Equality constrained minimization

Lecture 13, Convex Optimization

National Taiwan University

June 3, 2021

Table of contents

Algorithms for Convex Optimization Problems (1/2)

- Our goal: learn algorithms that solve convex optimization problems efficiently:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, 2, ..., m \\ & h_i(x) = 0, \quad i = 1, 2, ..., p \end{array}$$

where $f_i(x), 0 \leq i \leq m$ are convex functions and $h_i(x), 1 \leq i \leq p$ are affine functions.

Algorithms for Convex Optimization Problems (2/2)

- We have learned several descent methods for unconstrained convex optimization problems:

$$\text{minimize} \quad f_0(x)$$

  - In particular, the Newton's method has the best convergence properties among them.
- In the following, we will study methods for solving a convex optimization problem with equality constraints, including an extension of the Newton's method.

Equality Constrained Minimization Problems

- We will describe methods for solving a convex optimization problem with equality constraints,

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

  where $f : \mathbf{R}^n \to \mathbf{R}$ is convex and twice continuously differentiable, and $A \in \mathbf{R}^{p \times n}$ with rank $A = p < n$.
- We assume that an optimal solution x^* exists, and use p^* to denote the optimal value,

$$p^* = \inf \{f(x) \mid Ax = b\} = f(x^*).$$

Eliminating equality constraints (1/2)

- One general approach to solving the equality constrained problem is to eliminate the equality constraints, and then solve the resulting unconstrained problem using methods for unconstrained minimization.
- We first find a matrix $F \in \mathbf{R}^{n\times(n-p)}$ and vector $\hat{x} \in \mathbf{R}^n$ that parametrize the (affine) feasible set:

$$\{x \mid Ax = b\} = \left\{Fz + \hat{x} \mid z \in \mathbf{R}^{n-p}\right\}.$$

Eliminating equality constraints (2/2)

- Here $\hat{x}$ can be chosen as any particular solution of $Ax = b$, and $F \in \mathbf{R}^{n\times(n-p)}$ is any matrix whose range is the nullspace of A. We then form the reduced or eliminated optimization problem

$$\text{minimize} \quad \tilde{f}(z) = f(Fz + \hat{x}),$$

which is an unconstrained problem with variable $z \in \mathbf{R}^{n-p}$.

- From its solution z^*, we can find the solution of the equality constrained problem as $x^* = Fz^* + \hat{x}$.

Example – Optimal allocation with resource constraint (1/2)

- We consider the problem

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} f_i(x_i) \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = b, \end{aligned}$$

  where the functions $f_i : \mathbf{R} \to \mathbf{R}$ are convex and twice differentiable, and $b \in \mathbf{R}$ is a problem parameter.
- We interpret this as the problem of optimally allocating a single resource, with a fixed total amount b (the budget) to n otherwise independent activities.

Example – Optimal allocation with resource constraint (2/2)

- We can eliminate x_n using the parametrization $x_n = b - x_1 - \ldots - x_{n-1}$, which corresponds to the choices

$$\hat{x} = be_n, \quad F = \begin{bmatrix} I_{n-1} \\ -\mathbf{1}^T \end{bmatrix} \in \mathbf{R}^{n \times (n-1)}.$$

- The reduced problem is then

$$\text{minimize} \quad f_n(b - x_1 - \ldots - x_{n-1}) + \sum_{i=1}^{n-1} f_i(x_i),$$

with variables $x_1, \ldots, x_{n-1}$.

Choice of elimination matrix (1/2)

- There are many possible choices for the elimination matrix F, which can be chosen as any matrix in $\mathbf{R}^{n\times(n-p)}$ with $\mathcal{R}(F) = \mathcal{N}(A)$. [1]
- If F is one such matrix, and $T \in \mathbf{R}^{(n-p)\times(n-p)}$ is nonsingular, then $\tilde{F} = FT$ is also a suitable elimination matrix, since

$$\mathcal{R}(\tilde{F}) = \mathcal{R}(F) = \mathcal{N}(A).$$

- Conversely, if F and $\tilde{F}$ are any two suitable elimination matrices, then there is some nonsingular T such that $\tilde{F} = FT$.

[1] The notations $\mathcal{R}(\cdot)$ and $\mathcal{N}(\cdot)$ denote the range space (i.e., column space) and the null space, respectively, of a matrix.

Choice of elimination matrix (2/2)

- If we eliminate the equality constraints using F, we solve the unconstrained problem

$$\text{minimize } f(Fz + \hat{x}),$$

while if $\tilde{F}$ is used, we solve the unconstrained problem

$$\text{minimize } f(\tilde{F}\tilde{z} + \hat{x}) = f(F(T\tilde{z}) + \hat{x}).$$

- This problem is equivalent to the one above, and is simply obtained by the change of coordinates $z = T\tilde{z}$.
- In other words, changing the elimination matrix can be thought of as changing variables in the reduced problem.

Solving the KKT Conditions

- From KKT optimality conditions, a point $x^* \in \textbf{dom } f$ is optimal for the problem if and only if there is a $\nu^* \in \mathbf{R}^p$ such that

$$Ax^* = b, \quad \nabla f(x^*) + A^T \nu^* = 0,$$

which is a set of $n + p$ equations in the $n + p$ variables x^*, ν^*.
- The first set of equations, $Ax^* = b$, are called the **primal feasibility equations**, which are linear.
- The second set of equations, $\nabla f(x^*) + A^T \nu^* = 0$, are called the **dual feasibility equations**, and are in general nonlinear.
- We consider first the special case that these equations are also linear, that is, $\nabla f(x) = Px + q$ for some $P \in \mathbf{S}^n_+$ and $q \in \mathbf{R}^n$.

Equality constrained convex quadratic minimization (1/2)

- Consider the equality constrained convex quadratic minimization problem

$$\begin{array}{ll} \text{minimize} & f(x) = (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b, \end{array}$$

  where $P \in \mathbf{S}_+^n$ and $A \in \mathbf{R}^{p \times n}$.
- This problem is important on its own, and also because it forms the basis for an extension of Newton's method to equality constrained problems.

Equality constrained convex quadratic minimization (2/2)

- Here the optimality conditions are

$$Ax^* = b, \quad Px^* + q + A^T\nu^* = 0,$$

which we can write as

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

- This set of $n+p$ linear equations in the $n+p$ variables x^*, ν^* is called the **KKT system** for the equality constrained quadratic optimization problem.
- The coefficient matrix is called the **KKT matrix**.

Singularity of the KKT matrix (1/2)

- When the KKT matrix

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$$

is nonsingular, there is a unique optimal primal-dual pair (x^*, ν^*).

- If the KKT matrix is singular, but the KKT system

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

is solvable, any solution yields an optimal pair (x^*, ν^*).

- If the KKT system is not solvable, the quadratic optimization problem is unbounded below or infeasible.

Singularity of the KKT matrix (2/2)

- In this case there exist $v \in \mathbf{R}^n$ and $w \in \mathbf{R}^p$ such that

$$Pv + A^T w = 0, \quad Av = 0, \quad -q^T v + b^T w > 0.$$

- Let $\hat{x}$ be any feasible point. Then, the point $x = \hat{x} + tv$ is feasible for all t and [2]

$$\begin{aligned} f(\hat{x} + tv) &= f(\hat{x}) + t(v^T P\hat{x} + q^T v) + (1/2)t^2 v^T Pv \\ &= f(\hat{x}) + t(-\hat{x}^T A^T w + q^T v) - (1/2)t^2 w^T Av \\ &= f(\hat{x}) + t(-b^T w + q^T v), \end{aligned}$$

which decreases without bound as $t \to \infty$.

[2] Note: $f(x) = (1/2)x^T Px + q^T x + r$

Conditions on Nonsingularity of the KKT matrix

- Recall our assumption that $P \in \mathbf{S}_+^n$ and rank $A = p < n$. There are several conditions equivalent to nonsingularity of the KKT matrix:
  - $\mathcal{N}(P) \bigcap \mathcal{N}(A) = \{0\}$, i.e., P and A have no nontrivial common nullspace.
  - $Ax = 0$, $x \neq 0 \Longrightarrow x^T Px > 0$, i.e., P is positive definite on the nullspace of A.
  - $F^T PF \succ 0$, where $F \in \mathbf{R}^{n \times (n-p)}$ is a matrix for which $\mathcal{R}(F) = \mathcal{N}(A)$.
- As an important special case, we note that if $P \succ 0$, the KKT matrix must be nonsingular.

Newton's method with equality constraints

- In this section we describe an extension of Newton's method to include equality constraints.
- The method is almost the same as Newton's method without constraints, except for two differences:
  1. The initial point must be feasible (i.e., satisfy $x \in \textbf{dom } f$ and $Ax = b$).
  2. The definition of Newton step is modified to take the equality constraints into account.
- In particular, we make sure that the Newton step Δx_{nt} satisfies $A\Delta x_{\mathrm{nt}} = 0$. We say that $v \in \mathbf{R}^n$ is a **feasible direction** if $Av = 0$. This means that the Newton step Δx_{nt} is a feasible direction.

Newton Step Defined via 2nd-Order Approximation (1/3)

- To derive the Newton step Δx_{nt} for the equality constrained problem

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

at the feasible point x, we replace the objective with its second-order Taylor approximation (denoted $\hat{f}(x)$) near x, to form the problem

$$\begin{array}{ll} \text{minimize} & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b, \end{array}$$

with variable v, which is a (convex) quadratic minimization problem with equality constraints, and can be solved analytically.

Newton Step Defined via 2nd-Order Approximation (2/3)

- We define Δx_{nt}, the Newton step at x, as the solution of the convex quadratic problem

$$\begin{array}{ll} \text{minimize} & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b, \end{array}$$

  assuming the associated KKT matrix is nonsingular.
- Therefore, the Newton step Δx_{nt} is characterized by

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix},$$

  where w is the associated optimal dual variable for the quadratic problem.

Newton Step Defined via 2nd-Order Approximation (3/3)

- The Newton step Δx_{nt} is what must be added to x to solve the problem when the quadratic approximation is used in place of f.
- As in Newton's method for unconstrained problems, we observe that when the objective f is exactly quadratic, the Newton update $x + \Delta x_{\mathrm{nt}}$ exactly solves the equality constrained minimization problem, and in this case the vector w is the optimal dual variable for the original problem.
- This suggests, as in the unconstrained case, that when f is nearly quadratic, $x + \Delta x_{\mathrm{nt}}$ should be a very good estimate of the solution x^*, and w should be a good estimate of the optimal dual variable ν^*.

The Newton decrement (1/3)

- We define the Newton decrement for the equality constrained problem as
$$\lambda(x) = (\Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}})^{1/2},$$
which is exactly the same expression as in the unconstrained case.
- The Newton decrement $\lambda(x)$ is the norm of the Newton step, in the norm determined by the Hessian $\nabla^2 f(x)$, i.e.,
$$\lambda(x) = ||\Delta x_{\mathrm{nt}}||_{\nabla^2 f(x)}.$$

The Newton decrement (2/3)

- Let

$$\hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v$$

be the second-order Taylor approximation of f at x. The difference between $f(x)$ and the minimum of the second-order model satisfies

$$f(x) - \inf \left\{ \hat{f}(x+v) \mid A(x+v) = b \right\} = \lambda(x)^2/2,$$

exactly as in the unconstrained case.

The Newton decrement (3/3)

- This means that, as in the unconstrained case, $\lambda(x)^2/2$ gives an estimate of $f(x) - p^*$, based on the quadratic model at x, and also that $\lambda(x)$ (or a multiple of $\lambda(x)^2$) serves as the basis of a good stopping criterion.
- The Newton decrement comes up in the line search as well, since the directional derivative of f in the direction Δx_{nt} is

$$\frac{d}{dt} f(x + t\Delta x_{\mathrm{nt}})\bigg|_{t=0} = \nabla f(x)^T \Delta x_{\mathrm{nt}} = -\lambda(x)^2,$$

as in the unconstrained case.

Feasible descent direction

- Suppose that $Ax = b$. Recall that $v \in \mathbf{R}^n$ is a feasible direction if $Av = 0$.
- In this case, every point of the form $x + tv$ is also feasible, i.e., $A(x + tv) = b$.
- We say that v is a descent direction for f at x, if for small $t > 0$, $f(x + tv) < f(x)$.
- The Newton step is always a feasible descent direction (except when x is optimal, in which case $\Delta x_{\mathrm{nt}} = 0$).
- Indeed, the second set of equations that define Δx_{nt} are $A\Delta x_{\mathrm{nt}} = 0$, which shows it is a feasible direction; that it is a descent direction follows from

$$\frac{d}{dt} f(x + t\Delta x_{\mathrm{nt}}) \bigg|_{t=0} = \nabla f(x)^T \Delta x_{\mathrm{nt}} = -\lambda(x)^2.$$

Newton's method for equality constrained minimization

- **Algorithm 10.1** Newton's method for equality constrained minimization.

  given starting point $x \in \mathbf{dom}\ f$ with $Ax = b$, tolerance $\epsilon > 0$.

  repeat

  1. Compute the Newton step and decrement $\Delta x_{\mathrm{nt}}, \lambda(x)$.
  2. Stopping criterion. **quit** if $\lambda^2/2 \leq \epsilon$.
  3. Line search. Choose step size t by backtracking line search.
  4. Update. $x := x + t\Delta x_{\mathrm{nt}}$.

- The method is called a **feasible descent method**, since all the iterates are feasible, with $f(x^{(k+1)}) < f(x^{(k)})$ (unless $x^{(k)}$ is optimal).
- Newton's method requires that the KKT matrix be invertible at each x.

Newton's method and elimination (1/5)

- It can be shown that the iterates in Newton's method for the equality constrained problem

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

coincide with the iterates in Newton's method applied to the reduced problem

$$\text{minimize} \quad \tilde{f}(z) = f(Fz + \hat{x}),$$

- Suppose F satisfies $\mathcal{R}(F) = \mathcal{N}(A)$ and rank $F = n - p$, and $\hat{x}$ satisfies $A\hat{x} = b$.

Newton's method and elimination (2/5)

- The gradient and Hessian of the reduced objective function $\tilde{f}(z) = f(Fz + \hat{x})$ are

$$\nabla \tilde{f}(z) = F^T \nabla f(Fz + \hat{x}), \quad \nabla^2 \tilde{f}(z) = F^T \nabla^2 f(Fz + \hat{x}) F.$$

- From the Hessian expression, we see that the Newton step for the equality constrained problem is defined, i.e., the KKT matrix

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}$$

is invertible, if and only if the Newton step for the reduced problem is defined, i.e., $\nabla^2 \tilde{f}(z)$ is invertible.

Newton's method and elimination (3/5)

- The Newton step for the reduced problem is

$$\Delta z_{\mathrm{nt}} = -\nabla^2 \tilde{f}(z)^{-1} \nabla \tilde{f}(z) = -(F^T \nabla^2 f(x) F)^{-1} F^T \nabla f(x),$$

where $x = Fz + \hat{x}$.

- This search direction for the reduced problem corresponds to the direction

$$F \Delta z_{\mathrm{nt}} = -F(F^T \nabla^2 f(x) F)^{-1} F^T \nabla f(x)$$

for the original, equality constrained problem.

- We claim that $\Delta x_{\mathrm{nt}} = F \Delta z_{\mathrm{nt}}$.

Newton's method and elimination (4/5)

- To show this, we take $\Delta x_{\mathrm{nt}} = F\Delta z_{\mathrm{nt}}$, choose $w = -(AA^T)^{-1}A(\nabla f(x) + \nabla^2 f(x)\Delta x_{\mathrm{nt}})$, and verify that the equations defining the Newton step,

$$\nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w + \nabla f(x) = 0, \quad A\Delta x_{\mathrm{nt}} = 0,$$

  hold.

- The second equation, $A\Delta x_{\mathrm{nt}} = 0$, is satisfied because $AF = 0$. To verify the first equation, we observe that

$$\begin{aligned} & \begin{bmatrix} F^T \\ A \end{bmatrix} \left(\nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w + \nabla f(x) \right) \\ = & \begin{bmatrix} F^T \nabla^2 f(x)\Delta x_{\mathrm{nt}} + F^T A^T w + F^T \nabla f(x) \\ A\nabla^2 f(x)\Delta x_{\mathrm{nt}} + AA^T w + A\nabla f(x) \end{bmatrix} \\ = & \ 0. \end{aligned}$$

Newton's method and elimination (5/5)

- Since the matrix on the left of the first line is nonsingular, we conclude that the conditions

$$\nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w + \nabla f(x) = 0, \quad A\Delta x_{\mathrm{nt}} = 0,$$

  hold.
- In a similar way, the Newton decrement $\tilde{\lambda}(z)$ of $\tilde{f}$ at z and the Newton decrement of f at x turn out to be equal:

$$\begin{aligned} \tilde{\lambda}(z)^2 &= \Delta z_{\mathrm{nt}}^T \nabla^2 \tilde{f}(z) \Delta z_{\mathrm{nt}} \\ &= \Delta z_{\mathrm{nt}}^T F^T \nabla^2 f(x) F \Delta z_{\mathrm{nt}} \\ &= \Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}} \\ &= \lambda(x)^2. \end{aligned}$$