School of Mathematics

Introduction to Nonlinear Stochastic Programming

Jacek Gondzio
Email: J.Gondzio@ed.ac.uk
URL: http://www.maths.ed.ac.uk/~gondzio

Outline

- **Convexity**
  - convex sets, convex functions
  - local optimum, global optimum
- **Duality**
  - Lagrange duality
  - Wolfe duality
  - primal-dual pairs of LPs and QPs
- **Nonlinear Stochastic Program with Recourse**
- **Interior Point Methods for Optimization**
  $\Rightarrow$ unified view of Linear, Quadratic and Nonlinear programming

Consider the general optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \end{array}$$

where $x \in \mathcal{R}^n$, and $f : \mathcal{R}^n \mapsto \mathcal{R}$ and $g : \mathcal{R}^n \mapsto \mathcal{R}^m$
are convex, twice differentiable.

Basic Assumptions:

f and g are convex
$\Rightarrow$ If there exists a **local** minimum then it is a **global** one.

f and g are twice differentiable
$\Rightarrow$ We can use the **second order Taylor approximations** of them.

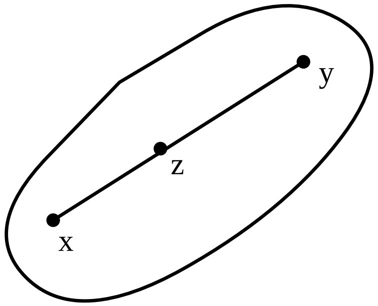
Convexity

Reading:

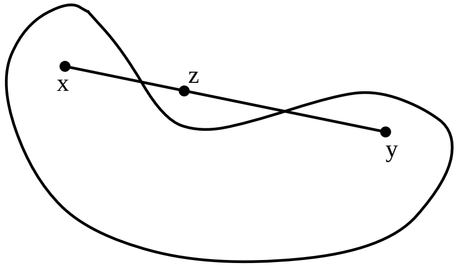
Bertsekas, D., *Nonlinear Programming*,
Athena Scientific, Massachusetts, 1995. ISBN 1-886529-14-0.

Convexity is a key property in optimization.

Def. A set $C \subset \mathcal{R}^n$ is *convex* if $\lambda x + (1-\lambda) y \in C, \ \forall x, y \in C, \ \forall \lambda \in [0,1]$.

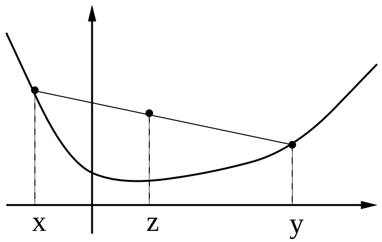


Convex set

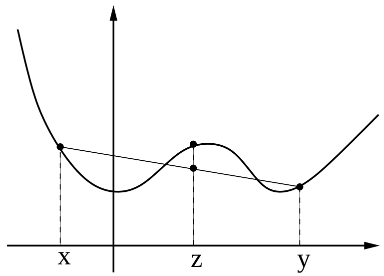


Nonconvex set

Def. Let C be a convex subset of $\mathcal{R}^n$. A function $f : C \mapsto \mathcal{R}$ is *convex* if
$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y), \ \forall x, y \in C, \ \forall \lambda \in [0,1]$.



Convex function



Nonconvex function

Convexity and Optimization

Consider a problem

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \end{array}$$

where X is a set of feasible solutions
and $f : X \to \mathcal{R}$ is an objective function.

Def. A vector $\hat{x}$ is a **local** minimum of f if

$$\exists \epsilon > 0 \text{ such that } f(\hat{x}) \leq f(x), \ \forall x \mid \|x - \hat{x}\| < \epsilon.$$

Def. A vector $\hat{x}$ is a **global** minimum of f if

$$f(\hat{x}) \leq f(x), \ \forall x \in X.$$

Lemma. If X is a convex set and $f : X \mapsto \mathcal{R}$ is a convex function, then a **local** minimum is a **global** minimum.

Proof.
Suppose that x is a local minimum, but not a global one. Then $\exists y \neq x$ such that $f(y) < f(x)$.
From convexity of f, for all $\lambda \in [0,1]$, we have

$$\begin{aligned} f((1-\lambda)x+\lambda y) &\leq (1-\lambda)f(x)+\lambda f(y) \\ &< (1-\lambda)f(x)+\lambda f(x) = f(x). \end{aligned}$$

In particular, for a sufficiently small λ, the point $z = (1-\lambda)x + \lambda y$ lies in the ϵ-neighbourhood of x and $f(z) < f(x)$. This contradicts the assumption that x is a local minimum.

Useful properties

1. For any collection $\{C_i \mid i \in I\}$ of convex sets, the intersection $\bigcap_{i \in I} C_i$ is convex.

2. If C is a convex set and $f : C \mapsto \mathcal{R}$ is a convex function, the level sets $\{x \in C \mid f(x) \leq \alpha\}$ and $\{x \in C \mid f(x) < \alpha\}$ are convex for all scalars α.

3. Let $C \in \mathcal{R}^n$ be a convex set and $f : C \mapsto \mathcal{R}$ be differentiable over C.
(a) The function f is convex if and only if

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

(b) If the inequality is strict for $x \neq y$, then f is strictly convex.

4. Let $C \in \mathcal{R}^n$ be a convex set and $f : C \mapsto \mathcal{R}$ be twice continuously differentiable over C.
(a) If $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$, then f is convex.
(b) If $\nabla^2 f(x)$ is positive definite for all $x \in C$, then f is strictly convex.
(c) If f is convex, then $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$.

Lemma. If $f : \mathcal{R}^n \mapsto \mathcal{R}$ and $g : \mathcal{R}^n \mapsto \mathcal{R}^m$ are convex, then the following general optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \le 0 \end{array}$$

is convex.
Proof. Since the objective function f is convex, we only need to prove that the feasible set of the above problem

$$X = \{x \in \mathcal{R}^n : g(x) \le 0\}$$

is convex. Define for $i = 1, 2, ..., m$

$$X_i = \{x \in \mathcal{R}^n : g_i(x) \le 0\}.$$

From Property 2, X_i is convex for all i.
We observe that

$$X = \{x \in \mathcal{R}^n : g_i(x) \le 0, \ \forall i = 1..m\} = \bigcap_i X_i.$$

i.e., X is an intersection of convex sets and from Property 1, X is a convex set.

Duality

Reading:

Bertsekas, D., *Nonlinear Programming*,
Athena Scientific, Massachusetts, 1995. ISBN 1-886529-14-0.

Consider a general optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \le 0, \\ & x \in X \subseteq \mathcal{R}^n, \end{array} \qquad (1)$$

where $f : \mathcal{R}^n \mapsto \mathcal{R}$ and $g : \mathcal{R}^n \mapsto \mathcal{R}^m$.
The set X is arbitrary; it may include, for example, an integrality constraint.

Let $\hat{x}$ be an optimal solution of (1) and define

$$\hat{f} = f(\hat{x}).$$

Introduce the Lagrange multiplier $y_i \ge 0$ for every inequality constraint $g_i(x) \le 0$.
Define $y = (y_1, \ldots, y_m)^T$ and the **Lagrangian**

$$L(x, y) = f(x) + y^T g(x),$$

y are also called *dual* variables.

Consider the problem

$$L_D(y) \;=\; \min_x L(x,y) \quad s.t. \quad x \in X \subseteq \mathcal{R}^n.$$

Its optimal solution x depends on y and so does the optimal objective $L_D(y)$.

Lemma. For any $y \geq 0$, $L_D(y)$ is a lower bound on $\hat{f}$ (the optimal solution of (1)), i.e.,

$$\hat{f} \geq L_D(y) \qquad \forall y \geq 0.$$

Proof.

$$\begin{aligned}
\hat{f} &= \min \; \{f(x) \mid g(x) \leq 0, x \in X\} \\
&\geq \min \; \left\{f(x) + y^T g(x) \mid g(x) \leq 0, y \geq 0, x \in X\right\} \\
&\geq \min \; \left\{f(x) + y^T g(x) \mid y \geq 0, x \in X\right\} \\
&= \; L_D(y).
\end{aligned}$$

Corollary.

$$\hat{f} \geq \max_{y \geq 0} L_D(y), \quad \text{i.e.,} \quad \hat{f} \geq \max_{y \geq 0} \min_{x \in X} L(x,y).$$

Lagrangian Duality

If $\exists i \; g_i(x) > 0$, then

$$\max_{y \geq 0} L(x,y) = +\infty$$

(we let the corresponding y_i grow to $+\infty$).
If $\forall i \; g_i(x) \leq 0$, then

$$\max_{y \geq 0} L(x,y) = f(x),$$

because $\forall i \; y_i g_i(x) \leq 0$ and the maximum is attained when

$$y_i g_i(x) = 0, \qquad \forall i = 1, 2, ..., m.$$

Hence the problem (1) is equivalent to the following **MinMax** problem

$$\min_{x \in X} \max_{y \geq 0} L(x,y),$$

which could also be written as follows:

$$\hat{f} = \min_{x \in X} \max_{y \geq 0} L(x,y).$$

Consider the following problem

$$\min \ \{f(x) \mid g(x) \leq 0, x \in X\},$$

where f, g and X are arbitrary.
With this problem we associate the **Lagrangian**

$$L(x,y) = f(x) + y^T g(x),$$

y are *dual* variables (Lagrange multipliers).
The **weak duality** always holds:

$$\min_{x \in X} \max_{y \geq 0} L(x,y) \ \geq \ \max_{y \geq 0} \min_{x \in X} L(x,y).$$

We have not made **any** assumption about functions f and g and set X.

If f and g are convex, X is convex and certain regularity conditions are satisfied, then

$$\min_{x \in X} \max_{y \geq 0} L(x,y) \ = \ \max_{y \geq 0} \min_{x \in X} L(x,y).$$

This is called the **strong duality**.

Duality and Convexity

The weak duality holds regardless of the form of functions f, g and set X:

$$\min_{x \in X} \max_{y \geq 0} L(x,y) \;\geq\; \max_{y \geq 0} \min_{x \in X} L(x,y).$$

What do we need for the *inequality* in the weak duality to become an *equation*?
If

- $X \subseteq \mathcal{R}^n$ is convex;
- f and g are convex;
- optimal solution is finite;
- some mysterious *regularity conditions* hold,

then strong duality holds. That is

$$\min_{x \in X} \max_{y \geq 0} L(x,y) \;=\; \max_{y \geq 0} \min_{x \in X} L(x,y).$$

An example of regularity conditions:
$\exists x \in int(X)$ such that $g(x) < 0$.

Lagrange duality does not need differentiability.
Suppose f and g are convex and differentiable. Suppose X is convex.
The **dual function**

$$L_D(y) = \min_{x \in X} L(x, y).$$

requires minimization with respect to x.
Instead of **minimization** with respect to x,
we ask for a **stationarity** with respect to x:

$$\nabla_x L(x, y) = 0.$$

Lagrange dual problem:

$$\max_{y \geq 0} L_D(y) \qquad \left(\text{i.e., } \max_{y \geq 0} \min_{x \in X} L(x, y) \right).$$

Wolfe dual problem:

$$\begin{array}{rr} \max & L(x, y) \\ \text{s.t.} & \nabla_x L(x, y) = 0 \\ & y \geq 0. \end{array}$$

Dual Linear Program

Consider a linear program

$$\begin{array}{lrcl} \min & c^T x & & \\ \text{s.t.} & Ax & = & b, \\ & x \geq 0, & & \end{array}$$

where $c, x \in \mathcal{R}^n, b \in \mathcal{R}^m$, $A \in \mathcal{R}^{m \times n}$.
We associate Lagrange multipliers $y \in \mathcal{R}^m$ and $s \in \mathcal{R}^n$ $(s \geq 0)$ with the constraints $Ax = b$ and $x \geq 0$, and write the **Lagrangian**

$$L(x, y, s) \ = \ c^T x - y^T (Ax - b) - s^T x.$$

To determine the *Lagrangian dual*

$$L_D(y, s) = \min_{x \in X} L(x, y, s)$$

we need stationarity with respect to x:

$$\nabla_x L(x, y, s) \ = \ c - A^T y - s \ = \ 0.$$

Hence

$$\begin{aligned} L_D(y,s) &= c^T x - y^T(Ax-b) - s^T x \\ &= b^T y + x^T(c - A^T y - s) = b^T y. \end{aligned}$$

and the **dual LP** has a form:

$$\begin{array}{rrcl} \max & b^T y & & \\ \text{s.t.} & A^T y \; + & s & = c, \\ & y \;\text{free}, & s \geq 0, & \end{array}$$

where $y \in \mathcal{R}^m$ and $s \in \mathcal{R}^n$.

Primal Problem	**Dual Problem**
$\min \; c^T x$	$\max \; b^T y$
s.t. $Ax = b,$	s.t. $A^T y + s = c,$
$x \geq 0;$	$s \geq 0.$

Dual Quadratic Program

Consider a quadratic program

$$\begin{array}{ll} \min & c^T x + \frac{1}{2} x^T Q\, x \\ \text{s.t.} & Ax = b, \\ & x \geq 0, \end{array}$$

where $c, x \in \mathcal{R}^n, b \in \mathcal{R}^m$, $A \in \mathcal{R}^{m \times n}$, $Q \in \mathcal{R}^{n \times n}$.
We associate Lagrange multipliers $y \in \mathcal{R}^m$ and $s \in \mathcal{R}^n$ $(s \geq 0)$ with the constraints $Ax = b$ and $x \geq 0$, and write the **Lagrangian**

$$L(x, y, s) \; = \; c^T x + \frac{1}{2} x^T Q\, x - y^T (Ax - b) - s^T x.$$

To determine the *Lagrangian dual*

$$L_D(y, s) = \min_{x \in X} L(x, y, s)$$

we need stationarity with respect to x:

$$\nabla_x L(x, y, s) \; = \; c + Qx - A^T y - s \; = \; 0.$$

Hence

$$\begin{aligned} L_D(y,s) &= c^T x + \tfrac{1}{2} x^T Q\, x - y^T(Ax-b) - s^T x \\ &= b^T y + x^T(c + Qx - A^T y - s) - \tfrac{1}{2} x^T Q\, x \\ &= b^T y \;-\; \tfrac{1}{2} x^T Q\, x, \end{aligned}$$

and the **dual QP** has the form:

$$\begin{array}{rc} \max & b^T y - \tfrac{1}{2} x^T Q\, x \\ \text{s.t.} & A^T y + s - Qx = c, \\ & x, s \geq 0, \end{array}$$

where $y \in \mathcal{R}^m$ and $x, s \in \mathcal{R}^n$.

Primal Problem	**Dual Problem**
$\min\ c^T x + \frac{1}{2} x^T Q x$	$\max\ b^T y - \frac{1}{2} x^T Q x$
$\text{s.t.}\ Ax = b,$	$\text{s.t.}\ A^T y + s - Qx = c,$
$x \geq 0;$	$s \geq 0.$

General Stochastic Program with Recourse

Reading:
Kall P. and S.W. Wallace., *Stochastic Programming*,
John Wiley & Sons, Chichester 1994, UK.

General Stochastic Program with Recourse

Consider a **deterministic** problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0, \;\; x \in X, \end{array}$$

where $f : \mathcal{R}^n \mapsto \mathcal{R}$ is an objective function, and
$g_i : \mathcal{R}^n \rightarrow \mathcal{R}, \;\; i = 1...m$ are constraints.

Consider its **stochastic** analogue

$$\begin{array}{ll} \text{“minimize”} & f(x, \xi) \\ \text{subject to} & g(x, \xi) \leq 0, \;\; x \in X, \end{array}$$

where ξ is a random vector varying over a set $\Xi \subset \mathcal{R}^k$,
$f : \mathcal{R}^n \times \Xi \mapsto \mathcal{R}$ is an objective function,
and $g_i : \mathcal{R}^n \times \Xi \rightarrow \mathcal{R}, \;\; i = 1...m$ are constraints.

Define

$$g_i^+(x,\xi) = \begin{cases} 0 & \text{if} \quad g_i(x,\xi) \le 0, \\ g_i(x,\xi) & \text{otherwise.} \end{cases}$$

and a nonlinear **recourse function**

$$Q(x,\xi) = \min\{q(y) | u_i(y) \ge g_i^+(x,\xi), i = 1..m, y \in Y \subset \mathcal{R}^{\bar{n}}\},$$

where $q : \mathcal{R}^{\bar{n}} \mapsto \mathcal{R}$ and $u_i : \mathcal{R}^{\bar{n}} \mapsto \mathcal{R}$ are given.

Replace original stochastic program by **stochastic program with recourse**

$$\min_{x \in X} \; E_{\tilde{\xi}}\{f(x,\tilde{\xi}) + Q(x,\tilde{\xi})\}.$$

Convexity of Recourse Problems

Lemma. If the functions $q(.)$ and $g_i(.,\xi), i = 1..m$ are convex and the functions $u_i(.), i = 1..m$ are concave, then the nonlinear recourse function

$$Q(x,\xi) = \min\{q(y) | u_i(y) \ge g_i^+(x,\xi), i = 1..m, y \in Y \subset \mathcal{R}^{\bar{n}}\} \tag{2}$$

is convex with respect to its first argument.

Proof.
Let y_1 and y_2 be the optimal solutions of the recourse problems for x_1 and x_2, respectively.
By the convexity of $g_i(.,\xi)$ and the concavity of $u_i(.)$ we have for any $\lambda \in [0,1]$:

$$\begin{aligned} g_i(\lambda x_1 + (1-\lambda)x_2, \xi) &\leq \lambda g_i(x_1,\xi) + (1-\lambda) g_i(x_2,\xi) \\ &\leq \lambda u_i(y_1) + (1-\lambda) u_i(y_2) \\ &\leq u_i(\lambda y_1 + (1-\lambda) y_2). \end{aligned}$$

Hence $\bar{y} = \lambda y_1 + (1-\lambda)y_2$ is feasible in (2) for $\bar{x} = \lambda x_1 + (1-\lambda)x_2$ and by convexity of $q(.)$

$$\begin{aligned} Q(\bar{x},\xi) &\leq q(\bar{y}) \\ &\leq \lambda q(y_1) + (1-\lambda) q(y_2) \\ &\leq \lambda Q(x_1,\xi) + (1-\lambda) Q(x_2,\xi), \end{aligned}$$

which completes the proof.

Lemma. If $f(.,\xi)$ and $Q(.,\xi)$ are convex in $x \;\; \forall \xi \in \Xi$, and if X is a convex set, then

$$\min_{x \in X} \; E_{\tilde{\xi}}\{f(x,\tilde{\xi}) + Q(x,\tilde{\xi})\}.$$

is a convex program.

Proof.

Take $x, y \in X, \lambda \in [0,1]$ and $z = \lambda x + (1-\lambda)y$. From convexity of $f(.,\xi)$ and $Q(.,\xi)$ with respect to their first argument, $\forall \xi \in \Xi$, we have:

$$f(z,\xi) \leq \lambda f(x,\xi) + (1-\lambda) f(y,\xi)$$

and

$$Q(z,\xi) \leq \lambda Q(x,\xi) + (1-\lambda) Q(y,\xi),$$

respectively.

Adding these inequalities, we obtain $\forall \xi \in \Xi$

$$f(z,\xi) + Q(z,\xi) \leq \lambda [f(x,\xi) + Q(x,\xi)] + (1-\lambda)[f(y,\xi) + Q(y,\xi)],$$

implying

$$E_{\tilde{\xi}}\{f(z,\tilde{\xi}) + Q(z,\tilde{\xi})\} \leq \lambda E_{\tilde{\xi}}\{f(x,\tilde{\xi}) + Q(x,\tilde{\xi})\} + (1-\lambda) E_{\tilde{\xi}}\{f(y,\tilde{\xi}) + Q(y,\tilde{\xi})\}.$$

Hence the objective in stochastic program with recourse is convex.

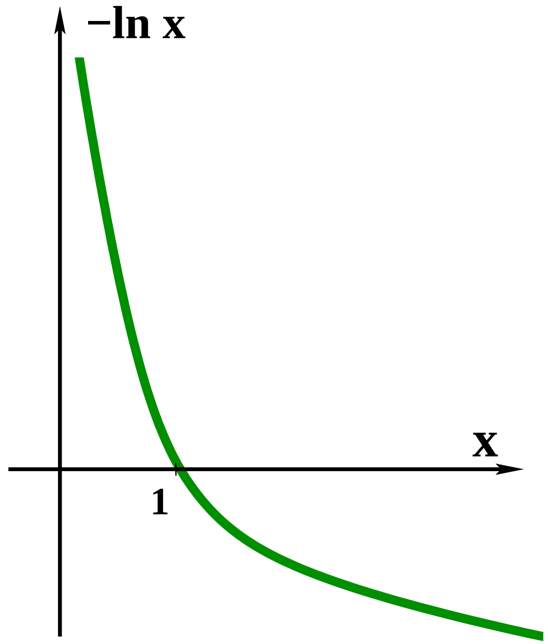
Interior Point Methods
→ from LP via QP to NLP

Reading:

Wright S., *Primal-Dual Interior-Point Methods*, SIAM, 1997.
Nocedal J. & S. Wright, *Numerical Optimization*, Springer-Verlag, 1999.
Conn A., N. Gould & Ph. Toint, *Trust-Region Methods*, SIAM, 2000.

Logarithmic barrier $\quad -\ln x_j$

"replaces" the inequality $\quad x_j \geq 0$.



Observe that

$$\min \ e^{-\sum_{j=1}^{n} \ln x_j} \quad \Longleftrightarrow \quad \max \prod_{j=1}^{n} x_j$$

The minimization of $-\sum_{j=1}^{n} \ln x_j$ is equivalent to the maximization of the product of distances from all hyperplanes defining the positive orthant: it prevents all x_j from approaching zero.

Use Logarithmic Barrier

Primal Problem

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \\ & x \geq 0; \end{array}$$

Dual Problem

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + s = c, \\ & s \geq 0. \end{array}$$

Primal Barrier Problem

$$\begin{array}{ll} \min & c^T x - \sum_{j=1}^{n} \ln x_j \\ \text{s.t.} & Ax = b, \end{array}$$

Dual Barrier Problem

$$\begin{array}{ll} \max & b^T y + \sum_{j=1}^{n} \ln s_j \\ \text{s.t.} & A^T y + s = c, \end{array}$$

Primal Barrier Problem:

$$\begin{array}{ll} \min & c^T x - \mu \sum_{j=1}^{n} \ln x_j \\ \text{s.t.} & Ax = b. \end{array}$$

Lagrangian: $$L(x, y, \mu) = c^T x - y^T (Ax - b) - \mu \sum_{j=1}^{n} \ln x_j,$$

Stationarity: $$\nabla_x L(x, y, \mu) = c - A^T y - \mu X^{-1} e = 0$$

Denote: $$s = \mu X^{-1} e, \quad \text{i.e.} \quad XSe = \mu e.$$

The **First Order Optimality Conditions** are:

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XSe &= \mu e \\ (x, s) &> 0. \end{aligned}$$

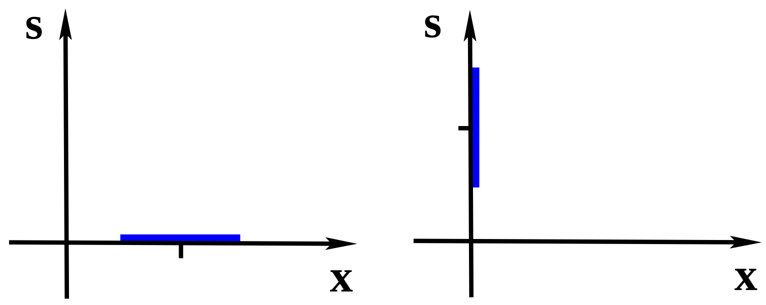
First Order Optimality Conditions

Simplex Method:

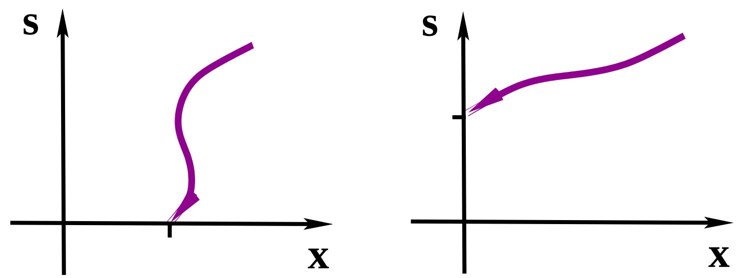
$$\begin{aligned} Ax &= b \\ A^T y + s &= c \\ XSe &= 0 \\ x,\, s &\geq 0. \end{aligned}$$

Interior Point Method:

$$\begin{aligned} Ax &= b \\ A^T y + s &= c \\ XSe &= \mu e \\ x,\, s &\geq 0. \end{aligned}$$



Basic: x > 0, s = 0 Nonbasic: x = 0, s > 0

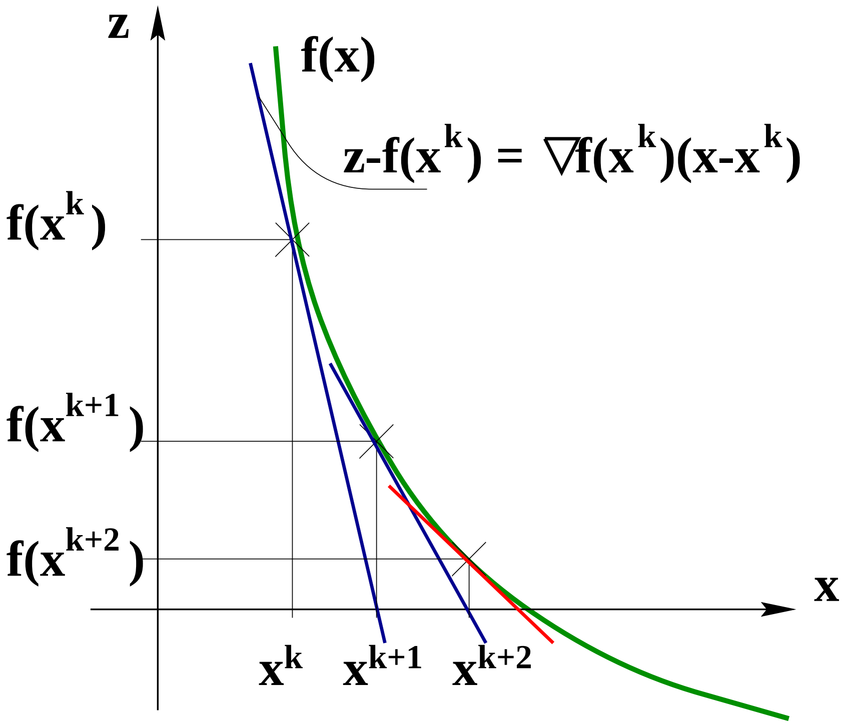


"Basic": x > 0, s = 0 "Nonbasic": x = 0, s > 0

Theory: IPMs converge in $\mathcal{O}(\sqrt{n})$ or $\mathcal{O}(n)$ iterations
Practice: IPMs converge in $\mathcal{O}(\log n)$ iterations
... but one iteration may be expensive!

Newton Method

We use **Newton Method** to find a stationary point of the barrier problem.



Find a root of a nonlinear equation

$$f(x) = 0.$$

A tangent line

$$z - f(x^k) = \nabla f(x^k) \cdot (x - x^k)$$

is a local approximation of the graph of the function $f(x)$.
Substite $z = 0$ to get a new point

$$x^{k+1} = x^k - (\nabla f(x^k))^{-1} f(x^k).$$

Newton Method

The first order optimality conditions for the barrier problem form a large system of nonlinear equations

$$F(x,y,s)=0,$$

where $F:\mathcal{R}^{2n+m}\mapsto\mathcal{R}^{2n+m}$ is an application defined as follows:

$$F(x,y,s)=\begin{bmatrix} Ax & - & b \\ A^Ty+s & - & c \\ XSe & - & \mu e \end{bmatrix}.$$

Actually, the first two terms of it are *linear*; only the last one, corresponding to the complementarity condition, is *nonlinear*.

For a given point (x,y,s) we find the Newton direction $(\Delta x,\Delta y,\Delta s)$ by solving the system of linear equations:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b-Ax \\ c-A^Ty-s \\ \mu e-XSe \end{bmatrix}.$$

IPM for LP

$$\begin{array}{lc} \min & c^T x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array} \quad \rightarrow \quad \begin{array}{lc} \min & c^T x - \mu \sum_{j=1}^{n} \ln x_j \\ \text{s.t.} & Ax = b, \end{array}$$

The first order conditions (for the barrier problem)

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XSe &= \mu e. \end{aligned}$$

Newton direction

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} \xi_p \\ \xi_d \\ \xi_\mu \end{bmatrix} = \begin{bmatrix} b - Ax \\ c - A^T y - s \\ \mu e - XSe \end{bmatrix}.$$

Augmented system

$$\begin{bmatrix} -\Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \xi_d - X^{-1} \xi_\mu \\ \xi_p \end{bmatrix}.$$

IPM for QP

$$\begin{array}{ll} \min & c^Tx + \frac{1}{2}x^T\mathbf{Q}\,x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array} \quad \longrightarrow \quad \begin{array}{ll} \min & c^Tx + \frac{1}{2}x^T\mathbf{Q}\,x - \mu\sum_{j=1}^n \ln x_j \\ \text{s.t.} & Ax = b, \end{array}$$

The first order conditions (for the barrier problem)

$$\begin{array}{rcl} Ax & = & b, \\ A^Ty + s - \mathbf{Q}x & = & c, \\ XSe & = & \mu e. \end{array}$$

Newton direction

$$\begin{bmatrix} A & 0 & 0 \\ -\mathbf{Q} & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} \xi_p \\ \xi_d \\ \xi_\mu \end{bmatrix} = \begin{bmatrix} b - Ax \\ c - A^Ty - s + \mathbf{Q}x \\ \mu e - XSe \end{bmatrix}.$$

Augmented system

$$\begin{bmatrix} -\mathbf{Q} - \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \xi_d - X^{-1}\xi_\mu \\ \xi_p \end{bmatrix}.$$

IPM for NLP

$$\begin{array}{llcll} \min & f(x) & \rightarrow & \min & f(x) - \mu \sum_{i=1}^{m} \ln z_i \\ \text{s.t.} & g(x) + z = 0 & & \text{s.t.} & g(x) + z = 0, \\ & z \geq 0. & & & \end{array}$$

Lagrangian: $L(x, y, z, \mu) = f(x) + y^T(g(x) + z) - \mu \sum_{i=1}^{m} \ln z_i.$

The first order conditions (for the barrier problem)

$$\begin{aligned} \nabla f(x) + \nabla g(x)^T y &= 0, \\ g(x) + z &= 0, \\ YZe &= \mu e. \end{aligned}$$

Newton direction

$$\begin{bmatrix} Q(x,y) & A(x)^T & 0 \\ A(x) & 0 & I \\ 0 & Z & Y \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} -\nabla f(x) - A(x)^T y \\ -g(x) - z \\ \mu e - YZe \end{bmatrix}.$$

Augmented system

$$\begin{bmatrix} Q(x,y) & A(x)^T \\ A(x) & -ZY^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) - A(x)^T y \\ -g(x) - \mu Y^{-1} e \end{bmatrix} \quad \text{where} \quad \begin{aligned} A(x) &= \nabla g \\ Q(x,y) &= \nabla^2_{xx} L \end{aligned}$$

Generic Interior-Point NLP Algorithm

Initialize

$k = 0$

(x^0, y^0, z^0) such that $y^0 > 0$ and $z^0 > 0$, $\quad \mu_0 = \frac{1}{m} \cdot (y^0)^T z^0$

Repeat until optimality

$k = k + 1$

$\mu_k = \sigma \mu_{k-1}$, where $\sigma \in (0, 1)$

Compute $A(x)$ and $Q(x, y)$

$\Delta =$ Newton direction towards μ-center

Ratio test:

$\alpha_1 := \max \{\alpha > 0 : y + \alpha \Delta y \geq 0\}, \quad \alpha_2 := \max \{\alpha > 0 : z + \alpha \Delta z \geq 0\}.$

Choose the step: $\alpha \leq \min \{\alpha_1, \alpha_2\}$ (use **trust region** or **line search**)

Make step:

$x^{k+1} = x^k + \alpha \Delta x,$

$y^{k+1} = y^k + \alpha \Delta y,$

$z^{k+1} = z^k + \alpha \Delta z.$