KTH Mathematics

2E5295/5B5749 Convex optimization with engineering applications

Lecture 8

Smooth convex unconstrained and equality-constrained minimization

Unconstrained convex program

Consider a convex optimization problem on the form

$$(CP) \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice continuously differentiable.

Proposition. *Let p^* denote the optimal value of (CP). Let x^* in $\mathbb{R}^n$, let $m = \eta_{\min}(\nabla^2 f(x))$ and let $M = \eta_{\max}(\nabla^2 f(x))$. Then,*

$$f(x) - \frac{1}{2m}\|\nabla f(x)\|_2^2 \le p^* \le f(x) - \frac{1}{2M}\|\nabla f(x)\|_2^2.$$

Note that $\tilde{y} = x - (1/m)\nabla f(x)$ minimizes $\nabla f(x)^T(y - x) + \frac{m}{2}\|y - x\|_2^2$.

The direction $-(1/m)\nabla f(x)$ is a *steepest descent* step.

Iterative methods

$$(CP) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in I\!R^n, \end{array}$$

where $f \in C^2$, f convex on $I\!R^n$.

An iterative method generates x_0, x_1, x_2, $\ldots$ such that $\lim_{k\to\infty} x_k = x^*$, where $\nabla f(x^*) = 0$.

Terminates when suitable convergence criteria is fulfilled, e.g., $\|\nabla f(x_k)\| < \epsilon$.

Linesearch methods

A *linesearch method* generates in each iteration a *search direction* and performs a *linesearch* along the search direction.

Iteration k takes the following form at x_k.

- Compute search direction p_k such that $\nabla f(x_k)^T p_k < 0$.
- Approximatively solve $\min_{\alpha \geq 0} f(x_k + \alpha p_k)$, which gives α_k.
- $x_{k+1} \leftarrow x_k + \alpha_k p_k$.

Different methods vary in choice of p_k and α_k.

Classes of linesearch methods

We will initially consider two fundamental methods.

- The *steepest-descent method*, where $p_k = -\nabla f(x_k)$, and
- *Newton's method*, where $\nabla^2 f(x_k) p_k = -\nabla f(x_k)$.

Steepest descent:	$+$	Search direction inexpensive to compute,
	$-$	Slow convergence.
Newton's method:	$-$	Search direction more expensive to compute,
	$+$	Faster convergence.

There are methods "in-between", e.g., *quasi-Newton methods* that aim at mimicking Newton's method without computing second derivatives.

Quadratic objective function

Consider model problem with quadratic objective function

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

where $H \succeq 0$.

Proposition. *The following holds for (QP) depending on H and c:*
(i) If $H \succ 0$ then (QP) has a unique minimizer x^* given by $Hx^* = -c$.
(ii) If $H \succeq 0$, $H \not\succ 0$ each x^* that fulfills $Hx^* = -c$ is a global minimizer to (QP). (There may possibly be no such x^*).

Proof. The condition $\nabla f(x^*) = 0$ gives the results. $\square$

We assume $H \succ 0$ in the discussion.

Linesearch method on quadratic objective function

Consider (QP) with $f(x) = \frac{1}{2} x^T H x + c^T x$, where $H \succ 0$. We obtain:

x^* minimizer to $(QP) \iff 0 = \nabla f(x^*) = Hx^* + c$.

Suppose search direction p_k satisfies $\nabla f(x_k)^T p_k < 0$.

Let $\varphi(\alpha) = f(x_k + \alpha p_k) = f(x_k) + \alpha \nabla f(x_k)^T p_k + \frac{\alpha^2}{2} p_k^T H p_k$.

Then $\alpha_k = -\dfrac{\nabla f(x_k)^T p_k}{p_k^T H p_k}$ gives the minimizer to $\min_{\alpha \ge 0} f(x_k + \alpha p_k)$,
i.e., we can perform *exact linesearch*.

Steepest descent on quadratic objective function

Assume that $f(x) = \frac{1}{2}x^T H x + c^T x$, where $H \succ 0$.

Further assume that steepest descent with exact linesearch is used, i.e., $p_k = -\nabla f(x_k)$ and $\alpha_k = -\dfrac{\nabla f(x_k)^T p_k}{p_k^T H p_k}$.

Then it can be shown that

$$f(x_{k+1}) - f(x^*) \le \left(\frac{\text{cond}(H) - 1}{\text{cond}(H) + 1}\right)^2 (f(x_k) - f(x^*)).$$

$\text{cond}(H) \gg 1 \Rightarrow \dfrac{\text{cond}(H) - 1}{\text{cond}(H) + 1} \approx 1$, i.e., slow *linear convergence.*

For a nonlinear function, we typically get slow linear convergence, where H is replaced by $\nabla^2 f(x_k)$.

Speed of convergence

Definition. *Assume that* $x_k \in I\!R^n, k = 0, 1, \ldots,$ *and assume that* $\lim_{k\to\infty} x_k = x^*$. *We say that* $\{x_k\}_{k=0}^{\infty}$ *converges to* x^* *with speed of convergence* r *if*

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^r} = C, \quad \textit{where } C < \infty.$$

We have $\|x_{k+1} - x^*\| \approx C \cdot \|x_k - x^*\|^r$.

We want r large (and C close to zero). Of interest:

- $r = 1$, $0 < C < 1$, linear convergence. (Steepest descent.)
- $r = 1$, $C = 0$, superlinear convergence. (Quasi-Newton.)
- $r = 2$, quadratic convergence. (Newton's method.)

Newton's method for solving a nonlinear equation

Consider solving the nonlinear equation $\nabla f(u) = 0$, where $f : I\!R^n \rightarrow I\!R$, $f \in C^2$.

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 $\nabla^2 f(u^*)$ is 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^*,

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

One-dimensional example for Newton's method

For a positive number d, consider computing $1/d$ by minimizing $f(u) = du - \ln u$.

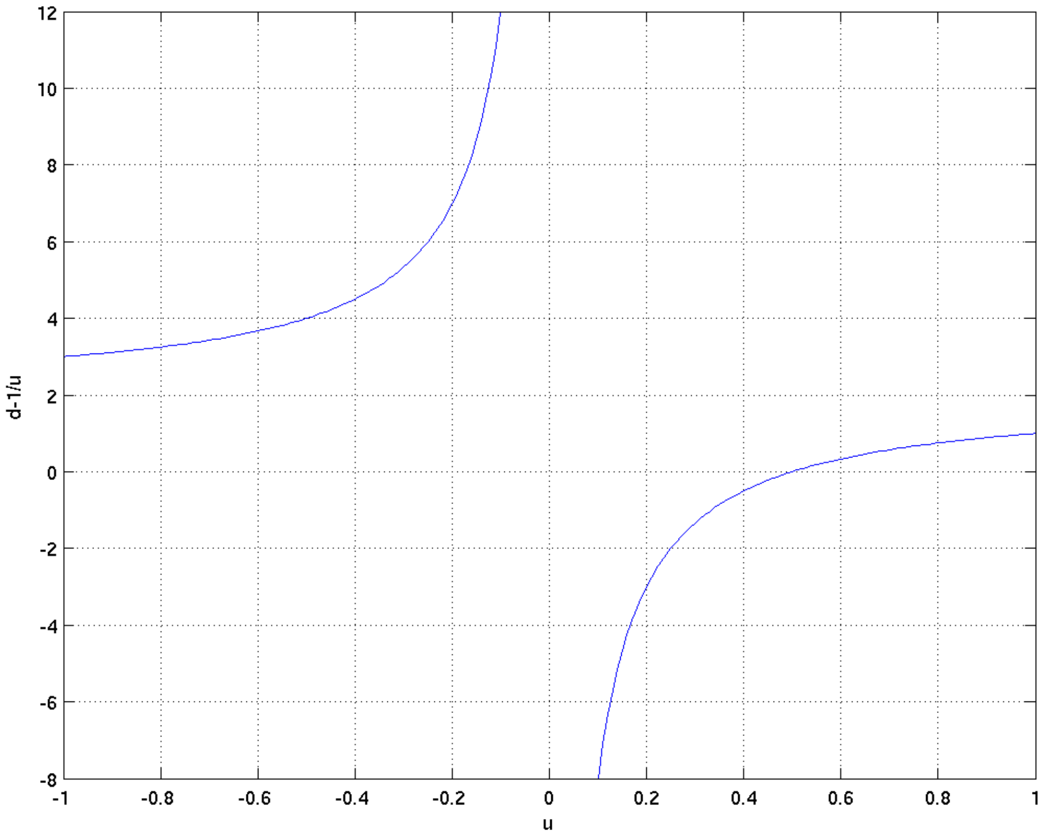
Then, $f'(u) = d - \dfrac{1}{u}$, $f''(u) = \dfrac{1}{u^2}$. We see that $u^* = \dfrac{1}{d}$.

$$u_{k+1} = u_k - \frac{f'(u_k)}{f''(u_k)} = u_k - \frac{d - \dfrac{1}{u_k}}{\dfrac{1}{u_k^2}} = 2u_k - u_k^2 d.$$

Then, $u_{k+1} - \dfrac{1}{d} = 2u_k - u_k^2 d - \dfrac{1}{d} = -d\left(u_k - \dfrac{1}{d}\right)^2$.

One-dimensional example for Newton's method, cont.

Graphical picture for $d = 2$.



Sufficient descent direction

For a direction p_k to ensure convergence is must be a *sufficient descent direction*. Typical conditions are

$$-\frac{\nabla f(x_k)^T p_k}{\|\nabla f(x_k)\| \|p_k\|} \geq \sigma, \quad \text{where } \sigma \text{ is a positive constant} \quad .$$

This means that p_k must be "sufficiently similar" to the negative gradient.

For search direction p_k from $B_k p_k = -\nabla f(x_k)$ this is required by ensuring that $\|B_k\| \leq M$ and $\|B_k^{-1}\| \leq m$, where m and M are positive constants.

In a modified Newton method modifications of $\nabla^2 f(x_k)$ can be made, if needed, by a modified Cholesky factorization.

Linesearch

In the linesearch α_k is determined as an approximate solution to $\min_{\alpha \geq 0} \varphi(\alpha)$, where $\varphi(\alpha) = f(x_k + \alpha p_k)$. We want $f(x_{k+1}) < f(x_k)$, i.e., $\varphi(\alpha_k) < \varphi(0)$. This is not sufficient to ensure convergence.

Example requirement for step not too long:

$\varphi(\alpha) \leq \varphi(0) + \mu \alpha \varphi'(0)$, i.e., (Armijo condition)

$f(x_k + \alpha p_k) \leq f(x_k) + \mu \alpha \nabla f(x_k)^T p_k$,

where $\mu \in (0, \frac{1}{2})$.

Example requirement for step not too short:

$|\varphi'(\alpha)| \leq -\eta \varphi'(0)$, i.e., (Wolfe condition)

$|\nabla f(x_k + \alpha p_k)^T p_k| \leq -\eta \nabla f(x_k)^T p_k$,

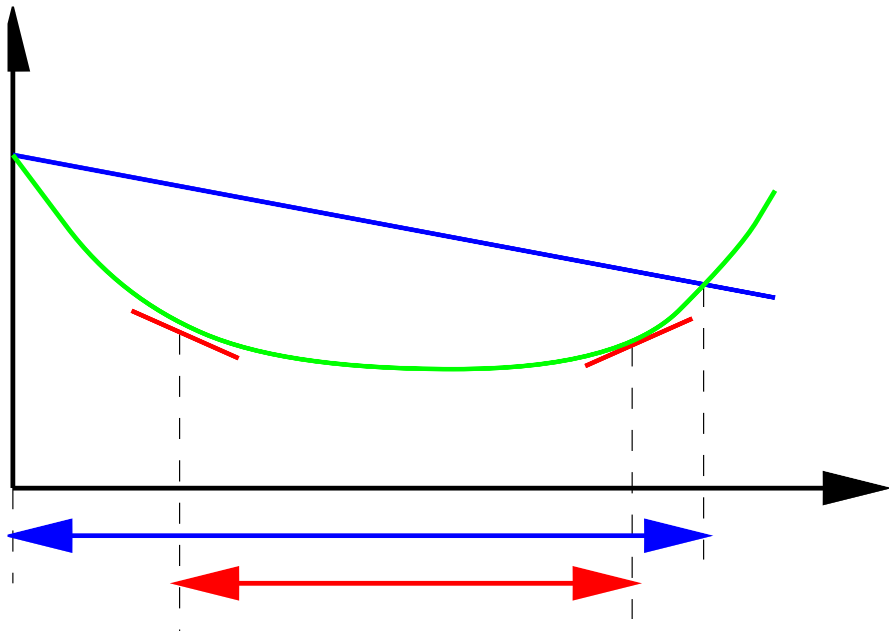
where $\eta \in (\mu, 1)$. Alternative requirement for not too short step:

Take smallest nonnegative integer i such that

$f(x_k + 2^{-i} p_k) \leq f(x_k) + \mu 2^{-i} \nabla f(x_k)^T p_k$. ("Backtracking")

Illustration of linesearch conditions

The linesearch conditions of Wolfe-Armijo type can be illustrated in the following picture.



Linesearch conditions

To find $\bar{\alpha}$ such that the Wolfe and Armijo conditions are fulfilled we may consider $\hat{\varphi}(\alpha) = \varphi(\alpha) - \varphi(0) - \mu\alpha\varphi'(0)$.

Then $\hat{\varphi}(0) = 0$ and $\hat{\varphi}'(0) < 0$. In addition, there must exist $\bar{\alpha} > 0$ such that $\hat{\varphi}(\bar{\alpha}) = 0$, otherwise φ is unbounded from below.

By the mean-value theorem there is an $\hat{\alpha} \in (0, \bar{\alpha})$ such that $\hat{\varphi}(\hat{\alpha}) < 0$ and $\hat{\varphi}'(\hat{\alpha}) = 0$.

Since $\mu < \eta$ we obtain $\varphi(\alpha) \leq \varphi(0) + \mu\alpha\varphi'(0)$ and $|\varphi'(\alpha)| \leq -\eta\varphi'(0)$ for α in a neighborhood of $\hat{\alpha}$.

For example, bisection in combination with polynomial interpolation can be used on $\hat{\varphi}$ to find a suitable α.

Newton's method and steepest descent

The steepest descent direction solves

$$\begin{array}{ll} \underset{p \in \mathbb{R}^n}{\text{minimize}} & \nabla f(x)^T p \\ \text{subject to} & p^T p \le 1. \end{array}$$

The Newton direction solves

$$\begin{array}{ll} \underset{p \in \mathbb{R}^n}{\text{minimize}} & \nabla f(x)^T p \\ \text{subject to} & p^T \nabla^2 f(x) p \le 1. \end{array}$$

The Newton step is a steepest-descent step in the norm defined by $\nabla^2 f(x)$, i.e.,

$$\|u\|_{\nabla^2 f(x)} = (u^T \nabla^2 f(x) u)^{1/2}.$$

Self-concordant functions

When proving polynomial complexity of *interior methods* for convex optimization, the notion of *self-concordant functions* is an important concept.

Definition. *A three times differentiable function* $f : C \to I\!R$*, which is convex on the convex set* C*, is* self-concordant *if* $|f'''(x)| \le 2 f''(x)^{3/2}$.

In essence, this means that the third derivatives are not "too large".

An important self-concordant function is $f(x) = -\ln x$ for $x > 0$.

Suggested reading

Suggested reading in the textbook:

- Sections 9.1–9.7.

Equality-constrained convex program

Consider a convex optimization problem on the form

$$(CP_=) \qquad \begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & Ax = b, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice continuously differentiable.

If the Lagrangian function is defined as $l(x, \lambda) = f(x) - \lambda^T (Ax - b)$, the first-order optimality conditions are $\nabla l(x, \lambda) = 0$. We write them as

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

Newton iteration

A Newton iteration on the optimality conditions takes the form

$$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\nu \end{pmatrix} = - \begin{pmatrix} \nabla f(x) - A^T \lambda \\ Ax - b \end{pmatrix}.$$

We may use

$$\left\| \begin{pmatrix} \nabla f(x) - A^T \lambda \\ Ax - b \end{pmatrix} \right\|^2$$

as *merit function*, i.e., to measure how “good” a point is.

Variable elimination

Note that for a feasible point $\bar{x}$, it holds that $A(x - \bar{x}) = 0$ for all feasible x. Let Z be a matrix whose columns form a basis for null(A). Then $x = \bar{x} + Zv$, with a one-to-one correspondence between x and v.

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

$$(CP'_=) \quad \underset{v \in I\!R^{n-m}}{\text{minimize}} \quad \varphi(v).$$

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

$Z^T \nabla f(x)$ is called the *reduced gradient* to f in x.

$Z^T \nabla^2 f(x) Z$ is called the *reduced Hessian* to f in x.

First-order optimality conditions as a system of equations, cont.

The resulting Newton system may equivalently be written as

$$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -(\lambda + \nu) \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -(Ax - b) \end{pmatrix},$$

alternatively

$$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -(Ax - b) \end{pmatrix}.$$

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

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

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, where Z is a matrix whose columns form a basis for null(A).

Newton iteration and equality-constrained quadratic program

Compare $\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -(Ax - b) \end{pmatrix}$

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

Identify:
$$\begin{array}{ccc} \nabla^2 f(x) & \longleftrightarrow & H \\ \nabla f(x) & \longleftrightarrow & c \\ A & \longleftrightarrow & A \\ -(Ax - b) & \longleftrightarrow & b. \end{array}$$

Newton iteration as a QP problem

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

$$
(QP_=) \qquad \begin{array}{ll} \text{minimize} & \frac{1}{2}p^T\nabla^2 f(x)p + \nabla f(x)^T p \\ \text{subject to} & Ap = -(Ax - b), \\ & p \in I\!R^n, \end{array}
$$

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 $Z^T\nabla^2 f(x)Z \succ 0$ and A has full row rank, where Z is a matrix whose columns form a basis for null(A).

An SQP iteration for problems with equality constraints

Given x, λ such that $Z^T\nabla^2 f(x)Z \succ 0$ and A has full row rank, a Newton iteration takes the following form.

- Compute optimal solution p and multiplier vector λ^+ to

$$
(QP_=) \qquad \begin{array}{ll} \text{minimize} & \frac{1}{2}p^T\nabla^2 f(x)p + \nabla f(x)^T p \\ \text{subject to} & Ap = -(Ax - b), \\ & p \in I\!R^n, \end{array}
$$

- $x \leftarrow x + p, \qquad \lambda \leftarrow \lambda^+.$

We call this method *sequential quadratic programming* (SQP).

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

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

Theorem. *Assume that* $f \in C^3$ *is convex on* R^n *and that* $A \in I\!R^{m\times n}$ *has full row rank. Further, assume that* x^* *is a minimizer of* $(CP_=)$ *such that* $Z^T\nabla^2 f(x)Z \succeq 0$*, where* Z *is a matrix whose columns form a basis for* null(A)*. If the SQP method (with steplength one) is started at a point sufficiently close to* x^**,* λ^**, then it is well defined and converges to* x^**,* λ^* *with convergence rate at least two.*

Proof. In a neighborhood of x^*, λ^* it holds that $Z^T\nabla^2 f(x)Z \succ 0$ and $\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix}$ is nonsingular. The subproblem $(QP_=)$ is hence well defined and the result follows from the quadratic rate of convergence of Newton's method. $\square$