Nonlinear Optimization: Algorithms 3: Interior-point methods

INSEAD, Spring 2006

Jean-Philippe Vert

Ecole des Mines de Paris

`Jean-Philippe.Vert@mines.org`

Outline

- Inequality constrained minimization
- Logarithmic barrier function and central path
- Barrier method
- Feasibility and phase I methods

Inequality constrained minimization

Setting

We consider the problem:

$$\begin{aligned} \textit{minimize} \quad & f(x) \\ \textit{subject to} \quad & g_i(x) \leq 0 \,, \quad i = 1, \ldots, m \,, \\ & Ax = b \,, \end{aligned}$$

- f and g are supposed to be *convex* and *twice continuously differentiable*.
- A is a $p \times n$ matrix of rank $p < n$ (i.e., fewer equality constraints than variables, and independent equality constraints).
- We assume f^* is finite and attained at x^*

Strong duality hypothesis

- We finally assume the problem is *strictly feasible*, i.e., there exists x with $g_i(x) < 0, i = 1, \ldots, m$, and $Ax = 0$. This means that Slater's constraint qualification holds $\implies$ strong duality holds and dual optimum is attained, i.e., there exists $\lambda^* \in \mathbb{R}^p$ and $\mu \in \mathbb{R}^m$ which together with x^* satisfy the KKT conditions:

$$
\begin{aligned}
Ax^* &= b \\
g_i(x^*) &\leq 0 , \quad i = 1, \ldots, m \\
\mu^* &\geq 0 \\
\nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla g_i(x^*) + A^\top \lambda^* &= 0 \\
\mu_i^* g_i(x^*) &= 0 , \quad i = 1, \ldots, m .
\end{aligned}
$$

Examples

Many problems satisfy these conditions, e.g.:

- LP, QP, QCQP
- Entropy maximization with linear inequality constraints

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} \quad & Fx \leq g \\
& Ax = b \,.
\end{aligned}
$$

Examples (cont.)

- To obtain differentiability of the objective and constraints we might reformulate the problem, e.g:

$$\text{minimize} \quad \max_{i=1,\dots,n} \left(a_i^\top x \right) + b_i$$

with nondifferentiable objective is equivalent to the LP:

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & a_i \top x + b \leq t \,, \quad i = 1, \dots, m \,. \\ & Ax = b \,. \end{aligned}$$

Overview

Interior-point methods solve the problem (or the KKT conditions) by applying Newton's method to a sequence of equality-constrained problems. They form another level in the *hierarchy of convex optimization algorithms*:

- Linear equality constrained quadratic problems (*LCQP*) are the simplest (set of linear equations that can be solved analytically)
- Newton's method: reduces linear equality constrained convex optimization problems (*LCCP*) with twice differentiable objective to a *sequence of LCQP*.
- Interior-point methods reduce a problem with linear equality and inequality constraints to a *sequence of LCCP*.

Logarithmic barrier function and central path

Problem reformulation

Our goal is to approximately formulate the *inequality constrained problem* as an *equality constrained problem* to which Newton's method can be applied. To this end we first hide the inequality constraint implicit in the objective:

$$\begin{aligned} \text{minimize} \quad & f(x) + \sum_{i=1}^{m} I_{-}\left(g_i(x)\right) \\ \text{subject to} \quad & Ax = b \,, \end{aligned}$$

where $I_{-} : \mathbb{R} \rightarrow \mathbb{R}$ is the *indicator function for nonpositive reals*:

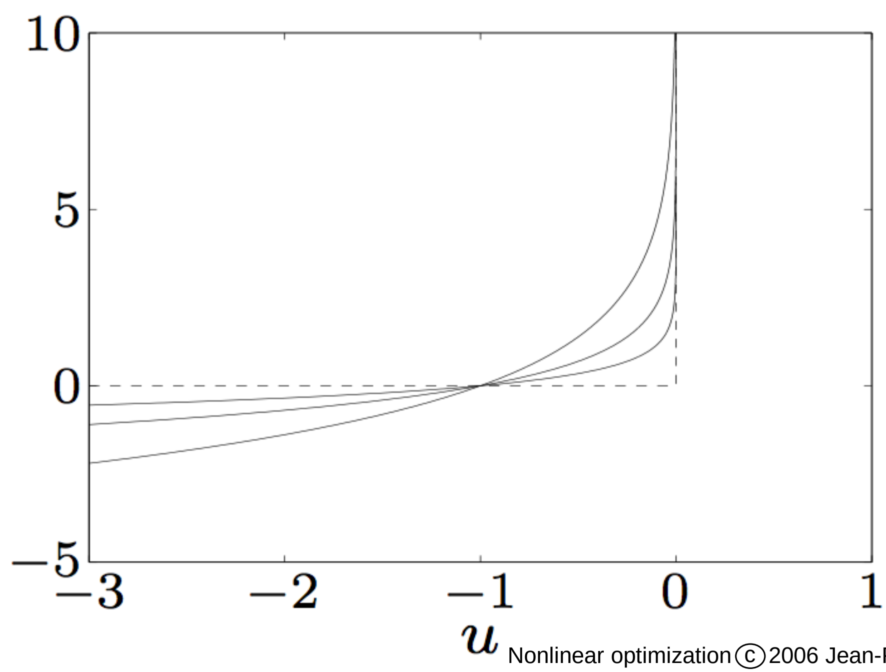
$$I_{-}(u) = \begin{cases} 0 & \text{if } u \leq 0 \,, \\ +\infty & \text{if } u > 0 \,. \end{cases}$$

Logarithmic barrier

The basic idea of the barrier method is to *approximate the indicator function* I_- by the *convex and differentiable* function

$$\hat{I}_-(u) = -\frac{1}{t}\log(-u)\,, \quad u < 0\,,$$

where $t > 0$ is a parameter that sets the accuracy of the prediction.



Problem reformulation

Subsituting $\hat{I}_-$ for I_- in the optimization problem gives the approximation:

$$\begin{array}{ll} \textit{minimize} & f(x) + \sum_{i=1}^{m} -\frac{1}{t} \log\left(-g_i(x)\right) \\ \textit{subject to} & Ax = b\,, \end{array}$$

The objective function of this problem is convex and twice differentiable, so *Newton's method* can be used to solve it. Of course this problem is just an *approximation* to the original problem. We will see that the quality of the approximation of the solution increases when t *increases*.

Logarithmic barrier function

The function

$$\phi(x) = -\sum_{i=1}^{m} \log\left(-g_i(x)\right)$$

is called the *logarithmic barrier* or *log barrier* for the original optimization problem. Its domain is the set of points that satisfy *all inequality constraints strictly*, and it grows without bound if $g_i(x) \to 0$ for any i. Its gradient and Hessian are given by:

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

$$\nabla^2 \phi(x) = \sum_{i=1}^{m} \frac{1}{g_i\left(x\right)^2} \nabla g_i(x) \nabla g_i(x)^\top + \sum_{i=1}^{m} \frac{1}{-g_i(x)} \nabla^2 g_i(x) \,.$$

Central path

Our approximate problem is therefore (equivalent to) the following problem:

$$\begin{aligned} \textit{minimize} \quad & t f(x) + \phi(x) \\ \textit{subject to} \quad & Ax = b\,. \end{aligned}$$

We assume for now that this problem can be solved via Newton's method, in particular that it has *a unique solution* $x^*(t)$ for each $t > 0$.

The *central path* is the set of solutions, i.e.:

$$\{x^*(t) \mid t > 0\} \ .$$

Characterization of the central path

A point $x^*(t)$ is on the central path *if and only if*:

- it is *strictly feasible*, i.e., satisfies:

$$Ax^*(t) = b\,, \quad g_i\left(x^*(t)\right) < 0\,, \quad i = 1, \ldots, m\,.$$

- there exists a $\hat{\lambda} \in \mathbb{R}^p$ such that:

$$\begin{aligned} 0 &= t\nabla f\left(x^*(t)\right) + \nabla\phi\left(x^*(t)\right) + A^\top\hat{\lambda} \\ &= t\nabla f\left(x^*(t)\right) + \sum_{i=1}^{m} \frac{1}{-g_i\left(x^*(t)\right)}\nabla g_i\left(x^*(t)\right) + A^\top\hat{\lambda}\,. \end{aligned}$$

Example: LP central path

The log barrier for a LP:

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax \leq b \,, \end{aligned}$$

is given by

$$\phi(x) = -\sum_{i=1}^{m} \log\left(b_i - a_i^\top x\right) \,,$$

where a_i is the ith row of A. Its derivatives are:

$$\nabla\phi(x) = \sum_{i=1}^{m} \frac{1}{b_i - a_i^\top x} a_i \,, \quad \nabla^2\phi(x) = \sum_{i=1}^{m} \frac{1}{\left(b_i - a_i^\top x\right)^2} a_i a_i^\top \,.$$

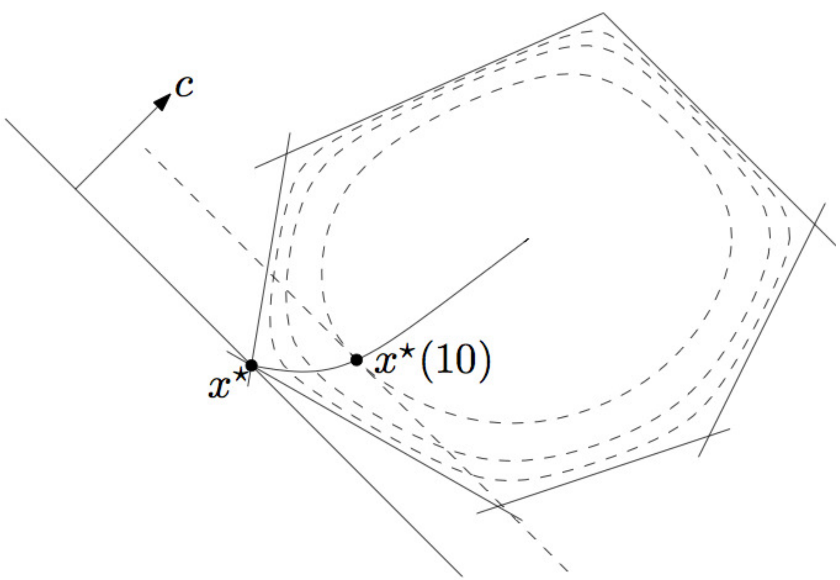
Example (cont.)

The derivatives can be rewritten more compactly:

$$\nabla \phi(x) = A^\top d \, , \quad \nabla^2 \phi(x) = A^\top diag(d)^2 A \, ,$$

where $d \in \mathbb{R}^m$ is defined by $d_i = 1/\left(b_i - a_i^\top x\right)$. The centrality condition for $x^*(t)$ is:

$$tc + A^\top d = 0$$



$\Longrightarrow$ at each point on the central path, $\nabla \phi(x)$ *is parallel to* $-c$.

Dual points on central path

Remember that $x = x^*(t)$ if there exists a w such that

$$t\nabla f\left(x^*(t)\right) + \sum_{i=1}^{m} \frac{1}{-g_i\left(x^*(t)\right)} \nabla g_i\left(x^*(t)\right) + A^\top \hat{\lambda} = 0\,, \quad Ax = b\,.$$

Let us now define:

$$\mu_i^*(t) = -\frac{1}{t g_i\left(x^*(t)\right)}\,, \quad i = 1, \ldots, m, \quad \lambda^*(t) = \frac{\hat{\lambda}}{t}\,.$$

We claim that *the pair $\lambda^*(t), \mu^*(t)$ is dual feasible.*

Dual points on central path (cont.)

Indeed:

- $\mu^*(t) > 0$ because $g_i\left(x^*(t)\right) < 0$
- $x^*(t)$ minimizes the Lagrangian

$$L\left(x, \lambda^*(t), \mu^*(t)\right) = f(x) + \sum_{i=1}^{m} \mu_i^*(t) g_i(x) + \lambda^*(t)^\top \left(Ax - b\right) \ .$$

Therefore the dual function $q\left(\mu^*(t), \lambda^*(t)\right)$ is finite and:

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

Convergence of the central path

From the equation:

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

we deduce that the *duality gap* associated with $x^*(t)$ and the dual feasible pair $\lambda^*(t), \mu^*(t)$ is simply *m/t*. As an important consequence we have:

$$f\left(x^*(t)\right) - f^* \leq \frac{m}{t}$$

This confirms the intuition that $f\left(x^*(t)\right) \rightarrow f^*$ *if* $t \rightarrow \infty$.

Interpretation via KKT conditions

We can rewrite the conditions for x to be on the central path by the existence of λ, μ such that:

1. Primal constraints: $g_i(x) \leq 0$, $Ax = b$
2. Dual constraints : $\mu \geq 0$
3. *approximate* complementary slackness: $-\mu_i g_i(x) = 1/t$
4. gradient of Lagrangian w.r.t. x vanishes:

$$\nabla f(x) + \sum_{i=1}^{m} \mu_i \nabla g_i(x) + A^\top \lambda = 0$$

The only difference with KKT is that 0 is replaced by $1/t$ in 3. For "large" t, the point on the central path "*almost*" satisfies the KKT conditions.

The barrier method

Motivations

We have seen that the point $x^*(t)$ is *m/t-suboptimal*. In order to solve the optimization problem with a guaranteed specific accuracy $\epsilon > 0$, it suffices to take $t = m/\epsilon$ and solve the equality constrained problem:

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

by Newton's method. However this only works for *small problems, good starting points and moderate accuracy*. It is rarely, if ever, used.

Barrier method

- *given* strictly feasible x, $t = t^{(0)} > 0, \mu > 1$, tolerance $\epsilon > 0$.
- *repeat*
  1. Centering step: compute $x^*(t)$ by minimizing $tf + \phi$, subject to $Ax = b$
  2. Update: $x := x^*(t)$.
  3. Stopping criterion: *quit* if $m/t < \epsilon$.
  4. Increase t: $t := \mu t$.

Barrier method: Centering

- Centering is usually done with Newton's method, *starting at current* x
- Inexact centering is possible, since the goal is only to obtain a sequence of points $x^{(k)}$ that converges to an optimal point. In practice, however, the cost of computing an extremely accurate minimizer of $tf_0 + \phi$ as compared to the cost of computing a good minimizer is only marginal.

Barrier method: choice of μ

The choice of μ involves a trade-off

- For small μ, the initial point of each Newton process is good and few Newton iterations are required; however, many outer loops (update of t) are required.
- For large μ, many Newton steps are required after each update of t, since the initial point is probably not very good. However few outer loops are required.
- In practice $\mu = 10 - 20$ works well.

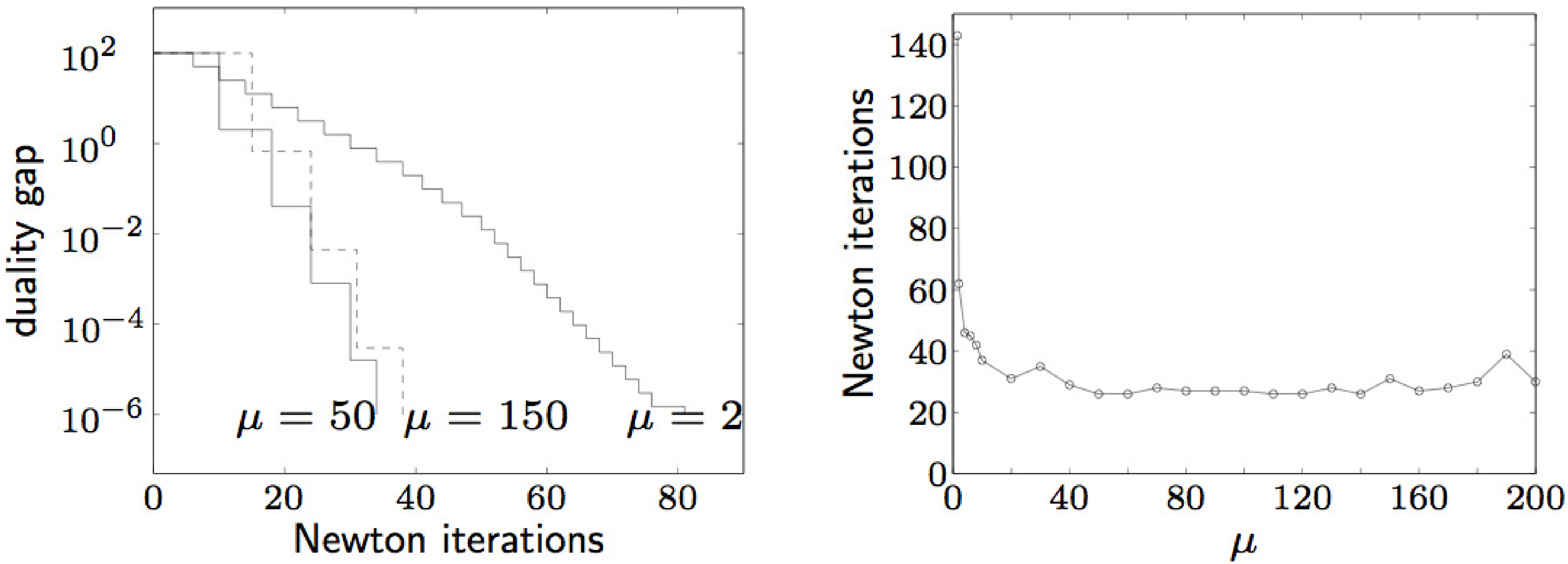
Barrier method: choice of $t^{(0)}$

The choice of $t^{(0)}$ involves a simple trade-off

- if $t^{(0)}$ is chosen too large, the first outer iteration will require too many Newton iterations
- if $t^{(0)}$ is chosen too small, the algorithm will require extra outer iterations

Several heuristics exist for this choice.

Example: LP in inequality form

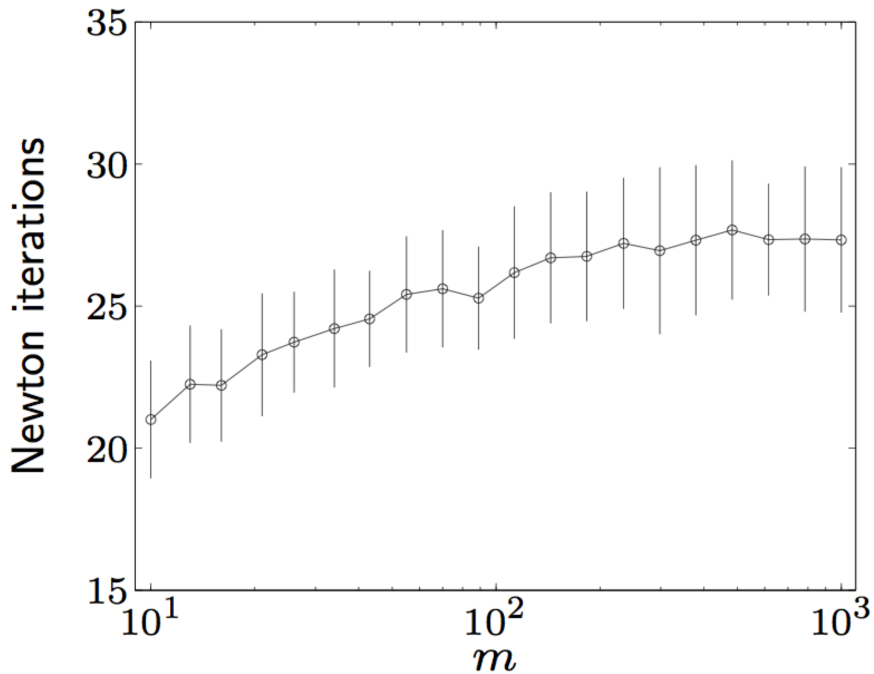


- $m = 100$ inequalities, $n = 50$ variables.
- start with x on central paht ($t^{(0)} = 1$, duality gap 100), terminates when $t = 10^8$ (gap 10^{-6})
- centering uses Newton's method with backtracking
- total number of Newton iterations not very sensitive for $\mu > 10$

Example: A family of standard LP

$$\text{minimize } c^\top x \quad \text{subject to } Ax = b, \; x \geq 0$$

for $A \in \mathbb{R}^{m \times 2m}$. Test for $m = 10, \ldots, 1000$:



The number of iterations grows *very slowly* as m ranges over a $100 : 1$ ratio.

Feasibility and phase I methods

The barrier method requires a *strictly feasible* starting point $x^{(0)}$:

$$g_i\left(x^{(0)}\right) < 0\ , i = 1, \ldots, m \quad Ax^{(0)} = 0\ .$$

When such a point is not known, the barrier method is preceded by a preliminary stage, called *phase I*, in which a strictly feasible point is computed.

Basic phase I method

$$
\begin{aligned}
\text{minimize} \quad & s \\
\text{subject to} \quad & g_i(x) \leq s \;, \quad i = 1, \ldots, m \;, \\
& Ax = b \;,
\end{aligned}
$$

- this problem is always strictly feasible (choose any x, and s large enough).
- apply the barrier method to this problem = *phase I optimization problem*.
- If x, s feasible with $s < 0$ then x is strictly feasible for the initial problem
- If $f^* > 0$ then the original problem is infeasible.

Primal-dual interior-point methods

A variant of the barrier method, more efficient when high accurary is needed

- update primal and dual variables at each iteration: no distinction between inner and outer iterations
- often exhibit superlinear asymptotic convergence
- search directions can be interpreted as Newton directions for modified KKT conditions
- can start at infeasible points
- cost per iteration same as barrier method