

ET4350
Applied Convex Optimization
Lecture 6

Equality constrained minimization

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

- f convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\text{rank } A = p$
- we assume p^* is finite and attained

optimality conditions: x^* is optimal iff there exists a ν^* such that

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

equality constrained quadratic minimization (with $P \in \mathbf{S}_+^n$)

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

optimality condition:

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

- coefficient matrix is called KKT matrix
- KKT matrix is nonsingular if and only if

$$Ax = 0, \quad x \neq 0 \quad \implies \quad x^T P x > 0$$

- equivalent condition for nonsingularity: $P + A^T A \succ 0$

Eliminating equality constraints

represent solution of $\{x \mid Ax = b\}$ as

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

- \hat{x} is (any) particular solution
- range of $F \in \mathbf{R}^{n \times (n-p)}$ is nullspace of A ($\text{rank } F = n - p$ and $AF = 0$)

reduced or eliminated problem

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

- an unconstrained problem with variable $z \in \mathbf{R}^{n-p}$
- from solution z^* , obtain x^* and ν^* as

$$x^* = Fz^* + \hat{x}, \quad \nu^* = -(AA^T)^{-1}A\nabla f(x^*)$$

example: optimal allocation with resource constraint

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) \\ \text{subject to} & x_1 + x_2 + \cdots + x_n = b \end{array}$$

eliminate $x_n = b - x_1 - \cdots - x_{n-1}$, *i.e.*, choose

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

reduced problem:

$$\text{minimize} \quad f_1(x_1) + \cdots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \cdots - x_{n-1})$$

(variables x_1, \dots, x_{n-1})

Newton step

Newton step Δx_{nt} of f at feasible x is given by solution v of

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

interpretations

- Δx_{nt} solves second order approximation (with variable v)

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

- Δx_{nt} equations follow from linearizing optimality conditions

$$\nabla f(x+v) + A^T w \approx \nabla f(x) + \nabla^2 f(x)v + A^T w = 0, \quad A(x+v) = b$$

Newton's method with equality constraints

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

repeat

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

- a feasible descent method: $x^{(k)}$ feasible and $f(x^{(k+1)}) < f(x^{(k)})$
- affine invariant

Newton's method and elimination

Newton's method for reduced problem

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

- variables $z \in \mathbf{R}^{n-p}$
- \hat{x} satisfies $A\hat{x} = b$; $\text{rank } F = n - p$ and $AF = 0$
- Newton's method for \tilde{f} , started at $z^{(0)}$, generates iterates $z^{(k)}$

Newton's method with equality constraints

when started at $x^{(0)} = Fz^{(0)} + \hat{x}$, iterates are

$$x^{(k+1)} = Fz^{(k)} + \hat{x}$$

hence, don't need separate convergence analysis

Solving KKT systems

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$

solution methods

- LDL^T factorization
- elimination (if H nonsingular)

$$AH^{-1}A^T w = h - AH^{-1}g, \quad Hv = -(g + A^T w)$$

- elimination with singular H : write as

$$\begin{bmatrix} H + A^T Q A & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g + A^T Q h \\ h \end{bmatrix}$$

with $Q \succeq 0$ for which $H + A^T Q A \succ 0$, and apply elimination

Network flow optimization

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n \phi_i(x_i) \\ \text{subject to} & Ax = b \end{array}$$

- directed graph with n arcs, $p + 1$ nodes
- x_i : flow through arc i ; ϕ_i : cost flow function for arc i (with $\phi_i''(x) > 0$)
- node-incidence matrix $\tilde{A} \in \mathbf{R}^{(p+1) \times n}$ defined as

$$\tilde{A}_{ij} = \begin{cases} 1 & \text{arc } j \text{ leaves node } i \\ -1 & \text{arc } j \text{ enters node } i \\ 0 & \text{otherwise} \end{cases}$$

- reduced node-incidence matrix $A \in \mathbf{R}^{p \times n}$ is \tilde{A} with last row removed
- $b \in \mathbf{R}^p$ is (reduced) source vector
- $\text{rank } A = p$ if graph is connected

KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$

- $H = \text{diag}(\phi_1''(x_1), \dots, \phi_n''(x_n))$, positive diagonal
- solve via elimination:

$$AH^{-1}A^T w = h - AH^{-1}g, \quad Hv = -(g + A^T w)$$

sparsity pattern of coefficient matrix is given by graph connectivity

$$\begin{aligned} (AH^{-1}A^T)_{ij} \neq 0 &\iff (AA^T)_{ij} \neq 0 \\ &\iff \text{nodes } i \text{ and } j \text{ are connected by an arc} \end{aligned}$$

Inequality constrained minimization

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned} \tag{1}$$

- f_i convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\text{rank } A = p$
- we assume p^* is finite and attained
- we assume problem is strictly feasible: there exists \tilde{x} with

$$\tilde{x} \in \text{dom } f_0, \quad f_i(\tilde{x}) < 0, \quad i = 1, \dots, m, \quad A\tilde{x} = b$$

hence, strong duality holds and dual optimum is attained

Examples

- LP, QP, QCQP, GP
- entropy maximization with linear inequality constraints

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Fx \preceq g \\ & Ax = b \end{array}$$

with $\text{dom } f_0 = \mathbf{R}_{++}^n$

- differentiability may require reformulating the problem, *e.g.*, piecewise-linear minimization or ℓ_∞ -norm approximation via LP
- SDPs and SOCPs are better handled as problems with generalized inequalities (see later)

Logarithmic barrier

reformulation of (1) via indicator function:

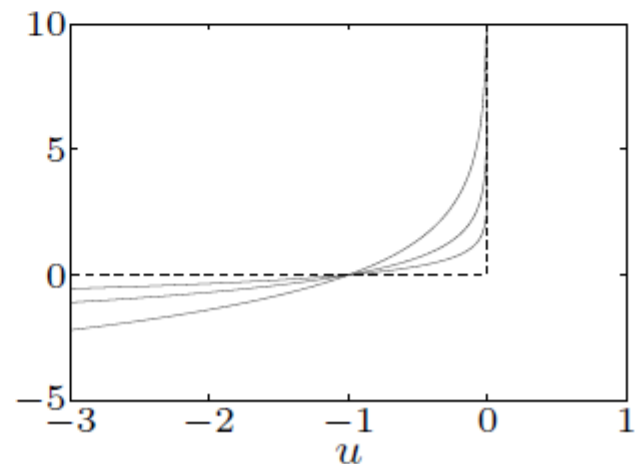
$$\begin{aligned} & \text{minimize} && f_0(x) + \sum_{i=1}^m I_-(f_i(x)) \\ & \text{subject to} && Ax = b \end{aligned}$$

where $I_-(u) = 0$ if $u \leq 0$, $I_-(u) = \infty$ otherwise (indicator function of \mathbf{R}_-)

approximation via logarithmic barrier

$$\begin{aligned} & \text{minimize} && f_0(x) - (1/t) \sum_{i=1}^m \log(-f_i(x)) \\ & \text{subject to} && Ax = b \end{aligned}$$

- an equality constrained problem
- for $t > 0$, $-(1/t) \log(-u)$ is a smooth approximation of I_-
- approximation improves as $t \rightarrow \infty$



logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^m \log(-f_i(x)), \quad \mathbf{dom} \phi = \{x \mid f_1(x) < 0, \dots, f_m(x) < 0\}$$

- convex (follows from composition rules)
- twice continuously differentiable, with derivatives

$$\nabla \phi(x) = \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x)$$

$$\nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla^2 f_i(x)$$

Central path

- for $t > 0$, define $x^*(t)$ as the solution of

$$\begin{array}{ll} \text{minimize} & t f_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

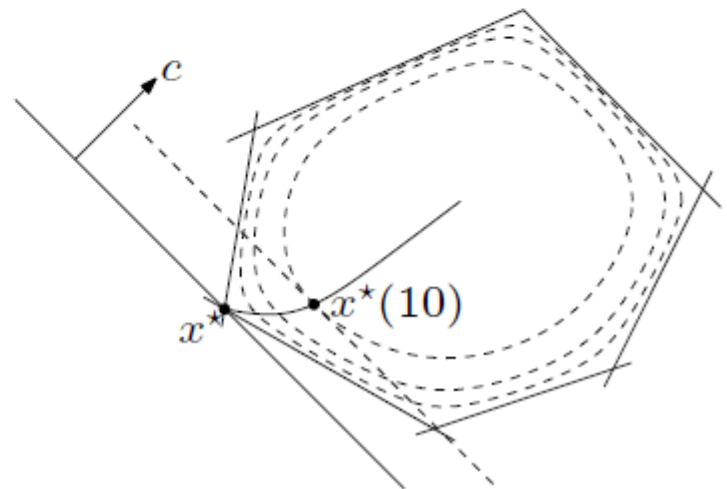
(for now, assume $x^*(t)$ exists and is unique for each $t > 0$)

- central path is $\{x^*(t) \mid t > 0\}$

example: central path for an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \dots, 6 \end{array}$$

hyperplane $c^T x = c^T x^*(t)$ is tangent to level curve of ϕ through $x^*(t)$



Interpretation via KKT conditions

$x = x^*(t)$, $\lambda = \lambda^*(t)$, $\nu = \nu^*(t)$ satisfy

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \dots, m$, $Ax = b$
2. dual constraints: $\lambda \succeq 0$
3. approximate complementary slackness: $-\lambda_i f_i(x) = 1/t$, $i = 1, \dots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T \nu = 0$$

difference with KKT is that condition 3 replaces $\lambda_i f_i(x) = 0$

Force field interpretation

centering problem (for problem with no equality constraints)

$$\text{minimize } tf_0(x) - \sum_{i=1}^m \log(-f_i(x))$$

force field interpretation

- $tf_0(x)$ is potential of force field $F_0(x) = -t\nabla f_0(x)$
- $-\log(-f_i(x))$ is potential of force field $F_i(x) = (1/f_i(x))\nabla f_i(x)$

the forces balance at $x^*(t)$:

$$F_0(x^*(t)) + \sum_{i=1}^m F_i(x^*(t)) = 0$$

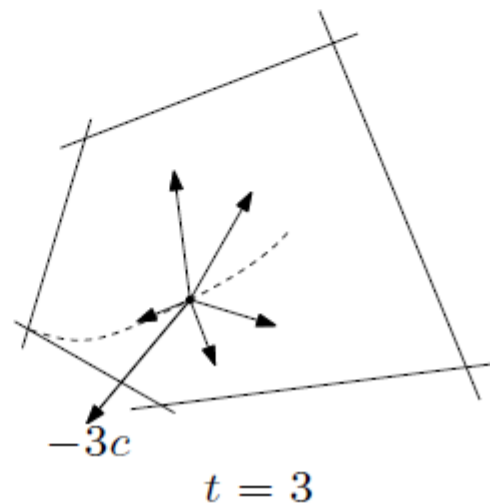
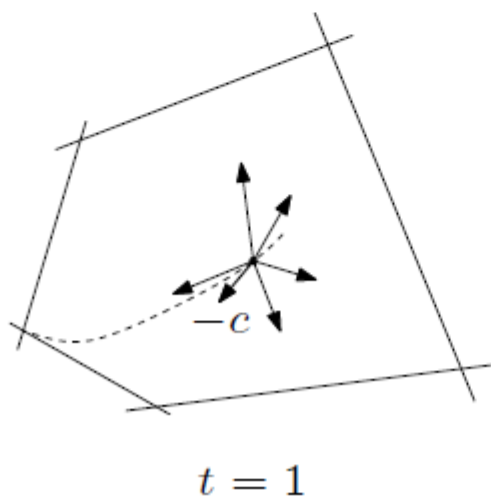
example

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && a_i^T x \leq b_i, \quad i = 1, \dots, m \end{aligned}$$

- objective force field is constant: $F_0(x) = -tc$
- constraint force field decays as inverse distance to constraint hyperplane:

$$F_i(x) = \frac{-a_i}{b_i - a_i^T x}, \quad \|F_i(x)\|_2 = \frac{1}{\text{dist}(x, \mathcal{H}_i)}$$

where $\mathcal{H}_i = \{x \mid a_i^T x = b_i\}$



Barrier method

given strictly feasible x , $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$.

repeat

1. *Centering step.* Compute $x^*(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$.
 2. *Update.* $x := x^*(t)$.
 3. *Stopping criterion.* **quit** if $m/t < \epsilon$.
 4. *Increase t .* $t := \mu t$.
-

- terminates with $f_0(x) - p^* \leq \epsilon$ (stopping criterion follows from $f_0(x^*(t)) - p^* \leq m/t$)
- centering usually done using Newton's method, starting at current x
- choice of μ involves a trade-off: large μ means fewer outer iterations, more inner (Newton) iterations; typical values: $\mu = 10\text{--}20$
- several heuristics for choice of $t^{(0)}$