Interior-point methods

Lecture 13, Nonlinear Programming

National Taiwan University

January 3, 2017

Table of contents

Inequality constrained minimization problems (1/4)

- In this lecture, we discuss **interior-point methods** for solving convex optimization problems that include inequality constraints,

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

  where $f_0, \ldots, f_m : \mathbf{R}^n \to \mathbf{R}$ are convex and twice continuously differentiable, and $A \in \mathbf{R}^{p \times n}$ with rank $A = p < n$.
- We assume that an optimal x^* exists and denote the optimal value $f_0(x^*)$ as p^*.

Inequality constrained minimization problems (2/4)

- We also assume that the problem is strictly feasible, i.e., there exists $x \in \textbf{relint } \mathcal{D}$ that satisfies $Ax = b$ and $f_i(x) < 0$ for $i = 1, \ldots, m$. (i.e., Slater's constraint qualification holds), so there exist dual optimal $\lambda^* \in \mathbf{R}^m$, $\nu^* \in \mathbf{R}^p$, which together with x^* satisfy the KKT conditions

$$\begin{aligned} Ax^* = b, \quad f_i(x^*) &\leq 0, \quad i = 1, \ldots, m \\ \lambda^* &\succeq 0 \\ \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + A^T \nu^* &= 0 \\ \lambda_i^* f_i(x^*) &= 0, \quad i = 1, \ldots, m. \end{aligned}$$

Inequality constrained minimization problems (3/4)

- Interior-point methods solve the optimization problem (or the corresponding KKT conditions) by applying Newton's method to a sequence of equality constrained problems, or to a sequence of modified versions of the KKT conditions.
- We will concentrate on a particular interior-point algorithm, the **barrier method**.
- We can view interior-point methods as another level in the hierarchy of convex optimization algorithms.
- Linear equality constrained quadratic problems are the simplest. For these problems the KKT conditions are a set of linear equations, which can be solved analytically.

Inequality constrained minimization problems (4/4)

- Newton's method is the next level in the hierarchy. We can think of Newton's method as a technique for solving a linear equality constrained optimization problem, with twice differentiable objective, by reducing it to a sequence of linear equality constrained quadratic problems.
- Interior-point methods form the next level in the hierarchy: They solve an optimization problem with linear equality and inequality constraints by reducing it to a sequence of linear equality constrained problems.

Examples (1/2)

- Many problems are already in the form of a convex optimization problem, and satisfy the assumption that the objective and constraint functions are twice differentiable.
- Obvious examples are LPs, QPs, QCQPs, and GPs in convex form; another example is linear inequality constrained entropy maximization,

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

with domain $\mathcal{D} = \mathbf{R}^n_{++}$.

Examples (2/2)

- Many other problems do not have the required form with twice differentiable objective and constraint functions, but can be reformulated in the required form.
- We have already seen many examples of this, such as the transformation of an unconstrained convex piecewise-linear minimization problem

$$\text{minimize} \quad \max_{i=1,\ldots,m} (a_i^T x + b_i)$$

(with nondifferentiable objective), to the LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, i = 1, \ldots, m \end{array}$$

(having twice differentiable objective and constraint functions).

Eliminating Inequality Constraints (1/2)

- Goal: approximately formulate the inequality constrained problem as an equality constrained problem to which Newton's method can be applied.
- Our first step is to rewrite the problem

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

making the inequality constraints implicit in the objective.

Eliminating Inequality Constraints (2/2)

- As a result, we obtain the following equivalent problem:

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

  where $I_- : \mathbf{R} \to \mathbf{R}$ is the indicator function for the nonpositive reals,

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

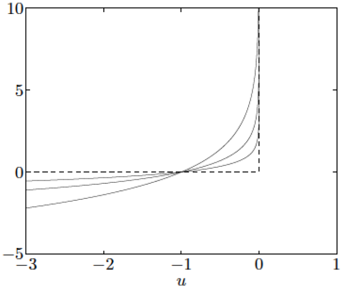
- The reformulated problem has no inequality constraints, but its objective function is not (in general) differentiable, so Newton's method cannot be applied.

Logarithmic barrier (1/4)

- We approximate the indicator function I_- by the function

$$\hat{I}_-(u) = -(1/t)\log(-u), \quad \mathbf{dom}\ \hat{I}_- = -\mathbf{R}_{++},$$

where $t > 0$ is a parameter for the approximation accuracy.



- Like I_-, the function $\hat{I}_-$ is convex and nondecreasing, and takes on the value ∞ for $u > 0$. However, unlike I_-, $\hat{I}_-$ is differentiable and closed: it increases to ∞ as u increases to 0.

Logarithmic barrier (2/4)

- As t increases, the approximation becomes more accurate.
- Substituting $\hat{I}_-$ for I_- gives the approximation

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

- The objective here is convex, since $-(1/t)\log(-u)$ is convex and increasing in u, and differentiable. So, Newton's method can be used to solve it, assuming an appropriate closedness condition holds.

Logarithmic barrier (3/4)

- The function $\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x))$, with **dom** $\phi = \{x \in \mathbf{R}^n \mid f_i(x) < 0, i = 1, ..., m\}$, is called the **logarithmic barrier** (or log barrier) for the original problem.
- Its domain is the set of points that strictly satisfy the inequality constraints of the original problem. The logarithmic barrier, $\phi(x)$, grows without bound if $f_i(x) \to 0$, for any i.
- Note that the problem

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

is just an approximation of the original problem. But it can be shown that the quality of the approximation improves as the parameter t grows.

Logarithmic barrier (4/4)

- On the other hand, when the parameter t is large, the function $f_0 + (1/t)\phi$ is difficult to minimize by Newton's method, since its Hessian varies rapidly near the boundary of the feasible set.
- We will see that this problem can be circumvented by solving a sequence of approximated problems, increasing the parameter t at each step, and starting each Newton minimization at the solution of the problem for the previous value of t.

Gradient and Hessian of the logarithmic barrier function

- For future reference, we note that the gradient and Hessian of the logarithmic barrier function ϕ are given by

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

and

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

Central path (1/3)

- We now consider in more detail the minimization problem

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

- To simplify notation, we multiply the objective by t, and consider the equivalent problem

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x) \\ \text{subject to} & Ax = b, \end{array}$$

which has the same minimizers.

Central path (2/3)

- We assume that the problem

$$\begin{aligned} \text{minimize} \quad & tf_0(x) + \phi(x) \\ \text{subject to} \quad & Ax = b, \end{aligned}$$

  can be solved via Newton's method, and that it has a unique solution for each $t > 0$.
- For $t > 0$ we define $x^*(t)$ as the solution of the above problem.
- The **central path** associated with the original problem is defined as the set of points $x^*(t)$, $t > 0$, which we call the **central points**.

Central path (3/3)

- Points on the central path are characterized by the following necessary and sufficient conditions: $x^*(t)$ is strictly feasible, i.e., satisfies

$$Ax^*(t) = b, \quad f_i(x^*(t)) < 0, \quad i = 1, \ldots, m,$$

and there exists a $\hat{\nu} \in \mathbf{R}^p$ such that

$$\begin{aligned} 0 &= t\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{\nu} \\ &= t\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t))}\nabla f_i(x^*(t)) + A^T\hat{\nu} \end{aligned}$$

holds.

Example – Inequality form linear programming (1/4)

- The logarithmic barrier function for an LP in inequality form,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b, \end{array}$$

is given by

$$\phi(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x), \textbf{dom } \phi = \{x \mid Ax \prec b\},$$

where $a_1^T, \ldots, a_m^T$ are the rows of A.

Example – Inequality form linear programming (2/4)

- The gradient and Hessian of the barrier function are

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

- By defining $d_i = 1/(b_i - a_i^T x)$, we reach a more compact form

$$\nabla \phi(x) = A^T d, \quad \nabla^2 \phi(x) = A^T \mathbf{diag}\,(d)^2 A,$$

where $d \in \mathbf{R}^m$.

- Since x is strictly feasible, we have $d \succ 0$, so the Hessian of ϕ is nonsingular if and only if A has rank n.

Example – Inequality form linear programming (3/4)

- The centrality condition

$$
\begin{aligned}
0 &= t\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{\nu} \\
&= t\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t))}\nabla f_i(x^*(t)) + A^T\hat{\nu}
\end{aligned}
$$

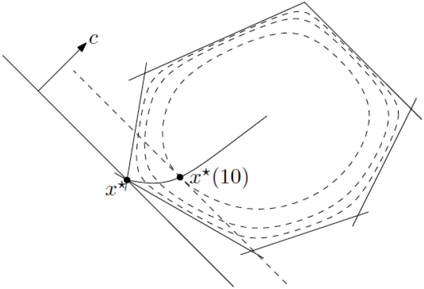
implies

$$
tc + \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i = tc + A^T d = 0.
$$

- We can give a simple geometric interpretation of the centrality condition as in the next slide.

Example – Inequality form linear programming (4/4)

- At a point $x^*(t)$ on the central path the gradient $\nabla\phi(x^*(t))$, which is normal to the level set of ϕ through $x^*(t)$, must be parallel to $-c$.
- In other words, the hyperplane $c^T x = c^T x^*(t)$ is tangent to the level set of ϕ through $x^*(t)$.



Dual points from central path (1/3)

- We can derive an important property of the central path: Every central point yields a dual feasible point, and hence a lower bound on the optimal value p^*.
- Specifically, define

$$\lambda_i^*(t) = -\frac{1}{tf_i(x^*(t))}, \quad i = 1, \ldots, m, \quad \nu^*(t) = \hat{\nu}/t.$$

- We claim that the pair $(\lambda^*(t), \nu^*(t))$ is dual feasible (i.e., $g(\lambda^*(t), \nu^*(t)) > -\infty$ and $\lambda^*(t) \succeq 0$).
- First, it is clear that $\lambda^*(t) \succ 0$ because $f_i(x^*(t)) < 0, \quad i = 1, \ldots, m.$

Dual points from central path (2/3)

- Next, note that the centrality condition can be expressed as

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

- We see that $x^*(t)$ minimizes the Lagrangian

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^T (Ax - b),$$

for $\lambda = \lambda^*(t)$ and $\nu = \nu^*(t)$, which means that $(\lambda^*(t), \nu^*(t))$ is a dual feasible pair.

Dual points from central path (3/3)

- Therefore, the dual function $g(\lambda^*(t), \nu^*(t))$ is finite, and

$$\begin{aligned} g(\lambda^*(t), \nu^*(t)) &= f_0(x^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t) f_i(x^*(t)) + \nu^*(t)^T (Ax^*(t) - b) \\ &= f_0(x^*(t)) - m/t. \end{aligned}$$

- In particular, the duality gap associated with $x^*(t)$ and the dual feasible pair $(\lambda^*(t), \nu^*(t))$ is simply m/t. As an important consequence, we have

$$f_0(x^*(t)) - p^* \leq m/t,$$

  i.e., $x^*(t)$ is no more than m/t-suboptimal.
- This confirms the intuitive idea that $x^*(t)$ converges to an optimal point as $t \to \infty$.

Example – Inequality form linear programming (1/2)

- Let us revisit the inequality form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b. \end{array}$$

- The dual problem the inequality form LP is

$$\begin{array}{ll} \text{maximize} & -b^T \lambda \\ \text{subject to} & A^T \lambda + c = 0 \\ & \lambda \succeq 0. \end{array}$$

Example – Inequality form linear programming (2/2)

- From the optimality conditions

$$tc + \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i = tc + A^T d = 0,$$

it is clear that

$$\lambda_i^*(t) = \frac{1}{t(b_i - a_i^T x^*(t))}, \quad i = 1, ..., m,$$

is dual feasible, with dual objective value

$$-b^T \lambda^*(t) = c^T x^*(t) + (Ax^*(t) - b)^T \lambda^*(t) = c^T x^*(t) - m/t.$$

Interpretation via KKT conditions (1/2)

- We can interpret the centrality conditions

$$\begin{aligned} 0 &= t\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{\nu} \\ &= t\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t))}\nabla f_i(x^*(t)) + A^T\hat{\nu} \end{aligned}$$

as a continuous deformation of the KKT optimality conditions.
- A point x equals $x^*(t)$ if and only if there exists λ, ν such that

$$\begin{aligned} Ax &= b, \\ f_i(x) &\leq 0, \quad i = 1, \dots, m \\ \lambda &\succeq 0 \\ \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T\nu &= 0 \\ -\lambda_i f_i(x) &= 1/t, \quad i = 1, \dots, m. \end{aligned}$$

Interpretation via KKT conditions (2/2)

- The only difference between the KKT conditions and the centrality conditions is that the complementarity condition $-\lambda_i f_i(x) = 0$ is replaced by the condition $-\lambda_i f_i(x) = 1/t$.
- In particular, for large t, $x^*(t)$ and the associated dual point $\lambda^*(t), \nu^*(t)$ 'almost' satisfy the KKT optimality conditions.

Force field interpretation (1/3)

- We can give a simple mechanics interpretation of the central path in terms of potential forces acting on a particle in the strictly feasible set C.
- For simplicity we assume that there are no equality constraints.
- We associate with each constraint the force

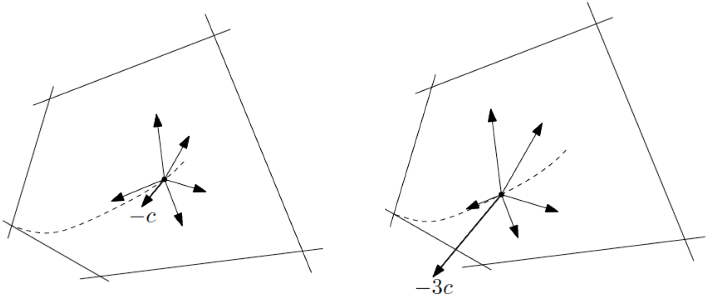
$$F_i(x) = -\nabla(-\log(-f_i(x))) = \frac{1}{f_i(x)} \nabla f_i(x)$$

  acting on the particle when it is at position x.
- Now we imagine another force acting on the particle, given by

$$F_0(x) = -t \nabla f_0(x),$$

  when the particle is at position x.

Force field interpretation (2/3)



- This objective force field acts to pull the particle in the negative gradient direction (i.e., toward smaller f_0).
- The central point $x^*(t)$ is the point where the constraint forces exactly balance the objective force felt by the particle.

Force field interpretation (3/3)

- As the parameter t increases, the particle is more strongly pulled toward the optimal point, but it is always trapped in C by the barrier potential, which becomes infinite as the particle approaches the boundary.

The barrier method (1/5)

- We have seen that the point $x^*(t)$ is m/t-suboptimal, and that a certificate of this accuracy is provided by the dual feasible pair $\lambda^*(t), \nu^*(t)$.
- This suggests a very straightforward method for solving the original problem with a guaranteed specified accuracy ϵ: We simply take $t = m/\epsilon$ and solve the equality constrained problem

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

using Newton's method.

The barrier method (2/5)

- This method could be called the unconstrained minimization method, since it allows us to solve the inequality constrained problem to a guaranteed accuracy by solving an unconstrained, or linearly constrained, problem.
- Although this method can work well for small problems, good starting points, and moderate accuracy (i.e., ϵ not too small), it does not work well in other cases. As a result, it is rarely used.

The barrier method (3/5)

- A simple extension of the unconstrained minimization method does work well.
- It is based on solving a sequence of unconstrained (or linearly constrained) minimization problems, using the last point found as the starting point for the next unconstrained minimization problem.
- In other words, we compute $x^*(t)$ for a sequence of increasing values of t, until $t \geq m/\epsilon$, which guarantees that we have an ϵ-suboptimal solution of the original problem.

The barrier method (4/5)

- A simple version of the method is as follows.

  Algorithm 11.1 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 $Ax = b$, starting at x.
  2. *Update.* $x := x^*(t)$.
  3. *Stopping criterion.* **quit** if $m/t < \epsilon$.
  4. *Increase t.* $t := \mu t$.

- At each iteration (except the first one) we compute the central point $x^*(t)$ starting from the previously computed central point, and then increase t by a factor $\mu > 1$.

The barrier method (5/5)

- The algorithm can also return $\lambda = \lambda^*(t)$, and $\nu = \nu^*(t)$, a dual ϵ-suboptimal point, or certificate for x.
- We refer to each execution of step 1 as a **centering step** (since a central point is being computed) or an outer iteration.
- Although any method for linearly constrained minimization can be used in step 1, we will assume that Newton's method is used.
- We refer to the Newton iterations or steps executed during the centering step as inner iterations.
- At each inner step, we have a primal feasible point; we have a dual feasible point, however, only at the end of each outer (centering) step.

Accuracy of centering (1/2)

- We should make some comments on the accuracy to which we solve the centering problems.
- Computing $x^*(t)$ exactly is not necessary since the central path has no significance beyond the fact that it leads to a solution of the original problem as $t \to \infty$; inexact centering will still yield a sequence of points $x^{(k)}$ that converges to an optimal point.
- Inexact centering, however, means that the points $\lambda^*(t), \nu^*(t)$, computed from

$$\lambda_i^*(t) = -\frac{1}{tf_i(x^*(t))}, \quad i = 1, \dots, m, \quad \nu^*(t) = \hat{\nu}/t,$$

are not exactly dual feasible.

Accuracy of centering (2/2)

- The cost of computing an extremely accurate minimizer of $tf_0 + \phi$, as compared to the cost of computing a good minimizer of $tf_0 + \phi$, is only marginally more, i.e., a few Newton steps at most.
- So it is usually reasonable to assume exact centering.

Choice of μ (1/3)

- The choice of the parameter μ involves a trade-off in the number of inner and outer iterations required.
- If μ is small (i.e., near 1) then at each outer iteration t increases by a small factor. As a result, the initial point for the Newton process, i.e., the previous iterate x, is a very good starting point, and the number of Newton steps needed to compute the next iterate is small.
- Thus for small μ we expect a small number of Newton steps per outer iteration, but a large number of outer iterations.
- In this case the iterates (and indeed, the iterates of the inner iterations as well) closely follow the central path.
- This explains the alternate name path-following method.

Choice of μ (2/3)

- On the other hand, if μ is large we have the opposite situation.
- After each outer iteration t increases a large amount, so the current iterate is probably not a very good approximation of the next iterate.
- Thus we expect many more inner iterations.
- This 'aggressive' updating of t results in fewer outer iterations, since the duality gap is reduced by the large factor μ at each outer iteration, but more inner iterations.
- With μ large, the iterates are widely separated on the central path; the inner iterates veer way off the central path.

Choice of μ (3/3)

- In practice, small values of μ (i.e., near one) result in many outer iterations, with just a few Newton steps for each outer iteration.
- For μ in a fairly large range, from around 3 to 100 or so, the two effects nearly cancel, so the total number of Newton steps remains approximately constant.
- This means that the choice of μ is not particularly critical; values from around 10 to 20 or so seem to work well.

Choice of $t^{(0)}$ (1/2)

- Another important issue is the choice of initial value of t.
- If $t^{(0)}$ is chosen too large, the first outer iteration will require too many iterations. If $t^{(0)}$ is chosen too small, the algorithm will require extra outer iterations, and possibly too many inner iterations in the first centering step.
- Since $m/t^{(0)}$ is the duality gap that will result from the first centering step, one reasonable choice is to choose $t^{(0)}$ so that $m/t^{(0)}$ is approximately of the same order as $f_0(x^{(0)}) - p^*$.
- For example, if a dual feasible point λ, ν is known, with duality gap $\eta = f_0(x^{(0)}) - g(\lambda, \nu)$, then we can take $t^{(0)} = m/\eta$. Thus, in the first outer iteration we simply compute a pair with the same duality gap as the initial primal and dual feasible points.

Choice of $t^{(0)}$ (2/2)

- Another possibility is suggested by the central path condition

$$0 = t\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{\nu}.$$

- We can interpret

$$\inf_{\nu} \left|\left| t\nabla f_0(x^{(0)}) + \nabla\phi(x^{(0)}) + A^T\nu \right|\right|_2$$

as a measure for the deviation of $x^{(0)}$ from the point $x^*(t)$, and choose for $t^{(0)}$ the value that minimizes the above expression (This value of t and ν can be found by solving a least-squares problem).

Linear programming in inequality form (1/7)

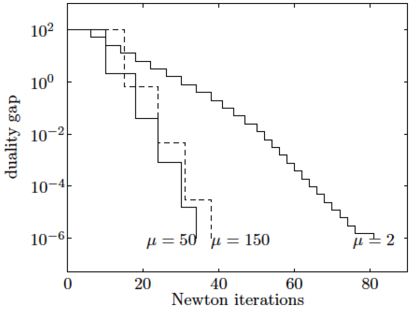
- Consider an example of a small LP in inequality form,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

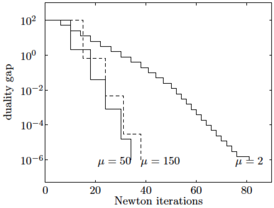
  with $A \in \mathbf{R}^{100\times 50}$.
- The data were generated randomly, so that the problem is strictly primal and dual feasible, with optimal value $p^* = 1$.
- The initial point $x^{(0)}$ is on the central path, with a duality gap of 100.
- The barrier method is terminated when the duality gap is less than 10^{-6}.
- The centering problems are solved by Newton's method with backtracking, using parameters $\alpha = 0.01, \beta = 0.5$.

Linear programming in inequality form (2/7)

- The stopping criterion for Newton's method is $\lambda(x)^2/2 \leq 10^{-5}$, where $\lambda(x)$ is the Newton decrement of the function $tc^T x + \phi(x)$.
- The progress of the barrier method, for three values of the parameter μ, is shown below.



Linear programming in inequality form (3/7)



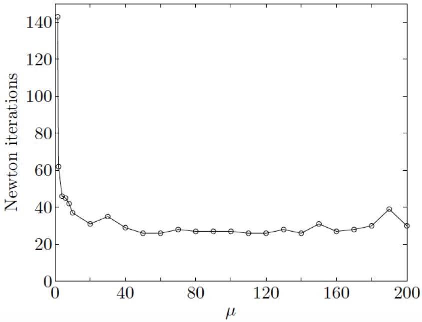
- Each of the plots has a staircase shape, with each stair associated with one outer iteration.
- The width of each stair tread (i.e., horizontal portion) is the number of Newton steps required for that outer iteration.
- The height of each stair riser (i.e., the vertical portion) is exactly equal to (a factor of) μ, since the duality gap is reduced by the factor μ at the end of each outer iteration.

Linear programming in inequality form (4/7)

- The following comments are noted.
- First of all, the method works very well, with approximately linear convergence of the duality gap.
- The plots clearly show the trade-off in the choice of μ. For $\mu = 2$, the treads are short; the number of Newton steps required to re-center is around 2 or 3. But the risers are also short, since the duality gap reduction per outer iteration is only a factor of 2.
- At the other extreme, when $\mu = 150$, the treads are longer, typically around 7 Newton steps, but the risers are also much larger, since the duality gap is reduced by the factor 150 in each outer iteration.

Linear programming in inequality form (5/7)

- The trade-off in choice of μ is further examined in the following figure.



Linear programming in inequality form (6/7)

- We use the barrier method to solve the LP, terminating when the duality gap is smaller than 10^{-3}, for 25 values of μ between 1.2 and 200. The plot shows the total number of Newton steps required to solve the problem, as a function of the parameter μ.
- This plot shows that the barrier method performs very well for a wide range of values of μ, from around 3 to 200.
- One interesting observation is that the total number of Newton steps does not vary much for values of μ larger than around 3. Thus, as μ increases over this range, the decrease in the number of outer iterations is offset by an increase in the number of Newton steps per outer iteration.

Linear programming in inequality form (7/7)

- For even larger values of μ, the performance of the barrier method becomes less predictable (i.e., more dependent on the particular problem instance). Since the performance does not improve with larger values of μ, a good choice is in the range $10 - 100$.

Geometric programming (1/4)

- We consider a geometric program in convex form,

$$\begin{array}{ll} \text{minimize} & \log\left(\sum_{k=1}^{K_0} \exp(a_{0k}^T x + b_{0k})\right) \\ \text{subject to} & \log\left(\sum_{k=1}^{K_i} \exp(a_{ik}^T x + b_{ik})\right) \leq 0, i = 1, \ldots, m, \end{array}$$

with variable $x \in \mathbf{R}^n$, and associated logarithmic barrier

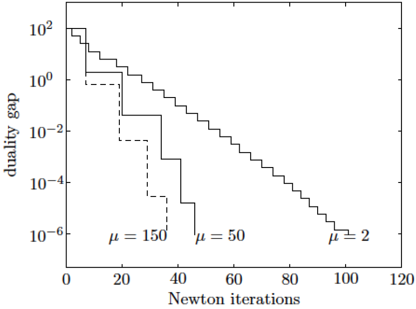
$$\phi(x) = -\sum_{i=1}^{m} \log\left(-\log \sum_{k=1}^{K_i} \exp(a_{ik}^T x + b_{ik})\right).$$

Geometric programming (2/4)

- The problem instance we consider has $n = 50$ variables and $m = 100$ inequalities.
- The objective and constraint functions all have $K_i = 5$ terms.
- The problem instance was generated randomly, in such a way that it is strictly primal and dual feasible, with optimal value one. We start with a point $x^{(0)}$ on the central path, with a duality gap of 100.
- The barrier method is used to solve the problem, with parameters $\mu = 2, \mu = 50$, and $\mu = 150$, and terminated when the duality gap is less than 10^{-6}. The centering problems are solved using Newton's method, with the same parameter values as in the LP example, i.e., $\alpha = 0.01, \beta = 0.5$, and stopping criterion $\lambda(x)^2/2 \leq 10^{-5}$.

Geometric programming (3/4)

- The following figure shows the duality gap versus cumulative number of Newton steps.



Geometric programming (4/4)

- We see an approximately constant number of Newton steps required per centering step, and therefore approximately linear convergence of the duality gap.
- The variation of the total number of Newton steps required to solve the problem, versus the parameter μ, is very similar to that in the LP example.
- For this GP, the total number of Newton steps required to reduce the duality gap below 10^{-3} is around 30 (ranging from around 20 to 40 or so) for values of μ between 10 and 200. So here, too, a good choice of μ is in the range $10 - 100$.

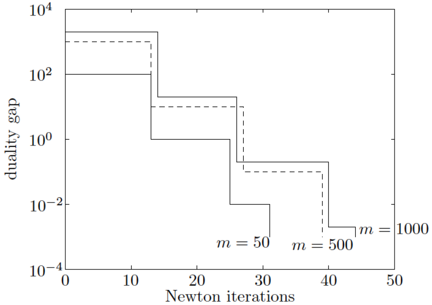
A family of standard form LPs (1/2)

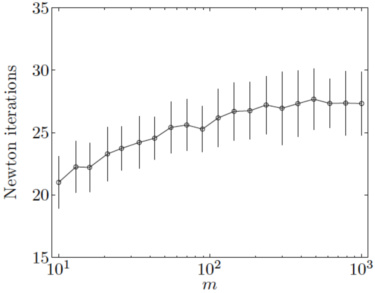
- We examine the performance of the barrier method as a function of the problem dimensions.
- We consider LPs in standard form,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

  with $A \in \mathbf{R}^{m \times n}$, and explore the total number of Newton steps required as a function of the number of variables n and number of equality constraints m, for a family of randomly generated problem instances.
- We take $n = 2m$, i.e., twice as many variables as constraints, and compare performance plots with various values of m.

A family of standard form LPs (2/2)





Left The duality gap v.s. iteration number for three problem instances, with dimensions $m = 50$, $m = 500$, and $m = 1000$.

Right The mean and standard deviation in the number of Newton steps, for each value of m.

Newton step for modified KKT equations (1/5)

- In the barrier method, the Newton step $\Delta x_{\rm nt}$, and associated dual variable are given by the linear equations

$$\begin{bmatrix} t\nabla^2 f_0(x) + \nabla^2\phi(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\rm nt} \\ \nu_{\rm nt} \end{bmatrix} = - \begin{bmatrix} t\nabla f_0(x) + \nabla\phi(x) \\ 0 \end{bmatrix}.$$

- Here we show how these Newton steps for the centering problem can be interpreted as Newton steps for directly solving the **modified KKT equations**

$$\begin{aligned} \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T\nu &= 0 \\ -\lambda_i f_i(x) &= 1/t, \quad i = 1, \ldots, m \\ Ax &= b \end{aligned}$$

in a particular way.

Newton step for modified KKT equations (2/5)

- To solve the modified KKT equations in the previous page, which is a set of $n+p+m$ nonlinear equations in the $n+p+m$ variables x, ν, and λ, we first eliminate the variables λ_i, using $\lambda_i = -1/(tf_i(x))$. This yields

$$\nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-tf_i(x)} \nabla f_i(x) + A^T \nu = 0, \quad Ax = b,$$

which is a set of $n+p$ equations in the $n+p$ variables x and ν.

Newton step for modified KKT equations (3/5)

- For v small, we have the Taylor approximation

$$\begin{aligned}
&\nabla f_0(x+v) + \sum_{i=1}^m \frac{1}{-tf_i(x+v)} \nabla f_i(x+v) \\
\approx \quad & \nabla f_0(x) + \sum_{i=1}^m \frac{1}{-tf_i(x)} \nabla f_i(x) + \nabla^2 f_0(x) v \\
&+ \sum_{i=1}^m \frac{1}{-tf_i(x)} \nabla^2 f_i(x) v + \sum_{i=1}^m \frac{1}{tf_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T v.
\end{aligned}$$

Newton step for modified KKT equations (4/5)

- The Newton step is obtained by replacing the nonlinear term in the modified KKT conditions by this Taylor approximation, which yields the linear equations

$$Hv + A^T \nu = -g, \quad Av = 0,$$

where

$$\begin{aligned} H &= \nabla^2 f_0(x) + \sum_{i=1}^m \frac{1}{-tf_i(x)} \nabla^2 f_i(x) + \sum_{i=1}^m \frac{1}{tf_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T \\ g &= \nabla f_0(x) + \sum_{i=1}^m \frac{1}{-tf_i(x)} \nabla f_i(x). \end{aligned}$$

- Now we observe that

$$H = \nabla^2 f_0(x) + (1/t)\nabla^2 \phi(x), \quad g = \nabla f_0(x) + (1/t)\nabla \phi(x).$$

- So, the Newton steps Δx_{nt} and ν_{nt} in the barrier method centering step satisfy

$$tH\Delta x_{\mathrm{nt}} + A^T \nu_{\mathrm{nt}} = -tg, \quad A\Delta x_{\mathrm{nt}} = 0.$$

Newton step for modified KKT equations (5/5)

- Comparing this with

$$Hv + A^T\nu = -g, \quad Av = 0$$

shows that

$$v = \Delta x_{\mathrm{nt}}, \quad \nu = (1/t)\nu_{\mathrm{nt}}.$$

- This shows that the Newton step for the centering problem

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

can be interpreted, after scaling the dual variable, as the Newton step for solving the modified KKT equations.