Convex Analysis and Optimization

Chapter 5 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

LAST UPDATE April 15, 2003

CHAPTER 5: SOLUTION MANUAL

5.1 (Second Order Sufficiency Conditions for Equality-Constrained Problems)

Define the Lagrangian function $L(x, \lambda)$ to be

$$L(x, \lambda) = f(x) + \lambda' h(x).$$

Assume that f and h are twice continuously differentiable, and let $x^* \in \Re^n$ and $\lambda^* \in \Re^m$ satisfy

$$\nabla_x L(x^*, \lambda^*) = 0, \qquad \nabla_\lambda L(x^*, \lambda^*) = 0,$$

$$y' \nabla^2_{xx} L(x^*, \lambda^*) y > 0, \qquad \forall\ y \neq 0 \text{ with } \nabla h(x^*)' y = 0.$$

Show that x^* is a strict local minimum of f subject to $h(x) = 0$.

Solution: We first prove the following lemma.

Lemma 5.1: Let P and Q be two symmetric matrices. Assume that Q is positive semidefinite and P is positive definite on the nullspace of Q, that is, $x'Px > 0$ for all $x \neq 0$ with $x'Qx = 0$. Then there exists a scalar $\overline{c}$ such that

$$P + cQ : \text{positive definite}, \qquad \forall\ c > \overline{c}.$$

Proof: Assume the contrary. Then for every integer k, there exists a vector x^k with $\|x^k\| = 1$ such that

$$x^{k\prime} P x^k + k x^{k\prime} Q x^k \leq 0.$$

Since $\{x^k\}$ is bounded, there is a subsequence $\{x^k\}_{k \in K}$ converging to some $\overline{x}$, and since $\|x^k\| = 1$ for all k, we have $\|\overline{x}\| = 1$. Taking the limit superior in the above inequality, we obtain

$$\overline{x}' P \overline{x} + \limsup_{k \to \infty,\, k \in K} (k x^{k\prime} Q x^k) \leq 0. \tag{5.1}$$

Since, by the positive semidefiniteness of Q, $x^{k\prime} Q x^k \geq 0$, we see that $\{x^{k\prime} Q x^k\}_K$ must converge to zero, for otherwise the left-hand side of the above inequality would be ∞. Therefore, $\overline{x}' Q \overline{x} = 0$ and since P is positive definite, we obtain $\overline{x}' P \overline{x} > 0$. This contradicts Eq. (5.1). **Q.E.D.**

Let us introduce now the *augmented Lagrangian* function

$$L_c(x, \lambda) = f(x) + \lambda' h(x) + \frac{c}{2} \|h(x)\|^2,$$

where c is a scalar. This is the Lagrangian function for the problem

$$\begin{aligned} &\text{minimize} \quad f(x) + \frac{c}{2}\|h(x)\|^2 \\ &\text{subject to} \quad h(x) = 0, \end{aligned}$$

which has the same local minima as our original problem of minimizing $f(x)$ subject to $h(x) = 0$. The gradient and Hessian of L_c with respect to x are

$$\nabla_x L_c(x, \lambda) = \nabla f(x) + \nabla h(x)\big(\lambda + ch(x)\big),$$

$$\nabla^2_{xx} L_c(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^{m} \big(\lambda_i + ch_i(x)\big)\nabla^2 h_i(x) + c\nabla h(x)\nabla h(x)'.$$

In particular, if x^* and λ^* satisfy the given conditions, we have

$$\nabla_x L_c(x^*, \lambda^*) = \nabla f(x^*) + \nabla h(x^*)\big(\lambda^* + ch(x^*)\big) = \nabla_x L(x^*, \lambda^*) = 0, \tag{5.2}$$

$$\begin{aligned} \nabla^2_{xx} L_c(x^*, \lambda^*) &= \nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) + c\nabla h(x^*)\nabla h(x^*)' \\ &= \nabla^2_{xx} L(x^*, \lambda^*) + c\nabla h(x^*)\nabla h(x^*)'. \end{aligned}$$

By assumption, we have that $y'\nabla^2_{xx} L(x^*, \lambda^*)y > 0$ for all $y \neq 0$ such that $y'\nabla h(x^*)\nabla h(x^*)'y = 0$, so by applying Lemma 5.1 with $P = \nabla^2_{xx} L(x^*, \lambda^*)$ and $Q = \nabla h(x^*)\nabla h(x^*)'$, it follows that there exists a $\overline{c}$ such that

$$\nabla^2_{xx} L_c(x^*, \lambda^*) : \text{positive definite}, \qquad \forall\ c > \overline{c}. \tag{5.3}$$

Using now the standard sufficient optimality condition for unconstrained optimization (see e.g., [Ber99a], Section 1.1), we conclude from Eqs. (5.2) and (5.3), that for $c > \overline{c}$, x^* is an unconstrained local minimum of $L_c(\cdot, \lambda^*)$. In particular, there exist $\gamma > 0$ and $\epsilon > 0$ such that

$$L_c(x, \lambda^*) \geq L_c(x^*, \lambda^*) + \frac{\gamma}{2}\|x - x^*\|^2, \qquad \forall\ x \text{ with } \|x - x^*\| < \epsilon.$$

Since for all x with $h(x) = 0$ we have $L_c(x, \lambda^*) = f(x)$, $\nabla_\lambda L(x^*, \lambda^*) = h(x^*) = 0$, it follows that

$$f(x) \geq f(x^*) + \frac{\gamma}{2}\|x - x^*\|^2, \qquad \forall\ x \text{ with } h(x) = 0, \text{ and } \|x - x^*\| < \epsilon.$$

Thus x^* is a strict local minimum of f over $h(x) = 0$.

5.2 (Second Order Sufficiency Conditions for Inequality-Constrained Problems)

Define the Lagrangian function $L(x, \lambda, \mu)$ to be

$$L(x, \lambda, \mu) = f(x) + \lambda' h(x) + \mu' g(x).$$

Assume that f, h, and g are twice continuously differentiable, and let $x^* \in \Re^n$, $\lambda^* \in \Re^m$, and $\mu^* \in \Re^r$ satisfy

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0, \qquad h(x^*) = 0, \qquad g(x^*) \leq 0,$$

$$\mu_j^* > 0, \quad \forall\, j \in A(x^*), \qquad \mu_j^* = 0, \quad \forall\, j \notin A(x^*),$$

$$y' \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) y > 0,$$

for all $y \neq 0$ such that

$$\nabla h_i(x^*)' y = 0, \quad \forall\, i = 1, \ldots, m, \qquad \nabla g_j(x^*)' y = 0, \quad \forall\, j \in A(x^*).$$

Show that x^* is a strict local minimum of f subject to $h(x) = 0$, $g(x) \leq 0$.

Solution: We prove this result by using a transformation to an equality-constrained problem together with Exercise 5.1. Consider the equivalent equality-constrained problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_1(x) = 0, \ldots, h_m(x) = 0, \\ & g_1(x) + z_1^2 = 0, \ldots, g_r(x) + z_r^2 = 0, \end{aligned} \qquad (5.4),$$

which is an optimization problem in variables x and $z = (z_1, \ldots, z_r)$. Consider the vector (x^*, z^*), where $z^* = (z_1^*, \ldots, z_r^*)$,

$$z_j^* = \big(-g_j(x^*)\big)^{1/2}, \qquad j = 1, \ldots, r.$$

We will show that (x^*, z^*) and (λ^*, μ^*) satisfy the sufficiency conditions of Exercise 5.1, thus showing that (x^*, z^*) is a strict local minimum of problem (5.4), proving that x^* is a strict local minimum of the original inequality-constrained problem.

Let $\overline{L}(x, z, \lambda, \mu)$ be the Lagrangian function for this problem, i.e.,

$$\overline{L}(x, z, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j \big(g_j(x) + z_j^2\big).$$

We have

$$\begin{aligned} \nabla_{(x,z)} \overline{L}(x^*, z^*, \lambda^*, \mu^*)' &= \big[\nabla_x \overline{L}(x^*, z^*, \lambda^*, \mu^*)',\ \nabla_z \overline{L}(x^*, z^*, \lambda^*, \mu^*)'\big] \\ &= \big[\nabla_x L(x^*, \lambda^*, \mu^*)',\ 2\mu_1^* z_1^*, \ldots, 2\mu_r^* z_r^*\big] \\ &= [0,\ 0], \end{aligned}$$

where the last equality follows since, by assumption, we have $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$, and $\mu_j^* = 0$ for all $j \notin A(x^*)$, whereas $z_j^* = \big(-g_j(x^*)\big)^{1/2} = 0$ for all $j \in A(x^*)$. We also have

$$\begin{aligned}\nabla_{(\lambda,\mu)} \overline{L}(x^*, z^*, \lambda^*, \mu^*)' &= \Big[h_1(x^*), \ldots, h_m(x^*), \\ &\qquad \big(g_1(x^*) + (z_1^*)^2\big), \ldots, \big(g_r(x^*) + (z_r^*)^2\big)\Big] \\ &= [0, 0].\end{aligned}$$

Hence the first order conditions of the sufficiency conditions for equality-constrained problems, given in Exercise 5.1, are satisfied.

We next show that for all $(y, w) \neq (0, 0)$ satisfying

$$\nabla h(x^*)'y = 0, \qquad \nabla g_j(x^*)'y + 2z_j^* w_j = 0, \qquad j = 1, \ldots, r, \tag{5.5}$$

we have

$$(\, y' \quad w' \,) \begin{pmatrix} \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) & & 0 & & \\ & 2\mu_1^* & 0 & \ldots & 0 \\ & 0 & 2\mu_2^* & \ldots & 0 \\ 0 & \vdots & \vdots & \vdots & \vdots \\ & 0 & 0 & \ldots & 2\mu_r^* \end{pmatrix} \begin{pmatrix} y \\ w \end{pmatrix} > 0. \tag{5.6}$$

The left-hand side of the preceding expression can also be written as

$$y' \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) y + 2 \sum_{j=1}^r \mu_j^* w_j^2. \tag{5.7}$$

Let $(y, w) \neq (0, 0)$ be a vector satisfying Eq. (5.5). We have that $z_j^* = 0$ for all $j \in A(x^*)$, so it follows from Eq. (5.5) that

$$\nabla h_i(x^*)'y = 0, \quad \forall\, i = 1, \ldots, m, \qquad \nabla g_j(x^*)'y = 0, \quad \forall\, j \in A(x^*).$$

Hence, if $y \neq 0$, it follows by assumption that

$$y' \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) y > 0,$$

which implies, by Eq. (5.7) and the assumption $\mu_j^* \geq 0$ for all j, that (y, w) satisfies Eq. (5.6), proving our claim.

If $y = 0$, it follows that $w_k \neq 0$ for some $k = 1, \ldots, r$. In this case, by using Eq. (5.5), we have

$$2z_j^* w_j = 0, \qquad j = 1, \ldots, r,$$

from which we obtain that z_k^* must be equal to 0, and hence $k \in A(x^*)$. By assumption, we have that

$$\mu_j^* > 0, \qquad \forall\, j \in A(x^*).$$

This implies that $\mu_k^* w_k^2 > 0$, and therefore

$$2 \sum_{j=1}^r \mu_j^* w_j^2 > 0,$$

showing that (y, w) satisfies Eq. (5.6), completing the proof.

5.3 (Sensitivity Under Second Order Conditions)

Let x^* and (λ^*, μ^*) be a local minimum and Lagrange multiplier, respectively, of the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h_1(x) = 0, \ldots, h_m(x) = 0, \\ &\qquad\qquad\quad g_1(x) \le 0, \ldots, g_r(x) \le 0, \end{aligned} \tag{5.8}$$

satisfying the second order sufficiency conditions of Exercise 5.2. Assume that the gradients $\nabla h_i(x^*)$, $i = 1, \ldots, m$, $\nabla g_j(x^*)$, $j \in A(x^*)$, are linearly independent. Consider the family of problems

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h(x) = u, \qquad g(x) \le v, \end{aligned} \tag{5.9}$$

parameterized by the vectors $u \in \Re^m$ and $v \in \Re^r$. Then there exists an open sphere S centered at $(u, v) = (0, 0)$ such that for every $(u, v) \in S$ there is an $x(u, v) \in \Re^n$ and $\lambda(u, v) \in \Re^m$, $\mu(u, v) \in \Re^r$, which are a local minimum and associated Lagrange multiplier vectors of problem (5.9). Furthermore, $x(\cdot, \cdot)$, $\lambda(\cdot, \cdot)$, and $\mu(\cdot, \cdot)$ are continuously differentiable in S and we have $x(0, 0) = x^*$, $\lambda(0, 0) = \lambda^*$, $\mu(0, 0) = \mu^*$. In addition, for all $(u, v) \in S$, there holds

$$\nabla_u p(u, v) = -\lambda(u, v),$$

$$\nabla_v p(u, v) = -\mu(u, v),$$

where $p(u, v)$ is the optimal cost parameterized by (u, v),

$$p(u, v) = f\big(x(u, v)\big).$$

Solution: We first prove the result for the special case of equality-constrained problems.

Proposition 5.3: Let x^* and λ^* be a local minimum and Lagrange multiplier, respectively, satisfying the second order sufficiency conditions of Exercise 5.1, and assume that the gradients $\nabla h_i(x^*)$, $i = 1, \ldots, m$, are linearly independent. Consider the family of problems

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h(x) = u, \end{aligned} \tag{5.10}$$

parameterized by the vector $u \in \Re^m$. Then there exists an open sphere S centered at $u = 0$ such that for every $u \in S$, there is an $x(u) \in \Re^n$ and a $\lambda(u) \in \Re^m$, which are a local minimum-Lagrange multiplier pair of problem (5.10). Furthermore, $x(\cdot)$ and $\lambda(\cdot)$ are continuously differentiable functions within S and we have $x(0) = x^*$, $\lambda(0) = \lambda^*$. In addition, for all $u \in S$ we have

$$\nabla p(u) = -\lambda(u),$$

where $p(u)$ is the optimal cost parameterized by u, that is,

$$p(u) = f\bigl(x(u)\bigr).$$

Proof: Consider the system of equations

$$\nabla f(x) + \nabla h(x)\lambda = 0, \qquad h(x) = u. \tag{5.11}$$

For each fixed u, this system represents $n+m$ equations with $n+m$ unknowns – the vectors x and λ. For $u = 0$ the system has the solution (x^*, λ^*). The corresponding $(n+m) \times (n+m)$ Jacobian matrix with respect to (x, λ) is given by

$$J = \begin{pmatrix} \nabla^2_{xx} L(x^*, \lambda^*) & \nabla h(x^*) \\ \nabla h(x^*)' & 0 \end{pmatrix}.$$

Let us show that J is nonsingular. If it were not, some nonzero vector $(y', z')'$ would belong to the nullspace of J, that is,

$$\nabla^2_{xx} L(x^*, \lambda^*) y + \nabla h(x^*) z = 0, \tag{5.12}$$

$$\nabla h(x^*)' y = 0. \tag{5.13}$$

Premultiplying Eq. (5.12) by y' and using Eq. (5.13), we obtain

$$y' \nabla^2_{xx} L(x^*, \lambda^*) y = 0.$$

In view of Eq. (5.13), it follows that $y = 0$, for otherwise our second order sufficiency assumption would be violated. Since $y = 0$, Eq. (5.12) yields $\nabla h(x^*) z = 0$, which in view of the linear independence of the columns $\nabla h_i(x^*)$, $i = 1, \ldots, m$, of $\nabla h(x^*)$, yields $z = 0$. Thus, we obtain $y = 0$, $z = 0$, which is a contradiction. Hence, J is nonsingular.

Returning now to the system (5.11), it follows from the nonsingularity of J and the Implicit Function Theorem that for all u in some open sphere S centered at $u = 0$, there exist $x(u)$ and $\lambda(u)$ such that $x(0) = x^*$, $\lambda(0) = \lambda^*$, the functions $x(\cdot)$ and $\lambda(\cdot)$ are continuously differentiable, and

$$\nabla f\bigl(x(u)\bigr) + \nabla h\bigl(x(u)\bigr)\lambda(u) = 0, \tag{5.14}$$

$$h\bigl(x(u)\bigr) = u.$$

For u sufficiently close to 0, the vectors $x(u)$ and $\lambda(u)$ satisfy the second order sufficiency conditions for problem (5.10), since they satisfy them by assumption for $u = 0$. This is straightforward to verify by using our continuity assumptions. [If it were not true, there would exist a sequence $\{u^k\}$ with $u^k \to 0$, and a sequence $\{y^k\}$ with $\|y^k\| = 1$ and $\nabla h\bigl(x(u^k)\bigr)' y^k = 0$ for all k, such that

$$y^{k\prime} \nabla^2_{xx} L\bigl(x(u^k), \lambda(u^k)\bigr) y^k \le 0, \qquad \forall\ k.$$

By taking the limit along a convergent subsequence of $\{y^k\}$, we would obtain a contradiction of the second order sufficiency condition at (x^*, λ^*).] Hence, $x(u)$ and $\lambda(u)$ are a local minimum-Lagrange multiplier pair for problem (5.10).

There remains to show that $\nabla p(u) = \nabla_u \{f\big(x(u)\big)\} = -\lambda(u)$. By multiplying Eq. (5.14) by $\nabla x(u)$, we obtain

$$\nabla x(u)\nabla f\big(x(u)\big) + \nabla x(u)\nabla h\big(x(u)\big)\lambda(u) = 0.$$

By differentiating the relation $h\big(x(u)\big) = u$, it follows that

$$I = \nabla_u \{h\big(x(u)\big)\} = \nabla x(u)\nabla h\big(x(u)\big), \tag{5.15}$$

where I is the $m \times m$ identity matrix. Finally, by using the chain rule, we have

$$\nabla p(u) = \nabla_u \{f\big(x(u)\big)\} = \nabla x(u)\nabla f\big(x(u)\big).$$

Combining the above three relations, we obtain

$$\nabla p(u) + \lambda(u) = 0, \tag{5.16}$$

and the proof is complete. **Q.E.D.**

We next use the preceding result to show the corresponding result for inequality-constrained problems. We assume that x^* and (λ^*, μ^*) are a local minimum and Lagrange multiplier, respectively, of the problem

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } \; h_1(x) = 0, \ldots, h_m(x) = 0, \\ &\qquad\qquad\quad g_1(x) \le 0, \ldots, g_r(x) \le 0, \end{aligned} \tag{5.17}$$

and they satisfy the second order sufficiency conditions of Exercise 5.2. We also assume that the gradients $\nabla h_i(x^*)$, $i = 1, \ldots, m$, $\nabla g_j(x^*)$, $j \in A(x^*)$ are linearly independent, i.e., x^* is regular. We consider the equality-constrained problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h_1(x) = 0, \ldots, h_m(x) = 0, \\ &\qquad\qquad\quad\; g_1(x) + z_1^2 = 0, \ldots, g_r(x) + z_r^2 = 0, \end{aligned} \tag{5.18}$$

which is an optimization problem in variables x and $z = (z_1, \ldots, z_r)$. Let z^* be a vector with

$$z_j^* = \big(-g_j(x^*)\big)^{1/2}, \qquad j = 1, \ldots, r.$$

It can be seen that, since x^* and (λ^*, μ^*) satisfy the second order assumptions of Exercise 5.2, (x^*, z^*) and (λ^*, μ^*) satisfy the second order assumptions of Exercise 5.1, thus showing that (x^*, z^*) is a strict local minimum of problem (5.18)(cf. proof of Exercise 5.2). It is also straightforward to see that since x^*

is regular for problem (5.17), (x^*, z^*) is regular for problem (5.18). We consider the family of problems

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = u_i, \qquad i = 1, \ldots, m, \\ & g_j(x) + z_j^2 = v_j, \qquad j = 1, \ldots, r, \end{aligned} \tag{5.19}$$

parametrized by u and v.

Using Prop. 5.3, given in the beginning of this exercise, we have that there exists an open sphere S centered at $(u, v) = (0, 0)$ such that for every $(u, v) \in S$ there is an $x(u, v) \in \Re^n$, $z(u, v) \in \Re^r$ and $\lambda(u, v) \in \Re^m$, $\mu(u, v) \in \Re^r$, which are a local minimum and associated Lagrange multiplier vectors of problem (5.19).

We claim that the vectors $x(u, v)$ and $\lambda(u, v) \in \Re^m$, $\mu(u, v) \in \Re^r$ are a local minimum and Lagrange multiplier vector for the problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = u_i, \qquad \forall\, i = 1, \ldots, m, \\ & g_j(x) \le v_j, \qquad \forall\, j = 1, \ldots, r. \end{aligned} \tag{5.20}$$

It is straightforward to see that $x(u, v)$ is a local minimum of the preceding problem. To see that $\lambda(u, v)$ and $\mu(u, v)$ are the corresponding Lagrange multipliers, we use the first order necessary optimality conditions for problem (5.19) to write

$$\nabla f\big(x(u,v)\big) + \sum_{i=1}^m \lambda_i(u,v) \nabla h_i\big(x(u,v)\big) + \sum_{j=1}^r \mu_j(u,v) \nabla g_j\big(x(u,v)\big) = 0,$$

$$2\mu_j(u,v) z_j(u,v) = 0, \qquad j = 1, \ldots, r.$$

Since $z_j(u,v) = \Big(v_j - g_j\big(x(u,v)\big)\Big)^{1/2} > 0$ for $j \notin A\big(x(u,v)\big)$, where

$$A\big(x(u,v)\big) = \Big\{ j \mid g_j\big(x(u,v)\big) = v_j \Big\},$$

the last equation can also be written as

$$\mu_j(u,v) = 0, \qquad \forall\, j \notin A\big(x(u,v)\big). \tag{5.21}$$

Thus, to show $\lambda(u, v)$ and $\mu(u, v)$ are Lagrange multipliers for problem (5.20), there remains to show the nonnegativity of $\mu(u, v)$. For this purpose we use the second order necessary condition for the equivalent equality constrained problem (5.19). It yields

$$(\, y' \quad w' \,) \begin{pmatrix} \nabla^2_{xx} L\big(x(u,v), \lambda(u,v), \mu(u,v)\big) & & 0 & & \\ & 2\mu_1(u,v) & 0 & \ldots & 0 \\ 0 & 0 & 2\mu_2(u,v) & \ldots & 0 \\ & \vdots & \vdots & \vdots & \vdots \\ & 0 & 0 & \ldots & 2\mu_r(u,v) \end{pmatrix} \begin{pmatrix} y \\ w \end{pmatrix} \ge 0, \tag{5.22}$$

for all $y \in \Re^n$ and $w \in \Re^r$ satisfying

$$\nabla h\big(x(u,v)\big)'y = 0, \qquad \nabla g_j\big(x(u,v)\big)'y + 2z_j(u,v)w_j = 0, \qquad j \in A\big(x(u,v)\big). \tag{5.23}$$

Next let us select, for every $j \in A\big(x(u,v)\big)$, a vector (y,w) with $y = 0$, $w_j \neq 0$, $w_k = 0$ for all $k \neq j$. Such a vector satisfies the condition of Eq. (5.23). By using such a vector in Eq. (5.22), we obtain $2\mu_j(u,v)w_j^2 \geq 0$, and

$$\mu_j(u,v) \geq 0, \qquad \forall\, j \in A\big(x(u,v)\big).$$

Furthermore, by Prop. 5.3 given in the beginning of this exercise, it follows that $x(\cdot,\cdot)$, $\lambda(\cdot,\cdot)$, and $\mu(\cdot,\cdot)$ are continuously differentiable in S and we have $x(0,0) = x^*$, $\lambda(0,0) = \lambda^*$, $\mu(0,0) = \mu^*$. In addition, for all $(u,v) \in S$, there holds

$$\nabla_u p(u,v) = -\lambda(u,v),$$
$$\nabla_v p(u,v) = -\mu(u,v),$$

where $p(u,v)$ is the optimal cost of problem (5.19), parameterized by (u,v), which is the same as the optimal cost of problem (5.20), completing our proof.

5.4 (General Sufficiency Condition)

Consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \qquad g_j(x) \leq 0, \quad j = 1,\ldots,r, \end{aligned}$$

where f and g_j are real valued functions on $\Re^n$, and X is a subset of $\Re^n$. Let x^* be a feasible point, which together with a vector $\mu^* = (\mu_1^*,\ldots,\mu_r^*)$, satisfies

$$\mu_j^* \geq 0, \qquad j = 1,\ldots,r,$$
$$\mu_j^* = 0, \qquad \forall\, j \notin A(x^*),$$

and minimizes the Lagrangian function $L(x,\mu^*)$ over $x \in X$:

$$x^* \in \arg\min_{x \in X} L(x,\mu^*).$$

Show that x^* is a global minimum of the problem.

Solution: We have

$$\begin{aligned} f(x^*) &= f(x^*) + \mu^{*\prime}g(x^*) \\ &= \min_{x \in X}\big\{f(x) + \mu^{*\prime}g(x)\big\} \\ &\leq \min_{x \in X,\, g(x) \leq 0}\big\{f(x) + \mu^{*\prime}g(x)\big\} \\ &\leq \min_{x \in X,\, g(x) \leq 0} f(x) \\ &\leq f(x^*), \end{aligned}$$

where the first equality follows from the hypothesis, which implies that $\mu^{*\prime}g(x^*) = 0$, the next-to-last inequality follows from the nonnegativity of μ^*, and the last inequality follows from the feasibility of x^*. It folows that equality holds throughout, and x^* is a optimal solution.

5.5

The purpose of this exercise is to work out an alternative proof of Lemma 5.3.1, assuming that $N = \{0\}$ [which corresponds to the case where there is no abstract set constraint ($X = \Re^n$)]. Let $a_0, \ldots, a_r$ be given vectors in $\Re^n$. Suppose that the set

$$M = \left\{ \mu \geq 0 \;\middle|\; a_0 + \sum_{j=1}^{r} \mu_j a_j = 0 \right\}$$

is nonempty, and let μ^* be the vector of minimum norm in M. For any $\gamma > 0$, consider the function

$$L_\gamma(d, \mu) = \left(a_0 + \sum_{j=1}^{r} \mu_j a_j \right)' d + \frac{\gamma}{2}\|d\|^2 - \frac{1}{2}\|\mu\|^2.$$

(a) Show that

$$\begin{aligned} -\tfrac{1}{2}\|\mu^*\|^2 &= \sup_{\mu \geq 0} \inf_{d \in \Re^n} L_0(d, \mu) \\ &\leq \inf_{d \in \Re^n} \sup_{\mu \geq 0} L_0(d, \mu) \\ &= \inf_{d \in \Re^n} \left\{ a_0' d + \tfrac{1}{2} \sum_{j=1}^{r} \big((a_j' d)^+ \big)^2 \right\}. \end{aligned} \tag{5.24}$$

(b) Use the lower bound of part (a) and the theory of Section 2.3 on the existence of solutions of quadratic programs to conclude that the infimum in the right-hand side above is attained for some $d^* \in \Re^n$.

(c) Show that for every $\gamma > 0$, L_γ has a saddle point (d^γ, μ^γ) such that

$$\mu_j^\gamma = (a_j' d^\gamma)^+, \qquad j = 1, \ldots, r.$$

Furthermore,

$$L_\gamma(d^\gamma, \mu^\gamma) = -\frac{\left\| a_0 + \sum_{j=1}^{r} \mu_j^\gamma a_j \right\|^2}{2\gamma} - \frac{1}{2}\|\mu^\gamma\|^2 \geq -\frac{1}{2}\|\mu^*\|^2.$$

(d) Use part (c) to show that $\|\mu^\gamma\| \leq \|\mu^*\|$, and use the minimum norm property of μ^* to conclude that as $\gamma \to 0$, we have $\mu^\gamma \to \mu^*$ and $L_\gamma(d^\gamma, \mu^\gamma) \to -(1/2)\|\mu^*\|^2$.

(e) Use part (d) and Eq. (5.24) to show that (d^*, μ^*) is a saddle point of L_0, and that

$$a_0' d^* = -\|\mu^*\|^2, \qquad (a_j' d^*)^+ = \mu_j^*, \quad j = 1, \ldots, r.$$

Solution: (a) We note that

$$\inf_{d \in \Re^n} L_0(d, \mu) = \begin{cases} -\frac{1}{2}\|\mu\|^2 & \text{if } \mu \in M, \\ -\infty & \text{otherwise,} \end{cases}$$

so since μ^* is the vector of minimum norm in M, we obtain for all $\gamma > 0$,

$$\begin{aligned} -\frac{1}{2}\|\mu^*\|^2 &= \sup_{\mu \geq 0} \inf_{d \in \Re^n} L_0(d, \mu) \\ &\leq \inf_{d \in \Re^n} \sup_{\mu \geq 0} L_0(d, \mu), \end{aligned}$$

where the inequality follows from the minimax inequality (cf. Chapter 2). For any $d \in \Re^n$, the supremum of $L_0(d, \mu)$ over $\mu \geq 0$ is attained at

$$\mu_j = (a_j' d)^+, \qquad j = 1, \ldots, r.$$

[to maximize $\mu_j a_j' d - (1/2)\mu_j^2$ subject to the constraint $\mu_j \geq 0$, we calculate the unconstrained maximum, which is $a_j' d$, and if it is negative we set it to 0, so that the maximum subject to $\mu_j \geq 0$ is attained for $\mu_j = (a_j' d)^+$]. Hence, it follows that, for any $d \in \Re^n$,

$$\sup_{\mu \geq 0} L_0(d, \mu) = a_0' d + \frac{1}{2} \sum_{j=1}^r \big((a_j' d)^+\big)^2,$$

which yields the desired relations.

(b) Since the infimum of the quadratic cost function $a_0' d + \frac{1}{2}\sum_{j=1}^r \big((a_j' d)^+\big)^2$ is bounded below, as given in part (a), it follows from the results of Section 2.3 that the infimum of this function is attained at some $d^* \in \Re^n$.

(c) From the Saddle Point Theorem, for all $\gamma > 0$, the coercive convex/concave quadratic function L_γ has a saddle point, denoted (d^γ, μ^γ), over $d \in \Re^n$ and $\mu \geq 0$. This saddle point is unique and can be easily characterized, taking advantage of the quadratic nature of L_γ. In particular, similar to part (a), the maximization over $\mu \geq 0$ when $d = d^\gamma$ yields

$$\mu_j^\gamma = (a_j' d^\gamma)^+, \qquad j = 1, \ldots, r. \tag{5.25}$$

Moreover, we can find $L_\gamma(d^\gamma, \mu^\gamma)$ by minimizing $L_\gamma(d, \mu^\gamma)$ over $d \in \Re^n$. To find the unconstrained minimum d^γ, we take the gradient of $L_\gamma(d, \mu^\gamma)$ and set it equal to 0. This yields

$$d^\gamma = -\frac{a_0 + \sum_{j=1}^r \mu_j^\gamma a_j}{\gamma}.$$

Hence,

$$L_\gamma(d^\gamma, \mu^\gamma) = -\frac{\Big\|a_0 + \sum_{j=1}^r \mu_j^\gamma a_j\Big\|^2}{2\gamma} - \frac{1}{2}\|\mu^\gamma\|^2.$$

We also have

$$
\begin{aligned}
-\frac{1}{2}\|\mu^*\|^2 &= \sup_{\mu\ge 0} \inf_{d\in\Re^n} L_0(d,\mu) \\
&\le \inf_{d\in\Re^n} \sup_{\mu\ge 0} L_0(d,\mu) \\
&\le \inf_{d\in\Re^n} \sup_{\mu\ge 0} L_\gamma(d,\mu) \\
&= L_\gamma(d^\gamma,\mu^\gamma),
\end{aligned}
\tag{5.26}
$$

where the first two relations follow from part (a), thus yielding the desired relation.

(d) From part (c), we have

$$
-\frac{1}{2}\|\mu^\gamma\|^2 \ge L_\gamma(d^\gamma,\mu^\gamma) = -\frac{\left\|a_0 + \sum_{j=1}^r \mu_j^\gamma a_j\right\|^2}{2\gamma} - \frac{1}{2}\|\mu^\gamma\|^2 \ge -\frac{1}{2}\|\mu^*\|^2. \tag{5.27}
$$

From this, we see that $\|\mu^\gamma\| \le \|\mu^*\|$, so that μ^γ remains bounded as $\gamma \to 0$. By taking the limit above as $\gamma \to 0$, we see that

$$
\lim_{\gamma\to 0}\left(a_0 + \sum_{j=1}^r \mu_j^\gamma a_j\right) = 0,
$$

so any limit point of μ^γ, call it $\overline{\mu}$, satisfies $-\left(a_0 + \sum_{j=1}^r \overline{\mu}_j a_j\right) = 0$. Since $\mu^\gamma \ge 0$, it follows that $\overline{\mu} \ge 0$, so $\overline{\mu} \in M$. We also have $\|\overline{\mu}\| \le \|\mu^*\|$ (since $\|\mu^\gamma\| \le \|\mu^*\|$), so by using the minimum norm property of μ^*, we conclude that any limit point $\overline{\mu}$ of μ^γ must be equal to μ^*. Thus, $\mu^\gamma \to \mu^*$. From Eq. (5.27), we then obtain

$$
L_\gamma(d^\gamma,\mu^\gamma) \to -\frac{1}{2}\|\mu^*\|^2. \tag{5.28}
$$

(e) Equations (5.26) and (5.28), together with part (b), show that

$$
L_0(d^*,\mu^*) = \inf_{d\in\Re^n}\sup_{\mu\ge 0} L_0(d,\mu) = \sup_{\mu\ge 0}\inf_{d\in\Re^n} L_0(d,\mu),
$$

[thus proving that (d^*,μ^*) is a saddle point of $L_0(d,\mu)$], and that

$$
a_0'd^* = -\|\mu^*\|^2, \qquad (a_j'd^*)^+ = \mu_j^*, \quad j = 1,\dots,r.
$$

5.6 (Strict Complementarity)

Consider the problem

$$
\begin{aligned}
&\text{minimize} \quad f(x) \\
&\text{subject to} \quad g_j(x) \le 0, \qquad j = 1,\dots,r,
\end{aligned}
$$

where $f : \Re^n \mapsto \Re$ and $g_j : \Re^n \mapsto \Re$ are smooth functions. A Lagrange multiplier $\{\mu_1^*, \ldots, \mu_r^*\}$, corresponding to a local minimum x^*, is said to satisfy *strict complementarity* if for all j such that $g_j(x^*) = 0$, we have $\mu_j^* > 0$. Show that a Lagrange multiplier that satisfies strict complementarity need not be informative, and conversely, a Lagrange multiplier that is informative need not satisfy strict complementarity.

Solution: Consider the following example

$$\begin{aligned} &\text{minimize} \quad x_1 + x_2 \\ &\text{subject to } x_1 \le 0, \; x_2 \le 0, \; -x_1 - x_2 \le 0. \end{aligned}$$

The only feasible vector is $x^* = (0, 0)$, which is therefore also the optimal solution of this problem. The vector $(1, 1, 2)'$ is a Lagrange multiplier vector which satisfies strict complementarity. However, it is not possible to find a vector that violates simultaneously all the constraints, showing that this Lagrange multiplier vector is not informative.

For the converse statement, consider the example of Fig. 5.1.3. The Lagrange multiplier vectors, that involve three nonzero components out of four, are informative, but they do not satisfy strict complementarity.

5.7

Consider the problem

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{n} f_i(x_i) \\ &\text{subject to} \quad x \in S, \qquad x_i \in X_i, \quad i = 1, \ldots, n, \end{aligned}$$

where $f_i : \Re \mapsto \Re$ are smooth functions, X_i are closed intervals of real numbers of $\Re^n$, and S is a subspace of $\Re^n$. Let x^* be a local minimum. Introduce artificial optimization variables $z_1, \ldots, z_n$ and the linear constraints $x_i = z_i$, $i = 1, \ldots, n$, while replacing the constraint $x \in S$ with $z \in S$, so that the problem becomes

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{n} f_i(x_i) \\ &\text{subject to} \quad z \in S, \qquad x_i \in X_i, \qquad x_i = z_i, \quad i = 1, \ldots, n. \end{aligned}$$

Show that there exists a Lagrange multiplier vector $\lambda^* = (\lambda_1^*, \ldots, \lambda_n^*)$ such that $\lambda^* \in S^\perp$ and

$$\big(\nabla f_i(x_i^*) + \lambda_i^*\big)(x_i - x_i^*) \ge 0, \qquad \forall\, x_i \in X_i, \; i = 1, \ldots, n.$$

Solution: Let x^* be a local minimum of the problem

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{n} f_i(x_i) \\ &\text{subject to} \quad x \in S, \qquad x_i \in X_i, \quad i = 1, \ldots, n, \end{aligned}$$

where $f_i : \Re \mapsto \Re$ are smooth functions, X_i are closed intervals of real numbers of $\Re^n$, and S is a subspace of $\Re^n$. We introduce artificial optimization variables $z_1, \ldots, z_n$ and the linear constraints $x_i = z_i$, $i = 1, \ldots, n$, while replacing the constraint $x \in S$ with $z \in S$, so that the problem becomes

$$\begin{aligned} &\text{minimize } \sum_{i=1}^{n} f_i(x_i) \\ &\text{subject to } \ z \in S, \qquad x_i \in X_i, \qquad x_i = z_i, \quad i = 1, \ldots, n. \end{aligned} \tag{5.29}$$

Let $a_1, \ldots, a_m$ be a basis for $S^\perp$, the orthogonal complement of S. Then, we can represent S as

$$S = \{y \mid a_j' y = 0,\ \forall\, j = 1, \ldots, m\}.$$

We also represent the closed intervals X_i as

$$X_i = \{y \mid c_i \leq y \leq d_i\}.$$

With the previous identifications, the constraint set of problem (5.29) can be described alternatively as

$$\begin{aligned} \text{minimize } & \sum_{i=1}^{n} f_i(x_i) \\ \text{subject to } & a_j' z = 0, \quad j = 1, \ldots, m, \\ & c_i \leq x_i \leq d_i, \quad i = 1, \ldots, n, \\ & x_i = z_i, \quad i = 1, \ldots, n \end{aligned}$$

[cf. extended representation of the constraint set of problem (5.29)]. This is a problem with linear constraints, so by Prop. 5.4.1, it admits Lagrange multipliers. But, by Prop. 5.6.1, this implies that the problem admits Lagrange multipliers in the original representation as well. We associate a Lagrange multiplier λ_i^* with each equality constraint $x_i = z_i$ in problem (5.29). By taking the gradient with respect to the variable x, and using the definition of Lagrange multipliers, we get

$$\big(\nabla f_i(x_i^*) + \lambda_i^*\big)(x_i - x_i^*) \geq 0, \qquad \forall\, x_i \in X_i,\ i = 1, \ldots, n,$$

whereas, by taking the gradient with respect to the variable z, we obtain $\lambda^* \in S^\perp$, thus completing the proof.

5.8

Show that if X is regular at x^* the constraint qualifications CQ5a and CQ6 are equivalent.

Solution: We first show that CQ5a implies CQ6. Assume CQ5a holds:

(a) There does not exist a nonzero vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*).$$

(b) There exists a $d \in N_X(x^*)^* = T_X(x^*)$ (since X is regular at x^*) such that

$$\nabla h_i(x^*)' d = 0, \quad i = 1, \ldots, m, \qquad \nabla g_j(x^*)' d < 0, \quad \forall\, j \in A(x^*).$$

To arrive at a contradiction, assume that CQ6 does not hold, i.e., there are scalars $\lambda_1, \ldots, \lambda_m$, $\mu_1, \ldots, \mu_r$, not all of them equal to zero, such that

(i)
$$-\left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right) \in N_X(x^*).$$

(ii) $\mu_j \geq 0$ for all $j = 1, \ldots, r$, and $\mu_j = 0$ for all $j \notin A(x^*)$.

In view of our assumption that X is regular at x^*, condition (i) can be written as

$$-\left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right) \in T_X(x^*)^*,$$

or equivalently,

$$\left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right)' y \geq 0, \qquad \forall\, y \in T_X(x^*). \tag{5.30}$$

Since not all the λ_i and μ_j are equal to 0, we conclude that $\mu_j > 0$ for at least one $j \in A(x^*)$; otherwise condition (a) of CQ5a would be violated. Since $\mu_j^* \geq 0$ for all j, with $\mu_j^* = 0$ for $j \notin A(x^*)$ and $\mu_j^* > 0$ for at least one j, we obtain

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*)' d + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)' d < 0,$$

where $d \in T_X(x^*)$ is the vector in condition (b) of CQ5a. But this contradicts Eq. (5.30), showing that CQ6 holds.

Conversely, assume that CQ6 holds. It can be seen that this implies condition (a) of CQ5a. Let H denote the subspace spanned by the vectors $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$, and let G denote the cone generated by the vectors $\nabla g_j(x^*)$, $j \in A(x^*)$. Then, the orthogonal complement of H is given by

$$H^\perp = \left\{ y \mid \nabla h_i(x^*)' y = 0,\ \forall\, i = 1, \ldots, m \right\},$$

whereas the polar of G is given by

$$G^* = \left\{ y \mid \nabla g_j(x^*)' y \leq 0,\ \forall\, j \in A(x^*) \right\},$$

(cf. the results of Section 3.1). The interior of G^* is the set

$$\mathrm{int}(G^*) = \big\{ y \mid \nabla g_j(x^*)'y < 0,\ \forall\, j \in A(x^*) \big\}.$$

Under CQ6, we have $\mathrm{int}(G^*) \neq \emptyset$, since otherwise the vectors $\nabla g_j(x^*)$, $j \in A(x^*)$ would be linearly dependent, contradicting CQ6. Similarly, under CQ6, we have

$$H^\perp \cap \mathrm{int}(G^*) \neq \emptyset. \tag{5.31}$$

To see this, assume the contrary, i.e., $H^\perp$ and $\mathrm{int}(G^*)$ are disjoint. The sets $H^\perp$ and $\mathrm{int}(G^*)$ are convex, therefore by the Separating Hyperplane Theorem, there exists some nonzero vector ν such that

$$\nu'x \le \nu'y, \qquad \forall\, x \in H^\perp,\ \forall\, y \in \mathrm{int}(G^*),$$

or equivalently,

$$\nu'(x-y) \le 0, \qquad \forall\, x \in H^\perp,\ \forall\, y \in G^*,$$

which implies, using also Exercise 3.4., that

$$\nu \in (H^\perp - G^*)^* = H \cap (-G).$$

But this contradicts CQ6, and proves Eq. (5.31).

Finally, we show that CQ6 implies condition (b) of CQ5a. Assume, to arrive at a contradiction, that condition (b) of CQ5a does not hold. This implies that

$$N_X(x^*)^* \cap H^\perp \cap \mathrm{int}\big(G^*\big) = \emptyset.$$

Since X is regular at x^*, the preceding is equivalent to

$$T_X(x^*) \cap H^\perp \cap \mathrm{int}\big(G^*\big) = \emptyset.$$

The regularity of X at x^* implies that $T_X(x^*)$ is convex. Similarly, since the interior of a convex set is convex and the intersection of two convex sets is convex, it follows that the set $H^\perp \cap \mathrm{int}\big(G^*\big)$ is convex. It is also nonempty by Eq. (5.31). Thus, by the Separating Hyperplane Theorem, there exists some vector $a \neq 0$ such that

$$a'x \le a'y, \qquad \forall\, x \in T_X(x^*),\ \forall\, y \in H^\perp \cap \mathrm{int}\big(G^*\big),$$

or equivalently,

$$a'(x-y) \le 0, \qquad \forall\, x \in T_X(x^*),\ \forall\, y \in H^\perp \cap G^*,$$

which implies that

$$a \in \big(T_X(x^*) - (H^\perp \cap G^*)\big)^*.$$

We have

$$\begin{aligned}
\big(T_X(x^*) - (H^\perp \cap G^*)\big)^* &= T_X(x^*)^* \cap \big(-(H^\perp \cap G^*)^*\big) \\
&= T_X(x^*)^* \cap \Big(-\big(\mathrm{cl}(H+G)\big)\Big) \\
&= T_X(x^*)^* \cap \big(-(H+G)\big) \\
&= N_X(x^*) \cap \big(-(H+G)\big),
\end{aligned}$$

where the second equality follows since $H^\perp$ and G^* are closed and convex, and the third equality follows since H and G are both polyhedral cones (cf. Chapter 3). Combining the preceding relations, it follows that there exists a nonzero vector a that belongs to the set

$$N_X(x^*) \cap \big(-(H+G)\big).$$

But this contradicts CQ6, thus completing our proof.

5.9 (Minimax Problems)

Derive Lagrange multiplier-like optimality conditions for the minimax problem

$$\begin{aligned} &\text{minimize} \quad \max\big\{f_1(x),\ldots,f_p(x)\big\} \\ &\text{subject to} \ \ x \in X, \end{aligned}$$

where X is a closed set, and the functions f_i are smooth. *Hint*: Convert the problem to the smooth problem

$$\begin{aligned} &\text{minimize} \quad z \\ &\text{subject to} \ \ x \in X, \quad f_i(x) \le z, \quad i = 1,\ldots,p, \end{aligned}$$

and show that CQ5 holds.

Solution: Let x^* be a local minimum of the minimax problem,

$$\begin{aligned} &\text{minimize} \quad \max\big\{f_1(x),\ldots,f_p(x)\big\} \\ &\text{subject to} \ \ x \in X. \end{aligned}$$

We introduce an additional scalar variable z and convert the preceding problem to the smooth problem

$$\begin{aligned} &\text{minimize} \quad z \\ &\text{subject to} \ \ x \in X, \quad f_i(x) \le z, \quad i = 1,\ldots,p, \end{aligned}$$

which is an optimization problem in the variables x and z and with an abstract set constraint $(x,z) \in X \times \Re$. Let

$$z^* = \max\big\{f_1(x^*),\ldots,f_p(x^*)\big\}.$$

It can be seen that (x^*,z^*) is a local minimum of the above problem.

It is straightforward to show that

$$N_{X\times\Re}(x^*,z^*) = N_X(x^*) \times \{0\}, \tag{5.32}$$

and

$$N_{X\times\Re}(x^*,z^*)^* = N_X(x^*)^* \times \Re. \tag{5.33}$$

Let $d = (0, 1)$. By Eq. (5.33), this vector belongs to the set $N_{X\times\Re}(x^*, z^*)^*$, and also

$$[\nabla f_i(x^*)',\ -1] \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -1 < 0, \qquad \forall\ i = 1, \ldots, p.$$

Hence, CQ5a is satisfied, which together with Eq. (5.32) implies that there exists a nonnegative vector $\mu^* = (\mu_1^*, \ldots, \mu_p^*)$ such that

(i) $-\left(\sum_{j=1}^p \mu_j^* \nabla f_i(x^*)\right) \in N_X(x^*)$.

(ii) $\sum_{j=1}^p \mu_j^* = 1$.

(iii) For all $j = 1, \ldots, p$, if $\mu_j^* > 0$, then

$$f_j(x^*) = \max\big\{f_1(x^*), \ldots, f_p(x^*)\big\}.$$

5.10 (Exact Penalty Functions)

Consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in C, \end{aligned} \tag{5.34}$$

where

$$C = X \cap \big\{x \mid h_1(x) = 0, \ldots, h_m(x) = 0\big\} \cap \big\{x \mid g_1(x) \le 0, \ldots, g_r(x) \le 0\big\},$$

and assume that f, h_i, and g_j are smooth functions. Let F_c be the exact penalty function, i.e.,

$$F_c(x) = f(x) + c\left(\sum_{i=1}^m |h_i(x)| + \sum_{j=1}^r g_j^+(x)\right),$$

where c is a positive scalar.

(a) Suppose that x^* is a local minimum of problem (5.34), and that for some given $c > 0$, x^* is also a local minimum of F_c over X. Show that there exists an R-multiplier vector (λ^*, μ^*) for problem (5.34) such that

$$|\lambda_i^*| \le c, \quad i = 1, \ldots, m, \qquad \mu_j^* \in [0, c], \quad j = 1, \ldots, r. \tag{5.35}$$

(b) Derive conditions that guarantee that if x^* is a local minimum of problem (5.34) and (λ^*, μ^*) is a corresponding Lagrange multiplier vector, then x^* is also a local minimum of F_c over X for all c such that Eq. (5.35) holds.

Solution: We consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in C, \end{aligned} \tag{5.36}$$

where

$$C = X \cap \big\{x \mid h_1(x) = 0, \ldots, h_m(x) = 0\big\} \cap \big\{x \mid g_1(x) \le 0, \ldots, g_r(x) \le 0\big\},$$

and the exact penalty function

$$F_c(x) = f(x) + c\left(\sum_{i=1}^m |h_i(x)| + \sum_{j=1}^r g_j^+(x)\right),$$

where c is a positive scalar.

(a) In view of our assumption that, for some given $c > 0$, x^* is also a local minimum of F_c over X, we have, by Prop. 5.5.1, that there exist $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_r$ such that

$$-\left(\nabla f(x^*) + c\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^r \mu_j \nabla g_j(x^*)\right)\right) \in N_X(x^*),$$

$$\lambda_i = 1 \quad \text{if } h_i(x^*) > 0, \qquad \lambda_i = -1 \quad \text{if } h_i(x^*) < 0,$$

$$\lambda_i \in [-1, 1] \quad \text{if } h_i(x^*) = 0,$$

$$\mu_j = 1 \quad \text{if } g_j(x^*) > 0, \qquad \mu_j = 0 \quad \text{if } g_j(x^*) < 0,$$

$$\mu_j \in [0, 1] \quad \text{if } g_j(x^*) = 0.$$

By the definition of R-multipliers, the preceding relations imply that the vector $(\lambda^*, \mu^*) = c(\lambda, \mu)$ is an R-multiplier for problem (5.36) such that

$$|\lambda_i^*| \le c, \quad i = 1, \ldots, m, \qquad \mu_j^* \in [0, c], \quad j = 1, \ldots, r. \tag{5.37}$$

(b) Assume that the functions f and the g_j are convex, the functions h_i are linear, and the set X is convex. Since x^* is a local minimum of problem (5.36), and (λ^*, μ^*) is a corresponding Lagrange multiplier vector, we have by definition that

$$\left(\nabla f(x^*) + \left(\sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right)\right)'(x - x^*) \ge 0, \qquad \forall\, x \in X.$$

In view of the convexity assumptions, this is a sufficient condition for x^* to be a local minimum of the function $f(x) + \sum_{i=1}^m \lambda_i^* h_i(x) + \sum_{j=1}^r \mu_j^* g_j(x)$ over $x \in X$. Since x^* is feasible for the original problem, and (λ^*, μ^*) satisfy Eq. (5.37), we have for all $x \in X$,

$$\begin{aligned}
F_C(x^*) &= f(x^*) \\
&\le f(x) + \sum_{i=1}^m \lambda_i^* h_i(x) + \sum_{j=1}^r \mu_j^* g_j(x) \\
&\le f(x) + c\left(\sum_{i=1}^m h_i(x) + \sum_{j=1}^r g_j(x)\right) \\
&\le f(x) + c\left(\sum_{i=1}^m |h_i(x)| + \sum_{j=1}^r g_j^+(x)\right) \\
&= F_c(x),
\end{aligned}$$

implying that x^* is a local minimum of F_c over X.

5.11 (Extended Representations)

This exercise generalizes Prop. 5.6.1 by including an additional set constraint in the extended representation. Assume that the set constraint can be described as

$$X = \bigl\{x \in \overline{X} \mid h_i(x) = 0,\, i = m+1, \ldots, \overline{m},\, g_j(x) \le 0,\, j = r+1, \ldots, \overline{r}\bigr\},$$

so that C is represented alternatively as

$$C = \overline{X} \cap \bigl\{x \mid h_1(x) = 0, \ldots, h_{\overline{m}}(x) = 0\bigr\} \cap \bigl\{x \mid g_1(x) \le 0, \ldots, g_{\overline{r}}(x) \le 0\bigr\}.$$

We call this the extended representation of C. Assuming that $\overline{X}$ is closed and that all the functions h_i and g_j are smooth, show the following:

(a) If the constraint set admits Lagrange multipliers in the extended representation, it admits Lagrange multipliers in the original representation.

(b) If the constraint set admits an exact penalty in the extended representation, it admits an exact penalty in the original representation.

Solution: (a) The hypothesis implies that for every smooth cost function f for which x^* is a local minimum there exist scalars $\lambda_1^*, \ldots, \lambda_{\overline{m}}^*$ and $\mu_1^*, \ldots, \mu_{\overline{r}}^*$ satisfying

$$\left(\nabla f(x^*) + \sum_{i=1}^{\overline{m}} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{\overline{r}} \mu_j^* \nabla g_j(x^*)\right)' y \ge 0, \qquad \forall\, y \in T_{\overline{X}}(x^*), \tag{5.38}$$

$$\mu_j^* \ge 0, \qquad \forall\, j = 1, \ldots, \overline{r},$$

$$\mu_j^* = 0, \qquad \forall\, j \notin \overline{A}(x^*),$$

where

$$\overline{A}(x^*) = \bigl\{j \mid g_j(x^*) = 0,\, j = 1, \ldots, \overline{r}\bigr\}.$$

Since $X \subset \overline{X}$, we have $T_X(x^*) \subset T_{\overline{X}}(x^*)$, so Eq. (5.38) implies that

$$\left(\nabla f(x^*) + \sum_{i=1}^{\overline{m}} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{\overline{r}} \mu_j^* \nabla g_j(x^*)\right)' y \ge 0, \qquad \forall\, y \in T_X(x^*). \tag{5.39}$$

Let $V(x^*)$ denote the set

$$\begin{aligned} V(x^*) = \bigl\{y \mid{}& \nabla h_i(x^*)'y = 0,\ i = m+1, \ldots, \overline{m}, \\ & \nabla g_j(x^*)'y \le 0,\ j = r+1, \ldots, \overline{r} \text{ with } j \in \overline{A}(x^*)\bigr\}. \end{aligned}$$

We claim that $T_X(x^*) \subset V(x^*)$. To see this, let y be a nonzero vector that belongs to $T_X(x^*)$. Then, there exists a sequence $\{x_k\} \subset X$ such that $x_k \ne x^*$ for all k and

$$\frac{x_k - x^*}{\|x_k - x^*\|} \to \frac{y}{\|y\|}.$$

Since $x_k \in X$, for all $i = m+1, \ldots, \overline{m}$ and k, we have

$$0 = h_i(x_k) = h_i(x^*) + \nabla h_i(x^*)'(x_k - x^*) + o(\|x_k - x^*\|),$$

which can be written as

$$\nabla h_i(x^*)' \frac{(x_k - x^*)}{\|x_k - x^*\|} + \frac{o(\|x_k - x^*\|)}{\|x_k - x^*\|} = 0.$$

Taking the limit as $k \to \infty$, we obtain

$$\nabla h_i(x^*)' y = 0, \qquad \forall\ i = m+1, \ldots, \overline{m}. \tag{5.40}$$

Similarly, we have for all $j = r+1, \ldots, \overline{r}$ with $j \in \overline{A}(x^*)$ and for all k

$$0 \geq g_j(x_k) = g_j(x^*) + \nabla g_j(x^*)'(x_k - x^*) + o(\|x_k - x^*\|),$$

which can be written as

$$\nabla g_j(x^*)' \frac{(x_k - x^*)}{\|x_k - x^*\|} + \frac{o(\|x_k - x^*\|)}{\|x_k - x^*\|} \leq 0.$$

By taking the limit as $k \to \infty$, we obtain

$$\nabla g_j(x^*)' y \leq 0, \qquad \forall\ j = r+1, \ldots, \overline{r} \text{ with } j \in \overline{A}(x^*).$$

Equation (5.40) and the preceding relation imply that $y \in V(x^*)$, showing that $T_X(x^*) \subset V(x^*)$.

Hence Eq. (5.39) implies that

$$\left(\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) \right)' y \geq 0, \qquad \forall\ y \in T_X(x^*),$$

and it follows that λ_i^*, $i = 1, \ldots, m$, and μ_j^*, $j = 1, \ldots, r$, are Lagrange multipliers for the original representation.

(b) Consider the exact penalty function for the extended representation:

$$\overline{F}_c(x) = f(x) + c \left(\sum_{i=1}^{\overline{m}} |h_i(x)| + \sum_{j=1}^{\overline{r}} g_j^+(x) \right).$$

We have $F_c(x) = \overline{F}_c(x)$ for all $x \in X$. Hence if $x^* \in C$ is a local minimum of $\overline{F}_c(x)$ over $x \in \overline{X}$, it is also a local minimum of $F_c(x)$ over $x \in X$. Thus, for a given $c > 0$, if x^* is both a strict local minimum of f over C and a local minimum of $\overline{F}_c(x)$ over $x \in \overline{X}$, it is also a local minimum of $F_c(x)$ over $x \in X$.