Lecture 16

Interior-Point Method

October 27, 2008

Outline

- Review of Self-concordance
- Overview of Newton's Methods for Equality Constrained Minimization
- Examples
- Interior-Point Method
    - Inequality constrained minimization
    - Logarithmic barrier function and central path
    - Barrier method
    - Feasibility and phase I methods
    - Complexity analysis via self-concordance

Equality Constrained Minimization

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

KKT Optimality Conditions imply that x^* is optimal if and only if there exists a λ^* such that $\quad Ax^* = b, \quad \nabla f(x^*) + A^T\lambda^* = 0$

- Newton's Method solves KKT conditions

$$\begin{bmatrix} \nabla^2 f(x_k) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d_k \\ w_k \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) \\ h_k \end{bmatrix}$$

where

- Feasible point method uses $h_k = 0$
- Infeasible point method uses $h_k = Ax_k - b$

with w_k being dual optimal for the minimization of the quadratic approximation of f at x_k

Equality Constrained Analytic Centering

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

Feasible point Newton's method: $g = \nabla f(x)$, $H = \nabla^2 f(x)$

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} d \\ w \end{bmatrix} = \begin{bmatrix} -g \\ 0 \end{bmatrix}, \quad g = \begin{bmatrix} -\frac{1}{x_1} \\ \vdots \\ -\frac{1}{x_n} \end{bmatrix}, \quad H = \operatorname{diag}\left[\frac{1}{x_1^2}, \ldots, \frac{1}{x_n^2}\right]$$

- The Hessian is positive definite
- KKT matrix first row: $Hd + A^T w = -g \;\Rightarrow\; d = -H^{-1}(g + A^T w)$ (1)
- KKT matrix second row, $Ad = 0$, and Eq. (1) $\Rightarrow\; AH^{-1}(g + A^T w) = 0$
- The matrix A has full row rank, thus $AH^{-1}A^T$ is invertible, hence

$$w = -\left(AH^{-1}A^T\right)^{-1} AH^{-1}g, \qquad H^{-1} = \operatorname{diag}\left[x_1^2, \ldots, x_n^2\right]$$

- The matrix $-AH^{-1}A^T$ is known as *Schur complement of H* (any H)

Network Flow Optimization

$$\begin{array}{ll} \text{minimize} & \sum_{l=1}^{n} \phi_l(x_l) \\ \text{subject to} & Ax = b \end{array}$$

- Directed graph with n arcs and $p + 1$ nodes
- Variable x_l: flow through arc l
- Cost ϕ_l: cost flow function for arc l, with $\phi_l''(t) > 0$
- Node-incidence matrix $\tilde{A} \in \mathbb{R}^{(p+1)\times n}$ defined as

$$\tilde{A}_{il} = \begin{cases} 1 & \text{arc } l \text{ originates at node } i \\ -1 & \text{arc } l \text{ ends at node } i \\ 0 & \text{otherwise} \end{cases}$$

- Reduced node-incidence matrix $A \in \mathbb{R}^{p\times n}$ is $\tilde{A}$ with last row removed
- $b \in \mathbb{R}^p$ is (reduced) source vector
- Rank $A = p$ when the graph is *connected*

KKT system for infeasible Newton's method

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

where $h = Ax - b$ is a measure of infeasibility at the current point x

- $g = [\phi_1'(x_1), \ldots, \phi_n(x_n)]^T$
- $H = \mathrm{diag}\,[\phi_1''(x_1), \ldots, \phi_n''(x_n)]$ with positive diagonal entries
- Solve via elimination:

$$w = (AH^{-1}A^T)^{-1}[h - AH^{-1}g], \qquad d = -H^{-1}(g + A^T w)$$

Interior-Point Methods

For solving inequality constrained problems of the form

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \le 0, \quad j = 1, \ldots, m \\ & Ax = b \end{aligned}$$

- The interior-point methods have been extensively studied since early 60's as a sub-class of penalty methods [penalizing the convex inequality constraints $g_j(x)$, $j = 1, \ldots, m$]
- Log-barrier is just one of the choices for penalty (convergence established by Fiacco and McCormick in 1965)
- Newton's method combined with Log-barrier penalty was analyzed by Karmarkar SSSR in 1984
    - As a new polynomial-time algorithm for linear programming problems
- Nesterov and Nemirovski in 1994 provided a new analysis of Karmarkar's approach for general convex optimization problem

Inequality Constrained Minimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_j(x) \le 0, \quad j = 1, \ldots, m \\ & Ax = b \end{array}$$

- **Assumptions**:
  - The functions f and g_j are convex and twice continuously differentiable (their domains are open)
  - The matrix $A \in \mathbb{R}^{p \times n}$ has rank p
  - The optimal value f^* is finite and attained
  - No duality gap and a dual optimal $(\mu^*, \lambda^*) \in \mathbb{R}^m \times \mathbb{R}^p$ exists $\tilde{x}$ with

$$\begin{array}{llll} g_j(\bar{x}) < 0, & j = 1, \ldots, m, & A\bar{x} = b, & \bar{x} \in \mathrm{dom} f, \\ g_j(\tilde{x}) \le 0, & j = 1, \ldots, m, & A\tilde{x} = b & \tilde{x} \in \mathrm{relint} dom f \end{array}$$

KKT Conditions

For the inequality constrained problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_j(x) \leq 0, \quad j = 1, \ldots, m \\ & Ax = b \end{array}$$

Under the assumption just stated, x^* is primal optimal if and only if there exist μ^* and λ^* such that the following KKT conditions hold:

- Primal feasibility: $Ax^* = b, \quad g_j(x^*) \leq 0$ for all j
- Dual feasibility: $\mu^* \succeq 0$
- Lagrangian optimality in x: $\nabla f(x^*) + \sum_{j=1}^m \mu_j^* \nabla g_j(x^*) + A^T \lambda^* = 0$
- Complementarity slackness: $\mu_j^* g_j(x^*) = 0$ for all j

The interior-point method solves these conditions

Our focus is on the *barrier* type method

Logarithmic Barrier Function

Based on reformulation of the constrained problem via indicator function:

$$\begin{array}{ll} \text{minimize} & f(x) + \sum_{j=1}^{m} I_{-}(g_j(x)) \\ \text{subject to} & Ax = b \end{array}$$

where I_{-} **is the indicator function of nonpositive reals**:

$$I_{-}(u) = 0 \text{ when } u \le 0, \quad \text{and} \quad I_{-}(u) = \infty \text{ otherwise}$$

- Consider a point-wise **approximation of I_{-} by a logarithmic barrier**
$$\phi_t(u) = -\frac{1}{t}\ln(-u) \quad \text{with } t > 0$$
- Thus: $\lim_{t\to\infty} \phi_t(u) = 0$ for $u < 0$, $\lim_{t\to\infty} \phi_t(u) = +\infty$ otherwise
- With $t > 0$, $\phi_t(u) = -\frac{1}{t}\ln(-u)$ is a smooth approximation of I_{-}, improving as $t \to \infty$
- Using this approximation, we have a family of equality constrained problems
$$\begin{array}{ll} \text{minimize} & f(x) - \frac{1}{t}\sum_{j=1}^{m} \ln(-g_j(x)) \\ \text{subject to} & Ax = b \end{array}$$

Family of Equality Constrained Problems

Consider the problems with $t > 0$:

$$\begin{array}{ll} \text{minimize} & f(x) - \frac{1}{t}\sum_{j=1}^{m} \ln(-g_j(x)) \\ \text{subject to} & Ax = b \end{array}$$

- Can be viewed as a sequence of penalized problems approximating the original problem
- The (penalty) parameter t drives the approximation accuracy: as t increases the approximation is more accurate
- Each of the approximate problems is convex and differentiable
- The function $\phi(x) = -\sum_{j=1}^{m} \ln(-g_j(x))$ is referred to as **logarithmic barrier** or **log barrier**
- Its domain is $\mathrm{dom}\phi = \{x \in \mathbb{R}^n \mid g_j(x) < 0,\ j = 1, \ldots, m\}$

Central Path

An equivalent family of problems:

$$\begin{array}{ll} \text{minimize} & tf(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array} \tag{1}$$

with $\phi(x) = -\sum_{j=1}^{m} \ln(-g_j(x))$

- The problem (1) has the same minimizers as when divided by t
- Assume a unique optimal $x^*(t)$ exists for each $t > 0$. When is this true?
- The set $\{x^*(t) \mid t > 0\}$ is referred to as **central path** for problem (1)
- Each point on the central path is characterized by the following (necessary and sufficient conditions):
  - Strict feasibility: $Ax^*(t) = b$ and $g_j(x^*(t)) < 0$ for all j
  - There exists $\hat{\lambda}(t)$ such that: $t\nabla f(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{\lambda}(t) = 0$ where

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

Important Property of Central Path

For an x with $g(x) \prec 0$, we have $x = x^*(t)$ if and only if there exists a $\hat{\lambda}(t)$ such that

$$\nabla f(x) + \sum_{j=1}^{m} \frac{1}{-tg_j(x)} \nabla g_j(x) + \frac{1}{t} A^T \hat{\lambda}(t) = 0, \qquad Ax = b$$

- Therefore, $x^*(t)$ minimizes the Lagrangian

$$L(x, \mu^*(t), \lambda^*(t)) = f(x) + \sum_{j=1}^{m} \mu_j^*(t) g_j(x) + \lambda^*(t)^T (Ax - b)$$

  where $\mu_j^*(t) = -\frac{1}{tg_j(x^*(t))}$ and $\lambda^*(t) = \frac{\hat{\lambda}(t)}{t}$
- Note that $L(x^*(t), \mu^*(t), \lambda^*(t)) = f(x^*(t)) - \frac{m}{t}$
- This confirms the intuitive idea that $f(x^*(t)) \to f^*$ as $t \to \infty$:

$$f(x^*(t)) - \frac{m}{t} = L(x^*(t), \mu^*(t), \lambda^*(t)) = q(\mu^*(t), \lambda^*(t)) \le f^*$$

- Provides a lower estimate for f^* (can serve as stopping criterion)

Interpretation via KKT Conditions

The vectors x, μ and λ are primal-dual optimal if and only if they satisfy

- Primal feasibility: $g_j(x) \le 0$, $j = 1, \ldots, m$, $Ax = b$
- Dual feasibility: $\mu \succeq 0$
- Lagrangian optimality in x:

$$\nabla f(x) + \sum_{j=1}^{m} \mu_j \nabla g_j(x) + A^T \lambda = 0$$

- Complementary slackness: $\mu^T g(x) = -\frac{m}{t}$

The difference with the KKT conditions for the original inequality constrained problem is that the last condition replaces $\mu^T g(x) = 0$

The last condition can be viewed as **"approximate" complementary slackness for the original problem**, which converges to the "original" complementary slackness as $t \to \infty$

KKT Conditions

Original problem:

minimize $f(x)$

subject to $g(x) \preceq 0,\ Ax = b$

x^* is optimal iff there is (μ^*, λ^*)

- $Ax^* = b, \quad g(x^*) \preceq 0, \quad \mu^* \succeq 0$
- $\nabla f(x^*) + \sum_{j=1}^m \mu_j^* \nabla g_j(x^*) + A^T \lambda^* = 0$
- $(\mu^*)^T g(x^*) = 0$

Penalized problem:

minimize $f(x) - \frac{1}{t} \sum_{j=1}^m \ln(-g_j(x))$

subject to $Ax = b$

x_t^* is optimal iff there is $\widehat{\lambda}$:

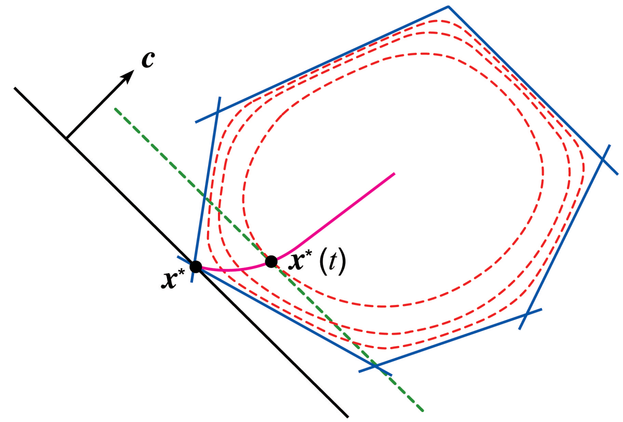
- $Ax_t^* = b, \quad g(x_t^*) \prec 0$
- $\nabla f(x_t^*) + \sum_{j=1}^m \frac{1}{-t g_j(x_t^*)} \nabla g_j(x_t^*) + A^T \widehat{\lambda}_t = 0$
- Letting $\widehat{\mu}_{jt} = \frac{1}{-t g_j(x_t^*)}$ for all j, we see that x_t^* and $(\widehat{\mu}_t, \widehat{\lambda}_t)$ satisfy KKTs for the original problem with approximate CS condition $\widehat{\mu}_t^T g(x_t^*) = -\frac{m}{t}$

Furthermore $\quad f^* \le f(x_t^*) \le f^* + \frac{m}{t}$

Central Path for LP

minimize $\quad c^T x$

subject to $\; a_j^T x \le b_j, \;\; j = 1, \ldots, m$



- The logarithmic barrier is given by
  $$\phi(x) = -\sum_{j=1}^m \ln\left(b_j - a_j^T x\right)$$
- The gradient and Hessian of the barrier function are
  $$\nabla\phi(x) = \sum_{j=1}^m \frac{1}{b_j - a_j^T x}\, a_j = A^T d \qquad \text{for } d \in \mathbb{R}^m \text{ with } d_j = \frac{1}{b_j - a_j^T x}$$
  $$\nabla^2\phi(x) = \sum_{j=1}^m \frac{1}{\left(b_j - a_j^T x\right)^2}\, a_j a_j^T = A^T \mathrm{diag}[d]^2 A$$
- Since $d \succ 0$ for $x \in \mathrm{dom}\,\phi$, the Hessian is nonsingular iff $\mathrm{rank}(A) = n$
- The centrality condition reduces to: $\quad tc + \sum_{j=1}^m \frac{1}{b_j - a_j^T x}\, a_j = 0$
- Interpretation: at $x^*(t)$ on the central path
  the gradient $\nabla\phi(x^*(t))$ is parallel to $-c$, i.e., the hyperplane $c^T x = c^T x^*(t)$ is tangent to the level set of ϕ through $x^*(t)$

Barrier Method

Given a strictly feasible x, $t := t_0 > 0$, $\beta > 1$, and a tolerance $\epsilon > 0$.

Repeat

1. *Centering step.* Compute $x^*(t)$ solving $\min_{Az=b}\{tf(z) + \phi(z)\}$

2. *Update.* Set $x := x^*(t)$.

3. *Stopping criterion.* **Quit** if $m/t < \epsilon$.

4. *Increase penalty.* Set $t := \beta t$.

- It terminates with $f(x) - f^* \leq \epsilon$ [follows from $f(x^*(t)) - f^* \leq m/t$]
- Centering steps are viewed as **outer iterations**
- Inside centering step, computing $x^*(t)$ is usually done using Newton's method starting at current x; these are **inner iterations**
- The methods of this form are also referred as *sequential minimization* techniques
- Choice of β involves a trade-off: large β means fewer outer iterations, more inner (Newton) iterations
- Choosing t large so as to have $m/t < \epsilon$ in one iteration causes problem with points close to the boundary of the feasible set: Hessian is unstable

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

Convergence Analysis: Outer Iterations

Number of outer (centering) iterations: to determine $x(t)$ such that $f(x(t)) - f^* \le \epsilon$, we need K such that

$$\frac{m}{\beta^K t_0} \le \epsilon$$

Thus, the number K of outer iterations for accuracy ϵ is given exactly by

$$\left\lceil \frac{\ln\left(\frac{m}{t_0 \epsilon}\right)}{\ln \beta} \right\rceil$$

Convergence Analysis: Initialization and Inner Iterations

To complete the analysis of the method, we need to address

- The **initial step of the method**:
  - Determining a **strictly feasible starting point** x

$$x \in \mathrm{dom} f, \qquad Ax = b, \qquad g_j(x) < 0,\ j = 1, \ldots, m$$

- The **centering step**: solving subproblems of the form

$$\begin{array}{ll} \text{minimize} & tf(z) + \phi(z) \\ \text{subject to} & Az = b \end{array}$$

  - The functions $tf + \phi$ must have closed level sets for all $t \geq t_0$
  - Analysis via self-concordance requires self-concordance of $tf + \phi$ [thus, three-times differentiable $tf + \phi$]

Feasibility and Phase I Methods

- The barrier method requires a strictly feasible starting point x
- When such a point is not available, the method is preceded by a preliminary stage (part of the initialization)
  - Phase I: computing a strictly feasible point or finding that the constraints are infeasible
  - The point found in Phase I serves as a starting iteration in the method
  - The later stage is referred as Phase II of the method
- Basic approach for determining a feasible x:

  Minimizing the maximum infeasibility
- Many variations of this basic approach exists, among them:

  Minimizing the sum of infeasibilities

Feasibility: Minimizing the Maximum Infeasibility

Basic phase I approach

$$\begin{aligned} \text{minimize} \quad & s \\ \text{subject to} \quad & g_j(x) \le s, \quad j = 1, \dots, m \\ & Ax = b \end{aligned} \tag{2}$$

NOTE:

- The minimization is with respect to (x, s) with $s \in \mathbb{R}$
- The domain of f has to be taken into account [added in constraints]

Let $\mathrm{dom} f = \mathbb{R}^n$ and $\widehat{f}^*$ be the optimal value of the feasibility problem

- The feasibility problem is strictly feasible: apply barrier method
- When (x, s) with $s < 0$ is feasible for problem (1), x is strictly feasible in original problem
- When $\widehat{f}^* > 0$, the original problem is infeasible
- When $\widehat{f}^* = 0$ and attained, the original problem is feasible [not strictly]
- When $\widehat{f}^* = 0$ and not attained, the original problem is infeasible

Note: In practice, we cannot exactly determine that $\widehat{f}^* = 0$;
typically, the feasibility algorithm terminates with $|\widehat{f}^*| \le \eta$

Alternative Approach: Minimize the Sum of Infeasibilities

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T s \\ \text{subject to} & g_j(x) \le s_j, \quad j = 1, \ldots, m \\ & Ax = b, \quad s \succeq 0 \end{array}$$

- For a fixed x, the optimal value of s_j is $\max\{g_j(x), 0\}$ - indeed maximizing the sum of infeasibilities
- The optimal value is $\widehat{f}^* = 0$ and attained *if and only if the original problem is feasible*
- Interesting property when the original problem is infeasible:
  - The approach produces a solution that satisfies many more constraints than the basic phase I approach [minimizing the max infeasibility]
  - Thus, it identifies a larger subset of constraints that are feasible

NOTE:

- However, this model produces a feasible point but NOT a strictly feasible

Analysis of the Inner Iterations

Analysis Assumption:

- The feasible level sets of the original problem are bounded

- The function $tf + \phi$ is closed and self-concordant for all $t \geq t_0$

When f and g_j are linear or quadratic, the function $tf - \sum_{j=1}^{m} \ln(-g_j)$ is self-concordant for all $t \geq 0$ (LPs, QPs and QCQPs)

NOTE:
The barrier method works well whether or not self-concordance is present

Inner Iterations

The iterations within centering step: accuracy fixed to ϵ_c

- Start with $x(t)$, a solution to $\min_{Az=b}\{tf(z) + \phi(z)\}$
- Use Newton's method and backtracking line search with parameters $\alpha = 1$, $\beta_c \in (0, 1)$ and $\sigma \in (0, 1/2)$
- Obtain $x(\beta t)$, an ϵ_c-solution to the problem $\min_{Az=b}\{\beta tf(z) + \phi(z)\}$

Recall: to minimize a closed (self-concordant function f, the number of Newton's iterations, with backtracking line search and starting iterate x_0, is bounded by

$$\frac{f(x_0) - f^*}{\Gamma} + \kappa$$

with $\Gamma = \frac{\sigma\beta_c(1-2\sigma)^2}{20-8\sigma}$ and $\kappa = \log_2 \log_2 \left(\frac{1}{\epsilon_c}\right)$

We apply this to the function $\beta tf(z) + \phi(z)$

Bound on the Number of Inner Iterations

Under the level set boundedness and self-concordance assumption on functions $tf + \phi$, $t \geq t_0$, the number of Newton's iterations within a centering step is estimated as follows:

- Under unrealistic assumptions of exact solutions: $x = x^*(t)$ and $x_\beta = x^*(\beta t)$ optimal for the penalized problems with penalties t and βt resp.
- We have: $\#$ Newton's steps $\leq \frac{\beta t f(x) + \phi(x) - \beta t f(x_\beta) - \phi(x_\beta)}{\Gamma} + \kappa$
  with $\Gamma = \frac{\sigma \beta_c (1-2\sigma)^2}{20 - 8\sigma}$ and $\kappa = \log_2 \log_2 \left(\frac{1}{\epsilon_c}\right)$

Therefore:

$$
\begin{aligned}
\text{numerator} &= \beta t f(x) - \beta t f(x_\beta) + \sum_{j=1}^m \ln[-\beta t \widehat{\mu}_j g_j(x_\beta)] - m \ln \beta \\
&\leq \beta t f(x) - \beta t f(x_\beta) - \beta t \sum_{j=1}^m \widehat{\mu}_j g_j(x_\beta) - m - m \ln \beta \\
&\leq \beta t f(x) - \beta t q(\widehat{\mu}, \widehat{\lambda}) - m - m \ln \beta \\
&= \beta t f(x) - \beta t f(x) + m\beta - m - m \ln \beta \\
&= m(\beta - 1 - \ln \beta)
\end{aligned}
$$

A Total Bound on Newton's Iterations

Combining the bound on total number of Newton's iterations:

$$N \le \left\lceil \frac{\ln\left(\frac{m}{t_0\epsilon}\right)}{\ln\beta} \right\rceil \left\lceil \frac{m(\beta - 1 - \ln\beta)}{\Gamma} + \kappa \right\rceil$$

- The bound depends strongly on the parameter β
- The function $\beta - 1 - \ln\beta$ is almost quadratic for small β, and grows linearly for large β
- Hence, for a small number of inner iterations, it is recommendable to use a smaller β [the number of inner iterations can be large for large β]
- The bound *does not depend on n and p* (size of x and # of rows in A)
- When $\beta = 1 + \frac{1}{\sqrt{m}}$ the bound on the total number of Newton's steps
$$N \le \left(\frac{1}{2\Gamma} + \kappa\right) \left\lceil \sqrt{m}\log_2\left(\frac{m}{t_0\epsilon}\right) \right\rceil$$

Pros and Cons

Advantages

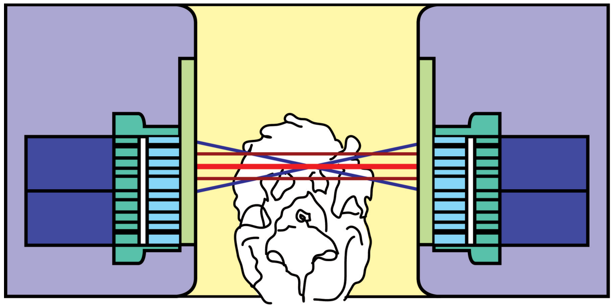
- Interior Point Method is successfully applied in practice
- Works very well for moderately large problems
- Even for nonconvex (issues with "trapped" at a local minimum)

Limitations

- High requirement on differentiability of f and g_js
- The method "is inefficient" when the size of the problem is very large (n in millions) and the Hessians are not sparse
- Not suitable for "distributed" computations in general

Example

Image Reconstruction in PET-scan [Ben-Tal, 2005]



- Maximum Likelihood Model results in convex optimization

$$\min_{x \geq 0,\ e'x \leq 1} \left\{ -\sum_{i=1}^{m} y_i \ln\left(\sum_{j=1}^{n} p_{ij} x_j \right) \right\}$$

  - x is a decision vector - size n
  - y models measured data (by PET detectors) - size m
  - p_{ij} probabilities modeling detections of emitted positrons
- The number n of decision variables ranges from $1/2 - 3$ millions
- The number m of data variables ranges from $3 - 25$ millions
- The Hessians are not sparse

Interior Point Method can be Inefficient

- For the PET Imaging problem [Ben-Tal, 2005]
  - Image resolution $64 \times 64 \times 64$ and $n = 262,144$, the CPU time per Newton's iteration is about 2.5 hours
  - Image resolution $128 \times 128 \times 128$ and $n = 2,097,152$ the CPU time per Newton's iteration is more than 13 days
- This motivates a re-newed interest in gradient-type methods
  - The complexity of a gradient-type step is linear in n
  - In practice, their accuracy is not high, BUT in large scale problems "high accuracy" is not required
- **Gradient-type methods are well suited for medium-accuracy solutions in extremely large convex problems**