1.5.1 Geometric Optimality Conditions

Definition 1.30 (Feasible Direction). *Let S be a nonempty set in $\mathbb{R}^{n_x}$. A vector $\mathbf{d} \in \mathbb{R}^{n_x}$, $\mathbf{d} \neq \mathbf{0}$, is said to be a* feasible direction *at $\bar{\mathbf{x}} \in \operatorname{cl}(S)$ if*

$$\exists \delta > 0 \text{ such that } \bar{\mathbf{x}} + \eta \mathbf{d} \in S \quad \forall \eta \in (0, \delta).$$

Moreover, the cone of feasible directions *at $\bar{\mathbf{x}}$, denoted by $\mathcal{D}(\bar{\mathbf{x}})$, is given by*

$$\mathcal{D}(\bar{\mathbf{x}}) := \{\mathbf{d} \neq \mathbf{0} : \exists \delta > 0 \text{ such that } \bar{\mathbf{x}} + \eta \mathbf{d} \in S \quad \forall \eta \in (0, \delta)\}.$$

From the above definition and Lemma 1.21, it is clear that a small movement from $\bar{\mathbf{x}}$ along a direction $\mathbf{d} \in \mathcal{D}(\bar{\mathbf{x}})$ leads to feasible points, whereas a similar movement along a direction $\mathbf{d} \in \mathcal{F}_0(\bar{\mathbf{x}})$ leads to solutions of improving objective value (see Definition 1.20). As shown in Theorem 1.31 below, a (geometric) necessary condition for local optimality is that: *"Every improving direction is not a feasible direction."* This fact is illustrated in Fig. 1.9., where both the half-space $\mathcal{F}_0(\bar{\mathbf{x}})$ and the cone $\mathcal{D}(\bar{\mathbf{x}})$ (see Definition A.10) are translated from the origin to $\bar{\mathbf{x}}$ for clarity.

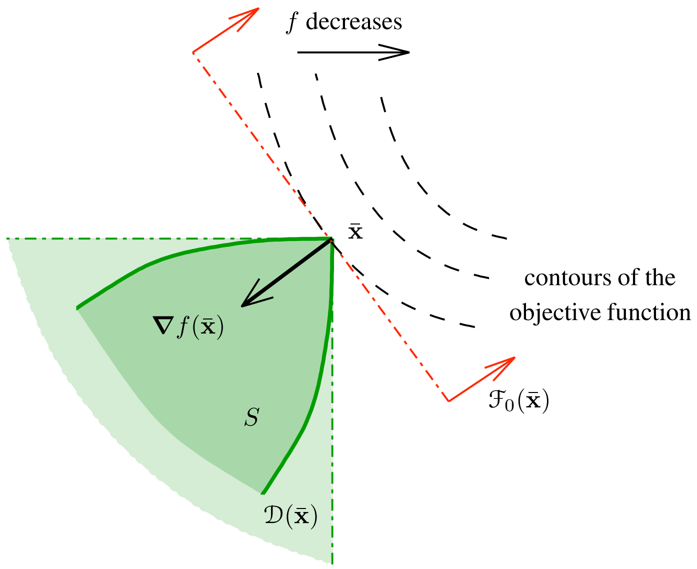

Figure 1.9. Illustration of the (geometric) necessary condition $\mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}(\bar{\mathbf{x}}) = \emptyset$.

Theorem 1.31 (Geometric Necessary Condition for a Local Minimum). *Let S be a nonempty set in $\mathbb{R}^{n_x}$, and let $f : \mathbb{R}^{n_x} \to \mathbb{R}$ be a differentiable function. Suppose that $\bar{\mathbf{x}}$ is a local minimizer of the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. Then, $\mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}(\bar{\mathbf{x}}) = \emptyset$.*

Proof. By contradiction, suppose that there exists a vector $\mathbf{d} \in \mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}(\bar{\mathbf{x}})$, $\mathbf{d} \neq \mathbf{0}$. Then, by Lemma 1.21,

$$\exists \delta_1 > 0 \text{ such that } f(\bar{\mathbf{x}} + \eta \mathbf{d}) < f(\bar{\mathbf{x}}) \quad \forall \eta \in (0, \delta_1).$$

Moreover, by Definition 1.30,

$$\exists \delta_2 > 0 \text{ such that } \bar{\mathbf{x}} + \eta \mathbf{d} \in S \quad \forall \eta \in (0, \delta_2).$$

Hence,

$$\exists \mathbf{x} \in \mathcal{B}_\eta(\bar{\mathbf{x}}) \cap S \text{ such that } f(\bar{\mathbf{x}} + \eta \mathbf{d}) < f(\bar{\mathbf{x}}),$$

for every $\eta \in (0, \min\{\delta_1, \delta_2\})$, which contradicts the assumption that $\bar{\mathbf{x}}$ is a local minimum of f on S (see Definition 1.9). □

1.5.2 KKT Conditions

We now specify the feasible region as

$$S := \{\mathbf{x} : g_i(\mathbf{x}) \leq 0 \quad \forall i = 1, \ldots, n_g\},$$

where $g_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_g$, are continuous functions. In the geometric optimality condition given by Theorem 1.31, $\mathcal{D}(\bar{\mathbf{x}})$ is the cone of feasible directions. From a practical viewpoint, it is desirable to convert this geometric condition into a more usable condition involving algebraic equations. As Lemma 1.33 below indicates, we can define a cone $\mathcal{D}_0(\bar{\mathbf{x}})$ in terms of the gradients of the *active constraints* at $\bar{\mathbf{x}}$, such that $\mathcal{D}_0(\bar{\mathbf{x}}) \subseteq \mathcal{D}(\bar{\mathbf{x}})$. For this, we need the following:

Definition 1.32 (Active Constraint, Active Set). *Let $g_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_g$, and consider the set $S := \{\mathbf{x} : g_i(\mathbf{x}) \leq 0, i = 1, \ldots, n_g\}$. Let $\bar{\mathbf{x}} \in S$ be a feasible point. For each $i = 1, \ldots, n_g$, the constraint g_i is said to* active *or* binding *at $\bar{\mathbf{x}}$ if $g_i(\bar{\mathbf{x}}) = 0$; it is said to be* inactive *at $\bar{\mathbf{x}}$ if $g_i(\bar{\mathbf{x}}) < 0$. Moreover,*

$$\mathcal{A}(\bar{\mathbf{x}}) \ := \ \{i : g_i(\bar{\mathbf{x}}) = 0\},$$

denotes the set of active constraints *at $\bar{\mathbf{x}}$.*

Lemma 1.33 (Algebraic Characterization of a Feasible Direction). *Let $g_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_g$ be differentiable functions, and consider the set $S := \{\mathbf{x} : g_i(\mathbf{x}) \leq 0, i = 1, \ldots, n_g\}$. For any feasible point $\bar{\mathbf{x}} \in S$, we have*

$$\mathcal{D}_0(\bar{\mathbf{x}}) \ := \ \{\mathbf{d} : \boldsymbol{\nabla} g_i(\bar{\mathbf{x}})^{\mathsf{T}} \mathbf{d} < 0 \quad \forall i \in \mathcal{A}(\bar{\mathbf{x}})\} \ \subseteq \ \mathcal{D}(\bar{\mathbf{x}}).$$

Proof. Suppose $\mathcal{D}_0(\bar{\mathbf{x}})$ is nonempty, and let $\mathbf{d} \in \mathcal{D}_0(\bar{\mathbf{x}})$. Since $\boldsymbol{\nabla} g_i(\bar{\mathbf{x}})\mathbf{d} < 0$ for each $i \in \mathcal{A}(\bar{\mathbf{x}})$, then by Lemma 1.21, $\mathbf{d}$ is a descent direction for g_i at $\bar{\mathbf{x}}$, i.e.,

$$\exists \delta_2 > 0 \text{ such that } g_i(\bar{\mathbf{x}} + \eta \mathbf{d}) < g_i(\bar{\mathbf{x}}) = 0 \quad \forall \eta \in (0, \delta_2), \ \forall i \in \mathcal{A}(\bar{\mathbf{x}}).$$

Furthermore, since $g_i(\bar{\mathbf{x}}) < 0$ and g_i is continuous at $\bar{\mathbf{x}}$ (for it is differentiable) for each $i \notin \mathcal{A}(\bar{\mathbf{x}})$,

$$\exists \delta_1 > 0 \text{ such that } g_i(\bar{\mathbf{x}} + \eta \mathbf{d}) < 0 \quad \forall \eta \in (0, \delta_1), \ \forall i \notin \mathcal{A}(\bar{\mathbf{x}}).$$

Furthermore, Overall, it is clear that the points $\bar{\mathbf{x}} + \eta \mathbf{d}$ are in S for all $\eta \in (0, \min\{\delta_1, \delta_2\})$. Hence, by Definition 1.30, $\mathbf{d} \in \mathcal{D}(\bar{\mathbf{x}})$. □

Remark 1.34. This lemma together with Theorem 1.31 directly leads to the result that $\mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}_0(\bar{\mathbf{x}}) = \emptyset$ for any local solution point $\bar{\mathbf{x}}$, i.e.,

$$\arg\min\{f(\mathbf{x}) : \mathbf{x} \in S\} \subset \{\mathbf{x} \in \mathbb{R}^{n_x} : \mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}_0(\bar{\mathbf{x}}) = \emptyset\}.$$

The foregoing geometric characterization of local solution points applies equally well to either interior points $\operatorname{int}(S) := \{\mathbf{x} \in \mathbb{R}^{n_x} : g_i(\mathbf{x}) < 0, \forall i = 1, \ldots, n_g\}$, or boundary points being at the boundary of the feasible domain. At an interior point, in particular, any direction is feasible, and the necessary condition $\mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}_0(\bar{\mathbf{x}}) = \emptyset$ reduces to $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, which gives the same condition as in unconstrained optimization (see Theorem 1.22).

Note also that there are several cases where the condition $\mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}_0(\bar{\mathbf{x}}) = \emptyset$ is satisfied by non-optimal points. In other words, this condition is necessary but *not sufficient* for a point $\bar{\mathbf{x}}$ to be a local minimum of f on S. For instance, any point $\bar{\mathbf{x}}$ with $\nabla g_i(\bar{\mathbf{x}}) = 0$ for some $i \in \mathcal{A}(\bar{\mathbf{x}})$ trivially satisfies the condition $\mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}_0(\bar{\mathbf{x}}) = \emptyset$. Another example is given below.

Example 1.35. Consider the problem

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^2} \ & f(\mathbf{x}) := x_1^2 + x_2^2 \\ \text{s.t.} \ & g_1(\mathbf{x}) := x_1 \leq 0 \\ & g_2(\mathbf{x}) := -x_1 \leq 0. \end{aligned} \tag{1.6}$$

Clearly, this problem is convex and $\mathbf{x}^\star = (0,0)^\mathsf{T}$ is the unique global minimum.

Now, let $\bar{\mathbf{x}}$ be any point on the line $\mathcal{C} := \{\mathbf{x} : x_1 = 0\}$. Both constraints g_1 and g_2 are active at $\bar{\mathbf{x}}$, and we have $\nabla g_1(\bar{\mathbf{x}}) = -\nabla g_2(\bar{\mathbf{x}}) = (1,0)^\mathsf{T}$. Therefore, no direction $\mathbf{d} \neq \mathbf{0}$ can be found such that $\nabla g_1(\bar{\mathbf{x}})^\mathsf{T}\mathbf{d} < 0$ and $\nabla g_2(\bar{\mathbf{x}})^\mathsf{T}\mathbf{d} < 0$ simultaneously, i.e., $\mathcal{D}_0(\bar{\mathbf{x}}) = \emptyset$. In turn, this implies that $\mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}_0(\bar{\mathbf{x}}) = \emptyset$ is trivially satisfied for any point on $\mathcal{C}$.

On the other hand, observe that the condition $\mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}(\bar{\mathbf{x}}) = \emptyset$ in Theorem 1.31 excludes all the points on $\mathcal{C}$, but the origin, since a feasible direction at $\bar{\mathbf{x}}$ is given, e.g., by $\mathbf{d} = (0,1)^\mathsf{T}$.

Next, we reduce the geometric necessary optimality condition $\mathcal{F}_0(\bar{\mathbf{x}}) \cap \mathcal{D}_0(\bar{\mathbf{x}}) = \emptyset$ to a statement in terms of the gradients of the objective function and of the active constraints. The resulting first-order optimality conditions are known as the *Karush-Kuhn-Tucker (KKT) necessary conditions*. Beforehand, we introduce the important concepts of a *regular point* and of a *KKT point*.

Definition 1.36 (Regular Point (for a Set of Inequality Constraints)). *Let $g_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_g$, be differentiable functions on $\mathbb{R}^{n_x}$ and consider the set $S := \{\mathbf{x} \in \mathbb{R}^{n_x} : g_i(\mathbf{x}) \leq 0, i = 1, \ldots, n_g\}$. A point $\bar{\mathbf{x}} \in S$ is said to be a* regular point *if the gradient vectors $\nabla g_i(\bar{\mathbf{x}})$, $i \in \mathcal{A}(\bar{\mathbf{x}})$, are linearly independent,*

$$\operatorname{rank}\left(\nabla g_i(\bar{\mathbf{x}}), i \in \mathcal{A}(\bar{\mathbf{x}})\right) = |\mathcal{A}(\bar{\mathbf{x}})|.$$

Definition 1.37 (KKT Point). *Let $f : \mathbb{R}^{n_x} \to \mathbb{R}$ and $g_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_g$ be differentiable functions. Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$. If a point $(\bar{\mathbf{x}}, \bar{\boldsymbol{\nu}}) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_g}$ satisfies the conditions:*

$$\nabla f(\bar{\mathbf{x}}) + \bar{\boldsymbol{\nu}}^\mathsf{T} \nabla \mathbf{g}(\bar{\mathbf{x}}) = \mathbf{0} \tag{1.7}$$

$$\bar{\boldsymbol{\nu}} \geq \mathbf{0} \tag{1.8}$$

$$\mathbf{g}(\bar{\mathbf{x}}) \leq \mathbf{0} \tag{1.9}$$

$$\bar{\boldsymbol{\nu}}^\mathsf{T} \mathbf{g}(\bar{\mathbf{x}}) = 0, \tag{1.10}$$

then $(\bar{\mathbf{x}}, \bar{\boldsymbol{\nu}})$ *is said to be a* KKT point.

Remark 1.38. The scalars ν_i, $i = 1, \ldots, n_g$, are called the *Lagrange multipliers*. The condition (1.7), i.e., the requirement that $\bar{\mathbf{x}}$ be feasible, is called the *primal feasibility* (PF) condition; the conditions (1.8) and (1.9) are referred to as the *dual feasibility* (DF) conditions; finally, the condition (1.10) is called the *complementarity slackness*[4] (CS) condition.

Theorem 1.39 (KKT Necessary Conditions). *Let* $f : \mathbb{R}^{n_x} \to \mathbb{R}$ *and* $g_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_g$ *be differentiable functions. Consider the problem to minimize* $f(\mathbf{x})$ *subject to* $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$. *If* $\mathbf{x}^\star$ *is a local minimum and a regular point of the constraints, then there exists a* unique *vector* $\boldsymbol{\nu}^\star$ *such that* $(\mathbf{x}^\star, \boldsymbol{\nu}^\star)$ *is a KKT point.*

Proof. Since $\mathbf{x}^\star$ solves the problem, then there exists no direction $\mathbf{d} \in \mathbb{R}^{n_x}$ such that $\nabla f(\bar{\mathbf{x}})^{\mathsf{T}}\mathbf{d} < 0$ and $\nabla g_i(\bar{\mathbf{x}})^{\mathsf{T}}\mathbf{d} < 0$, $\forall i \in \mathcal{A}(\mathbf{x}^\star)$ simultaneously (see Remark 1.34). Let $\mathsf{A} \in \mathbb{R}^{(|\mathcal{A}(\mathbf{x}^\star)|+1) \times n_x}$ be the matrix whose rows are $\nabla f(\bar{\mathbf{x}})^{\mathsf{T}}$ and $\nabla g_i(\bar{\mathbf{x}})^{\mathsf{T}}$, $i \in \mathcal{A}(\mathbf{x}^\star)$. Clearly, the statement $\{\exists \mathbf{d} \in \mathbb{R}^{n_x} : \mathsf{A}\mathbf{d} < \mathbf{0}\}$ is false, and by Gordan's Theorem 1.A.78, there exists a nonzero vector $\mathbf{p} \geq \mathbf{0}$ in $\mathbb{R}^{|\mathcal{A}(\mathbf{x}^\star)|+1}$ such that $\mathsf{A}^{\mathsf{T}}\mathbf{p} = \mathbf{0}$. Denoting the components of $\mathbf{p}$ by u_0 and u_i for $i \in \mathcal{A}(\mathbf{x}^\star)$, we get:

$$u_0 \nabla f(\mathbf{x}^\star) + \sum_{i \in \mathcal{A}(\mathbf{x}^\star)} u_i \nabla g_i(\mathbf{x}^\star) = \mathbf{0}$$

where $u_0 \geq 0$ and $u_i \geq 0$ for $i \in \mathcal{A}(\mathbf{x}^\star)$, and $(u_0, \mathbf{u}_{\mathcal{A}(\mathbf{x}^\star)}) \neq (0, \mathbf{0})$ (here $\mathbf{u}_{\mathcal{A}(\mathbf{x}^\star)}$ is the vector whose components are the u_i's for $i \in \mathcal{A}(\mathbf{x}^\star)$). Letting $u_i = 0$ for $i \notin \mathcal{A}(\mathbf{x}^\star)$, we then get the conditions:

$$\begin{aligned} u_0 \nabla f(\mathbf{x}^\star) + \mathbf{u}^{\mathsf{T}} \nabla \mathbf{g}(\mathbf{x}^\star) &= \mathbf{0} \\ \mathbf{u}^{\mathsf{T}} \mathbf{g}(\mathbf{x}^\star) &= \mathbf{0} \\ u_0, \mathbf{u} &\geq 0 \\ (u_0, \mathbf{u}) &\neq (0, \mathbf{0}), \end{aligned}$$

where $\mathbf{u}$ is the vector whose components are u_i for $i = 1, \ldots, n_g$. Note that $u_0 \neq 0$, for otherwise the assumption of linear independence of the active constraints at $\mathbf{x}^\star$ would be violated. Then, letting $\boldsymbol{\nu}^\star = \frac{1}{u_0}\mathbf{u}$, we obtain that $(\mathbf{x}^\star, \boldsymbol{\nu}^\star)$ is a KKT point. □

Remark 1.40. One of the major difficulties in applying the foregoing result is that we do not know *a priori* which constraints are active and which constraints are inactive, i.e., the active set is *unknown*. Therefore, it is necessary to investigate *all* possible active sets for finding candidate points satisfying the KKT conditions. This is illustrated in Example 1.41 below.

Example 1.41 (Regular Case). Consider the problem

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^3} \; & f(\mathbf{x}) := \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{s.t.} \; & g_1(\mathbf{x}) := x_1 + x_2 + x_3 + 3 \leq 0 \\ & g_2(\mathbf{x}) := x_1 \leq 0. \end{aligned} \tag{1.11}$$

[4] Often, the condition (1.10) is replaced by the equivalent conditions:

$$\bar{\nu}_i g_i(\bar{\mathbf{x}}) = 0 \quad \text{for } i = 1, \ldots, n_g.$$

Note that every feasible point is regular (the point (0,0,0) being infeasible), so $\mathbf{x}^\star$ must satisfy the dual feasibility conditions:

$$\begin{aligned} x_1^\star + \nu_1^\star + \nu_2^\star &= 0 \\ x_2^\star + \nu_1^\star &= 0 \\ x_3^\star + \nu_1^\star &= 0. \end{aligned}$$

Four cases can be distinguished:

(i) The constraints g_1 and g_2 are both *inactive*, i.e., $x_1^\star + x_2^\star + x_3^\star < -3$, $x_1^\star < 0$, and $\nu_1^\star = \nu_2^\star = 0$. From the latter together with the dual feasibility conditions, we get $x_1^\star = x_2^\star = x_3^\star = 0$, hence contradicting the former.

(ii) The constraint g_1 is *inactive*, while g_2 is *active*, i.e., $x_1^\star + x_2^\star + x_3^\star < -3$, $x_1^\star = 0$, $\nu_2^\star \geq 0$, and $\nu_1^\star = 0$. From the latter together with the dual feasibility conditions, we get $x_2^\star = x_3^\star = 0$, hence contradicting the former once again.

(iii) The constraint g_1 is *active*, while g_2 is *inactive*, i.e., $x_1^\star + x_2^\star + x_3^\star = -3$, $x_1^\star < 0$, and $\nu_1^\star \geq 0$, and $\nu_2^\star = 0$. Then, the point $(\mathbf{x}^\star, \boldsymbol{\nu}^\star)$ such that $x_1{}^\star = x_2{}^\star = x_3{}^\star = -1$, $\nu_1^\star = 1$ and $\nu_2^\star = 0$ is a KKT point.

(iv) The constraints g_1 and g_2 are both *active*, i.e., $x_1^\star + x_2^\star + x_3^\star = -3$, $x_1^\star = 0$, and $\nu_1^\star, \nu_2^\star > 0$. Then, we obtain $x_2{}^\star = x_3{}^\star = -\frac{3}{2}$, $\nu_1^\star = \frac{3}{2}$, and $\nu_2^\star = -\frac{3}{2}$, hence contradicting the dual feasibility condition $\nu_2^\star \geq 0$.

Overall, there is a *unique* candidate for a local minimum. Yet, it cannot be concluded as to whether this point is actually a global minimum, or even a local minimum, of (1.11). This question will be addressed later on in Example 1.45.

Remark 1.42 (Constraint Qualification). It is *very* important to note that for a local minimum $\mathbf{x}^\star$ to be a KKT point, an additional condition must be placed on the behavior of the constraint, i.e., **not every local minimum is a KKT point**; such a condition is known as a *constraint qualification*. In Theorem 1.39, it is shown that one possible constraint qualification is that $\mathbf{x}^\star$ be a regular point, which is the well known *linear independence constraint qualification* (LICQ). A weaker constraint qualification (i.e., implied by LICQ) known as the *Mangasarian-Fromovitz constraint qualification* (MFCQ) requires that there exits (at least) one direction $\mathbf{d} \in \mathcal{D}_0(\mathbf{x}^\star)$, i.e., such that $\nabla g_i(\mathbf{x}^\star)^\mathsf{T}\mathbf{d} < 0$, for each $i \in \mathcal{A}(\mathbf{x}^\star)$. Note, however, that the Lagrange multipliers are guaranteed to be unique if LICQ holds (as stated in Theorem 1.39), while this uniqueness property may be lost under MFCQ.

The following example illustrates the necessity of having a constraint qualification for a KKT point to be a local minimum point of an NLP.

Example 1.43 (Non-Regular Case). Consider the problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^2} \ & f(\mathbf{x}) := -x_1 \\ \text{s.t.} \ & g_1(\mathbf{x}) := x_2 - (1 - x_1)^3 \leq 0 \\ & g_2(\mathbf{x}) := -x_2 \leq 0. \end{aligned} \tag{1.12}$$

The feasible region is shown in Fig. 1.10. below. Note that a minimum point of (1.12) is $\mathbf{x}^\star = (1, 0)^\mathsf{T}$. The dual feasibility condition relative to variable x_1 reads:

$$-1 + 3\nu_1(1 - x_1)^2 \;=\; 0.$$

It is readily seen that this condition cannot be met at any point on the straight line $\mathcal{C} := \{\mathbf{x} : x_1 = 1\}$, including the minimum point $\mathbf{x}^\star$. In other words, the KKT conditions are not necessary in this example. This is because no constraint qualification can hold at $\mathbf{x}^\star$. In particular, $\mathbf{x}^\star$ not being a regular point, LICQ does not hold; moreover, the set $\mathcal{D}_0(\mathbf{x}^\star)$ being empty (the direction $\mathbf{d} = (-1, 0)^\mathsf{T}$ gives $\nabla g_1(\mathbf{x}^\star)^\mathsf{T}\mathbf{d} = \nabla g_2(\mathbf{x}^\star)^\mathsf{T}\mathbf{d} = 0$, while any other direction induces a violation of either one of the constraints), MFCQ does not hold at $\mathbf{x}^\star$ either.

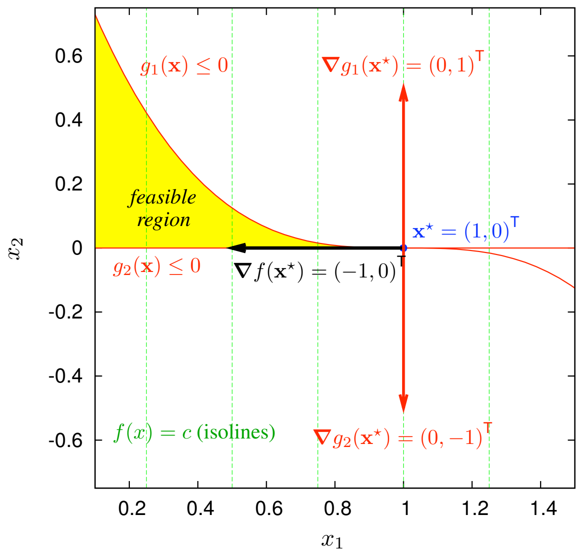

Figure 1.10. Solution of Example 1.43.

The following theorem provides a sufficient condition under which any KKT point of an inequality constrained NLP problem is guaranteed to be a global minimum of that problem.

Theorem 1.44 (KKT sufficient Conditions). *Let $f : \mathbb{R}^{n_x} \to \mathbb{R}$ and $g_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_g$, be convex and differentiable functions. Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$. If $(\mathbf{x}^\star, \boldsymbol{\nu}^\star)$ is a KKT point, then $\mathbf{x}^\star$ is a global minimum of that problem.*

Proof. Consider the function $\mathcal{L}(\mathbf{x}) := f(\mathbf{x}) + \sum_{i=1}^{n_g} \nu_i^\star g_i(\mathbf{x})$. Since f and g_i, $i = 1, \ldots, n_g$, are convex functions, and $\nu_i \geq 0$, $\mathcal{L}$ is also convex. Moreover, the dual feasibility conditions impose that we have $\nabla\mathcal{L}(\mathbf{x}^\star) = \mathbf{0}$. Hence, by Theorem 1.27, $\mathbf{x}^\star$ is a global minimizer for $\mathcal{L}$ on $\mathbb{R}^{n_x}$, i.e.,

$$\mathcal{L}(\mathbf{x}) \;\geq\; \mathcal{L}(\mathbf{x}^\star) \quad \forall \mathbf{x} \in \mathbb{R}^{n_x}.$$

In particular, for each $\mathbf{x}$ such that $g_i(\mathbf{x}) \leq g_i(\mathbf{x}^\star) = 0$, $i \in \mathcal{A}(\mathbf{x}^\star)$, we have

$$f(\mathbf{x}) - f(\mathbf{x}^\star) \ \geq \ - \sum_{i \in \mathcal{A}(\mathbf{x}^\star)} \mu_i^\star \left[g_i(\mathbf{x}) - g_i(\mathbf{x}^\star) \right] \ \geq \ 0.$$

Noting that $\{\mathbf{x} \in \mathbb{R}^{n_x} : g_i(\mathbf{x}) \leq 0, i \in \mathcal{A}(\mathbf{x}^\star)\}$ contains the feasible domain $\{\mathbf{x} \in \mathbb{R}^{n_x} : g_i(\mathbf{x}) \leq 0, i = 1, \ldots, n_g\}$, we therefore showed that $\mathbf{x}^\star$ is a global minimizer for the problem. □

Example 1.45. Consider the same Problem (1.11) as in Example 1.41 above. The point $(\mathbf{x}^\star, \boldsymbol{\nu}^\star)$ with $x_1{}^\star = x_2{}^\star = x_3{}^\star = -1$, $\nu_1^\star = 1$ and $\nu_2^\star = 0$, being a KKT point, and both the objective function and the feasible set being convex, by Theorem 1.44, $\mathbf{x}^\star$ is a global minimum for the Problem (1.11).

Both second-order necessary and sufficient conditions for inequality constrained NLP problems will be presented later on in §1.7.

1.6 PROBLEMS WITH EQUALITY CONSTRAINTS

In this section, we shall consider nonlinear programming problems with equality constraints of the form:

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) \ = \ 0 \quad i = 1, \ldots, n_h. \end{aligned}$$

Based on the material presented in §1.5, it is tempting to convert this problem into an inequality constrained problem, by replacing each equality constraints $h_i(\mathbf{x}) = 0$ by two inequality constraints $h_i^+(\mathbf{x}) = h_i(\mathbf{x}) \leq 0$ and $h_i^-(\mathbf{x}) = -h(\mathbf{x}) \leq 0$. Given a feasible point $\bar{\mathbf{x}} \in \mathbb{R}^{n_x}$, we would have $h_i^+(\bar{\mathbf{x}}) = h_i^-(\bar{\mathbf{x}}) = 0$ and $\boldsymbol{\nabla} h_i^+(\bar{\mathbf{x}}) = -\boldsymbol{\nabla} h_i^-(\bar{\mathbf{x}})$. Therefore, there could exist no vector $\mathbf{d}$ such that $\boldsymbol{\nabla} h_i^+(\bar{\mathbf{x}}) < 0$ and $\boldsymbol{\nabla} h_i^-(\bar{\mathbf{x}}) < 0$ simultaneously, i.e., $\mathcal{D}_0(\bar{\mathbf{x}}) = \emptyset$. In other words, the geometric conditions developed in the previous section for inequality constrained problems are satisfied by all feasible solutions and, hence, are not informative (see Example 1.35 for an illustration). A different approach must therefore be used to deal with equality constrained problems. After a number of preliminary results in §1.6.1, we shall describe the method of Lagrange multipliers for equality constrained problems in §1.6.2.

1.6.1 Preliminaries

An equality constraint $h(\mathbf{x}) = 0$ defines a set on $\mathbb{R}^{n_x}$, which is best view as a hypersurface. When considering $n_h \geq 1$ equality constraints $h_1(\mathbf{x}), \ldots, h_{n_h}(\mathbf{x})$, their intersection forms a (possibly empty) set $S := \{\mathbf{x} \in \mathbb{R}^{n_x} : h_i(\mathbf{x}) = 0, i = 1, \ldots, n_h\}$.

Throughout this section, we shall assume that the equality constraints are differentiable; that is, the set $S := \{\mathbf{x} \in \mathbb{R}^{n_x} : h_i(\mathbf{x}) = 0, i = 1, \ldots, n_h\}$ is said to be *differentiable manifold* (or *smooth manifold*). Associated with a point on a differentiable manifold is the *tangent set* at that point. To formalize this notion, we start by defining *curves* on a manifold. A curve $\boldsymbol{\xi}$ on a manifold S is a continuous application $\boldsymbol{\xi} : \mathcal{I} \subset \mathbb{R} \to S$, i.e., a family of points $\boldsymbol{\xi}(t) \in S$ continuously parameterized by t in an interval $\mathcal{I}$ of $\mathbb{R}$. A curve is said to pass through the point $\bar{\mathbf{x}}$ if $\bar{\mathbf{x}} = \boldsymbol{\xi}(\bar{t})$ for some $\bar{t} \in \mathcal{I}$; the *derivative* of a curve at $\bar{t}$, provided it exists, is defined as $\dot{\boldsymbol{\xi}}(\bar{t}) := \lim_{h \to 0} \frac{\boldsymbol{\xi}(\bar{t}+h) - \boldsymbol{\xi}(\bar{t})}{h}$. A curve is said to be *differentiable* (or *smooth*) if a derivative exists for each $t \in \mathcal{I}$.

Definition 1.46 (Tangent Set). *Let S be a (differentiable) manifold in $\mathbb{R}^{n_x}$, and let $\bar{\mathbf{x}} \in S$. Consider the collection of* all *the continuously differentiable curves on S passing through $\bar{\mathbf{x}}$. Then, the collection of* all *the vectors tangent to these curves at $\bar{\mathbf{x}}$ is said to be the* tangent set *to S at $\bar{\mathbf{x}}$, denoted by $\mathscr{T}(\bar{\mathbf{x}})$.*

If the constraints are *regular*, in the sense of Definition 1.47 below, then S is (locally) of dimension $n_x - n_h$, and $\mathscr{T}(\bar{\mathbf{x}})$ constitutes a subspace of dimension $n_x - n_h$, called the *tangent space*.

Definition 1.47 (Regular Point (for a Set of Equality Constraints)). *Let $h_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_h$, be differentiable functions on $\mathbb{R}^{n_x}$ and consider the set $S := \{\mathbf{x} \in \mathbb{R}^{n_x} : h_i(\mathbf{x}) = 0, i = 1, \ldots, n_h\}$. A point $\bar{\mathbf{x}} \in S$ is said to be a* regular point *if the gradient vectors $\nabla h_i(\bar{\mathbf{x}})$, $i = 1, \ldots, n_h$, are linearly independent, i.e.,*

$$\operatorname{rank}\begin{pmatrix} \nabla h_1(\bar{\mathbf{x}}) & \nabla h_2(\bar{\mathbf{x}}) & \cdots & \nabla h_{n_h}(\bar{\mathbf{x}}) \end{pmatrix} = n_h.$$

Lemma 1.48 (Algebraic Characterization of a Tangent Space). *Let $h_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_h$, be differentiable functions on $\mathbb{R}^{n_x}$ and consider the set $S := \{\mathbf{x} \in \mathbb{R}^{n_x} : h_i(\mathbf{x}) = 0, i = 1, \ldots, n_h\}$. At a regular point $\bar{\mathbf{x}} \in S$, the tangent space is such that*

$$\mathscr{T}(\bar{\mathbf{x}}) = \{\mathbf{d} : \nabla \mathbf{h}(\bar{\mathbf{x}})^{\mathsf{T}} \mathbf{d} = \mathbf{0}\}.$$

Proof. The proof is technical and is omitted here (see, e.g., [36, §10.2]). □

1.6.2 The Method of Lagrange Multipliers

The idea behind the method of Lagrange multipliers for solving equality constrained NLP problems of the form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) \ = \ 0 \quad i = 1, \ldots, n_h. \end{aligned}$$

is to restrict the search of a minimum on the manifold $S := \{\mathbf{x} \in \mathbb{R}^{n_x} : h_i(\mathbf{x}) = 0, \forall i = 1, \ldots, n_h\}$. In other words, we derive optimality conditions by considering the value of the objective function along curves on the manifold S passing through the optimal point.

The following Theorem shows that the tangent space $\mathscr{T}(\bar{\mathbf{x}})$ at a regular (local) minimum point $\bar{\mathbf{x}}$ is orthogonal to the gradient of the objective function at $\bar{\mathbf{x}}$. This important fact is illustrated in Fig. 1.11. in the case of a single equality constraint.

Theorem 1.49 (Geometric Necessary Condition for a Local Minimum). *Let $f : \mathbb{R}^{n_x} \to \mathbb{R}$ and $h_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_h$, be continuously differentiable functions on $\mathbb{R}^{n_x}$. Suppose that $\mathbf{x}^\star$ is a local minimum point of the problem to minimize $f(\mathbf{x})$ subject to the constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. Then, $\nabla f(\mathbf{x}^\star)$ is orthogonal to the tangent space $\mathscr{T}(\mathbf{x}^\star)$,*

$$\mathcal{F}_0(\mathbf{x}^\star) \cap \mathscr{T}(\mathbf{x}^\star) \ = \ \emptyset.$$

Proof. By contradiction, assume that there exists a $\mathbf{d} \in \mathscr{T}(\mathbf{x}^\star)$ such that $\nabla f(\mathbf{x}^\star)^{\mathsf{T}} \mathbf{d} \neq 0$. Let $\boldsymbol{\xi} : \mathcal{I} = [-a, a] \to S$, $a > 0$, be any smooth curve passing through $\mathbf{x}^\star$ with $\boldsymbol{\xi}(0) = \mathbf{x}^\star$

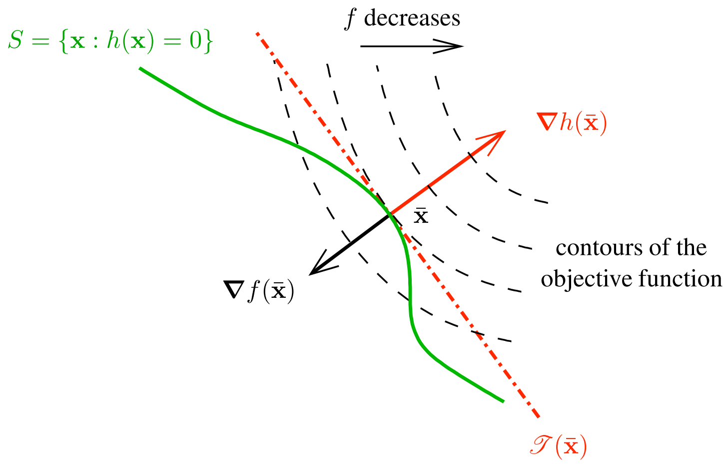

Figure 1.11. Illustration of the necessary conditions of optimality with equality constraints.

and $\dot{\boldsymbol{\xi}}(0) = \mathbf{d}$. Let also φ be the function defined as $\varphi(t) := f(\boldsymbol{\xi}(t))$, $\forall t \in \mathcal{I}$. Since $\mathbf{x}^\star$ is a local minimum of f on $S := \{\mathbf{x} \in \mathbb{R}^{n_x} : \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, by Definition 1.9, we have

$$\exists \eta > 0 \text{ such that } \varphi(t) = f(\boldsymbol{\xi}(t)) \geq f(\mathbf{x}^\star) = \varphi(0) \quad \forall t \in \mathcal{B}_\eta(0) \cap \mathcal{I}.$$

It follows that $t^\star = 0$ is an unconstrained (local) minimum point for φ, and

$$0 \;=\; \nabla\varphi(0) \;=\; \boldsymbol{\nabla} f(\mathbf{x}^\star)^{\mathsf{T}} \dot{\boldsymbol{\xi}}(0) \;=\; \boldsymbol{\nabla} f(\mathbf{x}^\star)^{\mathsf{T}} \mathbf{d}.$$

We thus get a contradiction with the fact that $\boldsymbol{\nabla} f(\mathbf{x}^\star)^{\mathsf{T}} \mathbf{d} \neq 0$. □

Next, we take advantage of the forgoing geometric characterization, and derive first-order necessary conditions for equality constrained NLP problems.

Theorem 1.50 (First-Order Necessary Conditions). *Let $f : \mathbb{R}^{n_x} \to \mathbb{R}$ and $h_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_h$, be continuously differentiable functions on $\mathbb{R}^{n_x}$. Consider the problem to minimize $f(\mathbf{x})$ subject to the constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. If $\mathbf{x}^\star$ is a local minimum and is a regular point of the constraints, then there exists a unique vector $\boldsymbol{\lambda}^\star \in \mathbb{R}^{n_h}$ such that*

$$\boldsymbol{\nabla} f(\mathbf{x}^\star) + \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^\star)^{\mathsf{T}} \boldsymbol{\lambda}^\star = \mathbf{0}.$$

Proof. [5] Since $\mathbf{x}^\star$ is a local minimum of f on $S := \{\mathbf{x} \in \mathbb{R}^{n_x} : \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, by Theorem 1.49, we have $\mathcal{F}_0(\mathbf{x}^\star) \cap \mathscr{T}(\mathbf{x}^\star) = \emptyset$, i.e., the system

$$\boldsymbol{\nabla} f(\mathbf{x}^\star)^{\mathsf{T}} \mathbf{d} < 0 \qquad \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^\star)^{\mathsf{T}} \mathbf{d} = 0,$$

is inconsistent. Consider the following two sets:

$$\begin{aligned} C_1 &:= \{(z_1, \mathbf{z}_2) \in \mathbb{R}^{n_h+1} : z_1 = \boldsymbol{\nabla} f(\mathbf{x}^\star)^{\mathsf{T}} \mathbf{d}, \quad \mathbf{z}_2 = \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^\star)^{\mathsf{T}} \mathbf{d}\} \\ C_2 &:= \{(z_1, \mathbf{z}_2) \in \mathbb{R}^{n_h+1} : z_1 < 0, \quad \mathbf{z}_2 = \mathbf{0}.\} \end{aligned}$$

[5] See also in Appendix of §1 for an alternative proof of Theorem 1.50 that does not use the concept of tangent sets.

Clearly, C_1 and C_2 are convex, and $C_1 \cap C_2 = \emptyset$. Then, by the separation Theorem A.9, there exists a nonzero vector $(\mu, \boldsymbol{\lambda}) \in \mathbb{R}^{n_h+1}$ such that

$$\mu \nabla f(\mathbf{x}^\star)^\mathsf{T} \mathbf{d} + \boldsymbol{\lambda}^\mathsf{T} [\nabla \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \mathbf{d}] \ \geq \ \mu z_1 + \boldsymbol{\lambda}^\mathsf{T} \mathbf{z}_2 \qquad \forall \mathbf{d} \in \mathbb{R}^{n_x},\ \forall (z_1, \mathbf{z}_2) \in C_2.$$

Letting $\mathbf{z}_2 = \mathbf{0}$ and since z_1 can be made an arbitrarily large negative number, it follows that $\mu \geq 0$. Also, letting $(z_1, \mathbf{z}_2) = (0, \mathbf{0})$, we must have $[\mu \nabla f(\mathbf{x}^\star) + \boldsymbol{\lambda}^\mathsf{T} \nabla \mathbf{h}(\mathbf{x}^\star)]^\mathsf{T} \mathbf{d} \geq \mathbf{0}$, for each $\mathbf{d} \in \mathbb{R}^{n_x}$. In particular, letting $\mathbf{d} = -[\mu \nabla f(\mathbf{x}^\star) + \boldsymbol{\lambda}^\mathsf{T} \nabla \mathbf{h}(\mathbf{x}^\star)]$, it follows that $-\|\mu \nabla f(\mathbf{x}^\star) + \boldsymbol{\lambda}^\mathsf{T} \nabla \mathbf{h}(\mathbf{x}^\star)\|^2 \geq 0$, and thus,

$$\mu \nabla f(\mathbf{x}^\star) + \boldsymbol{\lambda}^\mathsf{T} \nabla \mathbf{h}(\mathbf{x}^\star) = \mathbf{0} \qquad \text{with } (\mu, \boldsymbol{\lambda}) \neq (0, \mathbf{0}). \tag{1.13}$$

Finally, note that $\mu > 0$, for otherwise (1.13) would contradict the assumption of linear independence of $\nabla h_i(\mathbf{x}^\star)$, $i = 1, \ldots, n_h$. The result follows by letting $\boldsymbol{\lambda}^\star := \frac{1}{\mu} \boldsymbol{\lambda}$, and noting that the linear independence assumption implies the uniqueness of these Lagrangian multipliers. □

Remark 1.51 (Obtaining Candidate Solution Points). The first-order necessary conditions

$$\nabla f(\mathbf{x}^\star) + \nabla \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \boldsymbol{\lambda}^\star = \mathbf{0},$$

together with the constraints

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

give a total of $n_x + n_h$ (typically nonlinear) equations in the variables $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$. Hence, these conditions are complete in the sense that they determine, at least locally, a unique solution. However, as in the unconstrained case, a solution to the first-order necessary conditions need not be a (local) minimum of the original problem; it could very well correspond to a (local) maximum point, or some kind of saddle point. These consideration are illustrated in Example 1.54 below.

Remark 1.52 (Regularity-Type Assumption). It is important to note that for a local minimum to satisfy the foregoing first-order conditions and, in particular, for a unique Lagrange multiplier vector to exist, it is necessary that the equality constraint satisfy a regularity condition. In other word, the first-order conditions may not hold at a local minimum point that is non-regular. An illustration of these considerations is provided in Example 1.55.

There exists a number of similarities with the constraint qualification needed for a local minimizer of an inequality constrained NLP problem to be KKT point; in particular, the condition that the minimum point be a regular point for the constraints corresponds to LICQ (see Remark 1.42).

Remark 1.53 (Lagrangian). It is convenient to introduce the *Lagrangian* $\mathcal{L} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_h} \to \mathbb{R}$ associated with the constrained problem, by adjoining the cost and constraint functions as:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \ := \ f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T} \mathbf{h}(\mathbf{x}).$$

Thus, if $\mathbf{x}^\star$ is a local minimum which is regular, the first-order necessary conditions are written as

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) = \mathbf{0} \tag{1.14}$$

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) = \mathbf{0}, \tag{1.15}$$

the latter equations being simply a restatement of the constraints. Note that the solution of the original problem typically corresponds to a saddle point of the Lagrangian function.

Example 1.54 (Regular Case). Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) := x_1 + x_2 \tag{1.16}$$
$$\text{s.t.} \quad h(\mathbf{x}) := x_1^2 + x_2^2 - 2 = 0.$$

Observe first that every feasible point is a regular point for the equality constraint (the point (0,0) being infeasible). Therefore, every local minimum is a stationary point of the Lagrangian function by Theorem 1.50.

The gradient vectors $\nabla f(\mathbf{x})$ and $\nabla h(\mathbf{x})$ are given by

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathsf{T}} \quad \text{and} \quad \nabla h(\mathbf{x}) = \begin{pmatrix} 2x_1 & 2x_2 \end{pmatrix}^{\mathsf{T}},$$

so that the first-order necessary conditions read

$$\begin{aligned} 2\lambda x_1 &= -1 \\ 2\lambda x_2 &= -1 \\ x_1^2 + x_2^2 &= 2. \end{aligned}$$

These three equations can be solved for the three unknowns x_1, x_2 and λ. Two candidate local minimum points are obtained: (i) ${x_1}^\star = {x_2}^\star = -1$, $\lambda^\star = \frac{1}{2}$, and (ii) ${x_1}^\star = {x_2}^\star = 1$, $\lambda^\star = -\frac{1}{2}$. These results are illustrated on Fig. 1.12.. It can be seen that only the former actually corresponds to a local minimum point, while the latter gives a local maximum point.

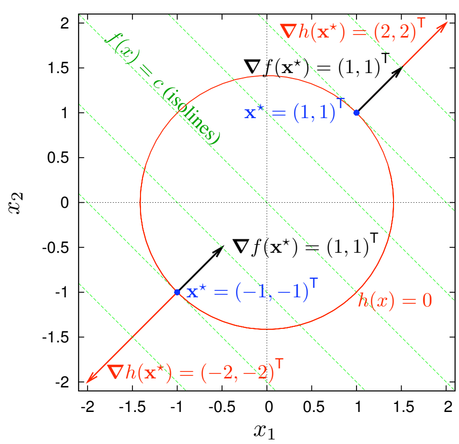

Figure 1.12. Solution of Example 1.54.

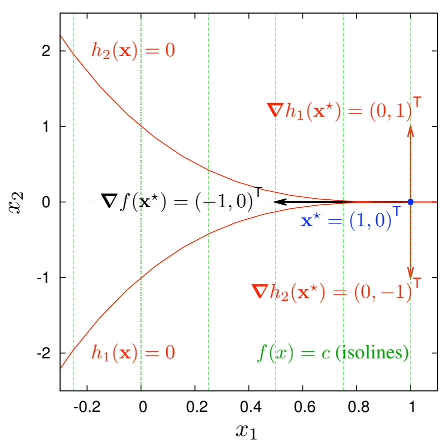

Figure 1.13. Solution of Example 1.55.

Example 1.55 (Non-Regular Case). Consider the problem

$$\begin{array}{rl} \min\limits_{\mathbf{x}\in\mathrm{IR}^2} & f(\mathbf{x}) := -x_1 \\ \text{s.t.} & h_1(\mathbf{x}) := (1-x_1)^3 + x_2 = 0 \\ & h_2(\mathbf{x}) := (1-x_1)^3 - x_2 = 0. \end{array} \tag{1.17}$$

As shown by Fig. 1.13., this problem has only one feasible point, namely, $\mathbf{x}^\star = (1,0)^\mathsf{T}$; that is, $\mathbf{x}^\star$ is also the unique global minimum of (1.17). However, at this point, we have

$$\nabla f(\mathbf{x}^\star) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \nabla h_1(\mathbf{x}^\star) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \nabla h_2(\mathbf{x}^\star) = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

hence the first-order conditions

$$\lambda_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

cannot be satisfied. This illustrates the fact that a minimum point may not be a stationary point for the Lagrangian if that point is non-regular.

The following theorem provides second-order necessary conditions for a point to be a local minimum of a NLP problem with equality constraints.

Theorem 1.56 (Second-Order Necessary Conditions). *Let $f : \mathrm{IR}^{n_x} \to \mathrm{IR}$ and $h_i : \mathrm{IR}^{n_x} \to \mathrm{IR}$, $i = 1, \ldots, n_h$, be twice continuously differentiable functions on IR^{n_x}. Consider the problem to minimize $f(\mathbf{x})$ subject to the constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. If $\mathbf{x}^\star$ is a local minimum and is a regular point of the constraints, then there exists a unique vector $\boldsymbol{\lambda}^\star \in \mathrm{IR}^{n_h}$ such that*

$$\nabla f(\mathbf{x}^\star) + \nabla \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \boldsymbol{\lambda}^\star = \mathbf{0},$$

and

$$\mathbf{d}^\mathsf{T} \left(\nabla^2 f(\mathbf{x}^\star) + \nabla^2 \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \boldsymbol{\lambda}^\star \right) \mathbf{d} \geq 0 \qquad \forall \mathbf{d} \textit{ such that } \nabla \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \mathbf{d} = 0.$$

Proof. Note first that $\nabla f(\mathbf{x}^\star) + \nabla \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \boldsymbol{\lambda}^\star = \mathbf{0}$ directly follows from Theorem 1.50.

Let $\mathbf{d}$ be an arbitrary direction in $\mathscr{T}(\mathbf{x}^\star)$; that is, $\nabla \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \mathbf{d} = 0$ since $\mathbf{x}^\star$ is a regular point (see Lemma 1.48). Consider any twice-differentiable curve $\boldsymbol{\xi} : \mathcal{I} = [-a, a] \to S$, $a > 0$, passing through $\mathbf{x}^\star$ with $\boldsymbol{\xi}(0) = \mathbf{x}^\star$ and $\dot{\boldsymbol{\xi}}(0) = \mathbf{d}$. Let φ be the function defined as $\varphi(t) := f(\boldsymbol{\xi}(t))$, $\forall t \in \mathcal{I}$. Since $\mathbf{x}^\star$ is a local minimum of f on $S := \{\mathbf{x} \in \mathrm{IR}^{n_x} : \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, $t^\star = 0$ is an unconstrained (local) minimum point for φ. By Theorem 1.25, it follows that

$$0 \;\leq\; \nabla^2 \varphi(0) \;=\; \dot{\boldsymbol{\xi}}(0)^\mathsf{T} \nabla^2 f(\mathbf{x}^\star) \dot{\boldsymbol{\xi}}(0) + \nabla f(\mathbf{x}^\star)^\mathsf{T} \ddot{\boldsymbol{\xi}}(0).$$

Furthermore, differentiating the relation $\mathbf{h}(\boldsymbol{\xi}(\mathbf{t}))^\mathsf{T} \boldsymbol{\lambda} = 0$ twice, we obtain

$$\dot{\boldsymbol{\xi}}(0)^\mathsf{T} \left(\nabla^2 \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \boldsymbol{\lambda} \right) \dot{\boldsymbol{\xi}}(0) + \left(\nabla \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \boldsymbol{\lambda} \right)^\mathsf{T} \ddot{\boldsymbol{\xi}}(0) \;=\; 0.$$

Adding the last two equations yields

$$\mathbf{d}^\mathsf{T} \left(\nabla^2 f(\mathbf{x}^\star) + \nabla^2 \mathbf{h}(\mathbf{x}^\star)^\mathsf{T} \boldsymbol{\lambda}^\star \right) \mathbf{d} \geq 0,$$

and this condition must hold for every $\mathbf{d}$ such that $\nabla\mathbf{h}(\mathbf{x}^\star)^{\mathsf{T}}\mathbf{d} = 0$. □

Remark 1.57 (Eigenvalues in Tangent Space). In the foregoing theorem, it is shown that the matrix $\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ restricted to the subspace $\mathscr{T}(\mathbf{x}^\star)$ plays a key role. Geometrically, the restriction of $\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ to $\mathscr{T}(\mathbf{x}^\star)$ corresponds to the projection $\mathcal{P}_{\mathscr{T}(\mathbf{x}^\star)}[\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)]$.

A vector $\mathbf{y} \in \mathscr{T}(\mathbf{x}^\star)$ is said to be an *eigenvector* of $\mathcal{P}_{\mathscr{T}(\mathbf{x}^\star)}[\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)]$ if there is a real number μ such that

$$\mathcal{P}_{\mathscr{T}(\mathbf{x}^\star)}[\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)]\mathbf{y} \;=\; \mu\mathbf{y};$$

the corresponding μ is said to be an *eigenvalue* of $\mathcal{P}_{\mathscr{T}(\mathbf{x}^\star)}[\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)]$. (These definitions coincide with the usual definitions of eigenvector and eigenvalue for real matrices.)

Now, to obtain a matrix representation for $\mathcal{P}_{\mathscr{T}(\mathbf{x}^\star)}[\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)]$, it is necessary to introduce a basis of the subspace $\mathscr{T}(\mathbf{x}^\star)$, say $\mathsf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_{n_x-n_h})$. Then, the eigenvalues of $\mathcal{P}_{\mathscr{T}(\mathbf{x}^\star)}[\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)]$ are the same as those of the $(n_x - n_h) \times (n_x - n_h)$ matrix $\mathsf{E}^{\mathsf{T}}\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\mathsf{E}$; in particular, they are independent of the particular choice of basis E.

Example 1.58 (Regular Case Continued). Consider the problem (1.16) addressed earlier in Example 1.54. Two candidate local minimum points, (i) $x_1{}^\star = x_2{}^\star = -1$, $\lambda^\star = \frac{1}{2}$, and (ii) $x_1{}^\star = x_2{}^\star = 1$, $\lambda^\star = -\frac{1}{2}$, were obtained on application of the first-order necessary conditions. The Hessian matrix of the Lagrangian function is given by

$$\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}, \lambda) \;=\; \nabla^2 f(\mathbf{x}) + \lambda\nabla^2 h(\mathbf{x}) \;=\; \lambda\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

and a basis of the tangent subspace at a point $\mathbf{x} \in \mathscr{T}(\mathbf{x})$, $\mathbf{x} \neq (0, 0)$, is

$$\mathsf{E}(\mathbf{x}) \;:=\; \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}.$$

Therefore,

$$\mathsf{E}^{\mathsf{T}}\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})\mathsf{E} \;=\; 2\lambda(x_1^2 + x_2^2).$$

In particular, for the candidate solution point (i), we have

$$\mathsf{E}^{\mathsf{T}}\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\mathsf{E} \;=\; 2 \;>\; 0,$$

hence satisfying the second-order necessary conditions (in fact, this point also satisfies the second-order sufficient conditions of optimality discussed hereafter). On the other hand, for the candidate solution point (ii), we get

$$\mathsf{E}^{\mathsf{T}}\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\mathsf{E} \;=\; -2 \;<\; 0$$

which does not satisfy the second-order requirement, so this point cannot be a local minimum.

The conditions given in Theorems 1.50 and 1.56 are necessary conditions that must hold at each local minimum point. Yet, a point satisfying these conditions may not be a local minimum. The following theorem provides sufficient conditions for a stationary point of the Lagrangian function to be a (local) minimum, provided that the Hessian matrix of the Lagrangian function is locally convex along directions in the tangent space of the constraints.

Theorem 1.59 (Second-Order Sufficient Conditions). *Let* $f : \mathbb{R}^{n_x} \to \mathbb{R}$ *and* $h_i : \mathbb{R}^{n_x} \to \mathbb{R}$, $i = 1, \ldots, n_h$, *be twice continuously differentiable functions on* $\mathbb{R}^{n_x}$. *Consider the problem to minimize* $f(\mathbf{x})$ *subject to the constraints* $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. *If* $\mathbf{x}^\star$ *and* $\boldsymbol{\lambda}^\star$ *satisfy*

$$\begin{aligned} \boldsymbol{\nabla}_{\mathbf{x}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) = 0 \\ \boldsymbol{\nabla}_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) = 0, \end{aligned}$$

and

$$\mathbf{y}^{\mathsf{T}} \boldsymbol{\nabla}^2_{\mathbf{xx}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) \mathbf{y} > 0 \quad \forall \mathbf{y} \neq \mathbf{0} \text{ such that } \boldsymbol{\nabla} \mathbf{h}(\mathbf{x}^\star)^{\mathsf{T}} \mathbf{y} = 0,$$

where $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{h}(\mathbf{x})$, *then* $\mathbf{x}^\star$ *is a strict local minimum.*

Proof. Consider the augmented Lagrangian function

$$\bar{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{h}(\mathbf{x}) + \frac{c}{2} \|\mathbf{h}(\mathbf{x})\|^2,$$

where c is a scalar. We have

$$\begin{aligned} \boldsymbol{\nabla}_{\mathbf{x}} \bar{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) &= \boldsymbol{\nabla}_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \bar{\boldsymbol{\lambda}}) \\ \boldsymbol{\nabla}^2_{\mathbf{xx}} \bar{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) &= \boldsymbol{\nabla}^2_{\mathbf{xx}} \mathcal{L}(\mathbf{x}, \bar{\boldsymbol{\lambda}}) + c \boldsymbol{\nabla} \mathbf{h}(\mathbf{x})^{\mathsf{T}} \boldsymbol{\nabla} \mathbf{h}(\mathbf{x}), \end{aligned}$$

where $\bar{\boldsymbol{\lambda}} = \boldsymbol{\lambda} + c\mathbf{h}(\mathbf{x})$. Since $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ satisfy the sufficient conditions and by Lemma 1.A.79, we obtain

$$\boldsymbol{\nabla}_{\mathbf{x}} \bar{\mathcal{L}}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) = 0 \quad \text{and} \quad \boldsymbol{\nabla}^2_{\mathbf{xx}} \bar{\mathcal{L}}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) \succ 0,$$

for sufficiently large c. $\bar{\mathcal{L}}$ being definite positive at $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$,

$$\exists \varrho > 0, \delta > 0 \text{ such that } \bar{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}^\star) \geq \bar{\mathcal{L}}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) + \frac{\varrho}{2} \|\mathbf{x} - \mathbf{x}^\star\|^2 \quad \text{for } \|\mathbf{x} - \mathbf{x}^\star\| < \delta.$$

Finally, since $\bar{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}^\star) = f(\mathbf{x})$ when $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, we get

$$f(\mathbf{x}) \geq f(\mathbf{x}^\star) + \frac{\varrho}{2} \|\mathbf{x} - \mathbf{x}^\star\|^2 \quad \text{if } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \|\mathbf{x} - \mathbf{x}^\star\| < \delta,$$

i.e., $\mathbf{x}^\star$ is a strict local minimum. □

Example 1.60. Consider the problem

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^3} \ & f(\mathbf{x}) := -x_1 x_2 - x_1 x_3 - x_2 x_3 \\ \text{s.t.} \ & h(\mathbf{x}) := x_1 + x_2 + x_3 - 3 = 0. \end{aligned} \tag{1.18}$$

The first-order conditions for this problem are

$$\begin{aligned} -(x_2 + x_3) + \lambda &= 0 \\ -(x_1 + x_3) + \lambda &= 0 \\ -(x_1 + x_2) + \lambda &= 0, \end{aligned}$$

together with the equality constraint. It is easily checked that the point $x_1^\star = x_2^\star = x_3^\star = 1$, $\lambda^\star = 2$ satisfies these conditions. Moreover,

$$\boldsymbol{\nabla}^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) = \boldsymbol{\nabla}^2 f(\mathbf{x}^\star) = \begin{pmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix},$$

and a basis of the tangent space to the constraint $h(\mathbf{x}) = 0$ at $\mathbf{x}^\star$ is

$$\mathsf{E} := \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ -1 & -1 \end{pmatrix}.$$

We thus obtain

$$\mathsf{E}^{\mathsf{T}}\boldsymbol{\nabla}^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\mathsf{E} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

which is clearly a definite positive matrix. Hence, $\mathbf{x}^\star$ is a strict local minimum of (1.18). (Interestingly enough, the Hessian matrix of the objective function itself is indefinite at $\mathbf{x}^\star$ in this case.)

We close this section by providing insight into the Lagrange multipliers.

Remark 1.61 (Interpretation of the Lagrange Multipliers). The concept of Lagrange multipliers allows to adjoin the constraints to the objective function. That is, one can view constrained optimization as a search for a vector $\mathbf{x}^\star$ at which the gradient of the objective function is a linear combination of the gradients of constraints.

Another insightful interpretation of the Lagrange multipliers is as follows. Consider the set of perturbed problems $v^\star(y) := \min\{f(\mathbf{x}) : h(\mathbf{x}) = y\}$. Suppose that there is a unique regular solution point for each y, and let $\{\boldsymbol{\xi}^\star(y)\} := \arg\min\{f(\mathbf{x}) : h(\mathbf{x}) = y\}$ denote the evolution of the optimal solution point as a function of the perturbation parameter y. Clearly,

$$v(0) = f(\mathbf{x}^\star) \quad \text{and} \quad \boldsymbol{\xi}(0) = \mathbf{x}^\star.$$

Moreover, since $h(\boldsymbol{\xi}(y)) = y$ for each y, we have

$$\nabla_y h(\boldsymbol{\xi}(y)) \; = \; 1 \; = \; \boldsymbol{\nabla}_{\mathbf{x}} h(\boldsymbol{\xi}(y))^{\mathsf{T}} \boldsymbol{\nabla}_y \boldsymbol{\xi}(y).$$

Denoting by $\lambda^\star$ the Lagrange multiplier associated to the constraint $h(\mathbf{x}) = 0$ in the original problem, we have

$$\nabla_y v(0) \; = \; \boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}^\star)^{\mathsf{T}} \boldsymbol{\nabla}_y \boldsymbol{\xi}(0) \; = \; -\lambda^\star \boldsymbol{\nabla}_{\mathbf{x}} h(\mathbf{x}^\star)^{\mathsf{T}} \boldsymbol{\nabla}_y \boldsymbol{\xi}(0) \; = \; -\lambda^\star.$$

Therefore, the Lagrange multipliers $\lambda^\star$ can be interpreted as the sensitivity of the objective function f with respect to the constraint h. Said differently, $\lambda^\star$ indicates how much the optimal cost would change, if the constraints were perturbed.

This interpretation extends straightforwardly to NLP problems having inequality constraints. The Lagrange multipliers of an active constraints $g(\mathbf{x}) \leq 0$, say $\nu^\star > 0$, can be interpreted as the sensitivity of $f(\mathbf{x}^\star)$ with respect to a change in that constraints, as $g(\mathbf{x}) \leq y$; in this case, the positivity of the Lagrange multipliers follows from the fact that by increasing y, the feasible region is relaxed, hence the optimal cost cannot increase. Regarding inactive constraints, the sensitivity interpretation also explains why the Lagrange multipliers are zero, as any infinitesimal change in the value of these constraints leaves the optimal cost value unchanged.