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

Inequality-constrained optimization, continued

17

Algorithms for linear inequality-constrained minimization

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

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, Cx \leq d \}, \tag{17.1}$$

- where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{r \times n}$, and $d \in \mathbb{R}^r$ are constants.
- We call the constraints $Cx \leq d$ **linear inequality constraints**.

Key issues

- Optimality conditions for **inequality-constrained problems** based on the results for equality-constrained problems,
- optimality conditions for **convex problems**,
- **transformations** of problems, and
- **duality** and **sensitivity analysis**.

17.1 Optimality conditions

17.1.1 First-order necessary conditions

17.1.1.1 Analysis

Theorem 17.1 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, C \in \mathbb{R}^{r\times n}, d \in \mathbb{R}^r$. Consider Problem (17.1):*

$$\min_{x\in\mathbb{R}^n} \{ f(x) | Ax = b, Cx \leq d\},$$

and a point $x^\star \in \mathbb{R}^n$. If $x^\star$ is a local minimizer of Problem (17.1) then:

$$\begin{aligned} \exists\lambda^\star \in \mathbb{R}^m, \exists\mu^\star \in \mathbb{R}^r \text{ such that: } \nabla f(x^\star) + A^\dagger\lambda^\star + C^\dagger\mu^\star &= \mathbf{0}; \\ M^\star(Cx^\star - d) &= \mathbf{0}; \\ Ax^\star &= b; \\ Cx^\star &\leq d; \text{ and} \\ \mu^\star &\geq \mathbf{0}, \end{aligned} \tag{17.2}$$

where $M^\star = \mathrm{diag}\{\mu^\star_\ell\} \in \mathbb{R}^{r\times r}$.

The vectors $\lambda^\star$ *and* $\mu^\star$ *satisfying the conditions (17.2) are called the vectors of Lagrange multipliers for the constraints* $Ax = b$ *and* $Cx \le d$*, respectively. The conditions that* $M^\star(Cx^\star - d) = \mathbf{0}$ *are called the* **complementary slackness conditions**. *They say that, for each* ℓ*, either the* ℓ*-th inequality constraint is binding or the* ℓ*-th Lagrange multiplier is equal to zero (or both).*

Proof The proof consists of several steps:

(i) showing that $x^\star$ is a local minimizer of the related equality-constrained problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, C_\ell x = d_\ell, \forall \ell \in \mathbb{A}(x^\star) \},$$

where C_ℓ is the ℓ-th row of C and the active inequality constraints at $x^\star$ for Problem (17.1) are included as equality constraints,

(ii) using the necessary conditions of the related equality-constrained problem to define $\lambda^\star$ and $\mu^\star$ that satisfy the first four lines of (17.2), and

(iii) proving that $\mu^\star \geq \mathbf{0}$ by showing that if a constraint ℓ, say, had a negative value of its Lagrange multiplier $\mu_\ell^\star < 0$ then the objective could be reduced by moving in a direction such that constraint ℓ becomes strictly feasible.

$\square$

17.1.1.2 Example

- Recall the example quadratic program, Problem (2.18):

$$\min_{x \in \mathbb{R}^2} \{ f(x) | Ax = b, Cx \leq d \}.$$

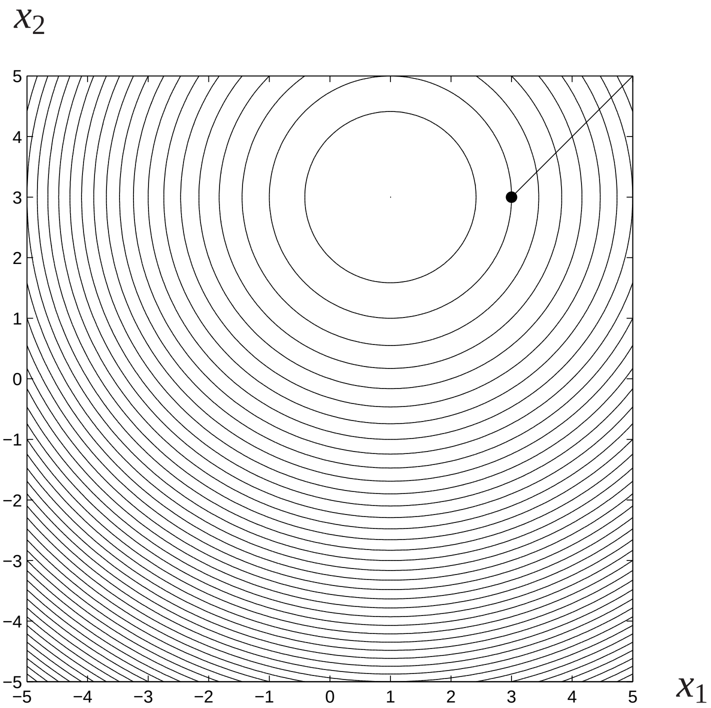

Fig. 17.1. Contour sets of objective function and feasible set for Problem (2.18). The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. The feasible set is the "half-line" starting at the point $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$, which is also the minimizer and is illustrated with a •.

Example, continued

- The objective and constraints are specified by:

$$\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\
A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\
b &= \begin{bmatrix} 0 \end{bmatrix}, \\
C &= \begin{bmatrix} 0 & -1 \end{bmatrix}, \\
d &= \begin{bmatrix} -3 \end{bmatrix}.
\end{aligned}$$

- In Section 2.3.2, we observed that the solution of this problem was $x^\star = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$.
- We claim that $x^\star = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ together with $\lambda^\star = [-4]$ and $\mu^\star = [4]$ satisfy (17.2) for Problem (2.18).

Example, continued

$$
\begin{aligned}
\forall x \in \mathbb{R}^2, \nabla f(x) &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} x + \begin{bmatrix} -2 \\ -6 \end{bmatrix}, \\
\nabla f(x^\star) + A^\dagger \lambda^\star + C^\dagger \mu^\star & \\
&= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} + \begin{bmatrix} -2 \\ -6 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} [-4] + \begin{bmatrix} 0 \\ -1 \end{bmatrix} [4], \\
&= \mathbf{0}; \\
\mu^\star (Cx^\star - d) &= [4] \left(\begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} - [-3] \right), \\
&= [0]; \\
Ax^\star &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \\
&= [0], \\
&= b;
\end{aligned}
$$

Example, continued

$$
\begin{aligned}
Cx^{\star} &= \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \\
&= [-3], \\
&\leq [-3], \\
&= d; \text{ and} \\
\mu^{\star} &= [4], \\
&\geq [0].
\end{aligned}
$$

17.1.1.3 Discussion

- The Lagrange multipliers adjust the unconstrained optimality conditions to balance the constraints against the objective.
- We will again refer to the equality and inequality constraints in (17.2) as *the* first-order necessary conditions, although we recognize that the first-order necessary conditions also include, strictly speaking, the other items in the hypothesis of Theorem 17.1.
- As previously, these conditions are also known as the **Kuhn–Tucker** (KT) or the **Karush–Kuhn–Tucker** (KKT) conditions and a point satisfying the conditions is called a **KKT point**.

17.1.1.4 Lagrangian

- Recall Definition 3.2 of the **Lagrangian**.
- For Problem (17.1) the Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}$ is defined by:

$$\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^\dagger (Ax - b) + \mu^\dagger (Cx - d).$$

- As in the equality-constrained case, define the gradients of $\mathcal{L}$ with respect to x, λ, and μ by, respectively, $\nabla_x \mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial x}\right]^\dagger$, $\nabla_\lambda \mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial \lambda}\right]^\dagger$, and $\nabla_\mu \mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial \mu}\right]^\dagger$.
- Evaluating the gradients with respect to x, λ, and μ, we have:

$$\begin{aligned}
\nabla_x \mathcal{L}(x, \lambda, \mu) &= \nabla f(x) + A^\dagger \lambda + C^\dagger \mu, \\
\nabla_\lambda \mathcal{L}(x, \lambda, \mu) &= Ax - b, \\
\nabla_\mu \mathcal{L}(x, \lambda, \mu) &= Cx - d.
\end{aligned}$$

- Setting the first two of these expressions equal to zero reproduces some of the first-order necessary conditions for the problem.

Lagrangian, continued

- As with equality-constrained problems, the Lagrangian provides a convenient way to remember the optimality conditions.
- However, unlike the equality-constrained case, in order to recover the first-order necessary conditions for Problem (17.1) we have to:
  - add the complementary slackness conditions; that is, $M^\star(Cx^\star - d) = \mathbf{0}$,
  - add the non-negativity constraints on μ, that is, $\mu \geq \mathbf{0}$, and
  - interpret the third expression as corresponding to inequality constraints; that is, $Cx \leq d$.
- If the hypotheses of Theorem 17.1 are satisfied and, additionally, f is convex then $x^\star$ is a global minimizer of $\mathcal{L}(\bullet, \lambda^\star, \mu^\star)$, where $\lambda^\star$ and $\mu^\star$ are the Lagrange multipliers.

17.1.2 Second-order sufficient conditions

17.1.2.1 Analysis

Theorem 17.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice partially differentiable with continuous second partial derivatives, $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, $C \in \mathbb{R}^{r \times n}, d \in \mathbb{R}^r$. Consider Problem (17.1):*

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, Cx \leq d \},$$

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

$$\begin{aligned}
\nabla f(x^\star) + A^\dagger \lambda^\star + C^\dagger \mu^\star &= \mathbf{0}, \\
M^\star (Cx^\star - d) &= \mathbf{0}, \\
Ax^\star &= b, \\
Cx^\star &\leq d, \\
\mu^\star &\geq \mathbf{0}, \text{ and}
\end{aligned}$$

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

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

where C_ℓ is the ℓ-th row of C and

$$\mathbb{A}_+(x^\star, \mu^\star) = \{\ell \in \{1, \dots, r\} | C_\ell x^\star = d_\ell, \mu_\ell^\star > 0\}.$$

Then $x^\star$ is a strict local minimizer of Problem (17.1). $\square$

- The conditions in the theorem are called the **second-order sufficient conditions** (or **SOSC**.)
- 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.

17.1.2.2 Example

- Recall again the example quadratic program, Problem (2.18).
- For this problem:

$$\begin{aligned} Cx^\star &= d, \\ \mu^\star &= [4], \\ \mathbb{A}_+(x^\star,\mu^\star) &= \{\ell \in \{1,\dots,r\} | C_\ell x^\star = d_\ell, \mu^\star_\ell > 0\}, \\ &= \{1\}, \end{aligned}$$

- since the only inequality constraint in this problem is binding and the corresponding Lagrange multiplier is non-zero.
- Consequently,

$$\begin{aligned} \mathcal{N}_+ &= \{\Delta x \in \mathbb{R}^n | A\Delta x = \mathbf{0}, C_\ell \Delta x = 0, \forall \ell \in \mathbb{A}_+(x^\star,\mu^\star)\}, \\ &= \{\Delta x \in \mathbb{R}^n | A\Delta x = 0, C\Delta x = 0\}, \\ &= \{\mathbf{0}\}, \end{aligned}$$

- and $\nabla^2 f(x^\star)$ is positive definite on this null space by definition.

17.1.2.3 Discussion

- The sets $\mathcal{N}_+$ and $\mathbb{A}_+(x^\star, \mu^\star)$ have analogous roles to their roles in the case of non-negativity constraints presented in Section 16.1.2.
- If $\nabla^2 f(x^\star)$ is positive definite on $\mathcal{N}_+$ then there can be no feasible descent directions for f at $x^\star$.
- As in the non-negatively constrained case, the set $\mathbb{A}_+(x^\star, \mu^\star)$ can be a *strict* subset of $\mathbb{A}(x^\star)$, since it omits those constraints ℓ for which $C_\ell x^\star = d_\ell$ and $\mu_\ell^\star = 0$.
- Therefore, the null space specified in Theorem 17.2:

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

- can strictly contain the null space corresponding to the equality constraints and the active inequality constraints.
- That is $\mathcal{N}_+$ can strictly contain the null space:

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

- As in the non-negatively constrained case, constraints for which $C_\ell x^\star = d_\ell$ and $\mu_\ell^\star = 0$ are called **degenerate constraints**.

17.1.2.4 Example of degenerate constraints

- Consider the following modified version of Problem (2.18):

$$\min_{x \in \mathbb{R}^2} \{ f(x) | Ax = b, Cx \leq \hat{d} \}. \tag{17.3}$$

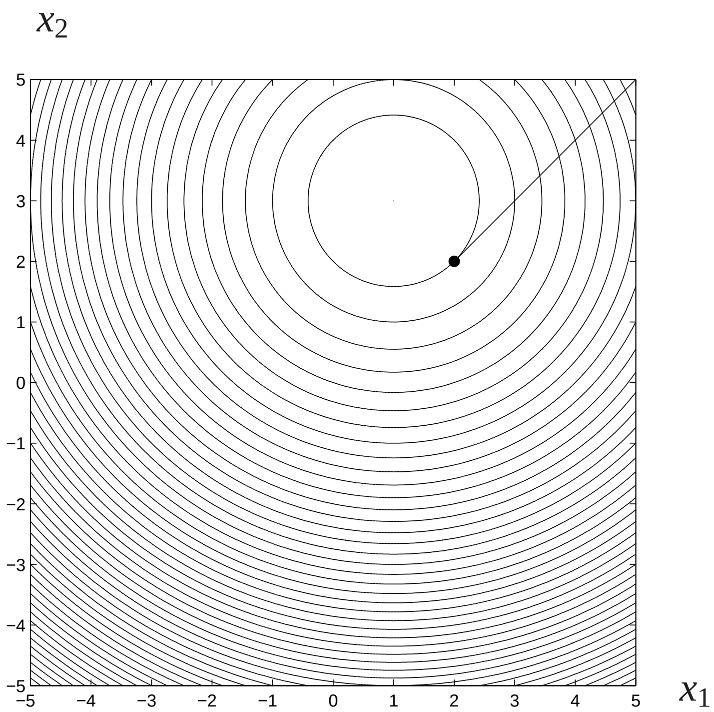

Fig. 17.2. Contour sets of objective function and feasible set for Problem (17.3). The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. The feasible set is the "half-line" starting at the point $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$, which is also the minimizer and is illustrated with a •.

Example of degenerate constraints, continued

- The objective and constraints are specified by:

$$\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\
A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\
b &= \begin{bmatrix} 0 \end{bmatrix}, \\
C &= \begin{bmatrix} 0 & -1 \end{bmatrix}, \\
\hat{d} &= \begin{bmatrix} -2 \end{bmatrix}.
\end{aligned}$$

- First consider the relaxation of Problem (17.3) where we neglect the inequality constraint.
- This relaxation yields Problem (2.13), which we first met in Section 2.3.2.2 and which has minimizer $x^\star = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

Example of degenerate constraints, continued

- Now notice that:

$$Cx^\star = [-2] \leq [-2] = \hat{d},$$

- so that $x^\star$ is feasible for Problem (17.3).
- By Theorem 3.10, $x^\star$ is a also minimizer of Problem (17.3).
- We claim that $x^\star$ together with $\lambda^\star = [-2]$ and $\mu^\star = [0]$ satisfy the first-order necessary conditions for Problem (17.3).

$$\begin{aligned}
\nabla f(x^\star) + A^\dagger \lambda^\star + C^\dagger \mu^\star & \\
&= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -2 \\ -6 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} [-2] + \begin{bmatrix} 0 \\ -1 \end{bmatrix} [0], \\
&= \mathbf{0}; \\
\mu^\star (Cx^\star - \hat{d}) &= [0] \left(\begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} - [-2] \right), \\
&= [0] \times [0], \\
&= [0];
\end{aligned}$$

Example of degenerate constraints, continued

$$
\begin{aligned}
Ax^\star &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \\
&= [0], \\
&= b; \\
Cx^\star &= \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \\
&= [-2], \\
&\leq [-2], \\
&= \hat{d}; \text{ and} \\
\mu^\star &= [0], \\
&\geq [0].
\end{aligned}
$$

- Notice that $Cx^\star = \hat{d}$ and $\mu^\star = [0]$, so that the constraint $Cx \leq \hat{d}$ is degenerate.

Example of degenerate constraints, continued

- For this problem:

$$\begin{aligned}
\mathbb{A}_+(x^\star,\mu^\star) &= \{\ell \in \{1,\ldots,r\} | C_\ell x^\star = \hat{d}_\ell, \mu_\ell^\star > 0\}, \\
&= \emptyset, \\
\mathcal{N}_+ &= \{\Delta x \in \mathbb{R}^2 | A\Delta x = \mathbf{0}, C_\ell \Delta x = 0, \forall \ell \in \mathbb{A}_+(x^\star,\mu^\star)\}, \\
&= \{\Delta x \in \mathbb{R}^2 | A\Delta x = \mathbf{0}\}, \\
&= \{\Delta x \in \mathbb{R}^2 | \Delta x_1 = \Delta x_2\}.
\end{aligned}$$

- We have that:

$$\forall x \in \mathbb{R}^2, \nabla^2 f(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

- which is positive definite on $\mathbb{R}^2$ and therefore also positive definite on $\mathcal{N}_+$.
- Therefore, the second-order sufficient conditions hold and, by Theorem 17.2, $x^\star$ is a strict local minimizer of Problem (17.3).

17.1.2.5 *Example of second-order sufficient conditions not holding*

- Consider the following modified version of Problem (17.3) from Section 17.1.2.4:

$$\min_{x \in \mathbb{R}^2} \{\phi(x) | Ax = b, Cx \leq \hat{d}\}, \tag{17.4}$$

- where $\phi : \mathbb{R}^2 \to \mathbb{R}$ is defined by:

$$\forall x \in \mathbb{R}^2, \phi(x) = -f(x).$$

- That is, we are minimizing $(-f)$ instead of f.
- We claim that $\hat{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ together with $\hat{\lambda} = [2]$ and $\hat{\mu} = [0]$ satisfy the first-order necessary conditions for Problem (17.4).

Example of second-order sufficient conditions not holding, continued

$$
\begin{aligned}
\forall x \in \mathbb{R}^2, \nabla\phi(x) &= \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} x + \begin{bmatrix} 2 \\ 6 \end{bmatrix}, \\
\nabla\phi(\hat{x}) + A^\dagger\hat{\lambda} + C^\dagger\hat{\mu} & \\
&= \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 6 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} [2] + \begin{bmatrix} 0 \\ -1 \end{bmatrix} [0], \\
&= \mathbf{0}; \\
\hat{\mu}(C\hat{x} - \hat{d}) &= [0] \left(\begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} - [-2] \right), \\
&= [0] \times [0], \\
&= [0]; \\
A\hat{x} &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \\
&= [0], \\
&= b;
\end{aligned}
$$

Example of second-order sufficient conditions not holding, continued

$$
\begin{aligned}
C\hat{x} &= \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \\
&= [-2], \\
&\leq [-2], \\
&= \hat{d}; \text{ and} \\
\hat{\mu} &= [0], \\
&\geq [0].
\end{aligned}
$$

- Notice that again $C\hat{x} = \hat{d}$ and $\hat{\mu} = [0]$.
- Therefore, if $\hat{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\hat{\mu} = [0]$ *were* the minimizer and the Lagrange multiplier corresponding to the constraint $Cx \leq \hat{d}$, then this constraint would be degenerate.

Example of second-order sufficient conditions not holding, continued

- For this problem:

$$
\begin{aligned}
\mathbb{A}_+(\hat{x},\hat{\mu}) &= \{\ell \in \{1,\dots,r\} | C_\ell \hat{x} = \hat{d}_\ell, \hat{\mu}_\ell > 0\}, \\
&= \emptyset, \\
\mathcal{N}_+ &= \{\Delta x \in \mathbb{R}^2 | A\Delta x = \mathbf{0}, C_\ell \Delta x = 0, \forall \ell \in \mathbb{A}_+(\hat{x},\hat{\mu})\}, \\
&= \{\Delta x \in \mathbb{R}^2 | A\Delta x = \mathbf{0}\}, \\
&= \{\Delta x \in \mathbb{R}^2 | \Delta x_1 = \Delta x_2\}.
\end{aligned}
$$

- However, we note that $\nabla^2\phi(\hat{x}) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$ is not positive definite on $\mathcal{N}_+$.
- Therefore the second-order sufficient conditions do not hold.
- In fact, $\hat{x}$ is not a minimizer of the problem, since the objective can be reduced by moving away from $\hat{x}$ along the equality constraint so as to make the inequality constraint strictly feasible.

Example of second-order sufficient conditions not holding, continued

- The fact that $\hat{x}$ is not a minimizer can be seen from Figure 17.2, on noting that the contours of ϕ are the same as those of f, except that the heights of the contours of ϕ decrease *away* from the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$.
- If we had *erroneously* considered the null space:

$$\begin{aligned} \mathcal{N} &= \{\Delta x \in \mathbb{R}^2 | A\Delta x = \mathbf{0}, C_\ell \Delta x = 0, \forall \ell \in \mathbb{A}(\hat{x})\}, \\ &= \{\Delta x \in \mathbb{R}^2 | \Delta x_1 = \Delta x_2, -\Delta x_2 = 0\}, \\ &= \{\mathbf{0}\}, \end{aligned}$$

- then we would not have realized that $\hat{x}$ is not a minimizer.

17.2 Convex problems

17.2.1 First-order sufficient conditions

17.2.1.1 Analysis

- If the constraints consist of linear equality and inequality constraints and if f is convex on the feasible set then the problem is convex.
- In this case, the first-order necessary conditions are also sufficient for optimality.

Theorem 17.3 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, C \in \mathbb{R}^{r\times n}, d \in \mathbb{R}^r$. Consider Problem (17.1):*

$$\min_{x\in\mathbb{R}^n} \{ f(x) | Ax = b, Cx \leq d \},$$

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

(i) *f is convex on $\{x \in \mathbb{R}^n | Ax = b, Cx \leq d\}$,*

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

(iii) $M^\star(Cx^\star - d) = \mathbf{0}$,

(iv) *$Ax^\star = b$ and $Cx^\star \leq d$, and*

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

Then $x^\star$ is a global minimizer of Problem (17.1).

Proof The proof is very similar to the proof of Theorem 16.3 in Chapter 16. □

- In addition to the first-order necessary conditions, the first-order sufficient conditions require that f is convex on the feasible set.

17.2.1.2 Example

- Again consider Problem (2.18) from Sections 2.3.2.3, 17.1.1.2, and 17.1.2.2.
- In Section 17.1.1.2, we observed that $x^\star = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, $\lambda^\star = [-4]$, and $\mu^\star = [4]$ satisfy the first-order necessary conditions for this problem.
- Moreover, f is twice continuously differentiable with continuous partial derivatives and the Hessian is positive definite.
- Therefore, f is convex and $x^\star$ is the global minimizer of the problem.

17.2.2 Duality

- As we discussed in Section 3.4 and as in the discussion of linear equality constraints in Section 13.2.2, we can define a dual problem where the role of variables and constraints is partly or fully swapped.
- We recall some of the discussion in Section 3.4 in the following sections.

17.2.2.1 Dual function

- We have observed in Section 17.1.1.4 that if f is convex then $x^\star$ is a global minimizer of $\mathcal{L}(\bullet, \lambda^\star, \mu^\star)$.
- Recall Definition 3.3 of the **dual function** and **effective domain**.
- For Problem (17.1), the dual function $\mathcal{D} : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R} \cup \{-\infty\}$ is defined by:

$$\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r}, \mathcal{D}(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu). \tag{17.5}$$

- The effective domain of $\mathcal{D}$ is:

$$\mathbb{E} = \left\{ \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r} \middle| \mathcal{D}(\lambda, \mu) > -\infty \right\}.$$

- Recall that by Theorem 3.12, $\mathbb{E}$ is convex and $\mathcal{D}$ is concave on $\mathbb{E}$.

Example

- We continue with Problem (2.18).
- The problem is:

$$\min_{x\in\mathbb{R}^2}\{f(x)|Ax=b, Cx\leq d\},$$

- where:

$$\begin{aligned}
\forall x\in\mathbb{R}^2, f(x) &= (x_1-1)^2+(x_2-3)^2,\\
A &= \begin{bmatrix}1 & -1\end{bmatrix},\\
b &= \begin{bmatrix}0\end{bmatrix},\\
C &= \begin{bmatrix}0 & -1\end{bmatrix},\\
d &= \begin{bmatrix}-3\end{bmatrix}.
\end{aligned}$$

- The Lagrangian $\mathcal{L}:\mathbb{R}^2\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ for this problem is defined by:

$$\begin{aligned}
&\forall x\in\mathbb{R}^2, \forall\lambda\in\mathbb{R}, \forall\mu\in\mathbb{R},\\
\mathcal{L}(x,\lambda,\mu) &= f(x)+\lambda^\dagger(Ax-b)+\mu^\dagger(Cx-d),\\
&= (x_1-1)^2+(x_2-3)^2+\lambda\begin{bmatrix}1 & -1\end{bmatrix}x+\mu(\begin{bmatrix}0 & -1\end{bmatrix}x+3).
\end{aligned}$$

Example, continued

- For any given λ and μ, the Lagrangian $\mathcal{L}(\bullet, \lambda, \mu)$ is strictly convex.
- By Corollary 10.6, the first-order necessary conditions $\nabla_x \mathcal{L}(x, \lambda, \mu) = 0$ are sufficient for minimizing $\mathcal{L}(\bullet, \lambda, \mu)$.
- Moreover, a minimizer exists, so that the inf in the definition of $\mathcal{D}$ can be replaced by min.
- Furthermore, there is a unique minimizer $x^{(\lambda,\mu)}$ corresponding to each value of λ and μ:

$$\forall x \in \mathbb{R}^2, \forall \lambda \in \mathbb{R}, \forall \mu \in \mathbb{R},$$
$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda, \mu) &= \nabla f(x) + A^\dagger \lambda + C^\dagger \mu, \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} x + \begin{bmatrix} -2 \\ -6 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} \lambda + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \mu, \\ \forall \lambda \in \mathbb{R}, \forall \mu \in \mathbb{R}, x^{(\lambda,\mu)} &= -\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \left[\begin{bmatrix} -2 \\ -6 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} \lambda + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \mu \right], \\ &= \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix} \lambda + \begin{bmatrix} 0 \\ 0.5 \end{bmatrix} \mu. \end{aligned} \tag{17.6}$$

Example, continued

- Consequently, the effective domain is $\mathbb{E} = \mathbb{R} \times \mathbb{R}$ and the dual function $\mathcal{D} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is given by:

$$
\begin{aligned}
\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^2, \mathcal{D}(\lambda, \mu) &= \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu), \\
&= \mathcal{L}(x^{(\lambda,\mu)}, \lambda, \mu), \text{ since } x^{(\lambda,\mu)} \text{ minimizes } \mathcal{L}(\bullet, \lambda, \mu), \\
&= (x_1^{(\lambda,\mu)} - 1)^2 + (x_2^{(\lambda,\mu)} - 3)^2 \\
&\quad + \lambda \begin{bmatrix} 1 & -1 \end{bmatrix} x^{(\lambda,\mu)} + \mu \left(\begin{bmatrix} 0 & -1 \end{bmatrix} x^{(\lambda,\mu)} + 3 \right), \\
&= -\frac{1}{2}(\lambda)^2 - \frac{1}{4}(\mu)^2 - 2\lambda - \frac{1}{2}\mu\lambda,
\end{aligned}
$$

- on substituting from (17.6) for $x^{(\lambda,\mu)}$.

Analysis

- As in the equality-constrained case, if the objective is convex on $\mathbb{R}^n$ then the minimum of Problem (17.1) is equal to $\mathcal{D}(\lambda^\star, \mu^\star)$, where $\lambda^\star$ and $\mu^\star$ are the Lagrange multipliers that satisfy the necessary conditions for Problem (17.1).
- As in the equality-constrained case, under certain conditions, the Lagrange multipliers can be found as the maximizer of the **dual problem**:

$$\max_{\begin{bmatrix}\lambda\\\mu\end{bmatrix}\in\mathbb{E}} \{\mathcal{D}(\lambda,\mu) | \mu \geq \mathbf{0}\}, \tag{17.7}$$

- where $\mathcal{D} : \mathbb{E} \to \mathbb{R}$ is the dual function defined in (17.5).
- Again, Problem (17.1) is called the **primal problem** to distinguish it from Problem (17.7).

Theorem 17.4 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is convex and partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{r\times n}$, and $d \in \mathbb{R}^r$. Consider the primal problem, Problem (17.1):*

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

Also, consider the dual problem, Problem (17.7). We have the following.

(i) *If the primal problem possesses a minimum then the dual problem possesses a maximum and the optima are equal. That is:*

$$\min_{x\in\mathbb{R}^n} \{ f(x) | Ax = b, Cx \leq d \} = \max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}} \{ \mathcal{D}(\lambda,\mu) | \mu \geq \mathbf{0} \}.$$

(ii) *If:*

- $\begin{bmatrix}\lambda\\ \mu\end{bmatrix} \in \mathbb{E}$,
- $\min_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda,\mu)$ *exists, and*
- f *is twice partially differentiable with continuous second partial derivatives and* $\nabla^2 f$ *is positive definite,*

then $\mathcal{D}$ is partially differentiable at $\begin{bmatrix} \lambda \\ \mu \end{bmatrix}$ *with continuous partial derivatives and:*

$$\begin{bmatrix} \nabla_\lambda \mathcal{D}(\lambda,\mu) \\ \nabla_\mu \mathcal{D}(\lambda,\mu) \end{bmatrix} = \nabla \mathcal{D}(\lambda,\mu) = \begin{bmatrix} Ax^{(\lambda,\mu)} - b \\ Cx^{(\lambda,\mu)} - d \end{bmatrix}, \qquad (17.8)$$

where $x^{(\lambda,\mu)}$ is the unique minimizer of $\min_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda,\mu)$.

Proof

(i) Suppose that Problem (17.1) possesses a minimum with minimizer $x^\star$. By Theorem 17.1,

$$\begin{aligned}\exists\lambda^\star \in \mathbb{R}^m, \exists\mu^\star \in \mathbb{R}^r_+ \text{ such that } \mathbf{0} &= \nabla f(x^\star) + A^\dagger\lambda^\star + C^\dagger\mu^\star, \\ &= \nabla_x\mathcal{L}(x^\star,\lambda^\star,\mu^\star),\end{aligned}$$

where we note that $\mathcal{L}(\bullet,\lambda^\star,\mu^\star)$ is convex and partially differentiable with continuous partial derivatives, so that, by Corollary 10.6, $x^\star$ is also a minimizer of $\mathcal{L}(\bullet,\lambda^\star,\mu^\star)$. Therefore,

$$\begin{aligned}\mathcal{D}(\lambda^\star,\mu^\star) &= \inf_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda^\star,\mu^\star), \\ &= \mathcal{L}(x^\star,\lambda^\star,\mu^\star), \text{ because } x^\star \text{ minimizes } \mathcal{L}(\bullet,\lambda^\star,\mu^\star), \\ &= f(x^\star) + [\lambda^\star]^\dagger(Ax^\star - b) + [\mu^\star]^\dagger(Cx^\star - d), \text{ by definition,} \\ &= f(x^\star), \text{ since } x^\star \text{ is feasible and, by Theorem 17.1,} \\ &\qquad\qquad \mu^\star_\ell(C_\ell x^\star - d_\ell) = 0, \forall \ell = 1,\ldots,r, \\ &\geq \mathcal{D}(\lambda,\mu), \forall\lambda \in \mathbb{R}^m, \forall\mu \in \mathbb{R}^r_+, \text{ by Theorem 3.13.}\end{aligned}$$

That is, $\begin{bmatrix}\lambda^\star\\ \mu^\star\end{bmatrix}$ maximizes the dual function over $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^r_+$:

$$\begin{aligned} f(x^\star) &= \max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}} \{\mathcal{D}(\lambda,\mu) | \mu \geq \mathbf{0}\}, \\ &= \mathcal{D}(\lambda^\star,\mu^\star). \end{aligned}$$

(ii) Proved in an exercise.

□

Discussion

- As in the equality-constrained case, it is possible for $\mathcal{D}$ to not be partially differentiable at a point $\begin{bmatrix}\lambda\\ \mu\end{bmatrix} \in \mathbb{E}$ if:

  $\mathcal{L}(\bullet,\lambda,\mu)$ is bounded below (so that $\inf_{x\in\mathbb{R}^n}\mathcal{L}(x,\lambda,\mu) \in \mathbb{R}$) yet the minimum $\min_{x\in\mathbb{R}^n}\mathcal{L}(x,\lambda,\mu)$ does not exist, or

  there are multiple minimizers of $\min_{x\in\mathbb{R}^n}\mathcal{L}(x,\lambda,\mu)$.

Corollary 17.5 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice partially differentiable with continuous second partial derivatives and with $\nabla^2 f$ positive definite, $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, C \in \mathbb{R}^{r\times n}, d \in \mathbb{R}^r$. Consider Problem (17.1):*

$$\min_{x\in\mathbb{R}^n} \{ f(x) | Ax = b, Cx \leq d\},$$

the Lagrangian of this problem, and the effective domain $\mathbb{E}$ of the dual function. If:

- *the effective domain $\mathbb{E}$ contains $\mathbb{R}^m \times \mathbb{R}^r_+$, and*
- *for each $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^r_+$, $\min_{x\in\mathbb{R}^n} \mathcal{L}(x, \lambda, \mu)$ exists,*

then necessary and sufficient conditions for $\begin{bmatrix} \lambda^\star \\ \mu^\star \end{bmatrix} \in \mathbb{R}^{m+r}$ to be the maximizer of the dual problem:

$$\max_{\left[\begin{smallmatrix}\lambda \\ \mu\end{smallmatrix}\right]\in\mathbb{E}} \{ \mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0}\},$$

are:

$$
\begin{aligned}
Ax^{(\lambda^\star,\mu^\star)} &= b;\\
Cx^{(\lambda^\star,\mu^\star)} - d &\leq \mathbf{0};\\
M^\star(Cx^{(\lambda^\star,\mu^\star)} - d) &= \mathbf{0}; \text{ and}\\
\mu^\star &\geq \mathbf{0},
\end{aligned}
$$

where $\{x^{(\lambda^\star,\mu^\star)}\} = \operatorname{argmin}_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda^\star,\mu^\star)$ *and* $M^\star = \operatorname{diag}\{\mu_\ell^\star\}$.

Moreover, if $\begin{bmatrix}\lambda^\star\\ \mu^\star\end{bmatrix}$ *maximizes the dual problem then the corresponding minimizer of the Lagrangian,* $x^{(\lambda^\star,\mu^\star)}$*, together with* $\lambda^\star$ *and* $\mu^\star$ *satisfy the first-order necessary conditions for Problem (17.1).*

Proof Note that the hypothesis implies that the dual function is finite for all $\lambda \in \mathbb{R}^m$ and all $\mu \in \mathbb{R}^r_+$ so that dual problem is a non-negatively constrained maximization of a real-valued function and, moreover, by Theorem 3.12, $-\mathcal{D}$ is convex and partially differentiable with continuous partial derivatives on the convex set:

$$\left\{ \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r} \middle| \mu \geq \mathbf{0} \right\}.$$

Moreover, by Theorem 17.4,

$$\begin{aligned} \nabla_\lambda \mathcal{D}(\lambda^\star, \mu^\star) &= Ax^{(\lambda^\star, \mu^\star)} - b, \\ \nabla_\mu \mathcal{D}(\lambda^\star, \mu^\star) &= Cx^{(\lambda^\star, \mu^\star)} - d. \end{aligned}$$

Applying Theorems 17.1 and 17.3 to the dual problem and some substitution yields the conclusion. □

Discussion

- Theorem 17.4 shows that an alternative approach to finding the minimum of Problem (17.1) involves finding the *maximum* of the dual function $\mathcal{D}(\lambda, \mu)$ over $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r, \mu \geq \mathbf{0}$.
- Theorem 3.12 shows that the dual function has at most one local maximum.
- To seek the maximum of $\mathcal{D}(\lambda, \mu)$ over $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r, \mu \geq \mathbf{0}$, we can, for example, utilize the value of the gradient of $\mathcal{D}$ from (17.8) as part of an active set or interior point algorithm.
- As in the equality-constrained case, under some circumstances, it is also possible to calculate the Hessian of $\mathcal{D}$.

Example

- Continuing with the dual of Problem (2.18), we recall that the effective domain is $\mathbb{E} = \mathbb{R} \times \mathbb{R}$ and the dual function $\mathcal{D} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is:

$$\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^2, \mathcal{D}(\lambda, \mu) = -\frac{1}{2}(\lambda)^2 - \frac{1}{4}(\mu)^2 - 2\lambda - \frac{1}{2}\mu\lambda,$$

- with unique minimizer of the Lagrangian specified by (17.6).
- The dual function is twice partially differentiable with continuous second partial derivatives.
- In particular,

$$\begin{aligned}
\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^2, \nabla\mathcal{D}(\lambda, \mu) &= \begin{bmatrix} -2 - \lambda - \mu/2 \\ -\lambda/2 - \mu/2 \end{bmatrix}, \\
\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^2, \nabla^2\mathcal{D}(\lambda, \mu) &= \begin{bmatrix} -1 & -0.5 \\ -0.5 & -0.5 \end{bmatrix}.
\end{aligned}$$

- We claim that $\begin{bmatrix} \lambda^\star \\ \mu^\star \end{bmatrix} = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$ maximizes the dual function over $\mu \geq [0]$.

Example, continued

- In particular $\nabla \mathcal{D}(\lambda^\star, \mu^\star) = \mathbf{0}$, $\mu^\star > [0]$, and $\nabla^2 \mathcal{D}$ is negative definite.
- Consequently, $\begin{bmatrix} \lambda^\star \\ \mu^\star \end{bmatrix}$ is the unique maximizer of dual Problem (17.7).
- We also observe that $\lambda^\star = [-4]$ and $\mu^\star = [4]$ satisfy the conditions specified in Corollary 17.5 for maximizing the dual.
- To see this, we first use (17.6) to evaluate $x^{(\lambda,\mu)}$ at $\lambda^\star = [-4]$ and $\mu^\star = [4]$.
- We obtain $x^{(\lambda^\star,\mu^\star)} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$.
- We will also show that the necessary and sufficient conditions in Corollary 17.5 for maximizing the dual are satisfied.

Example, continued

$$
\begin{aligned}
\mu^\star(Cx^{(\lambda^\star,\mu^\star)} - d) &= [4]\left(\begin{bmatrix} 0 & -1 \end{bmatrix}\begin{bmatrix} 3 \\ 3 \end{bmatrix} - [-3]\right), \\
&= [0]; \\
Ax^{(\lambda^\star,\mu^\star)} - b &= \begin{bmatrix} 1 & -1 \end{bmatrix}\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \\
&= [0]; \\
Cx^{(\lambda^\star,\mu^\star)} - d &= \begin{bmatrix} 0 & -1 \end{bmatrix}\begin{bmatrix} 3 \\ 3 \end{bmatrix} - [-3], \\
&= [0], \\
&\leq [0]; \text{ and} \\
\mu^\star &= [4], \\
&\geq [0].
\end{aligned}
$$

- Moreover, $x^{(\lambda^\star,\mu^\star)}$ together with $\lambda^\star$, and $\mu^\star$ satisfy the first-order necessary conditions for Problem (2.18).

Discussion

- As in the equality-constrained case, it is essential in Theorem 17.4 for f to be convex on the *whole* of $\mathbb{R}^n$, not just on the feasible set because the inner minimization of $\mathcal{L}(\bullet, \lambda, \mu)$ is taken over the whole of $\mathbb{R}^n$.
- Unfortunately, if f is not strictly convex then $\mathcal{L}(\bullet, \lambda, \mu)$ may have multiple minimizers over x for fixed λ and μ.
- In this case, it may turn out that some of the minimizers of $\mathcal{L}(\bullet, \lambda^\star, \mu^\star)$ do not actually minimize (17.1).
- Moreover, if there are multiple minimizers of $\mathcal{L}(\bullet, \lambda, \mu)$ then $\mathcal{D}(\lambda, \mu)$ may be not partially differentiable.
- The issues are similar to the equality-constrained case.

Discussion, continued

- In the particular cases of linear and of strictly convex quadratic programs, we can calculate the dual function and characterize the effective domain explicitly.
- This allows us to use duality for the not strictly convex case of linear programs.
- The dual problem is non-negatively constrained of the form of Problem (16.1) and we can apply essentially the same algorithms as we developed for Problem (16.1).
- We will take this approach in Section 17.3.2.

17.2.2.3 Dual of linear and quadratic programs

- In the case of linear and of strictly convex quadratic programs, we can characterize the effective domain and the dual function explicitly by solving the first-order necessary conditions for minimizing the Lagrangian:

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = \mathbf{0}.$$

- The approach parallels that of the Wolfe dual, described in Section 13.2.2.2.
- We first consider the case of linear objective and then strictly convex quadratic objective.

Linear program

$$\begin{aligned}
\forall x \in \mathbb{R}^n, f(x) &= c^\dagger x, \\
\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \mathcal{L}(x, \lambda, \mu) &= c^\dagger x + \lambda^\dagger (Ax - b) + \mu^\dagger (Cx - d), \\
\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \nabla_x \mathcal{L}(x, \lambda, \mu) &= c + A^\dagger \lambda + C^\dagger \mu.
\end{aligned}$$

- The first-order necessary and sufficient conditions for minimizing the Lagrangian are $c + A^\dagger \lambda + C^\dagger \mu = \mathbf{0}$.
- These conditions do not involve x, but also do not necessarily have a solution for all values of λ and μ.

Linear program, continued

If $c + A^{\dagger}\lambda + C^{\dagger}\mu \neq \mathbf{0}$ then $\mathcal{L}(\bullet, \lambda, \mu)$ is unbounded below and $\begin{bmatrix} \lambda \\ \mu \end{bmatrix} \notin \mathbb{E}$.

If $c + A^{\dagger}\lambda + C^{\dagger}\mu = \mathbf{0}$ then, after substituting, we find that:

$$\begin{aligned} \mathcal{D}(\lambda, \mu) &= -\lambda^{\dagger}b - \mu^{\dagger}d, \\ &> -\infty. \end{aligned}$$

- That is:

$$\begin{aligned} \mathbb{E} &= \left\{ \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r} \middle| c + A^{\dagger}\lambda + C^{\dagger}\mu = \mathbf{0} \right\}, \\ \forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{E}, \mathcal{D}(\lambda, \mu) &= -\lambda^{\dagger}b - \mu^{\dagger}d. \end{aligned}$$

- We now substitute the characterization of the dual function and effective domain into the definition of the dual problem and apply Theorem 17.4.
- We assume that $\min_{x \in \mathbb{R}^n} \{c^{\dagger}x | Ax = b, Cx \leq d\}$ possesses a minimum.

Linear program, continued

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \{c^\dagger x | Ax = b, Cx \leq d\} \\
&= \max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{E}} \{\mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0}\}, \text{ by Theorem 17.4,} \\
&= \max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{R}^{m+r}} \{\mathcal{D}(\lambda, \mu) | c + A^\dagger \lambda + C^\dagger \mu = \mathbf{0}, \mu \geq \mathbf{0}\}, \\
&\qquad \text{since } \mathbb{E} = \left\{ \begin{bmatrix}\lambda\\ \mu\end{bmatrix} \in \mathbb{R}^{m+r} \middle| c + A^\dagger \lambda + C^\dagger \mu = \mathbf{0} \right\}, \\
&= \max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{R}^{m+r}} \{-\lambda^\dagger b - \mu^\dagger d | c + A^\dagger \lambda + C^\dagger \mu = \mathbf{0}, \mu \geq \mathbf{0}\}, \\
&\qquad \text{since } \mathcal{D}(\lambda, \mu) = -\lambda^\dagger b - \mu^\dagger d \text{ for } c + A^\dagger \lambda + C^\dagger \mu = \mathbf{0}, \\
&= -\min_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{R}^{m+r}} \{\lambda^\dagger b + \mu^\dagger d | c + A^\dagger \lambda + C^\dagger \mu = \mathbf{0}, \mu \geq \mathbf{0}\}, \\
&= -\min_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{R}^{m+r}} \left\{ \begin{bmatrix}b\\ d\end{bmatrix}^\dagger \begin{bmatrix}\lambda\\ \mu\end{bmatrix} \middle| \begin{bmatrix}A\\ C\end{bmatrix}^\dagger \begin{bmatrix}\lambda\\ \mu\end{bmatrix} = -c, \mu \geq \mathbf{0} \right\}. \qquad (17.9)
\end{aligned}
$$

Linear program, continued

- The dual problem in the last line of (17.9) has a linear objective, linear equality constraints, and non-negativity constraints on the variables μ.
- Since the primal problem has a minimum, there is at least one point in the feasible set of the dual problem,

$$\mathbb{E}_+ = \left\{ \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r} \middle| \begin{bmatrix} A \\ C \end{bmatrix}^\dagger \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = -c, \mu \geq \mathbf{0} \right\},$$

  namely the Lagrange multipliers $\begin{bmatrix} \lambda^\star \\ \mu^\star \end{bmatrix}$ that correspond to the minimizer $x^\star$ of the primal problem.
- We say that the problem is **dual feasible**.
- We have transformed a primal problem with n variables, m equality constraints, and r inequality constraints into a dual problem with $m+r$ variables, n equality constraints, and r inequality constraints.
- The dual of a linear program is therefore also a linear program, but with non-negativity constraints instead of general linear inequalities.

Quadratic program

$$\begin{aligned}
\forall x \in \mathbb{R}^n, f(x) &= \tfrac{1}{2}x^\dagger Q x + c^\dagger x, \\
\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, & \\
\mathcal{L}(x,\lambda,\mu) &= \tfrac{1}{2}x^\dagger Q x + c^\dagger x + \lambda^\dagger(Ax-b) + \mu^\dagger(Cx-d), \\
\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, & \\
\nabla_x \mathcal{L}(x,\lambda,\mu) &= Qx + c + A^\dagger\lambda + C^\dagger\mu.
\end{aligned}$$

- The first-order necessary conditions for minimizing $\mathcal{L}(\bullet,\lambda,\mu)$ are that $Qx + c + A^\dagger\lambda + C^\dagger\mu = \mathbf{0}$.
- Assuming that Q is positive definite, these conditions have a solution for all values of λ and μ, namely $x = -Q^{-1}[c + A^\dagger\lambda + C^\dagger\mu]$, yielding:

$$\begin{aligned}
\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r}, & \\
\mathcal{D}(\lambda,\mu) &= -\frac{1}{2}[c + A^\dagger\lambda + C^\dagger\mu]^\dagger Q^{-1}[c + A^\dagger\lambda + C^\dagger\mu] - \lambda^\dagger b - \mu^\dagger d, \\
&> -\infty,
\end{aligned}$$

- so that $\mathbb{E} = \mathbb{R}^{m+r}$.

Quadratic program, continued

- If $\min_{x\in\mathbb{R}^n}\{\frac{1}{2}x^\dagger Qx + c^\dagger x | Ax = b, Cx \leq d\}$ possesses a minimum then by Theorem 17.4:

$$\begin{aligned}
&\min_{x\in\mathbb{R}^n}\{\frac{1}{2}x^\dagger Qx + c^\dagger x | Ax = b, Cx \leq d\} \\
&\quad = \max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}}\{\mathcal{D}(\lambda,\mu) | \mu \geq \mathbf{0}\}, \\
&\quad = \max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{R}^{m+r}}\left\{ -\frac{1}{2}[c + A^\dagger\lambda + C^\dagger\mu]^\dagger Q^{-1}[c + A^\dagger\lambda + C^\dagger\mu] - \lambda^\dagger b - \mu^\dagger d \,\middle|\, \mu \geq \mathbf{0}\right\}, \\
&\quad = -\min_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{R}^{m+r}}\left\{ \frac{1}{2}[c + A^\dagger\lambda + C^\dagger\mu]^\dagger Q^{-1}[c + A^\dagger\lambda + C^\dagger\mu] + \lambda^\dagger b + \mu^\dagger d \,\middle|\, \mu \geq \mathbf{0}\right\}.
\end{aligned} \tag{17.10}$$

- The dual problem in the last line of (17.10) has a quadratic objective and non-negativity constraints.

Quadratic program, continued

- We have transformed a primal problem with n variables, m equality constraints, and r inequality constraints into a dual problem with $m+r$ variables and r inequality constraints.
- The dual of a quadratic program is therefore also a quadratic program.
- Again, the form of the inequality constraints in the dual is simpler than in the primal problem since they are non-negativity constraints.
- If we solve the problem in the last line of (17.10) for optimal $\lambda^\star$ and $\mu^\star$ then the minimizer, $x^\star$, of the primal problem can be recovered as $x^\star = -Q^{-1}[c + A^\dagger\lambda^\star + C^\dagger\mu^\star]$.

Discussion

- There is considerable literature on the relationship between primal and dual linear programs and on primal and dual quadratic programs.
- The standard treatment of duality in linear programming differs from the way we have discussed it here, there are a variety of special cases, and we have omitted many details.
- For example, we have not discussed how to recover a minimizer of the primal problem from the solution of the dual of a linear program.
- Furthermore, **primal–dual algorithms** (including the primal–dual interior point algorithm described in Section 16.4.3.3) represent *both* the primal and dual variables and simultaneously solve for both the minimizer and the Lagrange multipliers.
- The primal–dual interior point algorithm is therefore essentially the same whether it is applied to the primal or dual problem.

Analysis

- We can define the **partial dual** with respect to some of the constraints.
- For example, define $\mathcal{D}_= : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$ and $\mathcal{D}_\leq : \mathbb{R}^r \to \mathbb{R} \cup \{-\infty\}$ by:

$$\begin{aligned}
\forall \lambda \in \mathbb{R}^m, \mathcal{D}_=(\lambda) &= \inf_{x \in \mathbb{R}^n} \{ f(x) + \lambda^\dagger (Ax - b) | Cx \leq d \}, \\
\forall \mu \in \mathbb{R}^r, \mathcal{D}_\leq(\mu) &= \inf_{x \in \mathbb{R}^n} \{ f(x) + \mu^\dagger (Cx - d) | Ax = b \}.
\end{aligned}$$

- The function $\mathcal{D}_=$ is called the partial dual with respect to the equality constraints, while $\mathcal{D}_\leq$ is called the partial dual with respect to the inequality constraints.

Theorem 17.6 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is convex and partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{r\times n}$, and $d \in \mathbb{R}^r$. Suppose that Problem (17.1) possesses a minimum. Then:*

$$\min_{x\in\mathbb{R}^n}\{f(x)|Ax = b, Cx \leq d\} = \max_{\lambda\in\mathbb{E}_=}\{\mathcal{D}_=(\lambda)\} = \max_{\mu\in\mathbb{E}_\leq}\{\mathcal{D}_\leq(\mu)|\mu \geq \mathbf{0}\},$$

where $\mathcal{D}_=$ is the partial dual with respect to the equality constraints and $\mathbb{E}_=$ is its effective domain and $\mathcal{D}_\leq$ is the partial dual with respect to the inequality constraints and $\mathbb{E}_\leq$ is its effective domain. □

- It is also possible to take a partial dual with respect to only some of the equality or some of the inequality constraints or some of both of the equality and inequality constraints, leaving the other constraints explicitly in the problem.

Separable problems

- To see an example of the usefulness of partial duality, consider the case where:

  f is separable and strictly convex, so that $f(x) = \sum_{k=1}^{n} f_k(x_k)$, and
  the inequality constraints consist only of upper and lower bound
  constraints $\underline{x} \leq x \leq \overline{x}$.

Separable problems, continued

$$
\begin{aligned}
\forall \lambda \in \mathbb{R}^m, \mathcal{D}_=(\lambda) \\
&= \min_{x \in \mathbb{R}^n} \{ f(x) + \lambda^\dagger (Ax - b) | Cx \leq d \}, \\
&= \min_{x \in \mathbb{R}^n} \{ f(x) + \lambda^\dagger (Ax - b) | \underline{x} \leq x \leq \overline{x} \}, \\
&= \min_{x \in \mathbb{R}^n} \left\{ \sum_{k=1}^{n} f_k(x_k) + \lambda^\dagger \sum_{k=1}^{n} A_k x_k - \lambda^\dagger b \,\middle|\, \underline{x}_k \leq x_k \leq \overline{x}_k, \forall k = 1, \ldots, n \right\}, \\
&\qquad \text{where } A_k \text{ is the } k\text{-th column of } A, \\
&= \min_{x \in \mathbb{R}^n} \left\{ \sum_{k=1}^{n} \left(f_k(x_k) + \lambda^\dagger A_k x_k \right) - \lambda^\dagger b \,\middle|\, \underline{x}_k \leq x_k \leq \overline{x}_k, \forall k = 1, \ldots, n \right\}, \\
&\qquad \text{on re-arranging}, \\
&= \sum_{k=1}^{n} \min_{x_k \in \mathbb{R}} \{ f_k(x_k) + \lambda^\dagger A_k x_k | \underline{x}_k \leq x_k \leq \overline{x}_k \} - \lambda^\dagger b, \qquad (17.11)
\end{aligned}
$$

- on swapping the minimum and the summation.

Separable problems, continued

- For a given value of λ, the dual with respect to the equality constraints is the sum of:

  a constant $(-\lambda^\dagger b)$, and

  n one-dimensional optimization sub-problems that can each be evaluated independently.

- The primal problem has been **decomposed** into a collection of sub-problems using the partial dual.
- For a problem with constraints that couple between sub-problems, by dualizing with respect to these **coupling constraints** we can decompose the problem into the sub-problems.
- If each sub-problem is simple enough, it may be possible to evaluate its minimizer and minimum explicitly without resorting to an iterative technique.
- This applies to the least-cost production case study from Section 15.1 and will be described in detail in Section 18.1.2.2.

17.3 Approaches to finding minimizers

- In this section we will show two basic ways in which inequality-constrained Problem (17.1) can be transformed into the form of Problem (16.1) from Chapter 16.
- We can then use the algorithmic development from Chapter 16 to solve Problem (17.1).

17.3.1 Primal algorithm

17.3.1.1 Transformation

Slack variables

- To handle the inequality constraints of the primal problem, we consider the following problem incorporating **slack variables** as introduced in Section 3.3.2:

$$\min_{x\in\mathbb{R}^n, w\in\mathbb{R}^r} \{f(x) | Ax = b, Cx + w = d, w \geq \mathbf{0}\}. \tag{17.12}$$

- The variables w are called the **slack variables** because they account for the "slack" in the constraints $Cx \leq d$.
- By Theorem 3.8, Problem (17.12) is equivalent to Problem (17.1).

Slack variables, continued

- In Problem (17.12), if we consider:

  $\begin{bmatrix} x \\ w \end{bmatrix} \in \mathbb{R}^{n+r}$ to be the decision vector,

  f to be the objective, and

  $\begin{bmatrix} A & \mathbf{0} \\ C & \mathbf{I} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}$ to be the equality constraints,

- then Problem (17.12) can be expressed in the form of Problem (16.1) (except that we have non-negativity constraints on just w and not on the whole of the decision vector $\begin{bmatrix} x \\ w \end{bmatrix}$.)
- The equivalent problem is:

$$\min_{x \in \mathbb{R}^n, w \in \mathbb{R}^r} \left\{ f(x) \,\middle|\, \begin{bmatrix} A & \mathbf{0} \\ C & \mathbf{I} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}, w \geq \mathbf{0} \right\}. \tag{17.13}$$

- In the next section, we will apply the primal–dual interior point algorithm from Section 16.4 to Problem (17.13).

17.3.1.2 *Primal–dual interior point algorithm*

Barrier objective and problem

- Given a barrier function $f_\text{b} : \mathbb{R}^r_{++} \to \mathbb{R}$ for the constraints $w \geq \mathbf{0}$ and a barrier parameter $t \in \mathbb{R}_{++}$, we form the **barrier objective** $\phi : \mathbb{R}^n \times \mathbb{R}^r_{++} \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}^n, \forall w \in \mathbb{R}^r_{++}, \phi(x,w) = f(x) + t f_\text{b}(w).$$

- Instead of solving (17.13), we will consider solving the **barrier problem**:

$$\min_{x \in \mathbb{R}^n, w \in \mathbb{R}^r} \left\{ \phi(x,w) \left| \begin{bmatrix} A & \mathbf{0} \\ C & \mathbf{I} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}, w > \mathbf{0} \right. \right\}. \qquad (17.14)$$

- We seek (approximate) minimizers of Problem (17.14) for a decreasing sequence of values of the barrier parameter.

Slater condition

- As in the case of non-negativity constraints described in Section 16.4.2.2, in order to apply the interior point algorithm effectively, we must assume that the **Slater condition** holds so that there are feasible points for Problem (17.14).
- That is, we assume that $\{x \in \mathbb{R}^n | Ax = b, Cx < d\} \neq \emptyset$.

Equality-constrained problem

- To solve Problem (17.14), we can take a similar approach to the primal–dual interior point algorithm for non-negativity constraints presented in Section 16.4 of Chapter 16.
- We partially ignore the inequality constraints and seek a solution to the following linear equality-constrained problem:

$$\min_{x\in\mathbb{R}^n, w\in\mathbb{R}^r} \left\{ \phi(x,w) \,\middle|\, \begin{bmatrix} A & \mathbf{0} \\ C & \mathbf{I} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix} \right\}, \tag{17.15}$$

- which has first-order necessary conditions:

$$\nabla f(x) + A^\dagger \lambda + C^\dagger \mu = \mathbf{0}, \tag{17.16}$$

$$Ax = b, \tag{17.17}$$

$$Cx + w = d, \tag{17.18}$$

$$t\nabla f_{\mathrm{b}}(w) + \mu = \mathbf{0}, \tag{17.19}$$

- where λ and μ are the dual variables on the constraints $Ax = b$ and $Cx + w = d$, respectively.

Equality-constrained problem, continued

- We can use the techniques for minimization of linear equality-constrained problems from Section 13.3.2 of Chapter 13 to solve Problem (17.15).
- In particular, in Section 17.3.1.3, we will consider the Newton–Raphson method for solving the first-order necessary conditions of Problem (17.15).

Logarithmic barrier function

- As in the primal–dual interior point algorithm for non-negativity constraints, we will use the logarithmic barrier function:

$$\begin{aligned}\forall w \in \mathbb{R}^r_{++}, f_{\mathrm{b}}(w) &= -\sum_{\ell=1}^{r} \ln(w_\ell), \\ \forall w \in \mathbb{R}^r_{++}, \nabla f_{\mathrm{b}}(w) &= -[W]^{-1}\mathbf{1},\end{aligned}$$

- where $W = \mathrm{diag}\{w_\ell\} \in \mathbb{R}^{r \times r}$ is a diagonal matrix with diagonal entries equal to $w_\ell, \ell = 1, \ldots, r$.
- Substituting the expression for ∇f_{b} into (17.19) and re-arranging, we obtain:

$$W\mu - t\mathbf{1} = \mathbf{0}. \tag{17.20}$$

- Note that (17.20) is analogous to (16.27) and can again be interpreted as approximating the complementary slackness constraints by a hyperbolic-shaped set.

17.3.1.3 Newton–Raphson method

Analysis

- The Newton–Raphson step direction to solve (17.20) and (17.16)–(17.18) is given by the solution of:

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

- where $M^{(\nu)} = \mathrm{diag}\{\mu_\ell^{(\nu)}\}$ and $W^{(\nu)} = \mathrm{diag}\{w_\ell^{(\nu)}\}$.

Analysis, continued

- As in the case of the primal–dual interior point algorithm for non-negativity constraints discussed in Section 16.4.3.3, we can re-arrange these equations to make them symmetric and use block pivoting on the top left-hand block of the matrix since it is diagonal.
- This results in a system that is similar to (13.35), except that a diagonal block of the form $[M^{(\nu)}]^{-1}W^{(\nu)}$ is added to the Hessian $\nabla^2 f(x^{(\nu)})$.
- Issues regarding solving the first-order necessary conditions, such as factorization of the indefinite coefficient matrix, approximate solution of the conditions, sparsity, and step-size selection, are similar to those described in Sections 16.4.3.3 and 13.3.2.3.

Example

- In this section, we will apply the primal–dual algorithm to the example quadratic program, Problem (2.18):

$$\min_{x \in \mathbb{R}^2} \{ f(x) | Ax = b, Cx \le d \},$$

- where:

$$\begin{aligned} \forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\ A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\ b &= \begin{bmatrix} 0 \end{bmatrix}, \\ C &= \begin{bmatrix} 0 & -1 \end{bmatrix}, \\ d &= \begin{bmatrix} -3 \end{bmatrix}. \end{aligned}$$

Example, continued

- The Newton–Raphson update for the corresponding barrier problem is:

$$\begin{bmatrix} \mu^{(\nu)} & 0 & 0 & 0 & w^{(\nu)} \\ 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta w^{(\nu)} \\ \Delta x^{(\nu)} \\ \Delta \lambda^{(\nu)} \\ \Delta \mu^{(\nu)} \end{bmatrix}$$

$$= \begin{bmatrix} -w^{(\nu)}\mu^{(\nu)} + t \\ -2(x_1^{(\nu)} - 1) - \lambda^{(\nu)} \\ -2(x_2^{(\nu)} - 3) + \lambda^{(\nu)} + \mu^{(\nu)} \\ -x_1^{(\nu)} + x_2^{(\nu)} \\ -3 + x_2^{(\nu)} - w^{(\nu)} \end{bmatrix}.$$

17.3.1.4 Other issues

Adjustment of barrier parameter

- To reduce the barrier parameter, we can use the approach described in Section 16.4.4 of Chapter 16.

Initial guess

- We can take an approach analogous to that in Section 16.4.5 to find an initial feasible guess for Problem (17.13) that is strictly feasible for the non-negativity constraints.

Stopping criterion

- $f(x^{(\nu)})$ will be within ε_f of the minimum of the non-negatively constrained problem if:

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

- where μ is the vector of dual variables corresponding to the constraints $w \geq \mathbf{0}$ (and corresponding to the constraints $Cx \leq d$.)

Non-negativity and lower and upper bound constraints on x

- If we add constraints of the form $x \geq \mathbf{0}$ to Problem (17.1) then we can also include them in the barrier function and Problem (17.14).
- Box constraints of the form $\underline{x}_\ell \leq x_\ell \leq \overline{x}_\ell$ can be treated with a barrier function of the form:

$$-t\left(\ln(x_\ell - \underline{x}_\ell) + \ln(\overline{x}_\ell - x_\ell)\right).$$

17.3.2 Dual algorithm

17.3.2.1 Inequality constraints

- We can take the dual with respect to some or all of the inequality constraints.
- Under convexity assumptions, the dual and primal problems have the same optima.
- If the objective is strictly convex, the minimizer of the primal problem can be recovered from the solution of the dual problem.
- Whereas Problem (17.1) has general linear inequality constraints, taking the dual with respect to all the constraints or with respect to the inequality constraints yields a dual problem where the inequality constraints are non-negativity constraints on variables only.
- We can apply algorithms developed for Problem (16.1).

17.3.2.2 Equality constraints

- Taking the dual with respect to the equality constraints yields a dual problem with no equality nor inequality constraints, but with inner problems having inequality constraints.
- To maximize the dual function, we can apply the algorithms developed in Section 10.2.
- Taking the dual with respect to only some of the equality constraints yields a dual problem with equality constraints.
- We can apply the algorithms developed in Section 13.5.2.
- Taking the partial dual of a problem with separable objectives can yield an inner problem with a simple structure.

17.3.2.3 Non-quadratic objectives

- Although the dual can be found for general non-quadratic objectives, it is often not as useful because the non-linearity of the optimality conditions in the definition of the dual function prevents us from simplifying the objective of the dual as in the linear and quadratic cases.
- If the primal problem is non-convex, we can still apply the algorithm to the dual problem.
- We must be more cautious about interpreting the results since the corresponding value of the primal variables may be infeasible or not optimal for the primal problem.

17.4 Sensitivity

17.4.1 Analysis

- In this section we will analyze a general and a special case of sensitivity analysis for Problem (17.1).
- For the general case, we imagine that we have solved the inequality-constrained minimization problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x;\chi) | A(\chi)x = b(\chi), C(\chi)x \leq d(\chi) \}, \tag{17.21}$$

- for a base-case value of the parameters, say $\chi = \mathbf{0}$, to find the base-case local minimizer $x^\star$ and the base-case Lagrange multipliers $\lambda^\star$ and $\mu^\star$.
- We consider the sensitivity of the local minimum of Problem (17.21) to variation of the parameters about $\chi = \mathbf{0}$.

Analysis, continued

- As well as considering the general case of the sensitivity of the local minimum of Problem (17.21) to χ, we also specialize to the case where only the right-hand sides of the equality and inequality constraints vary.
- That is, we now consider perturbations $\gamma \in \mathbb{R}^m$ and $\eta \in \mathbb{R}^r$ and the problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b - \gamma, Cx \leq d - \eta \}. \tag{17.22}$$

- For the parameter values $\gamma = \mathbf{0}$ and $\eta = \mathbf{0}$, Problem (17.22) is the same as Problem (17.1).
- We consider the sensitivity of the local minimum of Problem (17.22) to variation of the parameters about $\gamma = \mathbf{0}$ and $\eta = \mathbf{0}$.

Corollary 17.7 *Consider Problem (17.21) and suppose that $f : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ is twice partially differentiable with continuous second partial derivatives and that $A : \mathbb{R}^s \to \mathbb{R}^{m\times n}$, $b : \mathbb{R}^s \to \mathbb{R}^m$, $C : \mathbb{R}^s \to \mathbb{R}^{r\times n}$, and $d : \mathbb{R}^s \to \mathbb{R}^r$ are partially differentiable with continuous partial derivatives. Also consider Problem (17.22) and suppose that the function $f : \mathbb{R}^n \to \mathbb{R}$ is twice partially differentiable with continuous second partial derivatives. Suppose that $x^\star \in \mathbb{R}^n$, $\lambda^\star \in \mathbb{R}^m$, and $\mu^\star \in \mathbb{R}^r$ satisfy:*

- *the second-order sufficient conditions for Problem (17.21) for the value of parameters $\chi = \mathbf{0}$, and*
- *the second-order sufficient conditions for Problem (17.22) for the value of parameters $\gamma = \mathbf{0}$ and $\eta = \mathbf{0}$.*

In particular:

- *$x^\star$ is a local minimizer of Problem (17.21) for $\chi = \mathbf{0}$, and*
- *$x^\star$ is a local minimizer of Problem (17.22) for $\gamma = \mathbf{0}$ and $\eta = \mathbf{0}$,*

in both cases with associated Lagrange multipliers $\lambda^\star$ and $\mu^\star$. Moreover, suppose that the matrix $\hat{A}$ has linearly independent rows, where $\hat{A}$ is the matrix with rows consisting of:

- *the m rows of A (or $A(\mathbf{0})$), and*

- *those rows* C_ℓ *of* C *(or of* $C(\mathbf{0})$*) for which* $\ell \in \mathbb{A}(x^\star)$.

Furthermore, suppose that there are no degenerate constraints at the base-case solution.

Then, for values of χ *in a neighborhood of the base-case value of the parameters* $\chi = \mathbf{0}$*, there is a local minimum and corresponding local minimizer and Lagrange multipliers for Problem (17.21). Moreover, the local minimum, local minimizer, and Lagrange multipliers are partially differentiable with respect to* χ *and have continuous partial derivatives in this neighborhood. The sensitivity of the local minimum* $f^\star$ *to* χ*, evaluated at the base-case* $\chi = \mathbf{0}$*, is given by:*

$$\frac{\partial f^\star}{\partial \chi}(\mathbf{0}) = \frac{\partial \mathcal{L}}{\partial \chi}(x^\star, \lambda^\star, \mu^\star; \mathbf{0}),$$

where $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}$ *is the* **parameterized Lagrangian** *defined by:*

$$\begin{aligned}
&\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \forall \chi \in \mathbb{R}^s, \\
&\qquad \mathcal{L}(x, \lambda, \mu; \chi) \;=\; f(x;\chi) + \lambda^\dagger (A(\chi)x - b(\chi)) + \mu^\dagger (C(\chi)x - d(\chi)).
\end{aligned}$$

Furthermore, for values of γ *and* η *in a neighborhood of the base-case* *value of the parameters* $\gamma = \mathbf{0}$ *and* $\eta = \mathbf{0}$*, there is a local minimum and corresponding local minimizer and Lagrange multipliers for Problem (17.22). Moreover, the local minimum, local minimizer, and Lagrange multipliers are partially differentiable with respect to* γ *and* η *and have continuous partial derivatives. The sensitivities of the local minimum to* γ *and* η*, evaluated at the base-case* $\gamma = \mathbf{0}$ *and* $\eta = \mathbf{0}$*, are equal to* $[\lambda^\star]^\dagger$ *and* $[\mu^\star]^\dagger$*, respectively.* $\square$

Discussion

- The Lagrange multipliers yield the sensitivity of the objective to the right-hand side of the equality constraints and inequality constraints.
- Corollary 17.7 does not apply directly to linear programming problems; however, sensitivity analysis can also be applied to linear programming and, as with linear programming in general, the linearity of both objective and constraints leads to various special cases.

17.4.2 Example

- Consider Problem (2.18) from Sections 2.3.2.3, 17.1.1.2, …, 17.3.1.3, which has objective $f : \mathbb{R}^2 \to \mathbb{R}$ and constraints $Ax = b$ and $Cx \leq d$ defined by:

$$\begin{aligned} \forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\ A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\ b &= \begin{bmatrix} 0 \end{bmatrix}, \\ C &= \begin{bmatrix} 0 & -1 \end{bmatrix}, \\ d &= \begin{bmatrix} -3 \end{bmatrix}. \end{aligned}$$

- We have already verified that the second-order sufficient conditions hold at the base-case solution.
- The matrix:

$$\hat{A} = \begin{bmatrix} A \\ C \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix},$$

- has linearly independent rows, and, furthermore, the inequality constraint is not degenerate at the base-case solution.

Example, continued

- Suppose that the inequality constraint was changed to $Cx \leq d - \eta$.
- If η is small enough, then by Corollary 17.7 the minimum of the perturbed problem differs from the minimum of the original problem by approximately $[\mu^\star]^\dagger \eta$.

17.5 Summary

- In this chapter, we considered linear inequality-constrained problems and showed that they could be solved using the techniques developed for non-negatively constrained problems in two ways:
  - (i) using slack variables, and
  - (ii) using duality.
- We also considered sensitivity analysis.

18

Solution of the linear inequality-constrained case studies

Solution of the linear inequality-constrained case studies

- Least-cost production with capacity constraints (Section 18.1),
- Optimal routing in a data communications network (Section 18.2),
- Least absolute value estimation (Section 18.3), and
- Optimal margin pattern classification (Section 18.4).

18.1 Least-cost production with capacity constraints

18.1.1 Problem and analysis

- Recall Problem (15.1):

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

- where $A = -\mathbf{1}^\dagger$, $b = [-D]$.
- This problem has:
  - a convex separable objective,
  - one equality constraint, and
  - two inequality constraints for each variable.
- The inequality constraints are simple bounds on variables.
- We can solve this problem using slight modifications of the algorithms developed in Section 17.3.

18.1.2 Algorithms

18.1.2.1 Primal–dual interior point algorithm

- To enforce the bounds $\underline{x}_\ell \leq x_\ell \leq \overline{x}_\ell$, the corresponding term in the barrier objective is:

$$-t\left(\ln(x_\ell - \underline{x}_\ell) + \ln(\overline{x}_\ell - x_\ell)\right).$$

- Alternatively, we can represent the bound constraints as general linear inequalities in the form $Cx \leq d$.

18.1.2.2 Dual algorithm

- Taking the partial dual with respect to the equality constraints decomposes the problem into a set of sub-problems, one for each machine k, each with two bound constraints $\underline{x}_k \leq x_k \leq \overline{x}_k$.
- Suppose that for each k, the cost f_k of machine k is convex and quadratic and of the form defined in (12.6):

$$\forall x_k \in \underline{\mathbb{S}}_k, f_k(x_k) = \frac{1}{2}Q_{kk}(x_k)^2 + c_k x_k + d_k.$$

Dual algorithm, continued

- For any value of λ, we obtain constrained sub-problems:

$$\forall k = 1, \ldots, n, \min_{x_k \in \mathbb{R}} \left\{ \frac{1}{2} Q_{kk} (x_k)^2 + c_k x_k + d_k - \lambda x_k \middle| \underline{x}_k \leq x_k \leq \overline{x}_k \right\}.$$

- The *unconstrained* minimizer of the objective of each sub-problem is given by setting the derivative of the objective equal to zero, yielding:

$$x_k = \frac{1}{Q_{kk}} (\lambda - c_k).$$

- If the unconstrained minimizer is within the range allowed by the upper and lower bound constraints then, by Theorem 3.10, the unconstrained minimizer is also the minimizer of the constrained sub-problem.
- If the unconstrained minimizer lies outside the range allowed by the bound constraints then the minimizer of the sub-problem is the nearest bound.

Dual algorithm, continued

- For a given value of λ, the minimizer of the inner problem in the definition of the partial dual is $x^{(\lambda)}$, where:

$$\forall k = 1, \ldots, n, x_k^{(\lambda)} = \min\left\{\overline{x}_k, \max\left\{\underline{x}_k, \frac{1}{Q_{kk}}(\lambda - c_k)\right\}\right\}.$$

- Substituting the solution $x_k^{(\lambda)}$ into the expression for the dual, we obtain:

$$\forall \lambda \in \mathbb{R}, \mathcal{D}(\lambda) = \sum_{k=1}^{n} f_k(x_k^{(\lambda)}) + \lambda\left(D - \sum_{k=1}^{n} x_k^{(\lambda)}\right).$$

- The dual variable can be updated using a steepest ascent algorithm based on the satisfaction of the equality constraint according to:

$$\begin{aligned}
\Delta\lambda &= \nabla\mathcal{D}(\lambda), \\
&= Ax^{(\lambda)} - b, \\
&= D - \sum_{k=1}^{n} x_k^{(\lambda)}.
\end{aligned}$$

Dual algorithm, continued

- Since each machine cost function is strictly convex, the minimizer of the primal problem can be found from the solution of the dual algorithm.
- As in Section 13.5.3, we can interpret λ as the tentative price per unit of production.

18.1.3 Changes in demand and capacity

- Corollary 17.7 can be used to estimate the changes in costs due to a change in demand or capacity.

18.2 Optimal routing in a data communications network

18.2.1 Problem and analysis

- Recall Problem (15.6):

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

- where $f : \overline{\mathbb{S}} \to \mathbb{R}$, with $\overline{\mathbb{S}} = \{x \in \mathbb{R}^n | x \geq \mathbf{0}, Cx < \bar{y}\}$, was defined in (15.7):

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

- The delay function ϕ_{ij} in the objective increases without bound as a flow approaches its capacity.
- Consequently, assigning a flow to be arbitrarily close to the link capacity can never be optimal.
- The delay function has the same form as the **reciprocal barrier function**.

Problem and analysis, continued

- Because of the form of the delay function, the strict inequality constraints:

$$Cx < \bar{y},$$

- can be ignored so long as:
  - an initial feasible solution can be found that satisfies these constraints, and
  - a step-size is chosen at each iteration to avoid going outside the feasible region.
- We effectively have a problem with a barrier objective that enforces the strict inequality constraints $Cx < \bar{y}$ and that must be solved for a single fixed value of the barrier parameter.
- That is, to solve Problem (15.6) we can effectively solve the problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, x \geq \mathbf{0} \}. \tag{18.1}$$

18.2.2 Algorithms

- Problem (18.1) is non-negatively constrained and these constraints can be treated using an active set or interior point algorithm, so long as we ensure that the step-size is chosen at each iteration to also satisfy $Cx < \bar{y}$.
- A step-size rule analogous to that for the primal–dual interior point algorithm from Section 16.4.3.3 can be used to ensure satisfaction of the strict inequality constraints $Cx < \bar{y}$.

18.2.3 Changes in links and traffic

- Corollary 17.7 and extensions can be used to estimate the changes in optimal routing to respond to a change in traffic or link capacities.

18.3 Least absolute value estimation

18.3.1 Problem

- Recall Problem (15.10):

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

- This problem has a linear objective and linear inequality constraints.

18.3.2 Algorithms

- We can solve this problem using the primal or the dual algorithms developed in Section 17.3.
- The solution to the corresponding *least-squares* estimation problem can provide a suitable initial guess for $x^{(0)}$.

18.3.3 Changes in the number of points and data

- Corollary 17.7 and extensions can be used to estimate the changes in parameters specifying the affine fit if additional data points are added or if the data changes.

18.4 Optimal margin pattern classification

18.4.1 Problem and analysis

- Recall Problem (15.13):

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

- where $x=\begin{bmatrix}\beta\\\gamma\end{bmatrix}$.
- This problem has the drawback that its feasible set is not closed and may not be convex.
- Furthermore, the inequality constraints are non-linear.

Problem and analysis, continued

- Suppose that Problem (15.13) has maximizer:

$$\begin{bmatrix} z^{\star} \\ x^{\star\star} \end{bmatrix} = \begin{bmatrix} z^{\star} \\ \beta^{\star\star} \\ \gamma^{\star\star} \end{bmatrix},$$

and let $\kappa \in \mathbb{R}_{++}$.

- Then:

$$\begin{aligned} \begin{bmatrix} z^{\star} \\ x^{\star} \end{bmatrix} &= \begin{bmatrix} z^{\star} \\ \beta^{\star} \\ \gamma^{\star} \end{bmatrix}, \\ &= \begin{bmatrix} z^{\star} \\ \beta^{\star\star}/\kappa \\ \gamma^{\star\star}/\kappa \end{bmatrix} \end{aligned} \tag{18.2}$$

- is also a maximizer of Problem (15.13) with the same maximum.
- This is because the coefficients in the equation for a hyperplane can be scaled without changing the hyperplane.

18.4.1.1 First approach to transforming constraints

- Let $\kappa = \|\beta^{\star\star}\|_2$ in (18.2).
- If there is a maximizer to Problem (15.13) then there is a maximizer that satisfies $\beta^{\star} = \beta^{\star\star} / \|\beta^{\star\star}\|_2$, so that $\|\beta^{\star}\|_2 = 1$.
- That is, we can impose the additional constraint $\|\beta\|_2 = 1$ in Problem (15.13) without changing its maximum.
- Furthermore, since $\|\beta\|_2 = 1$ implies that $\beta \neq \mathbf{0}$, we can ignore the constraint $\beta \neq \mathbf{0}$.
- We can use Theorem 3.10 to show that if Problem (15.13) has a maximum then maximizing the objective over the "smaller" feasible set:

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

- will yield the same maximum and hyperplane as Problem (15.13).

First approach to transforming constraints, continued

- The smaller feasible set $\underline{\hat{\mathbb{S}}}$ is closed and bounded, which as we saw in Section 2.3.3 avoids the difficulties that non-closed and unbounded sets present.
- However, a constraint of the form $\|\beta\|_2 = 1$ is still difficult to handle directly because it defines a non-convex set.
- One way to deal with this is to convert the representation into polar coordinates.

First approach to transforming constraints, continued

- Instead, note that if Problem (15.13) has a strictly positive maximum then:

$$\begin{aligned}
& \max_{z\in\mathbb{R},x\in\mathbb{R}^n} \left\{ z \,\middle|\, \zeta(\ell)(\beta^\dagger\psi(\ell)+\gamma) \geq \|\beta\|_2 z, \forall \ell = 1,\ldots,r, \beta \neq \mathbf{0} \right\} \\
&= \max_{z\in\mathbb{R},x\in\mathbb{R}^n} \left\{ z \,\middle|\, \zeta(\ell)(\beta^\dagger\psi(\ell)+\gamma) \geq \|\beta\|_2 z, \forall \ell = 1,\ldots,r, \|\beta\|_2 = 1 \right\}, \\
&\qquad \text{by the argument above,} \\
&= \max_{z\in\mathbb{R},x\in\mathbb{R}^n} \left\{ z \,\middle|\, \zeta(\ell)(\beta^\dagger\psi(\ell)+\gamma) \geq z, \forall \ell = 1,\ldots,r, \|\beta\|_2 = 1 \right\}, \\
&\qquad \text{since } \|\beta\|_2 = 1, \\
&= \max_{z\in\mathbb{R},x\in\mathbb{R}^n} \left\{ z \,\middle|\, \zeta(\ell)(\beta^\dagger\psi(\ell)+\gamma) \geq z, \forall \ell = 1,\ldots,r, \|\beta\|_2 \leq 1 \right\},
\end{aligned}$$

- where any maximizer $\begin{bmatrix} z^\star \\ x^\star \end{bmatrix} = \begin{bmatrix} z^\star \\ \beta^\star \\ \gamma^\star \end{bmatrix}$ of the last problem will satisfy $\|\beta^\star\|_2 = 1$, since if $\|\beta^\star\|_2 < 1$ then we could find a feasible point having a larger objective by dividing both $z^\star$ and $x^\star$ by $\max\{0.5, \|\beta^\star\|_2\}$.

- The *relaxation* of the problem to having the larger feasible set with the constraint $\|\beta\|_2 \leq 1$ yields a convex feasible set with the same maximum as Problem (15.13) and its maximizer specifies the same hyperplane as a maximizer of Problem (15.13).
- Since $\|\beta\|_2$ is not smooth, we will use the equivalent condition $\|\beta\|_2^2 \leq 1$.
- By defining $C \in \mathbb{R}^{r \times n}$ to have ℓ-th row:

$$C_\ell = -\zeta(\ell) \left[\psi(\ell)^\dagger \;\; 1 \right],$$

- and noting that $z - \zeta(\ell)(\beta^\dagger \psi(\ell) + \gamma) = z + C_\ell x$, we can transform the problem to the equivalent problem:

$$\max_{z \in \mathbb{R}, x \in \mathbb{R}^n} \left\{ z \,\middle|\, \mathbf{1}z + Cx \leq \mathbf{0}, \|\beta\|_2^2 \leq 1 \right\}, \tag{18.3}$$

- where we have squared the norm of β to obtain a differentiable function.
- This problem has a linear objective, r linear inequality constraints, and one quadratic inequality constraint.
- We will treat the solution of this formulation of the problem in Section 20.1.

18.4.1.2 Second approach to transforming constraints

- Consider a maximizer $\begin{bmatrix} z^\star \\ x^{\star\star} \end{bmatrix} = \begin{bmatrix} z^\star \\ \beta^{\star\star} \\ \gamma^{\star\star} \end{bmatrix}$ of Problem (15.13).
- Suppose that $z^\star \in \mathbb{R}_{++}$ so that the margin is strictly positive.
- Since $\beta^{\star\star}$ is feasible, we have that $\beta^{\star\star} \neq \mathbf{0}$.
- We can choose $\kappa = \|\beta^{\star\star}\|_2 z^\star$ in (18.2).
- If there is a maximizer to Problem (15.13) with positive margin then there is a maximizer that satisfies $\beta^\star = \beta^{\star\star}/(\|\beta^{\star\star}\|_2 z^\star)$, so that $\|\beta^\star\|_2 z^\star = 1$.
- We can impose the additional constraint $\|\beta\|_2 z = 1$ in Problem (15.13) without changing its maximum.
- Furthermore, since $\|\beta\|_2 z = 1$ implies that $\beta \neq \mathbf{0}$, we can again ignore the constraint $\beta \neq \mathbf{0}$.

Second approach to transforming constraints, continued

- We can again use Theorem 3.10 to show that if Problem (15.13) has a maximizer and strictly positive maximum $z^\star$ then $z^\star$ will also be the maximum of a problem having the same objective but with “smaller” feasible set:

$$\underline{\mathbb{S}} = \left\{ \begin{bmatrix} z \\ x \end{bmatrix} \in \mathbb{R}^{n+1} \,\middle|\, \frac{\zeta(\ell) D(\psi(\ell))}{\|\beta\|_2} \geq z, \forall \ell = 1, \ldots, r, \|\beta\|_2 z = 1 \right\}.$$

- Moreover, if Problem (15.13) has a maximum and maximizer, then at least one of maximizers of the problem is an element of $\underline{\mathbb{S}}$.

- If Problem (15.13) has a maximum and the margin is strictly positive then:

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

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

by the argument above,

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

since $\|\beta\|_2 z=1$,

$$=\max_{z\in\mathbb{R},x\in\mathbb{R}^n}\left\{\frac{1}{\|\beta\|_2}\,\middle|\,\zeta(\ell)(\beta^\dagger\psi(\ell)+\gamma)\geq 1,\forall\ell=1,\ldots,r,\|\beta\|_2 z=1\right\},$$

since $z=1/\|\beta\|_2$,

$$=\max_{x\in\mathbb{R}^n}\left\{\frac{1}{\|\beta\|_2}\,\middle|\,\zeta(\ell)(\beta^\dagger\psi(\ell)+\gamma)\geq 1,\forall\ell=1,\ldots,r\right\},$$ by Corollary 3.7,

on eliminating the variable z using the constraint $\|\beta\|_2 z=1$.

- Also:

$$\max_{x \in \mathbb{R}^n} \left\{ \frac{1}{\|\beta\|_2} \middle| \zeta(\ell)(\beta^\dagger \psi(\ell) + \gamma) \geq 1, \forall \ell = 1, \ldots, r \right\}$$
$$= \left[\frac{1}{\min_{x \in \mathbb{R}^n} \{ \|\beta\|_2 | \zeta(\ell)(\beta^\dagger \psi(\ell) + \gamma) \geq 1, \forall \ell = 1, \ldots, r \}} \right],$$

- by Theorem 3.1, since the reciprocal function is monotonically decreasing.

- As in Section 18.4.1.1, by defining $C \in \mathbb{R}^{r \times n}$ to have ℓ-th row:

$$C_\ell = -\zeta(\ell) \begin{bmatrix} \psi(\ell)^\dagger & 1 \end{bmatrix},$$

- and defining $d = -\mathbf{1} \in \mathbb{R}^r$, we can transform the problem in the denominator to the equivalent problem:

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\beta\|_2^2 \middle| Cx \leq d \right\}, \tag{18.4}$$

- which has a quadratic objective and linear constraints and so is a quadratic program.
- If Problem (18.4) has a minimizer $x^\star = \begin{bmatrix} \beta^\star \\ \gamma^\star \end{bmatrix}$ and $\beta^\star \neq \mathbf{0}$ then the optimal margin is given by $1/\|\beta^\star\|_2$.

18.4.2 Algorithms

18.4.2.1 Primal algorithm

- Problem (18.4) has a convex quadratic objective, linear inequality constraints, and no equality constraints.
- If the number, r, of patterns is extremely large then a further relaxation of the problem may be much easier to solve.
- In particular, we can first solve the problem using only some of the patterns to find a tentative separating hyperplane.
- The feasible set using only some of the patterns is a relaxed version of the feasible set of Problem (18.4).
- Then the rest of the patterns are searched until a pattern is found that is not correctly identified by the tentative separating hyperplane.
- The problem is re-solved with the new pattern incorporated and the process repeated.
- If a separating hyperplane is found after only a modest number of patterns are added then we have avoided the computational effort of solving the problem will all r constraints explicitly represented.

18.4.2.2 Dual algorithm

- The dual of Problem (18.4) has a quadratic objective, non-negativity constraints, and one linear equality constraint.

18.4.3 Changes

- Adding a pattern would add an extra row to the inequality constraints $Cx \leq d$.
- The relaxation procedure described in Section 18.4.2.1 can be applied or the dual can be updated and solved.

19

Algorithms for non-linear inequality-constrained minimization

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

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, h(x) \leq \mathbf{0} \}. \tag{19.1}$$

Key issues

- The notion of a **regular point of constraints** as one characterization of suitable formulations of non-linear equality and inequality constraint functions,
- linearization of non-linear constraint functions,
- optimality conditions and the definition and interpretation of the **Lagrange multipliers**,
- the **Slater condition** as an alternative characterization of suitable formulation of constraint functions for convex problems,
- algorithms that seek points that satisfy the optimality conditions, and
- **sensitivity analysis**.

19.1 Geometry and analysis of constraints

- Our approach will be to linearize the constraint functions g and h about a current iterate and seek step directions.
- We must explore conditions under which this linearization yields a useful approximation to the original feasible set.
- The notion of a regular point, introduced in Section 14.1.1 for non-linear equality-constrained problems and suitably generalized here for non-linear inequality constraints, provides one such **constraint qualification**.

19.1.1 Regular point

Definition 19.1 Let $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^r$. Then we say that $x^\star$ is a **regular point** of the constraints $g(x) = \mathbf{0}$ and $h(x) \leq \mathbf{0}$ if:

(i) $g(x^\star) = \mathbf{0}$ and $h(x^\star) \leq \mathbf{0}$,

(ii) g and h are both partially differentiable with continuous partial derivatives, and

(iii) the matrix $\hat{A}$ has linearly independent rows, where $\hat{A}$ is the matrix with rows consisting of:

- the m rows of the Jacobian $J(x^\star)$ of g evaluated at $x^\star$, and
- those rows $K_\ell(x^\star)$ of the Jacobian K of h evaluated at $x^\star$ for which $\ell \in \mathbb{A}(x^\star)$.

The matrix $\hat{A}$ consists of the rows of $J(x^\star)$ together with those rows of $K(x^\star)$ that correspond to the active constraints. If there are no equality constraints then the matrix $\hat{A}$ consists of the rows of $K(x^\star)$ corresponding to active constraints. If there are no binding inequality constraints then $\hat{A} = J(x^\star)$. If there are no equality constraints and no binding inequality constraints then the matrix $\hat{A}$ has no rows and, by definition, it has linearly independent rows.

□

Regular point, continued

- Let $\hat{r}$ be the number of active inequality constraints at $x^\star$.
- For $x^\star$ to be a regular point of the constraints $g(x) = \mathbf{0}$ and $h(x) \leq \mathbf{0}$, we must have that $m + \hat{r} \leq n$, since otherwise the $m + \hat{r}$ rows of $\hat{A}$ cannot be linearly independent.
- If $x^\star$ is a regular point, then we can find a sub-vector $\omega \in \mathbb{R}^{m+\hat{r}}$ of x such that the $(m + \hat{r}) \times (m + \hat{r})$ matrix consisting of the corresponding $m + \hat{r}$ columns of $\hat{A}$ is non-singular.
- At a regular point of inequality constraints, linearization of the equality constraints and of the binding inequality constraints yields a useful approximation to the feasible set or its boundary, at least locally in the vicinity of the regular point.

19.1.2 Example

- Recall the **dodecahedron** from Section 2.3.2.3 and illustrated in Figure 2.14.

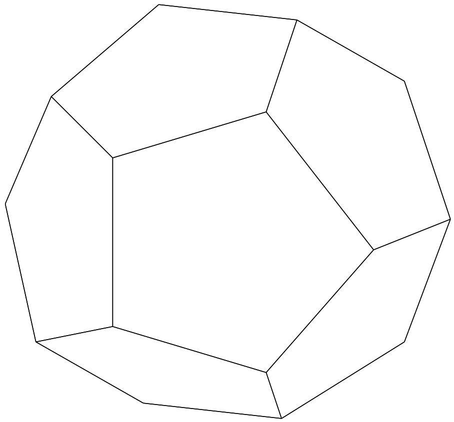

Fig. 19.1. The dodecahedron in $\mathbb{R}^3$ repeated from Figure 2.14.

Example, continued

- The dodecahedron can be described as the set of points satisfying the inequality constraints $h(x) \leq \mathbf{0}$, with $h : \mathbb{R}^3 \to \mathbb{R}^{12}$ affine:

$$\forall x \in \mathbb{R}^3, h(x) = Cx - d,$$

- where:

  $C \in \mathbb{R}^{12 \times 3}$ with each row of C not equal to the zero vector, and
  $d \in \mathbb{R}^{12}$.

- The Jacobian of h is $K = C$ and the ℓ-th row of K is the ℓ-th row of C, which we will denote by C_ℓ.
- If $h(x^\star) \not\leq \mathbf{0}$ so that $x^\star$ is not in the dodecahedron then $x^\star$ is not a regular point by definition.

Example, continued

- If $h(x^\star) \leq \mathbf{0}$ then consider the matrix $\hat{A}$ consisting of the rows C_ℓ of C for which $\ell \in \mathbb{A}(x^\star)$.

Various cases for $x^\star$

$x^\star$ is in the interior of the dodecahedron.

- That is, $h(x^\star) = Cx^\star - d < \mathbf{0}$,
- $\mathbb{A}(x^\star) = \emptyset$,
- $\hat{A}$ has no rows, and
- $x^\star$ is a regular point by definition.

$x^\star$ is on a face of the dodecahedron but not on an edge or vertex.

- That is, exactly one constraint ℓ is binding,
- $\mathbb{A}(x^\star) = \{\ell\}$,
- $\hat{A} = C_\ell$, where C_ℓ is the ℓ-th row of C.
- The single row of $\hat{A}$ is linearly independent, since it is a single row that is not equal to the zero vector.

Various cases for $x^\star$, continued

$x^\star$ is on an edge but not a vertex of the dodecahedron.

- That is, exactly two constraints ℓ, ℓ' are binding,
- $\mathbb{A}(x^\star) = \{\ell, \ell'\}$, and $\hat{A} = \begin{bmatrix} C_\ell \\ C_{\ell'} \end{bmatrix}$.
- Since the corresponding two faces of the dodecahedron are not parallel then the two corresponding rows of C, namely C_ℓ and $C_{\ell'}$, are linearly independent.

$x^\star$ is on a vertex of the dodecahedron.

- That is, exactly three constraints ℓ, ℓ', and ℓ'' are binding,
- $\mathbb{A}(x^\star) = \{\ell, \ell', \ell''\}$, and $\hat{A} = \begin{bmatrix} C_\ell \\ C_{\ell'} \\ C_{\ell''} \end{bmatrix}$.
- The corresponding three faces are oblique to each other and therefore the three corresponding rows of C are linearly independent.

- In summary, every feasible point is a regular point of the constraints $h(x) \leq \mathbf{0}$.

Example, continued

- Now add an additional, redundant inequality constraint corresponding to a plane that just grazes the dodecahedron at one of its vertices, say $x^\star$.
- We augment an additional row to C to form $\tilde{C} \in \mathbb{R}^{13\times 3}$ and augment an additional entry to d to form $\tilde{d} \in \mathbb{R}^{13}$.
- We define the function $\tilde{h} : \mathbb{R}^3 \to \mathbb{R}^{13}$ to consist of the entries of h together with a thirteenth entry $\tilde{h}_{13} : \mathbb{R}^3 \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}^3, \tilde{h}_{13}(x) = \tilde{C}_{13}x - \tilde{d}_{13}.$$

- We now have that $\{x \in \mathbb{R}^3 | h(x) \leq \mathbf{0}\} = \{x \in \mathbb{R}^3 | \tilde{h}(x) \leq \mathbf{0}\}$.
- The vertex $x^\star$ is not a regular point of the constraints $\tilde{h}(x) \leq \mathbf{0}$ because there are *four* constraints active at $x^\star$ and the four corresponding rows of $\tilde{C}$ cannot be linearly independent in $\mathbb{R}^3$.
- $\{x \in \mathbb{R}^3 | h(x) \leq \mathbf{0}\}$ and $\{x \in \mathbb{R}^3 | \tilde{h}(x) \leq \mathbf{0}\}$ represent the *same* set.
- Therefore, whether or not a point $x^\star$ is a regular point of the constraints depends on the choice of representation of the constraints.

19.2 Optimality conditions

19.2.1 First-order necessary conditions

19.2.1.1 Analysis

Theorem 19.1 *Suppose that the functions* $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, *and* $h : \mathbb{R}^n \to \mathbb{R}^r$ *are partially differentiable with continuous partial derivatives. Let* $J : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ *and* $K : \mathbb{R}^n \to \mathbb{R}^{r\times n}$ *be the Jacobians of* g *and* h, *respectively. Consider Problem (19.1):*

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

Suppose that $x^\star \in \mathbb{R}^n$ *is a regular point of the constraints* $g(x)=\mathbf{0}$ *and* $h(x)\le\mathbf{0}$.

If $x^\star$ is a local minimizer of Problem (19.1) then:

$$\begin{aligned} \exists\lambda^\star \in \mathbb{R}^m, \exists\mu^\star \in \mathbb{R}^r \textit{ such that: } \nabla f(x^\star) + J(x^\star)^\dagger\lambda^\star + K(x^\star)^\dagger\mu^\star &= \mathbf{0}; \\ M^\star h(x^\star) &= \mathbf{0}; \\ g(x^\star) &= \mathbf{0}; \\ h(x^\star) &\leq \mathbf{0}; \textit{ and} \\ \mu^\star &\geq \mathbf{0}, \end{aligned} \tag{19.2}$$

where $M^\star = \operatorname{diag}\{\mu_\ell^\star\} \in \mathbb{R}^{r\times r}$. The vectors $\lambda^\star$ and $\mu^\star$ satisfying the conditions (19.2) are called the vectors of Lagrange multipliers for the constraints $g(x) = \mathbf{0}$ and $h(x) \leq \mathbf{0}$, respectively. The conditions that $M^\star h(x^\star) = \mathbf{0}$ are called the **complementary slackness conditions**. *They say that, for each ℓ, either the ℓ-th inequality constraint is binding or the ℓ-th Lagrange multiplier is equal to zero (or both).* □

Discussion

- As previously, we refer to the equality and inequality constraints in (19.2) as the **first-order necessary conditions** (or **FONC**) or the **Karush–Kuhn–Tucker** conditions.
- As in the case of non-linear equality constraints, the condition that $x^\star$ be a regular point of the constraints is again called a **constraint qualification**.
- In Section 19.3.1, we will see an alternative constraint qualification for the case of convex problems.

19.2.1.2 Lagrangian

- Recall Definition 3.2 of the **Lagrangian**.
- Analogously to the discussion in Section 17.1.1.4, by defining the Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}$ by:

$$\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^\dagger g(x) + \mu^\dagger h(x),$$

- we can again reproduce some of the first-order necessary conditions as:

$$\begin{aligned} \nabla_x \mathcal{L}(x^\star, \lambda^\star, \mu^\star) &= \mathbf{0}, \\ \nabla_\lambda \mathcal{L}(x^\star, \lambda^\star, \mu^\star) &= \mathbf{0}, \\ \nabla_\mu \mathcal{L}(x^\star, \lambda^\star, \mu^\star) &\leq \mathbf{0}. \end{aligned}$$

19.2.1.3 Example

- Recall the example non-linear program, Problem (2.19), from Section 2.3.2.3:

$$\min_{x \in \mathbb{R}^3} \{ f(x) | g(x) = \mathbf{0}, h(x) \leq 0 \},$$

- where $f : \mathbb{R}^3 \to \mathbb{R}$, $g : \mathbb{R}^3 \to \mathbb{R}^2$, and $h : \mathbb{R}^3 \to \mathbb{R}$ are defined by:

$$\begin{aligned} \forall x \in \mathbb{R}^3, f(x) &= (x_1)^2 + 2(x_2)^2, \\ \forall x \in \mathbb{R}^3, g(x) &= \begin{bmatrix} 2 - x_2 - \sin(x_3) \\ -x_1 + \sin(x_3) \end{bmatrix}, \\ \forall x \in \mathbb{R}^3, h(x) &= [\sin(x_3) - 0.5]. \end{aligned}$$

- We claim that $x^\star = \begin{bmatrix} 0.5 \\ 1.5 \\ \pi/6 \end{bmatrix}$, $\lambda^\star = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$, and $\mu^\star = [5]$ satisfy the first-order necessary conditions in Theorem 19.1.

- First, $x^\star$ is feasible.

$$
\begin{aligned}
\forall x \in \mathbb{R}^3, \nabla f(x) &= \begin{bmatrix} 2x_1 \\ 4x_2 \\ 0 \end{bmatrix}, \\
\forall x \in \mathbb{R}^3, J(x) &= \begin{bmatrix} 0 & -1 & -\cos(x_3) \\ -1 & 0 & \cos(x_3) \end{bmatrix}, \\
J(x^\star) &= \begin{bmatrix} 0 & -1 & -\cos(\pi/6) \\ -1 & 0 & \cos(\pi/6) \end{bmatrix}, \\
\forall x \in \mathbb{R}^3, K(x) &= \begin{bmatrix} 0 & 0 & \cos(x_3) \end{bmatrix}, \\
K(x^\star) &= \begin{bmatrix} 0 & 0 & \cos(\pi/6) \end{bmatrix}.
\end{aligned}
$$

- Note that $\hat{A} = \begin{bmatrix} J(x^\star) \\ K(x^\star) \end{bmatrix}$ has linearly independent rows so that $x^\star$ is a regular point of the constraints.

$$
\begin{aligned}
\nabla f(x^\star) + J(x^\star)^\dagger \lambda^\star + K(x^\star)^\dagger \mu^\star \\
&= \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ -1 & 0 \\ -\cos(\pi/6) & \cos(\pi/6) \end{bmatrix} \begin{bmatrix} 6 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \cos(\pi/6) \end{bmatrix} 5, \\
&= \mathbf{0}; \\
\mu^\star h(x^\star) &= [5] \times [0], \\
&= [0]; \\
g(x^\star) &= \mathbf{0}; \\
h(x^\star) &= [0], \\
&\leq [0]; \text{ and} \\
\mu^\star &= [5], \\
&\geq [0].
\end{aligned}
$$

- That is, $x^\star, \lambda^\star$, and $\mu^\star$ satisfy the first-order necessary conditions.

19.2.2 Second-order sufficient conditions

19.2.2.1 Analysis

Theorem 19.2 *Suppose that the functions* $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, *and* $h : \mathbb{R}^n \to \mathbb{R}^r$ *are twice partially differentiable with continuous second partial derivatives. Let* $J : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ *and* $K : \mathbb{R}^n \to \mathbb{R}^{r\times n}$ *be the Jacobians of* g *and* h*, respectively. Consider Problem (19.1):*

$$\min_{x\in\mathbb{R}^n}\{f(x)|g(x) = \mathbf{0}, h(x) \leq \mathbf{0}\},$$

and points $x^\star \in \mathbb{R}^n, \lambda^\star \in \mathbb{R}^m$, *and* $\mu^\star \in \mathbb{R}^r$. *Let* $M^\star = \operatorname{diag}\{\mu^\star_\ell\}$.

Suppose that:

$$\begin{aligned}
\nabla f(x^\star) + J(x^\star)^\dagger \lambda^\star + K(x^\star)^\dagger \mu^\star &= \mathbf{0}, \\
M^\star h(x^\star) &= \mathbf{0}, \\
g(x^\star) &= \mathbf{0}, \\
h(x^\star) &\leq \mathbf{0}, \\
\mu^\star &\geq \mathbf{0}, \textit{ and}
\end{aligned}$$

$$\nabla^2 f(x^\star) + \sum_{\ell=1}^{m} \lambda_\ell^\star \nabla^2 g_\ell(x^\star) + \sum_{\ell=1}^{r} \mu_\ell^\star \nabla^2 h_\ell(x^\star)$$

is positive definite on the null space:

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

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

Then $x^\star$ *is a strict local minimizer of Problem (19.1).* $\square$

Discussion

- The function $\nabla^2_{xx}\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}^{n\times n}$ defined by:

$$\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r,$$

$$\nabla^2_{xx}\mathcal{L}(x,\lambda,\mu) = \nabla^2 f(x) + \sum_{\ell=1}^{m} \lambda_\ell \nabla^2 g_\ell(x) + \sum_{\ell=1}^{r} \mu_\ell \nabla^2 h_\ell(x),$$

- is called the **Hessian of the Lagrangian**.
- In addition to the first-order necessary conditions, the second-order sufficient conditions require that:

  f, g, and h are twice partially differentiable with continuous second partial derivatives, and

  the Hessian of the Lagrangian evaluated at the minimizer and corresponding Lagrange multipliers, $\nabla^2_{xx}\mathcal{L}(x^\star,\lambda^\star,\mu^\star)$, is positive definite on the null space $\mathcal{N}_+$ defined in the theorem.
- Constraints ℓ for which $\mu^\star_\ell = 0$ and $h_\ell(x^\star) = 0$ are called **degenerate constraints**.

19.2.2.2 Example

- Continuing with Problem (2.19) from Sections 2.3.2.3 and 19.2.1.3, we note that f, g, and h are twice partially differentiable with continuous second partial derivatives.
- By the discussion in Section 19.2.1.3, the first-order necessary conditions are satisfied by $x^\star = \begin{bmatrix} 0.5 \\ 1.5 \\ \pi/6 \end{bmatrix}$, $\lambda^\star = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$, and $\mu^\star = [5]$.
- Also, $\mathbb{A}(x^\star) = \mathbb{A}_+(x^\star, \mu^\star) = \{1\}$.
- That is, the constraint is not degenerate.

$$\begin{aligned} \mathcal{N}_+ &= \{\Delta x \in \mathbb{R}^n | J(x^\star)\Delta x = \mathbf{0}, K_\ell(x^\star)\Delta x = 0, \forall \ell \in \mathbb{A}_+(x^\star, \mu^\star)\}, \\ &= \{\Delta x \in \mathbb{R}^n | J(x^\star)\Delta x = \mathbf{0}, K_1(x^\star)\Delta x = 0\}, \\ &= \{\mathbf{0}\}, \end{aligned}$$

- so that the Hessian of the Lagrangian $\nabla^2_{xx}\mathcal{L}(x^\star, \lambda^\star, \mu^\star)$ is positive definite on the null space $\mathcal{N}_+$.
- That is $x^\star$, $\lambda^\star$, and $\mu^\star$ satisfy the second-order sufficient conditions.

19.3 Convex problems

- Consider affine $g : \mathbb{R}^n \to \mathbb{R}^m$ and convex $h : \mathbb{R}^n \to \mathbb{R}^r$:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, h(x) \leq \mathbf{0} \}, \tag{19.3}$$

- where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.
- If $f : \mathbb{R}^n \to \mathbb{R}$ is convex on the feasible set then Problem (19.3) is convex.

19.3.1 First-order necessary conditions

19.3.1.1 Slater condition

- In the case of affine g and convex h, we can obtain first-order necessary conditions with an alternative constraint qualification to the assumption of regular constraints.
- In particular, we will assume that:

$$\{x \in \mathbb{R}^n | Ax = b, h(x) < \mathbf{0}\} \neq \emptyset. \tag{19.4}$$

- This alternative constraint qualification is called the **Slater condition**.
- The Slater condition was first introduced in Section 16.4.2.3 in the context of the interior point algorithm for linear inequality-constrained problems.
- We will see in Section 19.4.1.2 that we also need to make a similar assumption for applying the interior point algorithm to non-linearly constrained problems.

19.3.1.2 Analysis

Theorem 19.3 *Suppose that* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $h : \mathbb{R}^n \to \mathbb{R}^r$ *are partially differentiable with continuous partial derivatives and with* h *convex,* $A \in \mathbb{R}^{m \times n}$*, and* $b \in \mathbb{R}^m$*. Let* $K : \mathbb{R}^n \to \mathbb{R}^{r \times n}$ *be the Jacobian of* h*. Consider Problem (19.3) and suppose that the Slater condition (19.4) holds. If* $x^\star \in \mathbb{R}^n$ *is a local minimizer of Problem (19.3) then:*

$$\begin{aligned}
\exists \lambda^\star \in \mathbb{R}^m, \exists \mu^\star \in \mathbb{R}^r \text{ such that: } \nabla f(x^\star) + A^\dagger \lambda^\star + K(x^\star)^\dagger \mu^\star &= \mathbf{0};\\
M^\star h(x^\star) &= \mathbf{0};\\
Ax^\star &= b;\\
h(x^\star) &\leq \mathbf{0}; \text{ and}\\
\mu^\star &\geq \mathbf{0},
\end{aligned}$$

where $M^\star = \mathrm{diag}\{\mu_\ell^\star\} \in \mathbb{R}^{r \times r}$. $\square$

19.3.2 First-order sufficient conditions

19.3.2.1 Analysis

Theorem 19.4 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^r$ are partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. Let $K : \mathbb{R}^n \to \mathbb{R}^{r\times n}$ be the Jacobian of h. Consider Problem (19.3) and points $x^\star \in \mathbb{R}^n$, $\lambda^\star \in \mathbb{R}^m$, and $\mu^\star \in \mathbb{R}^r$. Let $M^\star = \mathrm{diag}\{\mu_\ell^\star\}$. Suppose that:*

(i) *h is convex,*

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

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

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

(v) *$Ax^\star = b$ and $h(x^\star) \leq \mathbf{0}$, and*

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

Then $x^\star$ is a global minimizer of Problem (19.3).

Proof The proof is very similar to the proofs of Theorem 16.3 in Chapter 16 and of Theorem 17.3 in Chapter 17. □

19.3.2.2 Example

$$\begin{aligned}\forall x \in \mathbb{R}^2, f(x) &= x_1 + x_2,\\ \forall x \in \mathbb{R}^2, h(x) &= (x_1)^2 + (x_2)^2 - 2.\end{aligned}$$

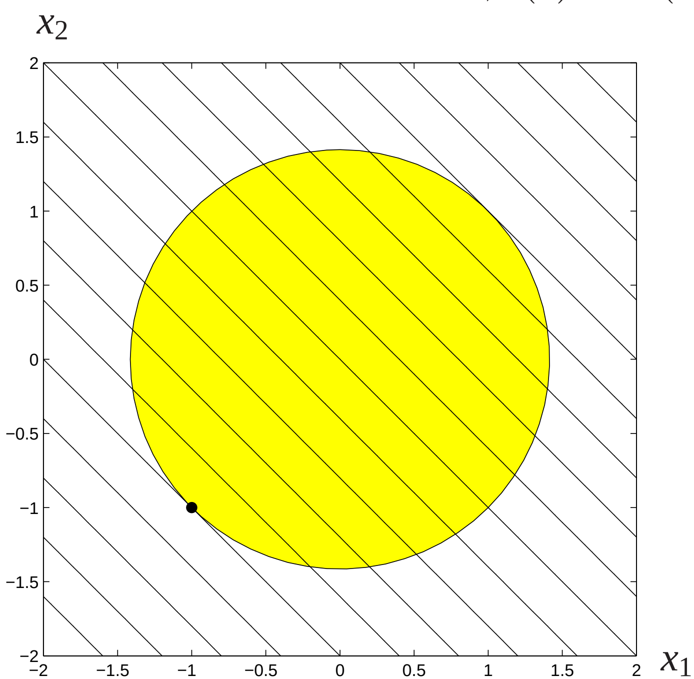

Fig. 19.2. Contour sets of objective function defined in Section 19.3.2.2 with feasible set shaded. The heights of the contours decrease to the left and down. The minimizer, $x^\star = -\mathbf{1}$, is indicated with the $\bullet$.

Example, continued

- f and h are partially differentiable with continuous partial derivatives and convex.
- We claim that $x^\star = -\mathbf{1}$ is the global minimizer with Lagrange multiplier $\mu^\star = [0.5]$:

$$
\begin{aligned}
\forall x \in \mathbb{R}^2, \nabla f(x) &= \mathbf{1}, \\
\forall x \in \mathbb{R}^2, K(x) &= \begin{bmatrix} 2x_1 & 2x_2 \end{bmatrix}, \\
K(x^\star) &= \begin{bmatrix} -2 & -2 \end{bmatrix}, \\
\nabla f(x^\star) + K(x^\star)^\dagger \mu^\star &= \mathbf{1} + \begin{bmatrix} -2 & -2 \end{bmatrix}^\dagger \times [0.5], \\
&= \mathbf{0}; \\
\mu^\star h(x^\star) &= 0; \\
h(x^\star) &= [0], \\
&\leq [0]; \text{ and} \\
\mu^\star &= [0.5], \\
&\geq [0].
\end{aligned}
$$

19.3.3 Duality

19.3.3.1 Dual function

Analysis

- If f and h are convex and g is affine then $\mathcal{L}(\bullet,\lambda,\mu)$ is convex for $\mu \geq \mathbf{0}$ and so $x^\star$ is a global minimizer of $\mathcal{L}(\bullet,\lambda^\star,\mu^\star)$.
- For Problem (19.3), the dual function $\mathcal{D} : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R} \cup \{-\infty\}$ is defined by:

$$\forall \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r, \mathcal{D}(\lambda,\mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x,\lambda,\mu). \tag{19.5}$$

- The effective domain of $\mathcal{D}$ is:

$$\mathbb{E} = \left\{ \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r} \middle| \mathcal{D}(\lambda,\mu) > -\infty \right\}.$$

- Recall that by Theorem 3.12, $\mathbb{E}$ is convex and $\mathcal{D}$ is concave on $\mathbb{E}$.

Example

- Continuing with the example problem from Section 19.3.2.2, the Lagrangian $\mathcal{L}:\mathbb{R}^2\times\mathbb{R}\to\mathbb{R}$ for this problem is defined by:

$$\begin{aligned}\forall x\in\mathbb{R}^2,\forall\mu\in\mathbb{R},\mathcal{L}(x,\mu) &= f(x)+\mu^\dagger h(x),\\ &= x_1+x_2+\mu((x_1)^2+(x_2)^2-2).\end{aligned}$$

- For $\mu>0$, the Lagrangian $\mathcal{L}(\bullet,\mu)$ is strictly convex and therefore, by Corollary 10.6, the first-order necessary conditions $\nabla_x\mathcal{L}(x,\mu)=0$ are sufficient for minimizing $\mathcal{L}(\bullet,\mu)$ and, moreover, a minimizer exists, so that the inf in the definition of $\mathcal{D}$ can be replaced by min.
- Furthermore, there is a unique minimizer $x^{(\mu)}$ corresponding to each value of $\mu>0$.

Example, continued

- In particular, we have:

$$
\begin{aligned}
\forall x \in \mathbb{R}^2, \forall \mu \in \mathbb{R}, \nabla_x \mathcal{L}(x,\mu) &= \nabla f(x) + K(x)^\dagger \mu, \\
&= \begin{bmatrix} 1+2\mu x_1 \\ 1+2\mu x_2 \end{bmatrix}, \\
\forall \mu \in \mathbb{R}_{++}, x^{(\mu)} &= \begin{bmatrix} -1/(2\mu) \\ -1/(2\mu) \end{bmatrix}, \\
\forall \mu \in \mathbb{R}_{++}, \mathcal{D}(\mu) &= -\frac{1}{2\mu} - 2\mu.
\end{aligned}
$$

- On the other hand, if $\mu \leq 0$ then the objective in the dual function is unbounded below.
- Consequently, the effective domain is $\mathbb{E} = \mathbb{R}_{++}$.

19.3.3.2 Dual problem

Analysis

- The **dual problem**:

$$\max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}} \{\mathcal{D}(\lambda,\mu) | \mu \geq \mathbf{0}\}, \tag{19.6}$$

- where $\mathcal{D} : \mathbb{E} \to \mathbb{R}$ is the dual function defined in (19.5).
- Problem (19.3) is called the **primal problem** in this context to distinguish it from Problem (19.6).

Theorem 19.5 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$ are convex and partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$. Consider the primal problem, Problem (19.3):*

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

and suppose that the Slater condition (19.4) holds. Also, consider the dual problem, Problem (19.6). We have that:

(i) *If the primal problem possesses a minimum then the dual problem possesses a maximum and the optima are equal. That is:*

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, h(x) \leq \mathbf{0} \} = \max_{\left[\begin{smallmatrix}\lambda \\ \mu\end{smallmatrix}\right] \in \mathbb{E}} \{ \mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0} \}.$$

(ii) *If:*

- $\begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{E}$,
- $\min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu)$ *exists, and*
- f *and* h *are twice partially differentiable with continuous*

second partial derivatives, $\nabla^2 f$ is positive definite, and $\nabla^2 h_\ell, \ell = 1, \ldots, r$, are all positive definite,

then $\mathcal{D}$ is partially differentiable at $\begin{bmatrix} \lambda \\ \mu \end{bmatrix}$ with continuous partial derivatives and:

$$\nabla \mathcal{D}(\lambda, \mu) = \begin{bmatrix} Ax^{(\lambda,\mu)} - b \\ h(x^{(\lambda,\mu)}) \end{bmatrix}. \tag{19.7}$$

$\square$

Discussion

- It is possible for $\mathcal{D}$ to not be partially differentiable at a point $\begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{E}$ if:

  $\mathcal{L}(\bullet, \lambda, \mu)$ is bounded below (so that $\inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu) \in \mathbb{R}$) yet the minimum $\min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu)$ does not exist, or

  there are multiple minimizers of $\min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu)$.

Corollary 19.6 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^r$ be twice partially differentiable with continuous second partial derivatives, $\nabla^2 f$ be positive definite, and $\nabla^2 h_\ell, \ell = 1, \ldots, r$, all be positive definite; $A \in \mathbb{R}^{m\times n}$; and $b \in \mathbb{R}^m$. Consider Problem (19.3):*

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

the Lagrangian of this problem, and the effective domain $\mathbb{E}$ of the dual function. If:

- *the effective domain $\mathbb{E}$ contains $\mathbb{R}^m \times \mathbb{R}^r_+$, and*
- *for each $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^r_+$, $\min_{x\in\mathbb{R}^n} \mathcal{L}(x, \lambda, \mu)$ exists,*

then necessary and sufficient conditions for $\begin{bmatrix} \lambda^\star \\ \mu^\star \end{bmatrix} \in \mathbb{R}^{m+r}$ to be the maximizer of the dual problem:

$$\max_{\left[\begin{smallmatrix}\lambda \\ \mu\end{smallmatrix}\right]\in\mathbb{E}} \{ \mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0} \},$$

are:

$$\begin{aligned} M^\star h(x^{(\lambda^\star,\mu^\star)}) &= \mathbf{0}; \\ Ax^{(\lambda^\star,\mu^\star)} &= b; \\ h(x^{(\lambda^\star,\mu^\star)}) &\leq \mathbf{0}; \textit{ and} \\ \mu^\star &\geq \mathbf{0}, \end{aligned}$$

where $\{x^{(\lambda^\star,\mu^\star)}\} = \operatorname{argmin}_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda^\star,\mu^\star)$ *and* $M^\star = \operatorname{diag}\{\mu^\star_\ell\}$. *Moreover, if* $\lambda^\star$ *and* $\mu^\star$ *maximize the dual problem then* $x^{(\lambda^\star,\mu^\star)}$, $\lambda^\star$, *and* $\mu^\star$ *satisfy the first-order necessary conditions for Problem (19.3).*

Proof The proof is very similar to the proof of Corollary 17.5 in Chapter 17. □

Discussion

- Theorem 19.5 shows that an alternative approach to finding the minimum of Problem (19.3) involves finding the *maximum* of the dual function over $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^r_+$.
- Theorem 3.12 shows that the dual function has at most one local maximum.
- To seek the maximum of $\mathcal{D}(\lambda, \mu)$ over $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r_+$, we can, for example, utilize the value of the gradient of $\mathcal{D}$ from (19.7) as part of an active set or interior point algorithm.

Example

- Continuing with the dual of the example problem from Sections 19.3.2.2 and 19.3.3.1, the effective domain is $\mathbb{E} = \mathbb{R}_{++}$ and the dual function $\mathcal{D} : \mathbb{R}_{++} \to \mathbb{R}$ is:

$$\begin{aligned}
\forall \mu \in \mathbb{R}_{++}, \mathcal{D}(\mu) &= -\frac{1}{2\mu} - 2\mu, \\
\forall \mu \in \mathbb{R}_{++}, \nabla\mathcal{D}(\mu) &= \frac{1}{2(\mu)^2} - 2, \\
\forall \mu \in \mathbb{R}_{++}, \nabla^2\mathcal{D}(\mu) &= -\frac{1}{4(\mu)^3}, \\
&< 0.
\end{aligned}$$

- We cannot apply Corollary 19.6 directly because $\mathbb{E} = \mathbb{R}_{++}$ does not contain $\mathbb{R}_{+}$.
- However, by inspection of $\mathcal{D}$, $\mu^\star = [0.5]$ maximizes the dual over $\mathbb{E}$.
- Moreover, the corresponding minimizer of the Lagrangian, $x^{(\mu^\star)}$, together with $\mu^\star$ satisfy the first-order necessary conditions for the primal problem.

Discussion

- It is essential in Theorem 19.5 for f and h to be convex on the *whole* of $\mathbb{R}^n$, not just on the feasible set.
- This is because the inner minimization of $\mathcal{L}(\bullet, \lambda, \mu)$ is taken over the whole of $\mathbb{R}^n$.
- We generally require strict convexity of f and h to ensure that there are not multiple minimizers of the Lagrangian.
- The issues are similar to the discussion in Section 17.2.2.2.
- Problem (19.6) is non-negatively constrained of the form of Problem (16.1) and so we can apply essentially the same algorithms as we developed for Problem (16.1).
- We will take this approach in Section 19.4.2.

19.3.3.3 *Partial duals*

- As in Section 17.2.2.4, it is also possible to take the partial dual with respect to some of the equality and some of the inequality constraints.

19.4 Approaches to finding minimizers

19.4.1 Primal algorithm

19.4.1.1 Transformation

- To handle the inequality constraints involving h we consider the following problem:

$$\min_{x\in\mathbb{R}^n, w\in\mathbb{R}^r} \{f(x) | g(x) = \mathbf{0}, h(x) + w = \mathbf{0}, w \geq \mathbf{0}\}. \tag{19.8}$$

- By Theorem 3.8, Problems (19.1) and (19.8) are equivalent.

19.4.1.2 Primal–dual interior point algorithm

Barrier objective and problem

- Given a barrier function $f_{\mathrm{b}} : \mathbb{R}^r_{++} \to \mathbb{R}$ and a barrier parameter $t \in \mathbb{R}_{++}$, we form the **barrier objective** $\phi : \mathbb{R}^n \times \mathbb{R}^r_{++} \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}^n, \forall w \in \mathbb{R}^r_{++}, \phi(x,w) = f(x) + t f_{\mathrm{b}}(w).$$

- Instead of solving Problem (19.8), we will consider solving the **barrier problem**:

$$\min_{x \in \mathbb{R}^n, w \in \mathbb{R}^r} \{\phi(x,w) | g(x) = \mathbf{0}, h(x) + w = \mathbf{0}, w > \mathbf{0}\}. \qquad (19.9)$$

- We then decrease the barrier parameter t.

Slater condition

- Analogously to the discussion in Sections 16.4.2.2 and 17.3.1.2, we must assume that Problem (19.9) is feasible.
- That is, we assume that $\{x \in \mathbb{R}^n | g(x) = \mathbf{0}, h(x) < \mathbf{0}\} \neq \emptyset$.
- We again call this the **Slater condition**.

Equality-constrained problem

- To solve Problem (19.9), we partially ignore the inequality constraints and the domain of the barrier function and seek a solution to the following non-linear equality-constrained problem:

$$\min_{x\in\mathbb{R}^n, w\in\mathbb{R}^r} \{\phi(x,w) | g(x) = \mathbf{0}, h(x) + w = \mathbf{0}\}, \tag{19.10}$$

- which has first-order necessary conditions:

$$\nabla f(x) + J(x)^\dagger \lambda + K(x)^\dagger \mu = \mathbf{0}, \tag{19.11}$$

$$g(x) = \mathbf{0}, \tag{19.12}$$

$$h(x) + w = \mathbf{0}, \tag{19.13}$$

$$t\nabla f_{\mathrm{b}}(w) + \mu = \mathbf{0}, \tag{19.14}$$

- where J and K are the Jacobians of g and h, respectively, and λ and μ are the dual variables on the constraints $g(x) = \mathbf{0}$ and $h(x) + w = \mathbf{0}$, respectively.

Logarithmic barrier function

- We again use the logarithmic barrier function:

$$\begin{aligned}\forall w \in \mathbb{R}^r_{++}, f_{\mathrm{b}}(w) &= -\sum_{\ell=1}^{r} \ln(w_\ell), \\ \forall w \in \mathbb{R}^r_{++}, \nabla f_{\mathrm{b}}(w) &= -[W]^{-1}\mathbf{1},\end{aligned}$$

- where $W = \mathrm{diag}\{w_\ell\} \in \mathbb{R}^{r\times r}$.
- Substituting the expression for ∇f_{b} into (19.14) and re-arranging, we again obtain:

$$W\mu - t\mathbf{1} = \mathbf{0}. \tag{19.15}$$

19.4.1.3 Newton–Raphson method

- The Newton–Raphson step direction to solve (19.15) and (19.11)–(19.13) is:

$$\begin{bmatrix} M^{(\nu)} & \mathbf{0} & \mathbf{0} & W^{(\nu)} \\ \mathbf{0} & \nabla^2_{xx}\mathcal{L}(x^{(\nu)},\lambda^{(\nu)},\mu^{(\nu)}) & J(x^{(\nu)})^\dagger & K(x^{(\nu)})^\dagger \\ \mathbf{0} & J(x^{(\nu)}) & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & K(x^{(\nu)}) & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta w^{(\nu)} \\ \Delta x^{(\nu)} \\ \Delta\lambda^{(\nu)} \\ \Delta\mu^{(\nu)} \end{bmatrix} = \begin{bmatrix} -W^{(\nu)}\mu^{(\nu)} + t\mathbf{1} \\ -\nabla f(x^{(\nu)}) - J(x^{(\nu)})^\dagger\lambda^{(\nu)} - K(x^{(\nu)})^\dagger\mu^{(\nu)} \\ -g(x^{(\nu)}) \\ -h(x^{(\nu)}) \end{bmatrix},$$

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

Newton–Raphson method, continued

- We can re-arrange the equations to make them symmetric and use block pivoting on the top left-hand block of the matrix since the top left-hand block is diagonal.
- This results in a system that is similar to (14.12), except that a diagonal block of the form $[M^{(\nu)}]^{-1}W^{(\nu)}$ is added to the Hessian of the Lagrangian.
- Issues regarding solving the first-order necessary conditions, such as factorization of the indefinite coefficient matrix, approximate solution of the conditions, sparsity, the merit function, step-size selection, and feasibility, are similar to those described in Sections 14.3.1 and 16.4.3.3.

19.4.1.4 Other issues

Adjustment of barrier parameter

- To reduce the barrier parameter, we can again use the approach described in Section 16.4.4 of Chapter 16.

Initial guess

- The effort to find a feasible initial guess may be significant.
- An alternative is to begin with $w^{(0)} > \mathbf{0}, x^{(0)}, \lambda^{(0)}, \mu^{(0)} > \mathbf{0}$ that do not necessarily satisfy the equality constraints $g(x) = \mathbf{0}$ nor $h(x) + w = \mathbf{0}$.
- Feasibility is approached during the course of iterations from this **infeasible start**.

Stopping criterion

- We can develop a stopping criterion based on duality using Theorem 3.13.
- If f or h are non-quadratic or g is non-linear, however, we can typically only approximately evaluate the dual function.

19.4.2 Dual algorithm

- Problem (19.6):

$$\max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}} \{\mathcal{D}(\lambda,\mu) | \mu \geq \mathbf{0}\},$$

- has non-negativity constraints.
- If the dual function can be evaluated conveniently, then the algorithms from Section 16.3 and 16.4 for non-negativity constraints can be applied to the dual problem.
- For example, if the objective and inequality constraint function are quadratic and strictly convex and the equality constraints are linear then the dual function can be evaluated through the solution of a linear equation.
- A dual algorithm can be particularly attractive if there are only a few constraints or if a **partial dual** is taken with respect to only some of the constraints.

19.5 Sensitivity

19.5.1 Analysis

- We consider a general and a special case of sensitivity analysis for Problem (19.1).
- For the general case, we suppose that the objective f, equality constraint function g, and inequality constraint function h are parameterized by a parameter $\chi \in \mathbb{R}^s$.
- We imagine that we have solved the non-linear inequality-constrained minimization problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x;\chi) | g(x;\chi) = \mathbf{0}, h(x;\chi) \leq \mathbf{0} \}, \tag{19.16}$$

- for a base-case value of the parameters, say $\chi = \mathbf{0}$, to find the base-case solution $x^\star$ and the base-case Lagrange multipliers $\lambda^\star$ and $\mu^\star$.
- We now consider the sensitivity of the minimum of Problem (19.16) to variation of the parameters about $\chi = \mathbf{0}$.

Analysis, continued

- We also specialize to the case where only the right-hand sides of the equality and inequality constraints vary.
- That is, we return to the special case where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^r$ are not explicitly parameterized.
- We now consider perturbations $\gamma \in \mathbb{R}^m$ and $\eta \in \mathbb{R}^r$ and the problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = -\gamma, h(x) \leq -\eta \}. \tag{19.17}$$

- For the parameter values $\gamma = \mathbf{0}$ and $\eta = \mathbf{0}$, Problem (19.17) is the same as Problem (19.1).
- We consider the sensitivity of the minimum of Problem (19.17) to variation of the parameters about $\gamma = \mathbf{0}$ and $\eta = \mathbf{0}$.

Corollary 19.7 *Consider Problem (19.16) and suppose that the functions $f : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$, and $h : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^r$ are twice partially differentiable with continuous second partial derivatives. Also consider Problem (19.17) and suppose that the functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^r$ are twice partially differentiable with continuous second partial derivatives. Suppose that $x^\star \in \mathbb{R}^n$, $\lambda^\star \in \mathbb{R}^m$, and $\mu^\star \in \mathbb{R}^r$ satisfy:*

- *the second-order sufficient conditions for Problem (19.16) for the value of parameters $\chi = \mathbf{0}$, and*
- *the second-order sufficient conditions for Problem (19.17) for the value of parameters $\gamma = \mathbf{0}$ and $\eta = \mathbf{0}$.*

In particular:

- *$x^\star$ is a local minimizer of Problem (19.16) for $\chi = \mathbf{0}$, and*
- *$x^\star$ is a local minimizer of Problem (19.17) for $\gamma = \mathbf{0}$ and $\eta = \mathbf{0}$,*

in both cases with associated Lagrange multipliers $\lambda^\star$ and $\mu^\star$. Moreover, suppose that $x^\star$ is a regular point of the constraints for the base-case problems and that there are no degenerate constraints at the base-case solution.

Then, for values of χ in a neighborhood of the base-case value of the parameters $\chi = \mathbf{0}$, there is a local minimum and corresponding local minimizer and Lagrange multipliers for Problem (19.16). Moreover, the local minimum, local minimizer, and Lagrange multipliers are partially differentiable with respect to χ and have continuous partial derivatives in this neighborhood. The sensitivity of the local minimum $f^\star$ to χ, evaluated at the base-case $\chi = \mathbf{0}$, is given by:

$$\frac{\partial f^\star}{\partial \chi}(\mathbf{0}) = \frac{\partial \mathcal{L}}{\partial \chi}(x^\star, \lambda^\star, \mu^\star; \mathbf{0}),$$

where $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}$ is the **parameterized Lagrangian** *defined by:*

$$\begin{aligned} &\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \forall \chi \in \mathbb{R}^s, \\ &\mathcal{L}(x, \lambda, \mu; \chi) \;=\; f(x;\chi) + \lambda^\dagger g(x;\chi) + \mu^\dagger h(x;\chi). \end{aligned}$$

Furthermore, for values of γ and η in a neighborhood of the base-case value of the parameters $\gamma = \mathbf{0}$ and $\eta = \mathbf{0}$, there is a local minimum and corresponding local minimizer and Lagrange multipliers for Problem (19.17). Moreover, the local minimum, local minimizer, and

Lagrange multipliers are partially differentiable with respect to γ *and* η *and have continuous partial derivatives. The sensitivities of the local minimum to* γ *and* η*, evaluated at the base-case* $\gamma = \mathbf{0}$ *and* $\eta = \mathbf{0}$*, are equal to* $[\lambda^\star]^\dagger$ *and* $[\mu^\star]^\dagger$*, respectively.* $\Box$

19.5.2 Discussion

- We can again interpret the Lagrange multipliers as the sensitivity of the minimum to the right-hand side of the equality constraints and inequality constraints.

19.5.3 Example

- Continuing with Problem (2.19) from Sections 2.3.2.3, 19.2.1.3, and 19.2.2.2, we have already verified that the second-order sufficient conditions are satisfied at the base-case solution, that $x^\star$ is a regular point of the constraints, and that there are no degenerate constraints.
- Suppose that the first entry in the equality constraint changed to $2 - x_2 - \sin(x_3) = -\gamma_1$ and that the inequality constraint changed to $\sin(x_3) - 0.5 \leq -\eta$.
- By Corollary 19.7, if γ_1 and η are small enough the change in the minimum is given approximately by $\lambda_1^\star \gamma_1 + \mu^\star \eta = 6\gamma_1 + 5\eta$.

19.6 Summary

- In this chapter we have considered problems with non-linear equality and inequality constraints, providing optimality conditions.
- We considered the convex case and sketched application of the primal–dual interior point method and dual algorithm to these problems.
- Finally, we provided sensitivity analysis.

20

Solution of the non-linear inequality-constrained case studies

- Optimal margin pattern classification (Section 20.1),
- Sizing of interconnects in integrated circuits (Section 20.2), and
- Optimal power flow (Section 20.3).

20.1 Optimal margin pattern classification

- The first transformation in Section 18.4.1.1 yielded the maximization Problem (18.3), which we recast into a minimization problem as:

$$\min_{z \in \mathbb{R}, x \in \mathbb{R}^n} \left\{ -z \,\middle|\, \mathbf{1}z + Cx \leq \mathbf{0}, \|\beta\|_2^2 \leq 1 \right\}. \tag{20.1}$$

- This problem has a linear objective, r linear inequality constraints, and one convex quadratic inequality constraint.
- This can be solved using the algorithms developed in Section 19.4.
- The dual of Problem (20.1) is equivalent to a quadratic program.

20.2 Sizing of interconnects in integrated circuits

20.2.1 Problem and analysis

- Problem (15.19):

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

- used the Elmore delay approximation $\tilde{h}$ to the actual delay h.
- This problem has a linear objective but has inequality constraints defined in terms of functions that are, in general, non-convex.
- However, as discussed in Section 15.5.4, the objective and constraint functions are **posynomial**.

Problem and analysis, continued

- Each posynomial function can be transformed into a convex function through a transformation involving the exponential of each entry of the decision vector and the logarithm of the function.
- The transformed problem is convex and therefore possesses at most one local minimum.
- Because the transformation of the decision vector is one-to-one and onto and the transformations of the objective and constraints are monotonically increasing then, by Theorems 3.1, 3.5, and 3.9, the original problem also possesses at most one local minimum.

20.2.2 Algorithms

20.2.2.1 Primal algorithm

- In principle, we can apply the optimization techniques developed in Section 19.4 to either the original problem or the transformed problem and be guaranteed that any local minimum is the global minimum.
- However, since the inequality constraint functions are not convex in the original problem, the Hessian of the Lagrangian for the original problem will typically not be positive definite and so we can expect that pivots will be modified significantly during factorization, potentially retarding the progress towards the minimizer.

20.2.2.2 Dual algorithm

- Since the transformed problem is convex, we can also dualize the transformed problem.
- Further transformation of the dual problem is possible to simplify the dual problem to having linear constraints.

20.2.2.3 Accurate delay model

- Recall Problem (15.20):

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

- which used the more accurate delay model h instead of the Elmore delay model $\tilde{h}$.
- In general, we cannot expect that h will have any particular functional form.
- However, $\tilde{h}$ may be a reasonable approximation of h.
- The algorithms we have described typically require both function evaluations *and* derivative evaluations.
- To solve the problem with the more accurate delay model, we can combine accurate delay values calculated according to h with approximate first and second derivatives calculated from the functional form of $\tilde{h}$.
- Furthermore, we can apply such an algorithm to the original problem or to the transformed problem.

20.2.3 Changes

- Corollary 19.7 and extensions can be used to estimate the changes in area and width due to changes in parameters and allowed delays.

20.3 Optimal power flow

- Recall Problem (15.23):

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

- This problem has non-linear objective and equality and inequality constraint functions.
- Under certain assumptions the problem is equivalent to a convex problem.
- We can use the primal–dual interior point algorithm sketched in Section 19.4.1 to solve it.
- Corollary 19.7 and extensions can be used to estimate the changes in costs due to changes in demand and changes in line and generator capacities.