Large-Scale Optimization

Interior-point methods

Jordi Castro
MSc in Statistics and Operations Research

Department of Statistics and Operations Research
School of Mathematics and Statistics
Universitat Politècnica de Catalunya
Barcelona

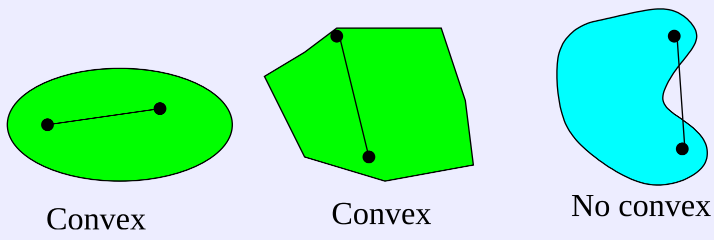
Contents

Convex set

- Set X is convex if for all $x_1, x_2 \in X$

$$\alpha x_1 + (1-\alpha)x_2 \in X \quad 0 \le \alpha \le 1$$

- Graphically, segment $\overline{x_1 x_2}$ belongs to X:



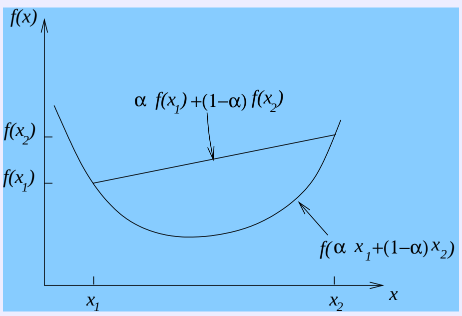
Convex function

- Function $f(x)$ is convex in convex set X if for all $x_1, x_2 \in X$

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha)f(x_2) \quad 0 \le \alpha \le 1$$

(strictly convex if $<$ instead of $\le$)

- Graphically:



- $f(x)$ is convex if Hessian $\nabla^2 f(x)$ is positive semidefinite (strictly convex if $\nabla^2 f(x)$ is positive definite).

Convex problem and global optimum

- Problem
$$\begin{array}{ll} \min & f(x) \\ \text{s.to} & x \in X \end{array}$$
is convex if $f(x)$ is convex function and X is convex set.
- If $g(x)$ is convex function then $X = \{x : g(x) \leq b\}$ is convex set (similarly, if $g(x)$ is concave function, $X = \{x : g(x) \geq b\}$ is convex set).
- It can be shown that: Any local optimum of a convex problem is also a global optimum. If $f(x)$ is strictly convex, then the optimum is unique.
- Are LPs a convex optimization problem?

Necessary optimality conditions

- Given problem
$$\begin{array}{lll} \min & f(x) & \\ \text{s.to} & h(x) = 0 & [h_i(x) = 0 \quad i = 1, \ldots, m] \\ & g(x) \leq 0 & [g_j(x) \leq 0 \quad j = 1, \ldots, p], \end{array}$$
and its Lagrangian function
$$L(x, \lambda, \mu) = f(x) + \lambda^\top h(x) + \mu^\top g(x),$$
- Linearly Independent Constraint Qualification (LICQ): gradients of active constraints at x^* are linearly independent (to characterize limiting directions of feasible sequences to x^*).
- **Necessary conditions.** If x^* is a regular (i.e., LICQ holds) local optimum then there are unique vectors $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that:

  First order conditions (KKT)

  (i) $h(x^*) = 0$, $g(x^*) \leq 0$ [primal feasibility]

  (ii) $\nabla_x L(x, \lambda, \mu) = \nabla f(x^*) + \lambda^\top \nabla h(x^*) + \mu^\top \nabla g(x^*) = 0$ [dual feasibility]

  (iii) $\mu \geq 0$ and $\mu^\top g(x^*) = 0$ (if $g_i(x^*)$ is inactive then $\mu_i = 0$) [complementarity]

  Second order conditions

  (iv) $y^\top \nabla^2_{xx} L(x^*, \lambda, \mu) y \geq 0$, for all y such that $\nabla h(x^*) y = 0$ and $\nabla g_i(x^*) y = 0, i \in \{j : g_j(x^*) = 0\}$

Sufficient optimality conditions

- **Sufficient conditions** Point x^* is local optimum if:

  First order conditions (KKT)

  (i) $h(x^*) = 0$, $g(x^*) \leq 0$ [primal feasibility]

  (ii) $\nabla_x L(x, \lambda, \mu) = \nabla f(x^*) + \lambda^\top \nabla h(x^*) + \mu^\top \nabla g(x^*) = 0$ [dual feasibility]

  (iii) $\mu \geq 0$ i $\mu^\top g(x^*) = 0$ (if $g_i(x^*)$ is inactive then $\mu_i = 0$) [complementarity]

  Second order conditions

  (iv) $y^\top \nabla^2_{xx} L(x^*, \lambda, \mu) y > 0$, for all y such that $\nabla h(x^*) y = 0$ and $\nabla g_i(x^*) y = 0, i \in \{j : g_j(x^*) = 0, \mu_j > 0\}$

- Main difference between 2nd order necessary and sufficient conditions:
  - regularity of x^* (LICQ) not needed
  - condition (iv):

    $y^\top \nabla^2_{xx} L(x^*, \lambda, \mu) y > 0$ [sufficient]

    $y^\top \nabla^2_{xx} L(x^*, \lambda, \mu) y \geq 0$ [necessary]

- If problem is convex, first order necessary conditions are also sufficient.

Optimality conditions: Example (I)

Problem
Given a segment of length a we want to divide it into two parts, such that the area of the squares whose side is each of these two parts is minimized.
Solution:

- Variables: x_1 i x_2
- Formulation:

$$\begin{array}{ll} \min & x_1^2 + x_2^2 \\ \text{s.to} & x_1 + x_2 = a \\ & x_1 \geq 0 \quad x_2 \geq 0 \end{array} \quad \equiv \quad \begin{array}{ll} \min & x_1^2 + x_2^2 \\ \text{s.to} & x_1 + x_2 = a \\ & -x_1 \leq 0 \quad - x_2 \leq 0 \end{array}$$

- Lagrangian:

$$L(x_1, x_2, \lambda, \mu) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - a) - \mu_1 x_1 - \mu_2 x_2$$

Optimality conditions: Example (II)

- First order necessary conditions:

$$\begin{array}{ll} (i) & x_1 + x_2 = a, \quad x_1 \geq 0, \quad x_2 \geq 0 \\ (ii.x_1) & \nabla_{x_1} L() = 2x_1 + \lambda - \mu_1 = 0 \\ (ii.x_2) & \nabla_{x_2} L() = 2x_2 + \lambda - \mu_2 = 0 \\ (iii) & \mu_i \geq 0, \mu_i = 0 \text{ si } x_i > 0, i = 1,2 \end{array}$$

- Four cases to be analyzed, depending on whether $x_1, x_2 \geq 0$ are active:
  - $x1 = 0$ $x_2 = 0$. Not possible, infeasible for $x_1 + x_2 = a$.
  - $x_1 > 0$, $x_2 = 0$. Then $x_1 = a$ i $\mu_1 = 0$. Solving

$$\begin{array}{ll} 2a + \lambda & = 0 \\ \lambda - \mu_2 & = 0 \end{array}$$

    we have $\mu_2 = \lambda = -2a$, which violates $\mu_2 \geq 0$.
  - $x_1 = 0$, $x_2 > 0$. Symmetric to previous case.
  - $x_1 > 0$, $x_2 > 0$. Then $\mu_1 = \mu_2 = 0$ and solving

$$\begin{array}{ll} x_1 + x_2 = a & \\ 2x_1 + \lambda & = 0 \\ 2x_2 + \lambda & = 0 \end{array}$$

    we have $x_1 = x_2 = a/2$, $\lambda = -a$, candidate to solution.

Optimality conditions: Example (III)

- Second order sufficient conditions:

$$\nabla^2_{xx} L() = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

  which is positive definite for all y, not only for y such that $\nabla(x^*)y = 0$.
- Therefore the candidate point is the optimal solution.
- And a global solution, since this problem is convex.

LP optimality conditions

- LP

$$\begin{array}{lll} \min & c^\top x & \\ \text{s.to} & Ax = b & [\pi] \\ & -x \leq 0 & [\mu] \end{array}$$

- Lagrangian

$$L(x, \pi, \mu) = c^\top x + \pi^\top (Ax - b) - \mu^\top x$$

- KKT conditions

$$\begin{array}{l} Ax = b, \quad x \geq 0 \\ c^\top + \pi^\top A - \mu^\top = 0 \\ \mu^\top x = 0, \quad \mu \geq 0 \end{array}$$

- Defining $\lambda = -\pi$ (no sign restriction) rewrite KKT:

$$\begin{array}{ll} Ax = b, x \geq 0 & \text{[primal feasibility]} \\ A^\top \lambda + \mu = c, \mu \geq 0 & \text{[dual feasibility]} \\ \mu^\top x = 0 & \text{[complementarity]} \end{array}$$

Simplex algorithm: a method for solving KKT

- KKT conditions

$$\begin{array}{l} Ax = b \\ A^\top \lambda + \mu = c \\ \mu^\top x = 0 \\ x \geq 0, \mu \geq 0 \end{array}$$

- In primal simplex μ named reduced cost and λ dual variables.
- Simplex looks for a (basic, nonbasic) partition satisfying KKT:

$$\begin{array}{l} x = [x_B \;\; x_N] \;\; x_N = 0 \;\; x_B > 0 \\ A = [B \;\; N], \;\; Ax = Bx_B = b \\ \mu_B = 0 \Rightarrow B^\top \lambda = -c_B \\ \mu_N = c_N - N^\top \lambda \quad \text{[reduced cost definition]} \\ \mu_N \geq 0 \Rightarrow \text{ simplex optimality condition} \end{array}$$

- Complementarity $\mu^\top x = 0$ guaranteed at each simplex iteration by construction: $\mu_B^\top x_B = 0, \mu_N^\top x_N = 0$
- Primal simplex violates $\mu_N \geq 0$. It iterates until achievement of this optimality condition.

QP optimality conditions

- QP

$$\begin{array}{ll} \min & c^\top x + \frac{1}{2} x^\top Q x \\ \text{s.to} & Ax = b \quad [\pi] \\ & -x \le 0 \quad [\mu] \end{array}$$

- Lagrangian

$$L(x, \pi, \mu) = c^\top x + \frac{1}{2} x^\top Q x + \pi^\top (Ax - b) - \mu^\top x$$

- KKT conditions

$$\begin{array}{l} Ax = b, \quad x \ge 0 \\ c + Qx + A^\top \pi - \mu = 0 \\ \mu^\top x = 0, \quad \mu \ge 0 \end{array}$$

- Using change $\lambda = -\pi$ (no sign constraint) rewrite KKT:

$$\begin{array}{l} Ax = b \\ A^\top \lambda - Qx + \mu = c \\ \mu^\top x = 0 \\ x \ge 0, \mu \ge 0 \end{array}$$

Dual problem

- **Primal problem**

$$\begin{array}{lll} \min\limits_x & f(x) & \\ \text{s.to} & h(x) = 0 & [h_i(x) = 0 \quad i = 1, \ldots, m] \\ & g(x) \le 0 & [g_j(x) \le 0 \quad j = 1, \ldots, p] \\ & x \in X & \end{array}$$

- Lagrangian function:

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

- Dual function $q(\lambda, \mu)$ is

$$q(\lambda, \mu) = \begin{array}{l} \min\limits_x \quad L(x, \lambda, \mu) \\ \qquad\;\; x \in X \end{array}$$

Constraints $h(x) = 0$ and $g(x) \le 0$ dualized, preserving $x \in X$. Depending what is dualized, different formulations obtained.

- **Dual problem**

$$\begin{array}{ll} \max\limits_{\lambda, \mu} & q(\lambda, \mu) \\ & \mu \ge 0 \end{array}$$

NOTE: although inf and sup preferred we will use min and max $q()$.

Dual problem: example

$$\begin{array}{lllll} \min & x_1^2 + x_2^2 & & & \\ \text{s.to} & x_1 + x_2 \geq 4 & \equiv & 4 - x_1 - x_2 \leq 0 \\ & x_1 \geq 0, x_2 \geq 0 & \equiv & -x_1 \leq 0, -x_2 \leq 0 \end{array}$$

Solution is $x_1^* = x_2^* = 2$, $f(x^*) = 8$.
Dual function is:

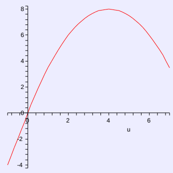
$$q(\mu) = \min_{x_1 \geq 0, x_2 \geq 0} x_1^2 + x_2^2 + \mu(4 - x_1 - x_2) = \min_{x_1 \geq 0, x_2 \geq 0} (x_1^2 - \mu x_1) + (x_2^2 - \mu x_2) + 4\mu$$

Problem is separable, with solution:

$$\begin{cases} x_1 = x_2 = 0 & \text{if } \mu < 0 \quad \text{since } x_i^2 - \mu x_i \geq 0 \\ x_1 = x_2 = \mu/2 & \text{if } \mu \geq 0 \quad \text{since solution of } \min x_i^2 - \mu x_i \end{cases}$$

Then $q(\mu)$ is the concave function:

$$q(\mu) = \begin{cases} 4\mu & \mu < 0 \\ -\mu^2/2 + 4\mu & \mu \geq 0 \end{cases}$$



Solution of dual problem $\max_{\mu \geq 0} q(\mu)$ is $\mu^* = 4$ and $q(\mu^*) = f(x^*) = 8$.

Geometric interpretation of the dual problem (I)

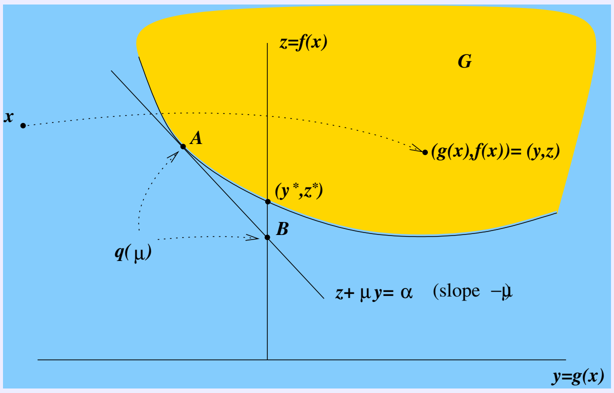
- Consider problem

$$\min_{x \in X} \quad z \triangleq f(x) \quad y \triangleq g(x) \leq 0 \qquad \text{(for instance } f(x) = x^2, g(x) = x)$$

- Let G be image of X for mapping (g, f):

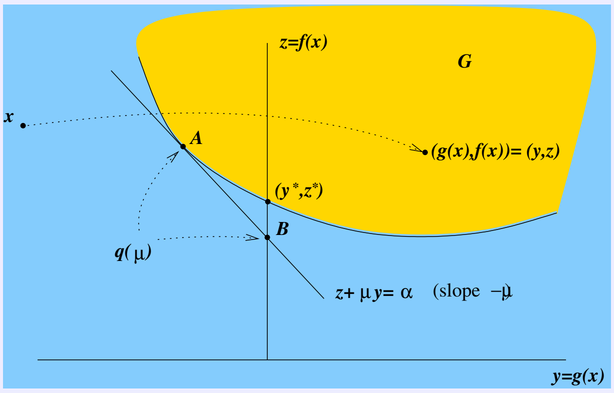
$$G = \{(y, z) : y = g(x), z = g(x) \text{ for some } x \in X\}$$

- The optimal solution is (y^*, z^*):



Geometric interpretation of the dual problem (II)

- The dual is
$$q(\mu) = \min_x f(x) + \mu g(x) = \min_x z + \mu y$$
with $\mu \geq 0$. For some level set α, the line with slope $-\mu \leq 0$ is
$$z + \mu y = \alpha \iff z = \alpha - \mu y.$$
- Fixing μ the solution for $q(\mu)$ is any point on the line AB in the plot:

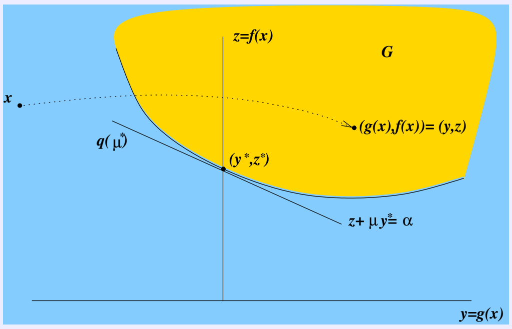


- For $y = 0$, $q(\mu) = z(B)$. If we want $max_\mu q(\mu)$ we have to change slope.

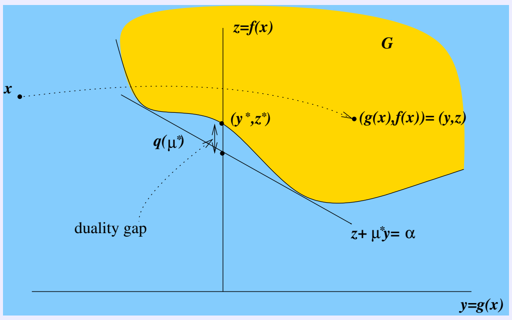
Duality

Geometric interpretation of the dual problem (III)

- The optimal slope is μ^* such that $q(\mu^*) = z^* = f(x^*)$. There is no duality gap:



- If problem is not convex, then the duality gap may be greater than 0:



Weak duality theorem

- Previous definition of dual problem valid for any optimization problem (continuous, discrete, convex, nonconvex...): **Lagrangian Duality**.
- Weak duality is also valid for any problem:
- **Weak duality theorem:**
  Let x be a feasible point of primal problem (i.e. $h(x) = 0$, $g(x) \le 0$, $x \in X$) and (λ, μ) a feasible point of dual problem (i.e., $\mu \ge 0$), then
  $$q(\lambda, \mu) \le f(x)$$
  Proof: From definition of $q(\lambda, \mu)$ and $\mu^\top g(x) \le 0$, we get:
  $$q(\lambda, \mu) = \min_t \{f(t) + \lambda^\top h(t) + \mu^\top g(t) : t \in X\} \le f(x) + \lambda^\top h(x) + \mu^\top g(x) \le f(x)$$
- **Corollary:**
  Dual problem provides a lower bound of primal problem.
- Distance between dual and primal is the **duality gap**.

Strong duality theorem

Primal problem

$$\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.to} & h(x) = 0 \\ & g(x) \le 0 \\ & x \in X \end{array}$$

Dual problem

$$\begin{array}{ll} \max\limits_{\lambda,\mu} & q(\lambda, \mu) \\ & \mu \ge 0 \end{array}$$

- For some problems $q(\lambda^*, \mu^*) = f(x^*)$:
- **Strong duality theorem:**
  If X is a convex set, $f(x)$ and $g(x)$ are convex function, $h(x) = Ax - b$ (affine function), under certain constraints qualifications (Slater constraint qualification) (see [Bazaraa, Sheraly, Shetty (2006), Bertsekas (1999)]), then:
  $$q(\lambda^*, \mu^*) = f(x^*)$$
- Strong duality is satisfied by LP, convex QP, and most convex problems.

Wolfe duality

- Lagrangian duality does not require differentiability. Wolfe duality assumes differentiability.
- If $f(x)$, $h(x)$ and $g(x)$ are convex and differentiable functions, a necessary and sufficient condition of optimality of the dual function

$$q(\lambda, \mu) = \min_x L(x, \lambda, \mu)$$

  is

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

- The dual problem

$$\begin{array}{ll} \max\limits_{\lambda,\mu} & q(\lambda, \mu) \\ & \mu \geq 0 \end{array}$$

  can thus be recast as

$$\begin{array}{ll} \max\limits_{x,\lambda,\mu} & L(x, \lambda, \mu) \\ & \nabla_x L(x, \lambda, \mu) = 0 \\ & \mu \geq 0 \end{array}$$

- This allows a simpler formulation of some problems: LP, QP

The dual of LP using Wolfe duality

- LP

$$\begin{array}{ll} \min & c^\top x \\ \text{s.to} & Ax = b \\ & x \geq 0 \equiv -x \leq 0 \end{array}$$

- Lagrangian function and gradient:

$$L(x, \lambda, \mu) = c^\top x + \lambda^\top (b - Ax) - \mu^\top x$$

$$\nabla_x L(x, \lambda, \mu) = c - A^\top \lambda - \mu$$

- Dual problem:

$$\begin{array}{ll} \max\limits_{x,\lambda,\mu} & c^\top x + \lambda^\top (b - Ax) - \mu^\top x \\ & c - A^\top \lambda - \mu = 0 \\ & \mu \geq 0 \end{array}$$

  Constraint $c - A^\top \lambda - \mu = 0$ means $c^\top x - \lambda^\top Ax - \mu^\top x = 0$, and replacing in objective we get the dual of LP:

$$\begin{array}{ll} \max\limits_{\lambda} & b^\top \lambda \\ & A^\top \lambda \leq c \end{array} \quad \equiv \quad \begin{array}{ll} \max\limits_{\lambda,\mu} & b^\top \lambda \\ & A^\top \lambda + \mu = c, \quad \mu \geq 0 \end{array}$$

The dual of QP using Wolfe duality

- QP

$$\begin{array}{ll} \min & c^\top x + 1/2x^\top Qx \\ \text{s.to} & Ax = b \\ & x \geq 0 \equiv -x \leq 0 \end{array}$$

- Lagrangian function and gradient:

$$L(x,\lambda,\mu) = c^\top x + 1/2x^\top Qx + \lambda^\top(b - Ax) - \mu^\top x$$

$$\nabla_x L(x,\lambda,\mu) = c + Qx - A^\top\lambda - \mu$$

- Dual problem:

$$\begin{array}{ll} \max\limits_{x,\lambda,\mu} & c^\top x + 1/2x^\top Qx + \lambda^\top(b - Ax) - \mu^\top x \\ & c + Qx - A^\top\lambda - \mu = 0 \\ & \mu \geq 0 \end{array}$$

  Constraint $c + Qx - A^\top\lambda - \mu = 0$ means $c^\top x + x^\top Qx - \lambda^\top Ax - \mu^\top x = 0$ and replacing in objective we get the dual of QP:

$$\begin{array}{ll} \max\limits_{\lambda,x} & b^\top\lambda - 1/2x^\top Qx \\ & A^\top\lambda - Qx \leq c \end{array} \quad \equiv \quad \begin{array}{ll} \max\limits_{\lambda,\mu,x} & b^\top\lambda - 1/2x^\top Qx \\ & A^\top\lambda - Qx + \mu = c, \quad \mu \geq 0 \end{array}$$

The dual of a Convex Problem using Wolfe duality

- Convex Problem ($f(x), g(x)$ convex)

$$\begin{array}{ll} \min & f(x) \\ \text{s.to} & h(x) = Ax - b = 0 \\ & g(x) \leq 0 \end{array}$$

- Lagrangian function and gradient:

$$L(x,\lambda,\mu) = f(x) + \lambda^\top(b - Ax) + \mu^\top g(x)$$

$$\nabla_x L(x,\lambda,\mu) = \nabla f(x) - \lambda^\top A + \mu^\top \nabla g(x)$$

- Dual problem

$$\begin{array}{ll} \max\limits_{x,\lambda,\mu} & f(x) + \lambda^\top(b - Ax) + \mu^\top g(x) \\ & \nabla f(x) - \lambda^\top A + \mu^\top \nabla g(x) = 0 \\ & \mu \geq 0 \end{array}$$

- If problem is QP (i.e., $f(x) = c^\top x + 1/2x^\top Qx$ and $g(x) = -x$) dual problem reduces to:

$$\begin{array}{ll} \max\limits_{\lambda,x} & b^\top\lambda - 1/2x^\top Qx \\ & A^\top\lambda - Qx \leq c \end{array} \quad \equiv \quad \begin{array}{ll} \max\limits_{\lambda,\mu,x} & b^\top\lambda - 1/2x^\top Qx \\ & A^\top\lambda - Qx + \mu = c, \quad \mu \geq 0 \end{array}$$

Computational complexity of algorithms

- Used to compute the number of arithmetic operations needed to solve some problem.
- Usually the computational cost is a function of the size of the input.
- For LP is not easy to determine the size of the input: n number of variables, m number of constraints? Sometimes L is used: the number of bits needed to code the problem.
- Total costs depends on the number of iterations and cost per iteration.
- In simplex, the worse case can be an exponential number of iterations.

Local convergence and rates of convergence

- Optimization methods generate sequences of points $x^0, x^1, \ldots, x^k, \ldots$
- Local convergence analyzes how fast we approach the optimal solution x^* when we are close to that point.
  - Linear convergence (Example: $x^k = 2^{-k}$)

    $$\exists r \in (0,1) : \frac{||x^{k+1} - x^*||}{||x^k - x^*||} \leq r \quad \forall k \text{ large enough}$$

  - Superlinear convergence (Example: $x^k = k^{-k}$)

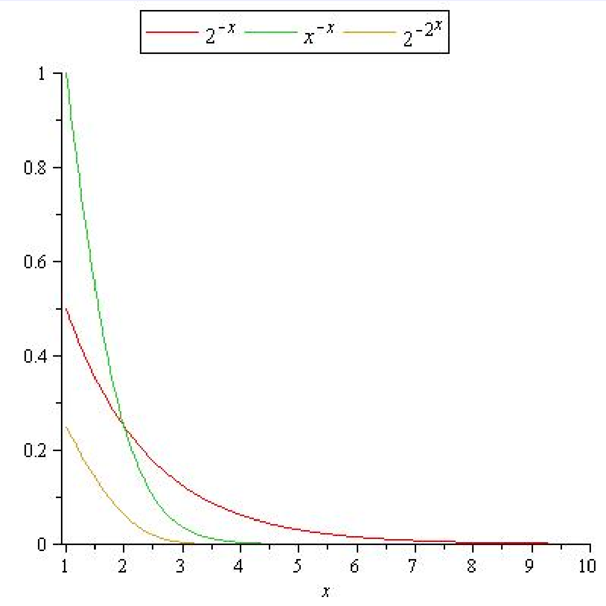
    $$\lim_{k \to \infty} \frac{||x^{k+1} - x^*||}{||x^k - x^*||} = 0$$

  - Quadratic convergence (Example: $x^k = 2^{-2^k}$)

    $$\exists M \in \mathbb{R} : \frac{||x^{k+1} - x^*||}{||x^k - x^*||^2} \leq M \quad \forall k \text{ suficientment gran}$$

  - Quadratic $\Rightarrow$ superlinear $\Rightarrow$ linear

Example rates of convergence



k	2^{-k}	k^{-k}	2^{-2^k}
1	0.500000	1.000000	0.250000
2	0.250000	0.250000	0.062500
3	0.125000	0.037037	0.003906
4	0.062500	0.003906	0.000015
5	0.031250	0.000320	0.000000
6	0.015625	0.000021	
7	0.007812	0.000001	
8	0.003906	0.000000	
9	0.001953		
10	0.000976		
11	0.000488		
12	0.000244		
13	0.000122		
14	0.000061		

Order notation

- Used for computational complexity and rates of convergence.
- Given to nonnegative sequences of scalars $\{a_k\}$, $\{b_k\}$:
  - $b_k = O(a_k)$ if $\exists C > 0 : \quad |b_k| \leq C|a_k| \quad \forall k$ large enough
  - $b_k = o(a_k)$ if $\lim_{k\to\infty} \dfrac{b_k}{a_k} = 0$

Newton's method

- In unconstrained optimization
$$\min_{x \in \mathbb{R}^n} f(x)$$
used for solving $\nabla f(x) = 0$.
- In constrained optimization (and IPMs) used for solving nonlinear system $F(x) = 0$ associated to KKT conditions.
- Using a linear approximation $M_k(x)$ of $F(x)$ based on a first-order Taylor series at point x^k:
$$F(x^k + \Delta) = F(x^k) + \nabla F(x^k)\Delta + o(||\Delta||^2)$$
$$M_k(x^k + \Delta) = F(x^k) + \nabla F(x^k)\Delta \approx F(x^k + \Delta)$$
$$M_k(x^k + \Delta) = 0 \quad \Rightarrow \quad \Delta = (\nabla F(x^k))^{-1}(-F(x^k))$$
Compute $x^{k+1} = x^k + \Delta$, $k := k + 1$ and iterate until solution.

Newton's method: basic properties

- Local quadratic convergence.
- No global convergence. Main drawbacks:
  - $\nabla F(x)$ may be singular.
    - **Solutions:** use modifications of Newton's method : quasi-Netwon updates, modification of $\nabla F(x)$.
  - It can iterate forever. **Example**: Newton's method to $F(x) = -x^5 + x^3 + 4x$, starting at $x^0 = 1$. We get $\Delta^0 = -2$ and $x^1 = -1$. But $\Delta^1 = 2$ and then $x^2 = 1 = x^0$.
    - **Solutions:** use line-search methods ($x^{k+1} = x^k + \alpha^k \Delta^k$), or a trust-region model ($\min_{\Delta} M_k(x^k + \Delta) : ||\Delta||_2 \leq \epsilon$).

General references

- M.S. Bazaraa, H.D. Sherali, C.M. Shetty, *Nonlinear Programming. Theory and Algorithms, 3rd Ed.*, Wiley, 2006.
- D.P. Bertsekas, *Nonlinear Programming, 2nd Ed.*, Athena Scientific, 1999.
- D.G. Luenberger, *Linear and Nonlinear Programming, 2nd Ed.*, Addison Wesley, 1984.
- J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer, 1999.