method to solve the necessary optimality conditions

$$\nabla L(x^*, \lambda^*) = 0.$$

- One iteration amounts to solve a quadratic problem.
- Enforce global convergence with a merit function.

We assume that the problem has only equality constraints

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

subject to

$$h(x) = 0 \;\; [h : \mathbb{R}^n \to \mathbb{R}^m]$$

Sequential quadratic programming

Lagrangian and derivatives:

$$L(x,\lambda) = f(x) + \lambda^T h(x).$$

$$\nabla L(x,\lambda) = \begin{pmatrix} \nabla_x L(x,\lambda) \\ h(x) \end{pmatrix},$$

$$\nabla^2 L(x,\lambda) = \begin{pmatrix} \nabla^2_{xx} L(x,\lambda) & \nabla h(x) \\ \nabla h(x)^T & 0 \end{pmatrix}.$$

Newton's method: at each iteration, find d such that

$$\nabla^2 L(x_k,\lambda_k) d = -\nabla L(x_k,\lambda_k),$$

Sequential quadratic programming

It can be shown that it is equivalent to solving the following quadratic problem

$$\min_d \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla^2_{xx} L(x_k, \lambda_k) d$$

subject to

$$\nabla h(x_k)^T d + h(x_k) = 0.$$

- An analytical solution can be derived for this problem.
- In practice, dedicated iterative algorithms are used.

Sequential quadratic programming

- Newton's method is not globally convergent.
- The same applies to the SQP method described above.
- Idea: apply similar globalization techniques than for unconstrained optimization (line search, trust region).
- Main concept: reject a candidate if it is not sufficiently better than the current one.
- But what does "better" mean?
- Two (potentially) conflicting objectives:
  - decrease $f(x)$
  - bring $h(x)$ close to 0.

Sequential quadratic programming

- Solution: combine them into a *merit* function

$$\phi_c(x) = f(x) + c\|h(x)\|_1 = f(x) + c\sum_{i=1}^{m} |h_i(x)|.$$

- For instance, use Wolfe's conditions on the merit function. But...
- technical difficulties: need to
  - guarantee that d is a descent direction for ϕ_c,
  - deal with the non differentiability of ϕ_c.

Sequential quadratic programming

Notes:

- Differentiable merit functions could also be used.
- They may involve singularities.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Readings

- Bierlaire (2006) Chapter 21.
- Bertsekas (1999) Section 4.3.