4 Optimality conditions

Proposition 4.1. *Let $J : \Omega \subset E \to \mathbb{R}$, with Ω open, and let $x^* \in \Omega$ be a local minimizer of J. Then:*

- *if J is differentiable at x^*, then*

$$DJ(x^*) = 0;$$

- *if J is twice differentiable at x^*, then*

$$D^2J(x^*)(h,h) \geq 0 \qquad \forall h \in E.$$

Proof. By the Taylor–Young formula, for every $h \in E$,

$$J(x^* + th) = J(x^*) + t\, DJ(x^*)(h) + o(t).$$

Since x^* is a local minimizer,

$$J(x^* + th) \geq J(x^*)$$

for all sufficiently small $t > 0$. Hence

$$DJ(x^*)(h) \geq 0 \qquad \text{and} \qquad DJ(x^*)(-h) \geq 0,$$

so $DJ(x^*)(h) = 0$ for every $h \in E$.

If J is twice differentiable at x^*, then

$$J(x^* + th) = J(x^*) + t\, DJ(x^*)(h) + \frac{t^2}{2} D^2J(x^*)(h,h) + o(t^2).$$

Using $DJ(x^*) = 0$ and the local minimality of x^* gives

$$D^2J(x^*)(h,h) \geq 0 \qquad \forall h \in E.$$

□

Examining the proof of the result above, we also obtain that in the constrained case the optimality conditions read as follows:

Proposition 4.2 (First-order condition under constraints)**.** *Let $J : \Omega \subset E \to \mathbb{R}$, with Ω open, and let $A \subset \Omega$. If $x^* \in A$ is a local minimizer of J over A and J is differentiable at x^*, then*

$$DJ(x^*)(h) \geq 0$$

for every $h \in E$ such that

$$[x^*, x^* + h] := \{x^* + th : t \in [0,1]\} \subset A. \tag{1}$$

Remark 4.1. *If A is convex, then every direction $h = x - x^*$ with $x \in A$ is admissible (in the sense that equation* (1) *holds). Hence a constrained local minimizer satisfies*

$$DJ(x^*)(x - x^*) \geq 0 \qquad \forall x \in A.$$

Definition 4.1 (Critical point)**.** *A point $x^* \in \Omega$ satisfying*

$$DJ(x^*) = 0$$

is called a critical point *of J.*

Remark 4.2. *The first- and second-order conditions above are* necessary *conditions for a local minimum. In general, however, they are not sufficient! For instance,*

$$J(x) = -x^4$$

has a critical point at $x = 0$, but $x = 0$ is a strict local maximum.

Theorem 4.1 (Second-order sufficient condition)**.** *Let $J : \Omega \subset E \to \mathbb{R}$, with Ω open, and let $x^* \in \Omega$ satisfy*

$$DJ(x^*) = 0.$$

Then the following holds:

- *Assume there exist $r > 0$ such that J is C^2 on $B_r(x^*)$ and*

  $$D^2J(x)(h,h) \geq 0 \qquad \forall x \in B_r(x^*),\ \forall h \in E.$$

  Then x^* is a local minimizer of J.

- *Assume J is twice differentiable at x^* and that there exist $\alpha > 0$*

  $$D^2J(x^*)(h,h) \geq \alpha \|h\|_E^2 \quad \forall h \in E\,.$$

  Then x^* is a strict local minimizer of J.

Proof sketch. Both results can be proven using Taylor formulas. We prove only the first point. By the second-order Taylor formula with integral remainder, for h sufficiently small,

$$J(x^* + h) = J(x^*) + DJ(x^*)(h) + \int_0^1 (1-t) D^2J(x^* + th)(h,h)\, dt.$$

Hence, since $DJ(x^*) = 0$,

$$J(x^* + h) - J(x^*) \geq 0,$$

for all sufficiently small h. □

Remark 4.3. *Both second-order conditions above are not necessary. For example,*

$$J(x) = x^4$$

has $DJ(0) = 0$ and $D^2J(0) = 0$, while $x = 0$ is nevertheless a strict local minimizer.

Exercise 4.1. *Consider*

$$J(x,y) = x^2 - xy, \qquad (x,y) \in \mathbb{R} \times [0,1].$$

Determine the constrained minimizer(s) using first-order conditions and discuss why the unconstrained second-order condition is not directly applicable at the boundary.

4.1 Convex functions

Definition 4.2 (Convex function)**.** *Let E be a vector space, $C \subset E$ a convex set, and $f : C \to \mathbb{R}$. We say that f is* convex *if*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \qquad \forall x, y \in C,\ \forall t \in [0,1]. \tag{2}$$

It is strictly convex *if the inequality is strict whenever $x \neq y$ and $t \in (0,1)$.*

Proposition 4.3. *Let $f : C \to \mathbb{R}$ be convex. Then:*

- *every local minimizer is a global minimizer;*
- *the set of global minimizers is convex;*
- *if f is strictly convex, it has at most one minimizer.*

Remark 4.4. *A minimizer need not exist, even for a convex function on a convex set.*

Proposition 4.4. *Let $C \subset E$ be open and convex, and let $f : C \to \mathbb{R}$ be differentiable. Then*

1. *f is convex if and only if*
$$f(y) \geq f(x) + Df(x)(y-x) \qquad \forall x, y \in C; \tag{3}$$
2. *f is strictly convex if and only if*
$$f(y) > f(x) + Df(x)(y-x) \qquad \forall x, y \in C, \quad x \neq y.$$

Proof sketch. For the forward implication in (1), apply convexity to the function

$$g(t) = f(x + t(y-x)), \qquad t \in [0,1],$$

and use this to estimate the derivative of f at x in the direction $y - x$. The strict version is analogous.

Conversely, assume that (3) holds and let $z = tx + (1-t)y$. Applying it at z gives

$$f(x) \geq f(z) + Df(z)(x-z), \qquad f(y) \geq f(z) + Df(z)(y-z).$$

Multiply the first inequality by t and the second by $1-t$ and add them. Since

$$t(x-z) + (1-t)(y-z) = 0,$$

we obtain

$$f(z) \leq tf(x) + (1-t)f(y).$$

The strict case is similar. □

Proposition 4.5. *Let $E = \mathbb{R}^d$, let $C \subset E$ be convex and open, and let $f : C \to \mathbb{R}$ be convex and differentiable. Then*

$$x^* \text{ is a global minimizer of } f \iff Df(x^*) = 0.$$

Remark 4.5. *The implication "global minimizer $\Rightarrow Df(x^*) = 0$" only uses the unconstrained first-order necessary condition; convexity is needed for the converse.*

5 Minimization with constraints

We consider now the case of minimization with constraints defined by a set of equalities or inequalities. The setting of this section is the following. Let E be a normed vector space, let $\Omega \subset E$ be open, and let

$$J : \Omega \to \mathbb{R}.$$

We consider minimization problems of the form

$$\inf\{J(x) : x \in A\}, \qquad A \subset \Omega,$$

where A is described by constraints. The points $x \in A$ are called *admissible.*

5.1 Equality constraints

Let $g_1, \dots, g_p : \Omega \to \mathbb{R}$ be given functions and let $c_1, \dots, c_p \in \mathbb{R}$. Define

$$A = \{x \in E : g_1(x) = c_1, \dots, g_p(x) = c_p\}.$$

Equivalently, with

$$g = (g_1, \dots, g_p) : \Omega \to \mathbb{R}^p, \qquad c = (c_1, \dots, c_p),$$

we have

$$A = \{x \in \Omega : g(x) = c\}.$$

Theorem 5.1 (First-order necessary conditions for equality constraints)**.** *Let E be a normed vector space, let $\Omega \subset E$ be open, and let $x^* \in A$ be a local minimizer of J over A. Assume that $J, g_1, \dots, g_p$ are C^1 in a neighbourhood of x^* and that the constraints are* qualified *at x^*, i.e.,*

$$Dg_1(x^*), \dots, Dg_p(x^*)$$

are linearly independent in $\mathcal{L}_c(E, \mathbb{R})$. Then there exist $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ such that

$$DJ(x^*) + \sum_{i=1}^{p} \lambda_i Dg_i(x^*) = 0.$$

Proof for the linear case. Suppose that $g_1, \dots, g_p$ are linear, and denote

$$g = (g_1, \dots, g_p).$$

By fixing a point $x_0 \in A$, the problem is equivalent to

$$\inf\{J(y + x_0) \ ; \ g(y) = 0\} \tag{4}$$

Note that problem (4) is an unconstrained problem on $\ker g$, therefore the first-order necessary conditions give

$$DJ(x^*)(h) = 0 \qquad \forall h \in \ker g.$$

Hence the condition above can be written as follows

$$\ker g \subset \ker DJ(x^*).$$

Since $DJ(x^*)$ vanishes on $\ker g$, there exists a linear functional $\ell : \mathbb{R}^p \to \mathbb{R}$ such that

$$DJ(x^*) = \ell \circ g.$$

Writing $\ell(y) = -\sum_{i=1}^p \lambda_i y_i$ gives

$$DJ(x^*) + \sum_{i=1}^p \lambda_i Dg_i(x^*) = 0.$$

□

Remark 5.1. *Note that in the linear case the linear-independence assumption just says that no constraints is redundant, and we did not really use it in the proof. If the assumption holds, however, the Lagrange multipliers (and also the map ℓ in the proof) are uniquely defined.*

In the nonlinear case, the necessary optimality conditions are a direct consequence of the implicit function theorem, recalled below. For simplicity, we only consider the finite dimensional setting.

Theorem 5.2 (Implicit function theorem)**.** *Let*

$$g : \Omega \subset \mathbb{R}^q \times \mathbb{R}^p \to \mathbb{R}^p$$

be C^1, and let $x^* = (x_1^*, x_2^*) \in \Omega$ satisfy

$$g(x^*) = y$$

and

$$\partial_2 g(x^*) \in \mathbb{R}^{p\times p}$$

be invertible. Then there exist open neighbourhoods U_1 of x_1^* and U_2 of x_2^* and a C^1 map

$$\psi : U_1 \to U_2$$

such that locally

$$\{x \in \Omega : g(x) = y\} = \{(x_1, \psi(x_1)) : x_1 \in U_1\}.$$

Moreover,

$$D\psi(x_1) = -\partial_2 g(x_1, \psi(x_1))^{-1} \partial_1 g(x_1, \psi(x_1)).$$

Remark 5.2. *Identifying $E = \mathbb{R}^{q+p}$, we may write*

$$Dg(x^*) = \begin{pmatrix} Dg_1(x^*) \\ \vdots \\ Dg_p(x^*) \end{pmatrix}.$$

The rows are linearly independent if and only if $Dg(x^*)$ has full rank p. Equivalently, after a suitable permutation of coordinates, one can choose p variables with respect to which the Jacobian $\partial_2 g(x^*)$ is invertible.

Proof for the nonlinear case in finite dimensions. We consider the case $E = \mathbb{R}^n$ with $n = p + q$. Suppose $x^* = (x_1^*, x_2^*)$ and assume that $\partial_2 g(x^*)$ is invertible (see Remark 5.2). Then, applying the implicit function theorem, the constraint set is locally given by the set of points (x_1, x_2) with $x_1 \in U_1$ and

$$x_2 = \psi(x_1)\,.$$

Hence x_1^* solves

$$\inf_{x_1 \in U_1} J(x_1, \psi(x_1))\,.$$

This is an unconstrained minimization problem and the first-order optimality conditions read

$$\partial_1 J(x_1^*, \psi(x_1^*))h_1 + \partial_2 J(x_1^*, \psi(x_1^*)) \circ D\psi(x_1^*)h_1 = 0$$

for all $h_1 \in \mathbb{R}^q$. Using the expression for $D\psi(x_1^*)$ this can also be written as follows

$$DJ(x^*)h = 0\,, \quad \forall\, h = (h_1, h_2) \text{ with } h_2 = -\partial_2 g(x^*)^{-1}\partial_1 g(x^*)h_1.$$

Now observe that

$$Dg(x^*)(h_1, h_2) = 0,$$

if and only if

$$h_2 = -\partial_2 g(x^*)^{-1}\partial_1 g(x^*)h_1.$$

Hence the first-order condition for the constrained problem becomes

$$DJ(x^*)h = 0 \qquad \forall h \in \ker Dg(x^*).$$

We conclude using the same linear algebra argument as in the linear case. □

Exercise 5.1. *Minimize*

$$f(x, y, z) = x - y + z$$

subject to

$$\begin{cases} x^2 + y^2 + z^2 = 4, \\ x + y + z = 1. \end{cases}$$

Check the constraint qualification, write the first-order optimality conditions, and determine the minimizer(s) of the problem.

Examining the proof above, we notice that the necessary conditions for the unconstrained case can be used to establish necessary conditions for our constrained minimization problem. This leads to the following result:

Theorem 5.3 (Second order sufficient condition for equality constraints)**.** *Let E be a normed vector space, $\Omega \subset E$ open, and A as above. Let $x^* \in A$, and suppose $J, g_1, \ldots, g_p$ are C^2 in a neighbourhood of x^*. Assume*

1. *$Dg_1(x^*), \ldots, Dg_p(x^*)$ are linearly independent;*
2. *there exist $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ such that*

$$DJ(x^*) + \sum_{i=1}^{p} \lambda_i Dg_i(x^*) = 0;$$

3. there exists $\alpha > 0$ such that

$$\left(D^2 J(x^*) + \sum_{i=1}^{p} \lambda_i D^2 g_i(x^*)\right)(h,h) \geq \alpha \|h\|^2 \qquad \forall h \in \ker(Dg(x^*)).$$

Then x^* is a strict local minimum of J on A.

5.2 Inequality constraints

Let $h_1, \ldots, h_q : E \to \mathbb{R}$ be C^1 functions and define

$$A = \{x \in E : h_i(x) \leq 0,\ i = 1, \ldots, q\}.$$

In this section, we assume for simplicity that E is a Hilbert space.

Definition 5.1 (Active constraints). *For $x \in A$, define the active set*

$$I(x) := \{i \in \{1, \ldots, q\} : h_i(x) = 0\}.$$

Definition 5.2 (Constraints qualification for inequality constraints). *We say that the inequality constraints are qualified at $x^* \in A$ if there exists $\bar{w} \in E$ such that*

$$Dh_i(x^*)(\bar{w}) < 0 \qquad \forall i \in I(x^*).$$

Remark 5.3. *If the active constraint gradients are linearly independent, then the inequality constraints are qualified. The converse is false in general.*

Definition 5.3 (Admissible directions). *For $x^* \in A$, the set $A(x^*)$ of admissible directions at x^* is defined as follows:*

$$A(x^*) := \{w \in E : Dh_j(x^*)(w) \leq 0 \quad \forall j \in I(x^*)\}.$$

Suppose the inequality constraints are qualified at x^*. Then every $w \in A(x^*)$ can be perturbed into a genuinely feasible direction: for every $w \in A(x^*)$ and every $\varepsilon > 0$ there exists $t_0 > 0$ such that

$$x^* + t(w + \varepsilon \bar{w}) \in A$$

for all $0 < t < t_0$.

Proof. Choose $\bar{w}$ as in the defintion of constraint qualification above, and fix $w \in A(x^*)$. If $j \notin I(x^*)$, then $h_j(x^*) < 0$, so by continuity,

$$h_j(x^* + t(w + \varepsilon \bar{w})) < 0$$

for all sufficiently small t.

If $j \in I(x^*)$, then

$$h_j(x^* + t(w + \varepsilon \bar{w})) = h_j(x^*) + tDh_j(x^*)(w) + t\varepsilon Dh_j(x^*)(\bar{w}) + o(t).$$

Since

$$h_j(x^*) = 0, \qquad Dh_j(x^*)(w) \leq 0, \qquad Dh_j(x^*)(\bar{w}) < 0,$$

we obtain

$$h_j(x^* + t(w + \varepsilon \bar{w})) \leq 0$$

for all sufficiently small t. □

Theorem 5.4 (Karush–Kuhn–Tucker conditions)**.** *Let E be a Hilbert space and let $J, h_1, \dots, h_q : E \to \mathbb{R}$ be C^1 in a neighbourhood of $x^* \in A$. Assume that x^* is a local minimizer of J over A and that the inequality constraints are qualified at x^*. Then there exist multipliers*

$$\mu_1, \dots, \mu_q \geq 0$$

such that

$$DJ(x^*) + \sum_{j=1}^{q} \mu_j Dh_j(x^*) = 0,$$

and

$$\mu_j = 0 \quad \textit{whenever} \quad h_j(x^*) < 0.$$

Equivalently,

$$\mu_j h_j(x^*) = 0 \qquad \forall j.$$

Proof sketch. By the first-order condition, for every direction w such that $[x^*, x^* + \varepsilon w] \in A$ for $\varepsilon > 0$ sufficiently small,

$$DJ(x^*)(w) \geq 0.$$

Since the constraints are qualified at x^*, using the perturbation lemma above, this inquality actually holds for all admissible directions in $A(x^*)$. Farkas' lemma (see below) yields the existence of nonnegative multipliers μ_j such that

$$DJ(x^*) + \sum_{j \in I(x^*)} \mu_j Dh_j(x^*) = 0.$$

$\square$

Theorem 5.5 (Farkas' lemma)**.** *Let $a_1, \dots, a_p, b \in E$, where E is a Hilbert space. The following are equivalent:*

1. $\langle a_j, w \rangle \geq 0\,,\ \forall j \implies \langle b, w \rangle \geq 0$;
2. *there exist $\lambda_1, \dots, \lambda_p \geq 0$ such that*

$$b = \sum_{j=1}^{p} \lambda_j a_j.$$

5.3 Equality and inequality constraints

Consider

$$A = \{x \in \Omega : g_1(x) = c_1, \dots, g_p(x) = c_p, h_1(x) \leq 0, \dots, h_q(x) \leq 0\}.$$

Here we assume that $J, g_1, \dots, g_p, h_1, \dots, h_q$ are C^1 in a neighborhood of x^*, and that x^* is a local minimizer of J on A. For this case, the notion of constraint qualification needs to be adapted to account for the presence of both equality and inequality constraint. We will use the following definition:

Definition 5.4 (Constraint qualification for equality and inequality constraints). *We say that the constraints are qualified at $x^* \in A$ if*

1. $(Dg_i(x^*))_i$ *are linearly independent;*

2. *there exists $\bar{w} \in \ker Dg(x^*)$ such that*

$$Dh_i(x^*)(\bar{w}) < 0\,, \qquad \forall i \in I(x^*).$$

Theorem 5.6 (KKT conditions with equality and inequality constraints). *Under the assumptions above, if x^* is a local minimizer of J and the constraints are qualified at x^*, there exist multipliers*

$$\lambda_1, \ldots, \lambda_p \in \mathbb{R}, \qquad \mu_1, \ldots, \mu_q \geq 0,$$

such that

$$DJ(x^*) + \sum_{i=1}^{p} \lambda_i Dg_i(x^*) + \sum_{j=1}^{q} \mu_j Dh_j(x^*) = 0,$$

with

$$\mu_j h_j(x^*) = 0 \qquad \forall j = 1, \ldots, q.$$

Remark 5.4. *In the particularly simple case in which all active equality and inequality constraint differentials at x^* are linearly independent, the constraints are qualified at x^* and the corresponding multipliers are uniquely determined. In fact, in this case, the map $T := (Dg(x^*), (Dh_j(x^*))_{j\in I(x^*)})$ is surjective so we can find a vector $\bar{w}$ such that $T(w) = (0, (-1, \ldots, -1))$.*

Proposition 5.1 (Convex case). *Let $\Omega \subset E$ be convex and let*

$$J, h_1, \ldots, h_q : \Omega \to \mathbb{R}$$

be convex and let

$$g_1, \ldots, g_p : \Omega \to \mathbb{R}$$

be affine. Suppose there exists $x^* \in A$ and multipliers

$$\lambda_1, \ldots, \lambda_p \in \mathbb{R}, \qquad \mu_j \geq 0 \quad \forall\, j \in I(x^*)\,,$$

such that

$$DJ(x^*) + \sum_{i=1}^{p} \lambda_i Dg_i(x^*) + \sum_{j \in I(x^*)} \mu_j Dh_j(x^*) = 0.$$

Then x^* is a global minimizer of J on A.

Proof. For every $x \in A$, convexity of J gives

$$J(x) \geq J(x^*) + DJ(x^*)(x - x^*).$$

Using the stationarity condition,

$$J(x) \geq J(x^*) - \sum_{i=1}^{p} \lambda_i Dg_i(x^*)(x - x^*)$$
$$- \sum_{j \in I(x^*)} \mu_j Dh_j(x^*)(x - x^*).$$

Since g_i is affine and both x and x^* satisfy the equality constraints,

$$Dg_i(x^*)(x - x^*) = 0.$$

For $j \in I(x^*)$, convexity of h_j gives

$$h_j(x) \geq h_j(x^*) + Dh_j(x^*)(x - x^*) = Dh_j(x^*)(x - x^*).$$

Because $h_j(x) \leq 0$ for $x \in A$, we obtain

$$Dh_j(x^*)(x - x^*) \leq 0,$$

and therefore

$$J(x) \geq J(x^*).$$

□

Example 5.1. *Consider*

$$\min_{x \in \mathbb{R}^2} J(x_1, x_2) = x_1 + x_2$$

subject to

$$\begin{cases} h_1(x) = x_1^2 + x_2^2 - 1 \leq 0, \\ h_2(x) = x_1 - x_2 \leq 0. \end{cases}$$

The feasible set is compact and convex, and the objective is continuous, so a global minimizer exists.

The constraints are qualified at every feasible boundary point. In fact, for $h_1(x) = 0$ and $h_1(x) \neq 0$,

$$Dh_1(x) = (2x_1, 2x_2) \neq 0,$$

and for $h_2(x) = 0$ and $h_1(x) \neq 0$,

$$Dh_2(x) = (1, -1) \neq 0.$$

At points where both constraints are active ($h_1(x) = h_2(x) = 0$), it is easy to check that the two gradients are also linearly independent.

The KKT system is

$$\begin{cases} 1 + 2\mu_1 x_1 + \mu_2 = 0, \\ 1 + 2\mu_1 x_2 - \mu_2 = 0, \\ \mu_1, \mu_2 \geq 0, \\ \mu_1(x_1^2 + x_2^2 - 1) = 0, \\ \mu_2(x_1 - x_2) = 0. \end{cases}$$

We know that any minimizer of the problem (and we know that there exists at least one) is necessarily a solution of this system!