Applied Optimization: Formulation and Algorithms for Engineering Systems Slides

Ross Baldick
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX 78712

Part V

Inequality-constrained optimization

15

Case studies of inequality-constrained optimization

(i) Production, at least-cost, of a commodity from machines that have minimum and maximum machine capacity constraints (Section 15.1),

(ii) Optimal routing in a data communications network (Section 15.2),

(iii) Least absolute value estimation (Section 15.3),

(iv) Optimal margin pattern classification (Section 15.4),

(v) Choosing the widths of interconnects between latches and gates in integrated circuits (Section 15.5), and

(vi) The optimal power flow problem in electric power systems (Section 15.6).

15.1 Least-cost production with capacity constraints

15.1.1 Motivation

- Recall the least-cost production case study discussed in Section 12.1.
- For that problem we ignored the minimum and maximum machine capacity constraints in order to formulate it as equality-constrained Problem (12.4), which we repeat here:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b \}.$$

- In this section, we will consider the case where the solution of Problem (12.4) does not satisfy all the minimum and maximum machine capacity constraints so that these constraints must be considered explicitly.

15.1.2 Formulation

15.1.2.1 Objective

$$\forall x \in \mathbb{R}^n, f(x) = \sum_{k=1}^{n} f_k(x_k).$$

15.1.2.2 Equality constraints

$$D = \sum_{k=1}^{n} x_k.$$

- We represented these constraints in the form $Ax = b$ with $A = -\mathbf{1}^\dagger \in \mathbb{R}^{1\times n}$ and $b = [-D] \in \mathbb{R}^1$.

15.1.2.3 Inequality constraints

$$\forall \ell = 1, \ldots, n, \underline{x}_\ell \leq x_\ell \leq \overline{x}_\ell.$$

- We summarize these constraints by writing $\underline{x} \leq x \leq \overline{x}$,
- where $\underline{x} \in \mathbb{R}^n$ and $\overline{x} \in \mathbb{R}^n$ are constant vectors with k-th entries $\underline{x}_k$ and $\overline{x}_k$, respectively.

15.1.2.4 Problem

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, \underline{x} \leq x \leq \overline{x} \}. \tag{15.1}$$

15.1.3 Changes in demand and capacity

- We may want to estimate the change in the costs due to a change in demand from D to $D + \Delta D$, say.
- If the capacity of a machine k changes or it fails then the corresponding entries $\overline{x}_k$ and $\underline{x}_k$ of $\overline{x}$ and $\underline{x}$ will change.

15.1.4 Problem characteristics

15.1.4.1 Objective

- If $\underline{x}_k > 0$ then, for typical cost functions, f_k is convex on $[\underline{x}_k, \overline{x}_k]$.

15.1.4.2 Equality constraints

- We have already discussed the equality constraint $D = \sum_{k=1}^{n} x_k$ in Section 12.1.2.4.

15.1.4.3 Inequality constraints and the feasible region

- The intersection of the box with the equality constraint restricts the feasible region to being a planar slice through the box.
- This is illustrated in Figure 15.1 for $n = 3$, $D = 10$, and:

$$\underline{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \overline{x} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

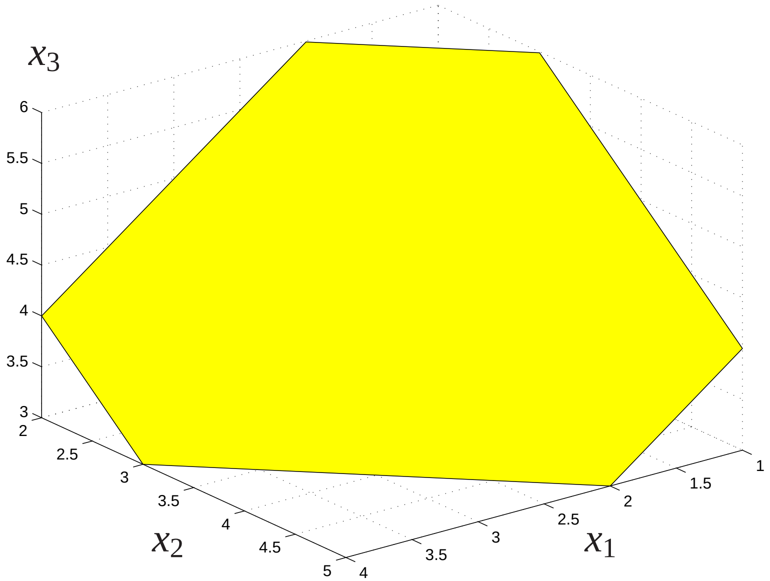

Fig. 15.1. Feasible set for least-cost production case study described in section 15.1.4.3.

15.1.4.4 Solvability

- Problem (15.1) is convex.
- It is certainly possible for there to be no feasible points for Problem (15.1).

15.2 Optimal routing in a data communications network

15.2.1 Motivation

- We consider a communications network consisting of communications **links** that join between **nodes**.
- Users desire to send data from **origin nodes** to **destination nodes** over links between the nodes.
- Each link has a maximum capacity to transmit data and several links may be incident to each node.
- Data is sent by users in **packets** of equal length.

Motivation, continued

- **Inter-arrival time** between packets is random, with **exponential distribution** that may differ from node to node.
- We assume that the probability distributions of the inter-arrival times do not vary over time.
- We can therefore consider the average traffic on each link due to:
  - the distributions of inter-arrival times, and
  - a **routing policy** that is, a decision process for choosing the links on which to send the data.
- We refer to the choice of links, with respect to a given criterion and for given traffic levels between origin–destination pairs, as **optimal routing**.
- We will see that our formulation of the objective only approximately captures the criterion we discuss and so we might better refer to our problem as **satisficing** routing.

15.2.2 Formulation

- We can represent the communications network as a graph.
- Each of the eight nodes in Figure 15.2 is shown as a bullet •, while each of the 12 links is shown as a line.
- As in previous case studies involving graphs, the typical number of links is far less than in a complete graph.

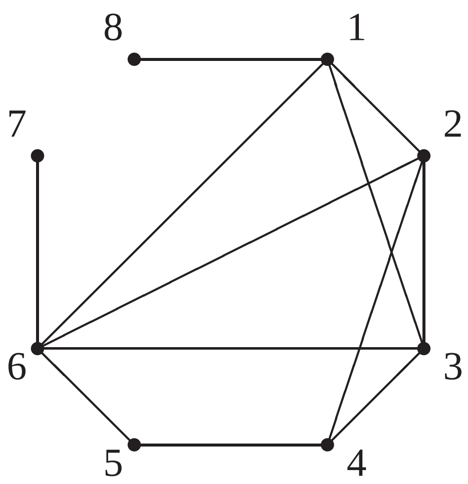

Fig. 15.2. Graphical representation of a data communications network with eight nodes and 12 links.

15.2.2.1 Links

- We write $\mathbb{L}$ for the set of all links in the network, where each link is represented by an ordered pair (i, j) of node numbers.

$$\begin{aligned}\mathbb{L} = \{&(1,8),(8,1),(1,2),(2,1),(1,3),(3,1),(1,6),(6,1),\\&(2,3),(3,2),(2,4),(4,2),(2,6),(6,2),(3,4),(4,3),\\&(3,6),(6,3),(4,5),(5,4),(5,6),(6,5),(6,7),(7,6)\}.\end{aligned}$$

- The capacity of link (i, j) is denoted by $\bar{y}_{ij} \in \mathbb{R}_{++}$.

15.2.2.2 Nodes

- Nodes have three roles, as follows.
  - Users put data into the network at nodes. These nodes can be thought of as the **origins** of data.
  - A node **switches** arriving data onto one of the links incident to it.
  - Users take data out of the network at nodes. These nodes can be thought of as the **destinations** of data.

15.2.2.3 Origin–destination pairs

- A user might put data into the network at node 7 and desire to transmit it to node 5:

  node 7 is the origin for the data and node 5 is the destination for the data.
- We assume that there are m origin–destination pairs and write $\mathbb{W}$ for the set of all origin–destination pairs.
- In our example, if $(7,5)$ and $(2,5)$ are the only origin–destination pairs then:

$$\mathbb{W} = \{(7,5),(2,5)\},$$

  with $m = 2$.
- In general, an origin–destination pair $(\ell,\ell') \in \mathbb{W}$ might or might not be joined directly by a link.
- If there is no link joining such an origin–destination pair then it is necessary for the data between this pair to traverse several successive links.

15.2.2.4 Paths

- A collection of successive links that joins an origin–destination pair is called a **path**.
- Two paths for the origin–destination pair $(7,5)$ are:
  - links $(7,6)$ and $(6,5)$, and
  - links $(7,6),(6,3),(3,4),(4,5)$.
- For each origin–destination pair $(\ell,\ell') \in \mathbb{W}$, we write $\mathbb{P}_{(\ell,\ell')}$ for the set of all allowable paths connecting ℓ to ℓ'.
- We index the paths with consecutive integers.
- For example, for the origin–destination pair $(7,5) \in \mathbb{W}$, we will denote:
  - the path consisting of links $(7,6)$ and $(6,5)$ as path 1, and
  - the path consisting of links $(7,6),(6,3),(3,4),(4,5)$ as path 2.
- For the origin–destination pair $(2,5) \in \mathbb{W}$, we will denote:
  - the path consisting of links $(2,4)$ and $(4,5)$ as path 3, and
  - the path consisting of links $(2,3),(3,4),(4,5)$ as path 4.

Paths, continued

- We summarize these assignments by $\mathbb{P}_{(7,5)} = \{1,2\}, \mathbb{P}_{(2,5)} = \{3,4\}$.
- We assign a different index k for each allowed path in the network and suppose that there are n paths in all.
- In our example, if we have described all the allowable paths then $n = 4$.

15.2.2.5 Variables

- To characterize the behavior of the network, we consider the **expected** or **average** flow of packets and ignore variance of the distribution of flow.
- We define $x_k, k = 1, \ldots, n$. to be the average flow of traffic, in packets per second, on path k.
- This flow represents the average amount of flow for a particular origin–destination pair that has been assigned to path k.
- We collect the set of all traffic assignments for all origin–destination pairs together into a vector $x \in \mathbb{R}^n$.

15.2.2.6 *Equality constraints*

- Let the input traffic arrival process for origin–destination pair $(\ell, \ell') \in \mathbb{W}$ have expected rate of arrival of $b_{(\ell,\ell')}$, in packets per second.
- In general, we must choose how to share the traffic amongst all the paths that join ℓ to ℓ'.

$$\forall (\ell, \ell') \in \mathbb{W}, \sum_{k \in \mathbb{P}_{(\ell,\ell')}} x_k = b_{(\ell,\ell')}.$$

- In our example, the constraints for the origin–destination pairs $(7,5)$ and $(2,5)$ are, respectively:

$$\begin{aligned} x_1 + x_2 &= b_{(7,5)}, \\ x_3 + x_4 &= b_{(2,5)}. \end{aligned}$$

- We collect the entries $b_{(\ell,\ell')}$ for $(\ell, \ell') \in \mathbb{W}$ into a vector $b \in \mathbb{R}^m$.

Equality constraints, continued

- Also, define $A \in \mathbb{R}^{m \times n}$ to be the path to origin–destination pair incidence matrix.
- That is, define:

$$\forall(\ell,\ell') \in \mathbb{W}, \forall k = 1,\ldots,n, A_{(\ell,\ell')k} = \begin{cases} 1, & \text{if } k \in \mathbb{P}_{(\ell,\ell')}, \\ 0, & \text{otherwise.} \end{cases}$$

- In our example:

$$\begin{aligned} A &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \\ b &= \begin{bmatrix} b_{(7,5)} \\ b_{(2,5)} \end{bmatrix}. \end{aligned}$$

- With these definitions, we can write the equality constraints as:

$$Ax = b. \tag{15.2}$$

15.2.2.7 *Objective*

Discussion

- Several criteria could be used to define an objective.
- Unlike the least-cost production case study in Sections 12.1 and 15.1, the operating cost of a data network is generally relatively constant.
- In *delivering service* to customers, however, the quality of service depends on a number of factors, including the delay between sending data and receiving it.

Delay

- The delay on a link depends on how much traffic is on the link.
- When the traffic is nearly as large as the capacity of the link, the delay is longer.
- We say that the link is **congested**.
- It is difficult to obtain an analytic model of the delay in a network because the packets interact as they traverse the links, so that the analysis of their statistics is complicated.

Delay, continued

- For example, consider an origin–destination pair (ℓ, ℓ') that is joined by one path, which consists of two successive links (ℓ, j) and (j, ℓ').
- The inter-arrival time at the origin ℓ is exponentially distributed.
- The inter-arrival time at node j cannot be exponentially distributed.
- The reason is that successive packets arriving at j must be separated in time by at least the packet transmission time for the first link and this violates the assumption of exponential distribution.

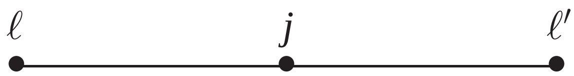

Fig. 15.3. A network with an origin–destination pair joined by a path consisting of two links.

Congestion model

- As a *proxy* to calculating the delay experienced by the packets in the network, we define a measure of the congestion on each link that is a convex function of the expected flow y_{ij} through the link.
- We will sum the congestion measure across all the links as a proxy to the average delay.
- Consider the function $\phi_{ij} : [0, \bar{y}_{ij}) \to \mathbb{R}_+$ defined by:

$$\forall y_{ij} \in [0, \bar{y}_{ij}), \phi_{ij}(y_{ij}) = \frac{y_{ij}}{\bar{y}_{ij} - y_{ij}} + \delta_{ij} y_{ij}, \tag{15.3}$$

- where δ_{ij} is the sum of the processing delay and the propagation delay through the router and link, and
- the term $\frac{y_{ij}}{\bar{y}_{ij} - y_{ij}}$ is due to queuing at the sending end of the link.
- The rapid rise in the congestion function as the expected flow approaches the capacity models the increase in the delay as the capacity is reached:
  - fluctuations about the expected value mean that the queue would become arbitrarily long if the expected flow equaled the capacity.

Flow

- The flow y_{ij} on the link is equal to the sum of the flows on all the paths that include link (i, j).
- We write $\mathbb{F}_{(i,j)}$ for the set of paths that include link (i, j), so that the flow y_{ij} can be expressed as:

$$\forall (i, j) \in \mathbb{L}, y_{ij} = \sum_{k \in \mathbb{F}_{(i,j)}} x_k.$$

- Define a matrix $C \in \mathbb{R}^{\mathbb{L} \times n}$ by:

$$\forall (i, j) \in \mathbb{L}, \forall k = 1, \ldots, n, C_{(i,j)k} = \begin{cases} 1, & \text{if } k \in \mathbb{F}_{(i,j)}, \\ 0, & \text{otherwise.} \end{cases}$$

- For each $(i, j) \in \mathbb{L}$, let $C_{(i,j)}$ be the (i, j)-th row of C.
- Then the flow y_{ij} can be expressed as $\forall (i, j) \in \mathbb{L}, y_{ij} = C_{(i,j)} x$.
- Let $y \in \mathbb{R}^{\mathbb{L}}$ be a vector with entries $y_{ij}, (i, j) \in \mathbb{L}$.
- Then $y = Cx$.

Additive congestion

- We have assumed that the congestion measure for each link can be added together to obtain an overall proxy for average delay through the network.
- Let $\mathbb{P} = \{y \in \mathbb{R}^{\mathbb{L}} | 0 \leq y_{ij} < \bar{y}_{ij}, \forall (i,j) \in \mathbb{L}\}$ and define the objective $\phi : \mathbb{P} \to \mathbb{R}$ by:

$$\forall y \in \mathbb{P}, \phi(y) = \sum_{(i,j)\in\mathbb{L}} \phi_{ij}(y_{ij}). \tag{15.4}$$

- Paths between various origin–destination pairs will typically have some links in common:
  - path 3 consists of the links $(2,4),(4,5)$, and
  - path 4 consists of the links $(2,3),(3,4),(4,5)$,
- and both of these paths are for the origin–destination pair $(2,5)$.
- Traffic on these paths must share the capacity of the link $(4,5)$ with traffic on path 2, which consists of links $(7,6),(6,3),(3,4),(4,5)$ for origin–destination pair $(7,5)$.
- This means that there will be an interaction between traffic between various origin–destination pairs.

Additive congestion, continued

- The objective captures the issue that increasing the flow on a path that is incident to a particular link will increase the average delay for *all* paths incident to that link.
- The objective does not exactly capture the average delay due to the flows on the paths.
- It is a proxy to the average delay that is designed to capture the *qualitative* dependence of average delay on the choice of routing.
- It may be sufficiently accurate to provide guidance to avoid bad routing decisions.

15.2.2.8 Inequality constraints and feasible set

- All traffic flows must be non-negative:

$$x \geq \mathbf{0}.$$

- Since the capacity of each link $(i, j) \in \mathbb{L}$ is $\overline{y}_{ij}$, the instantaneous flow on link (i, j) can never exceed $\overline{y}_{ij}$.
- Consequently, the average flow can never exceed $\overline{y}_{ij}$, suggesting constraints of the form:

$$\forall (i, j) \in \mathbb{L}, y_{ij} \leq \overline{y}_{ij}.$$

- However, as discussed in Section 15.2.2.7, the objective is unbounded if any y_{ij} were to equal $\overline{y}_{ij}$, so we must limit the values of the flows y_{ij} with constraints of the form:

$$\forall (i, j) \in \mathbb{L}, y_{ij} < \overline{y}_{ij}.$$

- We use the strict inequality because if the assigned flow were to equal the capacity then the congestion function would be unbounded.

Inequality constraints and feasible set, continued

- To represent these strict inequality constraints explicitly in terms of x, we note that:

$$\begin{aligned} \forall (i,j) \in \mathbb{L}, y_{ij} &= \sum_{k \in \mathbb{F}_{(i,j)}} x_k, \\ &= C_{(i,j)} x. \end{aligned}$$

- If we define $\overline{y} \in \mathbb{R}^{\mathbb{L}}$ to be a vector with entries $\overline{y}_{ij}, (i,j) \in \mathbb{L}$ then we can write the strict inequality constraints as:

$$Cx < \overline{y}. \tag{15.5}$$

- The inequality constraints for the problem therefore specify a set of the form:

$$\overline{\mathbb{S}} = \{x \in \mathbb{R}^n | x \geq \mathbf{0}, Cx < \overline{y}\}.$$

15.2.2.9 Problem

$$\min_{x \in \mathbb{R}^n} \{ f(x) \,|Ax = b, x \geq \mathbf{0}, Cx < \bar{y} \}, \tag{15.6}$$

- where $f : \overline{\mathbb{S}} \to \mathbb{R}$ is defined by:

$$\begin{aligned} \forall x \in \overline{\mathbb{S}}, f(x) &= \phi(Cx), \\ &= \sum_{(i,j) \in \mathbb{L}} \phi_{ij} \left(C_{(i,j)} x \right). \end{aligned} \tag{15.7}$$

15.2.3 Changes in links and traffic

- We would like to be able to change the routing to respond to changes in link capacity.
- Over time, we also expect that the traffic on the network would change.
- We would also like to be able to change the routing to respond to changes in traffic.

15.2.4 Problem characteristics

15.2.4.1 Objective

- The objective defined in (15.7) is convex and differentiable, since it is the composition of a linear function with the sum of functions ϕ_{ij}, which are themselves convex.
- The objective becomes arbitrarily large as the flow on any link approaches its capacity.

15.2.4.2 Equality constraints

- The equality constraints are indexed by ordered pairs $(\ell, \ell') \in \mathbb{W}$.
- This differs from our previous case studies were index sets were subsets of the integers.
- The equality constraints are affine and the coefficient matrix consists of only zeros and ones.

15.2.4.3 Inequality constraints

- There are non-negativity constraints and also strict inequality constraints due to the link capacities.
- The strict inequality constraints are indexed by the ordered pairs $(i, j) \in \mathbb{L}$.
- We discussed the potential difficulties with strict inequality constraints in Section 2.3.3.
- We will see in Section 18.2 that because of the form of the objective we can avoid explicit consideration of the strict inequality constraints.

15.2.4.4 Solvability

- There may be no feasible solution if there is not enough capacity in the network.

15.3 Least absolute value estimation

15.3.1 Motivation

- Recall the multi-variate linear regression problem introduction in Section 9.1, which was transformed into a least-squares problem in Section 11.1.1.
- The objective $f : \mathbb{R}^n \to \mathbb{R}$ was defined in Section 11.1.1 to be:

$$\forall x \in \mathbb{R}^n, f(x) = \frac{1}{2} \|Ax - b\|_2^2,$$

- where:

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} \in \mathbb{R}^{m \times n}, A_\ell = \begin{bmatrix} \psi(\ell)^\dagger & 1 \end{bmatrix} \in \mathbb{R}^{1 \times n}, \ell = 1, \dots, m,$$

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m, b_\ell = \zeta(\ell), \text{ and}$$

$(\psi(\ell), \zeta(\ell))$ are the ordered pairs of independent and dependent variables for trial ℓ.

Motivation, continued

- In some contexts, we may find the resulting solution is not **robust** to **outliers** in the data.
- That is, the quadratic objective allows data from a single trial to significantly affect the resulting estimate of the affine function that best represents the data
- For example, Figure 15.4 repeats the data from Figure 9.1, except that the data for one of the trials, $(\psi(6), \zeta(6))$, is significantly different, perhaps due to a gross failure of a measurement device.

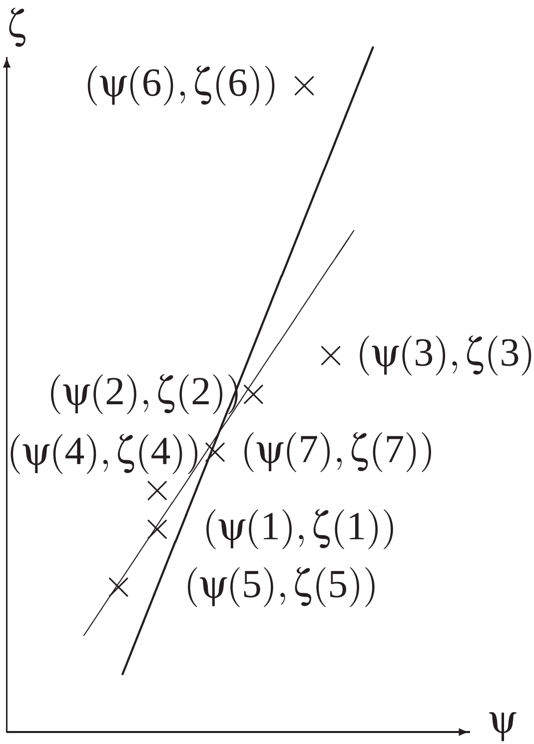

Fig. 15.4. The values of $(\psi(\ell),\zeta(\ell))$, including an outlier, (shown as $\times$) and least-squares fit (shown as a thick line). The thin line shows the least-squares fit if the data point $(\psi(6),\zeta(6))$ is ignored.

Motivation, continued

- The outlier $(\psi(6), \zeta(6))$ significantly affects the result of the least-squares problem.
- The least-squares fit to all of the points in Figure 15.4, including the outlier, is shown by the thick line.
- This least-squares fit is very different to the least-squares fit shown in Figure 9.1.
- If we ignore the point $(\psi(6), \zeta(6))$ then a least-squares fit to the rest of the points is shown as the thin line in Figure 15.4.
- The two least-squares fits are very different.
- That is, the fit is very sensitive to gross errors in individual data points.

Motivation, continued

- In these circumstances, we may prefer to use an objective that is less affected by outliers.
- This provides the motivation for **robust estimation**.
- One objective that is used to reduce the effect of outliers involves the L_1 norm of $Ax - b$ instead of the Euclidean norm.
- Instead of squaring the **residuals** $e_\ell = A_\ell x - b_\ell$, as in the least-squares problem, we take the absolute value of them.
- Outliers, which have large values of residual, will contribute relatively less to the objective when we use the absolute value rather than the square of the residual.

15.3.2 Formulation

15.3.2.1 Unconstrained problem

- Instead of the least-squares objective defined in Section 11.1.1, consider the L_1 norm objective $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by:

$$\begin{aligned} \forall x \in \mathbb{R}^n, \phi(x) &= \|Ax - b\|_1, \\ &= \sum_{\ell=1}^{m} |A_\ell x - b_\ell|, \end{aligned}$$

- where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are as defined in Section 11.1.1 and A_ℓ is the ℓ-th row of A.
- That is:

$$\begin{aligned} A_\ell &= \begin{bmatrix} \psi(\ell)^\dagger & 1 \end{bmatrix}, \\ b &= \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}. \end{aligned}$$

Unconstrained problem, continued

- We define an unconstrained problem:

$$\min_{x \in \mathbb{R}^n} \phi(x). \tag{15.8}$$

- As we saw in Section 3.1.4.4, the objective of this problem is non-differentiable because of the absolute values.

15.3.2.2 *Transformation*

- Problem (15.8) can be transformed into an inequality-constrained problem in several steps.
- As in Section 9.1.2.4, the residual, e_ℓ, for the ℓ-th measurement, is defined by:

$$\forall \ell = 1, \ldots, m, e_\ell = A_\ell x - b_\ell.$$

- Each absolute value of a residual can be obtained as:

$$|e_\ell| = \max\{e_\ell, -e_\ell\}, \ell = 1, \ldots, m. \tag{15.9}$$

- We then use a similar approach to that used in Theorem 3.4 to evaluate the maximum in (15.9).
- First we define variables $z_\ell, \ell = 1, \ldots, m$ and linear constraints:

$$\begin{aligned} z_\ell &\geq e_\ell, \ell = 1, \ldots, m, \\ z_\ell &\geq -e_\ell, \ell = 1, \ldots, m. \end{aligned}$$

- Then note that:

$$
\begin{aligned}
|e_\ell| &= \min_{z_\ell \in \mathbb{R}} \{z_\ell | z_\ell \geq e_\ell, z_\ell \geq -e_\ell\}, \\
\forall x \in \mathbb{R}^n, \phi(x) &= \sum_{\ell=1}^{m} |A_\ell x - b_\ell|, \\
&= \sum_{\ell=1}^{m} |e_\ell|, \text{ where } e_\ell = A_\ell x - b_\ell, \\
&= \sum_{\ell=1}^{m} \min_{z_\ell \in \mathbb{R}} \{z_\ell | z_\ell \geq e_\ell, z_\ell \geq -e_\ell\}.
\end{aligned}
$$

- Combining these observations, we consider the transformed problem:

$$
\min_{z \in \mathbb{R}^m, x \in \mathbb{R}^n, e \in \mathbb{R}^m} \{\mathbf{1}^\dagger z | Ax - b - e = \mathbf{0}, z \geq e, z \geq -e\}. \tag{15.10}
$$

- Problems (15.8) and (15.10) are equivalent.

15.3.3 Changes in the number of points and the data

- We could imagine adding a new trial and recalculating the estimate of the least absolute value fit without starting from scratch.
- We can also imagine modifying the data for a particular trial.

15.3.4 Problem characteristics

15.3.4.1 Objective

- The objective of Problem (15.8) is non-differentiable.
- Transformation into Problem (15.10) by representing each absolute value using two inequality constraints then yields a differentiable, in fact linear, objective.

15.3.4.2 Constraints

- The "cost" of making the objective differentiable is that we have introduced a large number of subsidiary constraints.
- There are m equality constraints and $2m$ inequality constraints in Problem (15.10), whereas Problem (15.8) was unconstrained.

15.3.4.3 Variables

- We have also increased the number of variables, from n to $n+2m$.

15.3.4.4 Solvability

- Problem (15.8) has a minimum and consequently Problem (15.10) also has a minimum.

15.3.4.5 Discussion

- If the number of trials m is extremely large then it may be unattractive to solve Problem (15.10).
- In this case, we may prefer to, for example:
  - solve Problem (15.8) using techniques of non-differentiable optimization,
  - approximate the objective of Problem (15.8) with a smooth function using the approach described in Section 3.1.4.4, or
  - use an iterative technique to successively approximate ϕ by smooth functions.

15.4 Optimal margin pattern classification

15.4.1 Motivation

- We will consider the problem of distinguishing between two classes of patterns on the basis of a linear decision function.
- Geometrically, we seek a hyperplane that separates the two classes of patterns.

15.4.2 Formulation

15.4.2.1 Classes and training set

- Label the two classes as class A and class B.
- We will consider how to find the coefficients that specify a linear decision function in such a way as to provide the best discrimination between classes A and B of patterns.
- In particular, we assume that we have r representatives in our **training set**.
- Potentially, r is very large.
- The training set is to be used to determine the best linear decision function to separate the classes.
- We index the representives in the training set as $\ell = 1, \ldots, r$.
- The ℓ-th representative consists of two items:
  - a **pattern**, namely a vector $\psi(\ell) \in \mathbb{R}^{n-1}$, and
  - a value $\zeta(\ell) \in \{-1, 1\}$.

$$\forall \ell = 1, \ldots, r, \zeta(\ell) = \begin{cases} 1, & \text{if } \psi(\ell) \text{ is of class A,} \\ -1, & \text{if } \psi(\ell) \text{ is of class B.} \end{cases}$$

Classes and training set, continued

- Patterns $\psi(1),\ldots,\psi(4)$ in the bottom half of Figure 15.5 are of class A, while the patterns $\psi(5),\ldots,\psi(7)$ in the top half of the figure are of class B.
- That is, $\zeta(1)=\zeta(2)=\zeta(3)=\zeta(4)=1$ and $\zeta(5)=\zeta(6)=\zeta(7)=-1$.

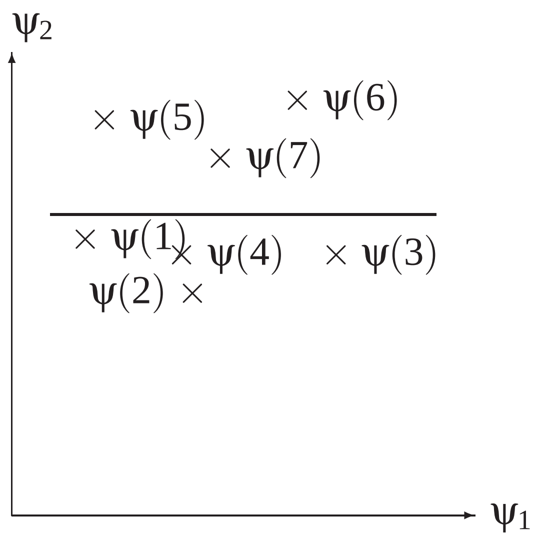

Fig. 15.5. Seven example patterns and hyperplane that separates them.

Classes and training set, continued

- The horizontal line in Figure 15.5 **perfectly discriminates** between classes A and B.
- The vectors representing each pattern may have a very large number of entries.
- That is, $n-1$ may be very large.

15.4.2.2 Feature space

- In a variation on this formulation, the patterns $\psi(\ell)$ are *transformed* versions of the ℓ-th original image.
- For the purposes of our discussion, it does not matter whether we think of the patterns as being "raw" images or transformed images in the feature space.

15.4.2.3 Decision function

- We consider an affine **decision function** $D : \mathbb{R}^{n-1} \to \mathbb{R}$ defined by:

$$\forall \psi \in \mathbb{R}^{n-1}, D(\psi) = \beta^\dagger \psi + \gamma,$$

- where the parameters $\beta \in \mathbb{R}^{n-1}$ and $\gamma \in \mathbb{R}$ are to be chosen so that:

$$\forall \ell = 1, \ldots, r, (D(\psi(\ell)) > 0) \Leftrightarrow (\zeta(\ell) = 1). \tag{15.11}$$

- There are many choices of parameters β and γ that will satisfy (15.11).
- Figure 15.5 shows a line, which is a hyperplane in $\mathbb{R}^{n-1} = \mathbb{R}^2$, of the form:

$$\{\psi \in \mathbb{R}^{n-1} | D(\psi) = 0\},$$

- that divides $\mathbb{R}^{n-1}$ into two half-spaces, one of which contains all the patterns in class A and the other one of which contains all the patterns in class B.

Decision function, continued

- The parameters β and γ are calculated using the training set and the function is then used to estimate the classes of new, unknown patterns for which we do not know the class.
- We must select a suitable criterion for choosing from amongst the values of β and γ that satisfy (15.11).
- If we know the functional form of the probability distribution of the patterns then we could estimate the parameters β and γ using a maximum likelihood criterion, as discussed in the multi-variate linear regression case study in Section 9.1.
- Unfortunately, we usually do not have a lot of information about the patterns that we must subsequently classify and do not know the functional form of the probability distribution from which they are drawn.
- Consequently, the criterion for choosing the parameters β and γ will be *ad hoc*, aimed at finding a satisficing solution.

Decision function, continued

- We will seek β and γ such that the corresponding hyperplane $\{\psi \in \mathbb{R}^{n-1} | D(\psi) = 0\}$ is as far as possible from all the patterns in the training set.
- That is, we will find the values of β and γ that:
  - maximize the minimum distance of any pattern from the hyperplane, and
  - allow classification of the two classes of patterns according to (15.11).
- We will use the notion of Euclidean distance to define distance.
- That is, we will use the norm $\|\bullet\|_2$.
- We are trying to find the hyperplane between the two classes that is at the middle of the thickest slab that separates the two sets of points.

15.4.2.4 Variables

- The decision vector for this problem consists of $\beta \in \mathbb{R}^{n-1}$ and $\gamma \in \mathbb{R}$.
- We collect these together into a vector $x = \begin{bmatrix} \beta \\ \gamma \end{bmatrix} \in \mathbb{R}^n$.
- That is, the parameters that specify the decision function D are the variables for the problem.

15.4.2.5 Objective

- We must evaluate the Euclidean distance of a pattern $\psi(\ell)$ from the closest point on the hyperplane:

$$\{\psi \in \mathbb{R}^{n-1} | D(\psi) = 0\}.$$

- This distance is given by:

$$\frac{|D(\psi(\ell))|}{\|\beta\|_2},$$

- assuming that $\beta \neq \mathbf{0}$.
- Define the set $\mathbb{P} \subset \mathbb{R}^n$ by:

$$\mathbb{P} = \left\{ \begin{bmatrix} \beta \\ \gamma \end{bmatrix} \in \mathbb{R}^n \middle| \beta \neq \mathbf{0} \right\}.$$

Objective, continued

- If the decision function D satisfies (15.11) then for each pattern $\psi(\ell)$ and classification $\zeta(\ell)$:

$$\zeta(\ell)D(\psi(\ell)) = |D(\psi(\ell))|.$$

- If $\beta \neq \mathbf{0}$ and γ satisfy (15.11) then the distance of $\psi(\ell)$ from the hyperplane is given by the function $\phi_\ell : \mathbb{P} \to \mathbb{R}$ defined by:

$$\begin{aligned} \forall x \in \mathbb{P}, \phi_\ell(x) &= \frac{|D(\psi(\ell))|}{\|\beta\|_2}, \\ &= \frac{\zeta(\ell)D(\psi(\ell))}{\|\beta\|_2}. \end{aligned}$$

- The minimum distance of any pattern $\psi(\ell)$ to the hyperplane, over all the patterns ℓ, is given by $\phi : \mathbb{R}^n \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{P}, \phi(x) = \min_{\ell=1,\ldots,r} \phi_\ell(x).$$

- We call this minimum distance the **margin** between the hyperplane and the patterns.

15.4.2.6 Constraint

- In order for the objective to be well-defined, we must restrict ourselves to choices of $x \in \mathbb{P}$; that is, we must require $\beta \neq \mathbf{0}$.
- This constraint is not in our standard form of either an equality or an inequality constraint.

15.4.2.7 Problem

- We seek the coefficients $\beta \neq \mathbf{0}$ and γ such that the margin is maximized.
- Our problem is therefore:

$$\max_{x \in \mathbb{R}^n} \{\phi(x) | \beta \neq \mathbf{0}\}. \tag{15.12}$$

- In the next section we will transform this problem to remove the minimization embedded in the definition of the objective.

15.4.2.8 Transformation

- By Theorem 3.4, we can remove the minimization in the definition of the objective ϕ by defining a subsidiary variable z:

$$
\begin{aligned}
&\max_{x\in\mathbb{R}^n}\{\phi(x)|\beta\neq\mathbf{0}\}\\
&\quad= \max_{x\in\mathbb{R}^n}\left\{\min_{\ell=1,\ldots,r}\phi_\ell(x)\,\middle|\,\beta\neq\mathbf{0}\right\},\\
&\quad= \max_{z\in\mathbb{R},x\in\mathbb{R}^n}\{z|\phi_\ell(x)\geq z,\forall\ell=1,\ldots,r,\beta\neq\mathbf{0}\},\ \text{by Theorem 3.4},\\
&\quad= \max_{z\in\mathbb{R},x\in\mathbb{R}^n}\left\{z\,\middle|\,\frac{\zeta(\ell)D(\psi(\ell))}{\|\beta\|_2}\geq z,\forall\ell=1,\ldots,r,\beta\neq\mathbf{0}\right\},\\
&\quad= \max_{z\in\mathbb{R},x\in\mathbb{R}^n}\left\{z\,\middle|\,\zeta(\ell)(\beta^\dagger\psi(\ell)+\gamma)\geq\|\beta\|_2 z,\forall\ell=1,\ldots,r,\beta\neq\mathbf{0}\right\}.
\end{aligned}
\tag{15.13}
$$

- If the maximum $z^\star$ of Problem (15.13) is strictly positive then the optimal margin is equal to $z^\star$ and is strictly positive.

15.4.3 Changes

- We could consider a change in the problem due to the addition of an extra pattern.

15.4.4 Problem characteristics

15.4.4.1 Objective

- The objective z of Problem (15.13) is linear.

15.4.4.2 Constraints

- The inequality constraints in Problem (15.13) are non-linear.
- Each binding inequality constraint at a solution to the problem corresponds to a pattern that is closest to the hyperplane.
- These are called the **supporting patterns** or **support vectors**.
- The constraint $\beta \neq \mathbf{0}$ in Problem (15.13) is not in the form of equality or inequality constraints.
- The feasible set of Problem (15.13):

$$\mathbb{S} = \left\{ \begin{bmatrix} z \\ x \end{bmatrix} \in \mathbb{R}^{n+1} \middle| \zeta(\ell)(\beta^{\dagger}\psi(\ell) + \gamma) \geq \|\beta\|_2 z, \forall \ell = 1, \ldots, r, \beta \neq \mathbf{0} \right\},$$

- is not closed and may not be convex.
- Feasible sets that are not closed can potentially present difficulties.
- We will consider further transformation of Problem (15.13) in Sections 18.4 and 20.1.

15.4.4.3 *Solvability*

- If there is no hyperplane that can separate the patterns then Problem (15.13) has a maximum that is zero or strictly negative.
- Algorithms for solving this problem are called **support vector machines**.

15.5 Sizing of interconnects in integrated circuits

15.5.1 Motivation

15.5.1.1 Hierarchical design

- The design of digital integrated circuits (ICs) is usually divided into a hierarchy of planning stages.
- For example, a specification of the functionality of the IC is translated into the logic required to meet the specification.
- The integrated components to implement the logic must then be laid out on the “floor-plan” of the chip.
- Once the layout is done, there are still various decisions to be made.
- For example, the widths of the “interconnects” that join one gate to another can be adjusted, within limits, to achieve performance goals.

15.5.1.2 Delay constraints

- One goal is to make sure that the propagation delay on each path from the output of one latch through combinational logic to the input of the next latch is within a limit.
- Adjusting the width of the interconnects affects the delay.
- Increasing the width of the interconnect decreases the resistance and increases the capacitance of an interconnect.
  - Decreasing resistance tends to reduce delay because the current from the driving latch or logic is increased.
  - Increasing capacitance tends to increase delay because the increased capacitance requires more current to charge or discharge.

15.5.1.3 Area of layout

- Another consideration besides delay is that the wider the interconnects, the more area may be required for the circuit.
- We will try to minimize chip area by adjusting the widths of the interconnects, while satisfying the delay constraints.

15.5.1.4 Other issues

- There are many other goals, such as minimizing power dissipation, and other constraints, such as guaranteeing noise immunity, that must be considered.
- In seeking a compromise between various goals, we are again seeking a satisficing solution.

15.5.1.5 Interaction between design levels

- At each level of the hierarchy, we take as fixed the decisions made at higher levels and seek to optimize the remaining decisions.

15.5.2 Formulation

15.5.2.1 Variables

Interconnect widths and lengths

- Latch a drives gate b through a piece of interconnect, labeled 1.
- Gate b drives a branching interconnect, labeled 2, 3, 4, 5, and 6, which in turn drives two more gates, labeled c and d.
- These gates drive the interconnect labeled 7 and 8, which in turn drive latches e and f.

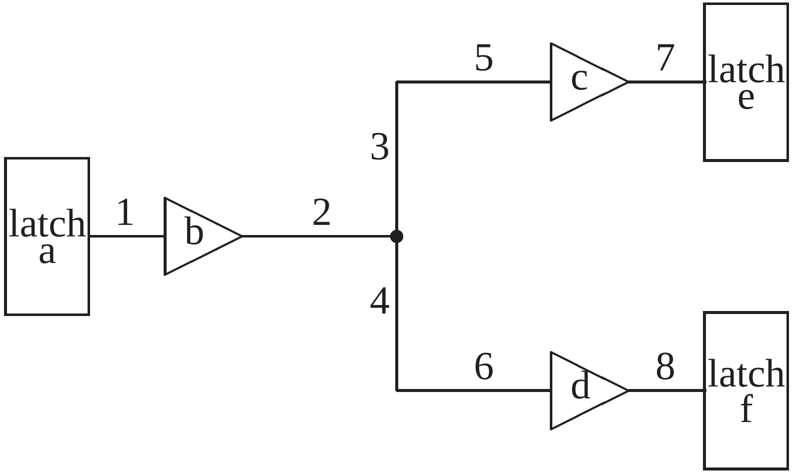

Fig. 15.6. Schematic diagram of gates and latches joined by interconnect.

Segments

- The interconnect can be thought of as consisting of **segments**, corresponding to the labeled pieces of interconnect shown in Figure 15.6.
- We assume that the interconnect can be partitioned into a set of n segments
- Let the k-th segment have width x_k, thickness T_k, and length L_k, as illustrated in Figure 15.7.

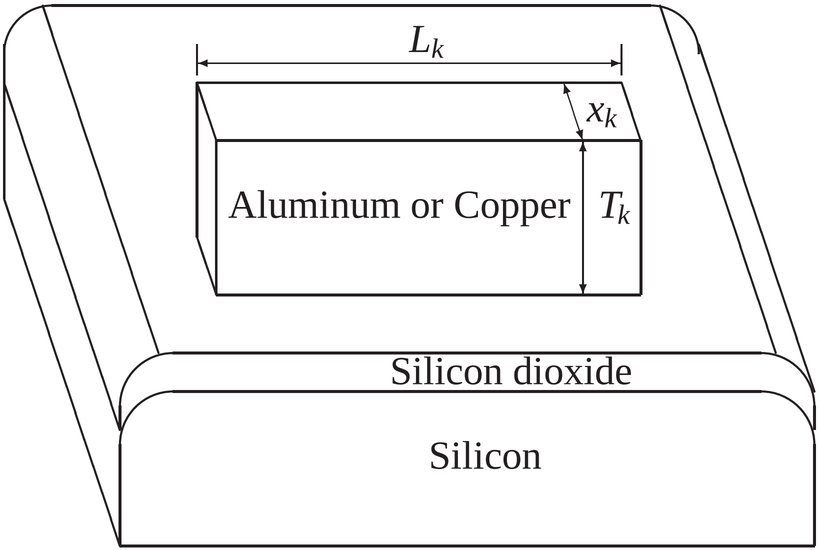

Fig. 15.7. Dimensions of k-th segment of interconnect. The figure is not to scale.

Discreteness

- Because we can only dimension features to be an integer multiple of the minimum feature size, x_k can only be chosen from a *discrete* set of alternatives.
- In general, optimizing over a discrete set of alternatives is much more difficult than optimizing over a continuous variable because in the discrete case we:
  - cannot use calculus to derive optimality conditions,
  - cannot obtain descent directions from purely local first derivative information, and
  - cannot make use of convexity to establish global optimality.
- In this case study, we will neglect discreteness and assume that the widths are continuously variable.

Alternative formulations

- As an alternative formulation, instead of optimizing over a continuous range of widths x_k for segment k, we consider a finite collection of possible widths, say $\{W_{k1},\ldots,W_{ks}\}$ for segment k.
- For example, these widths might correspond to the allowable integer multiples of the minimum feature size.
- A segment is then specified by a collection of sub-lengths L_{kj}, $j = 1,\ldots,s$ such that $\sum_{j=1}^{s} L_{kj} = L_k$. The value L_{kj} specifies how much of the total length of segment k is of width W_{kj}.
- This is an example of a radical transformation of a problem compared to its "natural" formulation.
- We will not pursue this formulation further.

15.5.2.2 Objective

- We have indicated that our goal is to minimize the area of interconnect.
- The area $f : \mathbb{R}^n \to \mathbb{R}$ is defined by:

$$\forall x \in \mathbb{R}^n, f(x) = \sum_{k=1}^{n} L_k x_k,$$

- where L_k is the length of the k-th segment.

15.5.2.3 Constraints

Upper and lower bounds

$$\forall k = 1, \ldots, n, \underline{x}_k \leq x_k \leq \overline{x}_k,$$

Bottlenecks

$$\sum_{k \in \mathbb{B}} x_k \leq \overline{x}_{\mathbb{B}}, \tag{15.14}$$

- where $\mathbb{B}$ is the set of segments involved in a particular bottleneck and $\overline{x}_{\mathbb{B}}$ is the maximum total width available for the segments in the set $\mathbb{B}$.

Delay constraints

- Consider a **path** from a latch through the combinational logic to the input of the next latch.
- We assume that the paths are labeled $\ell = 1, \ldots, r$.
- Our performance specification requires that, for each latch-to-latch path ℓ, a signal can propagate from:
  - the output of the latch at the beginning of path ℓ,
  - through the gates in path ℓ,
  - to the input of latch at the end of path ℓ,
- within a maximum allowed time delay that depends on:
  - the clock period,
  - the delay from the **clock edge** to when the outputs of latches become valid, and
  - the **set-up time** from the input of latches to the clock edge.
- Latch-to-latch delay on each path will depend on the widths of the segments.

Delay constraints, continued

- Therefore, the delay on the ℓ-th path is a function $h_\ell : \mathbb{R}^n \to \mathbb{R}$ depending on the widths and we require that:

$$\forall \ell = 1, \ldots, r, h_\ell(x) \leq \overline{h}_\ell, \tag{15.15}$$

- where $\overline{h}_\ell$ is the maximum allowed latch-to-latch delay on path ℓ.
- We collect the delay functions for each path together into a vector function $h : \mathbb{R}^n \to \mathbb{R}^r$.
- Similarly, we collect the maximum allowed delays into a vector $\overline{h} \in \mathbb{R}^r$.
- To evaluate the function h_ℓ we must define "delay" more carefully.
- Normatively, delay is the time difference between:

  (i) when the voltage at the output of the latch that is driving path ℓ can be considered to have changed state, and

  (ii) when the voltage at the input of the latch that is driven by path ℓ can be considered to have changed state.

Delay constraints, continued

- In practice, "changing state" is defined on an *ad hoc* basis as when, for example, the voltage waveform has risen to or fallen to within 50%, say, or 90%, say, of its final value.
- The delay is often approximated by a function $\tilde{h}_\ell$ that is easier to calculate.
- We will approximate the gate delays by constants neglecting the effect of the load of the interconnect on the delay through the combinational logic.
- We can then re-interpret h_ℓ as being the delay through the interconnect alone, neglecting the gate delays, and reduce the corresponding delay limit $\overline{h}_\ell$ by the sum of the gate delays on path ℓ.
- That is, we re-define each inequality in (15.15) by reducing the left-hand side and the right-hand side by the sum of the gate delays on path ℓ.
- A typical approximation used for the interconnect delay is the **Elmore delay**, which requires an electrical model of the interconnect.

Interconnect electrical model

- Each segment of the interconnect is a distributed resistive-capacitive transmission line.
- Segment k, for $k = 2, \ldots, 6$, has been represented by a series resistance R_k and shunt capacitance C_k, called an **L-segment**.

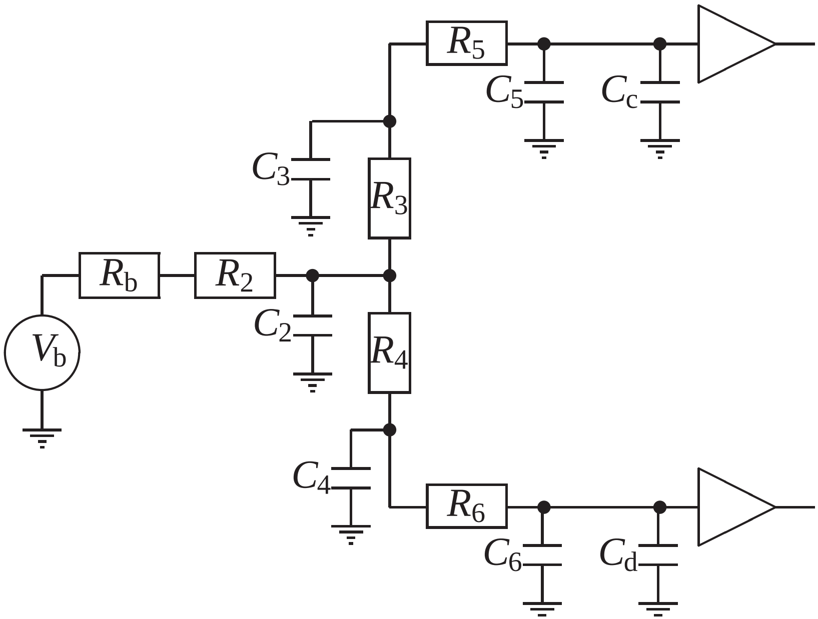

Fig. 15.8. Equivalent circuit of interconnect between gate b and gates c and d consisting of resistive–capacitive L-segments.

Interconnect electrical model, continued

- The resistance of segment k is determined by the **resistivity** ρ_k of the segment and its thickness, length, and width:

$$\begin{aligned} \forall k = 1, \ldots, n, R_k &= \rho_k L_k / (T_k x_k), \\ &= \kappa_{Rk} / x_k, \end{aligned} \tag{15.16}$$

- where $\kappa_{Rk} = \rho_k L_k / T_k$ is a parameter.
- The capacitance of segment k is determined approximately by the **sheet capacitance** per unit area κ_{Sk}, its **fringing capacitance** per unit length κ_{Fk}, and its height and width:

$$\begin{aligned} \forall k = 1, \ldots, n, C_k &= \kappa_{Sk} L_k x_k + \kappa_{Fk} L_k, \\ &= \kappa_{Ck} x_k + C_{Fk}, \end{aligned} \tag{15.17}$$

- where $\kappa_{Ck} = \kappa_{Sk} L_k$ and $C_{Fk} = \kappa_{Fk} L_k$ are parameters.

Gate model

- We can model the gate driving the interconnect by considering its output transistor.
- It can be approximately represented by a voltage source driving a resistance.
- The driving gate b is modeled in Figure 15.8 as the voltage source V_{b} and the driver resistance R_{b}.
- The load presented by **complementary metal-oxide semiconductor** (CMOS) gates at the sinks can be modeled by a capacitance.
- This is shown by C_{c} and C_{d} in Figure 15.8 for the inputs to gates c and d, respectively.

Elmore delay

- Consider a constant voltage source charging a capacitor C through a resistance R.
- The voltage across the capacitor will exponentially approach the driving voltage.
- The time-constant of the exponential is RC, so that a reasonable order-of-magnitude estimate for the rise time of the voltage across the capacitor is RC.
- The "Elmore delay" is an estimate of the time constant of a single exponential that approximates the true response.
- We use this time constant as an estimate of the delay; however, under certain conditions it can be a poor estimate of the delay.

Elmore delay, continued

- Given the lumped L-segment models, the Elmore delay is given by:

$$\forall \ell = 1, \ldots, r, \forall x \in \mathbb{R}^n, \tilde{h}_\ell(x) = \sum_{\mathbb{J} \in \mathbb{P}_\ell} \sum_{j \in \mathbb{J}} \left[R_j \sum_{k \in \mathbb{D}(j)} C_k \right],$$

- where:
  - $\mathbb{P}_\ell$ is the set of sets of **connected segments** on path ℓ. Two segments are connected if there is a path of segments between them. In a set of connected segments, each pair of segments is connected. For example, for the path ℓ from latch a to latch e in Figure 15.6, $\mathbb{P}_\ell = \{\{1\}, \{2,3,5\}, \{7\}\}$, since the path from latch a to latch e consists of three sets of connected segments, namely $\{1\}, \{2,3,5\}, \{7\}$. The connected segments are separated by the latches b and c on the path from latch a to latch e.
  - $\mathbb{D}(j)$ is the set of **downstream segments** including and between segment j and all sinks that are driven from segment j through connected segments. For example, in Figure 15.6, for $j = 2$, $\mathbb{D}(2) = \{2,3,4,5,6\}$. For $j = 3$, $\mathbb{D}(3) = \{3,5\}$.

Elmore delay, continued

- The Elmore delay is the sum of the resistive-capacitive time-constants of each segment, where:
  - the resistive-capacitive time-constant of a segment is equal to the product of the resistance of the segment and all the capacitive load on it, and
  - the capacitive load is defined to be the sum of the capacitances of all the downstream segments (including the input capacitance of all downstream gates and latches.)
- Using the lumped resistive-capacitive model (15.16)–(15.17) for each segment, we obtain:

$$\forall \ell = 1, \ldots, r, \forall x \in \mathbb{R}^n, \tilde{h}_\ell(x) = \sum_{\mathbb{J} \in \mathbb{P}_\ell} \sum_{j \in \mathbb{J}} \left[\frac{\kappa_{Rj}}{x_j} \sum_{k \in \mathbb{D}(j)} (\kappa_{Ck} x_k + C_{Fk}) \right]. \tag{15.18}$$

- We can collect the Elmore delay functions for each path together into a vector function $\tilde{h} : \mathbb{R}^n \to \mathbb{R}^r$, which we use to approximate the actual delay function $h : \mathbb{R}^n \to \mathbb{R}^r$.

15.5.2.4 Problem

- The approximate model for minimizing the area subject to the upper and lower constraints in segment widths and subject to the delay constraints can be written as:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | \tilde{h}(x) \leq \overline{h}, \underline{x} \leq x \leq \overline{x} \}. \tag{15.19}$$

- The more accurate delay model is:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | h(x) \leq \overline{h}, \underline{x} \leq x \leq \overline{x} \}. \tag{15.20}$$

15.5.3 Changes

- We could consider changes in parameters such as the sheet or fringe capacitance constants, due to a change in dielectric properties.
- We could also consider the effect of adding an additional gate in a path.

15.5.4 Problem characteristics

15.5.4.1 Objective

- The objective, $f(x)$, of both Problems (15.19)–(15.20) is linear.

15.5.4.2 Constraints

Upper and lower bounds

- The lower and upper bound constraints $\underline{x} \leq x \leq \overline{x}$ define a convex set.

Delay constraints

- We focus on Problem (15.19).
- The Elmore delay function is not convex.
- The constraint functions involve the sum of terms each of which is a positive constant times the product of powers of the entries in the decision vector.
- Such a function is called a **posynomial function**.

15.5.4.3 *Solvability*

- If there is no selection of widths that yield delays satisfying the delay constraints, then there may be no feasible solution.
- We may need to insert a buffer to break a long segment into two shorter pieces.

15.6 Optimal power flow

15.6.1 Motivation

15.6.1.1 Generalization of economic dispatch

- When applied to electric power systems, the problems described in Sections 12.1 and 15.1 are called **economic dispatch problems**.
- The equality constraint (12.3) requires that electric generation equal the demand; however, this does not fully characterize the situation in an electricity network.
- For example, if generators are remote from demand centers then there will be losses incurred in moving power along transmission lines.
- At the least, (12.3) should be modified to account for losses in this case.

15.6.1.2 Constraints on operation

- Transmission lines between generation and demand can also limit the feasible choices of generation.

15.6.1.3 Power flow equations

- To check whether or not the line flow and voltage constraints are satisfied, we must expand the detail of representation of the network by explicitly incorporating Kirchhoff's laws, as described in the electric power system case study in Section 6.2.2.4.

15.6.1.4 Other controllable elements

- Besides real power generations, we can also consider adjusting any controllable elements in the system so as to minimize costs and meet constraints.

15.6.2 Formulation

15.6.2.1 Variables

- In the decision vector, we need to represent:
  - real and reactive power generations at the generators, which we will collect together into the vectors P and Q,
  - any other controllable quantities in the system, such as the settings of **phase-shifting transformers** and capacitors,
  - the voltage magnitudes at every bus in the system, which we collect together into the vector u, and
  - the voltage angles at every bus in the system except for the reference bus, which we collect together into the vector θ. (The voltage angle at the reference bus is constant since, as previously, it represents an arbitrary time reference.)

Variables, continued

- We collect all the variables into the vector:

$$x = \begin{bmatrix} P \\ Q \\ u \\ \theta \end{bmatrix} \in \mathbb{R}^n.$$

- In the power flow case study in Section 6.2, the generations at the generators were fixed parameters, except at the reference bus.
- In this case study, the real and reactive power generations at all generator buses are variables.
- This is similar to the least-cost production case studies of Sections 12.1 and 15.1, where the real power generations were variables.
- This case study generalizes all of these earlier case studies and exemplifies the process of starting with only a few variables and many parameters and gradually re-interpreting the parameters to be variables.

15.6.2.2 Objective

- A typical objective is to minimize the total cost of power generation.
- Let $f : \mathbb{R}^n \to \mathbb{R}$ represent this cost.
- Typically:

  f depends only on the entries of x corresponding to real power generations; however, in some formulations f also depends somewhat on the entries of x corresponding to reactive power generations, and

  f is separable since the decisions at one generator do not usually affect the costs at any other generators.

15.6.2.3 Equality constraints

- We expressed Kirchhoff's laws as equations in the form:

$$\begin{aligned} \forall \ell, p_\ell(x) &= 0, \\ \forall \ell, q_\ell(x) &= 0, \end{aligned}$$

- where $p_\ell : \mathbb{R}^n \to \mathbb{R}$ and $q_\ell : \mathbb{R}^n \to \mathbb{R}$ were defined in (6.12)–(6.13):

$$\begin{aligned} \forall x \in \mathbb{R}^n, p_\ell(x) &= \sum_{k \in \mathbb{J}(\ell) \cup \{\ell\}} u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] - P_\ell, \\ \forall x \in \mathbb{R}^n, q_\ell(x) &= \sum_{k \in \mathbb{J}(\ell) \cup \{\ell\}} u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)] - Q_\ell, \end{aligned}$$

- where $\mathbb{J}(\ell)$ is the set of buses joined by a line to bus ℓ.

Equality constraints, continued

- We collect the equations together into a vector equation similar to the form of (6.14):

$$g(x) = \mathbf{0},$$

- where a typical entry of g is of the form of (6.12) or (6.13), but the decision vector x includes the real and reactive generations as well as the voltage magnitudes and angles.

15.6.2.4 Inequality constraints

- Limits on the entries in x:

$$\underline{x} \leq x \leq \overline{x}.$$

- A voltage magnitude limit at bus ℓ could be $0.95 = \underline{u}_\ell \leq u_\ell \leq \overline{u}_\ell = 1.05$.
- A generator real power limit could be $0.15 = \underline{P}_\ell \leq P_\ell \leq \overline{P}_\ell = 0.7$.
- There are also constraints involving functions of x.
- For example, there are typically angle difference constraints of the form:

$$\forall \ell, \forall k \in \mathbb{J}(\ell), -\pi/4 \leq \theta_\ell - \theta_k \leq \pi/4, \tag{15.21}$$

- and there might be limits on angle differences between buses that are not joined directly by a line.

Inequality constraints, continued

- In addition, transmission line flow constraints can be expressed via the power flow equations in terms of x.
- That is, we will also have functional constraints of the form:

$$\underline{h} \leq h(x) \leq \overline{h}.$$

- A typical constraint might limit the flow on a line that joins bus ℓ to bus k.
- Neglecting shunt elements in the line models, the line flow real and reactive power flow functions $p_{\ell k} : \mathbb{R}^n \to \mathbb{R}$ and $q_{\ell k} : \mathbb{R}^n \to \mathbb{R}$ are defined by:

$$\begin{aligned} \forall x \in \mathbb{R}^n, p_{\ell k}(x) &= u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] - (u_\ell)^2 G_{\ell k}, && (15.22) \\ \forall x \in \mathbb{R}^n, q_{\ell k}(x) &= u_\ell u_k [G_{\ell k} \sin(\theta_\ell - \theta_k) - B_{\ell k} \cos(\theta_\ell - \theta_k)] + (u_\ell)^2 B_{\ell k}. \end{aligned}$$

- If there is a real power flow limit of $\overline{p}_{\ell k}$ on the line joining bus ℓ and k then we represent this limit as an inequality constraint of the form $p_{\ell k}(x) \leq \overline{p}_{\ell k}$ in the inequality constraints $h(x) \leq \overline{h}$.

15.6.2.5 Problem

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, \underline{x} \le x \le \overline{x}, \underline{h} \le h(x) \le \overline{h} \}. \tag{15.23}$$

15.6.3 *Changes in demand, lines, and generators*

- We can consider changes in demand at buses and also consider changes in the system:
  - failure or return to service of a transmission line, and
  - failure or return to service of a generator.

15.6.4 Problem characteristics

15.6.4.1 Convexity

Objective

- As argued in the least-cost production case study in Section 12.1, the objective of this problem is typically convex.

Equality constraints

- Because the function g is non-linear, the set $\{x \in \mathbb{R}^n | g(x) = \mathbf{0}\}$ is not generally convex.
- We can argue from two perspectives that this non-convexity does not necessarily create multiple local minima of the problem.
- First, following the discussion in Section 8.2.4, we observe that the Jacobian J of g can often be well approximated by a constant; that is, the equations are approximately linear.
- Since the equations are approximately linear, the feasible set $\{x \in \mathbb{R}^n | g(x) = \mathbf{0}\}$ is not very different from a set defined by a linear equality constraint.

Equality constraints, continued

- Second, if we can "throw away" real and reactive power, then we can replace the power flow equalities with inequalities.

$$p_\ell(x) \leq 0, \tag{15.24}$$
$$q_\ell(x) \leq 0. \tag{15.25}$$

- That is, we have relaxed the constraints to requiring that the net power flowing out of a node is at most zero.
- That is, we allow power to flow into a node or to be generated at a node and be "thrown away."
- Consider solving the relaxed problem having inequality constraints as specified in (15.24) and (15.25) at each bus ℓ, but with all the other constraints as represented in Problem (15.23):

$$\min_{x\in\mathbb{R}^n}\{f(x)|g(x)\leq \mathbf{0}, \underline{x}\leq x\leq \overline{x}, \underline{h}\leq h(x)\leq \overline{h}\}. \tag{15.26}$$

Equality constraints, continued

- In Problem (15.26), the feasible set $\overline{\mathbb{S}} = \{x \in \mathbb{R}^n | g(x) \leq \mathbf{0}, \underline{x} \leq x \leq \overline{x}, \underline{h} \leq h(x) \leq \overline{h}\}$ is a relaxed version of the feasible set of Problem (15.23):

$$\mathbb{S} = \{x \in \mathbb{R}^n | g(x) = \mathbf{0}, \underline{x} \leq x \leq \overline{x}, \underline{h} \leq h(x) \leq \overline{h}\}.$$

- Suppose we obtain a solution $x^\star \in \overline{\mathbb{S}}$ to Problem (15.26) such that at bus ℓ we have $p_\ell(x^\star) < 0$ or $q_\ell(x^\star) < 0$.
- In this case, so long as we can dispose of real or reactive power at bus ℓ, then we can consider "throwing away" the difference and re-establishing equality to construct a solution $x^{\star\star} \in \mathbb{S}$ to the original equality-constrained problem with the same value of objective and all constraints satisfied.
- From a practical perspective, if there is a generator at ℓ then to "throw away" power at bus ℓ we can consider reducing the output of the generator to enable satisfaction of the constraint with equality.
- This would reduce the objective of the problem since costs typically increase with output.

Equality constraints, continued

- In summary, the inequality-constrained Problem (15.26) has essentially the same solution as Problem (15.23).
- Now we will show that the feasible set defined by the relaxed constraints (15.24) is convex under the assumption that all voltage magnitudes are constant.
- We will not consider the reactive power constraints (15.25) nor the case where voltage magnitudes can vary.

Equality constraints, continued

- Recall that p_ℓ is defined in (6.12) to be:

$$
\begin{aligned}
&\forall x \in \mathbb{R}^n, \\
&\quad p_\ell(x) = \sum_{k \in \mathbb{J}(\ell) \cup \{\ell\}} u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] - P_\ell, \\
&\quad\quad = \sum_{k \in \mathbb{J}(\ell)} u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] + (u_\ell)^2 G_{\ell\ell} - P_\ell, \\
&\quad\quad = \sum_{k \in \mathbb{J}(\ell)} \{u_\ell u_k [G_{\ell k} \cos(\theta_\ell - \theta_k) + B_{\ell k} \sin(\theta_\ell - \theta_k)] - (u_\ell)^2 G_{\ell k}\} \\
&\quad\quad\quad + (u_\ell)^2 \left(G_{\ell\ell} + \sum_{k \in \mathbb{J}(\ell)} G_{\ell k} \right) - P_\ell, \\
&\quad\quad\quad\quad \text{on adding and subtracting } (u_\ell)^2 \textstyle\sum_{k \in \mathbb{J}(\ell)} G_{\ell k}, \\
&\quad\quad = \sum_{k \in \mathbb{J}(\ell)} p_{\ell k}(x) + (u_\ell)^2 \left(G_{\ell\ell} + \sum_{k \in \mathbb{J}(\ell)} G_{\ell k} \right) - P_\ell,
\end{aligned}
$$

- where for each $k \in \mathbb{J}(\ell)$, the function $p_{\ell k}$ was defined in (15.22).

Equality constraints, continued

- Since all voltages are assumed constant, we can define functions $\hat{p}_{\ell k} : \mathbb{R} \to \mathbb{R}$ by:

$$\forall k \in \mathbb{J}(\ell), \forall \theta_{\ell k} \in \mathbb{R}, \hat{p}_{\ell k}(\theta_{\ell k}) = u_\ell u_k [G_{\ell k} \cos(\theta_{\ell k}) + B_{\ell k} \sin(\theta_{\ell k})] - (u_\ell)^2 G_{\ell k},$$

- and we obtain that:

$$\forall \ell, \forall x \in \mathbb{R}^n, p_\ell(x) = \sum_{k \in \mathbb{J}(\ell)} \hat{p}_{\ell k}(\theta_\ell - \theta_k) + (u_\ell)^2 \left(G_{\ell\ell} + \sum_{k \in \mathbb{J}(\ell)} G_{\ell k} \right) - P_\ell.$$

- That is, p_ℓ is equal to $\{(u_\ell)^2 (G_{\ell\ell} + \sum_{k \in \mathbb{J}(\ell)} G_{\ell k}) - P_\ell\}$ plus the sum of terms $\hat{p}_{\ell k}(\theta_\ell - \theta_k)$ each of which depends only on a linear function of two of the entries of x.
- We will find conditions for $\hat{p}_{\ell k}$ to be convex, which will therefore guarantee that p_ℓ is convex.
- We calculate the second derivative of $\hat{p}_{\ell k}$.

Equality constraints, continued

$$\forall \theta_{\ell k} \in \mathbb{R}, \frac{d^2 \hat{p}_{\ell k}}{d\theta_{\ell k}{}^2}(\theta_{\ell k}) = -u_\ell u_k [G_{\ell k} \cos(\theta_{\ell k}) + B_{\ell k} \sin(\theta_{\ell k})].$$

- Recalling that $G_{\ell k} < 0, B_{\ell k} > 0$ for $k \in \mathbb{J}(\ell)$, this expression is positive if:

$$\forall k \in \mathbb{J}(\ell), |G_{\ell k}| \cos(\theta_{\ell k}) - |B_{\ell k}| \sin(\theta_{\ell k}) \geq 0.$$

- This will be true if:

$$-\pi + \arctan\left(\frac{|G_{\ell k}|}{|B_{\ell k}|}\right) \leq \theta_\ell - \theta_k \leq \arctan\left(\frac{|G_{\ell k}|}{|B_{\ell k}|}\right).$$

- Considering power balance at bus $k \in \mathbb{J}(\ell)$ as well, the functions will be convex if for each line joining a bus ℓ to a bus k we have:

$$|\theta_\ell - \theta_k| \leq \min\left\{\arctan\left(\frac{|G_{\ell k}|}{|B_{\ell k}|}\right), \pi - \arctan\left(\frac{|G_{\ell k}|}{|B_{\ell k}|}\right)\right\}. \qquad (15.27)$$

Equality constraints, continued

- Typically, $|G_{\ell k}|/|B_{\ell k}| \approx 0.1$ so (15.27) requires that $|\theta_\ell - \theta_k| \leq 0.1\,\text{radian} \approx 6^\circ$.
- This is a little more restrictive than the angle restrictions (15.21) that we previously mentioned for stability limits in Section 15.6.2.4.

Inequality constraints

- Similarly, if a flow constraint between ℓ and k requires that $p_{\ell k}(x) \leq \overline{p}_{\ell k}$ then the constraint defines a convex set if (15.27) holds.

Discussion

- We have provided sufficient conditions under which the optimal power flow problem is convex.
- If these assumptions are violated then there may be multiple local minimizers.

15.6.4.2 Solvability

- There are a variety of constraints in the optimal power flow problem and it is easily possible for there to be no solution.

16

Algorithms for non-negatively constrained minimization

- In this chapter we will develop algorithms for constrained optimization problems of the form:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, x \geq \mathbf{0} \}, \tag{16.1}$$

- where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$.

Key issues

- Optimality conditions for **non-negatively constrained problems** based on the results for equality-constrained problems,
- the **complementary slackness conditions** in the optimality conditions,
- optimality conditions for **convex problems**, and
- **active set** and **interior point** algorithms to seek solutions.

16.1 Optimality conditions

16.1.1 First-order necessary conditions

16.1.1.1 Analysis

Theorem 16.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. Consider Problem (16.1),*

$$\min_{x\in\mathbb{R}^n}\{f(x)|Ax = b, x \geq \mathbf{0}\},$$

and a point $x^\star \in \mathbb{R}^n$. If $x^\star$ is a local minimizer of Problem (16.1) then:

$$\begin{aligned} \exists\lambda^\star \in \mathbb{R}^m, \exists\mu^\star \in \mathbb{R}^n \textit{ such that: } \nabla f(x^\star) + A^\dagger\lambda^\star - \mu^\star &= \mathbf{0}; \\ Ax^\star &= b; \\ M^\star x^\star &= \mathbf{0}; \\ x^\star &\geq \mathbf{0}; \textit{ and} \\ \mu^\star &\geq \mathbf{0}, \end{aligned} \qquad (16.2)$$

where $M^\star = \operatorname{diag}\{\mu^\star_\ell\}$ is a diagonal matrix with diagonal entries equal to $\mu^\star_\ell$. The vectors $\lambda^\star$ and $\mu^\star$ satisfying the conditions (16.2) are called the vectors of Lagrange multipliers for the constraints $Ax = b$ and $x \geq \mathbf{0}$, *respectively. The conditions that* $M^\star x^\star = \mathbf{0}$ *are called the* **complementary slackness conditions**. *The complementary slackness conditions together with the conditions* $x^\star \geq \mathbf{0}$ *and* $\mu^\star \geq \mathbf{0}$ *imply that, for each* ℓ*, either the* ℓ*-th non-negativity constraint* $x_\ell \geq 0$ *is binding or the* ℓ*-th Lagrange multiplier* $\mu_\ell^\star$ *is equal to zero (or both).*

Proof This is a special case of Theorem 17.1 to be presented in Chapter 17. We will only sketch the proof of this special case. Consider the *equality*-constrained problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, -x_\ell = 0, \forall \ell \in \mathbb{A}(x^\star) \}, \tag{16.3}$$

where $\mathbb{A}(x^\star) = \{\ell \in \{1, \ldots, n\} | x^\star_\ell = 0\}$ is the active set corresponding to the non-negativity constraints $x \geq \mathbf{0}$ for the point $x^\star$. That is, the equality-constrained Problem (16.3) includes as equality constraints the following:

- all of the equality constraints from Problem (16.1) and
- all of the non-negativity constraints of Problem (16.1) that were satisfied with equality by $x^\star$.

That is, the active non-negativity constraints from Problem (16.1) at its minimizer have been included as *equality* constraints in Problem (16.3). The representation of the constraint as $-x_\ell = 0$ rather than as $x_\ell = 0$ is for convenience in interpreting the Lagrange multipliers for equality-constrained Problem (16.3).

The proof involves applying our earlier results for *equality*-constrained problems to Problem (16.3) to prove the theorem. The proof is divided into three parts:

(i) showing that $x^\star$ is a local minimizer of equality-constrained Problem (16.3),

(ii) using the necessary conditions of the equality-constrained Problem (16.3) to define $\lambda^\star$ and $\mu^\star$ that satisfy the first four lines of (16.2), and

(iii) proving that $\mu^\star \geq \mathbf{0}$ by showing that if a particular Lagrange multiplier were negative, say $\mu_\ell^\star < 0$, then the objective could be reduced by moving in a direction such that x_ℓ increases and so becomes strictly feasible for the constraint $x_\ell \geq 0$. The intuition behind this observation is that if the second-order sufficient conditions held for Problem (16.3) at $x^\star$ then we could apply the sensitivity analysis Corollary 13.11. If we consider changing the constraint from $-x_\ell = 0$ to $-x_\ell = -\gamma$, with $\gamma > 0$, then, since $\mu_\ell^\star < 0$, Corollary 13.11 indicates that the minimum of the changed problem would be lower and x_ℓ would be strictly positive. This means that the constraint $x_\ell \geq 0$ could not have been binding at a minimizer of Problem (16.1) since a positive value of x_ℓ would reduce the objective.

□

16.1.1.2 Example

- Consider Problem (2.15):

$$\min_{x \in \mathbb{R}^2} \{x_1 - x_2 | x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0\}. \tag{16.4}$$

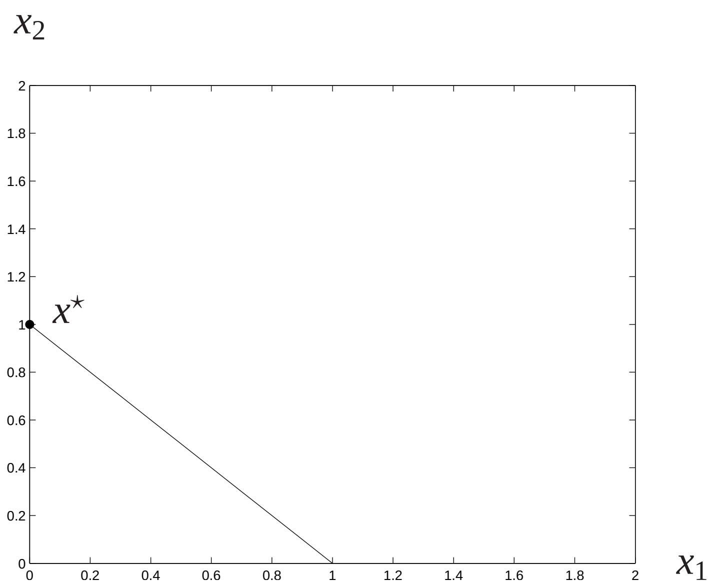

Fig. 16.1. Feasible set (shown as line) and minimizer $x^\star$ (shown as $\bullet$) for example problem.

- Consideration of the objective and inspection of Figure 16.1 shows that $x^\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is the unique minimizer of Problem (16.4).
- We apply Theorem 16.1 to this non-negatively constrained problem.
- The objective is linear, and hence partially differentiable with continuous partial derivatives.

$$\begin{aligned} \forall x \in \mathbb{R}^2, f(x) &= x_1 - x_2, \\ \forall x \in \mathbb{R}^2, \nabla f(x) &= \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \\ A &= \begin{bmatrix} 1 & 1 \end{bmatrix}, \\ &= \mathbf{1}^\dagger, \\ b &= [1]. \end{aligned}$$

- We claim that $\lambda^\star = [1]$ and $\mu^\star = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ satisfy (16.2).

Example, continued

$$
\begin{aligned}
\nabla f(x^\star) + A^\dagger \lambda^\star - \mu^\star &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} [1] - \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \\
&= \mathbf{0}; \\
M^\star x^\star &= \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \\
&= \mathbf{0}; \\
A x^\star &= \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \\
&= [1], \\
&= b; \\
x^\star &= \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \\
&\geq \mathbf{0}; \text{ and} \\
\mu^\star &= \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \\
&\geq \mathbf{0}.
\end{aligned}
$$

16.1.1.3 Discussion

- As in the equality-constrained case, the Lagrange multipliers adjust the unconstrained optimality conditions to balance the constraints against the objective.
- We will refer to the equality and inequality constraints specified in (16.2) as *the* first-order necessary conditions.
- There are inequality constraints on both the minimizer $x^\star$ and on the Lagrange multipliers $\mu^\star$ in the first-order conditions for inequality constraints.

Discussion, continued

- The complementary slackness conditions require that $M^\star x^\star = \mathbf{0}$.
- Consider linearizing Mx about $\mu^{(\nu)}$ and $x^{(\nu)}$ and using the linearized equations to construct an update.
- This approach is not effective unless we are careful to avoid the boundary of the set defined by $x \geq \mathbf{0}$ and $\mu \geq \mathbf{0}$.
- For example, suppose that at iteration ν we had $x_\ell^{(\nu)} = 0$.
- In this case, for the particular entry ℓ, linearizing the complementary slackness conditions involves linearizing $\mu_\ell x_\ell$ about $\mu_\ell^{(\nu)}$ and $x_\ell^{(\nu)}$.
- We obtain:

$$\begin{aligned}(\mu_\ell^{(\nu)} + \Delta\mu_\ell^{(\nu)})(x_\ell^{(\nu)} + \Delta x_\ell^{(\nu)}) &\approx \mu_\ell^{(\nu)} x_\ell^{(\nu)} + x_\ell^{(\nu)} \Delta\mu_\ell^{(\nu)} + \mu_\ell^{(\nu)} \Delta x_\ell^{(\nu)}, \\ &= \mu_\ell^{(\nu)} \Delta x_\ell^{(\nu)}, \text{ since } x_\ell^{(\nu)} = 0.\end{aligned}$$

- Setting this equal to zero yields $\Delta x_\ell^{(\nu)} = 0$.
- If we ever were at an iterate for which $x_\ell^{(\nu)} = 0$, then the Newton–Raphson update would prevent us from ever moving from this value.

Discussion, continued

- Linearizing the complementary slackness conditions does not yield a useful approximation in these cases.
- We will see in Section 16.4.3.3 that an effective linearization of this constraint requires us to carefully avoid the possibilities that $x_\ell^{(\nu)} = 0$ or $\mu_\ell^{(\nu)} = 0$.
- We will see that one way to do this is to *first* approximate the constraint $\mu_\ell x_\ell = 0$ by a hyperbola $\mu_\ell x_\ell = t$, where $t \in \mathbb{R}_{++}$, and then linearize the hyperbolic approximation.
- Then, we gradually reduce t.
- Figure 16.2 shows a hyperbolic approximation to the set of points satisfying the complementary slackness constraint for several values of t.
- As t is reduced, the set of ordered pairs $\begin{bmatrix} \mu_\ell \\ x_\ell \end{bmatrix}$ satisfying $\mu_\ell x_\ell = t$, $x_\ell \geq 0$, and $\mu_\ell \geq 0$ approaches the union of the non-negative μ_ℓ-axis and the non-negative x_ℓ-axis.

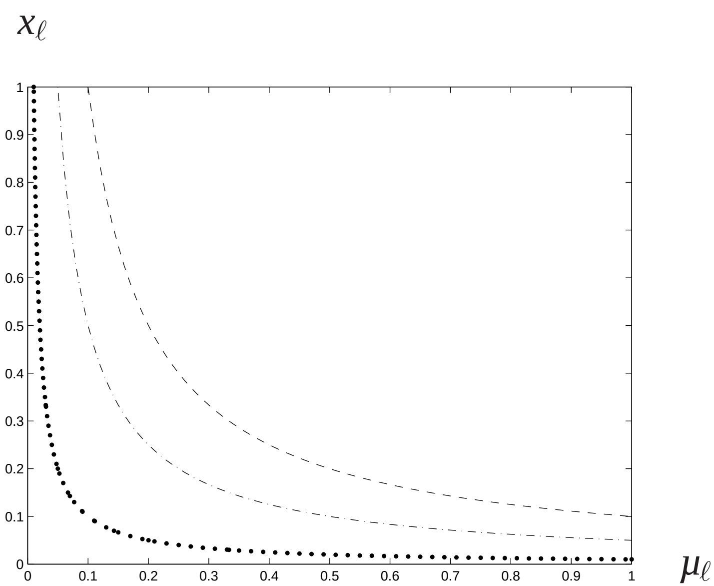

Fig. 16.2. The complementary slackness constraint for the entry ℓ requires that the point $\begin{bmatrix} \mu_\ell \\ x_\ell \end{bmatrix} \in \mathbb{R}^2$ lie either on the μ_ℓ-axis or on the x_ℓ-axis. The hyperbola $\mu_\ell x_\ell = t$ approximates the set of points satisfying the complementary slackness constraints. The dashed curve shows the hyperbola for $t = 0.1$; the dash-dot curve shows the hyperbola for $t = 0.05$; and the dotted curve shows the hyperbola for $t = 0.01$.

16.1.2 Second-order sufficient conditions

16.1.2.1 Analysis

Theorem 16.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. Consider Problem (16.1),*

$$\min_{x\in\mathbb{R}^n}\{f(x)|Ax=b, x\geq \mathbf{0}\},$$

and points $x^\star \in \mathbb{R}^n, \lambda^\star \in \mathbb{R}^m$, and $\mu^\star \in \mathbb{R}^n$. Let $M^\star = \mathrm{diag}\{\mu^\star_\ell\}$. Suppose that:

(i) *f is twice partially differentiable with continuous second partial derivatives,*

(ii)

$$\begin{aligned}
\nabla f(x^\star) + A^\dagger \lambda^\star - \mu^\star &= \mathbf{0};\\
M^\star x^\star &= \mathbf{0};\\
Ax^\star &= b;\\
x^\star &\geq \mathbf{0};\\
\mu^\star &\geq \mathbf{0};
\end{aligned}$$

and

(iii) $\nabla^2 f(x^\star)$ *is positive definite on the null space:*

$$\mathcal{N}_+ = \{\Delta x \in \mathbb{R}^n | A\Delta x = \mathbf{0}; \Delta x_\ell = 0, \forall \ell \in \mathbb{A}_+(x^\star, \mu^\star)\},$$

where $\mathbb{A}_+(x^\star, \mu^\star) = \{\ell \in \{1, \ldots, n\} | x^\star_\ell = 0, \mu^\star_\ell > 0\}$.

Then $x^\star$ *is a strict local minimizer of Problem (16.1).* □

- The conditions (i)–(iii) in the theorem are called the **second-order sufficient conditions**.
- In addition to the first-order necessary conditions, the second-order sufficient conditions require that:

  f is twice partially differentiable with continuous second partial derivatives, and

  $\nabla^2 f(x^\star)$ is positive definite on the null space $\mathcal{N}_+$ defined in the theorem.

16.1.2.2 Example

- Consider the objective $f : \mathbb{R}^2 \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}^2, f(x) = (x_1)^2 + (x_2 - 1)^2.$$

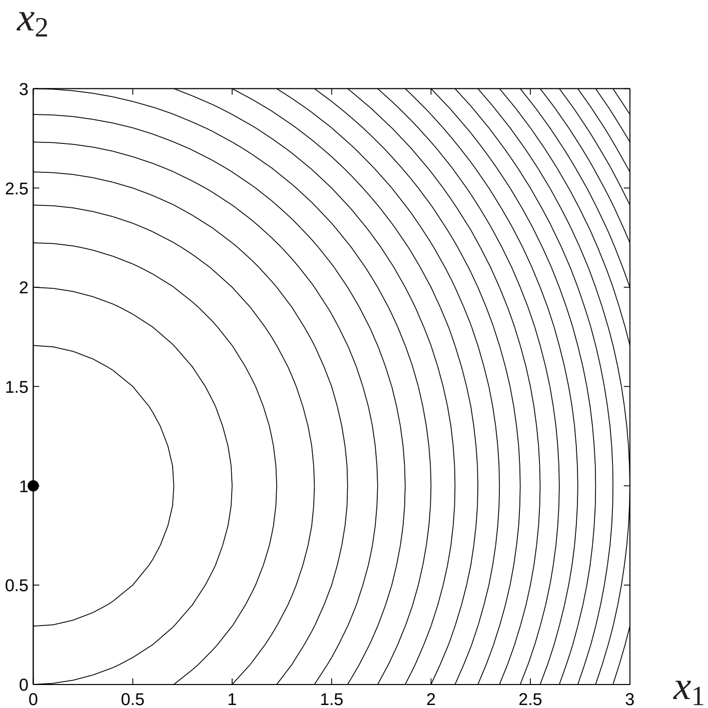

Fig. 16.3. Contour sets of objective function defined in section 16.1.2.2. The heights of the contours decrease towards the point $x^\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, which is indicated with the •.

Example, continued

- Consider the problem:

$$\min_{x \in \mathbb{R}^2} \{ f(x) | x \geq \mathbf{0} \}.$$

- The objective is twice partially differentiable with continuous second partial derivatives.
- We claim that the second-order sufficient conditions hold for $x^\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\mu^\star = \mathbf{0}$, which is illustrated as the $\bullet$ in Figure 16.3.

Example, continued

- The second-order sufficient conditions are that:

$$\begin{aligned}
\nabla f(x^\star) - \mu^\star &= \begin{bmatrix} 2x_1^\star \\ 2(x_2^\star - 1) \end{bmatrix} - \mu^\star, \\
&= \mathbf{0}; \\
M^\star x^\star &= \mathbf{0}; \\
x^\star &\geq \mathbf{0}; \\
\mu^\star &\geq \mathbf{0};
\end{aligned}$$

- and that $\nabla^2 f(x^\star) = 2\mathbf{I}$ is positive definite on the null space:

$$\mathcal{N}_+ = \{\Delta x \in \mathbb{R}^2 | \Delta x_\ell = 0, \forall \ell \in \mathbb{A}_+(x^\star, \mu^\star)\}.$$

Example, continued

$$
\begin{aligned}
\mathbb{A}(x^\star) &= \{1\}, \\
\mathbb{A}_+(x^\star, \mu^\star) &= \{\ell \in \{1,2\} | x_\ell^\star = 0, \mu_\ell^\star > 0\}, \\
&= \emptyset, \\
\mathcal{N}_+ &= \{\Delta x \in \mathbb{R}^2 | \Delta x_\ell = 0, \forall \ell \in \mathbb{A}_+(x^\star, \mu^\star)\}, \\
&= \{\Delta x \in \mathbb{R}^2 | \Delta x_\ell = 0, \forall \ell \in \emptyset\}, \\
&= \mathbb{R}^2, \\
\nabla^2 f(x^\star) &= 2\mathbf{I},
\end{aligned}
$$

- which is positive definite on $\mathcal{N}_+ = \mathbb{R}^2$.
- The second-order sufficient conditions hold at $x^\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\mu^\star = \mathbf{0}$.
- Note that $\mathbb{A}_+(x^\star, \mu^\star) = \emptyset$ is a strict subset of $\mathbb{A}(x^\star) = \{1\}$ for this example.

16.1.2.3 Discussion

- The set $\mathbb{A}_{+}(x^{\star},\mu^{\star})$ can be a *strict* subset of $\mathbb{A}(x^{\star})$ since $\mathbb{A}_{+}(x^{\star},\mu^{\star})$ omits those constraints ℓ for which $x_{\ell}^{\star}=0$ and $\mu_{\ell}^{\star}=0$.

$$\mathcal{N}_{+}=\{\Delta x\in\mathbb{R}^{n}|A\Delta x=\mathbf{0},\Delta x_{\ell}=0,\forall\ell\in\mathbb{A}_{+}(x^{\star},\mu^{\star})\},$$

- can strictly contain the null space:

$$\mathcal{N}=\{\Delta x\in\mathbb{R}^{n}|A\Delta x=\mathbf{0},\Delta x_{\ell}=0,\forall\ell\in\mathbb{A}(x^{\star})\},$$

- corresponding to the constraints of equality-constrained Problem (16.3):

$$\begin{aligned} Ax &= b, \\ -x_{\ell} &= 0,\forall\ell\in\mathbb{A}(x^{\star}). \end{aligned}$$

- By Corollary 13.4, if $x^{\star}$ satisfies the first-order necessary conditions for equality-constrained Problem (16.3) and if $\nabla^{2}f(x^{\star})$ is positive definite on the null space $\mathcal{N}$ then $x^{\star}$ is a strict local minimizer of equality-constrained Problem (16.3).
- However, this is *insufficient* to guarantee that $x^{\star}$ is a strict local minimizer of the corresponding inequality-constrained Problem (16.1) if there are any constraints ℓ for which both $x_{\ell}^{\star}=0$ and $\mu_{\ell}^{\star}=0$.

Discussion, continued

- Constraints for which $x_\ell^\star = 0$ and $\mu_\ell^\star = 0$ are called **degenerate constraints**.
- Intuitively, a degenerate constraint ℓ is only "just" binding.
- The sensitivity of the minimum to changes in x_ℓ is zero.
- There exist feasible movements Δx away from $x^\star$, namely those in which $\Delta x_\ell > 0$, for which the constraint $x_\ell \geq 0$ is no longer binding.
- Such feasible movements do not satisfy $\Delta x_\ell = 0$, so to guarantee that $x^\star$ is a minimizer of Problem (16.1) we must test for positive definiteness of the objective in the larger subspace that allows movement in directions Δx such that $\Delta x_\ell > 0$.
- If the Hessian is positive definite in these directions then the objective must increase in these directions as we move away from $x^\star$ and consequently we are indeed at a local minimizer of Problem (16.1).
- That is, if $\nabla^2 f(x^\star)$ is positive definite on $\mathcal{N}_+$ then there can be no feasible descent directions for f at $x^\star$.

16.1.2.4 *Example of not satisfying second-order sufficient conditions*

- Suppose that we have the objective $f : \mathbb{R}^2 \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}^2, f(x) = -(x_1)^3 + (x_2 - 1)^2.$$

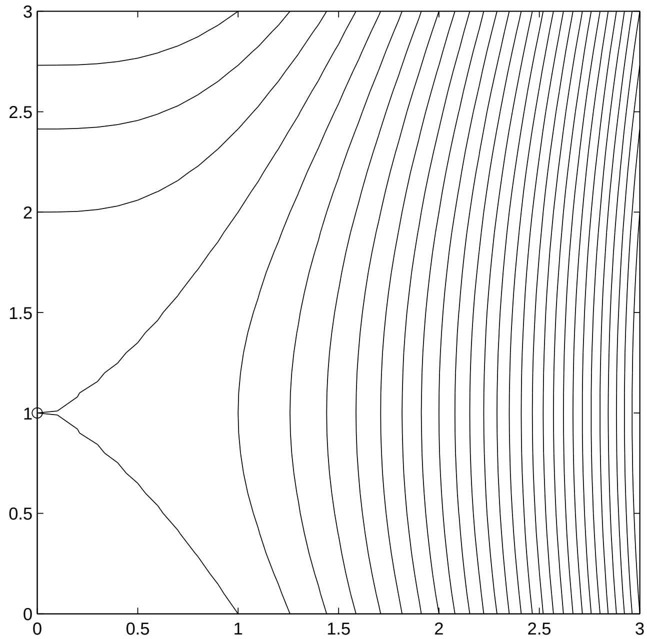

Fig. 16.4. Contour sets of objective function defined in section 16.1.2.4. The heights of the contours decrease away from the point $\hat{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, which is indicated with the $\circ$, in the direction of increasing values of x_1. The heights of the contours increase away from the point $\hat{x}$ in the direction of increasing or decreasing values of x_2.

Example of not satisfying second-order sufficient conditions, continued

- Consider the problem:

$$\min_{x\in\mathbb{R}^2}\{f(x)|x \geq \mathbf{0}\}.$$

- The problem is unbounded below on the feasible set and has no minimizer.
- However, consider the candidate minimizer $\hat{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and candidate value of Lagrange multipliers $\hat{\mu} = \mathbf{0}$:

$$\begin{aligned} \nabla f(\hat{x}) - \hat{\mu} &= \mathbf{0}, \\ \widehat{M}\hat{x} &= \mathbf{0}, \\ \hat{x} &\geq \mathbf{0}, \\ \hat{\mu} &\geq \mathbf{0}, \end{aligned}$$

- so that $\hat{x}$ and $\hat{\mu}$ satisfy the first-order necessary conditions, where $\widehat{M} = \mathrm{diag}\{\hat{\mu}_\ell\}$ is a diagonal matrix with diagonal entries equal to $\hat{\mu}_\ell$.

Example of not satisfying second-order sufficient conditions, continued

- The active set for $x \geq \mathbf{0}$ at $\hat{x}$ includes the first non-negativity constraint.

$$
\begin{aligned}
\mathbb{A}(\hat{x}) &= \{1\}, \\
&\neq \mathbb{A}_+(\hat{x},\hat{\mu}), \\
&= \{\ell \in \{1,2\} | x^\star_\ell = 0, \mu^\star_\ell > 0\}, \\
&= \emptyset.
\end{aligned}
$$

- Therefore, if $\hat{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\hat{\mu} = \mathbf{0}$ *were* the minimizer and corresponding Lagrange multipliers of this problem, then the constraint would be degenerate.

Example of not satisfying second-order sufficient conditions, continued

- The Hessian of the objective is:

$$\nabla^2 f(\hat{x}) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}.$$

- The subspace corresponding to the constraints of equality-constrained Problem (16.3) is:

$$\begin{aligned} \mathcal{N} &= \{\Delta x \in \mathbb{R}^2 | A\Delta x = \mathbf{0}, \Delta x_\ell = 0, \forall \ell \in \mathbb{A}(\hat{x})\}, \\ &= \{\Delta x \in \mathbb{R}^2 | \Delta x_\ell = 0, \forall \ell \in \{1\}\}, \\ &= \{\Delta x \in \mathbb{R}^2 | \Delta x_1 = \mathbf{0}\}. \end{aligned}$$

- Note that:

$$\begin{aligned} \forall \Delta x \in \mathcal{N}, (\Delta x \neq \mathbf{0}) &\Rightarrow (\Delta x_1 = 0, \Delta x_2 \neq 0), \\ &\Rightarrow (\Delta x^\dagger \nabla^2 f(\hat{x}) \Delta x = 2(\Delta x_2)^2 > 0). \end{aligned}$$

- The Hessian *is* positive definite on $\mathcal{N}$ and, by Corollary 13.4, $\hat{x}$ is a local minimizer of the equality-constrained problem $\min_{x \in \mathbb{R}^2}\{f(x) | -x_1 = 0\}$.

Example of not satisfying second-order sufficient conditions, continued

- But positive definiteness on $\mathcal{N}$ is *insufficient* to guarantee local optimality for Problem (16.1).
- In fact, $\nabla^2 f(\hat{x})$ is not positive definite on the null space $\mathcal{N}_+$ specified in Theorem 16.2:

$$\begin{aligned}
\mathcal{N}_+ &= \{\Delta x \in \mathbb{R}^2 | A\Delta x = \mathbf{0}, \Delta x_\ell = 0, \forall \ell \in \mathbb{A}_+(\hat{x}, \hat{\mu})\}, \\
&= \{\Delta x \in \mathbb{R}^2 | \Delta x_\ell = 0, \forall \ell \in \emptyset\}, \\
&= \mathbb{R}^2.
\end{aligned}$$

- The second-order sufficient conditions do not hold and $\hat{x}$ is not a minimizer of the problem.

16.2 Convex problems

16.2.1 First-order sufficient conditions

16.2.1.1 Analysis

Theorem 16.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. Consider Problem (16.1),*

$$\min_{x\in\mathbb{R}^n}\{f(x)|Ax = b, x \geq \mathbf{0}\},$$

and points $x^\star \in \mathbb{R}^n$, $\lambda^\star \in \mathbb{R}^m$, and $\mu^\star \in \mathbb{R}^n$. Let $M^\star = \mathrm{diag}\{\mu^\star_\ell\}$. Suppose that:

(i) *f is convex on $\{x \in \mathbb{R}^n|Ax = b, x \geq \mathbf{0}\}$,*

(ii) $\nabla f(x^\star) + A^\dagger\lambda^\star - \mu^\star = \mathbf{0}$,

(iii) $M^\star x^\star = \mathbf{0}$,

(iv) *$Ax^\star = b$ and $x^\star \geq \mathbf{0}$, and*

(v) $\mu^\star \geq \mathbf{0}$.

Then $x^\star$ is a global minimizer of Problem (16.1).

Proof By Item (iv), $x^\star$ is feasible. Consider any other feasible point $x \in \mathbb{R}^n$. That is, consider x such that:

$$Ax = b, x \geq \mathbf{0}.$$

We have $Ax = Ax^\star = b$, so $A(x - x^\star) = \mathbf{0}$ and:

$$[\lambda^\star]^\dagger A(x - x^\star) = \mathbf{0}. \tag{16.5}$$

We now consider constraints $\ell \in \mathbb{A}(x^\star)$ and constraints $\ell \notin \mathbb{A}(x^\star)$ separately.
For $\ell \notin \mathbb{A}(x^\star)$, we have that $x^\star_\ell > 0$. Consequently, Item (iii) implies that $\mu^\star_\ell = 0$. Therefore,

$$\forall \ell \notin \mathbb{A}(x^\star), \mu^\star_\ell(x_\ell - x^\star_\ell) = 0. \tag{16.6}$$

For $\ell \in \mathbb{A}(x^\star)$, we have that $x^\star_\ell = 0$. Moreover, since $x_\ell \geq 0$ for all ℓ, we have:

$$\begin{aligned} \forall \ell \in \mathbb{A}(x^\star), x_\ell - x^\star_\ell &= x_\ell - 0, \\ &\geq 0. \end{aligned}$$

Therefore, since $\mu_\ell^\star \geq 0$, we have:

$$\forall \ell \in \mathbb{A}(x^\star), \mu_\ell^\star (x_\ell - x_\ell^\star) \geq 0. \tag{16.7}$$

Combining (16.6) and (16.7), we have:

$$\begin{aligned}
[\mu^\star]^\dagger (x - x^\star) &= \sum_{\ell \in \mathbb{A}(x^\star)} \mu_\ell^\star (x_\ell - x_\ell^\star) + \sum_{\ell \notin \mathbb{A}(x^\star)} \mu_\ell^\star (x_\ell - x_\ell^\star), \\
&= \sum_{\ell \in \mathbb{A}(x^\star)} \mu_\ell^\star (x_\ell - x_\ell^\star), \text{ by (16.6)}, \\
&\geq 0, \text{ by (16.7)}. \qquad (16.8)
\end{aligned}$$

We have:

$$
\begin{aligned}
f(x) \;\geq\;& f(x^\star) + \nabla f(x^\star)^\dagger (x - x^\star), \text{ by Theorem 2.6, noting that:} \\
& \quad f \text{ is partially differentiable with continuous partial derivatives;} \\
& \quad f \text{ is convex on the convex set } \{x \in \mathbb{R}^n | Ax = b, x \geq \mathbf{0}\}, \\
& \qquad\qquad \text{by Item (i) of the hypothesis; and} \\
& \quad x, x^\star \in \{x \in \mathbb{R}^n | Ax = b, x \geq \mathbf{0}\}, \\
& \qquad\qquad \text{by Item (iv) of the hypothesis and construction,} \\
=\;& f(x^\star) - [A^\dagger \lambda^\star - \mu^\star]^\dagger (x - x^\star), \\
& \qquad\qquad \text{by Item (ii) of the hypothesis,} \\
=\;& f(x^\star) - [\lambda^\star]^\dagger A(x - x^\star) + [\mu^\star]^\dagger (x - x^\star), \\
=\;& f(x^\star) + [\mu^\star]^\dagger (x - x^\star), \text{ by (16.5),} \\
\geq\;& f(x^\star), \text{ by (16.8).}
\end{aligned}
$$

Therefore, $x^\star$ is a global minimizer of f on $\{x \in \mathbb{R}^n | Ax = b, x \geq \mathbf{0}\}$. □

16.2.1.2 Example

- Consider again the problem from Section 16.1.2.2:

$$\min_{x\in\mathbb{R}^2}\{f(x)|x\geq\mathbf{0}\},$$

- with objective $f:\mathbb{R}^2\to\mathbb{R}$ defined by:

$$\forall x\in\mathbb{R}^2, f(x)=(x_1)^2+(x_2-1)^2,$$

- The objective is partially differentiable with continuous partial derivatives and convex.
- We have already verified that $x^\star=\begin{bmatrix}0\\1\end{bmatrix}$ and $\mu^\star=[0]$ satisfy the first-order necessary conditions.
- By Theorem 16.3, $x^\star$ is a global minimizer of the problem.

16.3 Approaches to finding minimizers: active set method

- In the **active set method**, we consider a tentative list of the constraints that we believe are binding at the optimum.
- This tentative list is called the **working set** and typically consists of the indices of the binding inequalities at the current iterate.
- Since our tentative list may not be the correct list for the solution, we must consider how to change this tentative list, either by:
  - adding another constraint to the list, which is called **swapping in**, or
  - removing a constraint from the list, which is called **swapping out**.
- Geometrically, active set algorithms tend to step along the boundary of the region defined by the inequality constraints.

16.3.1 *Working set*

- We write $\mathbb{W}^{(\nu)}$ for the working set.
- The constraints in the working set are treated *temporarily* as equality constraints.
- A search direction is calculated that seeks the minimizer of an *equality*-constrained problem where the equality constraints consist of:
  - all the equality constraints in the original problem, and
  - the binding inequality constraints listed in $\mathbb{W}^{(\nu)}$.
- If $\mathbb{W}^{(\nu)}$ happens to coincide with the active set for the minimizer $x^\star$ of the inequality-constrained Problem (16.1) then, by the proof of Theorem 16.1, the solution of the equality-constrained problem using $\mathbb{W}^{(\nu)}$ will be $x^\star$.
- Inequality constraints are "swapped" in and out of the working set as calculations proceed.

16.3.2 Swapping in

16.3.2.1 Descent direction

- Consider iteration ν, the current value of the iterate $x^{(\nu)}$, and a working set $\mathbb{W}^{(\nu)}$.
- Suppose that $x^{(\nu)}$ is feasible with respect to all the constraints.
- We consider the problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, -x_\ell = 0, \forall \ell \in \mathbb{W}^{(\nu)} \}. \tag{16.9}$$

- We can use the algorithms from Chapter 13 to find a descent direction $\Delta x^{(\nu)}$ at $x^{(\nu)}$ for this equality-constrained problem.

16.3.2.2 Step-size

- We seek a step-size for the update that will maintain feasibility with respect to *all* of the constraints in Problem (16.1).
- In particular, consider any inequality constraint ℓ' that is not in the current working set.
- That is, consider $\ell' \notin \mathbb{W}^{(\nu)}$ so that $x_{\ell'}^{(\nu)} > 0$.
- For simplicity, first suppose that the objective function decreases along the descent direction $\Delta x^{(\nu)}$ for arbitrary step-sizes.

Step-size, continued

- Suppose that an update $\Delta x^{(\nu)}$ based on the current working set and a step-size of 1 would cause inequality constraint ℓ' to be violated because $x_{\ell'}^{(\nu)} + \Delta x_{\ell'}^{(\nu)} < 0$.
- Then:
  - the step-size $\alpha^{(\nu)}$ of the update should be chosen to make constraint ℓ' just binding at the next iterate $x_{\ell'}^{(\nu)} + \alpha^{(\nu)} \Delta x_{\ell'}^{(\nu)}$, and
  - the working set should be updated by including constraint ℓ' so that $\mathbb{W}^{(\nu+1)} = \mathbb{W}^{(\nu)} \cup \{\ell'\}$.
- We may find that the function evaluated at $x_{\ell'}^{(\nu)} + \alpha^{(\nu)} \Delta x_{\ell'}^{(\nu)}$ does not satisfy a sufficient decrease criterion.
- In this case, we should decrease the step-size further (and not add the constraint ℓ' to the working set.)

16.3.2.3 Example

- Consider the feasible set $\{x \in \mathbb{R}^3 | \mathbf{1}^\dagger x = 10, x \geq \mathbf{0}\}$.

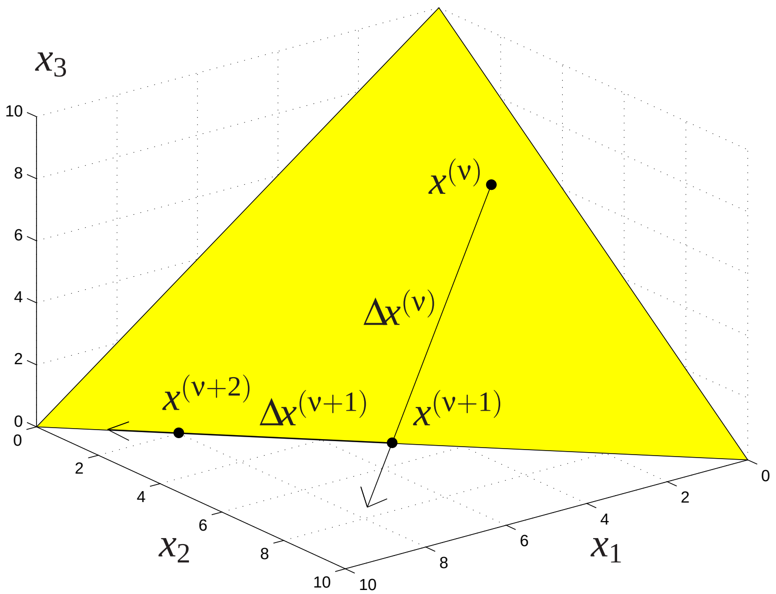

Fig. 16.5. Change in working set.

- This feasible set is an example of a set of the form:

$$\{x \in \mathbb{R}^n | Ax = b, x \geq \mathbf{0}\}, \tag{16.10}$$

- where $A = -\mathbf{1}^\dagger \in \mathbb{R}^{m \times n}$ and $b = [-10] \in \mathbb{R}^m$, for $m = 1$ and $n = 3$.
- This is the same form as the equality constraint in the least-cost production case study of Section 12.1 and we know from Section 12.1.4.2 that $Z = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ is a matrix with columns that form a basis for the null space of A.
- Also illustrated in Figure 16.5 is a current iterate $x^{(\nu)} = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix} \in \mathbb{R}^3_+$ that is feasible for the equality constraint.
- Since $x^{(\nu)} > \mathbf{0}$, we suppose that the current working set is empty, $\mathbb{W}^{(\nu)} = \emptyset$.

Example, continued

- Consider a partially differentiable objective $f : \mathbb{R}^3 \to \mathbb{R}$ such that
$\nabla f(x^{(\nu)}) = \begin{bmatrix} 2 \\ -1 \\ 11 \end{bmatrix}.$
- The vector:

$$\begin{aligned} \Delta x^{(\nu)} &= -ZZ^\dagger \nabla f(x^{(\nu)}), \\ &= \begin{bmatrix} 6 \\ 3 \\ -9 \end{bmatrix} \end{aligned}$$

- is a descent direction for f that lies in the null space of the equality constraint.
- Moving from $x^{(\nu)}$ in the direction $\Delta x^{(\nu)}$ will simultaneously:
  - improve the objective, and
  - maintain satisfaction of the equality constraint $\mathbf{1}^\dagger x = 10$.

Example, continued

- Suppose that the objective decreases along the direction $\Delta x^{(\nu)}$ for step-sizes up to at least 1:

$$\forall \alpha^{(\nu)} \in (0,1], f(x^{(\nu)} + \alpha^{(\nu)} \Delta x^{(\nu)}) < f(x^{(\nu)}).$$

- To maintain feasibility, the update cannot progress past the $x_3 = 0$ plane.
- We must choose $\alpha^{(\nu)}$ such that:

$$\begin{aligned} x^{(\nu+1)} &= x^{(\nu)} + \alpha^{(\nu)} \Delta x^{(\nu)}, \\ &= \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix} + \alpha^{(\nu)} \begin{bmatrix} 6 \\ 3 \\ -9 \end{bmatrix}, \\ &\geq \mathbf{0}. \end{aligned}$$

Example, continued

- To satisfy $x^{(\nu+1)} \geq \mathbf{0}$, a step-size of $\alpha^{(\nu)} = \frac{2}{3}$ is chosen so that $x^{(\nu+1)}$ satisfies $x_3^{(\nu+1)} = 0$ and, therefore, $x^{(\nu+1)} = \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix}$.
- Constraint 3 is added to the working set so that $\mathbb{W}^{(\nu+1)} = \{3\}$.

Example, continued

- We now consider movement in a direction $\Delta x^{(\nu+1)}$ such that:
  $\Delta x^{(\nu+1)}$ is a descent direction for the objective f at $x^{(\nu+1)}$,
  moving in the direction $\Delta x^{(\nu+1)}$ maintains feasibility for the equality constraint $\mathbf{1}^\dagger x = 10$, and
  moving in the direction $\Delta x^{(\nu+1)}$ maintains satisfaction of the equality constraint $-x_3 = 0$ implied by the current working set.
- Suppose that at $x^{(\nu+1)}$ the objective decreases with increasing values of x_1 and decreasing values of x_2.
- Then a suitable update direction is shown in Figure 16.5 as the arrow labelled $\Delta x^{(\nu+1)}$ having its tail at $x^{(\nu+1)}$ and pointing towards $x = \begin{bmatrix} 10 \\ 0 \\ 0 \end{bmatrix}$.

Example, continued

- We update along the direction $\Delta x^{(\nu+1)}$ until a sufficient decrease in the objective is achieved or another constraint becomes binding.
- In the former case, a point such as $x^{(\nu+2)}$ in Figure 16.5 would be obtained.
- In the latter case, another constraint would be added to the working set and the procedure would continue.
- The iterates typically lie on the boundary of the region defined by the inequality constraints.

16.3.3 *Swapping out*

16.3.3.1 *Descent direction*

- We can also consider swapping a constraint ℓ'' out of the feasible set.
- Suppose that for some $\ell'' \in \mathbb{W}^{(\nu)}$ we find that a Lagrange multiplier for the constraint $-x_{\ell''} = 0$ is negative for Problem (16.9).
- In this case, we can potentially reduce the objective by moving in a direction that makes the constraint non-binding and we should consider removing ℓ'' from the working set.
- This approach follows the proof of Theorem 16.1 where a negative value of a Lagrange multiplier corresponding to an inequality constraint allowed us to reduce the objective by moving in a direction such that the constraint became strictly feasible.

Descent direction, continued

- In practice, the equality-constrained problems may not be solved to optimality, so that the Lagrange multiplier estimate may be in error.
- Consequently, the working set approach can be prone to “zig-zagging” where constraints repeatedly move in and out of the active set without significant progress.
- Various strategies have been devised to avoid erroneously swapping a constraint out.
- Nevertheless, suppose that we choose to swap out constraint ℓ'' to update the working set at iteration ν.
- Then we revise the working set to be $\mathbb{W}^{(\nu)} \setminus \{\ell''\}$.
- That is, we remove ℓ'' from the working set. A descent direction is sought for the corresponding equality-constrained problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, -x_\ell = 0, \forall \ell \in \mathbb{W}^{(\nu)} \setminus \{\ell''\}\}.$$

16.3.3.2 Example

$$\min_{x_1, x_2 \in \mathbb{R}} \{x_1 - x_2 | x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0\}.$$

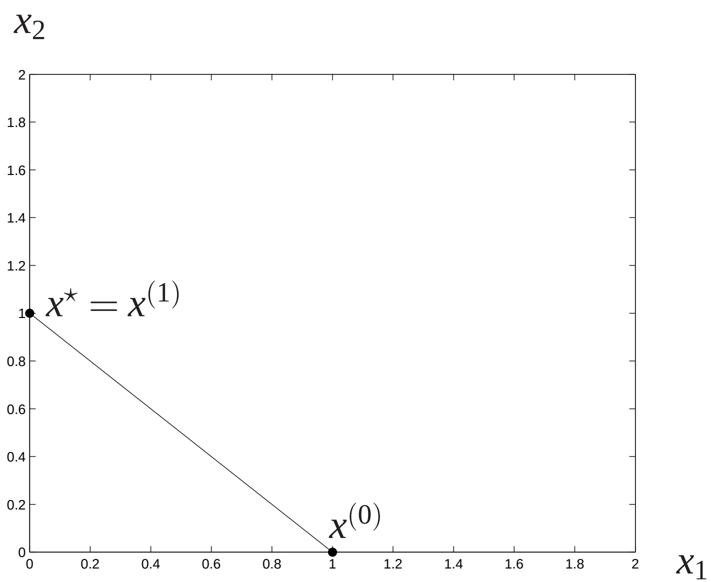

Fig. 16.6. Trajectory of iterates using active set algorithm for example problem. The feasible set is indicated by the solid line.

Example, continued

- Suppose that we start with the initial guess of $x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ for this problem.
- This initial guess is feasible with respect to all the constraints, is strictly feasible with respect to the inequality constraint $x_1 \geq 0$, and the inequality constraint $x_2 \geq 0$ is active at this initial guess.

Working set

- Since the inequality constraint $x_2 \geq 0$ is active for the initial guess, the initial working set is $\mathbb{W}^{(0)} = \{2\}$.

Descent direction at $x^{(0)}$

- We consider the equality-constrained problem:

$$
\begin{aligned}
\min_{x_1,x_2\in\mathbb{R}} \{x_1 - x_2 | x_1 + x_2 = 1, -x_\ell = 0, \forall \ell \in \mathbb{W}^{(0)}\} \\
= \min_{x_1,x_2\in\mathbb{R}} \{x_1 - x_2 | x_1 + x_2 = 1, -x_2 = 0\},
\end{aligned}
\tag{16.11}
$$

- and seek a descent direction for it.
- In fact, however, $x^{(0)}$ is optimal for this problem, but the sign of the Lagrange multiplier for the constraint $-x_2 = 0$ is negative.
- That is, we are at the minimizer of the equality-constrained problem but have not found the minimizer of inequality-constrained Problem (16.4).

Update working set

- We update the working set by removing constraint 2 from it.
- That is, we now have the revised working set $\mathbb{W}^{(0)} = \emptyset$.

Descent direction at $x^{(0)}$

- Since the objective increases with x_1 and decreases with x_2, a descent direction at $x^{(0)}$ for the objective that maintains feasibility for the equality constraints $x_1 + x_2 = 1$ is given by $\Delta x^{(0)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

16.3.4 Alternation of swapping in and out

- We must solve a sequence of problems, alternately swapping in and out.
- We continue with Problem (16.4), starting at $x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and using descent direction $\Delta x^{(0)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

16.3.4.1 Swapping in

- If we move along the descent direction according to $x^{(0)} + \alpha^{(0)} \Delta x^{(0)}$, we find that for $\alpha^{(0)} = 1$, the constraint $x_1 \geq 0$ becomes binding.
- We obtain the next iterate:

$$\begin{aligned} x^{(1)} &= x^{(0)} + \alpha^{(0)} \Delta x^{(0)}, \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \\ &= \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \end{aligned}$$

- and we update the working set to $\mathbb{W}^{(1)} = \{1\}$.

16.3.4.2 *Descent direction*

- We consider the equality-constrained problem corresponding to the working set $\mathbb{W}^{(1)} = \{1\}$:

$$\begin{aligned} & \min_{x_1,x_2\in\mathbb{R}} \{x_1 - x_2 | x_1 + x_2 = 1, -x_\ell = 0, \forall \ell \in \mathbb{W}^{(1)}\} \\ = & \min_{x_1,x_2\in\mathbb{R}} \{x_1 - x_2 | x_1 + x_2 = 1, -x_1 = 0\}, \end{aligned} \tag{16.12}$$

- and seek a descent direction for it.
- In fact, $x^{(1)}$ is the minimizer of this equality-constrained problem and the sign of the Lagrange multiplier for the constraint $-x_1 = 0$ is positive.
- That is, we are at the optimum of the equality-constrained problem and have also found the optimum of the inequality-constrained Problem (16.4).

16.3.4.3 Discussion

- Since there were only two inequality constraints we took just one swapping out operation and one swapping in operation to find the minimizer.
- In general we will find that we will have to successively swap in and out various of the constraints and solve several equality-constrained problems before reaching the minimizer of the original inequality-constrained problem.

16.3.5 *Finding an initial feasible guess*

- To find an initial feasible guess for Problem (16.1), we define another optimization problem that is *related* to Problem (16.1) and having the following properties:
  - it is easy to find an initial feasible guess for the related problem,
  - if Problem (16.1) is feasible, then a minimizer of the related problem yields a feasible initial guess for Problem (16.1), and
  - if Problem (16.1) is infeasible, then the minimum of the related problem signals this fact.
- The related problem includes the variables $x \in \mathbb{R}^n$ from Problem (16.1) and, additionally, includes **artificial variables** $w \in \mathbb{R}^m$.

Finding an initial feasible guess, continued

- Suppose that $b \geq \mathbf{0}$ (or, swap the sign of any negative entry in b and the signs of the entries in the corresponding row of A.)
- Consider the following problem, related to Problem (16.1):

$$\min_{x \in \mathbb{R}^n, w \in \mathbb{R}^m} \{\mathbf{1}^\dagger w | Ax + w = b, x \geq \mathbf{0}, w \geq \mathbf{0}\}. \tag{16.13}$$

- Note that $x^{(0)} = \mathbf{0}, w^{(0)} = b \geq \mathbf{0}$ satisfies the equality and inequality constraints of Problem (16.13) and is therefore a feasible initial guess for this problem that can be used by an active set method.
- We solve this problem using the active set method and this feasible initial guess.

Finding an initial feasible guess, continued

- Suppose that $\begin{bmatrix} x^\star \\ w^\star \end{bmatrix}$ is a minimizer of Problem (16.13) with $w^\star = \mathbf{0}$.
- Then the minimum of Problem (16.13) is $\mathbf{1}^\dagger w^\star = 0$ and $x^\star$ is a feasible initial guess for Problem (16.1), since:

$$\begin{aligned} b &= Ax^\star + w^\star, \text{ since } \begin{bmatrix} x^\star \\ w^\star \end{bmatrix} \text{ is feasible for Problem (16.13)}, \\ &= Ax^\star, \text{ since } w^\star = \mathbf{0}, \\ x^\star &\geq \mathbf{0}, \text{ since } \begin{bmatrix} x^\star \\ w^\star \end{bmatrix} \text{ is feasible for Problem (16.13)}. \end{aligned}$$

- If the minimum is non-zero (so that the minimizer $\begin{bmatrix} x^\star \\ w^\star \end{bmatrix}$ satisfies $w^\star \neq \mathbf{0}$) then Problem (16.1) is infeasible.
- The process of finding a feasible initial guess for Problem (16.1) is called **phase 1** of optimization.
- The feasible initial guess is then used as a starting point by an algorithm to minimize the objective of Problem (16.1) in **phase 2**.

16.3.6 Linear and quadratic objectives

16.3.6.1 Linear programming

Analysis

- Consider a non-negatively constrained linear programming problem:

$$\min_{x\in\mathbb{R}^n}\{c^\dagger x | Ax = b, x \geq \mathbf{0}\}. \tag{16.14}$$

- Except for:
  - the complementary slackness conditions $Mx = \mathbf{0}$, and
  - the inequalities $x \geq \mathbf{0}$ and $\mu \geq \mathbf{0}$,
- the necessary conditions are *linear* simultaneous equations.
- The linearity facilitates:
  - the calculation of descent directions for the corresponding equality-constrained problem,
  - avoiding zig-zagging, and
  - maintaining feasibility as successive iterates are calculated.

Analysis, continued

- The linear minimization Problem (16.14) is equivalent to maximizing the objective $-c^\dagger x$ over the same feasible set.
- By Theorem 2.5, there is a maximizer of $-c^\dagger x$ (and therefore a minimizer of $c^\dagger x$) that is an extreme point of the feasible set.
- We can restrict attention to points that are vertices of the feasible set and do not need to consider points such as $x^{(\nu+2)}$ in Figure 16.5 that are on the boundary but not at a vertex of the feasible set.
- Geometrically, contour sets of the objective are parallel hyperplanes.
- The minimum of the linear program corresponds to the hyperplane with minimum height that intersects the feasible set.
- The intersection will contain a vertex of the feasible set.

Discussion

- The active set strategy applied to linear programming problems represented in the form of Problem (16.14), together with various techniques to make the constraint swapping and calculation of descent directions more efficient, leads to the **simplex algorithm**.
- The simplex algorithm was developed in the 1940s by George Dantzig.
- The vertices of the feasible set of Problem (16.14) are points that satisfy equations of the form:

$$Ax = b, -x_\ell = 0, \forall \ell \in \mathbb{W},$$

- with $\mathbb{W}$ having $n - m$ members (for $A \in \mathbb{R}^{m \times n}$ having m linearly independent rows.)

Discussion, continued

- For example, for the feasible set illustrated in Figure 16.5, the vertices are:

$$\begin{bmatrix} 10 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 10 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 10 \end{bmatrix},$$

- corresponding, respectively, to the three choices:

$$\mathbb{W} = \{2,3\}, \mathbb{W} = \{1,3\}, \mathbb{W} = \{1,2\}.$$

- Each of these choices of working set has $n - m = 3 - 1 = 2$ members.
- The form of the feasible set leads to important simplifications for updating iterates and swapping in and swapping out.
- Swapping in and out is performed simultaneously and calculation of a descent direction is facilitated by maintaining and updating factors of an appropriate square sub-matrix of the coefficient matrix of the constraints $Ax = b, -x_\ell = 0, \forall \ell \in \mathbb{W}$.
- The MATLAB function `linprog` uses the simplex algorithm under some circumstances.

Discussion, continued

- For some problems, the simplex algorithm must examine a large proportion of the possible combinations of active inequalities.
- However, the simplex algorithm usually finds a solution of the problem in relatively few iterations.
- The simplex algorithm and its variants remain the most used and practical optimization algorithms.
- If an optimization problem can be formulated as a linear program then it is worthwhile to do so.
- Many special issues arise in linear programming that allow simplifications of hypotheses and sharpening of conclusions of the theory we have discussed.
- For example, some linear integer optimization problems have simple solutions in terms of linear programming.

16.3.6.2 Quadratic programming

- As with linear programming, there are also simplifications possible in the case of quadratic objectives.
- Moreover, there is a large body of active set-based software available to solve quadratic programming problems.
- The MATLAB function `quadprog` uses an active set algorithm under some circumstances.

16.3.6.3 Further details

- We have only introduced active set algorithms briefly here; however, much software written for optimization problems uses some form of active set algorithm.

16.4 Approaches to finding minimizers: interior point algorithm

- A very different approach to solving inequality-constrained problems is not based on identifying the active constraints directly.
- Conceptually, a “barrier” is erected that prevents violation of all the inequality constraints so that the sequence of iterates remains *strictly* feasible with respect to the inequality constraints.
- The iterates remain in the interior of the set defined by the inequality constraints.
- Ideally, the iterates step directly towards the minimizer across the interior of the feasible region, rather than stepping along its boundary as in the active set algorithm.
- For this reason, the technique is called an **interior point algorithm**.

16.4.1 Illustration

- To illustrate the interior point algorithm, consider the objective $f : \mathbb{R} \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}, f(x) = x,$$

- and a non-negativity constraint $x \geq 0$.
- We add a **barrier function** for the constraints $x \geq \mathbf{0}$ to the objective $f(x)$ to form the **barrier objective**, $\phi : \mathbb{R}_{++} \to \mathbb{R}$.
- The essential characteristic of the barrier function is that it is partially differentiable on the interior of the feasible set but becomes unbounded as the boundary of the feasible set is approached.
- Define the logarithmic barrier function $f_{\mathrm{b}} : \mathbb{R}_{++} \to \mathbb{R}$ for the constraint $x \geq 0$ by:

$$\forall x \in \mathbb{R}_{++}, f_{\mathrm{b}}(x) = -\ln(x).$$

- Let $t \in \mathbb{R}_{++}$ be a parameter, called the **barrier parameter**.

Illustration, continued

- Define the barrier objective $\phi : \mathbb{R}_{++} \to \mathbb{R}$ by:

$$\begin{aligned} \forall x \in \mathbb{R}_{++}, \phi(x) &= f(x) + t f_{\mathrm{b}}(x), \\ &= f(x) - t \ln x. \end{aligned}$$

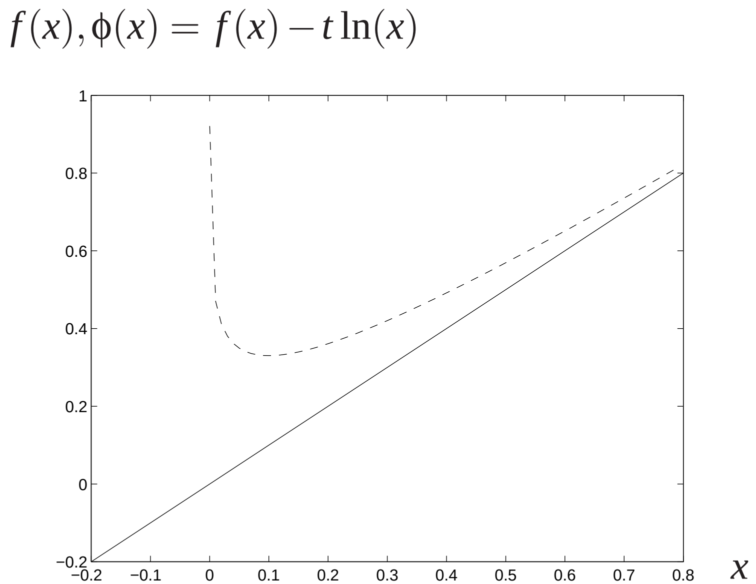

Fig. 16.7. Barrier objective for the constraint $x \geq 0, x \in \mathbb{R}$. The solid curve shows the objective f while the dashed curve shows the barrier objective ϕ for $t = 0.1$ on the interior of the feasible region.

Illustration, continued

- As $x \to 0^+$, $\phi(x) \to \infty$.
- An algorithm that is trying to minimize ϕ will avoid the vicinity of the boundary of the feasible region.
- That is, it will produce iterates that are interior to the set defined by the inequality constraint.

Illustration, continued

- For any fixed $x > 0$, the value of $-t\ln(x)$ approaches 0 as $t \to 0$.

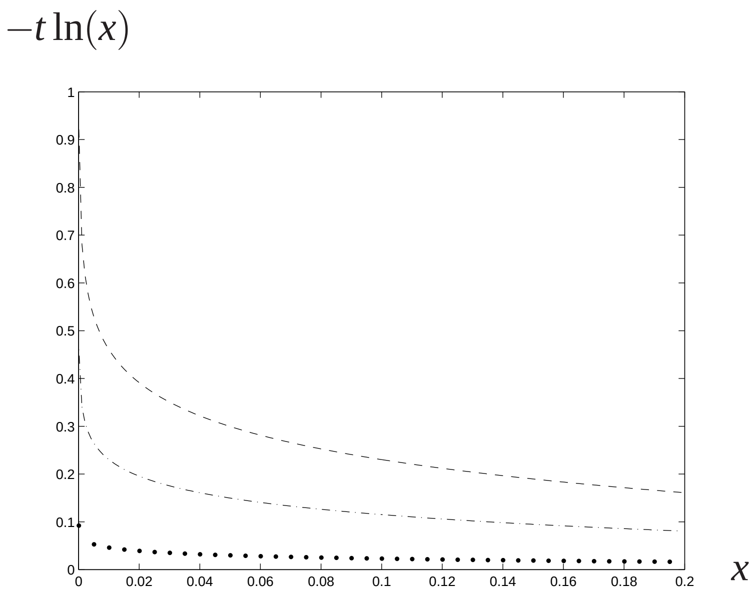

Fig. 16.8. Effect on barrier function for the constraint $x \geq 0$ as $t \to 0$. The dashed curve shows $-t\ln(x)$ for $t = 0.1$; the dash-dot curve shows $-t\ln(x)$ for $t = 0.05$; and the dotted curve shows $-t\ln(x)$ for $t = 0.01$.

16.4.2 Outline

16.4.2.1 Logarithmic barrier function

- We define the logarithmic barrier function $f_{\mathrm{b}} : \mathbb{R}^n_{++} \to \mathbb{R}$ for the constraints $x \geq \mathbf{0}$ by:

$$\forall x \in \mathbb{R}^n_{++}, f_{\mathrm{b}}(x) = -\sum_{\ell=1}^{n} \ln(x_\ell). \tag{16.15}$$

16.4.2.2 Barrier problem

- Given an objective $f : \mathbb{R}^n \to \mathbb{R}$, a barrier function $f_{\mathrm{b}} : \mathbb{R}^n_{++} \to \mathbb{R}$, and a barrier parameter $t \in \mathbb{R}_{++}$, we form the **barrier objective** $\phi : \mathbb{R}^n_{++} \to \mathbb{R}$ defined by:
$$\forall x \in \mathbb{R}^n_{++}, \phi(x) = f(x) + t f_{\mathrm{b}}(x).$$
- Instead of solving Problem (16.1), we will consider solving the **barrier problem**:
$$\min_{x \in \mathbb{R}^n} \{\phi(x) | Ax = b, x > \mathbf{0}\}. \tag{16.16}$$
- We discussed the potential disadvantages of an open feasible set such as $\{x \in \mathbb{R}^n | x > \mathbf{0}\}$ in Section 2.3.3.
- However, in practice, for suitable f, Problem (16.16) can be solved by a technique that considers only the *equality* constraints when seeking a descent direction.

16.4.2.3 Slater condition

- For Problem (16.16) to be useful in finding a solution of Problem (16.1), we need to assume that:

$$\{x \in \mathbb{R}^n | Ax = b, x > \mathbf{0}\} \neq \emptyset,$$

- so that Problem (16.16) has a non-empty feasible set.
- This is called the **Slater condition**.
- This condition requires the existence of a feasible point that is strictly feasible for the inequality constraints.
- That is, there must be a feasible interior point.
- Many constraint systems arising from physical systems satisfy the Slater condition.

Slater condition, continued

- A simple example of constraints that do not satisfy the Slater condition is defined by the following:

$$\begin{aligned} A &= \begin{bmatrix} 1 & 1 \end{bmatrix}, \\ b &= \begin{bmatrix} 0 \end{bmatrix}, \\ x &\geq \mathbf{0}. \end{aligned}$$

- The set $\{x \in \mathbb{R}^2 | Ax = b, x > \mathbf{0}\}$ is empty.

16.4.2.4 Solving the barrier problem

- To find the minimizer of Problem (16.16) for any particular value of t, we can start with an initial guess $x^{(0)}$ that satisfies $Ax = b$ and $x > \mathbf{0}$.
- We then search from $x^{(0)}$ using an iterative algorithm that seeks the value of x that minimizes $\phi(x)$ subject to $Ax = b$.
- Since the objective function ϕ of Problem (16.16) becomes arbitrarily large as its argument approaches the boundary of $x \geq \mathbf{0}$, we only need to prevent the iterates from going outside the region $x > \mathbf{0}$ by controlling the step-size appropriately.

16.4.2.5 Sequence of problems

- We solve Problem (16.16) not just at one value of t, but for a *sequence* of values of t that approach 0.
- The trajectory of minimizers of Problem (16.16) as a function of t is called the **central path**.

16.4.2.6 Example

- Consider again Problem (16.4), which we analyzed in Section 16.3.3.2:

$$\min_{x_1,x_2\in\mathbb{R}} \{x_1 - x_2 | x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0\}.$$

- The interior point algorithm involves solving the barrier problem, Problem (16.16), for a sequence of values of t that decrease towards zero.
- For Problem (16.4), the barrier problem is:

$$\min_{x_1,x_2\in\mathbb{R}} \{x_1 - x_2 - t\ln(x_1) - t\ln(x_2) | x_1 + x_2 = 1, x_1 > 0, x_2 > 0\}. \quad (16.17)$$

Example, continued

- We can calculate the minimizer of Problem (16.17) explicitly as a function of t.
- We can eliminate x_2 using the equality constraint to express the objective as a function of x_1:

$$2x_1 - 1 - t\ln(x_1) - t\ln(1 - x_1). \tag{16.18}$$

- We now have an unconstrained problem:

$$\min_{x_1 \in \mathbb{R}} \{2x_1 - 1 - t\ln(x_1) - t\ln(1 - x_1)\}.$$

- Differentiating (16.18), setting the derivative equal to zero, and re-arranging we find that:

$$(x_1)^2 - x_1(1 + t) + t/2 = 0, \tag{16.19}$$

- where we note that both x_1 and $x_2 = 1 - x_1$ must be greater than zero for the objective and derivative to be defined (and for the inequality constraints to be strictly satisfied.)

Example, continued

- The quadratic equation (16.19) has two solutions, both of which are positive.
- However, only one of the solutions:

$$x_1 = \frac{1+t-\sqrt{1+(t)^2}}{2}, \tag{16.20}$$

  yields a value of $x_2 = 1 - x_1$ that satisfies the strict non-negativity constraint for x_2.
- Substituting, we obtain:

$$x_2 = \frac{1-t+\sqrt{1+(t)^2}}{2}. \tag{16.21}$$

- In general, we may not be able to conveniently eliminate variables and solve for the minimizer of the barrier problem explicitly as a function of t as we have done for Problem (16.17).
- Nevertheless, we can think, in principle, of solving the barrier problem for a sequence of decreasing values of t.

Example, continued

- Figure 16.9 shows the minimizer given in (16.20) and (16.21) of Problem (16.17) versus t for $t = 1.0, 0.9, \ldots, 0.1$.
- The minimizers are always in the interior of the set $\{x \in \mathbb{R}^n | x \geq \mathbf{0}\}$.

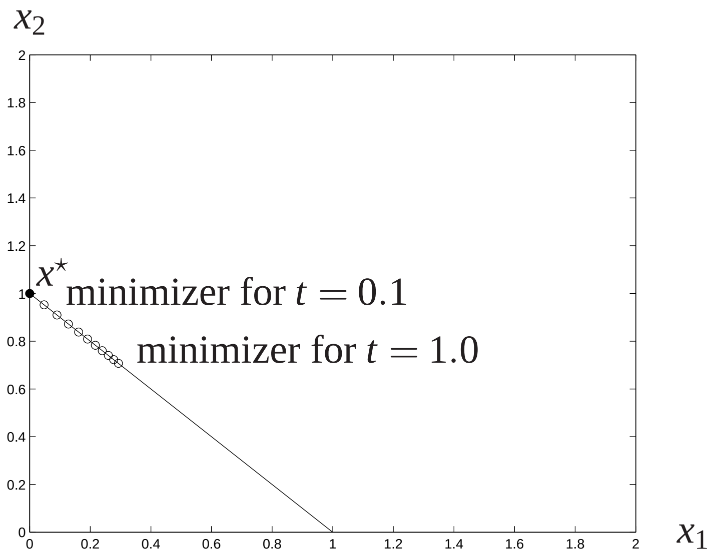

Fig. 16.9. The trajectory of the minimizers of Problem (16.17) versus t for $t = 1.0, 0.9, \ldots, 0.1$ shown as $\circ$. The minimizer $x^\star$ of Problem (16.4) is shown as a $\bullet$. The feasible set is indicated by the solid line.

Example, continued

- For large values of t, the minimizer of Problem (16.17) is far away from the minimizer $x^\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ of the inequality-constrained Problem (16.4).
- However, as t decreases towards zero, the minimizer of Problem (16.17) approaches $x^\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
- We will explicitly discuss a stopping criterion in Section 16.4.6.4.

16.4.2.7 *Reduction of barrier parameter*

- Because we evaluated the minimizer explicitly as a function of t, we could just pick $t = 10^{-10}$, say, and evaluate (16.20)–(16.21) to obtain:

$$x^\star \approx \begin{bmatrix} 5 \times 10^{-11} \\ 1.0000 \end{bmatrix}.$$

- In general, we cannot solve for the minimizer of Problem (16.16) explicitly and we will have to use an iterative algorithm.
- It is very difficult to solve Problem (16.16) from scratch for a small value of t because the initial guess that we can provide for the iterative algorithm leads to a poor update in seeking an unconstrained minimizer.
- Instead of trying to minimize the barrier problem from scratch for a small value of t, we start with a large value of t and use the Newton–Raphson update to seek a minimizer for this value of t.

16.4.3 Newton–Raphson update

16.4.3.1 Discussion of the barrier problem

- We seek a minimizer of the problem:

$$\min_{x\in\mathbb{R}^n}\{\phi(x)|Ax=b\}. \tag{16.22}$$

- By Theorem 13.2, the first-order necessary conditions of Problem (16.22) are:

$$\nabla\phi(x)+A^{\dagger}\lambda = 0, \tag{16.23}$$
$$Ax-b = 0. \tag{16.24}$$

16.4.3.2 Primal interior point algorithm

- We first investigate a straightforward approach to applying the Newton–Raphson update to solving the first-order necessary conditions (16.23)–(16.24).

Primal interior point algorithm, continued

- Consider the first term in (16.23):

$$
\begin{aligned}
\nabla\phi(x) &= \nabla[f(x)+t f_{\mathrm{b}}(x)], \\
&= \nabla f(x)+t\nabla f_{\mathrm{b}}(x), \\
&= \nabla f(x)+t \begin{bmatrix} \dfrac{\partial f_{\mathrm{b}}(x)}{\partial x_1} \\ \vdots \\ \dfrac{\partial f_{\mathrm{b}}(x)}{\partial x_n} \end{bmatrix}, \\
&= \nabla f(x)+t \begin{bmatrix} -\dfrac{1}{x_1} \\ \vdots \\ -\dfrac{1}{x_n} \end{bmatrix}, \\
&= \nabla f(x)-t[X]^{-1}\mathbf{1}, \\
\nabla^2\phi(x) &= \nabla^2 f(x)+t[X]^{-2}.
\end{aligned}
$$

- The Newton–Raphson update to solve (16.23)–(16.24) is given by:

$$\begin{bmatrix} \nabla^2\phi(x^{(\nu)}) & A^\dagger \\ A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta x^{(\nu)} \\ \Delta\lambda^{(\nu)} \end{bmatrix} = \begin{bmatrix} -\nabla\phi(x^{(\nu)}) - A^\dagger\lambda^{(\nu)} \\ b - Ax^{(\nu)} \end{bmatrix},$$

- or:

$$\begin{bmatrix} \nabla^2 f(x^{(\nu)}) + t[X^{(\nu)}]^{-2} & A^\dagger \\ A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta x^{(\nu)} \\ \Delta\lambda^{(\nu)} \end{bmatrix} = \begin{bmatrix} -\nabla f(x^{(\nu)}) + t[X^{(\nu)}]^{-1}\mathbf{1} - A^\dagger\lambda^{(\nu)} \\ b - Ax^{(\nu)} \end{bmatrix}. \tag{16.25}$$

- This update leads to the **primal interior point algorithm**.
- We are not going to investigate this algorithm further, except in Section 18.2.1 in the discussion of enforcement of the strict inequality constraints in the case study of optimal routing in a data communication network.
- Instead of discussing the primal interior point method, we will consider a variant in the next section.

16.4.3.3 Primal–dual interior point algorithm

- Instead of the primal interior point algorithm, we will describe an algorithm that incorporates linearization of a hyperbolic approximation to the complementary slackness constraints, as first introduced in Section 16.1.1.3.

New variable and equation

- We are going to introduce a new variable μ, which will turn out to correspond to the dual variables for the inequality constraints in Problem (16.1).
- We incorporate the equations:

$$\forall \ell = 1, \ldots, n, \mu_\ell x_\ell = t. \tag{16.26}$$

- The approximation in (16.26) allows $\begin{bmatrix} \mu_\ell \\ x_\ell \end{bmatrix}$ to lie on a hyperbolic-shaped set as shown in Figure 16.2.
- Linearization of (16.26), together with an explicit requirement to avoid the x_ℓ- and μ_ℓ-axes, yields a useful update that can approximately represent the kink in the complementary slackness conditions.
- We have remarked that we will solve Problem (16.16) for a sequence of decreasing values of t.
- As $t \to 0$, points that satisfy (16.26) will approach satisfaction of the complementary slackness conditions:

$$Mx = \mathbf{0}.$$

New variable and equation, continued

- We can re-write (16.26) as:

$$X\mu - t\mathbf{1} = \mathbf{0}, \tag{16.27}$$

- which we can re-arrange as $\mu = t[X]^{-1}\mathbf{1}$.
- Recall that $\nabla\phi : \mathbb{R}^n_{++} \to \mathbb{R}^n$ is defined by:

$$\forall x \in \mathbb{R}^n_{++}, \nabla\phi(x) = \nabla f(x) - t[X]^{-1}\mathbf{1}.$$

- Substituting the expression for $\nabla\phi$ into (16.23) and making the substitution $\mu = t[X]^{-1}\mathbf{1}$, we obtain:

$$\begin{aligned} \nabla f(x) + A^\dagger\lambda - \mu &= \mathbf{0}. && (16.28) \\ Ax &= b. && (16.29) \end{aligned}$$

- Equations (16.27)–(16.29) are equivalent to (16.23)–(16.24) in that:
  - a solution of (16.23)–(16.24) satisfies (16.28)–(16.29), given that μ is defined by (16.27), and
  - a solution of (16.27)–(16.29) satisfies (16.23)–(16.24).

New variable and equation, continued

- The hyperbolic approximation to the complementary slackness conditions together with (16.28) and (16.29) are equivalent to the first-order necessary conditions for minimizing Problem (16.22), ignoring the strict inequality constraints.
- Moreover, (16.28) and (16.29) are two of the lines of the first-order necessary conditions for Problem (16.1).
- The condition (16.27) becomes more nearly equivalent to the complementary slackness conditions for Problem (16.1) as $t \to 0$.
- Instead of seeking $x^\star$ and $\lambda^\star$ that satisfy (16.23)–(16.24), we will seek $x^\star$, $\lambda^\star$, and $\mu^\star$ that satisfy (16.27)–(16.29).

Step direction

- The Newton–Raphson step direction to solve (16.27)–(16.29) is given by:

$$\begin{bmatrix} X^{(\nu)} & M^{(\nu)} & \mathbf{0} \\ -\mathbf{I} & \nabla^2 f(x^{(\nu)}) & A^\dagger \\ \mathbf{0} & A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mu^{(\nu)} \\ \Delta x^{(\nu)} \\ \Delta\lambda^{(\nu)} \end{bmatrix} = \begin{bmatrix} -X^{(\nu)}\mu^{(\nu)} + t\mathbf{1} \\ -\nabla f(x^{(\nu)}) - A^\dagger\lambda^{(\nu)} + \mu^{(\nu)} \\ -Ax^{(\nu)} + b \end{bmatrix},$$

- where $M^{(\nu)} = \text{diag}\{\mu_\ell^{(\nu)}\}$ and $X^{(\nu)} = \text{diag}\{x_\ell^{(\nu)}\}$.

Symmetry

- The Newton–Raphson update equations have a coefficient matrix that is not symmetric.
- By multiplying the first block row of the equations through by $-[M^{(\nu)}]^{-1}$ on the left, we can create the symmetric system:

$$\begin{bmatrix} -[M^{(\nu)}]^{-1}X^{(\nu)} & -\mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \nabla^2 f(x^{(\nu)}) & A^\dagger \\ \mathbf{0} & A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mu^{(\nu)} \\ \Delta x^{(\nu)} \\ \Delta\lambda^{(\nu)} \end{bmatrix} = \begin{bmatrix} x^{(\nu)} - t[M^{(\nu)}]^{-1}\mathbf{1} \\ -\nabla f(x^{(\nu)}) - A^\dagger\lambda^{(\nu)} + \mu^{(\nu)} \\ -Ax^{(\nu)} + b \end{bmatrix}. \tag{16.30}$$

- This system is symmetric, but indefinite.
- In general, to factorize it we must make use of the special factorization algorithms for indefinite matrices as mentioned in Section 5.4.7.

Block pivoting of Jacobian and sparsity issues

- Unfortunately, the top left-hand block of the coefficient matrix of this system may have entries that are very large and entries that are very small, depending on whether or not the corresponding constraint $x_\ell \geq 0$ is binding.
- This means that the coefficient matrix can be ill-conditioned.
- We can deal analytically with the entries in the top left-hand block of the coefficient matrix because of its simple structure.
- We will do this by block factorizing the Jacobian using the diagonal matrix $-[M^{(\nu)}]^{-1}X^{(\nu)}$ as block pivot, noting that we can explicitly invert $-[M^{(\nu)}]^{-1}X^{(\nu)}$ to obtain $-[X^{(\nu)}]^{-1}M^{(\nu)}$.

Block pivoting of Jacobian and sparsity issues, continued

- We obtain:

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ -[X^{(\nu)}]^{-1}M^{(\nu)} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} -[M^{(\nu)}]^{-1}X^{(\nu)} & -\mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \nabla^2 f(x^{(\nu)}) & A^\dagger \\ \mathbf{0} & A & \mathbf{0} \end{bmatrix}$$

$$= \begin{bmatrix} -[M^{(\nu)}]^{-1}X^{(\nu)} & -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \nabla^2 f(x^{(\nu)}) + [X^{(\nu)}]^{-1}M^{(\nu)} & A^\dagger \\ \mathbf{0} & A & \mathbf{0} \end{bmatrix}. \quad (16.31)$$

Selection of step-size

- If we set:

$$\begin{bmatrix} \mu^{(\nu+1)} \\ x^{(\nu+1)} \\ \lambda^{(\nu+1)} \end{bmatrix} = \begin{bmatrix} \mu^{(\nu)} \\ x^{(\nu)} \\ \lambda^{(\nu)} \end{bmatrix} + \begin{bmatrix} \Delta\mu^{(\nu)} \\ \Delta x^{(\nu)} \\ \Delta\lambda^{(\nu)} \end{bmatrix},$$

- we may violate the non-negativity constraints on μ or x.
- To avoid this we may have to take a step that is less than the full step-size of one:

$$\begin{bmatrix} \mu^{(\nu+1)} \\ x^{(\nu+1)} \\ \lambda^{(\nu+1)} \end{bmatrix} = \begin{bmatrix} \mu^{(\nu)} \\ x^{(\nu)} \\ \lambda^{(\nu)} \end{bmatrix} + \alpha^{(\nu)} \begin{bmatrix} \Delta\mu^{(\nu)} \\ \Delta x^{(\nu)} \\ \Delta\lambda^{(\nu)} \end{bmatrix},$$

- The strict non-negativity constraints are somewhat problematic.
- For example, suppose that we implement the requirement of strict non-negativity by choosing a tolerance $\varepsilon > 0$ and requiring that the next iterate satisfies $x_\ell^{(\nu+1)} \geq \varepsilon, \forall \ell$ and $\mu_\ell^{(\nu+1)} \geq \varepsilon, \forall \ell$.

Selection of step-size, continued

- A serious drawback of this approach is that *a priori* we do not know how close the minimizer of Problem (16.22) is to the boundary.

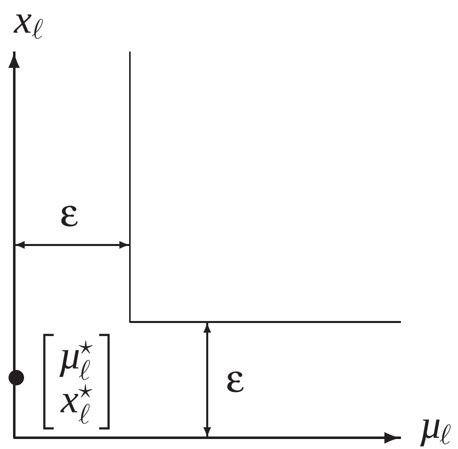

Fig. 16.10. Using a fixed tolerance to enforce non-negativity will prevent convergence to a minimizer.

Selection of step-size, continued

- We must adjust the tolerance so that iterates can, asymptotically, approach the boundary.
- One scheme is to pick $\alpha^{(\nu)} \leq 1$ at each iteration so that $\begin{bmatrix} \mu^{(\nu+1)} \\ x^{(\nu+1)} \end{bmatrix}$ is no closer than a fixed *fraction*, say 0.9995, of the distance from the current iterate $\begin{bmatrix} \mu^{(\nu)} \\ x^{(\nu)} \end{bmatrix}$ to the boundary of $x \geq \mathbf{0}, \mu \geq \mathbf{0}$ under the L_∞ norm.
- With this choice, $\mu^{(\nu)}$ and $x^{(\nu)}$ *can* approach any point that satisfies the complementary slackness condition.
- There are many variations on the choice of step-size.

Selection of step-size, continued

- It is also possible to use a different step-size for:
  - the *primal* variables x, and
  - the *dual* variables μ and λ.
- That is, we can update according to:

$$\begin{aligned} x^{(\nu+1)} &= x^{(\nu)} + \alpha^{(\nu)}_{\text{primal}} \Delta x^{(\nu)}, \\ \begin{bmatrix} \mu^{(\nu+1)} \\ \lambda^{(\nu+1)} \end{bmatrix} &= \begin{bmatrix} \mu^{(\nu)} \\ \lambda^{(\nu)} \end{bmatrix} + \alpha^{(\nu)}_{\text{dual}} \begin{bmatrix} \Delta\mu^{(\nu)} \\ \Delta\lambda^{(\nu)} \end{bmatrix}, \end{aligned}$$

- where:
  $\alpha^{(\nu)}_{\text{primal}}$ is chosen to preserve the strict non-negativity of x, and
  $\alpha^{(\nu)}_{\text{dual}}$ is chosen to preserve the strict non-negativity of μ.
- However, we will not take advantage of this flexibility.

16.4.3.4 Example

- Let us apply the primal–dual interior point algorithm to our example Problem (16.4).

Terms in update

$$\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= x_1 - x_2, \\
\forall x \in \mathbb{R}^2, \nabla f(x) &= \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \\
\forall x \in \mathbb{R}^2, \nabla^2 f(x) &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \\
A &= \begin{bmatrix} 1 & 1 \end{bmatrix}, \\
&= \mathbf{1}^\dagger, \\
b &= [1].
\end{aligned}$$

Factorization

$$
\begin{aligned}
\mathcal{A} &= \begin{bmatrix} -[M^{(\nu)}]^{-1}X^{(\nu)} & -\mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \nabla^2 f(x) & A^\dagger \\ \mathbf{0} & A & \mathbf{0} \end{bmatrix}, \\
&= \begin{bmatrix} -[M^{(\nu)}]^{-1}X^{(\nu)} & -\mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1}^\dagger & \mathbf{0} \end{bmatrix}.
\end{aligned}
$$

- This matrix is indefinite and, in general, we should use a special purpose factorization algorithm.
- Here, we will simply apply LU factorization, using the symbols $\mathcal{A}^{(j)}$ and $\mathcal{M}^{(j)}$ for the matrices created at the j-th stage of factorization.
- Note that $M^{(\nu)} = \mathrm{diag}\{\mu_\ell^{(\nu)}\}$.

Factorization, continued

- Block pivoting of $\mathcal{A}$ using its top-left block $-[M^{(\nu)}]^{-1}X^{(\nu)}$ as pivot yields $\mathcal{M}^{(1)}$ and $\mathcal{A}^{(1)}$ given by:

$$\begin{aligned}
\mathcal{M}^{(1)} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ -[X^{(\nu)}]^{-1}M^{(\nu)} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}, \\
\mathcal{A}^{(1)} &= \begin{bmatrix} -[M^{(\nu)}]^{-1}X^{(\nu)} & -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & [X^{(\nu)}]^{-1}M^{(\nu)} & \mathbf{1} \\ \mathbf{0} & \mathbf{1}^{\dagger} & \mathbf{0} \end{bmatrix}.
\end{aligned}$$

Factorization, continued

$$
\begin{aligned}
\mathcal{M}^{(2)} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1}^{\dagger}\left[M^{(\nu)}\right]^{-1} X^{(\nu)} & \mathbf{I} \end{bmatrix}, \\
&= \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\left[x^{(\nu)}\right]^{\dagger}\left[M^{(\nu)}\right]^{-1} & \mathbf{I} \end{bmatrix}, \\
\mathcal{A}^{(2)} &= \mathcal{M}^{(2)} \mathcal{A}^{(1)}, \\
&= \begin{bmatrix} -\left[M^{(\nu)}\right]^{-1} X^{(\nu)} & -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \left[X^{(\nu)}\right]^{-1} M^{(\nu)} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & -\mathbf{1}^{\dagger}\left[M^{(\nu)}\right]^{-1} X^{(\nu)} \mathbf{1} \end{bmatrix}, \\
&= \begin{bmatrix} -\left[M^{(\nu)}\right]^{-1} X^{(\nu)} & -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \left[X^{(\nu)}\right]^{-1} M^{(\nu)} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & -\left[\mu_1^{(\nu)}\right]^{-1} x_1^{(\nu)} - \left[\mu_2^{(\nu)}\right]^{-1} x_2^{(\nu)} \end{bmatrix},
\end{aligned}
$$

Factorization, continued

- so that we can factorize $\mathcal{A}$ into:

$$\mathcal{L} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \left[X^{(\nu)}\right]^{-1} M^{(\nu)} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \left[x^{(\nu)}\right]^{\dagger} \left[M^{(\nu)}\right]^{-1} & \mathbf{1}^{\dagger} \end{bmatrix},$$

$$\mathcal{U} = \begin{bmatrix} -\left[M^{(\nu)}\right]^{-1} X^{(\nu)} & -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \left[X^{(\nu)}\right]^{-1} M^{(\nu)} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & -\left[\mu_1^{(\nu)}\right]^{-1} x_1^{(\nu)} - \left[\mu_2^{(\nu)}\right]^{-1} x_2^{(\nu)} \end{bmatrix}.$$

Initial guess

- As an initial guess, we pick:

$$\begin{aligned}
x_1^{(0)} &= 0.5,\\
x_2^{(0)} &= 0.5,\\
\lambda^{(0)} &= 2,\\
t^{(0)} &= 0.25,\\
\mu_1^{(0)} &= t^{(0)}/x_1^{(0)} = 0.25/0.5 = 0.5,\\
\mu_2^{(0)} &= t^{(0)}/x_2^{(0)} = 0.25/0.5 = 0.5.
\end{aligned}$$

- The value of $t^{(0)}$ is large enough to yield a useful update direction for the initial guess $x^{(0)}, \lambda^{(0)}$, and $\mu^{(0)}$.
- We chose $x^{(0)}$ to satisfy $Ax^{(0)} = b$.
- However, $x^{(0)} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$ is in the "middle" of the region $Ax = b, x \geq \mathbf{0}$ and is not close to the minimizer of Problem (16.4).

Step direction

$$
\begin{aligned}
\mathcal{B} &= \begin{bmatrix} x^{(0)} - t\left[M^{(0)}\right]^{-1}\mathbf{1} \\ -\nabla f(x^{(0)}) - A^{\dagger}\lambda^{(0)} + \mu^{(0)} \\ -Ax^{(0)} + b \end{bmatrix}, \\
&= \begin{bmatrix} x_1^{(0)} - t^{(0)}\left[\mu_1^{(0)}\right]^{-1} \\ x_2^{(0)} - t^{(0)}\left[\mu_2^{(0)}\right]^{-1} \\ -1 - \lambda^{(0)} + \mu_1^{(0)} \\ 1 - \lambda^{(0)} + \mu_2^{(0)} \\ -x_1^{(0)} - x_2^{(0)} + 1 \end{bmatrix}, \\
&= \begin{bmatrix} 0 \\ 0 \\ -2.5 \\ -0.5 \\ 0 \end{bmatrix}.
\end{aligned}
$$

Step direction, continued

$$\mathcal{A}\begin{bmatrix}\Delta\mu_1^{(0)}\\ \Delta\mu_2^{(0)}\\ \Delta x_1^{(0)}\\ \Delta x_2^{(0)}\\ \Delta\lambda^{(0)}\end{bmatrix} = \mathcal{B},$$

$$\mathcal{Y} = \begin{bmatrix}0\\ 0\\ -2.5\\ -0.5\\ 3\end{bmatrix}.$$

Step direction, continued

$$\begin{bmatrix} \Delta\mu_1^{(0)} \\ \Delta\mu_2^{(0)} \\ \Delta x_1^{(0)} \\ \Delta x_2^{(0)} \\ \Delta\lambda^{(0)} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \\ -1.5 \end{bmatrix}.$$

First iterate

- If we set:

$$\begin{bmatrix} \mu^{(1)} \\ x^{(1)} \\ \lambda^{(1)} \end{bmatrix} = \begin{bmatrix} \mu^{(0)} \\ x^{(0)} \\ \lambda^{(0)} \end{bmatrix} + \begin{bmatrix} \Delta\mu^{(0)} \\ \Delta x^{(0)} \\ \Delta\lambda^{(0)} \end{bmatrix}, \tag{16.32}$$

- we will obtain:

$$\begin{bmatrix} \mu_1^{(1)} \\ \mu_2^{(1)} \\ x_1^{(1)} \\ x_2^{(1)} \\ \lambda^{(1)} \end{bmatrix} = \begin{bmatrix} 1.5 \\ -0.5 \\ -0.5 \\ 1.5 \\ 0.5 \end{bmatrix},$$

- which will not satisfy the non-negativity constraints on x or μ.

First iterate, continued

- Instead, we will update according to:

$$\begin{bmatrix} \mu^{(1)} \\ x^{(1)} \\ \lambda^{(1)} \end{bmatrix} = \begin{bmatrix} \mu^{(0)} \\ x^{(0)} \\ \lambda^{(0)} \end{bmatrix} + \alpha^{(0)} \begin{bmatrix} \Delta\mu^{(0)} \\ \Delta x^{(0)} \\ \Delta\lambda^{(0)} \end{bmatrix},$$

- where $0 < \alpha^{(0)} < 1$ is chosen to prevent the iterates from going outside $\mu > \mathbf{0}, x > \mathbf{0}$.
- For the initial guess $\begin{bmatrix} \mu^{(0)} \\ x^{(0)} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$, the boundary is 0.5 unit away in the L_∞ norm.

First iterate, continued

- Using the rule suggested in Section 16.4.3, we pick $\alpha^{(0)} \leq 1$ to come no closer than $(0.9995) \times 0.5$ units of the distance towards the boundary.
- We choose the largest $\alpha^{(0)}$ such that:

$$\alpha^{(0)} \begin{bmatrix} \Delta\mu_1^{(0)} \\ \Delta\mu_2^{(0)} \\ \Delta x_1^{(0)} \\ \Delta x_2^{(0)} \end{bmatrix} \geq -0.9995 \begin{bmatrix} \mu_1^{(0)} \\ \mu_2^{(0)} \\ x_1^{(0)} \\ x_2^{(0)} \end{bmatrix},$$

- which yields $\alpha^{(0)} = 0.49975$ and:

$$\begin{bmatrix} \mu_1^{(1)} \\ \mu_2^{(1)} \\ x_1^{(1)} \\ x_2^{(1)} \\ \lambda^{(1)} \end{bmatrix} = \begin{bmatrix} 0.99975 \\ 0.00025 \\ 0.00025 \\ 0.99975 \\ 1.250375 \end{bmatrix}.$$

16.4.4 Adjustment of the barrier parameter

16.4.4.1 Sequence of equality-constrained problems

- In principle, we could continue iterating with a fixed value $t = t^{(0)}$ until we approach a minimizer $x^{(0)\star}$ of equality-constrained Problem (16.22).
- We could then use $x^{(0)\star}$ as the starting point for the Newton–Raphson method for Problem (16.22) for a smaller value of t.
- That is, we would be accurately solving a sequence of equality-constrained problems for points that are on the central path.
- However, we want to reduce t as quickly as possible so that the iterates converge quickly to a minimizer of the inequality-constrained Problem (16.1).

16.4.4.2 Reduction of barrier parameter at every iteration

- The minimizer of Problem (16.1) can typically be approached more quickly by reducing t after *every* Newton–Raphson update.
- For Problem (16.4), we started far from its minimizer with an initial guess of $x^{(0)} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$ and used a relatively large value of $t = t^{(0)} = 0.25$.
- Nevertheless, $x^{(1)}$ is actually very close to the minimizer of inequality-constrained Problem (16.4).
- That is, $x^{(1)}$ can be thought of as being close to a minimizer of Problem (16.17) for a much smaller value of t than $t^{(0)}$.

16.4.4.3 Effective value of barrier parameter

- We would like a measure of how close the current iterate is to a minimizer of the original inequality-constrained problem and adjust t accordingly.
- Instead of interpreting $x^{(1)}$ as an *approximate* minimizer of Problem (16.22) for $t = t^{(0)}$, we will see if we can interpret $x^{(1)}$ as an *exact* (or nearly exact) minimizer of Problem (16.22) for some other, hopefully smaller, value of t.
- We think of this value of t as the effective value $t^{(1)}_{\text{effective}}$ for which $x^{(1)}$ is nearly the minimizer of Problem (16.22).
- We will then pick $t^{(1)} < t^{(1)}_{\text{effective}}$ for the value of t to apply in the next Newton–Raphson update to calculate $x^{(2)}$.
- By continuing in this way we will construct a sequence $\{t^{(\nu)}_{\text{effective}}\}_{\nu=0}^{\infty}$ and corresponding (approximate) minimizers $x^{(\nu)}$ of Problem (16.22) for $t = t^{(\nu)}_{\text{effective}}$.

Effective value of barrier parameter, continued

- If the sequence $\{t_{\text{effective}}^{(\nu)}\}_{\nu=0}^{\infty}$ converges to 0 then we have achieved our goal of a sequence of minimizers of Problem (16.22) with $t \to 0$.
- We will have avoided the effort of performing many iterations at each value of the barrier parameter t to solve Problem (16.22).
- To interpret the iterates as approximate minimizers of Problem (16.16) for a value of barrier parameter $t = t_{\text{effective}}^{(\nu)}$, recall that we have been trying to solve (16.27)–(16.29).
- We are going to interpret $\begin{bmatrix} \mu^{(1)} \\ x^{(1)} \\ \lambda^{(1)} \end{bmatrix}$ together with a value $t_{\text{effective}}^{(1)}$ as nearly satisfying (16.27)–(16.29).

Effective value of barrier parameter, continued

- We will assume that (16.28) and (16.29) are very nearly satisfied by $\mu^{(1)}$ and $x^{(1)}$.
- Let:

$$t^{(1)}_{\text{effective}} = \frac{\left[x^{(1)}\right]^{\dagger} \mu^{(1)}}{n}, \tag{16.33}$$

- where n is the length of x, so that that $t^{(1)}_{\text{effective}}$ is the average value of $x^{(1)}_{\ell} \mu^{(1)}_{\ell}$.
- If the values of $x^{(1)}_{\ell} \mu^{(1)}_{\ell}$ do not vary too much with ℓ, then:

$$X^{(1)} \mu^{(1)} - t^{(1)}_{\text{effective}} \mathbf{1} \approx \mathbf{0}.$$

- That is, $x^{(1)}$ and $\mu^{(1)}$ satisfy (16.27) approximately for $t = t^{(1)}_{\text{effective}}$.

16.4.4.4 Update of barrier parameter

- We now set:
$$t^{(1)} < t^{(1)}_{\text{effective}}.$$
- For example, we could choose:
$$\begin{aligned} t^{(1)} &= \frac{t^{(1)}_{\text{effective}}}{n}, \\ &= \frac{\left[x^{(1)}\right]^{\dagger} \mu^{(1)}}{n^2}. \end{aligned}$$
- For large n, this reduces t significantly at each step.
- We now must solve (or approximately solve) the barrier problem for the updated value $t = t^{(1)}$.
- As initial guess for the minimizer of the barrier problem for $t = t^{(1)}$ we can use $\mu^{(1)}, x^{(1)}, \lambda^{(1)}$.

- We calculate the Newton–Raphson step direction $\begin{bmatrix} \Delta\mu^{(1)} \\ \Delta x^{(1)} \\ \Delta\lambda^{(1)} \end{bmatrix}$, and update according to:

$$\begin{bmatrix} \mu^{(2)} \\ x^{(2)} \\ \lambda^{(2)} \end{bmatrix} = \begin{bmatrix} \mu^{(1)} \\ x^{(1)} \\ \lambda^{(1)} \end{bmatrix} + \alpha^{(1)} \begin{bmatrix} \Delta\mu^{(1)} \\ \Delta x^{(1)} \\ \Delta\lambda^{(1)} \end{bmatrix},$$

- where $\alpha^{(1)}$ is chosen to ensure that the $x^{(2)}$ and $\mu^{(2)}$ strictly satisfy the non-negativity constraints.

16.4.4.5 Adjustment of barrier parameter in example problem

- In Problem (16.4), since $n = 2$ is rather small, we will take an even more aggressive approach and set:

$$t^{(1)} = \frac{1}{10} t^{(1)}_{\text{effective}} = 2.499375 \times 10^{-5}.$$

- We solve $\mathcal{L}\mathcal{Y} = \mathcal{B}$, where:

$$\begin{aligned}
\mathcal{L} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ [X^{(1)}]^{-1} M^{(1)} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & [x^{(1)}]^{\dagger} [M^{(1)}]^{-1} & 1 \end{bmatrix}, \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 3999 & 0 & 1 & 0 & 0 \\ 0 & 2.501 \times 10^{-4} & 0 & 1 & 0 \\ 0 & 0 & 2.501 \times 10^{-4} & 3999 & 1 \end{bmatrix},
\end{aligned}$$

Adjustment of barrier parameter in example problem, continued

$$\mathcal{B} = \begin{bmatrix} x_1^{(1)} - t^{(1)}[\mu_1^{(1)}]^{-1} \\ x_2^{(1)} - t^{(1)}[\mu_2^{(1)}]^{-1} \\ -1 - \lambda^{(1)} + \mu_1^{(1)} \\ 1 - \lambda^{(1)} + \mu_2^{(1)} \\ -x_1^{(1)} - x_2^{(1)} + 1 \end{bmatrix},$$

$$= \begin{bmatrix} 2.250 \times 10^{-4} \\ 0.899775 \\ -1.251 \\ -0.250 \\ 0 \end{bmatrix},$$

$$\mathcal{Y} = \begin{bmatrix} 2.250 \times 10^{-4} \\ 0.899775 \\ -2.150 \\ -0.250 \\ 1001.050 \end{bmatrix}.$$

- Now we solve $\mathcal{U}\begin{bmatrix}\Delta\mu^{(1)}\\ \Delta x^{(1)}\\ \Delta\lambda^{(1)}\end{bmatrix} = \mathcal{Y}$, where:

$$\mathcal{U} = \begin{bmatrix} -[M^{(1)}]^{-1}X^{(1)} & -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & [X^{(1)}]^{-1}M^{(1)} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & -[\mu_1^{(1)}]^{-1}x_1^{(1)} - [\mu_2^{(1)}]^{-1}x_2^{(1)} \end{bmatrix},$$

$$= \begin{bmatrix} -2.501\times 10^{-4} & 0 & -1 & 0 & 0 \\ 0 & -3999 & 0 & -1 & 0 \\ 0 & 0 & 3999 & 0 & 1 \\ 0 & 0 & 0 & 2.501\times 10^{-4} & 1 \\ 0 & 0 & 0 & 0 & -3999 \end{bmatrix},$$

Adjustment of barrier parameter in example problem, continued

- so that:

$$\begin{bmatrix} \Delta\mu_1^{(1)} \\ \Delta\mu_2^{(1)} \\ \Delta x_1^{(1)} \\ \Delta x_2^{(1)} \\ \Delta\lambda^{(1)} \end{bmatrix} = \begin{bmatrix} 1.000 \\ -2.251 \times 10^{-4} \\ -4.751 \times 10^{-4} \\ 4.755 \times 10^{-4} \\ -0.25032 \end{bmatrix}.$$

Adjustment of barrier parameter in example problem, continued

- Solving for $\alpha^{(1)}$ to bring the next iterate no closer than 0.9995 of the distance to the boundary of $x \geq \mathbf{0}, \mu \geq \mathbf{0}$ we find $\alpha^{(1)} = 0.526$ and:

$$\begin{bmatrix} \mu_1^{(2)} \\ \mu_2^{(2)} \\ x_1^{(2)} \\ x_2^{(2)} \\ \lambda^{(2)} \end{bmatrix} = \begin{bmatrix} 1.525428 \\ 1.317 \times 10^{-4} \\ 3.056 \times 10^{-7} \\ 0.999999875 \\ 1.119 \end{bmatrix}.$$

- After only two iterations, $x^{(2)}$ is extremely close to the minimizer of Problem (16.4), which is $x^\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
- The optimal values of the other variables are: $\mu^\star = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and $\lambda^\star = [1]$.

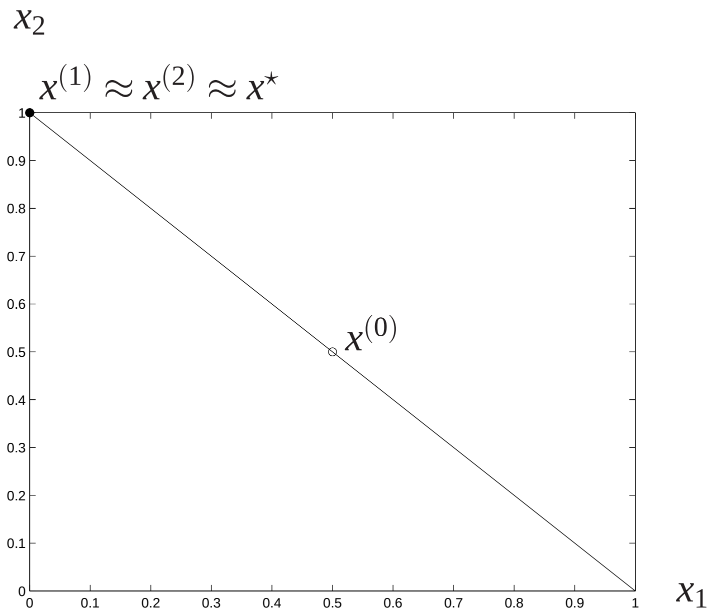

Fig. 16.11. Progress of primal–dual interior point algorithm in x coordinates for Problem (16.4). The feasible set is indicated by the solid line.

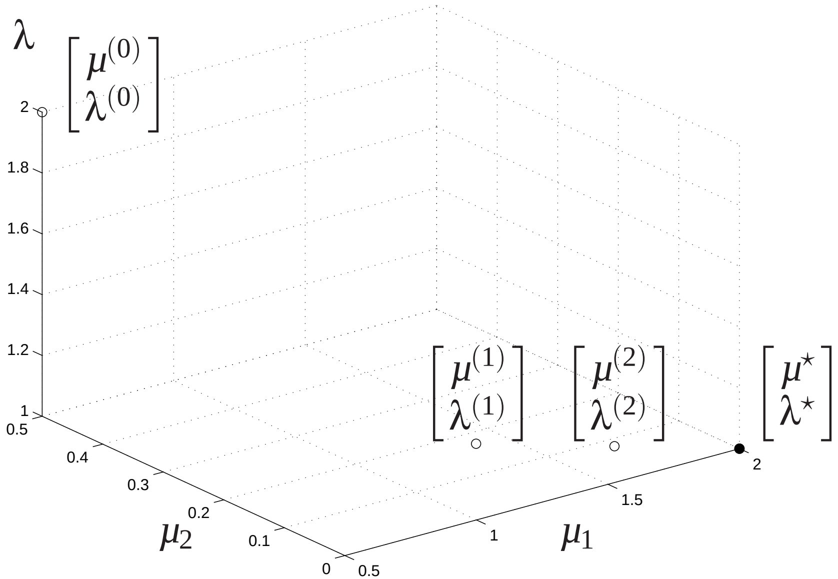

Fig. 16.12. Progress of primal–dual interior point algorithm in μ and λ coordinates for Problem (16.4).

16.4.4.6 Rate of convergence

- For larger and more complex problems, we should expect to take more iterations to approach an accurate answer and we might expect to use a less aggressive reduction of the barrier parameter t at each iteration.
- Empirically, however, even large problems usually take no more than a few tens of iterations to solve to high accuracy.
- Variants of this algorithm can be proven to converge super-linearly or quadratically for linear and quadratic programming problems and for some other types of convex objectives.

16.4.5 Finding an initial feasible guess

- As with the active set algorithm, we must find an initial feasible guess in **phase 1** before proceeding to minimize the objective in **phase 2.**
- We require that the initial guess for the primal–dual interior point algorithm satisfies $x > \mathbf{0}$ and $\mu > \mathbf{0}$.
- Again, we will define a problem related to Problem (16.1) that includes artificial variables and apply the primal–dual interior point algorithm to it.
- There are a number of possible ways to define the related problem.

Finding an initial feasible guess, continued

- For example let $x^{(0)} \in \mathbb{R}^n_{++}$, suppose A has linearly independent rows, define $\tilde{b} = b - Ax^{(0)}$, and consider the problem:

$$\min_{x \in \mathbb{R}^n, w \in \mathbb{R}} \{w | Ax + \tilde{b}w = b, x \geq \mathbf{0}, w \geq 0\}. \qquad (16.34)$$

- Note that $x^{(0)}$ and $w^{(0)} = 1$ satisfy the equality and strictly satisfies the inequality constraints of Problem (16.34) and is therefore a feasible initial guess for this problem that can be used by the primal–dual interior point algorithm.
- We solve this problem using the primal–dual interior point algorithm and this feasible initial guess.
- If $\begin{bmatrix} x^\star \\ w^\star \end{bmatrix}$ is a minimizer of Problem (16.34) with $w^\star = 0$ then $x^\star$ satisfies the equality and inequality constraints of Problem (16.1).
- If $x^\star > \mathbf{0}$ then the primal–dual interior point algorithm can then use $x^\star$ as an initial guess for solving Problem (16.1).
- If $w^\star > 0$ then Problem (16.1) is infeasible.

16.4.6 Summary

16.4.6.1 Initial guess

- The algorithm begins with an initial guess $\begin{bmatrix} \mu^{(0)} \\ x^{(0)} \\ \lambda^{(0)} \end{bmatrix}$ satisfying $Ax^{(0)} = b$, $\mu^{(0)} > \mathbf{0}, x^{(0)} > \mathbf{0}$, and with an initial barrier parameter $t^{(0)}$.

16.4.6.2 General iteration

Newton–Raphson step direction

- At the ν-th iteration we solve (16.30) for the Newton–Raphson step direction $\begin{bmatrix} \Delta\mu^{(\nu)} \\ \Delta x^{(\nu)} \\ \Delta\lambda^{(\nu)} \end{bmatrix}$.
- The coefficient matrix has been partially block factorized as shown in (16.31).
- The factorization should be completed by an algorithm for symmetric indefinite matrices as mentioned in Section 5.4.7.

Step-size

- The iterate is updated according to:

$$\begin{bmatrix} \mu^{(\nu+1)} \\ x^{(\nu+1)} \\ \lambda^{(\nu+1)} \end{bmatrix} = \begin{bmatrix} \mu^{(\nu)} \\ x^{(\nu)} \\ \lambda^{(\nu)} \end{bmatrix} + \alpha^{(\nu)} \begin{bmatrix} \Delta\mu^{(\nu)} \\ \Delta x^{(\nu)} \\ \Delta\lambda^{(\nu)} \end{bmatrix},$$

- where $\alpha^{(\nu)}$ is chosen so that $\mu^{(\nu+1)} > \mathbf{0}$ and $x^{(\nu+1)} > \mathbf{0}$, (and possibly also to satisfy a sufficient decrease criterion for the barrier objective ϕ.)
- One rule to guarantee non-negativity of $\mu^{(\nu+1)}$ and $x^{(\nu+1)}$ is to set:

$$\alpha^{(\nu)} = \min\left\{1.0, 0.9995 \times \left[\min_{\ell\in\{1,\dots,n\}} \left\{ \frac{\mu_\ell^{(\nu)}}{-\Delta\mu_\ell^{(\nu)}} \middle| \Delta\mu_\ell^{(\nu)} < 0 \right\}\right], \right.$$
$$\left. 0.9995 \times \left[\min_{\ell\in\{1,\dots,n\}} \left\{ \frac{x_\ell^{(\nu)}}{-\Delta x_\ell^{(\nu)}} \middle| \Delta x_\ell^{(\nu)} < 0 \right\}\right] \right\}.$$

- The step-size may have to be reduced further to satisfy the sufficient decrease criterion for the barrier objective ϕ.

16.4.6.3 Update of barrier parameter

- We then update the value of the barrier parameter using a rule such as:

$$t^{(\nu+1)} = \frac{\sum_{\ell=1}^{n} \mu_\ell^{(\nu+1)} x_\ell^{(\nu+1)}}{n^2}.$$

16.4.6.4 Stopping criteria

- The iterations continue until $t^{(\nu)}$ is sufficiently reduced, the change in iterates is small, and the first-order necessary conditions of Problem (16.1) are satisfied sufficiently accurately.
- In the case of linear and quadratic programs, we can use duality to develop a stopping criterion that guarantees closeness of $f(x^{(\nu)})$ to the minimum.
- Suppose that at each iteration ν we generate iterates $x^{(\nu)} > \mathbf{0}, \lambda^{(\nu)}$, and $\mu^{(\nu)} > \mathbf{0}$ that satisfy (16.28)–(16.29) then we can use duality to bound the error in the estimate of the infimum by:

$$f(x^{(\nu)}) - \inf_{x\in\mathbb{R}^n}\{f(x)|Ax=b, x\geq\mathbf{0}\} \leq [\mu^{(\nu)}]^{\dagger}x^{(\nu)}.$$

- If the problem has a minimum and we iterate until:

$$[\mu^{(\nu)}]^{\dagger}x^{(\nu)} \leq \varepsilon_f,$$

then $f(x^{(\nu)})$ will be within ε_f of the minimum.

16.4.7 Discussion and variations

- If f is quadratic then linearizing (16.28) introduces no error so that the Newton–Raphson update can exactly predict the changes necessary to satisfy (16.28)–(16.29).
- (16.27) is always non-linear and we neglect important terms when we linearize it.
- A development of the primal–dual algorithm we have described, called the primal–dual **predictor–corrector method**, uses the factorization of (16.30) for two successive updates, one of which is used to bring the iterates closer to being on the central path by reducing the variation of $x_\ell^{(\nu)} \mu_\ell^{(\nu)}$ with ℓ.
- If the problem formulation requires non-negativity constraints on only some of the entries of x, then the barrier function terms and the corresponding Lagrange multipliers can be omitted for the other, unconstrained, entries.

16.5 Summary

- We have described optimality conditions for non-negatively constrained minimization problems, considering also the special case of convex problems.
- We then considered active set algorithms briefly and interior point algorithms in more detail as algorithms to solve non-negatively constrained problems.