Interior-Point Method

Carrson C. Fung
Institute of Electronics
National Chiao Tung University

Unconstrained Minimization Problems

- In this chapter, we study *interior - point method* for solving convex optimization problems that include inequality constraints

$$
\begin{aligned}
\min_{\mathbf{x}}\ & f_0(\mathbf{x}) && (11.1)\\
\text{s.t. } & f_i(\mathbf{x}) \leq 0, \quad i = 1,\dots,m\\
& \mathbf{A}\mathbf{x} = \mathbf{b},
\end{aligned}
$$

where $f_0,\dots,f_m : \mathbb{R}^n \to \mathbb{R}$ are convex and twice continuously differentiable and $\mathbf{A} \in \mathbb{R}^{p\times n}$ with $\operatorname{rank}(\mathbf{A}) = p < n$.

- Assume that the problem is solvable, i.e. an optimal $\mathbf{x}^\star$ exists. We denote the optimal value $f_0(\mathbf{x}^\star)$ as $p^\star$.
- Assume that the problem is strictly feasible, i.e. there exists $\mathbf{x} \in \mathcal{D}$ that satisfies $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $f_i(\mathbf{x}) < 0$, for $i = 1,\dots,m$. This means that Slater's constraint qualification holds.

Inequality Constrained Minimization

- There exists dual optimal $\boldsymbol{\lambda}^\star \in \mathbb{R}^m$, $\mathbf{v}^\star \in \mathbb{R}^p$, which with $\mathbf{x}^\star$ satisfy the KKT conditions

$$
\begin{aligned}
\nabla f_0\left(\mathbf{x}^\star\right) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i\left(\mathbf{x}^\star\right) + \mathbf{A}^T \mathbf{v}^\star &= \mathbf{0} \\
f_i\left(\mathbf{x}^\star\right) &\le 0, \quad i = 1, \ldots, m \\
\mathbf{A}\mathbf{x}^\star &= \mathbf{b}, \\
\boldsymbol{\lambda}^\star &\succeq 0 \\
\lambda_i^\star f_i\left(\mathbf{x}^\star\right) &= 0, \quad i = 1, \ldots, m
\end{aligned}
\tag{11.2}
$$

- Many problems are already in the form (11.1), 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 equality constrained entropy maximization:

Inequality Constrained Minimization

$$\begin{aligned} \min_{\mathbf{x}} \; & \sum_{i=1}^{n} x_i \log x_i \\ \text{s.t. } & \mathbf{F}\mathbf{x} \preceq \mathbf{g} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned}$$

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

- Many other problems do not have the required form in (11.1), but can be reformulated in the required form. E.g. the unconstrained convex piecewise-linear minimization problem:
$$\min_{\mathbf{x}} \max_{i=1,\ldots,m} \left(\mathbf{a}_i^T \mathbf{x} + b_i\right)$$
can be recasted as
$$\begin{aligned} \min_{\mathbf{x},t} \; & t, \\ \text{s.t. } & \mathbf{a}_i^T \mathbf{x} + b_i \le t, \quad i = 1,\ldots,m \end{aligned}$$
- Other convex problems, such as SOCPs and SDPs are not readily recast in the required form, but can be handled by extensions of interior-point methods to problems with generalized equalities

Logarithm Barrier Function

- Our goal is to approximately formulate the inequality constrained problem (11.1) as an equality constrained problem to which Newton's method can be applied.
- We can rewrite (11.1) and make the inequality constraints implicit in the objective:

$$\begin{aligned} \min_{\mathbf{x}} \; & f_0(\mathbf{x}) + \sum_{i=1}^{m} I_-\left(f_i(\mathbf{x})\right) \\ \text{s.t. } & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{11.3}$$

where $I\_ : \mathbb{R} \to \mathbb{R}$ is the indicator function for the nonpositive reals,

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

- The problem (11.3) has no inequality constraints, but its objective function is not (in general) differentiable, so Newton's method cannot be applied.

Logarithm Barrier

- The basic idea of barrier method is to approximate the indicator function I_- by the function

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

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

- Function $\hat{I}_-$ is convex and nondecreasing, and (by our convention) takes on the value ∞ and $u > 0$. Most importantly, $\hat{I}_-$ is differentiable and closed: it increases to ∞ as u increases to 0.

- Fig. 11.1 shows the function I_- and the approximation $\hat{I}_-$, for several values of t. As t increases, the approximation becomes more accurate.

Logarithm Barrier

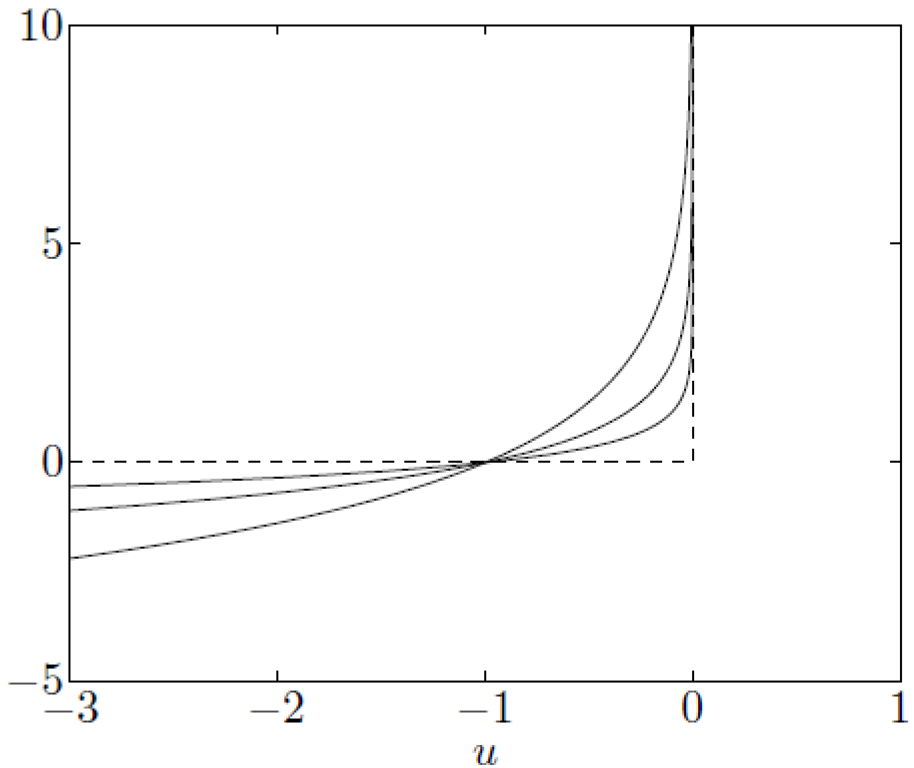

Figure 11.1 The dashed lines show the function $I_-(u)$, and the solid curves show $\widehat{I_-}(u) = -(1/t)\log(-u)$, for $t = 0.5,\ 1,\ 2$. The curve for $t = 2$ gives the best approximation.

Logarithmic Barrier

- Substituting $\hat{I}_-$ for I_- in (11.3) gives the approximation

$$\min_{\mathbf{x}} \; f_0(\mathbf{x}) + \sum_{i=1}^{m} -(1/t)\log\left(-f_i(\mathbf{x})\right)$$
$$\text{s.t. } \mathbf{Ax} = \mathbf{b}, \qquad (11.4)$$

The objective here is convex, since $-(1/t)\log(-u)$ is convex and increasing in u and differentiable.

- The function

$$\phi(\mathbf{x}) = -\sum_{i=1}^{m} \log\left(-f_i(\mathbf{x})\right), \qquad (11.5)$$

with $\operatorname{dom}\phi = \left\{\mathbf{x} \in \mathbb{R}^n \,\middle|\, f_i(\mathbf{x}) < 0,\, i = 1,\ldots,m\right\}$, is called the logarithm barrier or log barrier (11.1). Its domain is the set of points that satisfy the inequality constraints of (11.1) strictly.

Logarithm Barrier

- (11.4) is only an approximation of the original problem (11.3)
  - we will see how well solution of (11.4) will approximate that of (11.3)
- When 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.
  - Problem can be circumvented by solving a sequence of problems of the form (11.4), increasing the parameter t (and therefore the accuracy of the approximation) at each step, and starting each Newton minimization at the solution of the problem for the previous value of t
- Note that the gradient and Hessian of the log barrier function ϕ are given by

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

Central Path

- Consider the equivalent problem

$$\begin{aligned} \min_{\mathbf{x}} \; & t f_0(\mathbf{x}) + \phi(\mathbf{x}) \\ \text{s.t. } & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{11.6}$$

  which has the same minimizers as (11.4). We assume for now that the problem (11.6) can be solved via Newton's method, and, in particular, that it has a unique soution for each $t > 0$.

- For $t > 0$, we define $\mathbf{x}^\star$ as the solution of (11.6). The central path associated with (11.1) is defined as the set of points $\mathbf{x}^\star$, $t > 0$, which we call the central points.

Central Path

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

$$\mathbf{A}\mathbf{x}^\star = \mathbf{b}, \qquad f_i\left(\mathbf{x}^\star(t)\right) < 0, \quad i = 1, \ldots, m,$$

and there exists $\hat{\mathbf{v}} \in \mathbb{R}^p$ such that

$$\begin{aligned} \mathbf{0}_n &= t\nabla f_0\left(\mathbf{x}^\star(t)\right) + \nabla\phi\left(\mathbf{x}^\star(t)\right) + \mathbf{A}^T\hat{\mathbf{v}} \\ &= t\nabla f_0\left(\mathbf{x}^\star(t)\right) + \sum_{i=1}^{m} \frac{1}{-f_i\left(\mathbf{x}^\star(t)\right)} \nabla f_i\left(\mathbf{x}^\star(t)\right) + \mathbf{A}^T\hat{\mathbf{v}} \qquad (11.7) \end{aligned}$$

which is of course just the first KKT condition in (11.2). (11.7) is also called the centrality condition.

Example 11.1: Inequality Form LP

The log barrier function for an LP in inequality form

$$\begin{aligned} &\min_{\mathbf{x}} \ \mathbf{c}^T\mathbf{x} \\ &\text{s.t. } \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \end{aligned} \tag{11.8}$$

is given by $$\phi(\mathbf{x}) = -\sum_{i=1}^{m} \log\left(b_i - \mathbf{a}_i^T\mathbf{x}\right), \quad \operatorname{dom}\phi = \left\{\mathbf{x} \middle| \mathbf{A}\mathbf{x} \prec \mathbf{b}\right\},$$

where $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T$ are the rows of $\mathbf{A}$. The gradient and Hessian of the barrier function are

$$\nabla\phi(\mathbf{x}) = \sum_{i=1}^{m} \frac{1}{b_i - \mathbf{a}_i^T\mathbf{x}}\mathbf{a}_i, \qquad \nabla^2\phi(\mathbf{x}) = \sum_{i=1}^{m} \frac{1}{\left(b_i - \mathbf{a}_i^T\mathbf{x}\right)^2}\mathbf{a}_i\mathbf{a}_{\mathbf{i}}^T,$$

or more compactly

$$\nabla\phi(\mathbf{x}) = \mathbf{A}^T\mathbf{d}, \qquad \nabla^2\phi(\mathbf{x}) = \mathbf{A}^T\operatorname{Diag}(\mathbf{d})^2\mathbf{A},$$

where the elements of $\mathbf{d} \in \mathbb{R}^m$ are given by $d_i = \dfrac{1}{b_i - \mathbf{a}_i^T\mathbf{x}}$.

Example 11.1: Inequality Form LP

$$t\nabla f_0\left(\mathbf{x}^\star(t)\right)+\sum_{i=1}^{m}\frac{1}{-f_i\left(\mathbf{x}^\star(t)\right)}\nabla f_i\left(\mathbf{x}^\star(t)\right)+\mathbf{A}^T\hat{\mathbf{v}}=\mathbf{0} \qquad (11.7)$$

Since $\mathbf{x}$ is strictly feasible, we have $\mathbf{d}\succ 0$, so the Hessian of ϕ is nonsingular iff $\mathbf{A}$ has rank n. The centrality condition (11.7) is (i.e. 1st KKT condition in (11.2))

$$t\mathbf{c}+\sum_{i=1}^{m}\frac{1}{b_i-\mathbf{a}_i^T\mathbf{x}}\mathbf{a}_i=t\mathbf{c}+\mathbf{A}^T\mathbf{d}=\mathbf{0}_n. \qquad (11.9)$$

We can give a simple geometric interpretation of the centrality condition. At a point $\mathbf{x}^\star(t)$ on the central path the gradient $\nabla\phi\left(\mathbf{x}^\star(t)\right)$, which is normal to the level set of ϕ through $\mathbf{x}^\star(t)$, must be parallel to $-c$. In other words, the hyperplane $\mathbf{c}^T\mathbf{x}=\mathbf{c}^T\mathbf{x}^\star(t)$ is tangent to the level set of ϕ through $\mathbf{x}^\star(t)$. Fig. 11.2 shows an example with $m=6$ and $n=2$.

Example 11.1: Inequality Form LP

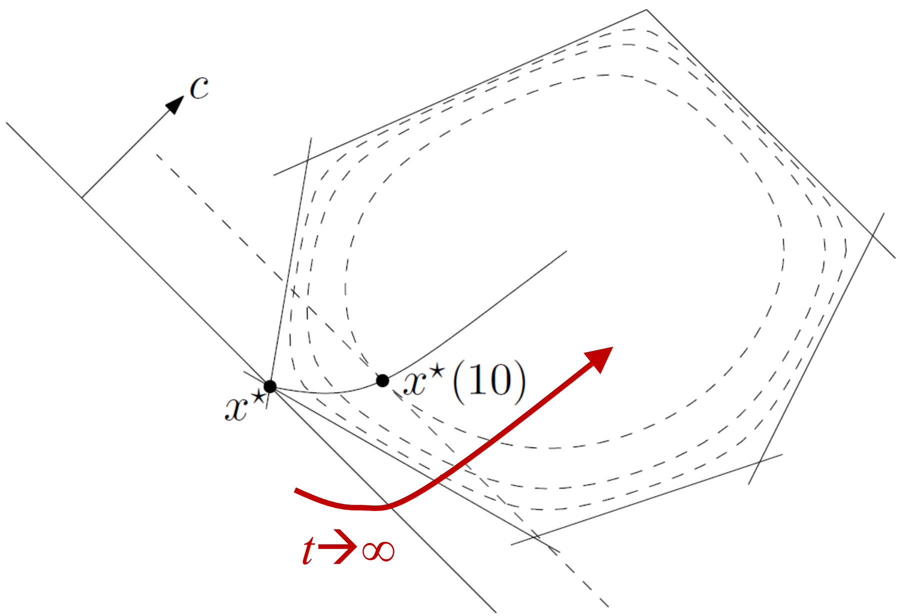

Figure 11.2 Central path for an LP with $n = 2$ and $m = 6$. The dashed curves show three contour lines of the logarithmic barrier function ϕ. The central path converges to the optimal point $x^\star$ as $t \to \infty$. Also shown is the point on the central path with $t = 10$. The optimality condition (11.9) at this point can be verified geometrically: The line $c^T x = c^T x^\star(10)$ is tangent to the contour line of ϕ through $x^\star(10)$.

Dual Points from Central Path

$$t\nabla f_0\left(\mathbf{x}^\star(t)\right)+\sum_{i=1}^{m}\frac{1}{-f_i\left(\mathbf{x}^\star(t)\right)}\nabla f_i\left(\mathbf{x}^\star(t)\right)+\mathbf{A}^T\hat{\mathbf{v}}=\mathbf{0} \qquad (11.7)$$

- From (11.7), can derive this important property: every central point yields a dual feasible point, and hence a lower bound on the optimal value $p^\star$. Define

$$\lambda_i^\star(t)\triangleq-\frac{1}{tf_i\left(\mathbf{x}^\star(t)\right)},\quad i=1,\ldots,m, \qquad \mathbf{v}^\star(t)\triangleq\frac{\hat{\mathbf{v}}}{t} \qquad (11.10)$$

  We claim that the pair $\boldsymbol{\lambda}^\star(t),\mathbf{v}^\star(t)$ is dual feasible.

- It is clear that $\boldsymbol{\lambda}^\star(t)\succ 0$ because $f_i\left(\mathbf{x}^\star(t)\right)<0,\ i=1,\ldots,m.$ By expressing the optimality conditions (11.7) as

$$\nabla f_0\left(\mathbf{x}^\star(t)\right)+\sum_{i=1}^{m}\lambda_i^\star(t)\nabla f_i\left(\mathbf{x}^\star(t)\right)+\mathbf{A}^T\mathbf{v}^\star(t)=\mathbf{0},$$

  we see that $\mathbf{x}^\star(t)$ minimizes the Lagrangian

Dual Points from Central Path

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \mathbf{v}^T(\mathbf{A}\mathbf{x} - \mathbf{b}),$$

for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^\star(t)$, $\mathbf{v} = \mathbf{v}^\star(t)$, which means $\boldsymbol{\lambda}^\star(t)$, $\mathbf{v}^\star(t)$ is a dual feasible pair. Therefore, the dual function $g\left(\boldsymbol{\lambda}^\star(t), \mathbf{v}^\star(t)\right)$ is finite, and

$$\begin{aligned} p^\star \geq g\left(\boldsymbol{\lambda}^\star(t), \mathbf{v}^\star(t)\right) &= f_0\left(\mathbf{x}^\star(t)\right) + \sum_{i=1}^{m} \lambda_i^\star(t) f_i\left(\mathbf{x}^\star(t)\right) + \mathbf{v}^\star(t)^T\left(\mathbf{A}\mathbf{x}^\star(t) - \mathbf{b}\right) \\ &= f_0\left(\mathbf{x}^\star(t)\right) - m/t, \end{aligned}$$

because $\lambda_i^\star(t) \triangleq -\dfrac{1}{t f_i\left(\mathbf{x}^\star(t)\right)}$ and $\mathbf{A}\mathbf{x}^\star(t) - \mathbf{b} = \mathbf{0}$ according to the KKT condition.

- The duality gap associated with $\mathbf{x}^\star(t)$ and the dual feasible pair $\boldsymbol{\lambda}^\star(t), \mathbf{v}^\star(t)$ is simply m/t. Hence
$$f_0\left(\mathbf{x}^\star(t)\right) - p^\star \leq m/t,$$
i.e. $\mathbf{x}^\star(t)$ is no more than m/t-suboptimal. Hence, as $t \to \infty$, $\mathbf{x}^\star(t)$ converges to optimal point.

Example 11.2: Inequality Form LP

$$\lambda_i^\star(t) \triangleq -\frac{1}{t f_i\left(\mathbf{x}^\star(t)\right)}, \quad i = 1, \ldots, m$$

$$p^\star \geq g\left(\boldsymbol{\lambda}^\star(t), \mathbf{v}^\star(t)\right) = f_0\left(\mathbf{x}^\star(t)\right) + \sum_{i=1}^{m} \lambda_i^\star(t) f_i\left(\mathbf{x}^\star(t)\right) + \mathbf{v}^\star(t)^T\left(\mathbf{A}\mathbf{x}^\star(t) - \mathbf{b}\right)$$
$$= f_0\left(\mathbf{x}^\star(t)\right) - m/t$$

The dual of the inequality form LP (11.8) is

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \ & -\mathbf{b}^T\boldsymbol{\lambda} \\ \text{s.t. } & \mathbf{A}^T\boldsymbol{\lambda} + \mathbf{c} = \mathbf{0} \\ & \boldsymbol{\lambda} \succeq 0. \end{aligned}$$

From the optimality conditions (11.9), it is clear that

$$\lambda_i^\star(t) = \frac{1}{t\left(b_i - \mathbf{a}_i^T\mathbf{x}^\star(t)\right)}, \quad i = 1, \ldots, m$$

is dual feasible, with dual objective

$$-\mathbf{b}^T\boldsymbol{\lambda}^\star(t) = \mathbf{c}^T\mathbf{x}^\star(t) + \left(\mathbf{A}\mathbf{x}^\star(t) - \mathbf{b}\right)^T\boldsymbol{\lambda}^\star(t) = \mathbf{c}^T\mathbf{x}^\star(t) - m/t$$

Interpretation via KKT Conditions

- We can also interpret the central path conditions (11.7) as continuous deformation of the the KKT optimality conditions (11.2). A point $\mathbf{x}$ is equal to $\mathbf{x}^\star(t)$ iff there exists $\boldsymbol{\lambda}, \mathbf{v}$ such that

$$\begin{aligned} \nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}) + \mathbf{A}^T \mathbf{v}^\star &= \mathbf{0} \\ \mathbf{A}\mathbf{x} = \mathbf{b}, \quad f_i(\mathbf{x}) \le 0, \quad i &= 1, \ldots, m \\ \boldsymbol{\lambda} &\succeq 0 \\ -\lambda_i f_i(\mathbf{x}) = 1/t, \quad i &= 1, \ldots, m \end{aligned} \tag{11.11}$$

- The only difference between the KKT conditions (11.2) and the centrality conditions (11.11) is that the complementary slackness condition $-\lambda_i f_i(\mathbf{x}) = 0$ is replaced by $-\lambda_i f_i(\mathbf{x}) = 1/t$.
- In particular, for large t, $\mathbf{x}^\star(t)$ and the associated dual point $\boldsymbol{\lambda}^\star(t), \mathbf{v}^\star(t)$ "almost" satisfy the KKT optimality conditions for (11.1).

Barrier Method

Recall from (11.6):
$$\min_{\mathbf{x}} \; tf_0(\mathbf{x}) + \phi(\mathbf{x})$$
$$\text{s.t. } \mathbf{Ax} = \mathbf{b},$$

- We can solve this problem by solving a sequence of unconstrained (or linearly constrained) minimizaton problems, using the last point found as the starting point for the next unconstrained minimization problem
- In other words, compute $\mathbf{x}^\star(t)$ 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

---

Algorithm 11.1 *Barrier method.*

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

repeat

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

Barrier Method

- At each iteration, except for the first, compute the central point $\mathbf{x}^\star(t)$ starting from the previously computed central point, and then increase t by a factor of $\mu > 1$.
  - the algorithm can also return $\boldsymbol{\lambda} = \boldsymbol{\lambda}^\star(t)$, and $\mathbf{v} = \mathbf{v}^\star(t)$, a dual ϵ-suboptimal point or certificate for $\mathbf{x}$
- Step 1 is called centering step since a central point is being computed
  - first centering step computes $\mathbf{x}^\star\left(t^{(0)}\right)$
  - centering usually done using Newton's method
- the dual feasible point will be computed at the end of the outer iteration
- terminates with $f_0(\mathbf{x}) - p^\star \le \epsilon$ (stopping criterion follows from $f\left(\mathbf{x}^\star(t)\right) - p^\star \le m/t$
- choice of μ involves a trade-off: large μ means fewer outer iterations, more inner (Newton) iterations; typical values: $\mu = 10 - 20$
- several heuristics for choice of $t^{(0)}$

Example: LP in Inequality Form

The first example is a small LP in inequality form

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{c}^T \mathbf{x} \\ \text{s.t. } & \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \end{aligned}$$

with $\mathbf{A} \in \mathbb{R}^{100\times 50}$. The data were randomly generated, in such a way that the problem is strictly primal and dual feasible, with optimal value $p^\star = 1$

- The initial point $\mathbf{x}^{(0)}$ is on the central path, with a duality gap of 100. The barrier method is used to solve the problem, and terminated when the duality gap is less than 10^{-6}. The centering problems are solved by Newton's method with backtracking, using parameters $\alpha = 0.01, \beta = 0.5$.
- The stopping criterion for Newton's method is $\lambda(\mathbf{x})^2 / 2 \leq 10^{-5}$, where $\lambda(\mathbf{x})$ is the Newton decrement of the function $t\mathbf{c}^T\mathbf{x} + \phi(\mathbf{x})$.

Example: LP in Inequality Form

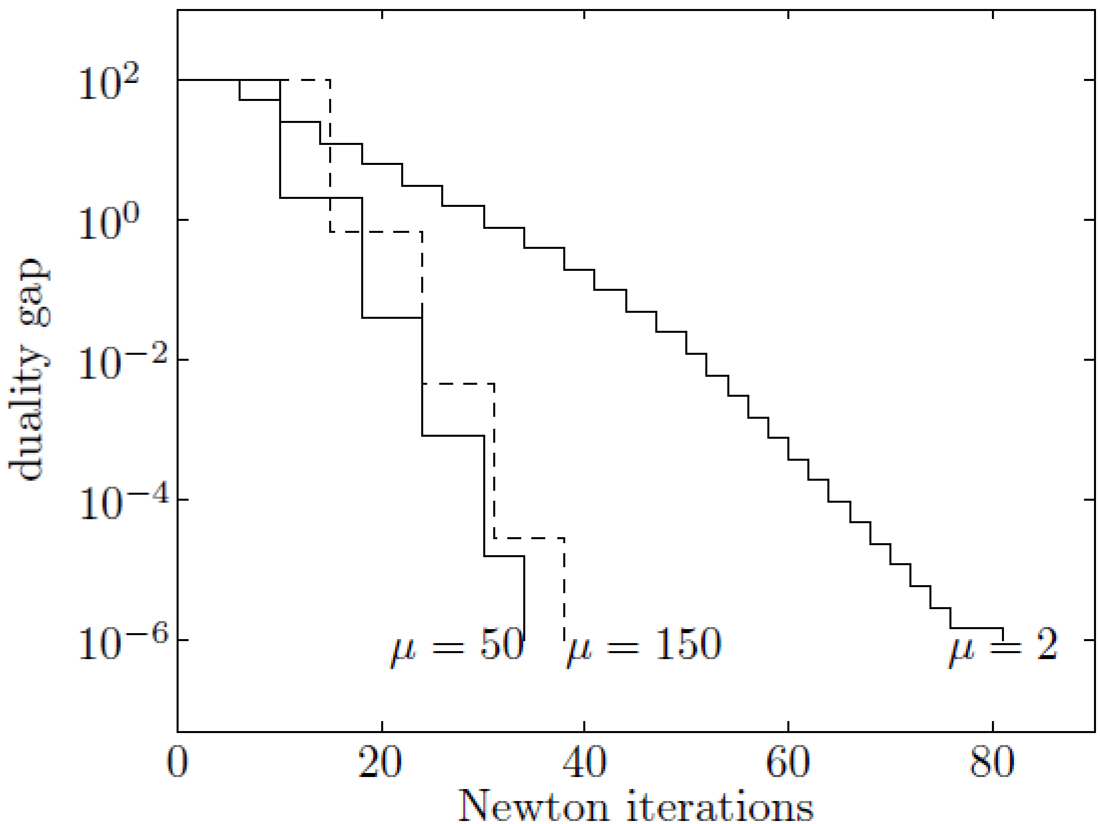

Figure 11.4 Progress of barrier method for a small LP, showing duality gap versus cumulative number of Newton steps. Three plots are shown, corresponding to three values of the parameter μ: 2, 50, and 150. In each case, we have approximately linear convergence of duality gap.

Example: Geometric Programming

- Consider a geometric program in convex form,

$$\min_{\mathbf{x}} \log\left(\sum_{k=1}^{K_0} \exp\left(\mathbf{a}_{0k}^T \mathbf{x} + b_{0k}\right)\right)$$

$$\text{s.t.} \log\left(\sum_{k=1}^{K_i} \exp\left(\mathbf{a}_{ik}^T \mathbf{x} + b_{ik}\right)\right) \le 0, \quad i = 1, \ldots, m$$

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

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

- The problem instance we consider has $n = 50$ variables and $m = 100$ inequalities (like the small LP considered above). 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 $p^\star = 1$.
- We start with a point $\mathbf{x}^{(0)}$ on the central path, with a duality gap of 100.

Example: Geometric Programming

- The barrier method is used to solve the problem, with parameters $\mu = 2, 50$ and 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 $\lambda(\mathbf{x})^2 / 2 \le 10^{-5}$.

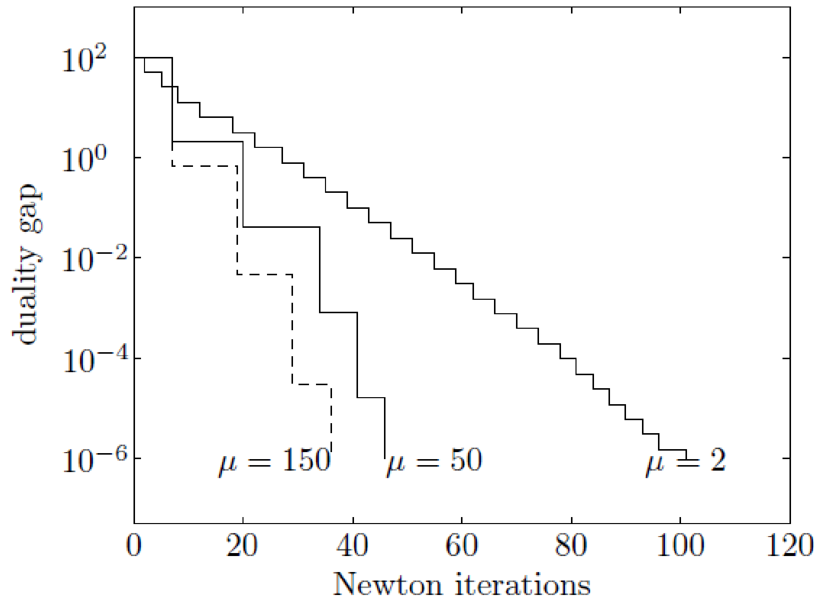

Figure 11.6 Progress of barrier method for a small GP, showing duality gap versus cumulative number of Newton steps. Again we have approximately linear convergence of duality gap.

Convergence Analysis of Barrier Method

- Assuming that $tf_0 + \phi$ can be minimized by Newton's method for $t = t^{(0)}, \mu t^{(0)}, \mu^2 t^{(0)}, \ldots,$ the duality gap after the initial centering step, and k additional centering steps, is $m / \left(\mu^k t^{(0)}\right)$. Therefore the desired accuracy ϵ is achieved after exactly

$$\left\lceil \frac{\log\left(m / \left(\epsilon t^{(0)}\right)\right)}{\log \mu} \right\rceil$$

centering steps, plus the initial centering step (to compute $\mathbf{x}^\star\left(t^{(0)}\right)$)

- Centering problem: $\min_{\mathbf{x}} tf_0\left(\mathbf{x}\right) + \phi\left(\mathbf{x}\right)$

See convergence analysis of Newton's method

- For standard Newton method, it suffices that for $t \geq t^{(0)}$, the function $tf_0 + \phi$ satisfies the condition that initial sublevel set (i.e. $t \geq t^{(0)}$) is closed, the associated inverse KKT matrix is bounded, and the Hessian satisfies Lipschitz condition (pp. 62, lec06_unconst_min.pdf)

Newton Step for Modified KKT Equations

- In the barrier method, the Newton step $\Delta\mathbf{x}_{\text{nt}}$, and associated dual variable are given by the linear equations

$$\begin{bmatrix} t\nabla^2 f_0(\mathbf{x}) + \nabla^2\phi(\mathbf{x}) & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_{\text{nt}} \\ \mathbf{v}_{\text{nt}} \end{bmatrix} = -\begin{bmatrix} t\nabla f_0(\mathbf{x}) + \nabla\phi(\mathbf{x}) \\ \mathbf{0} \end{bmatrix} \qquad (11.14)$$

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

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

$$-\lambda_i f_i(\mathbf{x}) = 1/t, \quad i = 1,\ldots,m$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \qquad (11.15)$$

in a particular way. To solve the modified KKT equations (11.15), which is a set of $n+p+m$ nonlinear equations in the $n+p+m$ variables $\mathbf{x}$, $\mathbf{v}$, $\boldsymbol{\lambda}$, we first eliminate the variables λ_i using the definition $\lambda_i = -1/\left(tf_i(\mathbf{x})\right)$. This yields

Newton Step for Modified KKT Equations

$$\nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{-tf_i(\mathbf{x})} \nabla f_i(\mathbf{x}) + \mathbf{A}^T \mathbf{v} = \mathbf{0}, \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{11.16}$$

which is a set of $n+p$ equations in the $n+p$ variables $\mathbf{x}$ and $\mathbf{v}$.

- To find the Newton step for solving the set of nonlinear equations (11.16), we form the Taylor approximation for the nonlinear term occurring in the first equation. For $\mathbf{v}$ small, we have the Taylor approximation

$$\nabla f_0(\mathbf{x}+\mathbf{v}) + \sum_{i=1}^{m} \frac{1}{-tf_i(\mathbf{x}+\mathbf{v})} \nabla f_i(\mathbf{x}+\mathbf{v})$$

$$\approx \nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{-tf_i(\mathbf{x})} \nabla f_i(\mathbf{x}) + \nabla^2 f_0(\mathbf{x})\mathbf{v}$$

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

Newton Step for Modified KKT Equations

- The Newton step is obtained by replacing the nonlinear term in equation (11.16) by this Taylor approximation, which yields the linear equations

$$\mathbf{Hv} + \mathbf{A}^T\mathbf{v} = -\mathbf{g}, \qquad \mathbf{Av} = \mathbf{0}, \qquad (11.17)$$

where

$$\mathbf{H} = \nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{-tf_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{tf_i(\mathbf{x})^2} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^T$$

$$\mathbf{g} = \nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{-tf_i(\mathbf{x})} \nabla f_i(\mathbf{x}).$$

so, from (11.14), the Newton steps $\Delta\mathbf{x}_{\text{nt}}$ and $\mathbf{v}_{\text{nt}}$ in the barrier method centering step satisfy

$$t\mathbf{H}\Delta\mathbf{x}_{\text{nt}} + \mathbf{A}^T\mathbf{v}_{\text{nt}} = -t\mathbf{g}, \qquad \mathbf{Av}_{\text{nt}} = \mathbf{0}.$$

Comparing this with (11.17) shows that $\mathbf{v} = \Delta\mathbf{x}_{\text{nt}}, \qquad \mathbf{v}_{\text{nt}} = (1/t)\mathbf{v}_{\text{nt}}$.

This shows the Newton step for the centering problem (11.6) can be interpreted, after scaling the dual variable, as the Newton step for solving the modified KKT equations (11.16).

Feasibility and Phase I Methods

- The barrier method requires a strictly feasible starting point $\mathbf{x}^{(0)}$. 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: We consider a set of inequalities and equalities in the variables $\mathbf{x} \in \mathbb{R}^n$

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m, \qquad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{11.18}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ are convex, with continuous second derivatives. We assume that we are given a point $\mathbf{x}^{(0)} \in \mathrm{dom}\, f_1 \bigcap \cdots \bigcap \mathrm{dom}\, f_m$ with $\mathbf{A}\mathbf{x}^{(0)} = \mathbf{b}$.

Phase I Method

- Feasibility problem: find $\mathbf{x}$ such that

$$f_i(\mathbf{x}) \le 0, \quad i = 1, \dots, m, \qquad \mathbf{A}\mathbf{x} = \mathbf{b} \qquad (*)$$

Phase I: Computes strictly feasible starting point for barrier method

- Basic Phase I Method

$$\begin{aligned} \min_{\mathbf{x}, s} \ & s \\ \text{s.t. } & f_i(\mathbf{x}) \le s, \quad i = 1, \dots, m \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \qquad (11.19)$$

  - if $\mathbf{x}$, s are feasible, with $s < 0$, then $\mathbf{x}$ is strictly feasible for (*)
  - if optimal value $\overline{p}^\star$ of (11.19) is positive, then problem (*) is infeasible
  - if $\overline{p}^\star = 0$ and attained, then problem (*) is feasible (but not strictly)

    if $\overline{p}^\star = 0$ and not attained, then problem (*) is infeasible

Phase I Method: Sum of Infeasibilities

- 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

$$\begin{aligned} &\min_{\mathbf{x},\,\mathbf{s}} \ \mathbf{1}^T \mathbf{s} \\ &\text{s.t.}\ f_i(\mathbf{x}) \le s_i, \quad i = 1, \ldots, m \\ &\quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{s} \succeq 0 \end{aligned} \tag{11.20}$$

- For fixed $\mathbf{x}$, the optimal value of s_i is $\max\left\{ f_i(\mathbf{x}), 0 \right\}$, so in this problem we are minimizing the sum of the infeasibilities. The optimal value of (11.20) is zero and achieved iff 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 (11.19) is infeasible. In this case, the optimal point for the phase I problem (11.20) often violates only a small number, say r, of the inequalities

Phase I Method: Sum of Infeasibilities

- 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.

Example 11.4: Comparison of Phase I Methods

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

$$\begin{aligned} &\min_{\mathbf{x},\, s}\ s \\ &\text{ s.t. } \mathbf{Ax} \preceq \mathbf{b} + \mathbf{1}s, \end{aligned}$$

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

$$\begin{aligned} &\min_{\mathbf{x},\, \mathbf{s}}\ \mathbf{1}^T\mathbf{s} \\ &\text{ s.t. } \mathbf{Ax} \preceq \mathbf{b} + \mathbf{s}, \\ &\qquad\quad \mathbf{s} \succeq 0. \end{aligned}$$

Fig. 11.9 shows the distributions of the infeasibilities $b_i - \mathbf{a}_i^T\mathbf{x}$ for these two values of $\mathbf{x}$, denotes as $\mathbf{x}_{\text{max}}$ and $\mathbf{x}_{\text{sum}}$, respectively. The point $\mathbf{x}_{\text{max}}$ satisfies 30 of the 100 inequalities whereas the point $\mathbf{x}_{\text{sum}}$ satisfies 79 of the inequalities.

Example 11.4: Comparison of Phase I Methods

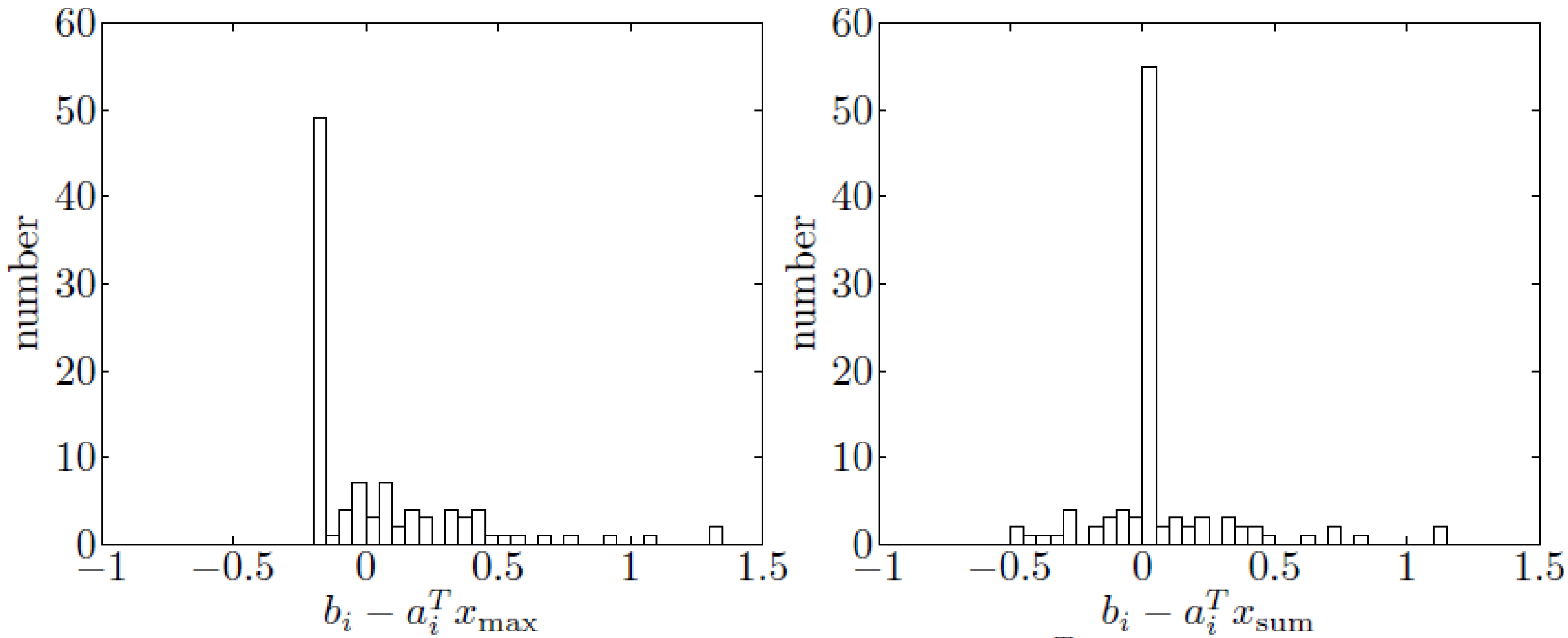

Figure 11.9 Distributions of the infeasibilities $b_i - a_i^T x$ for an infeasible set of 100 inequalities $a_i^T x \leq b_i$, with 50 variables. The vector $x_{\max}$ used in the left plot was obtained by the basic phase I algorithm. It satisfies 39 of the 100 inequalities. In the right plot the vector x_{sum} was obtained by minimizing the sum of the infeasibilities. This vector satisfies 79 of the 100 inequalities.

Phase I via Infeasible Start Newton Method

Problem domain: $\mathcal{D} = \left\{ \bigcap_{i=0}^{m} \operatorname{dom} f_i \right\} \bigcap \left\{ \bigcap_{i=1}^{p} \operatorname{dom} h_i \right\}$

We can also carry out the phase I stage using an infeasible start Newton method, applied to a modified version of the original problem

$$\begin{aligned} \min_{\mathbf{x}} \ & f_0(\mathbf{x}) \\ \text{s.t.} \ & f_i(\mathbf{x}) \le 0, \quad i = 1, \ldots, m \\ & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned}$$

We first express the problem in the (obviously equivalent) form

$$\begin{aligned} \min_{\mathbf{x},s} \ & f_0(\mathbf{x}) \\ \text{s.t.} \ & f_i(\mathbf{x}) \le s, \quad i = 1, \ldots, m \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \quad s = 0, \end{aligned}$$

with the additional variable $s \in \mathbb{R}$. To start the barrier method, we use an infeasible start Newton method to solve

$$\begin{aligned} \min_{\mathbf{x}} \ & t^{(0)} f_0(\mathbf{x}) - \sum_{i=1}^{m} \log\left(s - f_i(\mathbf{x})\right) \\ \text{s.t.} \ & \mathbf{A}\mathbf{x} = \mathbf{b}, \quad s = 0. \end{aligned}$$

This can be initialized with any $\mathbf{x} \in \mathcal{D}$, and any $s > \max_i f_i(\mathbf{x})$.

Example

- We consider a family of linear feasibility problems,

$$\mathbf{A}\mathbf{x} \preceq \mathbf{b}\left(\gamma\right),$$

where $\mathbf{A} \in \mathbb{R}^{50\times 20}$ and $\mathbf{b}\left(\gamma\right) = \mathbf{b} + \gamma\Delta\mathbf{b}$. The problem data are chosen so that the inequalities are strictly feasible for $\gamma > 0$ and infeasible for $\gamma < 0$. For $\gamma = 0$, the problem is feasible but not strictly feasible.

- Fig. 11.10 shows the total number of Newton steps required to find a strictly feasible point, or a certificate of infeasibility, for 40 values of γ in [-1,1].
- We use the basic phase I method, i.e. for each value of γ, we form the LP

$$\begin{aligned} &\min_{\mathbf{x},\, s} s \\ &\text{s.t. } \mathbf{A}\mathbf{x} \preceq \mathbf{b}\left(\gamma\right) + \mathbf{1}s. \end{aligned}$$

The barrier method is used with $\mu = 10$, and starting point $\mathbf{x} = \mathbf{0}$, $s = -\min_i b_i\left(\gamma\right) + 1$.

Example

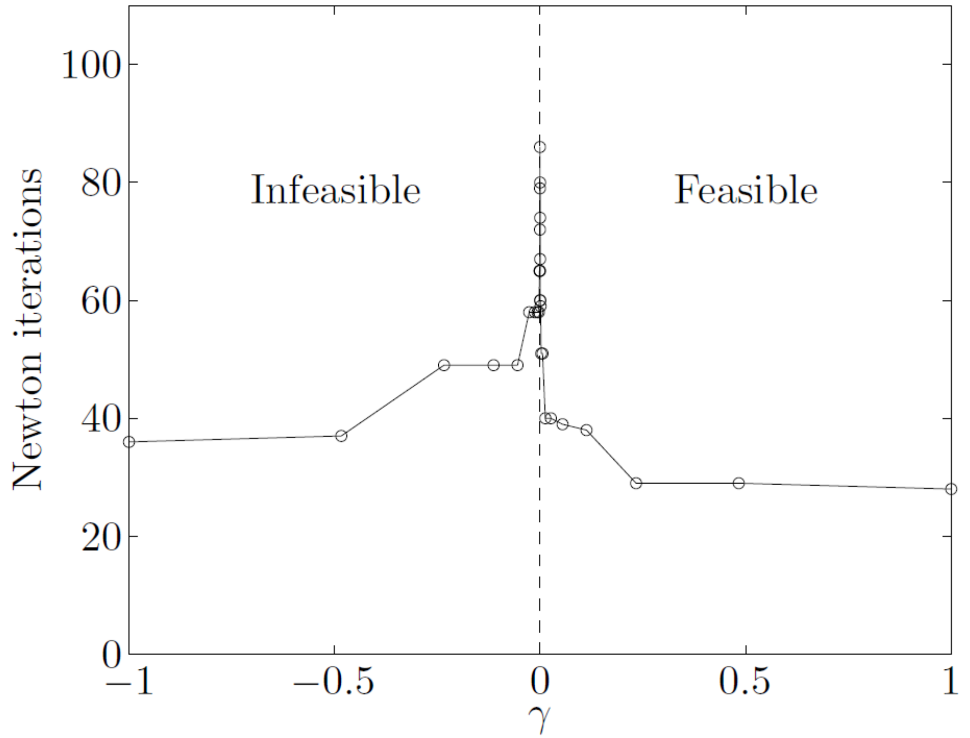

Figure 11.10 Number of Newton iterations required to detect feasibility or infeasibility of a set of linear inequalities $Ax \preceq b + \gamma\Delta b$ parametrized by $\gamma \in \mathbf{R}$. The inequalities are strictly feasible for $\gamma > 0$, and infeasible for $\gamma < 0$. For γ larger than around 0.2, about 30 steps are required to compute a strictly feasible point; for γ less than -0.5 or so, it takes around 35 steps to produce a certificate proving infeasibility. For values of γ in between, and especially near zero, more Newton steps are required to determine feasibility.

Problems with Generalized Inequalities

- We consider the problem

$$\begin{aligned} \min_{\mathbf{x}} \ & f_0(\mathbf{x}) \\ \text{s.t.}\ & f_i(\mathbf{x}) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{11.38}$$

where $f_0 : \mathbb{R}^n \to \mathbb{R}$ is convex, $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$, $i = 1, \ldots, k$ are K_i-convex, and $K_i \subseteq \mathbb{R}^{k_i}$ are proper cones. We assume f_i's are twice continuously differentiable, that $\mathbf{A} \in \mathbb{R}^{p \times n}$ with $\operatorname{rank}(\mathbf{A}) = p$, and that the problem is solvable

Problems with Generalized Inequalities

- The KKT conditions for (11.38) are

$$\begin{aligned} \mathbf{A}\mathbf{x}^\star &= \mathbf{b}, \\ f_i\left(\mathbf{x}^\star\right) &\preceq_{K_i} 0, \quad i=1,\ldots,m \\ \boldsymbol{\lambda}_i^\star &\succeq_{K_i^\star} 0, \quad i=1,\ldots,m \\ \nabla f_0\left(\mathbf{x}^\star\right)+\sum_{i=1}^{m} Df_i\left(\mathbf{x}^\star\right)^T \boldsymbol{\lambda}_i^\star + \mathbf{A}^T\mathbf{v}^\star &= \mathbf{0} \\ \boldsymbol{\lambda}_i^{\star T} f_i\left(\mathbf{x}^\star\right) &= 0, \quad i=1,\ldots,m, \end{aligned} \tag{11.39}$$

where $Df_i\left(\mathbf{x}^\star\right) \in \mathbb{R}^{k_i \times n}$ is the derivative of f_i at $\mathbf{x}^\star$.

- We will assume the problem (11.38) is strictly feasible, so the KKT conditions are necessary and sufficient conditions for optimality of $\mathbf{x}^\star$.
- The development of the method is parallel to the case with scalar constraints. In [Sec. 11.6, BV04], a generalization of the logarithm function that applies to general proper cones are developed, and the log barrier function for (11.38) is defined. Then the barrier method for (11.38) is essentially the same as in the scalar case.

Primal-dual Interior Point Methods

- Primal-dual interior-point methods are very similar to the barrier method, with some differences
  - 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
  - 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 logarithm 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
  - In a primal-dual interior-point method, the primal and dual iterates are not necessarily feasible

Primal-dual Interior Point Methods

- Primal-dual IPMs 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 IPMs are still a topic of active research, but show great promise
- Another advantage of prima-dual algorithms over the barrier method is that they can work when the problem is feasible, but not strictly feasible (although we not pursue this).

Primal-dual Search Direction

$$\nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}) + \mathbf{A}^T \mathbf{v} = \mathbf{0}$$
$$-\lambda_i f_i(\mathbf{x}) = 1/t, \quad i = 1,\ldots,m$$
$$\mathbf{Ax} = \mathbf{b}, \qquad (11.15)$$

- As in the barrier method, we start with the modified KKT conditions (11.15), expressed as $\mathbf{r}_t(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = \mathbf{0}$, where we define

$$\mathbf{r}_t(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = \begin{bmatrix} \nabla f_0(\mathbf{x}) + Df(\mathbf{x})^T \boldsymbol{\lambda} + \mathbf{A}^T \mathbf{v} \\ -\mathrm{Diag}(\boldsymbol{\lambda}) \boldsymbol{f}(\mathbf{x}) - (1/t)\mathbf{1} \\ \mathbf{Ax} - \mathbf{b} \end{bmatrix} \qquad (11.53)$$

where $t > 0$. Here $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ and its derivative matrix $D\boldsymbol{f}$ are given by

$$\boldsymbol{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}, \qquad \nabla \boldsymbol{f}(\mathbf{x}) = D\boldsymbol{f}(\mathbf{x})^T = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla f_1(\mathbf{x})^T \\ \vdots \\ \nabla f_m(\mathbf{x})^T \end{bmatrix}.$$

If $\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}$ satisfy $\mathbf{r}_t(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = \mathbf{0}$ (and $f_i(\mathbf{x}) < 0$), then $\mathbf{x} = \mathbf{x}^\star(t)$, $\boldsymbol{\lambda} = \boldsymbol{\lambda}^\star(t)$, and $\mathbf{v} = \mathbf{v}^\star(t)$. In particular, $\mathbf{x}$ is primal feasible, and $\boldsymbol{\lambda}, \mathbf{v}$ are dual feasible, with duality gap m/t. The first block component of $\mathbf{r}_t$,

Primal-dual Search Direction

$$\mathbf{r}_{\text{dual}} = \nabla f_0(\mathbf{x}) + D\boldsymbol{f}(\mathbf{x})^T \boldsymbol{\lambda} + \mathbf{A}^T \mathbf{v},$$

is called the dual residual, and the last block, $\mathbf{r}_{\text{pri}} = \mathbf{A}\mathbf{x} - \mathbf{b}$ is called the primal residual. The middle block,

$$\mathbf{r}_{\text{cent}} = -\text{Diag}(\boldsymbol{\lambda})\,\boldsymbol{f}(\mathbf{x}) - (1/t)\mathbf{1},$$

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

Now consider the Newton step for solving the nonlinear equations $\mathbf{r}_t(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = \mathbf{0}$ for fixed t, (without first eliminating $\boldsymbol{\lambda}$), at a point $(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v})$ that satisfies $f(\mathbf{x}) \prec 0$, $\boldsymbol{\lambda} \succ 0$. We will denote the current point and Newton step as

$$\mathbf{y} = \begin{bmatrix} \mathbf{x}^T & \boldsymbol{\lambda}^T & \mathbf{v}^T \end{bmatrix}^T, \qquad \Delta\mathbf{y} = \begin{bmatrix} \Delta\mathbf{x}^T & \Delta\boldsymbol{\lambda}^T & \Delta\mathbf{v}^T \end{bmatrix}^T,$$

respectively. The Newton step is characterized by the linear equations

$$\mathbf{r}_t(\mathbf{y} + \Delta\mathbf{y}) \approx \mathbf{r}_t(\mathbf{y}) + D\mathbf{r}_t(\mathbf{y})\Delta\mathbf{y} = \mathbf{0}, \qquad \Rightarrow \qquad \Delta\mathbf{y} = -D\mathbf{r}_t^{-1}\mathbf{r}_t(\mathbf{y})$$

Primal-dual Search Direction

- In terms of $\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}$, we have

$$\begin{bmatrix} \nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(\mathbf{x}) & D\boldsymbol{f}(\mathbf{x})^T & \mathbf{A}^T \\ -\mathrm{Diag}(\boldsymbol{\lambda}) D\boldsymbol{f}(\mathbf{x}) & -\mathrm{Diag}(\boldsymbol{f}(\mathbf{x})) & \mathbf{0} \\ \mathbf{A} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\boldsymbol{\lambda} \\ \Delta\mathbf{v} \end{bmatrix} = -\begin{bmatrix} \Delta\mathbf{r}_{\text{dual}} \\ \Delta\mathbf{r}_{\text{cent}} \\ \Delta\mathbf{r}_{\text{pri}} \end{bmatrix} \qquad (11.54)$$

  The primal-dual search direction $\Delta\mathbf{y}_{\text{pd}} = \begin{bmatrix} \Delta\mathbf{x}_{\text{pd}}^T & \Delta\boldsymbol{\lambda}_{\text{pd}}^T & \Delta\mathbf{v}_{\text{pd}}^T \end{bmatrix}^T$ is defined as the solution for (11.54).

- The primal and dual search directions are coupled, both through the coefficient matrix and the residuals. For example, the primal search direction $\Delta\mathbf{x}_{\text{pd}}$ depends on the current value of the dual variables $\boldsymbol{\lambda}$ and $\mathbf{v}$, as well as $\mathbf{x}$.
- We note also that if $\mathbf{x}$ satisfies $\mathbf{Ax} = \mathbf{b}$, i.e. the primal feasibility residual $\mathbf{r}_{\text{pri}}$ is zero, then we have $\mathbf{A}\Delta\mathbf{x}_{\text{pd}} = \mathbf{0}$, so $\Delta\mathbf{x}_{\text{pd}}$ defines a (primal) feasible direction: for any s, $\mathbf{x} + s\Delta\mathbf{x}_{\text{pd}}$ will satisfy $\mathbf{A}\left(\mathbf{x} + s\Delta\mathbf{x}_{\text{pd}}\right) = \mathbf{b}$.

Primal-dual Search Direction

$$\begin{bmatrix} \nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(\mathbf{x}) & D\boldsymbol{f}(\mathbf{x})^T & \mathbf{A}^T \\ -\mathrm{Diag}(\boldsymbol{\lambda}) D\boldsymbol{f}(\mathbf{x}) & -\mathrm{Diag}(\boldsymbol{f}(\mathbf{x})) & \mathbf{0} \\ \mathbf{A} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\boldsymbol{\lambda} \\ \Delta\boldsymbol{\nu} \end{bmatrix} = -\begin{bmatrix} \Delta\mathbf{r}_{\mathrm{dual}} \\ \Delta\mathbf{r}_{\mathrm{cent}} \\ \Delta\mathbf{r}_{\mathrm{pri}} \end{bmatrix} \qquad (11.54)$$

- **Comparison with barrier method search directions**: The primal-dual search directions are closely related to the search directions used in the barrier method, but not the same. We start with the linear equations (11.54) that define the primal-dual search directions. We eliminate the variable $\Delta\boldsymbol{\lambda}_{\mathrm{pd}}$, using the second block of equations with $\Delta\boldsymbol{\lambda}_{\mathrm{pd}} = \Delta\boldsymbol{\lambda}$, $\Delta\mathbf{x}_{\mathrm{pd}} = \Delta\mathbf{x}$:

$$-\mathrm{Diag}(\boldsymbol{\lambda}) D\boldsymbol{f}(\mathbf{x}) \Delta\mathbf{x}_{\mathrm{pd}} - \mathrm{Diag}(\boldsymbol{f}(\mathbf{x})) \Delta\boldsymbol{\lambda}_{\mathrm{pd}} = -\mathbf{r}_{\mathrm{cent}}$$

$$\Leftrightarrow \Delta\boldsymbol{\lambda}_{\mathrm{pd}} = -\mathrm{Diag}(\boldsymbol{f}(\mathbf{x}))^{-1} \mathrm{Diag}(\boldsymbol{\lambda}) D\boldsymbol{f}(\mathbf{x}) \Delta\mathbf{x}_{\mathrm{pd}} + \mathrm{Diag}(\boldsymbol{f}(\mathbf{x}))^{-1} \mathbf{r}_{\mathrm{cent}}.$$

Primal-dual Search Direction

$$\Delta\boldsymbol{\lambda}_{\mathrm{pd}} = -\mathrm{Diag}\left(\boldsymbol{f}(\mathbf{x})\right)^{-1}\mathrm{Diag}(\boldsymbol{\lambda})D\boldsymbol{f}(\mathbf{x})\Delta\mathbf{x}_{\mathrm{pd}} + \mathrm{Diag}\left(\boldsymbol{f}(\mathbf{x})\right)^{-1}\mathbf{r}_{\mathrm{cent}}.$$

Substituting this into the first block of equations

$$\left[\nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(\mathbf{x})\right]\Delta\mathbf{x}_{\mathrm{pd}} + D\boldsymbol{f}(\mathbf{x})^T\Delta\boldsymbol{\lambda}_{\mathrm{pd}} + \mathbf{A}^T\Delta\mathbf{v}_{\mathrm{pd}} = -\mathbf{r}_{\mathrm{dual}}$$

and using the fact that $D\boldsymbol{f}(\mathbf{x})^T = \begin{bmatrix} \dfrac{\partial \boldsymbol{f}(\mathbf{x})}{\partial x_1} & \cdots & \dfrac{\partial \boldsymbol{f}(\mathbf{x})}{\partial x_m} \end{bmatrix} = \begin{bmatrix} \nabla f_1(\mathbf{x})^T \\ \vdots \\ \nabla f_m(\mathbf{x})^T \end{bmatrix}$ gives

$$\Leftrightarrow \left[\nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(\mathbf{x})\right]\Delta\mathbf{x}_{\mathrm{pd}} - D\boldsymbol{f}(\mathbf{x})^T\mathrm{Diag}\left(\boldsymbol{f}(\mathbf{x})\right)^{-1}\mathrm{Diag}(\boldsymbol{\lambda})D\boldsymbol{f}(\mathbf{x})\Delta\mathbf{x}_{\mathrm{pd}} + \mathbf{A}^T\Delta\mathbf{v}_{\mathrm{pd}}$$

$$= -\left(\mathbf{r}_{\mathrm{dual}} + D\boldsymbol{f}(\mathbf{x})^T\mathrm{Diag}\left(\boldsymbol{f}(\mathbf{x})\right)^{-1}\mathbf{r}_{\mathrm{cent}}\right)$$

$$= \left[\nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(\mathbf{x})\right]\Delta\mathbf{x}_{\mathrm{pd}} + \left[\sum_{i=1}^m \frac{\lambda_i}{-f_i(\mathbf{x})}\nabla f_i(\mathbf{x})\nabla f_i(\mathbf{x})^T\right]\Delta\mathbf{x}_{\mathrm{pd}} + \mathbf{A}^T\Delta\mathbf{v}_{\mathrm{pd}}$$

Primal-dual Search Direction

which in matrix form becomes

$$\begin{bmatrix} \mathbf{H}_{\text{pd}} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{pd}} \\ \Delta \mathbf{v}_{\text{pd}} \end{bmatrix} = -\begin{bmatrix} \mathbf{r}_{\text{dual}} + D\boldsymbol{f}\left(\mathbf{x}\right)^T \text{Diag}\left(\boldsymbol{f}\left(\mathbf{x}\right)\right)^{-1} \mathbf{r}_{\text{cent}} \\ \mathbf{r}_{\text{pri}} \end{bmatrix},$$

where $\mathbf{H}_{\text{pd}} = \nabla^2 f_0\left(\mathbf{x}\right) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i\left(\mathbf{x}\right) + \sum_{i=1}^{m} \frac{\lambda_i}{-f_i\left(\mathbf{x}\right)} \nabla f_i\left(\mathbf{x}\right) \nabla f_i\left(\mathbf{x}\right)^T$

Since $\mathbf{r}_{\text{cent}} = -\text{Diag}\left(\boldsymbol{\lambda}\right) \boldsymbol{f}\left(\mathbf{x}\right) - (1/t)\mathbf{1}$ and $\mathbf{r}_{\text{dual}} = \nabla f_0\left(\mathbf{x}\right) + D\boldsymbol{f}\left(\mathbf{x}\right)^T \boldsymbol{\lambda} + \mathbf{A}^T \mathbf{v}$, then it becomes

$$\begin{bmatrix} \mathbf{H}_{\text{pd}} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{pd}} \\ \Delta \mathbf{v}_{\text{pd}} \end{bmatrix} = -\begin{bmatrix} \nabla f_0\left(\mathbf{x}\right) + \left(1/t\right) \sum_{i=1}^{m} \frac{1}{-f_i\left(\mathbf{x}\right)} \nabla f_i\left(\mathbf{x}\right) + \mathbf{A}^T \mathbf{v} \\ \mathbf{r}_{\text{pri}} \end{bmatrix} \qquad (11.55)$$

Primal-dual Search Direction

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

Recall that the barrier method: the Newton step $\Delta\mathbf{x}_{\text{nt}}$, and associated dual variable are given by

$$\begin{bmatrix} t\nabla^2 f_0(\mathbf{x}) + \nabla^2 \phi(\mathbf{x}) & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_{\text{nt}} \\ \boldsymbol{\nu}_{\text{nt}} \end{bmatrix} = -\begin{bmatrix} t\nabla f_0(\mathbf{x}) + \nabla \phi(\mathbf{x}) \\ \mathbf{0} \end{bmatrix} \qquad (11.14)$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{H}_{\text{bar}} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_{\text{bar}} \\ \boldsymbol{\nu}_{\text{bar}} \end{bmatrix} = -\begin{bmatrix} t\nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}) \\ \mathbf{r}_{\text{pri}} \;\; (= \mathbf{0} \text{ when } \mathbf{x} \text{ feasible}) \end{bmatrix}, \qquad (11.57)$$

where $\mathbf{H}_{\text{bar}} = t\nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{f_i(\mathbf{x})^2} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^T + \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x}) \qquad (11.58)$

(11.57) is the general form for infeasible Newton step. If current $\mathbf{x}$ is feasible, i.e. $\mathbf{r}_{\text{pri}} = \mathbf{0}$, then $\Delta\mathbf{x}_{\text{bar}} = \Delta\mathbf{x}_{\text{nt}}$, i.e. feasible Newton step in (11.14)

Primal-dual Search Direction

$$\begin{bmatrix} \mathbf{H}_{\text{pd}} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{pd}} \\ \Delta \mathbf{v}_{\text{pd}} \end{bmatrix} = -\begin{bmatrix} \nabla f_0(\mathbf{x}) + (1/t) \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}) + \mathbf{A}^T \mathbf{v} \\ \mathbf{r}_{\text{pri}} \end{bmatrix} \quad (11.55)$$

$$\begin{bmatrix} \mathbf{H}_{\text{bar}} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{bar}} \\ \mathbf{v}_{\text{bar}} \end{bmatrix} = -\begin{bmatrix} t\nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}) \\ \mathbf{r}_{\text{pri}} \end{bmatrix} \quad (11.57)$$

- Coefficient matrix in (11.55) and (11.57) have similar structure, $\mathbf{H}_{\text{pd}}$ and $\mathbf{H}_{\text{bar}}$ are both positive linear combinations of the matrices

$$\nabla^2 f_0(\mathbf{x}), \quad \nabla^2 f_1(\mathbf{x}), \ldots, \nabla^2 f_m(\mathbf{x}), \quad \nabla f_1(\mathbf{x}) \nabla f_1(\mathbf{x})^T, \ldots, \nabla f_m(\mathbf{x}) \nabla f_m(\mathbf{x})^T,$$

implying that the same method can be used to compute the primal-dual search directions and the barrier method Newton step

Primal-dual Search Direction

$$\begin{bmatrix} \mathbf{H}_{\text{bar}} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{bar}} \\ \mathbf{v}_{\text{bar}} \end{bmatrix} = -\begin{bmatrix} t\nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}) \\ \mathbf{r}_{\text{pri}} \end{bmatrix} \quad (11.57)$$

- Suppose we divide the first block of (11.57) by t, and define the variable $\Delta \mathbf{v}_{\text{bar}} \triangleq (1/t)\mathbf{v}_{\text{bar}} - \mathbf{v}$ (where $\mathbf{v}$ is arbitrary), then we have

$$\begin{bmatrix} (1/t)\mathbf{H}_{\text{bar}} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{bar}} \\ \Delta \mathbf{v}_{\text{bar}} \end{bmatrix} = -\begin{bmatrix} \nabla f_0(\mathbf{x}) + (1/t)\sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}) \\ \mathbf{r}_{\text{pri}} \end{bmatrix}$$

  then righthand side is identical to the righthand side of the primal-dual equations (evaluated at the same $\mathbf{x}$, $\boldsymbol{\lambda}$, $\mathbf{v}$).
- The coefficient matrices differ only in the 1,1 block

$$\mathbf{H}_{\text{pd}} = \nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(\mathbf{x}) + \sum_{i=1}^{m} \frac{\lambda_i}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^T$$

$$(1/t)\mathbf{H}_{\text{bar}} = \nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{-tf_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x}) + \sum_{i=1}^{m} \frac{1}{tf_i(\mathbf{x})^2} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^T.$$

  When $\mathbf{x}$ and $\boldsymbol{\lambda}$ satisfy $-f_i(\mathbf{x}) = 1/t$, the coefficient matrices, and therefore also the search directions, coincide.

Surrogate Duality Gap

In the primal-dual IPM, the iterates $\mathbf{x}^{(k)}, \boldsymbol{\lambda}^{(k)}, \mathbf{v}^{(k)}$ 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 $\mathbf{x}$ that satisfies $f(\mathbf{x}) \prec \mathbf{0}$ and $\boldsymbol{\lambda} \succeq \mathbf{0}$, as

$$\hat{\eta}(\mathbf{x}, \boldsymbol{\lambda}) = -f(\mathbf{x})^T \boldsymbol{\lambda}. \tag{11.59}$$

- The surrogate gap $\hat{\eta}$ would the duality gap, if $\mathbf{x}$ were primal feasible and $\boldsymbol{\lambda}, \mathbf{v}$ were dual feasible, i.e. if $\mathbf{r}_{\text{pri}} = \mathbf{0}$ and $\mathbf{r}_{\text{dual}} = \mathbf{0}$.
- 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

Algorithm 11.2 *Primal-dual interior-point method.*

given x that satisfies $f_1(x) < 0, \ldots, f_m(x) < 0$, $\lambda \succ 0$, $\mu > 1$, $\epsilon_{\text{feas}} > 0$, $\epsilon > 0$.

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 $y := y + s\Delta y_{\text{pd}}$.

until $\|r_{\text{pri}}\|_2 \leq \epsilon_{\text{feas}}$, $\|r_{\text{dual}}\|_2 \leq \epsilon_{\text{feas}}$, and $\hat{\eta} \leq \epsilon$.

Primal-dual Interior-Point Method

- 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 $\hat{\eta}$.
- If $\mathbf{x}$, $\boldsymbol{\lambda}$, $\mathbf{v}$ 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 μ on the order of 10 appear to work well
- The primal-dual IPM terminates when $\mathbf{x}$ is primal feasible and $\boldsymbol{\lambda}, \mathbf{v}$ are dual feasible (within tolerance ϵ_{feas}) and $\hat{\eta}$ is smaller than the tolerance ϵ.
- Since the primal-dual IPM often has faster than linear convergence, it is common to choose ϵ_{feas} and ϵ small.

Primal-dual Interior-Point Method

- **Line search**: The line search in the primal-dual method is a standard backtracking line search, based on the norm of the residual and modified to ensure that $\boldsymbol{\lambda} \succ \mathbf{0}$ and $f(\mathbf{x}) \prec \mathbf{0}$. We denote the current iterate as $\mathbf{x}$, $\boldsymbol{\lambda}$, $\mathbf{v}$, and the next iterate as $\mathbf{x}^+$, $\boldsymbol{\lambda}^+$, $\mathbf{v}^+$, i.e.,

$$\mathbf{x}^+ = \mathbf{x} + s\Delta\mathbf{x}_{\mathrm{pd}}, \qquad \boldsymbol{\lambda}^+ = \boldsymbol{\lambda} + s\Delta\boldsymbol{\lambda}_{\mathrm{pd}}, \qquad \mathbf{v}^+ = \mathbf{v} + s\Delta\mathbf{v}_{\mathrm{pd}}.$$

  The residual, evaluated at $\mathbf{y}^+$, will be denoted as $\mathbf{r}^+$

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

$$s^{\max} = \sup\left\{s \in [0,1] \,\middle|\, \boldsymbol{\lambda} + s\Delta\boldsymbol{\lambda} \succeq \mathbf{0}\right\} = \min\left\{1,\ \min\left\{-\frac{\lambda_i}{\Delta\lambda_i} \,\middle|\, \Delta\lambda_i < 0\right\}\right\}.$$

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

$$\left\|\mathbf{r}_t\left(\mathbf{x}^+, \boldsymbol{\lambda}^+, \mathbf{v}^+\right)\right\|_2 = (1-\alpha s)\left\|\mathbf{r}_t\left(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}\right)\right\|_2.$$

- Values for α and β are same as those for Newton's method: $\alpha \in [0.01, 0.1]$, $\beta \in [0.3, 0.8]$.