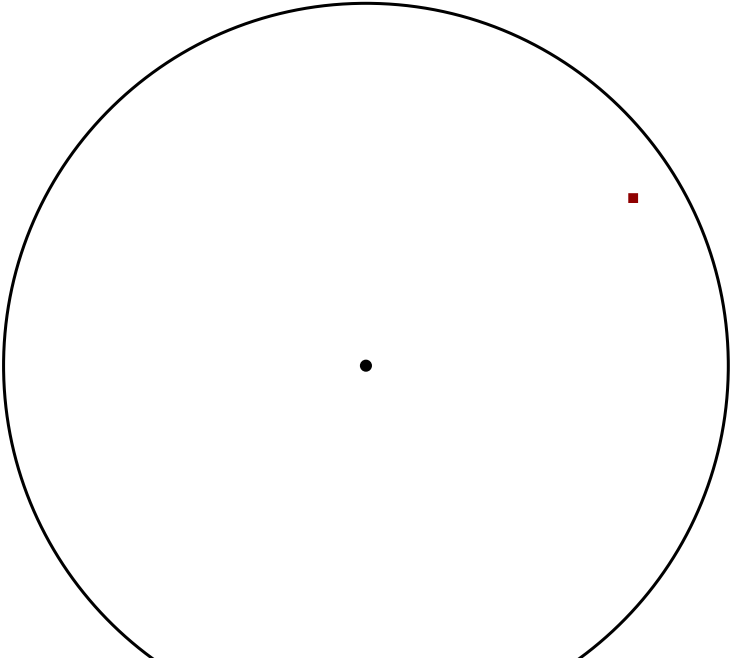
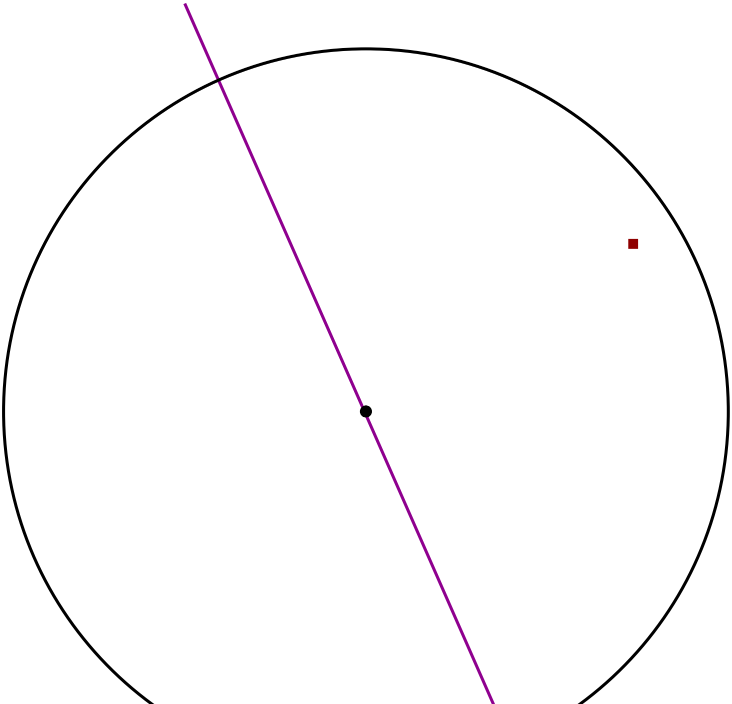
Interior Point Methods in Mathematical Programming

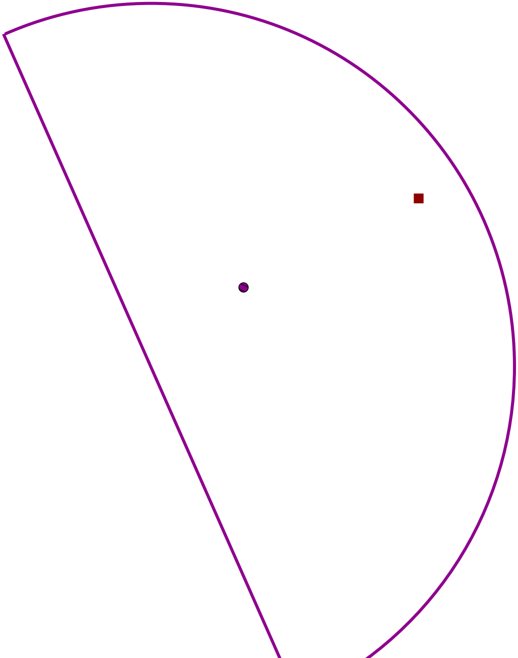
Clóvis C. Gonzaga

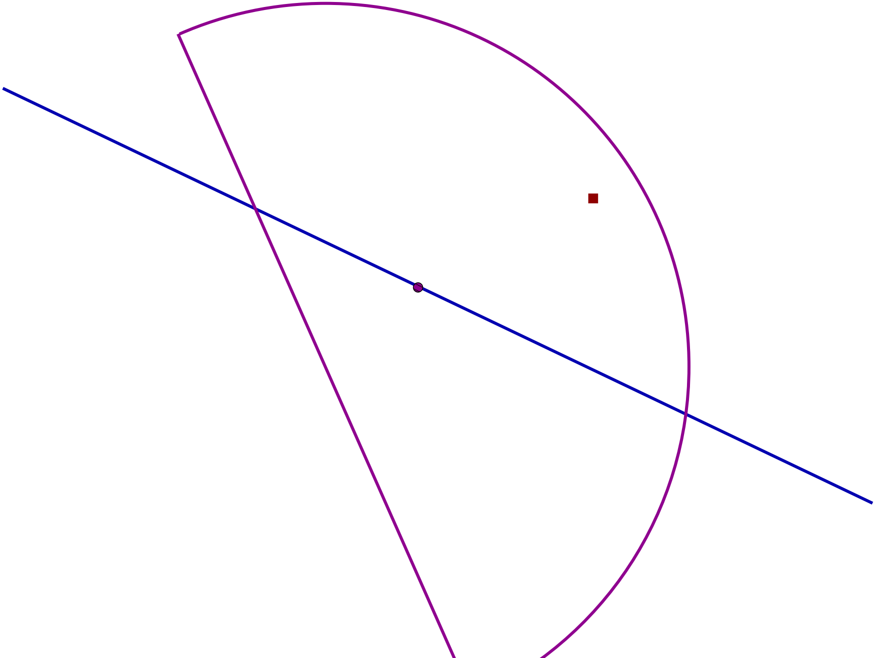
Federal University of Santa Catarina, Brazil

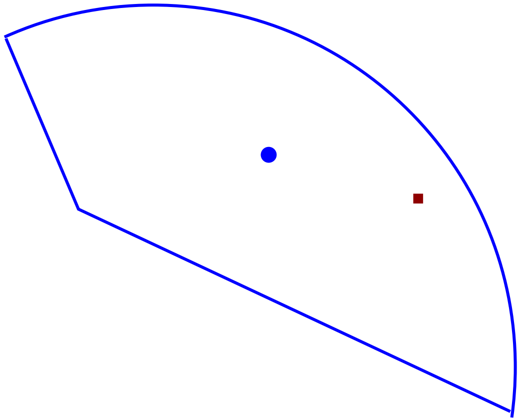
Journées en l'honneur de Pierre Huard
Paris, novembre 2008

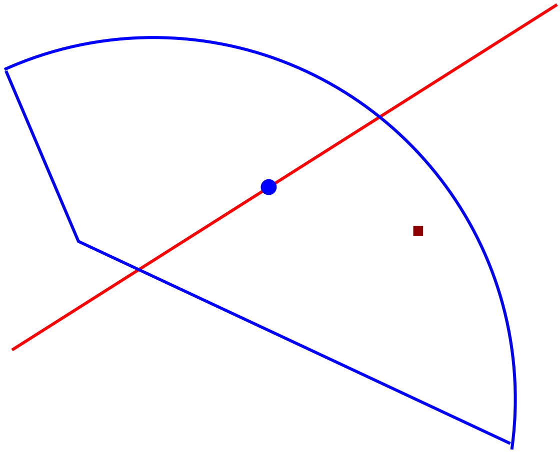


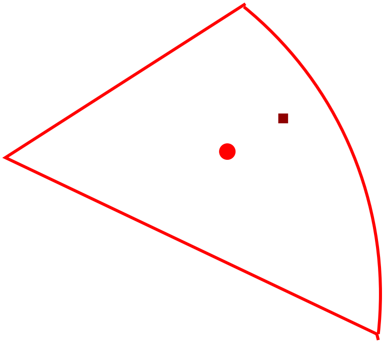


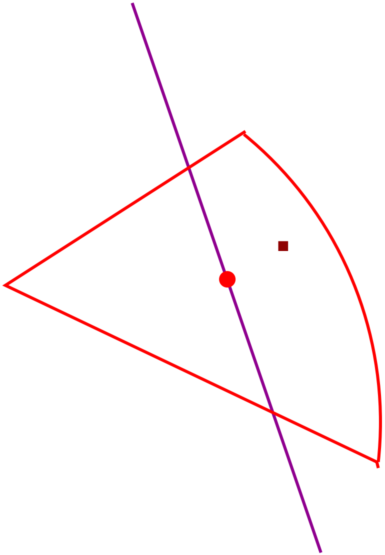


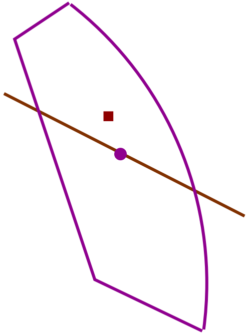


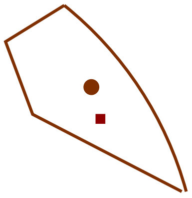


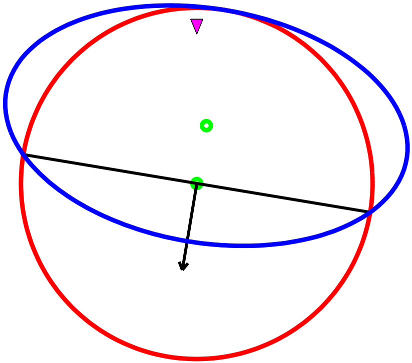


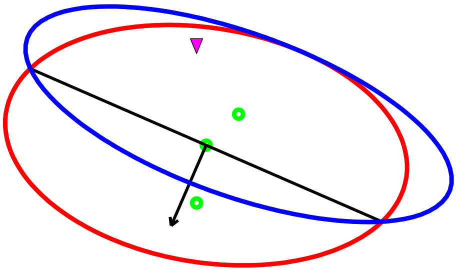


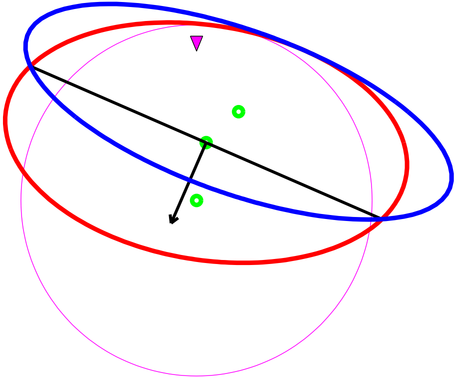


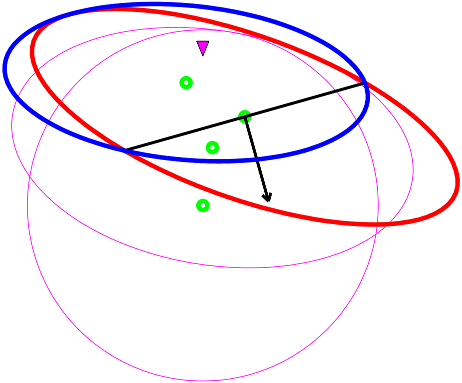


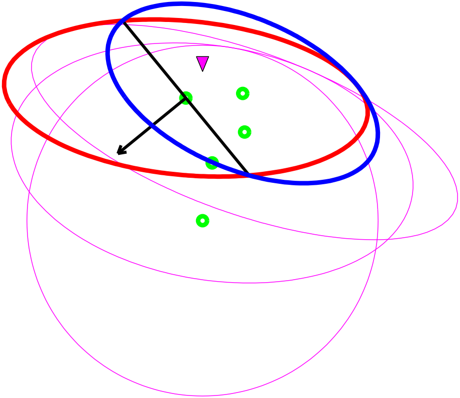


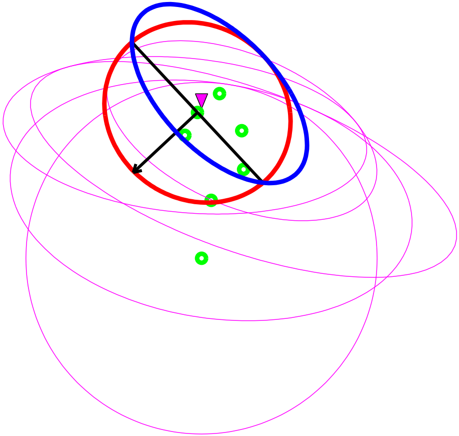


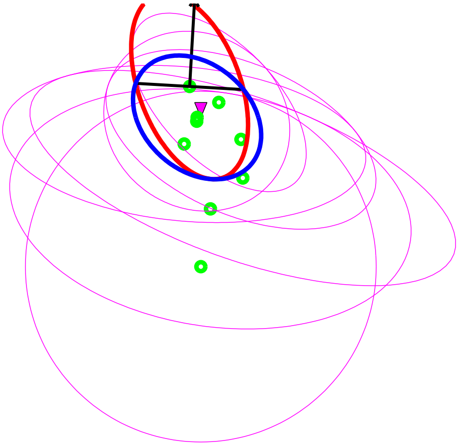












Definition of interior point

General non linear programming problem

Equality and Inequality

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & g(x) \leq 0 \\ & h(x) = 0 \end{aligned}$$

Equality and non-negativity

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f(x) = 0 \\ & x_I \geq 0 \end{aligned}$$

- Interior point: $x > 0$.
- Solution: $f(x) = 0$.
- Feasible solution: $f(x) = 0$, $x_I \geq 0$.
- Interior (feasible) solution: $f(x) = 0$, $x_I > 0$.

Definition of interior point

General non linear programming problem

Equality and Inequality

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & g(x) \leq 0 \\ & h(x) = 0 \end{aligned}$$

Equality and non-negativity

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f(x) = 0 \\ & x_I \geq 0 \end{aligned}$$

- Interior point: $x > 0$.
- Solution: $f(x) = 0$.
- Feasible solution: $f(x) = 0$, $x_I \geq 0$.
- Interior (feasible) solution: $f(x) = 0$, $x_I > 0$.

Definition of interior point

General non linear programming problem

Equality and Inequality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & g(x) \le 0 \\ & h(x) = 0 \end{array}$$

Equality and non-negativity

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f(x) = 0 \\ & x_I \ge 0 \end{array}$$

- Interior point: $x > 0$.
- Solution: $f(x) = 0$.
- Feasible solution: $f(x) = 0$, $x_I \ge 0$.
- Interior (feasible) solution: $f(x) = 0$, $x_I > 0$.

Definition of interior point

General non linear programming problem

Equality and Inequality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & g(x) \le 0 \\ & h(x) = 0 \end{array}$$

Equality and non-negativity

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f(x) = 0 \\ & x_I \ge 0 \end{array}$$

- Interior point: $x > 0$.
- Solution: $f(x) = 0$.
- Feasible solution: $f(x) = 0$, $x_I \ge 0$.
- Interior (feasible) solution: $f(x) = 0$, $x_I > 0$.

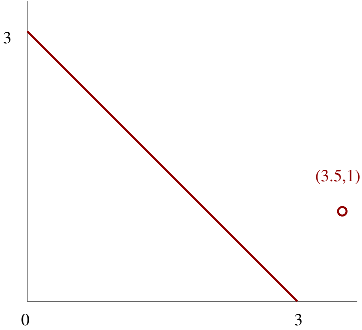
Example of interior infeasible point

Inequality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \leq 3 \\ & x \geq 0 \end{array}$$

Equality and non-negativity

$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & x_1 + x_2 = 3 \\ & x_1, x_2 \geq 0 \end{array}$$



The Affine-Scaling direction

Projection matrix

Given $c \in \mathbb{R}^n$ and a matrix A,
c can be decomposed as

$$c = P_A c + A^T y,$$

where $P_A c \in \mathcal{N}(A)$ is the projection of c into $\mathcal{N}(A)$.

The Affine-Scaling direction

Linearly constrained problem:

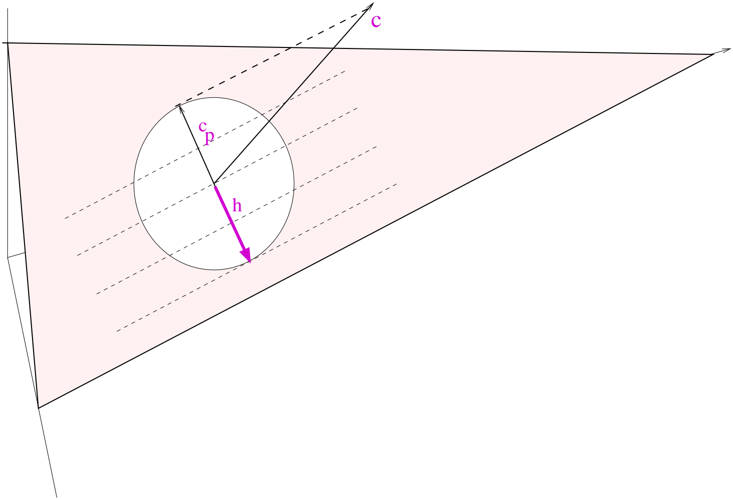
$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

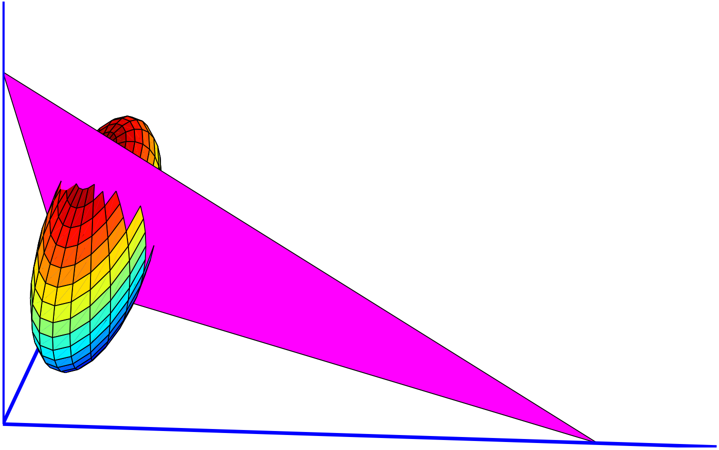
Define $c = \nabla f(x^0)$. The projected gradient (Cauchy) direction is

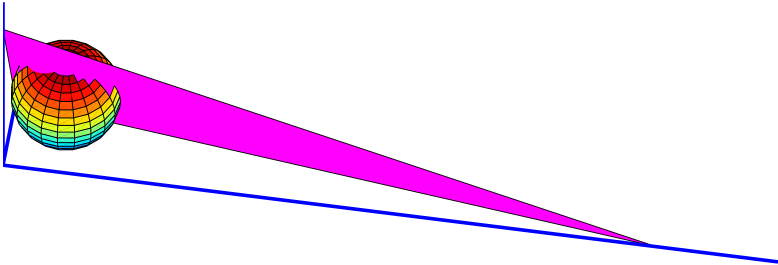
$$c_P = P_A c,$$

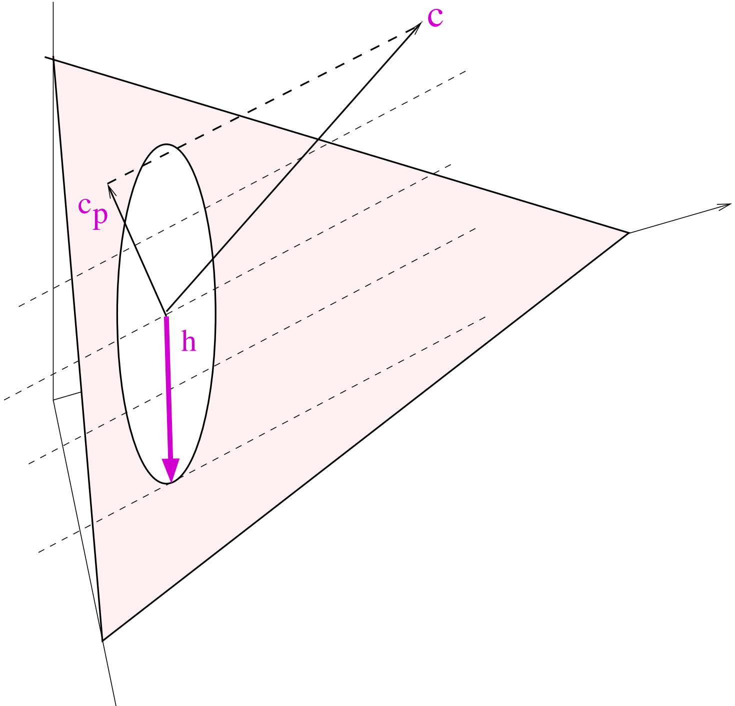
and the steepest descent direction is $d = -c_P$. It solves the trust region problem

$$\text{minimize}\{c^T h \mid Ah = 0, \ \|d\| \leq \Delta\}.$$


c
c
p
h






c
c p
h

The Affine-Scaling direction

Given a feasible point x_0, $X = diag(x_0)$ and $c = \nabla f(x_0)$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \qquad \begin{array}{rcl} x & = & X\bar{x} \\ d & = & X\bar{d} \end{array} \qquad \begin{array}{ll} \text{minimize} & (Xc)^T \bar{x} \\ \text{subject to} & AX\bar{x} = b \\ & \bar{x} \geq 0 \end{array}$$

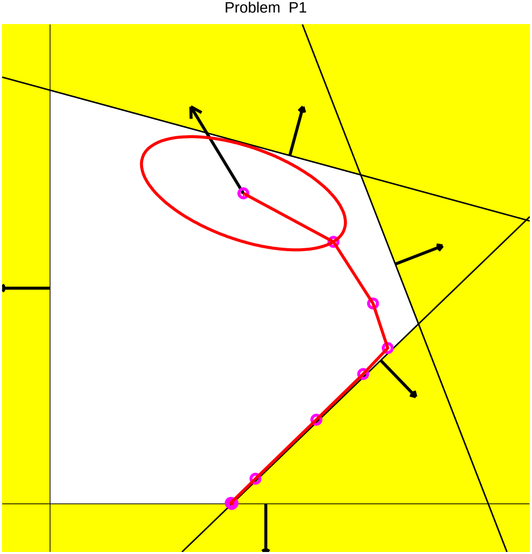
Scaled steepest descent direction:

$$\begin{array}{rcl} \bar{d} & = & -P_{AX}Xc \\ d & = & X\bar{d} = -XP_{AX}Xc \end{array}$$

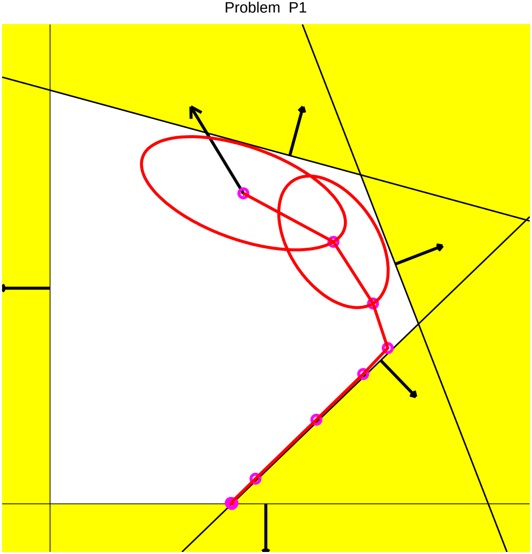
Dikin's direction:

$$\begin{array}{rcl} \bar{d} & = & -P_{AX}Xc \\ d & = & -X\bar{d}/\|\bar{d}\|. \end{array}$$

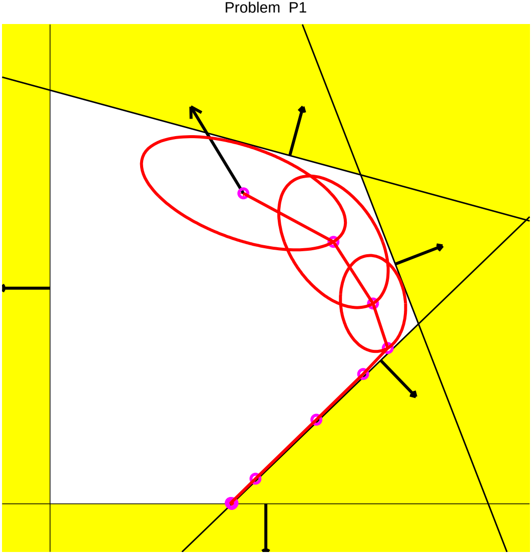
Dikin's algorithm



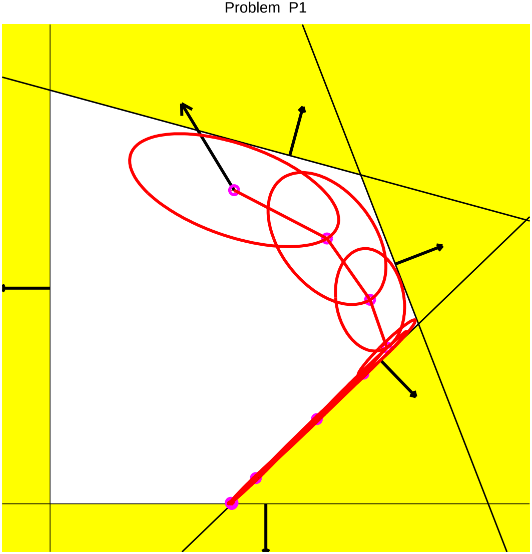
Dikin’s algorithm



Dikin's algorithm

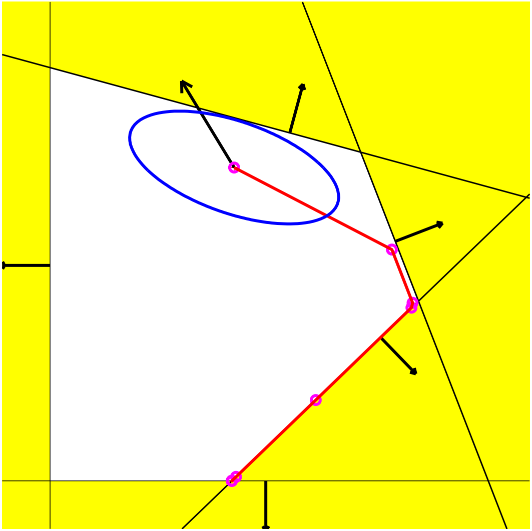


Dikin’s algorithm



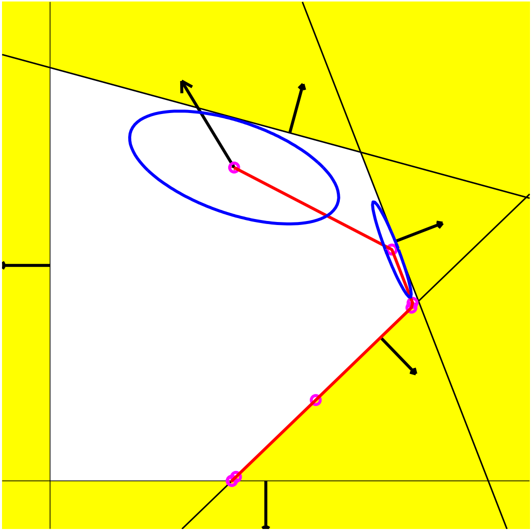
Affine scaling algorithm

Problem P1



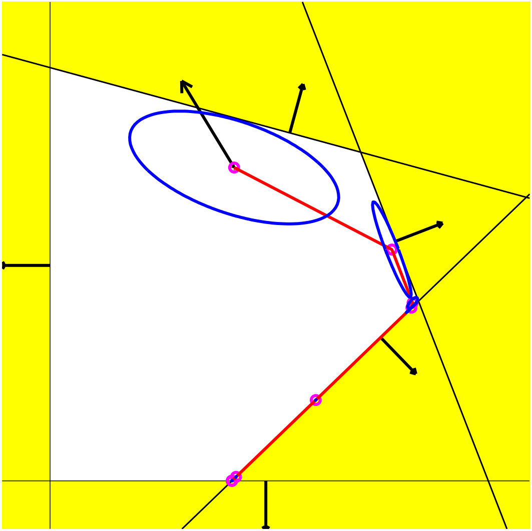
Affine scaling algorithm

Problem P1



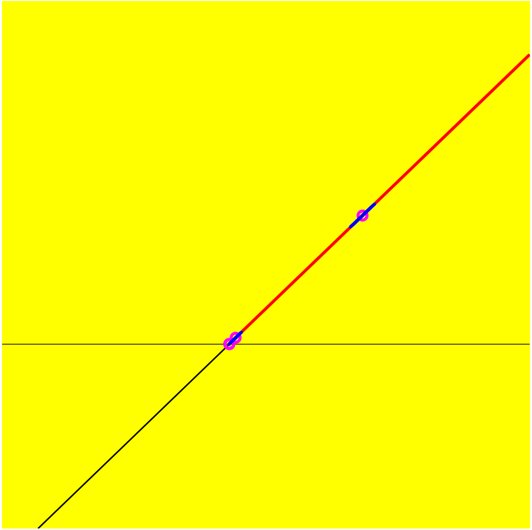
Affine scaling algorithm

Problem P1



Affine scaling algorithm

Problem P1



The logarithmic barrier function

$$x \in \mathbb{R}^n_{++} \mapsto p(x) = -\sum_{i=1}^{n} \log x_i.$$

Scaling: for a diagonal matrix $D > 0$

$$\begin{aligned} p(Dx) &= p(x) + \text{ constant}, \\ p(Dx) - p(Dy) &= p(x) - p(y). \end{aligned}$$

Derivatives:

$$\begin{aligned} \nabla p(x) &= -x^{-1} & \qquad \nabla p(e) &= -e \\ \nabla^2 p(x) &= X^{-2} & \qquad \nabla^2 p(e) &= I. \end{aligned}$$

At $x = e$, the Hessian matrix is the identity, and hence the Newton direction coincides with the Cauchy direction.

The logarithmic barrier function

$$x \in \mathbb{R}^n_{++} \mapsto p(x) = -\sum_{i=1}^{n} \log x_i.$$

Scaling: for a diagonal matrix $D > 0$

$$\begin{aligned} p(Dx) &= p(x) + \text{ constant}, \\ p(Dx) - p(Dy) &= p(x) - p(y). \end{aligned}$$

Derivatives:

$$\begin{aligned} \nabla p(x) &= -x^{-1} & \qquad \nabla p(e) &= -e \\ \nabla^2 p(x) &= X^{-2} & \qquad \nabla^2 p(e) &= I. \end{aligned}$$

At $x = e$, the Hessian matrix is the identity, and hence the Newton direction coincides with the Cauchy direction.

At any $x > 0$, the affine scaling direction coincides with the Newton direction.

The penalized function in linear programming

For $x > 0$, $\mu > 0$ and $\alpha = 1/\mu$,

$$f_\alpha(x) = \alpha c^T x + p(x) \quad \text{or} \quad f_\mu(x) = c^T x + \mu p(x)$$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \qquad \begin{array}{ll} \text{minimize} & c^T x + p(x) \\ \text{subject to} & Ax = b \\ & \bar{x} \geq 0 \end{array}$$

- For $\alpha \geq 0$ f_α is strictly convex and grows indefinitely near the boundary of the feasible set.
- Whenever the minimizers exist, they are defined uniquely by

$$x_\alpha = \text{argmin}_{x \in \Omega} f_\alpha(x).$$

- In particular, if Ω is bounded, x_0 is the analytic center of Ω
- If the optimal face of the problem is bounded, then the curve

$$\alpha > 0 \mapsto x_\alpha$$

is well defined and is called the primal central path.

The penalized function in linear programming

For $x > 0$, $\mu > 0$ and $\alpha = 1/\mu$,

$$f_\alpha(x) = \alpha c^T x + p(x) \quad \text{or} \quad f_\mu(x) = c^T x + \mu p(x)$$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \qquad \begin{array}{ll} \text{minimize} & c^T x + p(x) \\ \text{subject to} & Ax = b \\ & \bar{x} \geq 0 \end{array}$$

- For $\alpha \geq 0$ f_α is strictly convex and grows indefinitely near the boundary of the feasible set.
- Whenever the minimizers exist, they are defined uniquely by

$$x_\alpha = \text{argmin}_{x \in \Omega} f_\alpha(x).$$

- In particular, if Ω is bounded, x_0 is the analytic center of Ω
- If the optimal face of the problem is bounded, then the curve

$$\alpha > 0 \mapsto x_\alpha$$

is well defined and is called the primal central path.

The penalized function in linear programming

For $x > 0$, $\mu > 0$ and $\alpha = 1/\mu$,

$$f_\alpha(x) = \alpha c^T x + p(x) \quad \text{or} \quad f_\mu(x) = c^T x + \mu p(x)$$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \qquad \begin{array}{ll} \text{minimize} & c^T x + p(x) \\ \text{subject to} & Ax = b \\ & \bar{x} \geq 0 \end{array}$$

- For $\alpha \geq 0$ f_α is strictly convex and grows indefinitely near the boundary of the feasible set.
- Whenever the minimizers exist, they are defined uniquely by

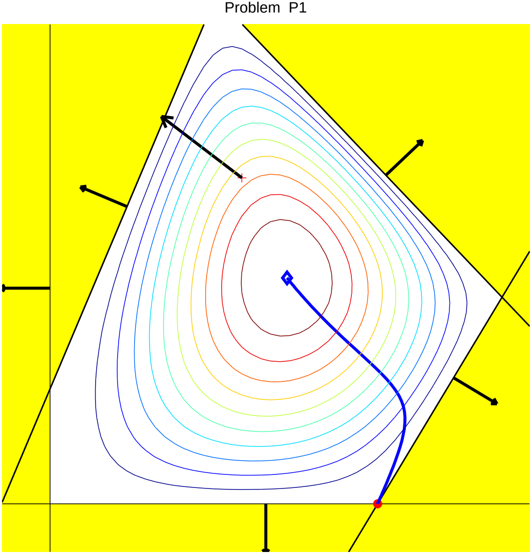
$$x_\alpha = \text{argmin}_{x \in \Omega} f_\alpha(x).$$

- In particular, if Ω is bounded, x_0 is the analytic center of Ω
- If the optimal face of the problem is bounded, then the curve

$$\alpha > 0 \mapsto x_\alpha$$

is well defined and is called the primal central path.

The central path



Equivalent definitions of the central path

There are four equivalent ways of defining central points:

- Minimizers of the penalized function:

$$\text{argmin}_{x\in\Omega} f_\alpha(x).$$

- Analytic centers of constant cost slices

$$\text{argmin}_{x\in\Omega}\{p(x) \mid c^T x = K\}$$

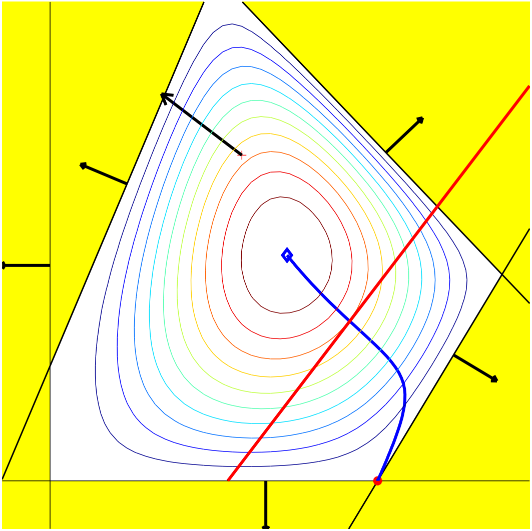
- Renegar centers: Analytic centers of Ω with an extra constraint $c^T x \leq$.

$$\text{argmin}_{x\in\Omega}\{p(x) - \log(K - c^T x) \mid c^T x < K\}$$

- Primal-dual central points (seen ahead).

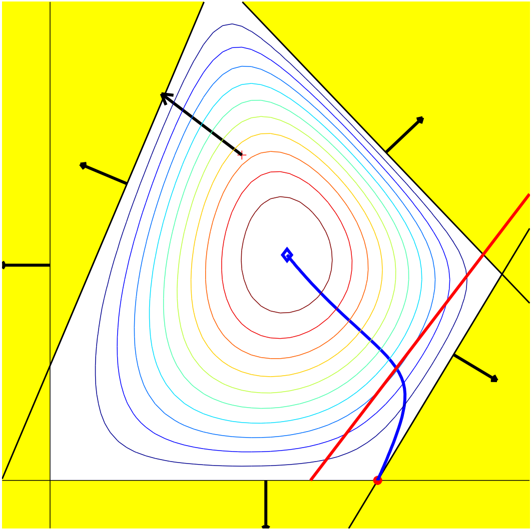
Constant cost slices

Enter the new cut position (one point) and then the initial point



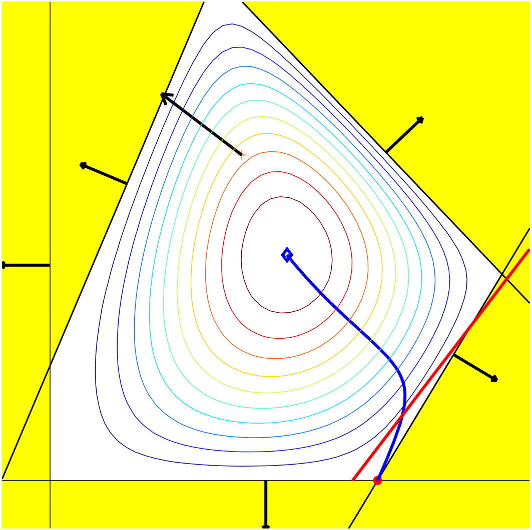
Constant cost slices

Enter the new cut position (one point) and then the initial point

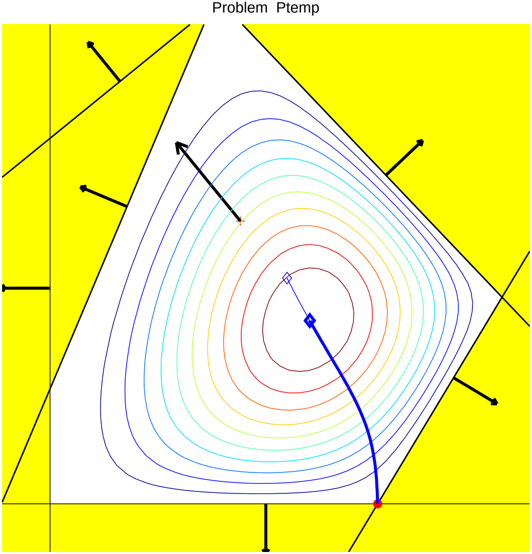


Constant cost slices

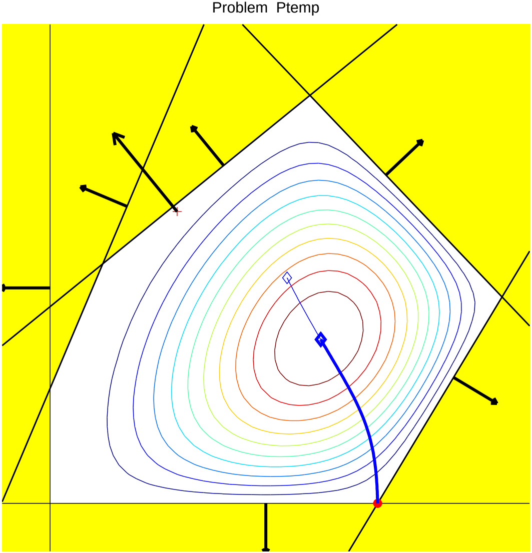
Enter the new cut position (one point) and then the initial point



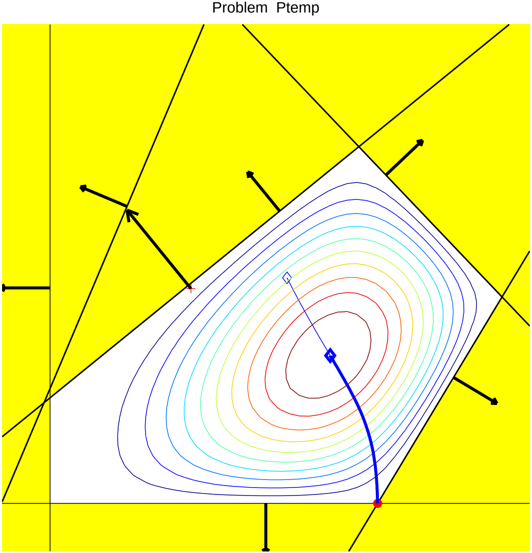
Renegar cuts



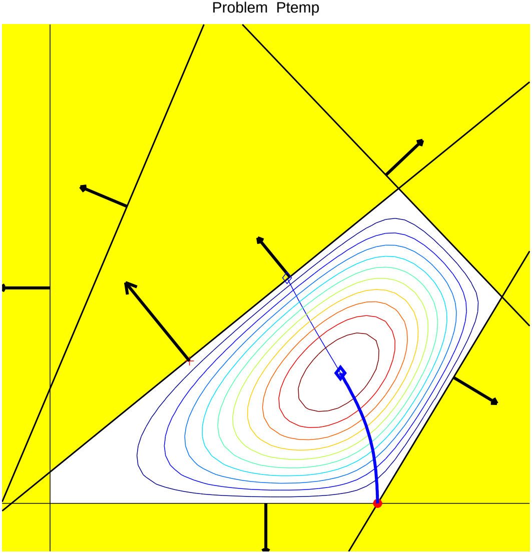
Renegar cuts



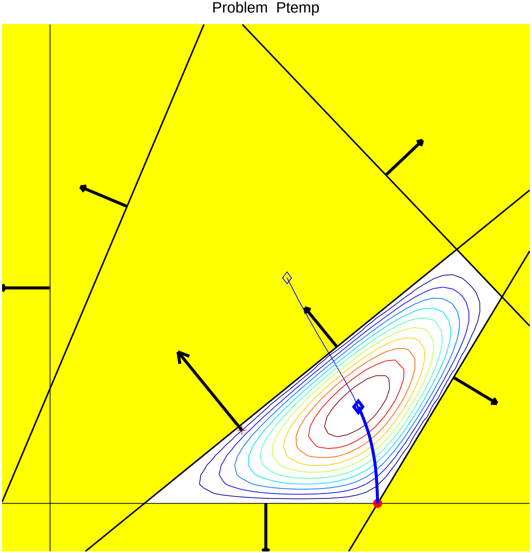
Renegar cuts



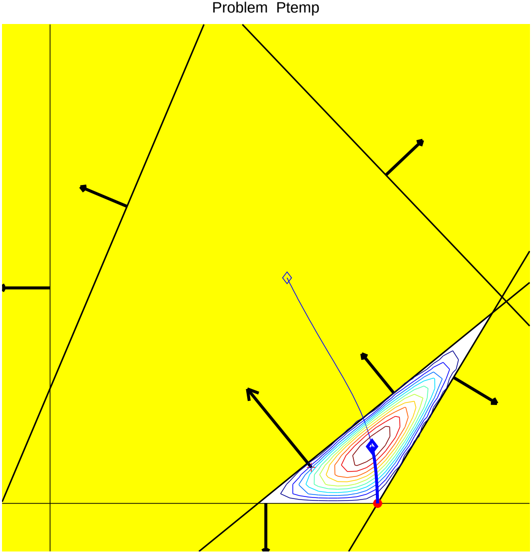
Renegar cuts



Renegar cuts



Renegar cuts



Centering

The most important problem in interior point methods is the following:

Centering problem

Given a feasible interior point x^0 and a value $\alpha \geq 0$, solve approximately the problem

$$\text{minimize}_{x \in \Omega^0} \, \alpha c^T x + p(x).$$

The Newton direction from x^0 coincides with the affine-scaling direction, and hence is the best possible. It is given by

$$\begin{aligned} d &= X\bar{d}, \\ \bar{d} &= -P_{AX}X(\alpha c - x^{-1}) = -\alpha P_{AX}Xc + P_{AX}e. \end{aligned}$$

Centering

The most important problem in interior point methods is the following:

Centering problem

Given a feasible interior point x^0 and a value $\alpha \geq 0$, solve approximately the problem

$$\text{minimize}_{x \in \Omega^0} \, \alpha c^T x + p(x).$$

The Newton direction from x^0 coincides with the affine-scaling direction, and hence is the best possible. It is given by

$$\begin{aligned} d &= X\bar{d}, \\ \bar{d} &= -P_{AX}X(\alpha c - x^{-1}) = -\alpha P_{AX}Xc + P_{AX}e. \end{aligned}$$

Efficiency of the Newton step for centering

Newton direction:

$$
\begin{aligned}
d &= X\bar{d}, \\
\bar{d} &= -P_{AX}X(\alpha c - x^{-1}) = -\alpha P_{AX}Xc + P_{AX}e.
\end{aligned}
$$

We define the Proximity to the central point as

$$\delta(x,\alpha) = \|\bar{d}\| = \| -\alpha P_{AX}Xc + P_{AX}e\|.$$

The following important theorem says how efficient it is:

Theorem

Consider a feasible point x and a parameter α. Let $x^+ = x + d$ be the point resulting from a Newton centering step. If $\delta(x,\alpha) = \delta < 1$, then $\delta(x^+,\alpha) < \delta^2$.

If $\delta(x,\alpha) \leq 0.5$, then this value is a very good approximation to the euclidean distance between e and $X^{-1}x_\alpha$, i. e., between x and x_α in the scaled space.

Efficiency of the Newton step for centering

Newton direction:

$$\begin{aligned} d &= X\bar{d}, \\ \bar{d} &= -P_{AX}X(\alpha c - x^{-1}) = -\alpha P_{AX}Xc + P_{AX}e. \end{aligned}$$

We define the Proximity to the central point as

$$\delta(x,\alpha) = \|\bar{d}\| = \| -\alpha P_{AX}Xc + P_{AX}e\|.$$

The following important theorem says how efficient it is:

Theorem

Consider a feasible point x and a parameter α. Let $x^+ = x + d$ be the point resulting from a Newton centering step. If $\delta(x,\alpha) = \delta < 1$, then $\delta(x^+,\alpha) < \delta^2$.

If $\delta(x,\alpha) \leq 0.5$, then this value is a very good approximation to the euclidean distance between e and $X^{-1}x_\alpha$, i. e., between x and x_α in the scaled space.

Primal results as we saw are important to give a geometrical meaning to the procedures, and to develop the intuition. Also, these results can be generalized to a large class of problems, by generalizing the idea of barrier functions.

From now on we shall deal with primal-dual results, which are more efficient for linear and non-linear programming.

The Linear Programming Problem

LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

LD

$$\begin{aligned} \text{maximize} \quad & b^T y \\ \text{subject to} \quad & A^T y \leq c \end{aligned}$$

The Linear Programming Problem

LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

LD

$$\begin{aligned} \text{maximize} \quad & b^T y \\ \text{subject to} \quad & A^T y + s = c \\ & s \geq 0 \end{aligned}$$

The Linear Programming Problem

LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

KKT: multipliers y, s

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

LD

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y + s = c \\ & s \geq 0 \end{array}$$

The Linear Programming Problem

LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

LD

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y + s = c \\ & s \geq 0 \end{array}$$

KKT: multipliers y, s

$$\begin{array}{rcl} A^T y + s & = & c \\ Ax & = & b \\ xs & = & 0 \\ x, s & \geq & 0 \end{array}$$

KKT: multipliers x

$$\begin{array}{rcl} A^T y + s & = & c \\ Ax & = & b \\ xs & = & 0 \\ x, s & \geq & 0 \end{array}$$

The Linear Programming Problem

LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

LD

$$\begin{aligned} \text{maximize} \quad & b^T y \\ \text{subject to} \quad & A^T y + s = c \\ & s \geq 0 \end{aligned}$$

Primal-dual optimality

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

Duality gap

For x, y, s feasible,

$$c^T x - b^T y = x^T s \geq 0$$

The Linear Programming Problem

LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

LD

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y + s = c \\ & s \geq 0 \end{array}$$

Primal-dual optimality

$$\begin{array}{rcl} A^T y + s & = & c \\ Ax & = & b \\ xs & = & 0 \\ x, s & \geq & 0 \end{array}$$

Duality gap

For x, y, s feasible,

$$c^T x - b^T y = x^T s \geq 0$$

(LP) has solution x and (LD) has solution y, s if and only if the optimality conditions have solution x, y, s.

Primal-dual centering

Let us write the KKT conditions for the centering problem (now using μ instead of $\alpha = 1/\mu$).

$$\begin{array}{ll} \text{minimize} & c^T x - \mu \sum \log x_i \\ \text{subject to} & Ax = b \\ & x > 0 \end{array}$$

A feasible point x is a minimizer if and only if the gradient of the objective function is orthogonal to the null space of A, which means

$$c - \mu x^{-1} = -A^T y,$$

for some $y \in \mathbb{R}^m$. Defining $s = \mu x^{-1}$, we get the conditions for a primal-dual center:

Primal-dual center

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

Primal-dual centering

Let us write the KKT conditions for the centering problem (now using μ instead of $\alpha = 1/\mu$).

$$\begin{array}{ll} \text{minimize} & c^T x - \mu \sum \log x_i \\ \text{subject to} & Ax = b \\ & x > 0 \end{array}$$

A feasible point x is a minimizer if and only if the gradient of the objective function is orthogonal to the null space of A, which means

$$c - \mu x^{-1} = -A^T y,$$

for some $y \in \mathbb{R}^m$. Defining $s = \mu x^{-1}$, we get the conditions for a primal-dual center:

Primal-dual center

$$\begin{array}{rcl} xs & = & \mu e \\ A^T y + s & = & c \\ Ax & = & b \\ x, s & > & 0 \end{array}$$

Generalization

Let us write the KKT conditions for the convex quadratic programming problem

$$\begin{array}{ll} \text{minimize} & c^T x + \frac{1}{2} x^T H x \\ \text{subject to} & Ax = b \\ & x > 0 \end{array}$$

The first KKT condition is written as

$$c + Hx - A^T y - s = 0$$

To obtain a symmetrical formulation for the problem, we may multiply this equation by a matrix B whose rows for a basis for the null space of A. Then $BA^T y = 0$, and we obtain the following conditions conditions:

$$\begin{array}{rcl} xs & = & 0 \\ -BHx + Bs & = & Bc \\ Ax & = & b \\ x, s & \geq & 0 \end{array}$$

Horizontal linear complementarity problem

In any case, the problem can be written as

$$\begin{aligned} xs &= 0 \\ Qx + Rs &= b \\ x, s &\geq 0 \end{aligned}$$

This is a linear complementarity problem, which includes linear and quadratic programming as particular problems. The techniques studied here apply to these problems, as long as the following monotonicity condition holds:

For any feasible pair of directions (u, v) such that $Qu + Rv = 0$, we have $u^T v \geq 0$.

The optimal face: the optimal solutions must satisfy $x_i s_i = 0$ for $i = 1, \dots, n$. This is a combinatorial constraint, responsible for all the difficulty in the solution.

Horizontal linear complementarity problem

In any case, the problem can be written as

$$\begin{aligned} xs &= 0 \\ Qx + Rs &= b \\ x, s &\geq 0 \end{aligned}$$

This is a linear complementarity problem, which includes linear and quadratic programming as particular problems. The techniques studied here apply to these problems, as long as the following monotonicity condition holds:

For any feasible pair of directions (u, v) such that $Qu + Rv = 0$, we have $u^T v \geq 0$.

The optimal face: the optimal solutions must satisfy $x_i s_i = 0$ for $i = 1, \ldots, n$. This is a combinatorial constraint, responsible for all the difficulty in the solution.

Primal-dual centering: the Newton step

Given x, s feasible and $\mu > 0$, find

$$\begin{aligned} x^+ &= x + u \\ s^+ &= x + v \end{aligned}$$

such that

$$\begin{aligned} x^+ s^+ &= \mu e \\ Qx^+ + Rs^+ &= b \end{aligned} \qquad \begin{aligned} xs + su + xv + uv &= \mu e \\ Qu + Rv &= 0 \end{aligned}$$

Newton step

$$\begin{aligned} Xv + Su &= -xs + \mu e \\ Qu + Rv &= 0 \end{aligned}$$

Solving this linear system is all the work. In the case of linear programming one should keep the multipliers y and simplify the resulting system of equations

Primal-dual centering: the Newton step

Given x, s feasible and $\mu > 0$, find

$$\begin{aligned} x^+ &= x+u \\ s^+ &= x+v \end{aligned}$$

such that

$$\begin{aligned} x^+ s^+ &= \mu e \\ Qx^+ + Rs^+ &= b \end{aligned} \qquad \begin{aligned} xs+su+xv+uv &= \mu e \\ Qu+Rv &= 0 \end{aligned}$$

Newton step

$$\begin{aligned} Xv+Su &= -xs+\mu e \\ Qu+Rv &= 0 \end{aligned}$$

Solving this linear system is all the work. In the case of linear programming one should keep the multipliers y and simplify the resulting system of equations

Primal-dual centering: the Newton step

Given x, s feasible and $\mu > 0$, find

$$\begin{aligned} x^+ &= x + u \\ s^+ &= x + v \end{aligned}$$

such that

$$\begin{aligned} x^+ s^+ &= \mu e \\ Q x^+ + R s^+ &= b \end{aligned} \qquad \begin{aligned} xs + su + xv + uv &= \mu e \\ Qu + Rv &= 0 \end{aligned}$$

Newton step

$$\begin{aligned} Xv + Su &= -xs + \mu e \\ Qu + Rv &= 0 \end{aligned}$$

Solving this linear system is all the work. In the case of linear programming one should keep the multipliers y and simplify the resulting system of equations

Primal-dual centering: Proximity measure

Given x, s feasible and $\mu > 0$, we want

$$xs = \mu e \text{ or equivalently } \frac{xs}{\mu} - e = 0$$

The actual error in this equation gives the proximity measure:

Proximity measure

$$x, s, \mu \mapsto \delta(x, s, \mu) = \|\frac{xs}{\mu} - e\|.$$

Theorem

Given a feasible pair (x, s) and a parameter μ, Let $x^+ = x + u$ and $s^+ = s + v$be the point resulting from a Newton centering step. If $\delta(x, s, \mu) = \delta < 1$, then

$$\delta(x^+, s^+, \mu) < \frac{1}{\sqrt{8}} \frac{\delta^2}{1-\delta}.$$

In particular, if $\delta \leq 0.7$, then $\delta(x^+, s^+, \mu) < \delta^2$.

Primal-dual centering: Proximity measure

Given x, s feasible and $\mu > 0$, we want

$$xs = \mu e \text{ or equivalently } \frac{xs}{\mu} - e = 0$$

The actual error in this equation gives the proximity measure:

Proximity measure

$$x, s, \mu \mapsto \delta(x, s, \mu) = \|\frac{xs}{\mu} - e\|.$$

Theorem

Given a feasible pair (x, s) and a parameter μ, Let $x^+ = x + u$ and $s^+ = s + v$be the point resulting from a Newton centering step. If $\delta(x, s, \mu) = \delta < 1$, then

$$\delta(x^+, s^+, \mu) < \frac{1}{\sqrt{8}} \frac{\delta^2}{1 - \delta}.$$

In particular, if $\delta \leq 0.7$, then $\delta(x^+, s^+, \mu) < \delta^2$.

Primal-dual path following: Traditional approach

- Assume that we have x, s, μ such that (x, s) is feasible and $\delta(x, s, \mu) \leq \alpha < 1$
- Choose $\mu^+ = \gamma\mu$, with $\gamma < 1$.
- Use Newton's algorithm (with line searches to avoid infeasible points) to find (x^+, s^+) such that $\delta(x^+, s^+, \mu^+) \leq \alpha$

Neighborhood of the central path

Given $\beta \in (0, 1)$, we define the neighborhood $\eta(\alpha)$ as the set of all feasible pairs (x, s) such that for some $\mu > 0$

$$\delta(x, s, \mu) \leq \beta$$

The methods must ensure that all points are in such a neighborhood, using line searches

Primal-dual path following: Traditional approach

- Assume that we have x, s, μ such that (x,s) is feasible and $\delta(x,s,\mu) \leq \alpha < 1$
- Choose $\mu^+ = \gamma\mu$, with $\gamma < 1$.
- Use Newton's algorithm (with line searches to avoid infeasible points) to find (x^+, s^+) such that $\delta(x^+, s^+, \mu^+) \leq \alpha$

Neighborhood of the central path

Given $\beta \in (0,1)$, we define the neighborhood $\eta(\alpha)$ as the set of all feasible pairs (x,s) such that for some $\mu > 0$

$$\delta(x,s,\mu) \leq \beta$$

The methods must ensure that all points are in such a neighborhood, using line searches

Primal-dual path following: Traditional approach

- Assume that we have x, s, μ such that (x, s) is feasible and $\delta(x, s, \mu) \leq \alpha < 1$
- Choose $\mu^+ = \gamma\mu$, with $\gamma < 1$.
- Use Newton's algorithm (with line searches to avoid infeasible points) to find (x^+, s^+) such that $\delta(x^+, s^+, \mu^+) \leq \alpha$

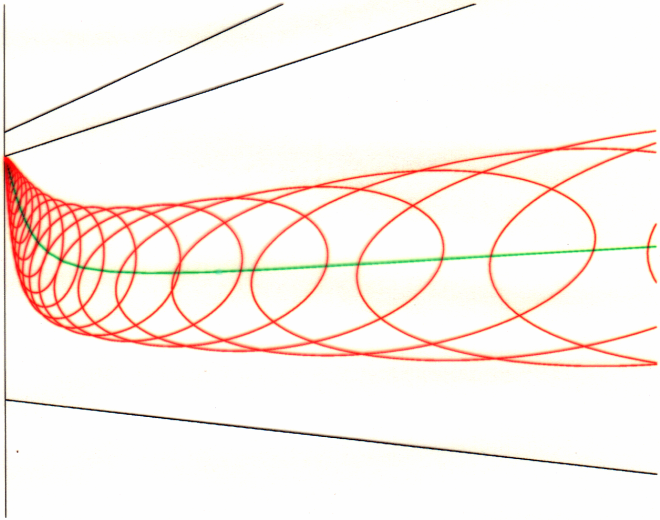
Neighborhood of the central path

Given $\beta \in (0, 1)$, we define the neighborhood $\eta(\alpha)$ as the set of all feasible pairs (x, s) such that for some $\mu > 0$

$$\delta(x, s, \mu) \leq \beta$$

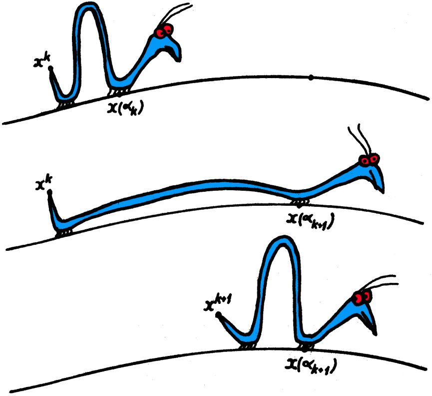
The methods must ensure that all points are in such a neighborhood, using line searches

A neighborhood of the central path

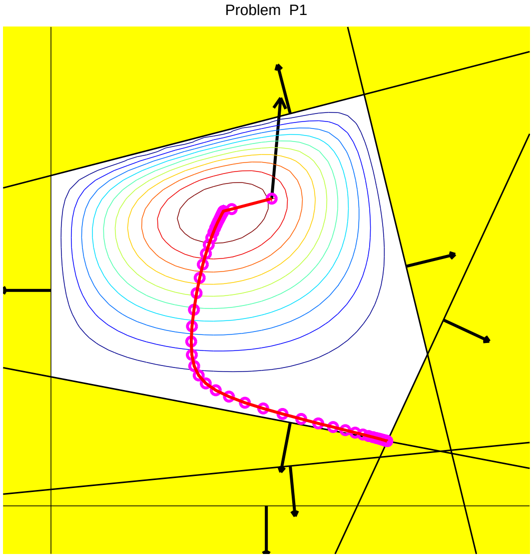


Short steps

Using γ near 1, we obtain short steps. With $\gamma = 0.4/\sqrt{n}$, only one Newton step is needed at each iteration, and the algorithm is polynomial: it finds a solution with precision 2^{-L} in $O(\sqrt{n}L)$ iterations.

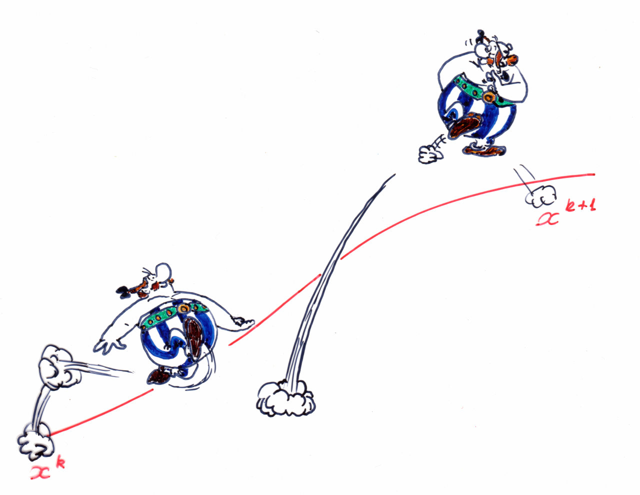


Short steps

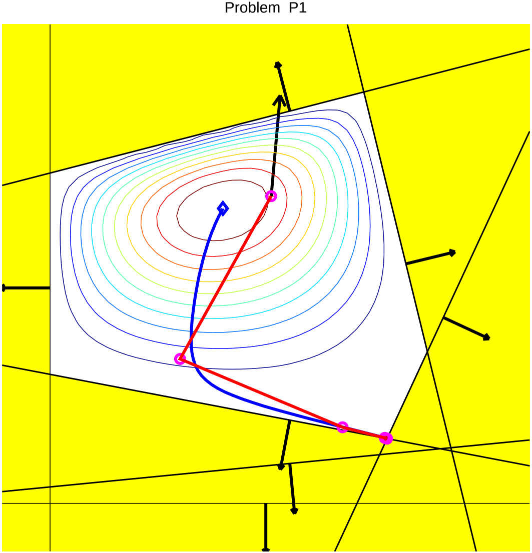


Large steps

Using γ small, say $\gamma = 0.1$, we obtain large steps. This uses to work well in practice, but some sort of line search is needed, to avoid leaving the neighborhood. Predictor-corrector methods are better, as we shall see.



Large steps

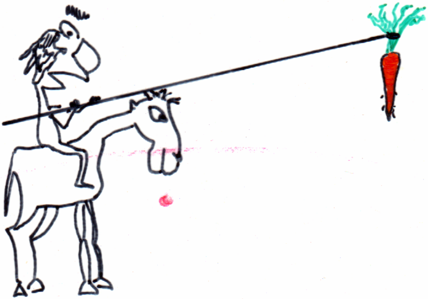


Adaptive methods

Assume that (x, s) feasible is given in $\eta(\beta)$, but no value of μ is given. Then we know:

- if (x, s) is a central point, then $xs = \mu e$ implies $x^T s = n\mu$. Hence the best choice for μ is $\mu = s^T s/n$.
- If (x, s) is not a central point, the value $\mu(x, s) = x^T s/n$ gives a parameter value which in a certain sense is the best possible.
- An adaptive algorithm does not use a value of μ coming from a former iteration: it computes $\mu(x, s)$ and then chooses a value $\gamma\mu(x, s)$ as new target.
- The target may be far. Compute a direction (u, v) and follow it until

$$\delta(x + \lambda u, s + \lambda v, \mu(x + \lambda u, s + \lambda v)) = \beta$$



Predictor-corrector methods

Alternate two kinds of iterations:

- Predictor: An iteration starts with (x, s) near the central path, and computes a Newton step (u, v) with goal $\gamma\mu(x, s)$, γ small.
- Follow it until

$$\delta(x+\lambda u, s+\lambda v, \mu(x+\lambda u, s+\lambda v)) = \beta$$

- Corrector: Set $x^+ = x+\lambda u$, $s^+ = s+\lambda v$, compute $\mu = \mu(x^+, s^+)$ and take a Newton step with target μ

- If the predictor uses $\gamma = 0$, it is called the affine scaling step. It has no centering, and tries to solve the original problem in one step.
- Using a neighborhood with $\beta = 0.5$, the resulting algorithm (the Mizuno-Todd-Ye algorithm) converges quadratically to an optimal solution, keeping the complexity at its best value of $O(\sqrt{n}L)$ iterations.

Predictor-corrector methods

Alternate two kinds of iterations:

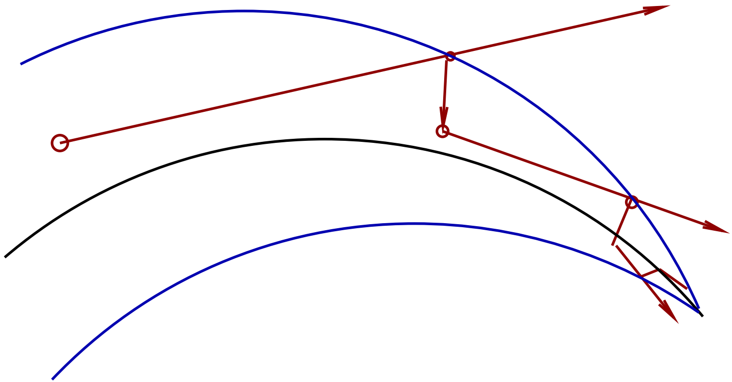
- Predictor: An iteration starts with (x, s) near the central path, and computes a Newton step (u, v) with goal $\gamma\mu(x, s)$, γ small.
- Follow it until

$$\delta(x+\lambda u, s+\lambda v, \mu(x+\lambda u, s+\lambda v)) = \beta$$

- Corrector: Set $x^+ = x + \lambda u$, $s^+ = s + \lambda v$, compute $\mu = \mu(x^+, s^+)$ and take a Newton step with target μ

- If the predictor uses $\gamma = 0$, it is called the affine scaling step. It has no centering, and tries to solve the original problem in one step.
- Using a neighborhood with $\beta = 0.5$, the resulting algorithm (the Mizuno-Todd-Ye algorithm) converges quadratically to an optimal solution, keeping the complexity at its best value of $O(\sqrt{n}L)$ iterations.

Predictor-corrector



Mehrotra Predictor-corrector method: second order

When computing the Newton step, we eliminated the nonlinear term uv in the equation

$$\begin{aligned} xs + su + xv + uv &= \mu e \\ Qu + Rv &= 0 \end{aligned}$$

The second order method corrects the values of u, v by estimating the value of the term uv by a predictor step.

- Predictor: An iteration starts with (x, s) near the central path, and computes a Newton step (u, v) with goal μ^+, small. The first equation is

$$xv + su = -xs + \mu^+ e$$

- Compute a correction $(\Delta u, \Delta v)$ by

$$x\Delta v + s\Delta u = -uv.$$

- Line search: Set $x^+ = x + \lambda u + \lambda^2 \Delta u$, $s^+ = s + \lambda v + \lambda^2 \Delta v$, by a line search so that $\delta(x^+, s^+, \mu(x^+, s^+)) = \beta$.

Mehrotra Predictor-corrector