Nonlinear optimization

Anders Forsgren

Optimization and Systems Theory
Department of Mathematics
Royal Institute of Technology (KTH)
Stockholm, Sweden

eVITA Winter School 2009
Geilo, Norway
January 11–15, 2009

Outline

1. Background on nonlinear optimization
2. Linear programming
3. Quadratic programming
4. General nonlinear programming
5. Applications

Nonlinear optimization

A nonlinear optimization problem takes the form

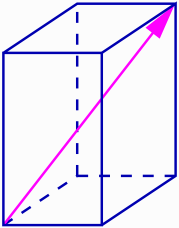
$$\begin{array}{lll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) & \\ \text{subject to} & g_i(x) \geq 0, & i \in \mathcal{I}, \\ & g_i(x) = 0, & i \in \mathcal{E}, \end{array} \qquad \begin{array}{l} \mathcal{I} \bigcup \mathcal{E} = \{1, \dots, m\}, \\ \mathcal{I} \bigcap \mathcal{E} = \emptyset. \end{array}$$

where f and g_i, $i = 1, \dots, m$, are nonlinear smooth functions from $\mathbb{R}^n$ to $\mathbb{R}$.

The feasible region is denoted by F. In our case

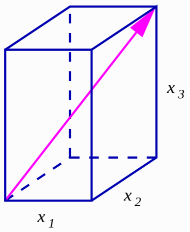
$$F = \{x \in \mathbb{R}^n : g_i(x) \geq 0,\ i \in \mathcal{I},\ g_i(x) = 0,\ i \in \mathcal{E}\}.$$

Example problem



Construct a box of volume 1 m^3 so that the space diagonal is minimized. What does it look like?

Formulation of example problem



- Introduce variables x_i, $i = 1, \ldots 3$. We obtain

$$(P) \qquad \begin{array}{ll} \underset{x \in \mathbb{R}^3}{\text{minimize}} & x_1^2 + x_2^2 + x_3^2 \\ \text{subject to} & x_1 \cdot x_2 \cdot x_3 = 1, \\ & x_i \geq 0, \quad i = 1, 2, 3. \end{array}$$

- The problem is not convex.

Alternative formulation of example problem

- We have the formulation

$$(P) \qquad \begin{array}{ll} \underset{x \in \mathbb{R}^3}{\text{minimize}} & x_1^2 + x_2^2 + x_3^2 \\ \text{subject to} & x_1 \cdot x_2 \cdot x_3 = 1, \\ & x_i \geq 0, \quad i = 1, 2, 3. \end{array}$$

- Replace $x_i \geq 0$, $i = 1, \dots, 3$ by $x_i > 0$, $i = 1, \dots, 3$.
- Let $y_i = \ln x_i$, $i = 1, 2, 3$, which gives

$$(P') \qquad \begin{array}{ll} \underset{y \in \mathbb{R}^3}{\text{minimize}} & e^{2y_1} + e^{2y_2} + e^{2y_3} \\ \text{subject to} & y_1 + y_2 + y_3 = 0. \end{array}$$

- This problem is convex.
- Is this a simpler problem to solve?

Applications of nonlinear optimization

- Nonlinear optimization arises in a wide range of areas.
- Two application areas will be menioned in this talk:
  - Radiation therapy.
  - Telecommunications.
- The optimization problems are often very large.
- Problem structure is highly important.

Problem classes in nonlinear optimization

Important problem classes in nonlinear optimization:

- Linear programming.
- Quadratic programming.
- General nonlinear programming.
- ...

Some comments:

- Convexity is a very useful poperty.
- Nonlinear (nonconvex) constraints cause increased difficulty.

Convex program

Proposition

Let $F = \{x \in \mathbb{R}^n : g_i(x) \ge 0,\ i \in \mathcal{I},\ g_i(x) = 0,\ i \in \mathcal{E}\}$. Then F is a convex set if g_i, $i \in \mathcal{I}$, are concave functions on $\mathbb{R}^n$ and g_i, $i \in \mathcal{E}$, are affine functions on $\mathbb{R}^n$.

We refer to the problem

$$
\begin{array}{lll}
\text{minimize} & f(x) & \\
\text{subject to} & g_i(x) \ge 0, \quad i \in \mathcal{I}, & \qquad \mathcal{I} \bigcup \mathcal{E} = \{1, \dots, m\}, \\
 & g_i(x) = 0, \quad i \in \mathcal{E}, & \qquad \mathcal{I} \bigcap \mathcal{E} = \emptyset, \\
 & x \in \mathbb{R}^n, &
\end{array}
$$

as a convex program if f and $-g_i$, $i \in \mathcal{I}$, are convex functions on $\mathbb{R}^n$, and $-g_i$, $i \in \mathcal{I}$, are affine functions on $\mathbb{R}^n$.

Optimality conditions for nonlinear programs

Consider a nonlinear program

$$(P) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in F \subseteq \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2.$$

Definition

A direction p is a feasible direction to F at x^* if there is an $\bar{\alpha} > 0$ such $x^* + \alpha p \in F$ for $\alpha \in [0, \bar{\alpha}]$.

Definition

A direction p is a descent direction to f at x^* if $\nabla f(x^*)^T p < 0$.

Definition

A direction p is a direction of negative curvature to f at x^* if $p^T \nabla^2 f(x^*)^T p < 0$.

Optimality conditions for unconstrained problems

Consider an unconstrained problem

$$(P) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2.$$

Theorem (First-order necessary optimality conditions)

If x^* is a local minimizer to (P) then $\nabla f(x^*) = 0$.

Theorem (Second-order necessary optimality conditions)

If x^* is a local minimizer to (P) then $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succeq 0$.

Theorem (Second-order sufficient optimality conditions)

If $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succ 0$ then x^* is a local minimizer to (P).

Optimality conditions, linear equality constraints

Consider an equality-constrained problem

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \; x \in \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2, A \text{ full row rank.}$$

Let $F = \{x \in \mathbb{R}^n : Ax = b\}$. Assume that $\bar{x}$ is a known point in F, and let x be an arbitrary point in F. Then, $A(x - \bar{x}) = 0$, i.e. $x - \bar{x} \in \text{null}(A)$.

If Z denotes a matrix whose columns form a basis for null(A), it means that $x - \bar{x} = Zv$ for some $v \in \mathbb{R}^{n-m}$.

For example, if $A = (B \; N)$, where B is $m \times m$ and invertible, we may choose $\bar{x} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$ and $Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}$.

Optimality conditions, linear equality constraints, cont.

Let $\varphi(v) = f(\bar{x} + Zv)$. We may rewrite the problem according to

$$(P'_{=}) \quad \begin{array}{ll} \text{minimize} & \varphi(v) \\ \text{subject to} & v \in \mathbb{R}^{n-m}. \end{array}$$

Differentiation gives $\nabla\varphi(v) = Z^T \nabla f(\bar{x} + Zv)$,
$\nabla^2\varphi(v) = Z^T \nabla^2 f(\bar{x} + Zv) Z$.

This is an unconstrained problem, where we know the optimality conditions.

We may apply them and identify $x^* = \bar{x} + Zv^*$, where v^* is associated with $(P'_{=})$.

$Z^T \nabla f(x)$ is called the reduced gradient of f at x.

$Z^T \nabla^2 f(x) Z$ is called the reduced Hessian of f at x.

Necessary optimality conditions, equality constraints

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \;\; x \in \mathbb{R}^n, \end{array}$$ where $f \in C^2$, A full row rank.

Theorem (First-order necessary optimality conditions)

If x^* is a local minimizer to $(P_=)$, then

(i) $Ax^* = b$, *and*

(ii) $Z^T \nabla f(x^*) = 0$.

Theorem (Second-order necessary optimality conditions)

If x^* is a local minimizer to $(P_=)$, then

(i) $Ax^* = b$,

(ii) $Z^T \nabla f(x^*) = 0$, *and*

(iii) $Z^T \nabla^2 f(x^*) Z \succeq 0$.

Sufficient optimality conditions, equality constraints

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \; x \in \mathbb{R}^n, \end{array}$$ where $f \in C^2$, A full row rank.

Theorem (Second-order sufficient optimality conditions)

If

(i) $Ax^* = b$,

(ii) $Z^T \nabla f(x^*) = 0$, *and*

(iii) $Z^T \nabla^2 f(x^*) Z \succ 0$,

then x^* *is a local minimizer to* $(P_=)$.

Lagrange multipliers

Proposition

Let $A \in \mathbb{R}^{m \times n}$. The null space of A and the range space of A^T are orthogonal spaces that together span R^n.

- We have $Z^T c = 0 \iff c = A^T \lambda$ for some λ.
- In particular, let $c = \nabla f(x^*)$.
- We have $Z^T \nabla f(x^*) = 0$ if and only if $\nabla f(x^*) = A^T \lambda^*$ for some $\lambda^* \in \mathbb{R}^m$.
- We call λ^* Lagrange multiplier vector.

Necessary optimality conditions, equality constraints

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b,\ x \in \mathbb{R}^n, \end{array}$$ where $f \in C^2$, A full row rank.

Theorem (First-order necessary optimality conditions)

If x^* is a local minimizer to $(P_=)$, then
- *(i)* $Ax^* = b$, *and*
- *(ii)* $\nabla f(x^*) = A^T\lambda^*$ *for some* $\lambda^* \in \mathbb{R}^m$.

Theorem (Second-order necessary optimality conditions)

If x^* is a local minimizer to $(P_=)$, then
- *(i)* $Ax^* = b$,
- *(ii)* $\nabla f(x^*) = A^T\lambda^*$ *for some* $\lambda^* \in \mathbb{R}^m$, *and*
- *(iii)* $Z^T\nabla^2 f(x^*)Z \succeq 0$.

Sufficient optimality conditions, equality constraints

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \ x \in \mathbb{R}^n, \end{array}$$ where $f \in C^2$, A full row rank.

Theorem (Second-order sufficient optimality conditions)

If
- *(i)* $Ax^* = b$,
- *(ii)* $\nabla f(x^*) = A^T\lambda^*$ *for some* $\lambda^* \in \mathbb{R}^m$, *and*
- *(iii)* $Z^T\nabla^2 f(x^*)Z \succ 0$,

then x^* *is a local minimizer to* $(P_=)$.

Optimality conditions, linear equality constraints, cont.

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \; x \in \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2, \ A \text{ full row rank.}$$

If we define the Lagrangian $\mathcal{L}(x, \lambda) = f(x) - \lambda^T(Ax - b)$, the first-order optimality conditions are equivalent to

$$\begin{pmatrix} \nabla_x \mathcal{L}(x^*, \lambda^*) \\ \nabla_\lambda \mathcal{L}(x^*, \lambda^*) \end{pmatrix} = \begin{pmatrix} \nabla f(x^*) - A^T \lambda^* \\ b - Ax^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Alternatively, the requirement is $Ax^* = b$ where the problem

$$\begin{array}{ll} \text{minimize} & \nabla f(x^*)^T p \\ \text{subject to} & Ap = 0, \; p \in \mathbb{R}^n, \end{array}$$

has optimal value zero.

Optimality conditions, linear inequality constraints

Assume that we have inequality constraints according to

$$(P_{\geq}) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \geq b, \ x \in \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2.$$

Consider a feasible point x^*. Partition $A = \begin{pmatrix} A_A \\ A_I \end{pmatrix}$,

$b = \begin{pmatrix} b_A \\ b_I \end{pmatrix}$, where $A_A x^* = b_A$ and $A_I x^* > b_I$.

The constraints $A_A x \geq b_A$ are active at x^*.

The constraints $A_I x \geq b_I$ are inactive at x^*.

Optimality conditions, linear inequality constraints

If x^* is a local minimizer to $(P_{\geq})$ there must not exist a feasible descent direction in x^*. Thus the problems

$$\begin{array}{ll} \text{minimize} & \nabla f(x^*)^T p \\ \text{subject to} & A_A p \geq 0, \end{array} \qquad \begin{array}{ll} \text{maximize} & 0^T \lambda_A \\ \text{subject to} & A_A^T \lambda_A = \nabla f(x^*),\ \lambda_A \geq 0, \end{array}$$

must have optimal value zero. (The second problem is the LP-dual of the first one.) Consequently, there is $\lambda_A^* \geq 0$ such that $A_A^T \lambda_A^* = \nabla f(x^*)$.

Necessary optimality conditions, linear ineq. cons.

$$(P_{\geq}) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \geq b, \ x \in \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2.$$

Theorem (First-order necessary optimality conditions)

If x^* is a local minimizer to $(P_{\geq})$ it holds that

(i) $Ax^* \geq b$, *and*

(ii) $\nabla f(x^*) = A_A^T \lambda_A^*$ *for some* $\lambda_A^* \geq 0$,

where A_A is associated with the active constraints at x^*.

The first-order necessary optimality conditions are often referred to as the KKT conditions.

Necessary optimality conditions

$$(P_{\geq}) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \geq b, \ x \in \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2.$$

Theorem (Second-order necessary optimality conditions)

If x^* is a local minimizer to $(P_{\geq})$ it holds that

(i) $Ax^* \geq b$, *and*

(ii) $\nabla f(x^*) = A_A^T \lambda_A^*$ *for some* $\lambda_A^* \geq 0$,

(iii) $Z_A^T \nabla^2 f(x^*) Z_A \succeq 0$,

where A_A is associated with the active constraints at x^* and Z_A is a matrix whose columns form a basis for $\text{null}(A_A)$.

Condition *(iii)* corresponds to replacing $Ax \geq b$ by $A_A x = b_A$.

Sufficient optimality conditions for linear constraints

$$(P_{\geq}) \qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \geq b, \; x \in \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2.$$

Theorem (Second-order sufficient optimality conditions)

If

(i) $Ax^* \geq b$,

(ii) $\nabla f(x^*) = A_A^T \lambda_A^*$ *for some* $\lambda_A^* > 0$, *and*

(iii) $Z_A^T \nabla^2 f(x^*) Z_A \succ 0$,

then x^* *is a local minimizer to* $(P_{\geq})$, *where* A_A *is associated with the active constraints at* x^* *and* Z_A *is a matrix whose columns form a basis for* $\text{null}(A_A)$.

(Slightly more complicated if $\lambda_A^* \geq 0$, $\lambda_A^* \not> 0$.)

Necessary optimality conditions

$$(P_{\geq}) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \geq b, \ x \in \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2.$$

The first-order necessary optimality conditions are often stated with an m-dimensional Lagrange-multiplier vector λ^*.

Theorem (First-order necessary optimality conditions)

If x^* is a local minimizer to $(P_{\geq})$ then x^* and some $\lambda^* \in \mathbb{R}^m$ satisfy

(i) $Ax^* \geq b$,

(ii) $\nabla f(x^*) = A^T \lambda^*$,

(iii) $\lambda^* \geq 0$, *and*

(iv) $\lambda_i^*(a_i^T x^* - b_i) = 0, \ i = 1, \dots, m.$

Necessary optimality conditions, linear constraints

$$(P) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & a_i^T x \geq b_i, \; i \in \mathcal{I}, \\ & a_i^T x = b_i, \; i \in \mathcal{E}, \\ & x \in \mathbb{R}^n, \end{array} \quad \text{where } f \in C^2.$$

Theorem (First-order necessary optimality conditions)

If x^* is a local minimizer to $(P_{\geq})$ then x^* and some $\lambda^* \in \mathbb{R}^m$ satisfy

(i) $a_i^T x^* \geq b_i$, $i \in \mathcal{I}$, $a_i^T x^* = b_i$, $i \in \mathcal{E}$,

(ii) $\nabla f(x^*) = A^T \lambda^*$,

(iii) $\lambda_i^* \geq 0$, $i \in \mathcal{I}$, *and*

(iv) $\lambda_i^* (a_i^T x^* - b_i) = 0$, $i \in \mathcal{I}$.

Optimality conditions for nonlinear equality constraints

Consider an equality-constrained nonlinear program

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) = 0, \end{array} \quad \text{where } f, g \in C^2,\ g : \mathbb{R}^n \to \mathbb{R}^m.$$

Let $A(x) = \begin{pmatrix} \nabla g_1(x)^T \\ \vdots \\ \nabla g_m(x)^T \end{pmatrix}$.

The linearization of the constraints has to be "sufficiently good" at x^* to get optimality conditions analogous to those for linear constraints.

Definition (Regularity for equality constraints)

A point $x^* \in F$ is regular to $(P_=)$ if $A(x^*)$ has full row rank, i.e., if $\nabla g_i(x^*)$, $i = 1, \ldots, m$, are linearly independent.

Regularity allows generalization to nonlinear constraints.

Necessary optimality conditions, equality constraints

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) = 0, \end{array} \quad \text{where } f, g \in C^2,\ g : \mathbb{R}^n \to \mathbb{R}^m.$$

Theorem (First-order necessary optimality conditions)

If x^* is a regular point and a local minimizer to $(P_=)$, then

(i) $g(x^*) = 0$, *and*

(ii) $\nabla f(x^*) = A(x^*)^T \lambda^*$ *for some* $\lambda^* \in \mathbb{R}^m$.

Necessary optimality conditions, equality constraints

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) = 0, \end{array} \quad \text{where } f, g \in C^2,\ g : \mathbb{R}^n \to \mathbb{R}^m.$$

Theorem (Second-order necessary optimality conditions)

If x^* is a regular point and a local minimizer to $(P_=)$, then

(i) $g(x^*) = 0$, *and*

(ii) $\nabla f(x^*) = A(x^*)^T \lambda^*$ *for some* $\lambda^* \in \mathbb{R}^m$, *and*

(iii) $Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*) \succeq 0$.

Note that (*iii*) involves the Lagrangian $\mathcal{L}(x, \lambda) = f(x) - \lambda^T g(x)$, not the objective function.

Sufficient optimality conditions, equality constraints

$$(P_=) \qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) = 0, \end{array} \quad \text{where } f, g \in C^2,\ g : \mathbb{R}^n \to \mathbb{R}^m.$$

Theorem (Second-order sufficient optimality conditions)

If

(i) $g(x^*) = 0$,

(ii) $\nabla f(x^*) = A(x^*)^T \lambda^*$ *for some* $\lambda^* \in \mathbb{R}^m$, *and*

(iii) $Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*) \succ 0$,

then x^* *is a local minimizer to* $(P_=)$.

Necessary optimality conditions, inequality constraints

Assume that we have an inequality-constrained problem

$$(P_{\geq}) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \geq 0, \; x \in \mathbb{R}^n, \end{array} \quad \text{where } f, g \in C^2, \; g : \mathbb{R}^n \to \mathbb{R}^m.$$

Consider a feasible point x^*. Partition $g(x^*) = \begin{pmatrix} g_A(x^*) \\ g_I(x^*) \end{pmatrix}$,

where $g_A(x^*) = 0$ and $g_I(x^*) > 0$. Partition $A(x^*)$ analogously.

Definition (Regularity for inequality constraints)

A point $x^* \in \mathbb{R}^n$ which is feasible to $(P_{\geq})$ is regular to $(P_{\geq})$ if $A_A(x^*)$ has full row rank, i.e., if $\nabla g_i(x^*)$, $i \in \{l : g_l(x^*) = 0\}$, are linearly independent.

Necessary optimality conditions, inequality constraints

$$(P_{\geq}) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \geq 0, \ x \in \mathbb{R}^n, \end{array} \quad \text{where } f, g \in C^2, \ g : \mathbb{R}^n \to \mathbb{R}^m.$$

Theorem (First-order necessary optimality conditions)

If x^* is a regular point and a local minimizer to $(P_{\geq})$, then

(i) $g(x^*) \geq 0$, *and*

(ii) $\nabla f(x^*) = A_A(x^*)^T \lambda_A^*$ *for some* $\lambda_A^* \geq 0$.

where $A_A(x^*)$ corresponds to the active constraints at x^*.

Necessary optimality conditions, inequality constraints

$$(P_{\geq}) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \geq 0, \ x \in \mathbb{R}^n, \end{array}$$ where $f, g \in C^2$, $g : \mathbb{R}^n \to \mathbb{R}^m$.

Theorem (Second-order necessary optimality conditions)

If x^* is a regular point and a local minimizer to $(P_{\geq})$, then

(i) $g(x^*) \geq 0$, *and*

(ii) $\nabla f(x^*) = A_A(x^*)^T \lambda_A^*$ *for some* $\lambda_A^* \geq 0$, *and*

(iii) $Z_A(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z_A(x^*) \succeq 0$,

where $A_A(x^*)$ corresponds to the active constraints at x^* and $Z_A(x^*)$ is a matrix whose columns form a basis for $\text{null}(A_A(x^*))$.

Condition *(iii)* corresponds to replacing $g(x) \geq 0$ with $g_A(x) = 0$.

Sufficient optimality conditions, inequality constraints

$$(P_{\geq}) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \geq 0, \; x \in \mathbb{R}^n, \end{array}$$ where $f, g \in C^2$, $g : \mathbb{R}^n \to \mathbb{R}^m$.

Theorem (Second-order sufficient optimality conditions)

If

(i) $g(x^*) \geq 0$,

(ii) $\nabla f(x^*) = A_A(x^*)^T \lambda_A^*$ *for some* $\lambda_A^* > 0$, *and*

(iii) $Z_A(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z_A(x^*) \succ 0$,

then x^* *is a local minimizer to* $(P_{\geq})$, *where* $A_A(x^*)$ *corresponds to the active constraints at* x^*, *and* $Z_A(x^*)$ *is a matrix whose columns form a basis for* $\text{null}(A_A(x^*))$.

(Slightly more complicated if $\lambda_A^* \geq 0$, $\lambda_A^* \not> 0$.)

First-order necessary optimality conditions

$$
(P) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \geq 0, \ i \in \mathcal{I}, \\ & g_i(x) = 0, \ i \in \mathcal{E}, \\ & x \in \mathbb{R}^n, \end{array} \quad \text{where } f, g \in C^2,\ g : \mathbb{R}^n \to \mathbb{R}^m.
$$

Theorem (First-order necessary optimality conditions)

If x^* is a regular point and a local minimizer to (P), there is a $\lambda^* \in \mathbb{R}^m$ such that x^* and λ^* satisfy

(i) $g_i(x^*) \geq 0,\ i \in \mathcal{I},\ \ g_i(x^*) = 0,\ i \in \mathcal{E}$,

(ii) $\nabla f(x^*) = A(x^*)^T \lambda^*$,

(iii) $\lambda_i^* \geq 0,\ i \in \mathcal{I}$, *and*

(iv) $\lambda_i^* g_i(x^*) = 0,\ \ i \in \mathcal{I}$.

Convexity gives global optimality

$$(P) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \geq 0,\ i \in \mathcal{I}, \quad g_i(x) = 0,\ i \in \mathcal{E}, \quad x \in \mathbb{R}^n, \end{array}$$

where $f, g \in C^2$, $g : \mathbb{R}^n \to \mathbb{R}^m$.

Theorem

Assume that g_i, $i \in \mathcal{I}$, are concave functions on $\mathbb{R}^n$ and g_i, $i \in \mathcal{E}$, are affine functions on $\mathbb{R}^n$. Assume that f is a convex function on the feasible region of (P). If $x^* \in \mathbb{R}^m$ and $\lambda^* \in \mathbb{R}^m$ satisfy

(i) $g_i(x^*) \geq 0$, $i \in \mathcal{I}$, $g_i(x^*) = 0$, $i \in \mathcal{E}$,

(ii) $\nabla f(x^*) = A(x^*)^T \lambda^*$,

(iii) $\lambda_i^* \geq 0$, $i \in \mathcal{I}$, *and*

(iv) $\lambda_i^* g_i(x^*) = 0$, $i \in \mathcal{I}$,

then x^* is a global minimizer to (P).

Nonlinear programming is a wide problem class

Consider a binary program (IP) in the form

$$(IP) \qquad \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \geq b, \\ & x_j \in \{0,1\}, \quad j = 1,\dots,n. \end{array}$$

This problem is NP-hard. (Difficult.)

An equivalent formulation of (IP) is

$$(NLP) \qquad \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \geq b, \\ & x_j(1 - x_j) = 0, \quad j = 1,\dots,n. \end{array}$$

To find a global minimizer of (NLP) is equally hard.

Linear program

A linear program is a convex optimization problem on the form

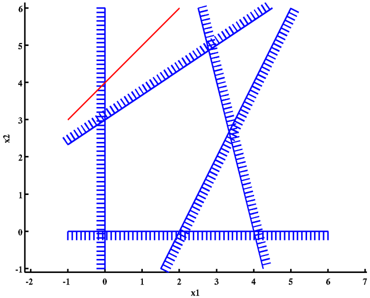
$$(LP) \qquad \begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b, \\ & x \geq 0. \end{array}$$

May be written on many (equivalent) forms.

The feasible set is a polyhedron, i.e., given by the intersection of a finite number of hyperplanes in $\mathbb{R}^n$.

Example linear program

$$\begin{array}{ll} \min & -x_1 + x_2 \\ \text{subject to} & -2x_1 + x_2 \geq -4, \\ & 2x_1 - 3x_2 \geq -9, \\ & -4x_1 - x_2 \geq -16, \\ & x_1 \geq 0, \\ & x_2 \geq 0. \end{array}$$



Example linear program, cont.

Equivalent linear programs.

$$\begin{array}{ll} \text{minimize} & -x_1 + x_2 \\ \text{subject to} & -2x_1 + x_2 \geq -4, \\ & 2x_1 - 3x_2 \geq -9, \\ & -4x_1 - x_2 \geq -16, \\ & x_1 \geq 0, \\ & x_2 \geq 0. \end{array}$$

$$\begin{array}{ll} \text{minimize} & -x_1 + x_2 \\ \text{subject to} & -2x_1 + x_2 - x_3 = -4, \\ & 2x_1 - 3x_2 - x_4 = -9, \\ & -4x_1 - x_2 - x_5 = -16, \\ & x_j \geq 0, \quad j = 1, \dots, 5. \end{array}$$

Methods for linear programming

We will consider two type of methods for linear programming.

- The simplex method.
  - Combinatoric in its nature.
  - The iterates are extreme points of the feasible region.
- Interior methods.
  - Approximately follow a trajectory created by a perturbation of the optimality conditions.
  - The iterates belong to the relative interior of the feasible region.

Linear program and extreme points

Definition

Let S be a convex set. Then x is an extreme point to S if $x \in S$ and there are no $y \in C$, $z \in C$, $y \neq x$, $z \neq x$, and $\alpha \in (0, 1)$ such that $x = (1 - \alpha)y + \alpha z$.

$$(LP) \qquad \begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b, \\ & x \geq 0. \end{array}$$

Theorem

Assume that (LP) *has at least one optimal solution. Then, there is an optimal solution which is an extreme point.*

One way of solving a linear program is to move from extreme point to extreme point, requiring decrease in the objective function value. (The simplex method.)

Linear program extreme points

Proposition

Let $S = \{x \in \mathbb{R}^n : Ax = b \text{ where } A \in \mathbb{R}^{m \times n} \text{ of rank } m\}$.
Then, if x is an extreme point of S, we may partition $A = (B\ N)$ (column permuted), where B is $m \times m$ and invertible, and x conformally, such that

$$\begin{pmatrix} B & N \\ 0 & I \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}, \text{ with } x_B \geq 0.$$

Note that $x_B = B^{-1}b$, $x_N = 0$.

We refer to B as a basis matrix.

Extreme points are referred to as basic feasible solutions.

Optimality of basic feasible solution

Assume that we have a basic feasible solution

$$\begin{pmatrix} B & N \\ 0 & I \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}.$$

Proposition

The basic feasible solution is optimal if $c^T p^i \geq 0$, $i = 1, \ldots, n-m$, where p^i is given by

$$\begin{pmatrix} B & N \\ 0 & I \end{pmatrix} \begin{pmatrix} p^i_B \\ p^i_N \end{pmatrix} = \begin{pmatrix} 0 \\ e_i \end{pmatrix}, \quad i = 1, \ldots, n-m.$$

Proof.

If $\widetilde{x}$ is feasible, it must hold that $\widetilde{x} - x = \sum_{i=1}^{n-m} \gamma_i p^i$, where $\gamma_i \geq 0$, $i = 1, \ldots, n-m$. Hence, $c^T(\widetilde{x} - x) \geq 0$. □

Test of optimality of basic feasible solution

Note that $c^T p^i$ may be written as

$$c^T p^i = \begin{pmatrix} c_B^T & c_N^T \end{pmatrix} \begin{pmatrix} B & N \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ e_i \end{pmatrix}.$$

Let y and s_N solve $\begin{pmatrix} B^T & 0 \\ N^T & I \end{pmatrix} \begin{pmatrix} y \\ s_N \end{pmatrix} = \begin{pmatrix} c_B \\ c_N \end{pmatrix}$.

Then $c^T p^i = \begin{pmatrix} y^T & s_N^T \end{pmatrix} \begin{pmatrix} 0 \\ e_i \end{pmatrix} = (s_N)_i$.

We may compute $c^T p^i$, $i = 1, \ldots, n - m$, by solving one system of equations.

An iteration in the simplex method

- Compute simplex mulipliers y and reduced costs s from
$$\begin{pmatrix} B^T & 0 \\ N^T & I \end{pmatrix} \begin{pmatrix} y \\ s_N \end{pmatrix} = \begin{pmatrix} c_B \\ c_N \end{pmatrix}.$$
- If $(s_N)_t < 0$, compute search direction p from
$$\begin{pmatrix} B & N \\ 0 & I \end{pmatrix} \begin{pmatrix} p_B \\ p_N \end{pmatrix} = \begin{pmatrix} 0 \\ e_t \end{pmatrix}.$$
- Compute maximum steplength $\alpha_{\max}$ and limiting constraint r from
$$\alpha_{\max} = \min_{i:(p_B)_i<0} \frac{(x_B)_i}{-(p_B)_i}, \quad r = \operatorname*{argmin}_{i:(p_B)_i<0} \frac{(x_B)_i}{-(p_B)_i}.$$
- Let $x = x + \alpha_{\max} p$.
- Replace $(x_N)_t = 0$ by $(x_B)_r = 0$ among the active constraints.

An iteration in the simplex method, alternatively

- Compute simplex mulipliers y and reduced costs s from

$$B^T y = c_B, \quad s_N = c_N - N^T y.$$

- If $(s_N)_t < 0$, compute search direction p from

$$p_N = e_t, \quad Bp_B = -N_t.$$

- Compute maximum steplength $\alpha_{\max}$ and limiting constraint r from

$$\alpha_{\max} = \min_{i:(p_B)_i<0} \frac{(x_B)_i}{-(p_B)_i}, \quad r = \operatorname{argmin}_{i:(p_B)_i<0} \frac{(x_B)_i}{-(p_B)_i}.$$

- Let $x = x + \alpha_{\max} p$.
- Replace $(x_N)_t = 0$ by $(x_B)_r = 0$ among the active constraints.

Optimality conditions for linear programming

We want to solve the linear program

$$(LP) \quad \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \\ & x \geq 0. \end{array}$$

Proposition

A vector $x \in \mathbb{R}^n$ is optimal to (LP) if and only if there are $y \in \mathbb{R}^m$, $s \in \mathbb{R}^n$ such that

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

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

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

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:

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

where $X = \operatorname{diag}(x)$, $S = \operatorname{diag}(s)$ 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

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

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

Optimality conditions for (P_μ):

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

Dual point of view

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

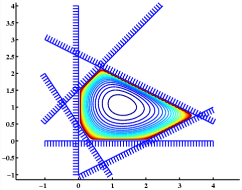
$$(D_\mu) \qquad \begin{array}{ll} \text{maximize} & b^T y + \mu \sum_{j=1}^{n} \ln s_j \\ \text{subject to} & A^T y + s = c, \quad s > 0, \end{array}$$

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

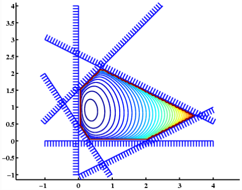
Optimality conditions for (D_μ):

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

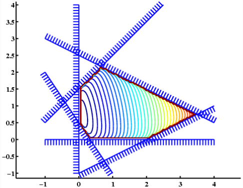
Primal barrier function for example linear program



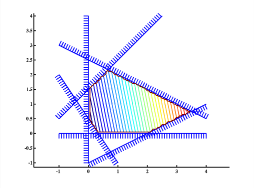
$\mu = 5$



$\mu = 1$



$\mu = 0.3$



$\mu = 10^{-16}$

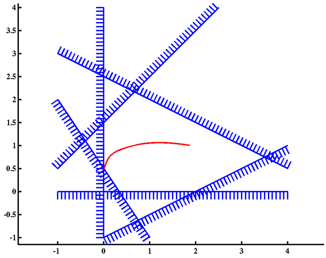
The barrier trajectory

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*

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

where $\mathcal{B} = \{i : \tilde{x}_i > 0$ *for some optimal solution* $\tilde{x}$ *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

1. Choose μ.
2. Compute Δx, Δy and Δs from

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

3. Find maximum steplength $\alpha_{\max}$ from $x + \alpha \Delta x \geq 0$, $s + \alpha \Delta s \geq 0$.
4. Let $\alpha = \min\{1, 0.999 \cdot \alpha_{\max}\}$.
5. 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.

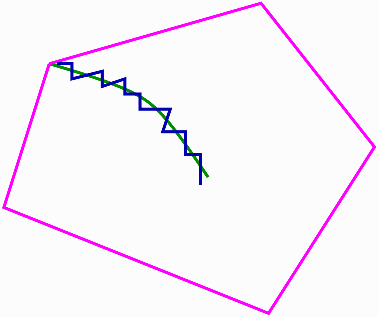
Three main strategies:

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

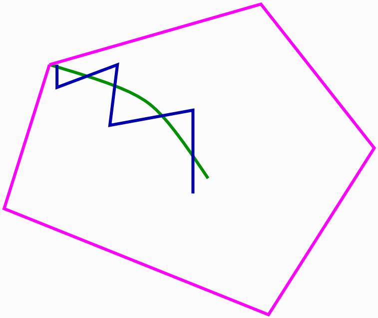
The iterates remain close to the trajectory.



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.

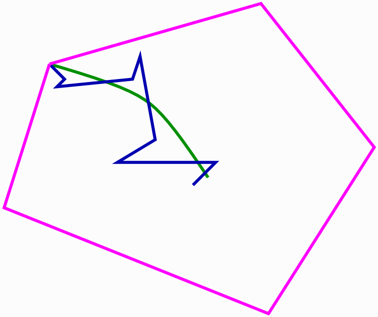


Polynomial complexity.

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.

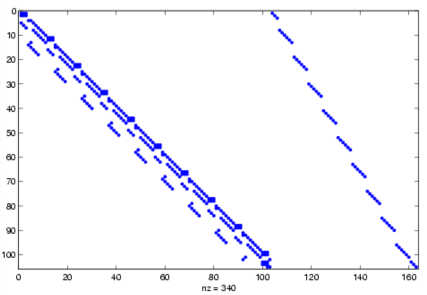


Polynomial complexity.

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

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

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

Quadratic programming with equality constraints

Look at model problem with quadratic objective function,

$$(EQP) \quad \begin{array}{ll} \text{minimize} & f(x) = \frac{1}{2}x^T H x + c^T x \\ \text{subject to} & Ax = b, \\ & x \in \mathbb{R}^n. \end{array}$$

We assume that $A \in \mathbb{R}^{m \times n}$ with rank m.

The first-order optimality conditions become

$$\begin{aligned} Hx + c &= A^T \lambda, \\ Ax &= b. \end{aligned}$$

This is a system of linear equations.

Optimality conditions, quadratic program

The first-order necessary optimality conditions may be written

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ -\lambda \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}.$$

Let Z be a matrix whose columns form a basis for null(A).

Proposition

A point $x^* \in \mathbb{R}^n$ is a global minimizer to (EQP) if and only if there exists a $\lambda^* \in \mathbb{R}^m$ such that

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ -\lambda^* \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix} \quad \text{and} \quad Z^T H Z \succeq 0.$$

Quadratic programming with equality constraints

Alternatively, let x be a given point and p the step to optimum,

$$(EQP') \quad \begin{array}{ll} \text{minimize} & f(x+p) = \frac{1}{2}(x+p)^T H(x+p) + c^T(x+p) \\ \text{subject to} & Ap = b - Ax, \\ & p \in \mathbb{R}^n. \end{array}$$

Proposition

A point $x + p^* \in \mathbb{R}^n$ is a global minimizer to (EQP) if and only if there is $\lambda^* \in \mathbb{R}^m$ such that

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p^* \\ -\lambda^* \end{pmatrix} = - \begin{pmatrix} Hx + c \\ Ax - b \end{pmatrix} \quad \textit{and} \quad Z^T H Z \succeq 0.$$

Note! Same λ^* as previously.

The KKT matrix

The matrix $K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}$ is called the KKT matrix.

Proposition

If $A \neq 0$, then $K \nsucceq 0$.

This means that K is an indefinite matrix.

Proposition

If $Z^T H Z \succ 0$ and $\operatorname{rank}(A) = m$ then K is nonsingular.

If $Z^T H Z \succ 0$ and $\operatorname{rank}(A) = m$ then x^* and λ^* are unique.

We assume that $Z^T H Z \succ 0$ and $\operatorname{rank}(A) = m$ for the equality-constrained case.

How do we compute x^* and λ^*?

We prefer (EQP') to (EQP).

Observation related to inequality constraints

Assume that $x^* = x + p^*$ and associated λ^* form optimal solution to

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^THx + c^Tx \\ \text{subject to} & Ax = b, \\ & x \in \mathbb{R}^n, \end{array}$$

where $H \succ 0$. If $\lambda^* \geq 0$ then x^* is also an optimal solution to

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^THx + c^Tx \\ \text{subject to} & Ax \geq b, \\ & x \in \mathbb{R}^n. \end{array}$$

This observation is the basis for an active-set method for solving inequality-constrained quadratic programs.

Inequality-constrained quadratic programming

Consider the inequality-constrained quadratic program

$$(IQP) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^T H x + c^T x \\ \text{subject to} & Ax \geq b, \\ & x \in \mathbb{R}^n. \end{array}$$

We assume that $H \succ 0$. The problem is then convex.

We have previously considered equality-constrained problems.

Now we must determine the active constraints at the solution.

We will consider two types of method:

- Active-set methods. ("Hard" choice.)
- Interior methods. ("Soft" choice.)

Background to active-set method

An active-set method generates feasible points.

Assume that we know a feasible point $\bar{x}$. (Solve LP.)

Guess that the constraints active at $\bar{x}$ are active at x^* too.

Let $\mathcal{A} = \{l : a_l^T \bar{x} = b_l\}$. The active constraints at $\bar{x}$.

Let $\mathcal{W} \subseteq \mathcal{A}$ be such that $A_{\mathcal{W}}$ has full row rank.

Keep (temporarily) the constraints in $\mathcal{W}$ active, i.e., solve

$$
(EQP_{\mathcal{W}}) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}(\bar{x}+p)^T H(\bar{x}+p) + c^T(\bar{x}+p) \\ \text{subject to} & A_{\mathcal{W}} p = 0, \\ & p \in \mathbb{R}^n. \end{array}
$$

Solution of equality-constrained subproblem

The problem

$$(EQP_{\mathcal{W}}) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}(\bar{x}+p)^T H(\bar{x}+p) + c^T(\bar{x}+p) \\ \text{subject to} & A_{\mathcal{W}} p = 0, \\ & p \in \mathbb{R}^n. \end{array}$$

has, from above, optimal solution p^* and associate multiplier vector $\lambda_{\mathcal{W}}^*$ given by

$$\begin{pmatrix} H & A_{\mathcal{W}}^T \\ A_{\mathcal{W}} & 0 \end{pmatrix} \begin{pmatrix} p^* \\ -\lambda_{\mathcal{W}}^* \end{pmatrix} = - \begin{pmatrix} H\bar{x} + c \\ 0 \end{pmatrix}.$$

Optimal x^* associated with $(EQP_{\mathcal{W}})$ is given by $x^* = \bar{x} + p^*$.

What have we ignored?

When solving ($EQP_{\mathcal{W}}$) instead of (IQP) we have ignored two things:

1. We have ignored all inactive constraints, i.e., we must require $a_i^T x \geq b_i$ for $i \notin \mathcal{W}$.
2. We have ignored that the active constraints are inequalities, i.e., we have required $A_{\mathcal{W}} x = b_{\mathcal{W}}$ instead of $A_{\mathcal{W}} x \geq b_{\mathcal{W}}$.

How are these requirements included?

Inclusion of inactive constraints

We have started in $\bar{x}$ and computed search direction p^*.

If $A(\bar{x} + p^*) \geq b$ then $\bar{x} + p^*$ satisfies all constraints.

Otherwise we can compute the maximum step length $\alpha_{\max}$ such that $A(\bar{x} + \alpha_{\max} p^*) \geq b$ holds.

The condition is $\alpha_{\max} = \min_{i:a_i^T p^* < 0} \dfrac{a_i^T \bar{x} - b_i}{-a_i^T p^*}$.

Two cases:

- $\alpha_{\max} \geq 1$. We let $\widetilde{x} \leftarrow \bar{x} + p^*$.
- $\alpha_{\max} < 1$. We let $\widetilde{x} \leftarrow \bar{x} + \alpha_{\max} p^*$ and $\mathcal{W} \leftarrow \mathcal{W} \bigcup \{l\}$, where $a_l^T(\bar{x} + \alpha_{\max} p^*) = b_l$.

The point $\bar{x} + p^*$ is of interest when $\alpha_{\max} \geq 1$.

Inclusion of inequality requirement

We assume that $\alpha_{\max} \geq 1$, i.e., $A\widetilde{x} \geq b$, where $\widetilde{x} = \bar{x} + p^*$.

When solving $(EQP_{\mathcal{W}})$ we obtain p^* and $\lambda^*_{\mathcal{W}}$. Two cases:

- $\lambda^*_{\mathcal{W}} \geq 0$. Then $\widetilde{x}$ is the optimal solution to

$$(IQP_{\mathcal{W}}) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^THx + c^Tx \\ \text{subject to} & A_{\mathcal{W}}x \geq b_{\mathcal{W}}, \quad x \in \mathbb{R}^n, \end{array}$$

  and hence an optimal solution to (IQP).

- $\lambda^*_k < 0$ for some k. If $A_{\mathcal{W}}p = e_k$ then $(H\widetilde{x} + c)^Tp = {\lambda^*_{\mathcal{W}}}^T A_{\mathcal{W}}p = \lambda^*_k < 0$. Therefore, let $\mathcal{W} \leftarrow \mathcal{W}\backslash\{k\}$.

An iteration in an active-set method for solving (*IQP*)

Given feasible $\bar{x}$ and $\mathcal{W}$ such that $A_{\mathcal{W}}$ has full row rank and $A_{\mathcal{W}}\bar{x} = b_{\mathcal{W}}$.

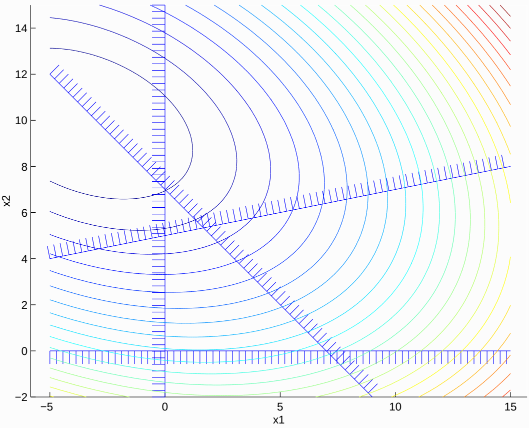
1. Solve $\begin{pmatrix} H & A_{\mathcal{W}}^T \\ A_{\mathcal{W}} & 0 \end{pmatrix} \begin{pmatrix} p^* \\ -\lambda_{\mathcal{W}}^* \end{pmatrix} = -\begin{pmatrix} H\bar{x} + c \\ 0 \end{pmatrix}$.
2. $l \leftarrow$ index for constraint first becomes violated along p^*.
3. $\alpha_{\max} \leftarrow$ maximum step length along p^*.
4. If $\alpha_{\max} < 1$, let $\bar{x} \leftarrow \bar{x} + \alpha_{\max} p^*$ and $\mathcal{W} \leftarrow \mathcal{W} \bigcup \{l\}$. New iteration.
5. Otherwise, $\alpha_{\max} \geq 1$. Let $\bar{x} \leftarrow \bar{x} + p^*$.
6. If $\lambda_{\mathcal{W}}^* \geq 0$ then $\bar{x}$ is optimal. Done!
7. Otherwise, $\lambda_k^* < 0$ for some k. Let $\mathcal{W} \leftarrow \mathcal{W} \backslash \{k\}$. New iteration.

Example problem

Consider the following two-dimensional example problem.

$$
\begin{array}{ll}
\text{minimize} & x_1^2 + x_1 x_2 + 2x_2^2 - 3x_1 - 36x_2 \\
\text{subject to} & x_1 \geq 0, \\
& x_2 \geq 0, \\
& -x_1 - x_2 \geq -7, \\
& x_1 - 5x_2 \geq -25.
\end{array}
$$

Geometric illustration of example problem



Optimal solution to example problem

Assume that we want to solve the example problem by an active-set method.

Initial point $x = (5\ 0)^T$.

We may initially choose $\mathcal{W} = \{2\}$ or $\mathcal{W} = \{0\}$.

Optimal solution $x^* = \begin{pmatrix} \frac{15}{32} & 5\frac{3}{32} \end{pmatrix}^T$ with $\lambda^* = \begin{pmatrix} 0 & 0 & 0 & 3\frac{1}{32} \end{pmatrix}^T$.

Comments on active-set method

Active-set method for quadratic programming:

- "Inexpensive" iterations. Only one constraint is added to or deleted from $\mathcal{W}$.
- $A_{\mathcal{W}}$ maintains full row rank.
- Straightforward modification to the case $H \succeq 0$. (For $H = 0$ we get the simplex method if the initial point is a vertex.)
- May potentially require an exponential number of iterations.
- May cycle (in theory). Anti-cycling strategy as in the simplex method.
- May be "warm started" efficiently if the initial point has "almost correct" active constraints.

Interior method for quadratic programming

$$(IQP) \qquad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^T H x + c^T x \\ \text{subject to} & Ax \geq b, \\ & x \in \mathbb{R}^n. \end{array}$$

We assume that $H \succeq 0$. Then, the problem is convex.

- An interior method for solving (IQP) approximately follows the barrier trajectory, which is created by a perturbation of the optimality conditions.
- To understand the method, we first consider the trajectory.
- Thereafter we study the method.
- The focus is on primal-dual interior methods.

Optimality conditions for (IQP)

$$
(IQP) \qquad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^T H x + c^T x \\ \text{subject to} & Ax \geq b, \\ & x \in \mathbb{R}^n. \end{array}
$$

We assume that $H \succeq 0$. Then, the problem is convex.

The optimality conditions for (IQP) may be written as

$$
\begin{aligned}
Ax - s &= b, \\
Hx - A^T\lambda &= -c, \\
s_i \lambda_i &= 0, \quad i = 1, \dots, m, \\
s &\geq 0, \\
\lambda &\geq 0.
\end{aligned}
$$

The primal-dual nonlinear equations

If the complementarity conditions $s_i \lambda_i = 0$ are perturbed to $s_i \lambda_i = \mu$ for a positive parameter μ, we obtain the primal-dual nonlinear equations

$$\begin{aligned} Ax - s &= b, \\ Hx - A^T \lambda &= -c, \\ s_i \lambda_i &= \mu, \quad i = 1, \dots, m. \end{aligned}$$

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

The parameter μ is called the barrier parameter.

Proposition

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

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

The primal-dual nonlinear equations, cont.

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

$$\begin{aligned} Ax - s &= b, \\ Hx - A^T\lambda &= -c, \\ S\Lambda e &= \mu e, \end{aligned}$$

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

Primal point of view

Primal point of view: $x(\mu), s(\mu)$ solve

$$(P_\mu) \qquad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^T Hx + c^T x - \mu \sum_{i=1}^{m} \ln s_i \\ \text{subject to} & Ax - s = b, \quad s > 0, \end{array}$$

with $\lambda(\mu)$ as Lagrange multipliers of $Ax - s = b$.

Optimality conditions for (P_μ):

$$\begin{aligned} Ax - s &= b, \\ Hx + c &= A^T\lambda, \\ -\frac{\mu}{s_i} &= -\lambda_i, \quad i = 1, \dots, m, \\ s &> 0. \end{aligned}$$

The barrier trajectory

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

We prefer the primal-dual nonlinear equations to the primal. A pure primal point of view gives high nonlinearity.

Theorem

If the barrier trajectory is well defined, it holds that $\lim_{\mu\to 0} x(\mu) = x^*$, $\lim_{\mu\to 0} s(\mu) = s^*$, $\lim_{\mu\to 0} \lambda(\mu) = \lambda^*$, where x^* is an optimal solution to (IQP), and λ^* is the associated Lagrange multiplier vector.

Hence, the barrier trajectory converges to an optimal solution.

Example problem

Consider the following two-dimensional example problem.

$$
\begin{array}{ll}
\text{minimize} & x_1^2 + x_1 x_2 + 2x_2^2 - 3x_1 - 36x_2 \\
\text{subject to} & x_1 \geq 0, \\
& x_2 \geq 0, \\
& -x_1 - x_2 \geq -7, \\
& x_1 - 5x_2 \geq -25.
\end{array}
$$

Geometric illustration of example problem

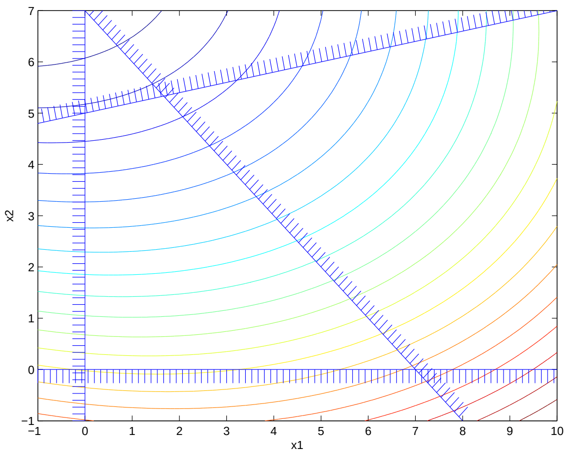


Illustration of primal barrier problem

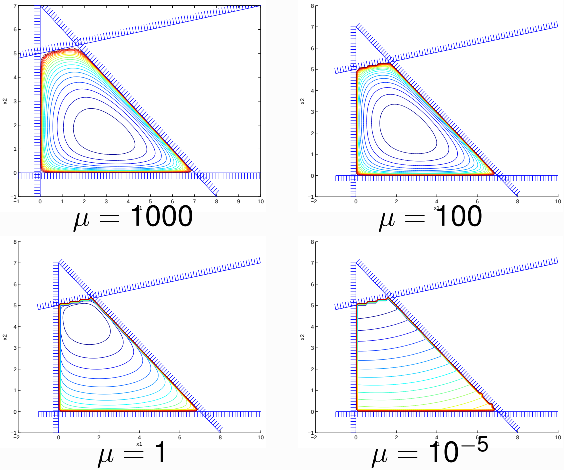
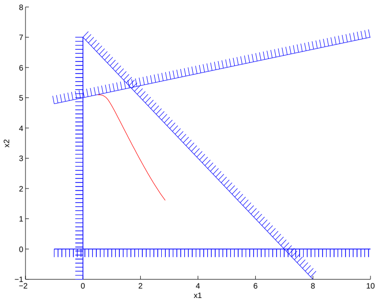
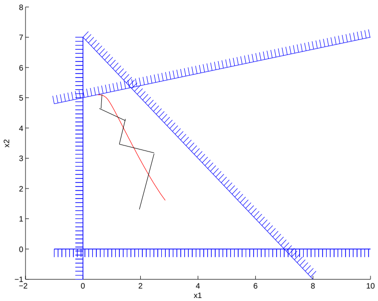


Illustration of primal part of barrier trajectory

An interior method approximately follows the barrier trajectory.



The trajectory



Generated iterates

A primal-dual interior method

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

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

$$\begin{pmatrix} H & 0 & -A^T \\ A & -I & 0 \\ 0 & \Lambda & S \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta \lambda \end{pmatrix} = - \begin{pmatrix} Hx + c - A^T\lambda \\ Ax - s - b \\ S\Lambda e - \mu e \end{pmatrix}.$$

Note that $Ax - s = b$ and $Hx - A^T\lambda = -c$ need not be satisfied at the initial point. Satisfied at $x + \Delta x$, $s + \Delta s$, $\lambda + \Delta\lambda$.

An iteration in a primal-dual interior method

1. Select a value for μ.
2. Compute the directions Δx, Δs and $\Delta\lambda$ from

$$\begin{pmatrix} H & 0 & -A^T \\ A & -I & 0 \\ 0 & \Lambda & S \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} Hx + c - A^T\lambda \\ Ax - s - b \\ S\Lambda e - \mu e \end{pmatrix}.$$

3. Compute the maximum steplength $\alpha_{\max}$ from $s + \alpha\Delta s \geq 0$, $\lambda + \alpha\Delta\lambda \geq 0$.
4. Let α be a suitable step, $\alpha = \min\{1, \eta\alpha_{\max}\}$, where $\eta < 1$.
5. Let $x = x + \alpha\Delta x$, $s = s + \alpha\Delta s$, $\lambda = \lambda + \alpha\Delta\lambda$.

Behavior of interior method

Normally rather few iterations on a quadratic program. (Depends on the strategy for reducing μ). The number of iterations does typically not increase significantly with problem size.

The Newton iteration may be written

$$\begin{pmatrix} H & A^T \\ A & -S\Lambda^{-1} \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta\lambda \end{pmatrix} = - \begin{pmatrix} Hx + c - A^T\lambda \\ Ax - b - \mu\Lambda^{-1}e \end{pmatrix}.$$

Symmetric indefinite matrix. Sparse matrix if H and A are sparse.

Unclear how to “warm start” the method efficiently.

Solution methods

- Solution methods are typically iterative methods that solve a sequence of simpler problems.
- Methods differ in terms of how complex subproblems that are formed.
- Many methods exist, e.g., interior methods, sequential quadratic programming methods etc.
- Rule of thumb: Second-derivatives are useful.

Two important classes of solution methods

- Sequential-quadratic programming (SQP) methods.
  - Local quadratic models of the problem are made.
  - Subproblem is a constrained quadratic program.
  - “Hard” prediction of active constraints.
  - Subproblem may be warmstarted.
- Interior methods.
  - Linearizations of perturbed optimality conditions are made.
  - Subproblem is a system of linear equations.
  - “Soft” prediction of active constraints.
  - Warm start is not easy.

Derivative information

- First-derivative methods are often not efficient enough.
- SQP methods and interior methods are second-derivative methods.
- An alternative to exact second derivatives are quasi-Newton methods.
- Stronger convergence properties for exact second derivatives.
- Exact second derivatives expected to be more efficient in practice.
- Exact second derivatives requires handling of nonconvexity.

Optimality conditions for nonlinear programs

Consider an equality-constrained nonlinear programming problem

$$(P_=) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) = 0, \end{array} \quad \text{where } f, g \in C^2,\ g : \mathbb{R}^n \to \mathbb{R}^m.$$

If the Lagrangian function is defined as $\mathcal{L}(x,\lambda) = f(x) - \lambda^T g(x)$, the first-order optimality conditions are $\nabla \mathcal{L}(x,\lambda) = 0$. We write them as

$$\begin{pmatrix} \nabla_x \mathcal{L}(x,\lambda) \\ -\nabla_\lambda \mathcal{L}(x,\lambda) \end{pmatrix} = \begin{pmatrix} \nabla f(x) - A(x)^T \lambda \\ g(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where $A(x)^T = \begin{pmatrix} \nabla g_1(x) & \nabla g_2(x) & \cdots & \nabla g_m(x) \end{pmatrix}$.

Newton's method for solving a nonlinear equation

Consider solving the nonlinear equation $\nabla f(u) = 0$, where $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$.

Then, $\nabla f(u + p) = \nabla f(u) + \nabla^2 f(u)p + o(\|p\|)$.

Linearization given by $\nabla f(u) + \nabla^2 f(u)p$.

Choose p so that $\nabla f(u) + \nabla^2 f(u)p = 0$, i.e., solve $\nabla^2 f(u)p = -\nabla f(u)$.

A Newton iteration takes the following form for a given u.

- p solves $\nabla^2 f(u)p = -\nabla f(u)$.
- $u \leftarrow u + p$.

(The nonlinear equation need not be a gradient.)

Speed of convergence for Newton's method

Theorem

Assume that $f \in C^3$ and that $\nabla f(u^*) = 0$ with $\nabla^2 f(u^*)$ nonsingular. Then, if Newton's method (with steplength one) is started at a point sufficiently close to u^*, then it is well defined and converges to u^* with convergence rate at least two, i.e., there is a constant C such that $\|u_{k+1} - u^*\| \le C\|u_k - u^*\|^2$.

The proof can be given by studying a Taylor-series expansion,

$$\begin{aligned} u_{k+1} - u^* &= u_k - \nabla^2 f(u_k)^{-1} \nabla f(u_k) - u^* \\ &= \nabla^2 f(u_k)^{-1}(\nabla f(u^*) - \nabla f(u_k) - \nabla^2 f(u_k)(u^* - u_k)). \end{aligned}$$

For u_k sufficiently close to u^*,

$$\|\nabla f(u^*) - \nabla f(u_k) - \nabla^2 f(u_k)(u^* - u_k)\| \le \bar{C}\|u_k - u^*\|^2.$$

First-order optimality conditions

The first-order necessary optimality conditions may be viewed as a system of $n+m$ nonlinear equations with $n+m$ unknowns, x and λ, according to

$$\begin{pmatrix} \nabla f(x) - A(x)^T\lambda \\ g(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

A Newton iteration takes the form $\begin{pmatrix} x^+ \\ \lambda^+ \end{pmatrix} = \begin{pmatrix} x \\ \lambda \end{pmatrix} + \begin{pmatrix} p \\ \nu \end{pmatrix}$,

where

$$\begin{pmatrix} \nabla^2{}_{xx}\mathcal{L}(x,\lambda) & -A(x)^T \\ A(x) & 0 \end{pmatrix} \begin{pmatrix} p \\ \nu \end{pmatrix} = \begin{pmatrix} -\nabla f(x) + A(x)^T\lambda \\ -g(x) \end{pmatrix},$$

for $\mathcal{L}(x,\lambda) = f(x) - \lambda^T g(x)$.

First-order optimality conditions, cont.

The resulting Newton system may equivalently be written as

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & -A(x)^T \\ A(x) & 0 \end{pmatrix} \begin{pmatrix} p \\ \lambda + \nu \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -g(x) \end{pmatrix},$$

alternatively

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & A(x)^T \\ A(x) & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -g(x) \end{pmatrix}.$$

We prefer the form with λ^+, since it can be directly generalized to problems with inequality constraints.

Quadratic programming with equality constraints

Compare with an equality-constrained quadratic programming problem

$$(EQP) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2} p^T H p + c^T p \\ \text{subject to} & Ap = b, \\ & p \in \mathbb{R}^n, \end{array}$$

where the unique optimal solution p and multiplier vector λ^+ are given by

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix},$$

if $Z^T H Z \succ 0$ and A has full row rank.

Newton iteration and equality-constrained QP

Compare $\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & A(x)^T \\ A(x) & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -g(x) \end{pmatrix}$

with $\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}$.

Identify:

$$\begin{array}{ccc} \nabla^2_{xx}\mathcal{L}(x,\lambda) & \longleftrightarrow & H \\ \nabla f(x) & \longleftrightarrow & c \\ A(x) & \longleftrightarrow & A \\ -g(x) & \longleftrightarrow & b. \end{array}$$

Newton iteration as a QP problem

A Newton iteration for solving the first-order necessary optimality conditions to $(P_=)$ may be viewed as solving the QP problem

$$(QP_=) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}p^T \nabla^2_{xx}\mathcal{L}(x,\lambda)p + \nabla f(x)^T p \\ \text{subject to} & A(x)p = -g(x), \\ & p \in \mathbb{R}^n, \end{array}$$

and letting $x^+ = x + p$, and λ^+ are given by the multipliers of $(QP_=)$.

Problem $(QP_=)$ is well defined with unique optimal solution p and multiplier vector λ^+ if $Z(x)^T \nabla^2_{xx}\mathcal{L}(x,\lambda)Z(x) \succ 0$ and $A(x)$ has full row rank, where $Z(x)$ is a matrix whose columns form a basis for null$(A(x))$.

An SQP iteration for problems with equality constraints

Given x, λ such that $Z(x)^T \nabla^2_{xx} \mathcal{L}(x, \lambda) Z(x) \succ 0$ and $A(x)$ has full row rank, a Newton iteration takes the following form.

1. Compute optimal solution p and multiplier vector λ^+ to

$$(QP_=) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2} p^T \nabla^2_{xx} \mathcal{L}(x, \lambda) p + \nabla f(x)^T p \\ \text{subject to} & A(x) p = -g(x), \\ & p \in \mathbb{R}^n, \end{array}$$

2. $x \leftarrow x + p, \quad \lambda \leftarrow \lambda^+$.

We call this method sequential quadratic programming (SQP).

Note! $(QP_=)$ is solved by solving a system of linear equations.

Note! x and λ have given numerical values in $(QP_=)$.

SQP method for equality-constrained problems

So far we have discussed SQP for $(P_=)$ in an "ideal" case.
Comments:

- If $Z(x)^T \nabla^2_{xx} \mathcal{L}(x, \lambda) Z(x) \not\succ 0$ we may replace $\nabla^2_{xx} \mathcal{L}(x, \lambda)$ by B in $(QP_=)$, where B is a symmetric approximation of $\nabla^2_{xx} \mathcal{L}(x, \lambda)$ that satisfies $Z(x)^T B Z(x) \succ 0$.
- A quasi-Newton approximation B of $\nabla^2_{xx} \mathcal{L}(x, \lambda)$ may be used.
- If $A(x)$ does not have full row rank $A(x)p = -g(x)$ may lack solution. This may be overcome by introducing "elastic" variables. This is not covered here.
- We have shown local convergence properties. To obtain convergence from an arbitrary initial point we may utilize a merit function and use linesearch.

Enforcing convergence by a linesearch strategy

Compute optimal solution p and multiplier vector λ^+ to

$$(QP_=) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2} p^T \nabla^2{}_{xx} \mathcal{L}(x,\lambda) p + \nabla f(x)^T p \\ \text{subject to} & A(x)p = -g(x), \\ & p \in \mathbb{R}^n, \end{array}$$

$x \leftarrow x + \alpha p$, where α is determined in a linesearch to give sufficient decrease of a merit function.

(Ideally, $\alpha = 1$ eventually.)

Example of merit function for SQP on $(P_=)$

A merit function typically consists of a weighting of optimality and feasibility. An example is the augmented Lagrangian merit function $M_\mu(x) = f(x) - \lambda(x)^T g(x) + \dfrac{1}{2\mu} g(x)^T g(x)$, where μ is a positive parameter and $\lambda(x) = (A(x)A(x)^T)^{-1}A(x)\nabla f(x)$. (The vector $\lambda(x)$ is here the least-squares solution of $A(x)^T\lambda = \nabla f(x)$.)

Then the SQP solution p is a descent direction to M_μ at x if μ is sufficiently close to zero and $Z(x)^T B Z(x) \succ 0$.

We may then carry out a linesearch on M_μ in the x-direction and define $\lambda(x) = (A(x)A(x)^T)^{-1}A(x)\nabla f(x)$.

Ideally the step length is chosen as $\alpha = 1$.

We consider the "pure" method, where $\alpha = 1$ and λ^+ is given by $(QP_=)$.

SQP for inequality-constrained problems

- In the SQP subproblem $(QP_=)$, the constraints are approximated by a linearization around x, i.e., the requirement on p is $g_i(x) + \nabla g_i(x)^T p = 0$, $i = 1, \dots, m$.
- For an inequality constraint $g_i(x) \ge 0$ this requirement may be generalized to $g_i(x) + \nabla g_i(x)^T p \ge 0$.
- An SQP method gives in each iteration a prediction of the active constraints in (P) by the constraints that are active in the SQP subproblem.
- The QP subproblem gives nonnegative multipliers for the inequality constraints.

The SQP subproblem for a nonlinear program

The problem

$$
(P)\quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \geq 0,\ i \in \mathcal{I}, \\ & g_i(x) = 0,\ i \in \mathcal{E}, \\ & x \in \mathbb{R}^n, \end{array}
$$

where $f, g \in C^2$, $g : \mathbb{R}^n \to \mathbb{R}^m$,

has, at a certain point x, λ, an SQP subproblem

$$
(QP)\quad \begin{array}{ll} \text{minimize} & \frac{1}{2} p^T \nabla^2{}_{xx} \mathcal{L}(x, \lambda) p + \nabla f(x)^T p \\ \text{subject to} & \nabla g_i(x)^T p \geq -g_i(x),\ i \in \mathcal{I}, \\ & \nabla g_i(x)^T p = -g_i(x),\ i \in \mathcal{E}, \\ & p \in \mathbb{R}^n, \end{array}
$$

which has optimal solution p and Lagrange multiplier vector λ^+.

An SQP iteration for nonlinear optimization problem

Given x, λ such that $\nabla^2_{xx}\mathcal{L}(x,\lambda) \succ 0$, an SQP iteration for (P) takes the following form.

1. Compute optimal solution p and multiplier vector λ^+ to

$$(QP) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}p^T\nabla^2_{xx}\mathcal{L}(x,\lambda)p + \nabla f(x)^T p \\ \text{subject to} & \nabla g_i(x)^T p \geq -g_i(x), \ i \in \mathcal{I}, \\ & \nabla g_i(x)^T p = -g_i(x), \ i \in \mathcal{E}, \\ & p \in \mathbb{R}^n. \end{array}$$

2. $x \leftarrow x + p, \quad \lambda \leftarrow \lambda^+$.

Note that $\lambda_i \geq 0$, $i \in \mathcal{I}$, is maintained since λ^+ are Lagrange multipliers to (QP).

SQP method for nonlinear optimization

We have discussed the "ideal" case. Comments:

- If $\nabla^2_{xx}\mathcal{L}(x,\lambda) \nsucc 0$, we may replace $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ by B in (QP), where B is a symmetric approximation of $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ that satisfies $B \succ 0$.
- A quasi-Newton approximation B of $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ may be used. (Example SQP quasi-Newton solver: SNOPT.)
- The QP subproblem may lack feasible solutions. This may be overcome by introducing "elastic" variables. This is not covered here.
- We have shown local convergence properties. To obtain convergence from an arbitrary initial point we may utilize a merit function and use linesearch or trust-region strategy.

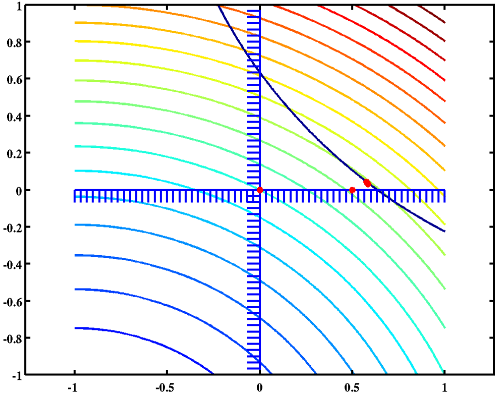
Example problem

Consider small example problem

$$
(P) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}(x_1+1)^2 + \frac{1}{2}(x_2+2)^2 \\ \text{subject to} & -3(x_1+x_2-2)^2 - (x_1-x_2)^2 + 6 = 0, \\ & x_1 \geq 0, \\ & x_2 \geq 0, \\ & x \in \mathbb{R}^2. \end{array}
$$

Optimal solution $x^* \approx (0.5767\ 0.0431)^T$, $\lambda_1^* \approx 0.2185$.

Graphical illustration of example problem



Optimal solution $x^* \approx (0.5767\ 0.0431)^T$, $\lambda_1^* \approx 0.2185$.

Barrier function for general nonlinear problem

Consider an inequality-constrained problem

$$(P_{\geq}) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \geq 0, \end{array} \quad \text{where } f, g \in C^2, \; g : \mathbb{R}^n \to \mathbb{R}^m.$$

We assume $\{x \in \mathbb{R}^n : g(x) > 0\} \neq \emptyset$ and require $g(x) > 0$ "implicitly".

For a positive parameter μ, form the logarithmic barrier function

$$B_\mu(x) = f(x) - \mu \sum_{i=1}^{m} \ln g_i(x).$$

Necessary conditions for a minimizer of $B_\mu(x)$ are $\nabla B_\mu(x) = 0$, where

$$\nabla B_\mu(x) = \nabla f(x) - \mu \sum_{i=1}^{m} \frac{1}{g_i(x)} \nabla g_i(x) = \nabla f(x) - \mu A(x)^T G(x)^{-1} e,$$

with $G(x) = \operatorname{diag}(g(x))$ and $e = (1 \; 1 \; \ldots 1)^T$.

Barrier function for general nonlinear problem, cont.

If $x(\mu)$ is a local minimizer of $\min_{x:g(x)>0} B_\mu(x)$ it holds that $\nabla f(x(\mu)) - \mu A(x(\mu))^T G(x(\mu))^{-1} e = 0$.

Proposition

Let $x(\mu)$ be a local minimizer of $\min_{x:g(x)>0} B_\mu(x)$. Under suitable conditions, it holds that

$$\lim_{\mu \to 0} x(\mu) = x^*, \quad \lim_{\mu \to 0} \mu G(x(\mu))^{-1} e = \lambda^*,$$

where x^* is a local minimizer of $(P_\geq)$ and λ^* is the associated Lagrange multiplier vector.

Note! It holds that $g(x(\mu)) > 0$.

Barrier function for general nonlinear problem, cont.

Let $\lambda(\mu) = \mu G(x(\mu))^{-1} e$, i.e., $\lambda_i(\mu) = \dfrac{\mu}{g_i(x(\mu))}$, $i = 1, \dots, m$.

Then, $\nabla B_\mu(x(\mu)) = 0 \iff \nabla f(x(\mu)) - A(x(\mu))^T \lambda(\mu) = 0$.

This means that $x(\mu)$ and $\lambda(\mu)$ solve the nonlinear equation

$$\begin{aligned} \nabla f(x) - A(x)^T \lambda &= 0, \\ \lambda_i - \frac{\mu}{g_i(x)} &= 0, \quad i = 1, \dots, m, \end{aligned}$$

where we in addition require $g(x) > 0$ and $\lambda > 0$. If the second block of equations is multiplied by $G(x)$ we obtain

$$\begin{aligned} \nabla f(x) - A(x)^T \lambda &= 0, \\ g_i(x)\lambda_i - \mu &= 0, \quad i = 1, \dots, m. \end{aligned}$$

A perturbation of the first-order necessary optimality conditions.

Barrier function method

A barrier function method approximately finds $x(\mu)$, $\lambda(\mu)$ for decreasing values of μ.

A primal-dual method takes Newton iterations on the primal-dual nonlinear equations

$$\begin{aligned} \nabla f(x) - A(x)^T\lambda &= 0, \\ G(x)\lambda - \mu e &= 0. \end{aligned}$$

The Newton step Δx, $\Delta\lambda$ is given by

$$\begin{pmatrix} \nabla^2{}_{xx}\mathcal{L}(x,\lambda) & A(x)^T \\ \Lambda A(x) & -G(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta\lambda \end{pmatrix} = -\begin{pmatrix} \nabla f(x) - A(x)^T\lambda \\ G(x)\lambda - \mu e \end{pmatrix},$$

where $\Lambda = \text{diag}(\lambda)$.

An iteration in a primal-dual barrier function method

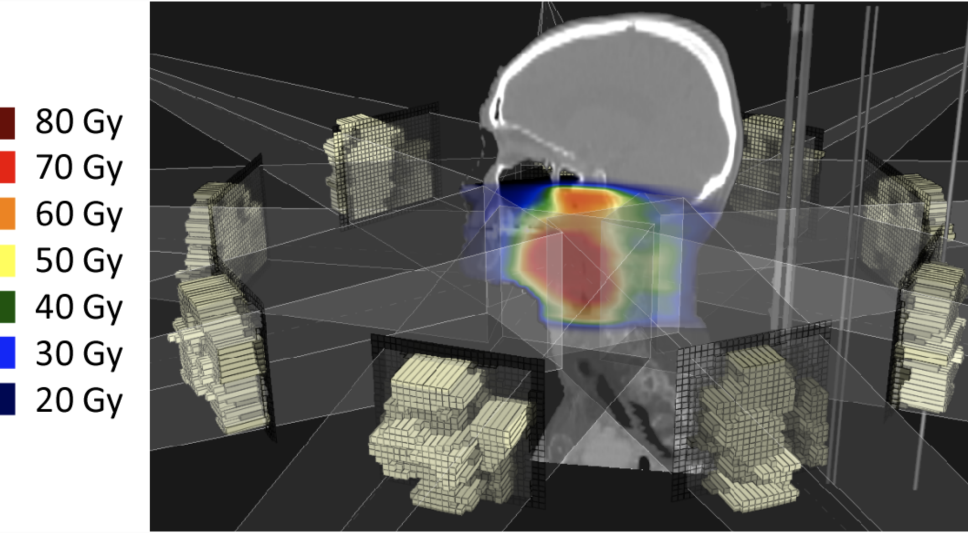
An iteration in a primal-dual barrier function method takes the following form, given $\mu > 0$, x such that $g(x) > 0$ and $\lambda > 0$.

1. Compute Δx, $\Delta\lambda$ from
$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & A(x)^T \\ \Lambda A(x) & -G(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta\lambda \end{pmatrix} = -\begin{pmatrix} \nabla f(x) - A(x)^T\lambda \\ G(x)\lambda - \mu e \end{pmatrix}.$$
2. Choose "suitable" steplength α such that $g(x + \alpha\Delta x) > 0$, $\lambda + \alpha\Delta\lambda > 0$.
3. $x \leftarrow x + \alpha\Delta x,\;\; \lambda \leftarrow \lambda + \alpha\Delta\lambda$.
4. If (x, λ) "sufficiently close" to $(x(\mu), \lambda(\mu))$, reduce μ.

Radiation therapy

- Treatment of cancer is a very important task.
- Radiation therapy is one of the most powerful methods of treatment. In Sweden 30% of the cancer patients are treated with radiation therapy.
- The radiation may be optimized to improve performance of radiation.
- Hence, behind this important medical application is an optimization problem.

Radiation treatment



Aim of radiation

- The aim of the radiation is typically to give a treatment that leads to a desirable dose distribution in the patient.
- Typically, high dose is desired in the tumor cells, and low dose in the other cells.
- In particular, certain organs are very sensitive to radiation and must have a low dose level, e.g., the spine.
- Hence, a desired dose distribution can be specified, and the question is how to achieve this distribution.
- This is an inverse problem in that the desired result of the radiation is known, but the treatment plan has to be designed.

Formulation of optimization problem

- A radiation treatment is typically given as a series of radiations.
- For an individual treatment, the performance depends on
  - the beam angle of incidence, which is governed by the supporting gantry; and
  - the intensity modulation of the beam, which is governed by the treatment head.
- One may now formulate an optimization problem, where the variables are the beam angles of incidence and the intensity modulations of the beams.
- In this talk, we assume that the beam angles of incidence are fixed.

Optimization of radiation therapy

Joint research project between

KTH and RaySearch Laboratories AB.

Financially supported by the Swedish Research Council.

Previous industrial graduate student: Fredrik Carlsson. (PhD April 2008)

Current industrial graduate students: Rasmus Bokrantz and Albin Fredriksson.

Solution method

A simplified bound-constrained problem may be posed as

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & l \leq x \leq u. \end{array}$$

- Large-scale problem solved in few (˜20) iterations using a quasi-Newton SQP method.
- Difficulty: “Jagged” solutions for more accurate plans.
- Idea: Use second-derivatives and an interior method to obtain fast convergence and smooth solutions.
  - Good news: Faster convergence.
  - Bad news: Increased jaggedness.
- Not following the folklore.

Radiation therapy and the conjugate-gradient method

- Why does a quasi-Newton sequential quadratic programming method do so well on these problems?
- The answer lies in the problem structure.
- Simplify further, consider a quadratic approximation of the objective function and eliminate the constraints.

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \qquad \tfrac{1}{2} x^T H x + c^T x$$

  where $H = H^T \succeq 0$.
- Quasi-Newton methods and the conjugate-gradient method are equivalent on this problem.
- The conjugate-gradient method minimizes in directions corresponding to large eigenvalues first.

Radiation therapy and the conjugate-gradient method

- The conjugate-gradient method minimizes in directions corresponding to large eigenvalues first.
- Our simplified problem has few large eigenvalues, corresponding to smooth solutions.
- Many small eigenvalues that correspond to jagged solutions.
- The conjugate-gradient method takes a desirable path to the solution.
- Additional properties of the solution, not seen in the formulation, are important.

Behavior of the conjugate gradient subproblems

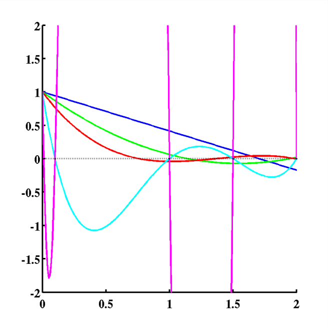
$$
\begin{array}{ll}
\underset{\xi\in\mathbb{R}^n,\zeta\in\mathbb{R}^k}{\text{minimize}} & \frac{1}{2}\sum_{i=1}^{n}\lambda_i\xi_i^2 \\
\text{subject to} & \xi_i = \prod_{l=1}^{k}\left(1-\frac{\lambda_i}{\zeta_l}\right)\xi_i^{(0)}, \quad i = 1,\dots,n,
\end{array}
$$

The optimal solution $\xi^{(k)}$ will tend to have smaller components $\xi_i^{(k)}$ for i such that λ_i is large and/or $\xi_i^{(0)}$ is large.

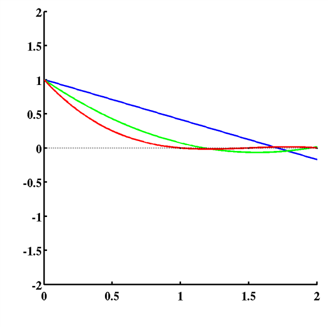
Nonlinear dependency of $\xi^{(k)}$ on λ and $\xi^{(0)}$.

We are interested in the ill-conditioned case, when H has relatively few large eigenvalues.

Polynomials for ill-conditioned example problem



Polynomials for problem with $\lambda = (2, 1.5, 1, 0.1, 0.01)^T$ and $\xi^{(0)} = (1, 1, 1, 1, 1)^T$.



Polynomials for problem with $\lambda = (2, 1.5, 1)^T$ and $\xi^{(0)} = (1, 1, 1)^T$.

Optimization approaches to distributed multi-cell radio resource management

Research project within the

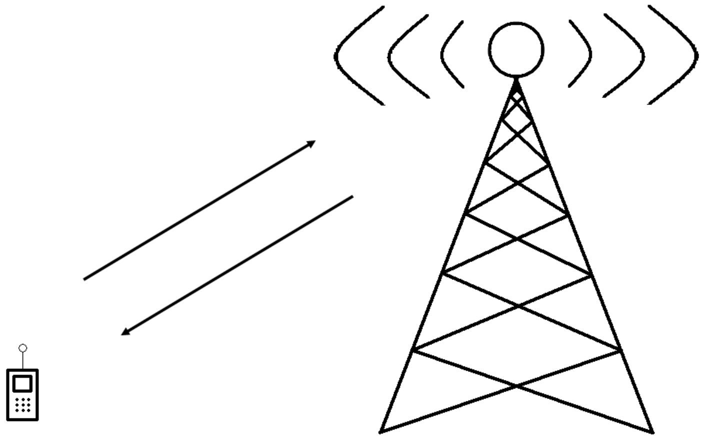
KTH Center for Industrial and Applied Mathematics (CIAM).

Industrial partner: Ericsson.

Financially supported by the
Swedish Foundation for Strategic Research.

Graduate student: Mikael Fallgren.

Radio resource management



Optimization problem

- Maximize throughput.
- Nonconvex problem.
- Convexification possible.
- Leads to loss of separability.

Question: How is this problem best solved?

Research in progress.

Some personal comments

A personal view on nonlinear optimization.

- Methods are very important.
- Applications give new challenges.
- Often two-way communication between method and application.
- Collaboration with application experts extremely important.

Thank you for your attention!

Conference announcement

3rd Nordic Optimization Symposium
March 13–14 2009
KTH, Stockholm

See http://www.math.kth.se/optsyst/3nos

Welcome!