KTH Mathematics

2E5295/5B5749 Convex optimization with engineering applications

Lecture 9

Interior methods

Nonlinearly constrained convex program

Consider a convex optimization problem on the form

$$
(CP) \qquad \begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g_i(x) \geq 0, \quad i \in \mathcal{I}, & \mathcal{I} \cup \mathcal{E} = \{1, \ldots, m\}, \\ & a_i^T x - b_i = 0, \quad i \in \mathcal{E}, & \mathcal{I} \cap \mathcal{E} = \emptyset, \\ & x \in I\!R^n, & \end{array}
$$

where $f : I\!R^n \to I\!R$ and $-g_i : I\!R^n \to I\!R$ are convex and twice continuously differentiable on $I\!R^n$.

The inequality constraints give an added combinatorial problem of identifying the constraints that are active at the solution.

One way of dealing with inequality constraints is via a *barrier transformation*.

Barrier transformation

Consider replacing an inequality constraint $g_i(x) \geq 0$ by a *logarithmic barrier term* $-\ln(g_i(x))$ added to the objective function.

Other barrier terms are possible. The logarithmic barrier term is the "canonic" choice.

The effect is a perturbed problem where an infinite cost is incurred as $g_i(x) \to 0$.

The weight which we put on the barrier term is denoted by μ and referred to as the *barrier parameter*.

The combinatorial effect is removed at the expense of a perturbation of the original problem.

The problem resulting from a barrier transformation

For a positive barrier parameter μ, the barrier transformed problem becomes

$$(CP_\mu) \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) - \mu \sum_{i \in \mathcal{I}} \ln(g_i(x)) \qquad \mathcal{I} \bigcup \mathcal{E} = \{1, \ldots, m\},$$
$$a_i^T x - b_i = 0, \quad i \in \mathcal{E}, \qquad \mathcal{I} \bigcap \mathcal{E} = \emptyset.$$

Note that convexity is preserved.

Proposition. *Let $g_i : \mathbb{R}^n \to \mathbb{R}$ be a concave function on $\mathbb{R}^n$. Then, $-\ln(g_i(x))$ is a convex function on the convex set $\{x \in \mathbb{R}^n : g_i(x) > 0\}$.*

The barrier trajectory

Under suitable assumptions, the barrier transformed problem has a unique optimal solution $x(\mu)$ and corresponding Lagrange multipliers $\lambda_i(\mu)$, $i \in \mathcal{E}$, for each $\mu > 0$.

In this situation, the *barrier trajectory* is defined as the set $\{x(\mu) : \mu > 0\}$. The barrier trajectory is sometimes referred to as the *central path*.

Theorem. *Under suitable assumptions, the barrier trajectory is well defined and it holds that* $\lim_{\mu \to 0} x(\mu) = x^*$, $\lim_{\mu \to 0} \mu / g_i(x(\mu)) = \lambda_i^*$, $i \in \mathcal{I}$, *and* $\lim_{\mu \to 0} \lambda_i(\mu) = \lambda_i^*$, $i \in \mathcal{E}$, *where* x^* *is an optimal solution to* (CP), *and* λ^* *is an associated Lagrange multiplier vector.*

Hence, the barrier trajectory converges to an optimal solution.

A primal approach: Sequential unconstrained minimization

We may now apply the methods outlined previously for (approximately) solving a sequence of unstrained minimization problems for decreasing values of the barrier parameter μ.

This is sometimes referred to as sequential unconstrained minimization techniques. We shall refer to this approach as a primal approach. The method is an *interior method*, i.e., it generates points that lie in the (relative) interior of the feasible set.

Let $f_\mu(x) = f(x) - \mu \sum_{i \in \mathcal{I}} \ln(g_i(x))$. Then,

$$\nabla^2 f_\mu(x) = \nabla^2 f(x) - \sum_{i \in \mathcal{I}} \frac{\mu}{g_i(x)} \nabla^2 g_i(x) + \sum_{i \in \mathcal{I}} \frac{\mu}{(g_i(x))^2} \nabla g_i(x) \nabla g_i(x)^T.$$

As $\mu \to 0$, $\nabla^2 f_\mu(x(\mu))$ becomes increasingly ill-conditioned in general, with the condition number tending to infinity.

Optimality conditions for the barrier transformed problem

We may write the necessary optimality conditions for (P_μ) as

$$\nabla f(x) - \sum_{i\in\mathcal{I}} \nabla g_i(x) \frac{\mu}{g_i(x)} - \sum_{i\in\mathcal{E}} a_i\lambda_i = 0,$$
$$a_i^T x - b_i = 0, \quad i \in \mathcal{E},$$

where $g_i(x) > 0$, $i \in \mathcal{I}$, is required implicitly.

If auxiliary variables λ_i, $i \in \mathcal{I}$ are introduced, as defined by $\lambda_i = \mu/g_i(x)$, $i \in \mathcal{I}$, we obtain

$$\nabla f(x) - \sum_{i\in\mathcal{I}} \nabla g_i(x)\lambda_i - \sum_{i\in\mathcal{E}} a_i\lambda_i = 0,$$
$$a_i^T x - b_i = 0, \quad i \in \mathcal{E},$$
$$\lambda_i - \frac{\mu}{g_i(x)} = 0, \quad i \in \mathcal{I}.$$

A primal-dual reformulation of the optimality conditions

Note that if $g_i(x) > 0$, then $\lambda_i - \mu/(g_i(x)) = 0$ if and only if $g_i(x)\lambda_i - \mu = 0$,

The optimality conditions for (CP_μ) are thus equivalent to

$$\begin{aligned} \nabla f(x) - \sum_{i\in\mathcal{I}} \nabla g_i(x)\lambda_i - \sum_{i\in\mathcal{E}} a_i\lambda_i &= 0, \\ a_i^T x - b_i &= 0, \quad i \in \mathcal{E}, \\ g_i(x)\lambda_i - \mu &= 0, \quad i \in \mathcal{I}, \end{aligned}$$

with the implicit requirement $g_i(x) > 0$, $i \in \mathcal{I}$.

We will refer to this form of nonlinear equations defining the optimality conditions for (CP_μ) as the *primal-dual nonlinear equations*, and a method based on approximately solving these equations as a *primal-dual interior method*.

Perturbed optimality conditions

We may view the primal-dual nonlinear equations

$$\begin{aligned} \nabla f(x) - \sum_{i \in \mathcal{I}} \nabla g_i(x) \lambda_i - \sum_{i \in \mathcal{E}} a_i \lambda_i &= 0, \\ a_i^T x - b_i &= 0, \quad i \in \mathcal{E}, \\ g_i(x) \lambda_i - \mu &= 0, \quad i \in \mathcal{I}, \end{aligned}$$

with the implicit requirement $g_i(x) > 0$, $i \in \mathcal{I}$, as a perturbation of the optimality conditions for (CP),

$$\begin{aligned} \nabla f(x) - \sum_{i \in \mathcal{I}} \nabla g_i(x) \lambda_i - \sum_{i \in \mathcal{E}} a_i \lambda_i &= 0, \\ a_i^T x - b_i &= 0, \quad i \in \mathcal{E}, \\ g_i(x) \geq 0, \quad \lambda_i \geq 0, &\quad i \in \mathcal{I}, \\ g_i(x) \lambda_i &= 0, \quad i \in \mathcal{I}. \end{aligned}$$

Interior methods

- The term *interior methods* is used as a common name for methods of barrier type for nonlinear optimization.
- Barrier methods in primal form are from the 60s. They have some less desirable properties due to ill-conditioning. The methods were revived in 1984 for linear programming.
- Primal-dual interior methods are methods of the 90s. They have "better" behavior.

We will consider the special case of linear programming, for simplicity.

Interior methods for linear programming

We want to solve the linear programs

$$(PLP) \quad \begin{array}{ll} \text{min} & c^T x \\ \text{subject to} & Ax = b, \\ & x \geq 0, \end{array} \qquad (DLP) \quad \begin{array}{ll} \text{max} & b^T y \\ \text{subject to} & A^T y + s = c, \\ & s \geq 0. \end{array}$$

The primal-dual nonlinear equations

If the complementarity condition $x_j s_j = 0$ is perturbed to $x_j s_j = \mu$ for a positive barrier parameter μ, we obtain a nonlinear equation on the form

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ x_j s_j &= \mu, \quad j = 1, \ldots, n. \end{aligned}$$

The inequalities $x \geq 0$, $s \geq 0$ are kept "implicitly".

Proposition. *The primal-dual nonlinear equations are well defined and have a unique solution with $x > 0$ and $s > 0$ for all $\mu > 0$ if $\{x : Ax = b, x > 0\} \neq \emptyset$ and $\{(y, s) : A^T y + s = c, s > 0\} \neq \emptyset$.*

We refer to this solution as $x(\mu)$, $y(\mu)$ and $s(\mu)$.

The primal-dual nonlinear equations, cont.

The primal-dual nonlinear equations may be written in vector form:

$$Ax = b,$$
$$A^T y + s = c,$$
$$XSe = \mu e,$$

where $X = \text{diag}(x)$, $S = \text{diag}(s)$ and $e = (1, 1, \ldots, 1)^T$.

Proposition. *A solution $(x(\mu), y(\mu), s(\mu))$ is such that $x(\mu)$ is feasible to (PLP) and $y(\mu)$, $s(\mu)$ is feasible to (DLP) with duality gap $n\mu$.*

Primal point of view

Primal point of view: $x(\mu)$ solves

$$(P_\mu) \qquad \begin{aligned} \text{minimize} \quad & c^T x - \mu \sum_{j=1}^{n} \ln x_j \\ \text{subject to} \quad & Ax = b, \quad x > 0, \end{aligned}$$

with $y(\mu)$ as Lagrange multiplier vector of $Ax = b$.

Optimality conditions for (P_μ):

$$\begin{aligned} c_j - \frac{\mu}{x_j} &= A_j^T y, \quad j = 1, \ldots, n, \\ Ax &= b, \\ x &> 0. \end{aligned}$$

Dual point of view

Dual point of view: $y(\mu)$ and $s(\mu)$ solve

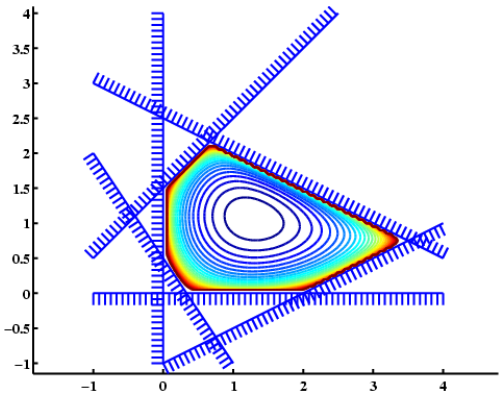
$$(D_\mu) \qquad \begin{aligned} \text{maximize} \quad & b^T y + \mu \sum_{j=1}^{n} \ln s_j \\ \text{subject to} \quad & A^T y + s = c, \quad s > 0, \end{aligned}$$

with $x(\mu)$ as Lagrange multiplier vector of $A^T y + s = c$.

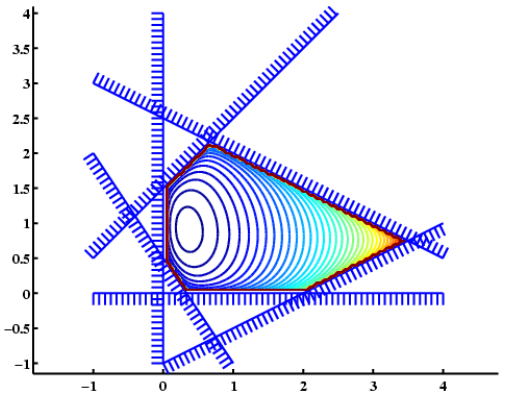
Optimality conditions for (D_μ):

$$\begin{aligned} b &= Ax, \\ \frac{\mu}{s_j} &= x_j, \quad j = 1, \ldots, n, \\ A^T y + s &= c, \\ s &> 0. \end{aligned}$$

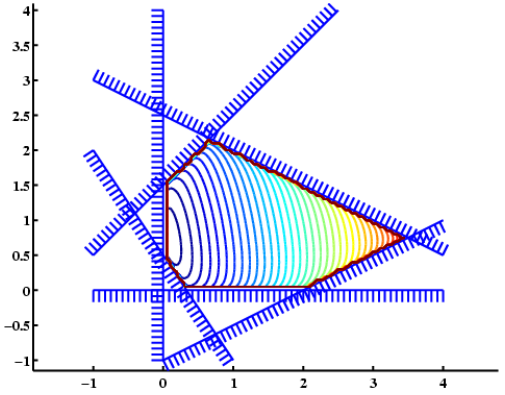
Primal barrier function for example linear program



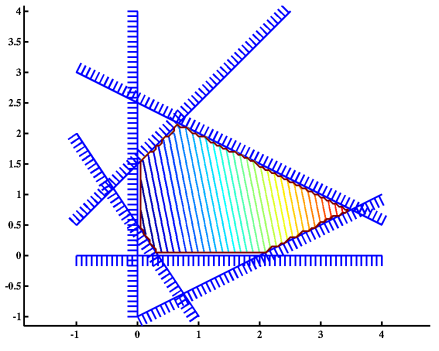
$\mu = 5$



$\mu = 1$



$\mu = 0.3$



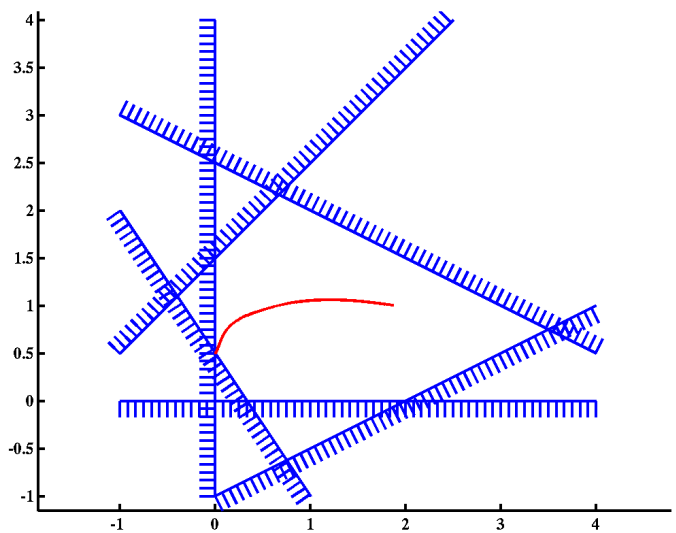
$\mu = 10^{-16}$

The barrier trajectory

The *barrier trajectory* is defined as the set $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$.

The primal-dual system of nonlinear equations is to prefer. Pure primal and pure dual point of view gives high nonlinearity.

Example of primal part of barrier trajectory:



Properties of the barrier trajectory

Theorem. *If the barrier trajectory is well defined, then* $\lim_{\mu\to 0} x(\mu) = x^*$, $\lim_{\mu\to 0} y(\mu) = y^*$, $\lim_{\mu\to 0} s(\mu) = s^*$, *where* x^* *is an optimal solution to* (PLP), *and* y^*, s^* *are optimal solutions to* (DLP).

Hence, the barrier trajectory converges to an optimal solution.

Theorem. *If the barrier trajectory is well defined, then* $\lim_{\mu\to 0} x(\mu)$ *is the optimal solution to the problem*

$$\begin{aligned} \text{min} \quad & -\textstyle\sum_{i\in\mathcal{B}} \ln x_i \\ \text{subject to} \quad & \textstyle\sum_{i\in\mathcal{B}} A_i x_i = b, \quad x_i > 0,\ i \in \mathcal{B}, \end{aligned}$$

where $\mathcal{B} = \{i : \widetilde{x}_i > 0$ *for some optimal solution* $\widetilde{x}$ *of* $(PLP)\}$.

Thus, the barrier trajectory converges to an extreme point only if (PLP) has unique optimal solution.

Primal-dual interior method

A primal-dual interior method is based on Newton-iterations on the perturbed optimality conditions.

For a given point x, y, s, with $x > 0$ and $s > 0$ a suitable value of μ is chosen. The Newton-iteration then becomes

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

Common choice $\mu = \sigma \dfrac{x^T s}{n}$ for some $\sigma \in [0, 1]$.

Note that $Ax = b$ and $A^T y + s = c$ need not be satisfied at the initial point. It will be satisfied at $x + \Delta x$, $y + \Delta y$, $s + \Delta s$.

An iteration in a primal-dual interior method

- Choose μ.
- Compute Δx, Δy and Δs from

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

- Find maximum steplength $\alpha_{\max}$ from $x + \alpha \Delta x \geq 0$, $s + \alpha \Delta s \geq 0$.
- Let $\alpha = \min\{1, 0.999 \cdot \alpha_{\max}\}$.
- Let $x = x + \alpha \Delta x$, $y = y + \alpha \Delta y$, $s = s + \alpha \Delta s$.

(This steplength rule is simplified, and is not guaranteed to ensure convergence.)

Strategies for choosing σ

Proposition. *Assume that* x *satisfies* $Ax = b$, $x > 0$, *and assume that* y, s *satisfies* $A^T y + s = c$, $s > 0$, *and let* $\mu = \sigma x^T s/n$. *Then*

$$(x + \alpha \Delta x)^T (s + \alpha \Delta s) = (1 - \alpha(1 - \sigma)) x^T s.$$

It is desirable to have σ small and α large. These goals are in general contradictory.

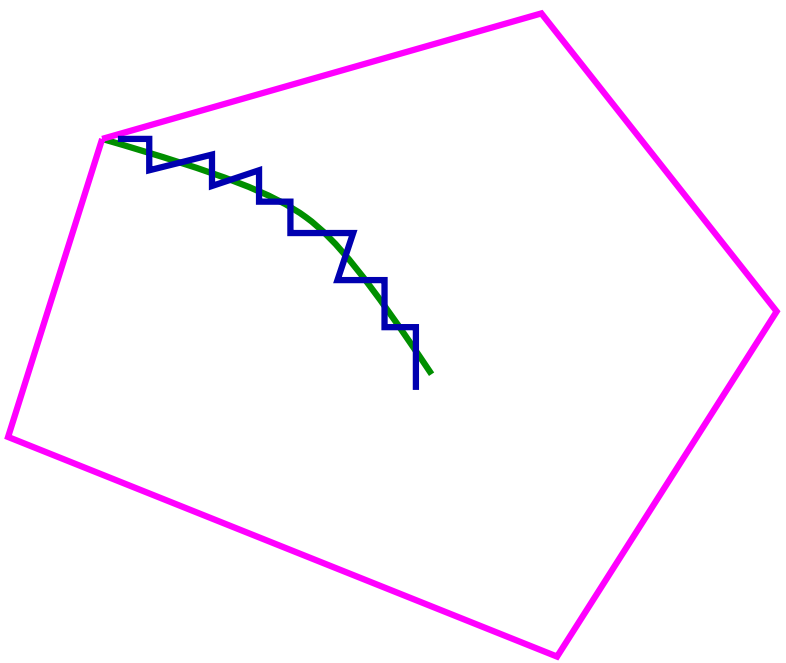
Three main strategies:

- Short-step method, σ close to 1.
- Long-step method, σ significantly smaller than 1.
- Predictor-corrector method, $\sigma = 0$ each even iteration and $\sigma = 1$ each odd iteration.

Short-step method

We may choose $\sigma^k = 1 - \delta/\sqrt{n}$, $\alpha^k = 1$.

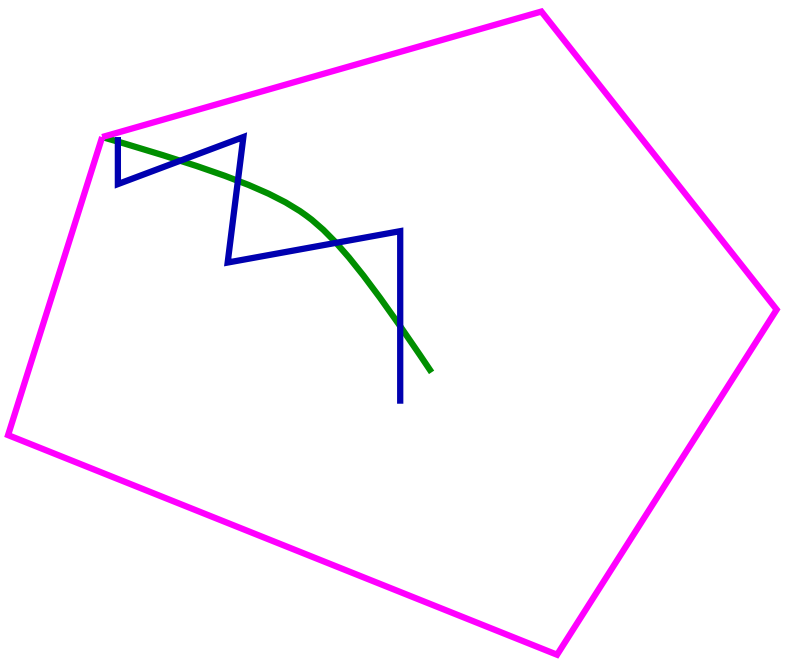
The iterates remain close to the trajectory.



Polynomial complexity. In general not efficient enough.

Long-step method

We may choose $\sigma^k = 0.1$, α^k given by proximity to the trajectory.

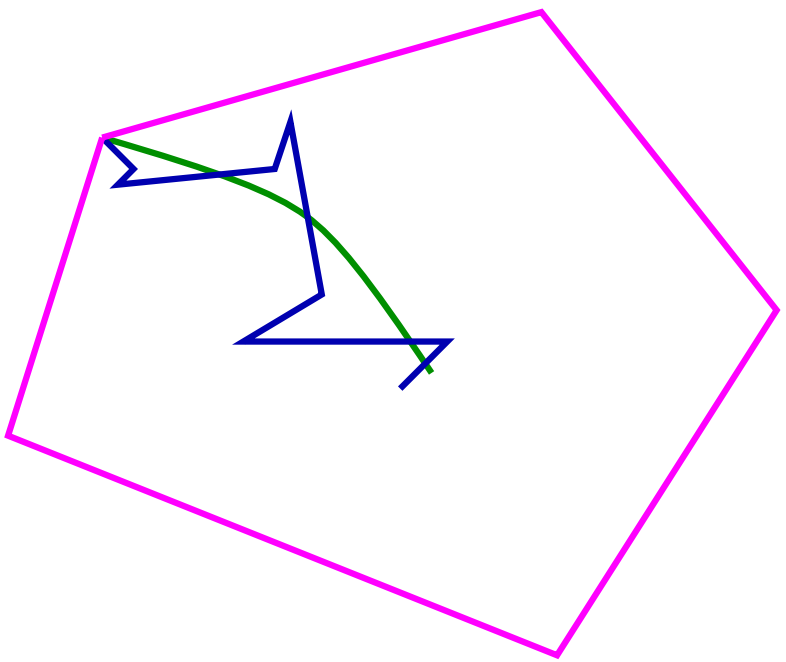


Polynomial complexity.

Predictor-corrector method

$\sigma^k = 0$, α^k given by proximity to the trajectory for k even.

$\sigma^k = 1$, $\alpha^k = 1$ for k odd.

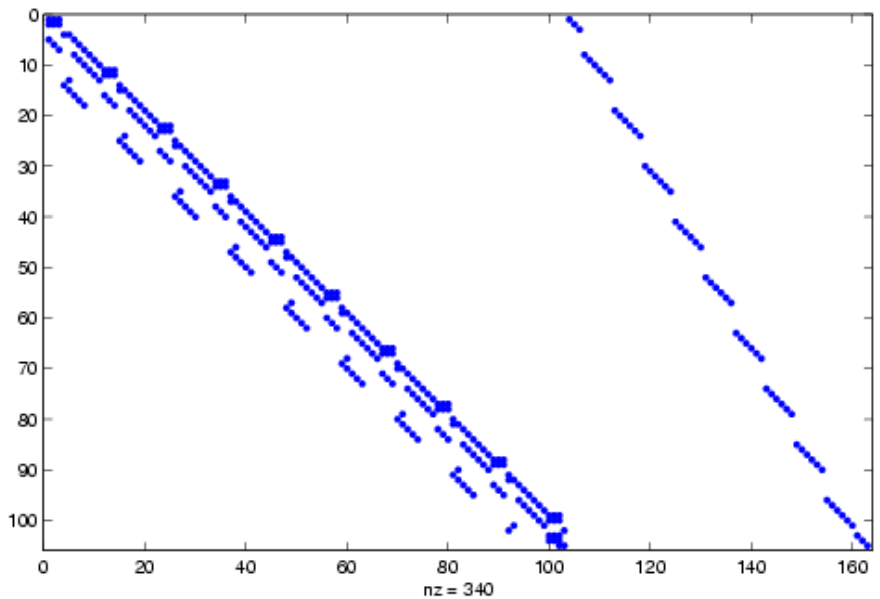


Polynomial complexity.

Behavior of interior method for linear programming

Normally few iterations, in the order or 20. Typically does not grow with problem size.

Sparse systems of linear equations. Example A:



The iterates become more computationally expensive as problem size increases.

Not clear how to "warm start" the method efficiently.

On the solution of the linear systems of equation that arise

The aim is to compute Δx, Δy and Δs from

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

One may for example solve

$$\begin{pmatrix} X^{-1}S & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = - \begin{pmatrix} c - \mu X^{-1} e - A^T y \\ Ax - b \end{pmatrix},$$

or, alternatively

$$AXS^{-1}A^T \Delta y = AXS^{-1}(c - \mu X^{-1} e - A^T y) + b - Ax.$$

Suggested reading

Suggested reading in the textbook:

- Chapter 10.
- Chapter 11.

Be selective in your reading.

Interior methods for convex programs

A primal-dual interior method approximately solves

$$\nabla f(x) - \sum_{i \in \mathcal{I}} \nabla g_i(x) \lambda_i - \sum_{i \in \mathcal{E}} a_i \lambda_i = 0,$$

$$a_i^T x - b_i = 0, \quad i \in \mathcal{E},$$

$$g_i(x)\lambda_i - \mu = 0, \quad i \in \mathcal{I},$$

with the implicit requirement $g_i(x) > 0$, $\lambda_i > 0$, $i \in \mathcal{I}$.

Newton iteration takes the form

$$\begin{pmatrix} \nabla^2_{xx} l(x,\lambda) & A_{\mathcal{E}}^T & A_{\mathcal{I}}(x)^T \\ A_{\mathcal{E}} & 0 & 0 \\ \Lambda_{\mathcal{I}} A_{\mathcal{I}}(x) & 0 & -G_{\mathcal{I}}(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta\lambda_{\mathcal{E}} \\ -\Delta\lambda_{\mathcal{I}} \end{pmatrix} = - \begin{pmatrix} \nabla_x l(x,\lambda) \\ A_{\mathcal{E}} x - b_{\mathcal{E}} \\ G_{\mathcal{I}}(x)\lambda_{\mathcal{I}} - \mu e \end{pmatrix},$$

where $G_{\mathcal{I}}(x) = \text{diag}(g_{\mathcal{I}}(x))$ and $\Lambda_{\mathcal{I}} = \text{diag}(\lambda_{\mathcal{I}})$.

Slater's constraint qualification

The barrier trajectory defined by the primal-dual nonlinear equations is well defined if some *constraint quailification* holds.

For example, Slater's constraint qualification, which holds if there exists an $x \in I\!R^n$ such that $A_{\mathcal{E}}x - b_{\mathcal{E}} = 0$ and $g_{\mathcal{I}}(x) > 0$.

Another constraint qualification is $g_{\mathcal{I}}$ affine.

Constraint qualification satisfied means that strong duality holds.

An iteration in a primal-dual interior method

An iteration in a primal-dual interior method takes the following form, given $\mu > 0$, x such that $g_{\mathcal{I}}(x) > 0$ and λ such that $\lambda_{\mathcal{I}} > 0$.

- Compute Δx and $\Delta\lambda$ from the Newton equation.
- Choose "suitable" steplength α such that $g_{\mathcal{I}}(x + \alpha\Delta x) > 0$, $\lambda_{\mathcal{I}} + \alpha\Delta\lambda_{\mathcal{I}} > 0$.
- $x \leftarrow x + \alpha\Delta x, \;\; \lambda \leftarrow \lambda + \alpha\Delta\lambda$.
- If (x, λ) "sufficiently close to" $(x(\mu), \lambda(\mu))$, reduce μ.

In general, suitable choices give overall method with polynomial complexity. (Convexity is not quite enough.)

Adding slack variables

Requiring $g_{\mathcal{I}}(x) > 0$ may be beneficial or complicating.

Alternative is to add slack variables $s_{\mathcal{I}}$ and write the primal-dual nonlinear equations as

$$\begin{aligned} \nabla f(x) - A_{\mathcal{I}}(x)^T \lambda_{\mathcal{I}} - A_{\mathcal{E}}^T \lambda_{\mathcal{E}} &= 0 \\ A_{\mathcal{E}} x - b_{\mathcal{E}} &= 0, \\ g_{\mathcal{I}}(x) - s_{\mathcal{I}} &= 0, \\ s_{\mathcal{I}} \lambda_{\mathcal{I}} - \mu e &= 0. \end{aligned}$$

with the implicit requirement $s_{\mathcal{I}} > 0$, $\lambda_{\mathcal{I}} > 0$.

Adding slack variables, cont.

Newton iteration may be written

$$\begin{pmatrix} \nabla^2_{xx} l(x,\lambda) & A_{\mathcal{E}}^T & A_{\mathcal{I}}(x)^T \\ A_{\mathcal{E}} & 0 & 0 \\ \Lambda_{\mathcal{I}} A_{\mathcal{I}}(x) & 0 & -S_{\mathcal{I}} \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta\lambda_{\mathcal{E}} \\ -\Delta\lambda_{\mathcal{I}} \end{pmatrix} = - \begin{pmatrix} \nabla_x l(x,\lambda) \\ A_{\mathcal{E}} x - b_{\mathcal{E}} \\ \Lambda_{\mathcal{I}} g_{\mathcal{I}}(x) - \mu e \end{pmatrix},$$

where $S_{\mathcal{I}} = \text{diag}(s_{\mathcal{I}})$ and $\Lambda_{\mathcal{I}} = \text{diag}(\lambda_{\mathcal{I}})$.

Then, $\Delta s_{\mathcal{I}} = -s_{\mathcal{I}} + g_{\mathcal{I}}(x) + A_{\mathcal{I}}(x)\Delta x$.

Ensure that $s_{\mathcal{I}} > 0$ and $\lambda_{\mathcal{I}} > 0$ in the linesearch. We obtain $g_{\mathcal{I}}(x) \geq 0$ asymptotically.

Primal and dual semidefinite programs

Primal and dual semidefinite programs may be written as

$$(PSDP) \qquad \begin{array}{ll} \underset{X \in \mathcal{S}^n}{\text{minimize}} & \text{trace}(CX) \\ \text{subject to} & \text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m, \\ & X = X^T \succeq 0. \end{array}$$

and

$$(DSDP) \qquad \begin{array}{ll} \underset{y \in I\!R^m, S \in \mathcal{S}^n}{\text{maximize}} & b^T y \\ \text{subject to} & \sum_{i=1}^m A_i y_i + S = C, \\ & S = S^T \succeq 0, \end{array}$$

where $\mathcal{S}^n = \{X \in I\!R^{n \times n} : X = X^T\}$.

Without loss of generality: $C \in \mathcal{S}^n$, $A_i \in \mathcal{S}^n$, $i = 1, \ldots, n$.

The relative interiors

Let $F(PSDP)$ denote the feasible set of $(PSDP)$, i,e,.

$$F(PSDP) = \{X \in \mathcal{S}^n : \text{trace}(A_i X) = b_i,\ i = 1, \ldots, m,\ X \succeq 0\}.$$

The *relative interior* of $(PSDP)$ is defined as

$$F^0(PSDP) = \{X \in F(PSDP) : X \succ 0\}.$$

Analogously, with

$$F(DSDP) = \{(y, S) \in I\!R^m \times \mathcal{S}^n : \sum_{i=1}^{m} A_i y_i + S = C,\ S \succeq 0\},$$

the *relative interior* of $(DSDP)$ is defined as

$$F^0(DSDP) = \{(y, S) \in F(DSDP) : S \succ 0\}.$$

Nonempty strict relative interior gives strong duality

Theorem. *If $F^0(PSDP) \neq \emptyset$, then the duality gap is zero and $(DSDP)$ has an optimal solution.*

The proof can be given by the separating hyperplane theorem (in a slightly different form from what what was presented earlier).

Corollary. *If $F^0(DSDP) \neq \emptyset$, then the duality gap is zero and $(PSDP)$ has an optimal solution.*

Corollary. *If $F^0(PSDP) \neq \emptyset$ and $F^0(DSDP) \neq \emptyset$, then the duality gap is zero, and both $(PSDP)$ and $(DSDP)$ have optimal solutions.*

Alternative formulation of a semidefinite program

A semidefinite program on standard form may be rewritten as

$$\begin{aligned} \text{minimize} \quad & \text{trace}(CX) \\ \text{subject to} \quad & \text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m, \\ & X = X^T, \\ & u^T X u \geq 0 \ \ \forall\ u \in I\!R^n : \|u\|_2 = 1. \end{aligned}$$

We obtain an infinite number of linear inequality constraints. Hence, more complex than a linear program. (A *semi-infinite program*.)

How do we solve $(PSDP)$? The generalization of the simplex method is not obvious. Interior methods may be generalized in a rather straightforward way.

Alternative formulation of a semidefinite program

$$
(PSDP) \qquad \begin{array}{ll} \text{minimize} & \text{trace}(CX) \\ \text{subject to} & \text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m, \\ & X = X^T \succeq 0. \end{array}
$$

We may equivalently formulate $(PSDP)$ as

$$
(PSDP') \qquad \begin{array}{ll} \text{minimize} & \text{trace}(CX) \\ \text{subject to} & \text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m, \\ & X = X^T, \\ & \eta_j(X) \geq 0, \quad j = 1, \ldots, n, \end{array}
$$

where $\eta_j(X)$ denotes the jth eigenvalue of X.

A drawback is that $(PSDP')$ is nondifferentiable.

A barrier transformation for semidefinite programming

$$
(PSDP_\mu) \qquad \begin{array}{ll} \text{minimize} & \text{trace}(CX) - \mu \sum_{j=1}^{n} \ln(\eta_j(X)) \\ \text{subject to} & \text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m, \\ & X = X^T \succ 0. \end{array}
$$

Note that $\sum_{j=1}^{n} \ln(\eta_j(X)) = \ln \prod_{j=1}^{n} \eta_j(X) = \ln \det(X)$, which gives

$$
(PSDP_\mu) \qquad \begin{array}{ll} \text{minimize} & \text{trace}(CX) - \mu \ln \det(X) \\ \text{subject to} & \text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m, \\ & X = X^T \succ 0. \end{array}
$$

This is a differentiable problem.

The derivative of the barrier term

We may write $X = \sum_{k=1}^{n} \sum_{l=k}^{n} x_{kl} E_{kl}$,

where $E_{kk} = e_k e_k^T$ och $E_{kl} = e_k e_l^T + e_l e_k^T,\ k < l.$

Proposition. *If* $X \in \mathcal{S}^n$, $X \succ 0$, *then* $\dfrac{\partial \ln \det(X)}{\partial x_{kl}} = \operatorname{trace}(X^{-1} E_{kl})$.

Proof. We obtain $\ln \det(X + tE_{kl}) = \ln \det(X) + \ln \det(I + tX^{-1}E_{kl})$.
Let η_j, $j = 1, \ldots, n$, denote the eigenvalues of $X^{-1}E_{kl}$. Then,

$$\ln \det(I + tX^{-1}E_{kl}) = \ln \prod_{j=1}^{n} (1 + t\,\eta_j) = \sum_{j=1}^{n} \ln(1 + t\,\eta_j)$$

$$= t \sum_{j=1}^{n} \eta_j + o(t) = t \operatorname{trace}(X^{-1}E_{kl}) + o(t).$$

The result follows by letting $t \to 0$. □

The second derivative of the barrier term

Proposition. *If* $X \in \mathcal{S}^n$, $X \succ 0$, *then*

$$\frac{\partial^2 \ln\det(X)}{\partial x_{ij}\partial x_{kl}} = \operatorname{trace}(X^{-1}E_{ij}X^{-1}E_{kl}).$$

Proof. The identity $XX^{-1} \equiv I$ gives $\dfrac{\partial X^{-1}}{\partial x_{ij}} = -X^{-1}E_{ij}X^{-1}$.

Hence, $\dfrac{\partial \operatorname{trace}(X^{-1}E_{kl})}{\partial x_{ij}} = -\operatorname{trace}(X^{-1}E_{ij}X^{-1}E_k l)$. □

Corollary. *The function* $-\ln\det(X)$ *is convex on* $\{X \in \mathcal{S}^n : X \succ 0\}$.

Proof. For $P \in \mathcal{S}^n$, we have

$$-\sum_{i=1}^{n}\sum_{j=i}^{n}\sum_{k=1}^{n}\sum_{l=k}^{n} p_{ij}\frac{\partial^2 \ln\det(X)}{\partial x_{ij}\partial x_{kl}}p_{kl} = \operatorname{trace}(X^{-1}PX^{-1}P)$$
$$= \operatorname{trace}(X^{-1/2}PX^{-1}PX^{-1/2}) \geq 0.$$

□

Optimality conditions for the semidefinite barrier problem

The optimality conditions for $(PSDP_\mu)$ are given by

$$\text{trace}(CE_{kl}) - \mu\,\text{trace}(X^{-1}E_{kl}) - \sum_{i=1}^{n}\text{trace}(A_iE_{kl})y_i = 0,\ 1 \le k \le l \le n,$$

$$\text{trace}(A_iX) = b_i, \quad i = 1,\dots,m,$$

$$X = X^T.$$

Equivalently,

$$\sum_{i=1}^{n} A_iy_i + \mu X^{-1} = C,$$

$$\text{trace}(A_iX) = b_i, \quad i = 1,\dots,m,$$

$$X = X^T.$$

Primal-dual form of optimality conditions

Let $S = \mu X^{-1}$, which gives

$$\sum_{i=1}^{n} A_i y_i + S = C,$$
$$\text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m,$$
$$XS = \mu I.$$

This is the primal-dual system of equations which defines the barrier trajectory, where we in addition implicitly require $X \succ 0$ and $S \succ 0$.

Note that the solution gives $X \in \mathcal{S}^n$ and $S \in \mathcal{S}^n$ since $A_i \in \mathcal{S}^n$, $i = 1, \ldots, m$, and $C \in \mathcal{S}^n$.

The solution is primal and dual feasible, respectively, with difference in objective function value given by $n\mu$.

Nonempty relative interior ensures existence of barrier trajectory

The primal-dual form of the barrier trajectory is

$$\sum_{i=1}^{n} A_i y_i + S = C,$$
$$\text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m,$$
$$XS = \mu I, \quad X \succ 0, \quad S \succ 0.$$

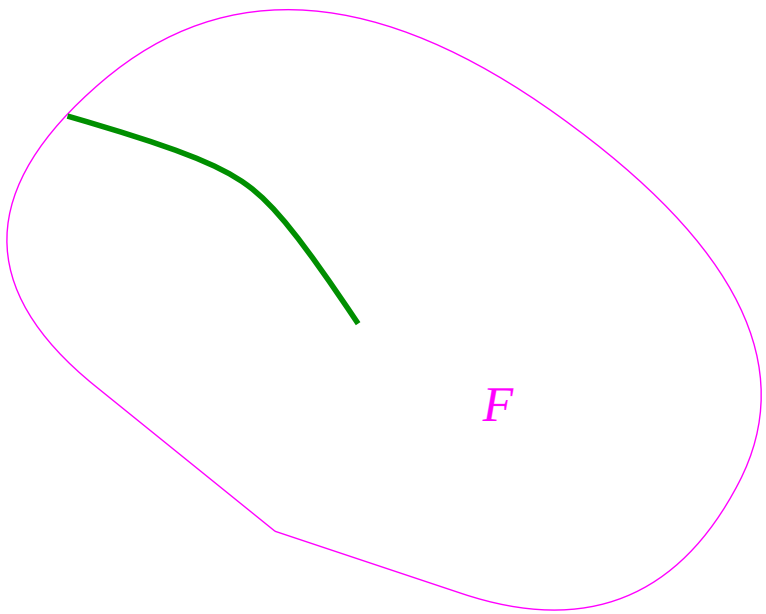
Theorem. *If $F^0(PSDP) \neq \emptyset$, $F^0(DSDP) \neq \emptyset$ and $\{A_i\}_{i=1}^m$ are linearly independent, then the barrier trajectory is well defined for $\mu > 0$.*

The proof can be given by showing that for a positive μ, a solution to the above system is equivalent to optimality conditions to the problem

$$\underset{X \in F^0(PSDP)}{\text{minimize}} \quad \text{trace}(CX) - \mu \ln \det X.$$

The barrier trajectory

Given suitable conditions, the barrier trajectory leads to the optimal solution.



Interior methods follow the trajectory approximately.

Primal-dual interior method

Assume that X is feasible to $(PSDP)$ with $X \succ 0$, and assume that y, S are feasible to $(DSDP)$ with $S \succ 0$.

Let $\mu = \dfrac{\text{trace}(XS)}{n}$ be the estimate of the barrier parameter.

Let $\sigma \in [0, 1]$ denote the desired reduction of duality gap.

This means that we aim at solving

$$\sum_{i=1}^{m} A_i y_i + S = C,$$

$$\text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m,$$

$$XS = \sigma \mu I.$$

An iteration in a primal-dual interior method

- Given X^k, y^k and S^k such that $X^k \succ 0$ and $S^k \succ 0$;
- $\mu^k \leftarrow \dfrac{\text{trace}(X^k S^k)}{n}$, $\sigma^k \in [0, 1]$;
- Compute ΔX^k, Δy^k and ΔS^k;
- Compute steplength α^k;
- $X^{k+1} \leftarrow X^k + \alpha^k \Delta X^k$, $y^{k+1} \leftarrow y^k + \alpha^k \Delta y^k$, $S^{k+1} \leftarrow S^k + \alpha^k \Delta S^k$;

We require $X^{k+1} \succ 0$ and $S^{k+1} \succ 0$.

Methods differ in the choice of σ^k, choice of step and choice of steplength.

Straightforward Newton iterations

The last block of equations may be written as $XS - \sigma\mu I = 0$, which gives the Newton iterations

$$\begin{aligned}
\textstyle\sum_{i=1}^m A_i \Delta y_i + \Delta S &= C - \textstyle\sum_{i=1}^m A_i y_i - S, \\
\mathsf{trace}(A_i \Delta X) &= b_i - \mathsf{trace}(A_i X), \quad i = 1, \ldots, m, \\
X\Delta S + \Delta X S &= \sigma\mu I - XS,
\end{aligned}$$

or as $SX - \sigma\mu I = 0$, which gives

$$\begin{aligned}
\textstyle\sum_{i=1}^m A_i \Delta y_i + \Delta S &= C - \textstyle\sum_{i=1}^m A_i y_i - S, \\
\mathsf{trace}(A_i \Delta X) &= b_i - \mathsf{trace}(A_i X), \quad i = 1, \ldots, m, \\
S\Delta X + \Delta S X &= \sigma\mu I - SX.
\end{aligned}$$

In general, the solutions differ, and they are nonsymmetric.

Reduction of duality gap

Assume that ΔX, Δy, ΔS satisfy

$$\begin{aligned} \textstyle\sum_{i=1}^m A_i \Delta y_i + \Delta S &= 0, \\ \mathsf{trace}(A_i \Delta X) &= 0, \quad i = 1, \ldots, m, \\ X\Delta S + \Delta X S &= \sigma\mu I - XS. \end{aligned}$$

Then $\mathsf{trace}(\Delta X \Delta S) = -\sum_{i=1}^m \mathsf{trace}(A_i \Delta X)\Delta y_i = 0$.

Hence,

$$\begin{aligned} \mathsf{trace}((X + \alpha\Delta X)(S + \alpha\Delta S)) &= \mathsf{trace}(XS) + \alpha\,\mathsf{trace}(X\Delta S + \Delta X S) \\ &= (1 - \alpha(1 - \sigma))\,\mathsf{trace}(XS). \end{aligned}$$

Same result for both linearizations.

Strategies for choosing σ

The relation

$$\text{trace}((X + \alpha\Delta X)(S + \alpha\Delta S)) = (1 - \alpha(1 - \sigma))\,\text{trace}(XS)$$

implies that it is desirable to have σ small and α large. These goals are in general contradictory.

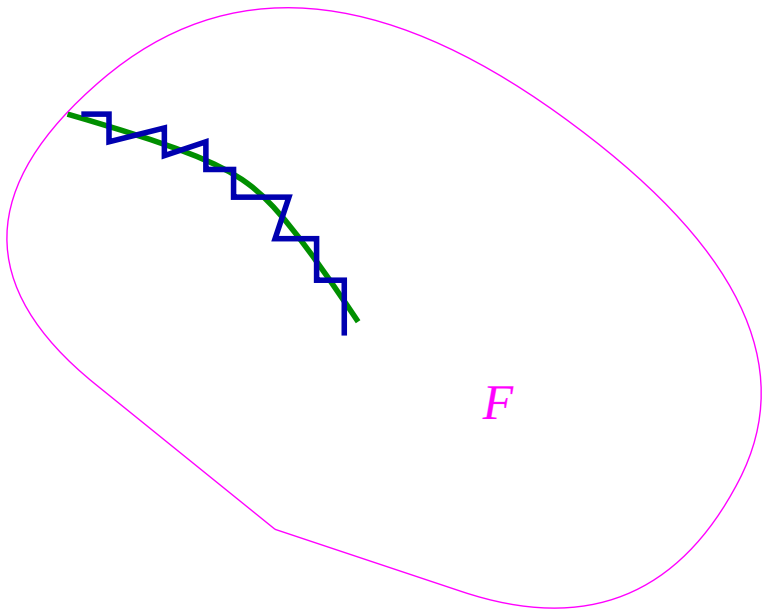
Three main strategies:

- Short-step method, σ close to 1.
- Long-step method, σ significantly smaller than 1.
- Predictor-corrector method, $\sigma = 0$ each even iteration and $\sigma = 1$ each odd iteration.

Short-step method

We may choose $\sigma^k = 1 - \delta/\sqrt{n}$, $\alpha^k = 1$.

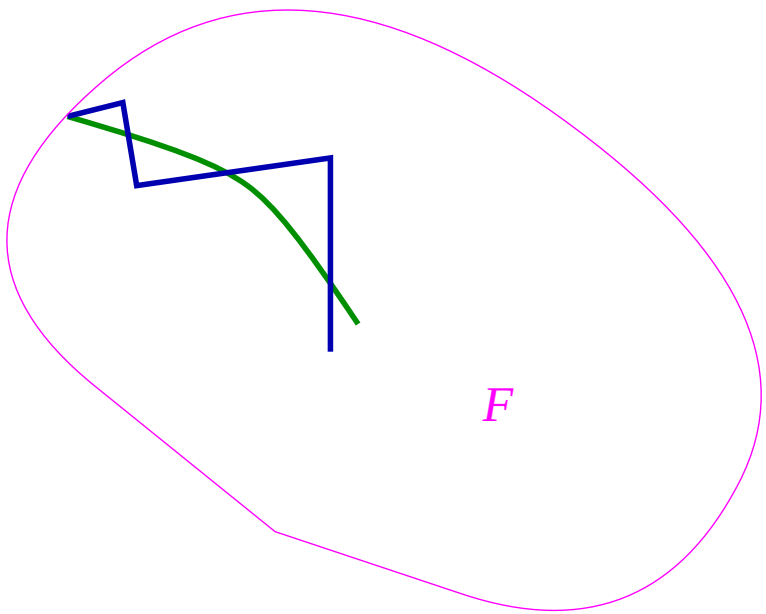
The iterates remain close to the trajectory.



Polynomial complexity for a given accuracy. In general not efficient enough.

Long-step method

We may choose $\sigma^k = 0.1$, α^k given by proximity to the trajectory.

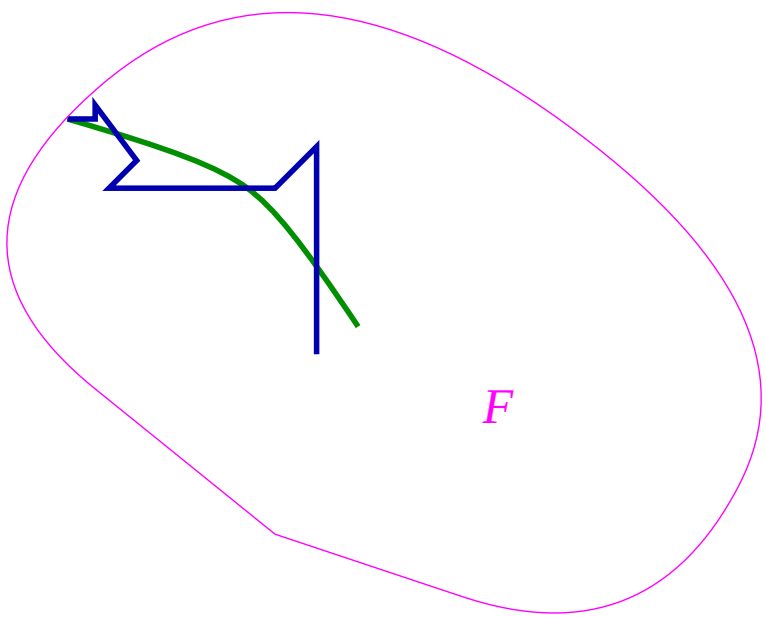


Polynomial complexity for a given accuracy.

Predictor-corrector method

$\sigma^k = 0$, α^k given by proximity to the trajectory for k even.

$\sigma^k = 1$, $\alpha^k = 1$ for k odd.



Polynomial complexity for a given accuracy.

Symmetrization of Newton equations

There are different ways of symmetrizing the last block of equations.

One example is the Alizadeh-Haeberly-Overton direction (AHO direction), where

$$\frac{1}{2}XS + \frac{1}{2}SX - \sigma\mu I = 0.$$

The corresponding block of Newton equations become

$$\frac{1}{2}X\Delta S + \frac{1}{2}\Delta XS + \frac{1}{2}S\Delta X + \frac{1}{2}\Delta SX = \sigma\mu I - \frac{1}{2}XS + \frac{1}{2}SX.$$

The solution ΔX, ΔS is symmetric.

More generally we may introduce $\mathcal{S}_P(U) = \frac{1}{2}\left(PUP^{-1} + (PUP^{-1})^T\right)$ for a nonsingular $n \times n$ matrix P and consider $\mathcal{S}_P(XS) = \sigma\mu I$.

Comments

- Polynomial complexity can be proved under the general framework of self-concordance, considering $-\ln\det(X)$. Define a *distance measure* to the trajectory.
- Note that the dimension may become a problem. If $X \in \mathcal{S}^n$, then the number of variables is $n(n+1)/2$. The polynomial complexity may be with a rather high exponent.
- Many interesting applications in areas such as systems and control, signal processing and combinatorial optimization.