Interior-point methods (II)

Lecture 14, Nonlinear Programming

National Taiwan University

January 5, 2017

Table of contents

Feasibility and phase I methods

- The barrier method requires a strictly feasible starting point $x^{(0)}$.
- 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 (or the constraints are found to be infeasible).
- The strictly feasible point found during phase I is then used as the starting point for the barrier method, which is called the phase II stage.

Basic phase I method (1/4)

- We consider a set of inequalities and equalities in the variables $x \in \mathbf{R}^n$,

$$f_i(x) \leq 0, i = 1, ..., m, Ax = b,$$

  where $f_i : \mathbf{R}^n \rightarrow \mathbf{R}$ are convex, with continuous second derivatives.
- We assume that we are given a point $x^{(0)} \in \mathbf{dom}\ f_1 \bigcap ... \bigcap \mathbf{dom}\ f_m$, with $Ax^{(0)} = b$.
- Our goal is to find a strictly feasible solution of these inequalities and equalities, or determine that none exists.

Basic phase I method (2/4)

- To do this we form the following optimization problem:

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

  in the variables $x \in \mathbf{R}^n$, $s \in \mathbf{R}$.

- The variable s can be interpreted as a bound on the maximum infeasibility of the inequalities; the goal is to drive the maximum infeasibility below zero.
- This problem is always strictly feasible, since we can choose $x^{(0)}$ as starting point for x, and for s, we can choose any number larger than $\max_{i=1,\ldots,m} f_i(x^{(0)})$.

Basic phase I method (3/4)

- We can therefore apply the barrier method to solve the problem in the previous page, called the **phase I optimization problem** associated with the inequality and equality system.
- We can distinguish three cases depending on the sign of the optimal value $\bar{p}^*$ of the phase I problem.
  1. If $\bar{p}^* < 0$, then $f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad Ax = b$ has a strictly feasible solution. (i.e., the original problem is strictly feasible.) We do not need to solve the phase I optimization problem with high accuracy; we can terminate when $s < 0$.
  2. If $\bar{p}^* > 0$, then the original problem is infeasible. We also do not need to solve the phase I optimization problem to high accuracy; we can terminate when a dual feasible point is found with positive dual objective (which proves that $\bar{p}^* > 0$).
  3. If $\bar{p}^* = 0$ and the minimum is attained at x^* and $s^* = 0$, then the set of inequalities is feasible, but not strictly feasible.

Basic phase I method (4/4)

- If $\bar{p}^* = 0$ and the minimum is not attained, then the inequalities are infeasible.
- In practice it is impossible to determine exactly that $\bar{p}^* = 0$.
- Instead, an optimization algorithm applied to the basic phase I problem will terminate with the conclusion that $|\bar{p}^*| < \epsilon$ for some small, positive ϵ.
- This allows us to conclude that the inequalities $f_i(x) \leq -\epsilon$ are infeasible, while the inequalities $f_i(x) \leq \epsilon$ are feasible.

Sum of infeasibilities (1/3)

- There are many variations on the basic phase I method just described.
- One method is based on minimizing the sum of the infeasibilities, instead of the maximum infeasibility.
- We form the problem

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T s \\ \text{subject to} & f_i(x) \leq s_i, \quad i = 1, \ldots, m \\ & Ax = b \\ & s \succeq 0. \end{array}$$

- For fixed x, the optimal value of s_i is $\max\{f_i(x), 0\}$, so in this problem we are minimizing the sum of the infeasibilities.

Sum of infeasibilities (2/3)

- The optimal value of the sum-of-infeasibilities problem is zero and achieved if and only if the original set of equalities and inequalities is feasible. This sum of infeasibilities phase I method has a very interesting property when the system of equalities and inequalities

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

  is infeasible.
- In this case, the optimal point for the sum-of-infeasibilities phase I problem often violates only a small number, say r, of the inequalities.

Sum of infeasibilities (3/3)

- Therefore, we have computed a point that satisfies many $(m - r)$ of the inequalities, i.e., we have identified a large subset of inequalities that is feasible.
- In this case, the dual variables associated with the strictly satisfied inequalities are zero, so we have also proved infeasibility of a subset of the inequalities.
- This is more informative than finding that the m inequalities, together, are mutually infeasible.

Comparison of phase I methods (1/2)

- We apply two phase I methods to an infeasible set of inequalities $Ax \preceq b$ with dimensions $m = 100, n = 50$.
- The first method is the basic phase I method

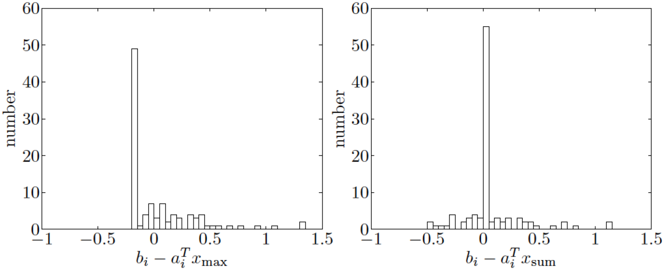
$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & Ax \preceq b + \mathbf{1}s, \end{array}$$

  which minimizes the maximum infeasibility.
- The second method minimizes the sum of the infeasibilities, i.e., solves the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T s \\ \text{subject to} & Ax \preceq b + s \\ & s \succeq 0. \end{array}$$

- The figure in the following page shows the distributions of the infeasibilities $b_i - a_i^T x$ for these two values of x, denoted x_{max} and x_{sum}, respectively.

Comparison of phase I methods (2/2)



- The point x_{max} satisfies 39 of the 100 inequalities, whereas the point x_{sum} satisfies 79 of the inequalities.

Termination near the phase II central path (1/2)

- A simple variation on the basic phase I method, using the barrier method, has the property that (when the equalities and inequalities are strictly feasible) the central path for the phase I problem intersects the central path for the original optimization problem.
- We assume a point $x^{(0)} \in \mathcal{D} = \textbf{dom } f_0 \cap \textbf{dom } f_1 \cap \ldots \cap \textbf{dom } f_m$, with $Ax^{(0)} = b$ is given.
- We form the phase I optimization problem

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

where M is a constant chosen to be larger than $\max\left\{f_0(x^{(0)}), p^*\right\}$.

Termination near the phase II central path (2/2)

- We assume now that the original problem is strictly feasible, so the optimal value $\bar{p}^*$ of the phase I problem in the previous page is negative.
- The central path of the phase I problem is characterized by

$$\sum_{i=1}^{m} \frac{1}{s - f_i(x)} = \bar{t}, \quad \frac{1}{M - f_0(x)} \nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{s - f_i(x)} \nabla f_i(x) + A^T \nu = 0,$$

  where $\bar{t}$ is the parameter.
- If (x, s) is on the central path and $s = 0$, then x and ν satisfy $t \nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) + A^T \nu = 0$ for $t = 1/(M - f_0(x))$.
- This means that x is on the central path for the original optimization problem, with associated duality gap

$$m(M - f_0(x)) \leq m(M - p^*).$$

Primal-dual interior-point methods (1/2)

- In this section, we describe a basic primal-dual interior-point method, which is very similar to the barrier method, with some differences.
  1. There is only one loop or iteration, i.e., there is no distinction between inner and outer iterations as in the barrier method. At each iteration, both the primal and dual variables are updated.
  2. The search directions in a primal-dual interior-point method are obtained from Newton's method, applied to modified KKT equations (i.e., the optimality conditions for the logarithmic barrier centering problem). The primal-dual search directions are similar to, but not quite the same as, the search directions that arise in the barrier method.
  3. In a primal-dual interior-point method, the primal and dual iterates are not necessarily feasible.

Primal-dual interior-point methods (2/2)

- Primal-dual interior-point methods are often more efficient than the barrier method, especially when high accuracy is required, since they can exhibit better than linear convergence.
- For several basic problem classes, such as linear, quadratic, second-order cone, geometric, and semidefinite programming, customized primal-dual methods outperform the barrier method.
- For general nonlinear convex optimization problems, primal-dual interior-point methods are still a topic of active research, but show great promise.

Primal-dual search direction (1/3)

- As in the barrier method, we start with the modified KKT conditions

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

expressed as $r_t(x, \lambda, \nu) = 0$, where we define

$$
r_t(x, \lambda, \nu) = \begin{bmatrix} \nabla f_0(x) + Df(x)^T \lambda + A^T \nu \\ -\mathbf{diag}\,(\lambda) f(x) - (1/t)\mathbf{1} \\ Ax - b \end{bmatrix},
$$

and $t > 0$. Here $f : \mathbf{R}^n \to \mathbf{R}^m$ and its derivative matrix Df are

$$
f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}, Df(x) = \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{bmatrix}.
$$

Primal-dual search direction (2/3)

- If x, λ, ν satisfy $r_t(x, \lambda, \nu) = 0$ (and $f_i(x) < 0$), then $x = x^*(t), \lambda = \lambda^*(t)$, and $\nu = \nu^*(t)$.
- In particular, x is primal feasible, and λ, ν are dual feasible, with duality gap m/t.
- The first block component of r_t,

$$r_{\text{dual}} = \nabla f_0(x) + Df(x)^T \lambda + A^T \nu,$$

  is called the **dual residual**, and the last block component, $r_{\text{pri}} = Ax - b$, is called the **primal residual**.
- The middle block,

$$r_{\text{cent}} = -\textbf{diag}\,(\lambda) f(x) - (1/t)\mathbf{1},$$

  is the **centrality residual**, i.e., the residual for the modified complementarity condition.

Primal-dual search direction (3/3)

- Now consider the Newton step for solving the nonlinear equations $r_t(x, \lambda, \nu) = 0$, for fixed t at a point (x, λ, ν) that satisfies $f(x) \prec 0, \lambda \succ 0$.
- The Newton step is characterized by the linear equations $r_t(y + \Delta y) \approx r_t(y) + Dr_t(y)\Delta y = 0$, i.e., $\Delta y = -Dr_t(y)^{-1} r_t(y)$.
- In terms of x, λ, and ν, we have

$$\begin{bmatrix} \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) & Df(x)^T & A^T \\ -\textbf{diag}\,(\lambda) Df(x) & -\textbf{diag}\,(f(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} r_{\mathrm{dual}} \\ r_{\mathrm{cent}} \\ r_{\mathrm{pri}} \end{bmatrix}.$$

- We note also that if x satisfies $Ax = b$, i.e., the primal feasibility residual r_{pri} is zero, then we have $A\Delta x_{\mathrm{pd}} = 0$, so Δx_{pd} defines a (primal) feasible direction: for any $s, x + s\Delta x_{\mathrm{pd}}$ will satisfy $A(x + s\Delta x_{\mathrm{pd}}) = b$.

Comparison with barrier method search directions (1/5)

- We compare the primal-dual search directions with the search directions used in the barrier method.
- We start with the linear equations

$$\begin{bmatrix} \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) & Df(x)^T & A^T \\ -\mathbf{diag}\,(\lambda) Df(x) & -\mathbf{diag}\,(f(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{pd}} \\ \Delta \lambda_{\mathrm{pd}} \\ \Delta \nu_{\mathrm{pd}} \end{bmatrix} = - \begin{bmatrix} r_{\mathrm{dual}} \\ r_{\mathrm{cent}} \\ r_{\mathrm{pri}} \end{bmatrix}.$$

  that define the primal-dual search directions.
- We eliminate the variable $\Delta\lambda_{\mathrm{pd}}$, using

$$\Delta\lambda_{\mathrm{pd}} = -\mathbf{diag}\,(f(x))^{-1} \left[\mathbf{diag}\,(\lambda) Df(x) \Delta x_{\mathrm{pd}} - r_{\mathrm{cent}}\right],$$

  which comes from the second block of equations.

Comparison with barrier method search directions (2/5)

- Substituting this into the first block of equations gives

$$\begin{bmatrix} H_{\mathrm{pd}} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{pd}} \\ \Delta \nu_{\mathrm{pd}} \end{bmatrix}$$

$$= \quad -\begin{bmatrix} r_{\mathrm{dual}} + Df(x)^T \mathbf{diag}\,(f(x))^{-1} r_{\mathrm{cent}} \\ r_{\mathrm{pri}} \end{bmatrix}$$

$$= \quad -\begin{bmatrix} \nabla f_0(x) + (1/t)\sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) + A^T \nu \\ r_{\mathrm{pri}} \end{bmatrix},$$

where

$$H_{\mathrm{pd}} = \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) + \sum_{i=1}^m \frac{\lambda_i}{-f_i(x)} \nabla f_i(x) \nabla f_i(x)^T.$$

Comparison with barrier method search directions (3/5)

- On the other hand, the equation for the centering problem in the barrier method with parameter t can be written as

$$\begin{aligned}
\begin{bmatrix} H_{\mathrm{bar}} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{bar}} \\ \nu_{\mathrm{bar}} \end{bmatrix}
&= -\begin{bmatrix} t\nabla f_0(x) + \nabla \phi(x) \\ r_{\mathrm{pri}} \end{bmatrix} \\
&= -\begin{bmatrix} t\nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) \\ r_{\mathrm{pri}} \end{bmatrix},
\end{aligned}$$

where

$$H_{\mathrm{bar}} = t\nabla^2 f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla^2 f_i(x) + \sum_{i=1}^m \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T.$$

Comparison with barrier method search directions (4/5)

- We define the variable $\Delta\nu_{\mathrm{bar}} = (1/t)\nu_{\mathrm{bar}} - \nu$ (where ν is arbitrary). Then we obtain

$$\begin{bmatrix} (1/t)H_{\mathrm{bar}} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{bar}} \\ \Delta\nu_{\mathrm{bar}} \end{bmatrix}$$

$$= - \begin{bmatrix} \nabla f_0(x) + (1/t)\sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) + A^T\nu \\ r_{\mathrm{pri}} \end{bmatrix}$$

.

Comparison with barrier method search directions (5/5)

- In this form, the righthand side is identical to the righthand side of the primal-dual equations (evaluated at the same x, λ, and ν). The coefficient matrices differ only in the 1, 1 block:

$$
\begin{aligned}
H_{\mathrm{pd}} &= \nabla^2 f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(x) + \sum_{i=1}^{m} \frac{\lambda_i}{-f_i(x)} \nabla f_i(x) \nabla f_i(x)^T, \\
(1/t) H_{\mathrm{bar}} &= \nabla^2 f_0(x) + \sum_{i=1}^{m} \frac{1}{-t f_i(x)} \nabla^2 f_i(x) + \sum_{i=1}^{m} \frac{1}{t f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T.
\end{aligned}
$$

- When x and λ satisfy $-f_i(x)\lambda_i = 1/t$, the coefficient matrices, and therefore also the search directions, coincide.

The surrogate duality gap

- In the primal-dual interior-point method the iterates $x^{(k)}, \lambda^{(k)}$, and $\nu^{(k)}$ are not necessarily feasible, except in the limit as the algorithm converges.
- This means that we cannot easily evaluate a duality gap $\eta^{(k)}$ associated with step k of the algorithm, as we do in (the outer steps of) the barrier method.
- Instead we define the **surrogate duality gap**, for any x that satisfies $f(x) \prec 0$ and $\lambda \succeq 0$, as

$$\hat{\eta}(x, \lambda) = -f(x)^T \lambda.$$

- The surrogate gap $\hat{\eta}$ would be the duality gap, if x were primal feasible and λ, ν were dual feasible, i.e., if $r_{\mathrm{pri}} = 0$ and $r_{\mathrm{dual}} = 0$. Note that the value of the parameter t that corresponds to the surrogate duality gap $\hat{\eta}$ is $m/\hat{\eta}$.

Primal-dual interior-point method (1/2)

- **Algorithm 11.2 Primal-dual interior-point method.**
  given x that satisfies
  $f_1(x) < 0, \ \ldots, \ f_m(x) < 0, \ \lambda \succ 0, \ \mu > 1, \epsilon_{feas} > 0, \epsilon > 0.$
  repeat
  1. Determine t. Set $t := \mu m/\hat{\eta}$.
  2. Compute primal-dual search direction Δy_{pd}.
  3. Line search and update.
     Determine step length $s > 0$ and set $y := y + s\Delta y_{\mathrm{pd}}$.

  until $||r_{\mathrm{pri}}||_2 \leq \epsilon_{feas}, ||r_{\mathrm{dual}}||_2 \leq \epsilon_{feas}$, and $\hat{\eta} \leq \epsilon$.
- In step 1, the parameter t is set to a factor μ times $m/\hat{\eta}$, which is the value of t associated with the current surrogate duality gap $\hat{\eta}$.

Primal-dual interior-point method (2/2)

- If x, λ, and ν were central, with parameter t (and therefore with duality gap m/t), then in step 1 we would increase t by the factor μ, which is exactly the update used in the barrier method.
- Values of the parameter μ on the order of 10 appear to work well.
- The primal-dual interior-point algorithm terminates when x is primal feasible and λ, ν are dual feasible (within the tolerance ϵ_{feas}) and the surrogate gap is smaller than the tolerance ϵ.
- Since the primal-dual interior-point method often has faster than linear convergence, it is common to choose ϵ_{feas} and ϵ small.

Line search (1/3)

- The line search in the primal-dual interior point method is a standard backtracking line search, based on the norm of the residual, and modified to ensure that $\lambda \succ 0$ and $f(x) \prec 0$.
- We denote the current iterate as x, λ, and ν, and the next iterate as x^+, λ^+, and ν^+, i.e., $x^+ = x + s\Delta x_{\mathrm{pd}}$, $\lambda^+ = \lambda + s\Delta\lambda_{\mathrm{pd}}$, $\nu^+ = \nu + s\Delta\nu_{\mathrm{pd}}$.
- The residual, evaluated at y^+, will be denoted r^+.

Line search (2/3)

- We first compute the largest positive step length, not exceeding one, that gives $\lambda^+ \succeq 0$, i.e.,

$$
\begin{aligned}
s^{max} &= \sup\{s \in [0,1] \mid \lambda + s\Delta\lambda \succeq 0\} \\
&= \min\{1, \min\{-\lambda_i/\Delta\lambda_i \mid \Delta\lambda_i < 0\}\}.
\end{aligned}
$$

  We start the backtracking with $s = 0.99s^{max}$, and multiply s by $\beta \in (0,1)$ until we have $f(x^+) \prec 0$.

- We continue multiplying s by β until we have

$$
||r_t(x^+, \lambda^+, \nu^+)||_2 \le (1 - \alpha s)||r_t(x, \lambda, \nu)||_2.
$$

Line search (3/3)

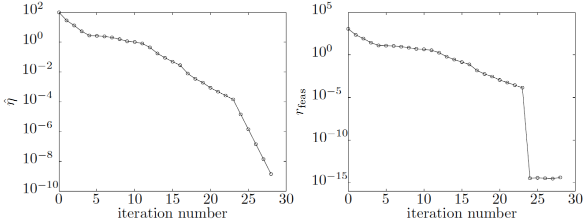
- Common choices for the backtracking parameters α and β are the same as those for Newton's method: α is typically chosen in the range 0.01 to 0.1, and β is typically chosen in the range 0.3 to 0.8.
- One iteration of the primal-dual interior-point algorithm is the same as one step of the infeasible Newton method, applied to solving $r_t(x, \lambda, \nu) = 0$, but modified to ensure $\lambda \succ 0$ and $f(x) \prec 0$ (or, equivalently, with **dom** r_t restricted to $\lambda \succ 0$ and $f(x) \prec 0$).

Examples

- We illustrate the performance of the primal-dual interior-point method for the same problems considered in the previous lecture.
- The only difference is that instead of starting with a point on the central path, we start the primal-dual interior-point method at a randomly generated $x^{(0)}$, that satisfies $f(x) \prec 0$, and take $\lambda_i^{(0)} = -1/f_i(x^{(0)})$, so the initial value of the surrogate gap is $\hat{\eta} = 100$.
- The parameter values we use for the primal-dual interior-point method are $\mu = 10,\ \beta = 0.5,\ \epsilon = 10^{-8},\ \alpha = 0.01$.

Small LP and GP (1/2)

- We first consider the small LP used in the previous lecture, with $m = 100$ inequalities and $n = 50$ variables.
- The following figure shows the progress of the primal-dual interior-point method.



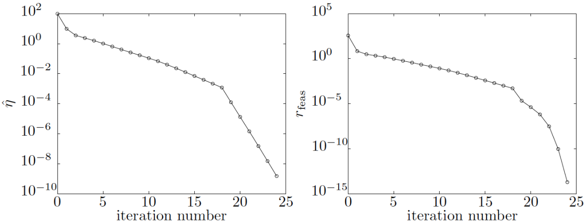
- Two plots are shown: the surrogate gap $\hat{\eta}$, and the norm of the primal and dual residuals,

$$r_{feas} = \left(||r_{\rm pri}||_2^2 + ||r_{\rm dual}||_2^2\right)^{1/2},$$

versus iteration number.

Small LP and GP (2/2)

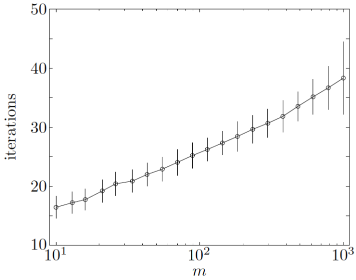
- Compared to the barrier method, the primal-dual interior-point method is faster.
- The following figure shows the progress of the primal-dual interior-point method on the GP considered in the previous lecture.



- The convergence is similar to the LP example.

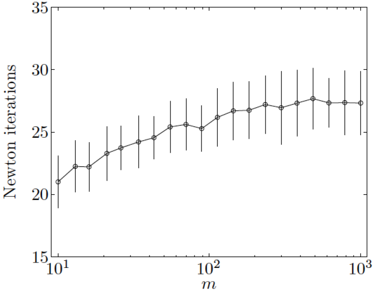
A family of LPs (1/2)

- We consider the same family of standard form LPs considered in the previous lecture, and use the primal-dual interior-point method to solve the same 2000 instances, consisting of 100 instances for each value of m.
- The primal-dual algorithm is started at $x^{(0)} = \mathbf{1}$, $\lambda^{(0)} = \mathbf{1}$, $\nu^{(0)} = 0$, and terminated using tolerance $\epsilon = 10^{-8}$.
- The following figure shows the average, and standard deviation, of the number of iterations required versus m.



A family of LPs (2/2)

- The number of iterations ranges from 15 to 35, and grows approximately as the logarithm of m.



- Comparing with the results for the barrier method shown in the above figure, we see that the number of iterations in the primal-dual method is only slightly higher, despite the fact that we start at infeasible starting points, and solve the problem to a much higher accuracy.