JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 23, No. 4, DECEMBER 1977

Approximation Procedures Based on the Method of Multipliers[1]

D. P. BERTSEKAS[2]

Communicated by D. G. Luenberger

Abstract. In this paper, we consider a method for solving certain optimization problems with constraints, nondifferentiabilities, and other ill-conditioning terms in the cost functional by approximating them by well-behaved optimization problems. The approach is based on methods of multipliers. The convergence properties of the methods proposed can be inferred from corresponding properties of multiplier methods with partial elimination of constraints. A related analysis is provided in this paper.

Key Words. Approximation, multiplier methods, nonlinear programming, minimax problems, Chebyshev approximation.

1. Introduction

Many optimization problems of interest can be written as

$$\begin{aligned}&\text{minimize } f(x)+\sum_{i=1}^{m}\gamma_i[g_i(x)],\\&\text{subject to } x\in X,\end{aligned}\tag{1}$$

where

$$f: R^n \to R$$

is a continuously differentiable function on R^n (n-dimensional Euclidean space),

$$g_i: R^n \to R^{r_i}, \qquad i=1,\ldots,m,$$

[1] This work was supported in part by the Joint Services Electronics Program (US Army, US Navy, and US Air Force) under Contract No. DAAB-07-72-C-0259, and by the National Science Foundation under Grant No. ENG-74-19332.

[2] Associate Professor, Department of Electrical Engineering and Coordinated Science Laboratory, University of Illinois, Urbana, Illinois.

are continuously differentiable mappings from R^n to R^{r_i}, respectively (r_i is a positive integer for each i), and X is a given subset of R^n. The functions

$$\gamma_i: R^{r_i} \to (-\infty, +\infty], \qquad i=1, \ldots, m,$$

are assumed to be *extended real-valued, convex, lower semicontinuous convex functions with* $\gamma_i(t) < +\infty$ *for at least one* $t \in R^{r_i}$.

We are primarily interested in the case where the presence of the functions γ_i introduces difficulties in the numerical solution of problem (1), in the sense that, if the functions γ_i were replaced by some real-valued and continuously differentiable functions $\tilde{\gamma}_i$, then problem (1) could be solved in a relatively easy manner. For example, the functions γ_i may induce constraints, nondifferentiabilities, or ill-conditioning in problem (1). Here are some examples. In the first five examples, γ_i is a function defined on the real line R ($r_i = 1$).

Example 1.1. *Equality Constraints*:

$$\gamma_i(t) = \begin{cases} 0, & \text{if } t=0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2}$$

Here, the presence of $\gamma_i[g_i(x)]$ in problem (1) is equivalent to an additional equality constraint

$$g_i(x) = 0.$$

Example 1.2. *One-Sided Inequality Constraints*:

$$\gamma_i(t) = \begin{cases} 0, & \text{if } t \le 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3}$$

Here, the presence of $\gamma_i[g_i(x)]$ induces the constraint

$$g_i(x) \le 0.$$

Example 1.3. *Two-Sided Inequality Constraints*:

$$\gamma_i(t) = \begin{cases} 0, & \text{if } \alpha_i \le t \le \beta_i, \\ +\infty, & \text{otherwise,} \end{cases} \tag{4}$$

where α_i, β_i are scalars with $\alpha_i < \beta_i$.

Example 1.4. *Polyhedral Functions*:

$$\gamma_i(t) = \max[0, t], \tag{5}$$

$$\gamma_i(t) = |t|, \tag{6}$$

$$\gamma_i(t) = \begin{cases} |t|, & \text{if } |t| \le \alpha, \\ +\infty, & \text{otherwise,} \end{cases} \tag{7}$$

$$\gamma_i(t) = \begin{cases} \max_{j=1,\ldots,r} \{\gamma_j t + \delta_j\}, & \text{if } \alpha \le t \le \beta, \\ +\infty, & \text{otherwise,} \end{cases} \tag{8}$$

where α, β, γ_j, δ_j are given scalars.

Example 1.5. *Ill-Conditioning Terms*:

$$\gamma_i(t) = \tfrac{1}{2} s t^2, \tag{9}$$

$$\gamma_i(t) = \alpha \exp(\beta t) \tag{10}$$

where s, α, β are given scalars with $s > 0$, $\alpha > 0$. The term (9) may induce ill-conditioning in problem (1) if s is very large, while the term (10) may induce ill-conditioning in problem (1) if β is very large. More generally, if the second derivatives or third derivatives of γ_i are very large, relative to other terms in the cost functional, the numerical solution of problem (1) may run into serious difficulties.

Example 1.6. *Minimax Problems*: For $t = (t_1, t_2, \ldots, t_{r_i}) \in R^{r_i}$, consider

$$\gamma_i(t) = \max\{t_1, t_2, \ldots, t_{r_i}\}, \tag{11}$$

$$\gamma_i(t) = \max\{|t_1|, |t_2|, \ldots, |t_{r_i}|\}, \tag{12}$$

$$\gamma_i(t) = \max_{\|z-\alpha\|^2 \le 1} t'z, \tag{13}$$

where $\|\cdot\|$ is the usual Euclidean norm in R^{r_i} and α is a given vector in R^{r_i}.

This paper presents an approach for solving numerically problems of the type described above. The approach consists of approximation of problem (1) by a sequence of optimization problems which involve relatively well-behaved objective functions. The approximation is effected by introducing additional variables and constraints in problem (1), thus forming an equivalent constrained minimization problem. This problem is subsequently handled by a suitable method of multipliers (see Refs. 1–3 for analysis and references on multiplier methods). For the case of Examples 1.1 and 1.2, our approach turns out to be identical to standard multiplier methods. However, for other cases, our approach results in computational and storage savings over standard multiplier methods due to the fact that the number of scalar multipliers utilized is significantly reduced. For many cases of interest, our algorithms may be viewed in effect as multiplier methods with partial elimination of constraints. Aside from the case of convex programming problems, there is no convergence analysis available for multiplier methods of this type. For this reason, we provide in the last section of this paper a convergence and rate-of-convergence result for such methods. This result can be used in turn to provide corresponding convergence results for specific algorithms. Throughout the paper, we have restricted ourselves exclusively to first-order multiplier methods (i.e., steepest-ascent methods for solving a related dual problem in the sense described in Ref. 1). Second-order methods of the Newton type (see Section 6 and Proposition 6 of Ref. 2 for related description and analysis) could also be used.

2. Approximation Procedure

It is clear that problem (1) is equivalent to the following problem:

$$\begin{aligned} &\text{minimize } f(x)+\sum_{i=1}^{m}\gamma_i[g_i(x)-u_i],\\ &\text{subject to } x\in X, \qquad u_i=0, \qquad i=1,\ldots,m, \end{aligned} \tag{14}$$

where we have introduced the additional vectors

$$u_i\in R^{r_i}, \qquad i=1,\ldots,m.$$

A method of multipliers for the problem above is based on sequential minimization over x, $u_1,\ldots,u_m$ of the form

$$\begin{aligned} &\text{minimize } f(x)+\sum_{i=1}^{m}\{\gamma_i[g_i(x)-u_i]+y_k^{i\prime}u_i+\tfrac{1}{2}c_k\|u_i\|^2\},\\ &\text{subject to } x\in X, \end{aligned} \tag{15}$$

where $\|\cdot\|$ denotes the usual Euclidean norm on R^{r_i}, y_k^i are multiplier vectors in R^{r_i}, c_k is a positive scalar penalty parameter, and prime denotes transposition. Equivalently, problem (15) is written as

$$\begin{aligned} &\text{minimize } f(x)+\sum_{i=1}^{m}p_{c_k}^i[g_i(x),y_k^i],\\ &\text{subject to } x\in X, \end{aligned} \tag{16}$$

where

$$p_{c_k}^i[g_i(x),y_k^i]=\min_{u_i}\{\gamma_i[g_i(x)-u_i]+y_k^{i\prime}u_i+\tfrac{1}{2}c_k\|u_i\|^2\}. \tag{17}$$

The initial multiplier vectors y_0^i, $i = 1, \ldots, m$, are arbitrary; and, after each minimization (16), the multiplier vectors y_k^i are updated by means of

$$y_{k+1}^i = y_k^i + c_k u_i^k, \qquad i = 1, \ldots, m, \tag{18}$$

where u_i^k, $i = 1, \ldots, m$, solve (15), together with some vector x_k. Alternate methods could be obtained by using a nonquadratic penalty function in (15); in fact, in some cases, the use of such nonquadratic penalty functions is essential. In order to keep the exposition simple, we will restrict ourselves for the moment to quadratic penalty functions.

It is important to note that the function $p_{c_k}^i$ of (17) is *both real-valued and continuously differentiable* in x, provided the function g_i is continuously differentiable; hence, problem (16) can be solved by the powerful methods available for differentiable functions whenever f, g_i are differentiable. These properties of the function $p_{c_k}^i$ can be inferred from the following lemma.

Lemma 2.1. Let $\gamma: R^r \to (-\infty, +\infty]$ be a lower semicontinuous, convex function, and assume that $\gamma(t) < +\infty$ for at least one vector $t \in R^r$. Let also λ be any vector in R^r and $c > 0$ be a scalar. Then, the function $p_c(\cdot, \lambda)$ defined by

$$p_c(t, \lambda) = \inf_u \{\gamma(t-u) + \lambda' u + \tfrac{1}{2}c\|u\|^2\} \tag{19}$$

is real-valued, convex, and continuously differentiable in t. Furthermore, the infimum with respect to u in (19) is attained at a unique point for every $t \in R^r$.

Proof. The function $p_c(\cdot, t)$ is the infimal convolution (Ref. 4) of the convex function γ and the quadratic convex function $h: R^r \to R$ defined by

$$h(u) = \lambda' u + \tfrac{1}{2}c\|u\|^2. \tag{20}$$

Since

$$h(u) \to \infty \qquad \text{as } \|u\| \to \infty,$$

it follows from Corollary 9.2.2 of Ref. 4 that $p_c(\cdot, \lambda)$ is convex and the infimum in (19) is attained for each t by some u. Since h is strictly convex and real-valued, it follows that $p_c(\cdot, \lambda)$ is also real-valued and the minimum is attained at a single point. Also, h is a smooth function (Ref. 4, p. 251); and, from Corollary 26.3.2 of Ref. 4, it follows that $p_c(\cdot, \lambda)$ is an essentially smooth convex function. Since it is also real-valued, it is continuously differentiable.

The interpretation of $p_c(\cdot, \lambda)$ in the proof above as the infimal convolution of γ and h defined by (20) is useful in visualizing the form of $p_c(\cdot, \lambda)$.

The epigraph of $p_c(\cdot, \lambda)$ is obtained as the vector sum of the epigraph of the functions $\gamma(\cdot)$ and $h(\cdot)$ (see Ref. 4, Theorem 5.4).

In some cases, it is useful to work with a dual expression for the function $p_c(\cdot, \lambda)$ of (19). This expression is given in the following lemma, the proof of which follows by straightforward application of Theorem 31.2 of Ref. 4 (Fenchel's duality theorem).

Lemma 2.2. The function $p_c(\cdot, \lambda)$ of (19) is also given by

$$p_c(t, \lambda) = \sup_{u^*}\{t'u^* - \gamma^*(u^*) - \tfrac{1}{2}c\|u^* - \lambda\|^2\}, \tag{21}$$

where

$$\gamma^*(u^*) = \sup_{u}\{u'u^* - \gamma(u)\}$$

is the convex conjugate function of γ. Furthermore, the supremum in (21) is attained at a unique point $u^*(t, \lambda, c)$, and we have

$$u^*(t, \lambda, c) = \lambda + cu(t, \lambda, c) = \nabla_t p_c(t, \lambda), \tag{22}$$

where $u(t, \lambda, c)$ is the unique point attaining the infimum in (19) and $\nabla_t p_c$ is the gradient of p_c with respect to t.

The correspondence between Eqs. (22) and (18) is often convenient in the analysis of specific cases.

It is to be noted that, even though we employed the additional vectors $u_1, \ldots, u_m$ in order to introduce the algorithm, the numerical computation itself need not involve these vectors, since, in the cases of interest to us, the functions $p^i_{c_k}$ of (16)–(17) can be obtained in explicit form. Furthermore, the minimizing vectors u^k_i of (18) can be expressed directly in terms of minimizing vectors x_k in problem (16), since u^k_i is uniquely defined in terms of x_k, c_k, y^i_k as the minimizing vector in (17). We provide the corresponding analysis for the examples given in the previous section.

Example 2.1. For the case where $\gamma_i(t)$ is given by (2), we obtain from (17)–(18):

$$p^i_{c_k}[g_i(x), y^i_k] = y^i_k g_i(x) + \tfrac{1}{2}c_k[g_i(x)]^2,$$
$$y^i_{k+1} = y^i_k + c_k g_i(x_k), \qquad i = 1, \ldots, m.$$

In this case, the iteration reduces to the ordinary multiplier iteration with quadratic penalty function as proposed for equality constraints by Hestenes (Ref. 5).

Example 2.2. For $\gamma_i(t)$ given by (3), we obtain from (17)–(18) by straightforward calculation:

$$p^i_{c_k}[g_i(x), y^i_k] = (1/2c_k)\{(\max[0, y^i_k + c_k g_i(x)])^2 - (y^i_k)^2\},$$
$$y^i_{k+1} = \max[0, y^i_k + c_k g_i(x_k)], \qquad i = 1, \ldots, m.$$

The algorithm reduces to the multiplier method proposed for inequality constraints by Rockafellar (Ref. 6).

Example 2.3. For the case of two-sided inequality constraints, $\gamma_i(t)$ given by (4), we obtain

$$p^i_{c_k}[g_i(x), y^i_k] = \begin{cases} y^i_k[g_i(x) - \beta_i] + \frac{1}{2}c_k[g_i(x) - \beta_i]^2, & \text{if } \beta_i - y^i_k/c_k \le g_i(x), \\ y^i_k[g_i(x) - \alpha_i] + \frac{1}{2}c_k[g_i(x) - \alpha_i]^2, & \text{if } g_i(x) \le \alpha_i - y^i_k/c_k, \\ -(y^i_k)^2/2c_k, & \text{otherwise;} \end{cases}$$

$$y^i_{k+1} = \begin{cases} y^i_k + c_k[g_i(x_k) - \beta_i], & \text{if } \beta_i - y^i_k/c_k \le g_i(x_k), \\ y^i_k + c_k[g_i(x_k) - \alpha_i], & \text{if } g_i(x_k) \le \alpha_i - y^i_k/c_k \\ 0, & \text{otherwise.} \end{cases}$$

Notice that this algorithm involves a single multiplier per two-sided constraint as compared with two multipliers per constraint required by a method of multipliers which would separate the constraint

$$\alpha_i \le t \le \beta_i$$

into two constraints

$$\alpha_i \le t \quad \text{and } t \le \beta_i.$$

On the other hand, we note that it is possible to convert the two-sided constraint

$$\alpha_i \le t \le \beta_i$$

into the one-sided constraint

$$(t - \alpha_i)(t - \beta_i) \le 0$$

and subsequently employ the method of multipliers of Example 2.2, thus also utilizing a single multiplier per constraint. This possibility was suggested by an anonymous referee. A possible objection to this approach is that a constraint of the form

$$[g_i(x) - \alpha_i][g_i(x) - \beta_i] \le 0$$

involves a more complex nonlinearity than the equivalent constraint

$$\alpha_i \leq g_i(x) \leq \beta_i$$

(for example, it is nonlinear even if the function g_i is linear). However, the relative merit (if any) of the method proposed here versus the approach of replacing the constraint

$$\alpha_i \leq g_i(x) \leq \beta_i$$

by the constraint

$$[g_i(x) - \alpha_i][g_i(x) - \beta_i] \leq 0$$

can only be determined by extensive numerical experimentation on a variety of test problems, which has not been undertaken.

Example 2.4. Consider the case where $\gamma_i(t) = \max[0, t]$. Then, by straightforward calculation, we obtain

$$p^i_{c_k}[g_i(x), y^i_k] = \begin{cases} g_i(x) - (1 - y^i_k)^2/2c_k, & \text{if } g_i(x) \geq (1 - y^i_k)/c_k, \\ -(y^i_k)^2/2c_k, & \text{if } g_i(x) \leq -y^i_k/c_k, \\ y^i_k g_i(x) + \frac{1}{2}c_k[g_i(x)]^2, & \text{if } -y^i_k/c_k \leq g_i(x) \leq (1 - y^i_k)/c_k; \end{cases}$$

$$y^i_{k+1} = \begin{cases} 1, & \text{if } g_i(x_k) \geq (1 - y^i_k)/c_k, \\ 0, & \text{if } g_i(x_k) \leq -y^i_k/c_k, \\ y^i_k + c_k g_i(x_k), & \text{if } -y^i_k/c_k \leq g_i(x_k) \leq (1 - y^i_k)/c_k. \end{cases}$$

The corresponding algorithm is a special case of an approximation method for nondifferentiable optimization given in Ref. 7. Notice that a single multiplier per term $\gamma_i[g_i(x)]$ is utilized. If one were to convert the problem to a nonlinear programming problem of the form

$$\text{minimize } f(x) + \sum_{i=1}^{m} z_i,$$

$$\text{subject to } g_i(x) \leq z_i, \qquad 0 \leq z_i, \qquad i = 1, \ldots, m,$$

where z_i are additional variables, then two multipliers per term $\gamma_i[g_i(x)]$ would be required in order for the problem to be solved by the method of multipliers.

The case where $\gamma_i(t)$ is given by (6) can be converted to the earlier case by writing

$$|t| = -t + \max[0, 2t].$$

Let $\gamma_i(t)$ be given by (7), where α is some positive number. Such terms appear, for example, in the cost functional of minimum-fuel problems. We have by straightforward calculation

$$p^i_{c_k}[g_i(x), y^i_k] = \begin{cases} \alpha + y^i_k[g_i(x)+\alpha] + \frac{1}{2}c_k[g_i(x)+\alpha]^2, & \text{if } g_i(x) \le -\alpha-(1+y^i_k)/c_k, \\ -g_i(x)-(1+y^i_k)^2/2c_k, & \text{if } -\alpha-(1+y^i_k)/c_k \le g_i(x) \le -(1+y^i_k)/c_k, \\ y^i_k g_i(x) + \frac{1}{2}c_k[g_i(x)]^2, & \text{if } -(1+y^i_k)/c_k \le g_i(x) \le (1-y^i_k)/c_k, \\ g_i(x)-(1-y^i_k)^2/2c_k, & \text{if } (1-y^i_k)/c_k \le g_i(x) \le \alpha+(1-y^i_k)/c_k \\ \alpha + y^i_k[g_i(x)-\alpha] + \frac{1}{2}c_k[g_i(x)-\alpha]^2, & \text{if } \alpha+(1-y^i_k)/c_k \le g_i(x); \end{cases}$$

and iteration (18) takes the form

$$y^i_{k+1} = \begin{cases} y^i_k + c_k[g_i(x_k)+\alpha], & \text{if } g_i(x_k) \le -\alpha-(1+y^i_k)/c_k, \\ -1, & \text{if } -\alpha-(1+y^i_k)/c_k \le g_i(x_k) \le -(1+y^i_k)/c_k, \\ y^i_k + c_k g_i(x), & \text{if } -(1+y^i_k)/c_k \le g_i(x_k) \le (1-y^i_k)/c_k, \\ 1, & \text{if } (1-y^i_k)/c_k \le g_i(x_k) \le \alpha+(1-y^i_k)/c_k, \\ y^i_k + c_k[g_i(x_k)-\alpha], & \text{if } \alpha+(1-y^i_k)/c_k \le g_i(x_k). \end{cases}$$

Notice that a single multiplier per term γ_i is utilized in place of four multipliers per term γ_i for the ordinary method of multipliers.

Similarly, one may obtain the function $p^i_{c_k}$ and the iteration (18) in explicit form for the function $\gamma_i(t)$ given by (8). Again, only one multiplier per term is required in place of r multipliers for the ordinary multiplier method.

Example 2.5. Let $\gamma_i(t)$ be given by (9). Then, we have by straightforward calculation:

$$p^i_{c_k}[g_i(x), y^i_k] = [s_i/(s_i+c_k)]\{\tfrac{1}{2}c_k[g_i(x)]^2 + y^i_k g_i(x) - (y^i_k)^2/2s_i\},$$

and the iteration (18) takes the form

$$y^i_{k+1} = y^i_k + c_k[s_i g_i(x_k) - y^i_k]/(s_i + c_k).$$

Notice that the second derivative of $p^i_{c_k}(\cdot\,, y^i_k)$ given above is $s_i c_k/(s_i + c_k)$ and can be made arbitrarily small by choosing c_k sufficiently small.

The case where $\gamma_i(t)$ is given by (10) requires a slightly different approximation method and a nonquadratic penalty function. It will be examined after we provide algorithms for the minimax cases (see next example).

Example 2.6. Let $\gamma_i(t)$ be given by (11). From Eq. (21), we have

$$p_c(t, \lambda) = \sup_{u^*}\{t'u^* - \gamma^*(u^*) - \tfrac{1}{2}c\|u^* - \lambda\|^2\},$$

where the convex conjugate function of γ can be easily calculated as

$$\gamma^*(u^*) = \begin{cases} 0, & \text{if } \sum_{i=1}^r u_i^* = 1,\ u_i^* \geq 0,\ i = 1, \ldots, r, \\ +\infty, & \text{otherwise.} \end{cases}$$

Hence,

$$p_c(t, \lambda) = \max_{\substack{\sum_{i=1}^r u_i^* = 1 \\ u_i^* \geq 0 \\ i = 1, \ldots, r}} \{t'u^* - (1/2c)\|u^* - \lambda\|^2\}. \tag{23}$$

By introducing a Lagrange multiplier μ corresponding to

$$\sum_{i=1}^r u_i^* = 1$$

and carrying out the straightforward optimization in (23), we obtain

$$p_c(t, \lambda) = (\tfrac{1}{2}c) \sum_{i=1}^r \{(\max\{0, \lambda_i + c[t_i - \mu(t, \lambda, c)]\})^2 - \lambda_i^2\} + \mu(t, \lambda, c).$$

The maximizing vector $\bar{u}^*$ in (23) has coordinates given by

$$\bar{u}_i^* = \max\{0, \lambda_i + c[t_i - \mu(t, \lambda, c)]\}, \qquad i = 1, \ldots, r, \tag{24}$$

and the Lagrange multiplier $\mu(t, \lambda, c)$ corresponding to t, λ, c is determined from

$$\sum_{i=1}^{r} \max\{0, \lambda_i + c[t_i - \mu(t, \lambda, c)]\} = 1.$$

In terms of problem (1), a term of the form

$$\gamma[g(x)] = \max\{g^1(x), g^2(x), \ldots, g^r(x)\}$$

is approximated by

$$p_{c_k}[g(x), y_k] = (\tfrac{1}{2}c_k) \sum_{i=1}^{r} \{(\max\{0, y_k^i + c_k[g_i(x) - \mu[g(x), y_k, c_k]]\})^2 - (y_k^i)^2\} + \mu[g(x), y_k, c_k].$$

The gradient with respect to x of the expression above is obtained from (22) and (24):

$$\nabla p_{c_k}[g(x), y_k] = \sum_{i=1}^{r} \nabla g_i(x) \max\{0, y_k^i + c_k[g_i(x) - \mu[g(x), y_k, c_k]]\}.$$

The scalar $\mu[g(x), y_k, c_k]$ is determined from

$$\textstyle\sum_{i=1}^{r} \max\{0, y_k^i + c_k[g_i(x) - \mu[g(x), y_k, c_k]]\} = 1.$$

It is easy to see that the computer can determine the value of $\mu[g(x), y_k, c_k]$ from the relation above with very little effort.

Regarding the multiplier iteration, we have [see (18), (22), (24)]

$$y_{k+1}^i = \max\{0, y_k^i + c_k[g_i(x_k) - \mu[g(x_k), y_k, c_k]]\}, \qquad i = 1, \ldots, r.$$

The algorithm described above was tested on the following problem.

Problem 2.1. This is a minimax problem, described by

$$\min_x \max\{g_1(x), g_2(x), g_3(x), g_4(x)\},$$

where

$$g_1(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 + 21x_3 + 7x_4 + 44,$$
$$g_2(x) = 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5,$$
$$g_3(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8,$$
$$g_4(x) = x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10.$$

This problem was obtained by alteration of the well-known Rosen–Suzuki problem. The optimal value is zero, the optimal solution is

$$x^* = (0, 1, 2, -1),$$

and the final multiplier vector is

$$y^* = (0.25, 0.50, 0.25, 0).$$

We solved the problem with the method described above by using for unconstrained minimization the Fletcher–Powell method available on the IBM-360 as the FMFP scientific subroutine. The parameter EPS of the subroutine, which controls minimization accuracy, was set at 10^{-5}; and the parameter EST was set at -1. The starting multiplier vector was

$$y_0 = (0, 0, 0, 0);$$

and the starting vector x for the first minimization was

$$x = (0, 0, 0, 0).$$

The starting vector x in each subsequent minimization was the final vector in the previous minimization. The penalty parameter sequence was

$$c_k = 4^k.$$

The optimal solution and multiplier vector and the optimal value function were all obtained within five significant digits of accuracy in a total of 47 iterations of the Fletcher–Powell method and a total of five unconstrained minimization cycles.

For the case where

$$\gamma[g(x)] = \max\{|g_1(x)|, |g_2(x)|, \ldots, |g_r(x)|\},$$

a very similar calculation as the one for the previous case yields the following:

$$p_{c_k}[g(x), y_k] = \sum_{i=1}^{r} \tilde{p}^i_{c_k}[g(x), y_k] + \mu[g(x), y_k, c_k],$$

where

$$\tilde{p}^i_{c_k}[g(x), y_k] = \begin{cases} y^i_k[g_i(x) - \mu[g(x), y_k, c_k]] + \frac{1}{2}c_k[g_i(x) - \mu[g(x), y_k, c_k]]^2, \\ \qquad\qquad \text{if } y^i_k + c_k[g_i(x) - \mu[g(x), y_k, c_k]] \geq 0, \\ y^i_k[g_i(x) + \mu[g(x), y_k, c_k]] + \frac{1}{2}c_k[g_i(x) + \mu[g(x), y_k, c_k]]^2, \\ \qquad\qquad \text{if } y^i_k + c_k[g_i(x) + \mu[g(x), y_k, c_k]] \leq 0, \\ -(y^i_k)^2/2c_k, \qquad \text{otherwise.} \end{cases}$$

The gradient of $p_{c_k}[g(x), y_k]$with respect to x is given by

$$\nabla p_{c_k}[g(x), y_k] = \sum_{i=1}^{r} \nabla g_i(x)\bar{u}_i^*(x, y_k, c_k),$$

where

$$\bar{u}_i^*(x, y_k, c_k) = \begin{cases} y_k^i + c_k[g_i(x) - \mu[g(x), y_k, c_k]], & \text{if } y_k^i + c_k[g_i(x) \\ & \quad - \mu[g(x), y_k, c_k]] \geq 0, \\ y_k^i + c_k[g_i(x) + \mu[g(x), y_k, c_k]], & \text{if } y_k^i + c_k[g_i(x) \\ & \quad + \mu[g(x), y_k, c_k]] \leq 0, \\ 0, & \text{otherwise.} \end{cases}$$

The scalar $\mu[g(x), y_k, c_k]$ is determined from

$$\mu[g(x), y_k, c_k] = 0, \qquad \text{if } \sum_{i=1}^{r} \max\{0, |y_k^i + c_k g_i(x)|\} \leq 1,$$

$$\sum_{i=1}^{r} \max\{0, |y_k^i + c_k g_i(x)| - c_k\mu[g(x), y_k, c_k]\} = 1, \qquad \text{otherwise.}$$

The multiplier iteration is given by

$$y_{k+1}^i = \bar{u}_i^*(x_k, y_k, c_k), \qquad i = 1, \ldots, r,$$

where $\bar{u}_i^*$ is defined above.

The case where $\gamma_i(t)$ is given by (13) and other related cases where γ is the support function of a relatively simple set can be handled in a similar manner.

A Variation of the Approximation Procedure. A slightly different type of algorithm may be obtained when the functions γ_i in problem (1) are *monotonically nondecreasing*. Then, we may consider the following problem which is equivalent to problem (1):

$$\text{minimize } f(x) + \sum_{i=1}^{m} \gamma_i[g_i(x) - u_i],$$

$$\text{subject to } x \in X, \qquad u_i \leq 0, \qquad i = 1, \ldots, m,$$

and we may utilize a multiplier method for solving the problem above. An example is provided for the case where $\gamma_i(t)$ is given by (10), with $\alpha > 0$, $\beta > 0$. For this case, it is convenient to use a multiplier method with exponential penalty function proposed in 1972 by Kort and the author (Ref.

8). The penalty for this method is $(1/c)y\,[\exp(cu)-1]$, and we have

$$p^i_{c_k}[g_i(x), y^i_k] = \min_u\{\alpha\exp[\beta(y_i(x)-u)] + (1/c_k)y^i_k[\exp(c_k u)-1]\}$$
$$= (\alpha\beta)^{c_k/(c_k+\beta)}(y^i_k)^{\beta/(c_k+\beta)}$$
$$\times[(c_k+\beta)/c_k\beta]\exp\{[c_k\beta/(c_k+\beta)]g_i(x)\}.$$

The initial multiplier vector y_0 must have positive coordinates. The multiplier iteration is

$$y^i_{k+1} = y^i_k\exp(c_k u_k),$$

which yields

$$y^i_{k+1} = (y^i_k)^{\beta/(c_k+\beta)}(\alpha\beta)^{c_k/c_k+\beta)}\exp\{[c_k\beta/(c_k+\beta)]g_i(x_k)\}.$$

An application of this method, together with some computational results, is given elsewhere (Ref. 9), where an algorithm is described for solving electric network problems involving physical or ideal diodes. The voltage–current characteristics of physical diodes are typically described by sharply rising exponential functions, which cause formidable difficulties during the computational solution of the related network problem. By using the multiplier method with exponential penalty function described above, all computational difficulties due to steep diode characteristics are bypassed.

We also note that the same exponential penalty function when used in connection with the function

$$\gamma[g(x)] = \max\{g^1(x), g^2(x), \dots, g^r(x)\}$$

yields the approximating function

$$p_{c_k}[g(x), y_k] = (1/c_k)\log\left\{\sum_{i=1}^r y^i_k\exp[c_k g^i(x)]\right\}$$

and the multiplier iteration

$$y^i_{k+1} = \frac{y^i_k\exp[c_k g^i(x_k)]}{\sum_{j=1}^r y^j_k\exp[c_k g^j(x_k)]}.$$

We close this section by mentioning that we expect that the main idea of the approximation procedure provided here should find application in problems involving ill-conditioning terms other than the ones considered here. One such application has been described recently in a paper by Gabay and Mercier (Ref. 10). The approximation method could also prove useful in generalized versions of problem (1), such as, for example, problems of the form

$$\text{minimize } f(x) + h[x, \gamma_1[g_1(x)], \dots, \gamma_m[g_m(x)]],$$
$$\text{subject to } x \in X,$$

where γ_i, $i = 1, \ldots, m$, are real-valued and

$$h: R^{m+n} \to R$$

is a continuously differentiable function. The terms γ_i could be approximated in the same way as earlier, and the same multiplier iterations could be used. One such algorithm has been described in Ref. 7, and favorable computational results have been reported. Other related algorithms are described in Ref. 11, together with encouraging computational experience. A local duality theory as well as analysis relating to the behavior of approximation algorithms for the problem above where the function h is nonlinear can be found in Ref. 12.

3. Convergence Analysis: Multiplier Methods with Partial Elimination of Constraints

From the discussion given above, it should be evident that some of the algorithms of the type that we have discussed are multiplier methods where only some of the constraints (i.e., the constraints $u_i = 0$) are eliminated by means of a penalty function. For example, in the problem involving two-sided inequality constraints:

$$\text{minimize } f(x),$$
$$\text{subject to } x \in X, \quad \alpha_i \le g_i(x) - u_i \le \beta_i, \quad u_i = 0, \quad i = 1, \ldots, m,$$

only the constraints $u_i = 0$ are eliminated by means of a penalty function. The approximate minimization problem takes the form

$$\text{minimize } f(x) + \sum_{i=1}^{m} (y_k^i u_i + \tfrac{1}{2} c_k u_i^2),$$
$$\text{subject to } x \in X, \quad \alpha_i \le g_i(x) - u_i \le \beta_i, \quad i = 1, \ldots, m.$$

This problem may also be converted to the following problem involving equality constraints by introducing additional variables z_i, w_i, $i = 1, \ldots, m$,

$$\text{minimize } f(x) + \sum_{i=1}^{m} (y_k^i u_i + \tfrac{1}{2} c_k u_i^2),$$
$$\text{subject to } x \in X, \quad \alpha_i + z_i^2 = g_i(x) - u_i, \quad g_i(x) - u_i + w_i^2 = \beta_i, \quad i = 1, \ldots, m.$$

Under second-order sufficiency assumptions that we will introduce shortly, the equality constrained problem and the inequality constrained problem above are equivalent, in the sense that there is a one-to-one correspondence between their Kuhn–Tucker pairs. Thus, even though we shall restrict ourselves to the case of equality constraints, the analysis applies to problems with inequality constraints as well. This analytically convenient device has been utilized in the past in connection with ordinary multiplier methods (Ref. 2).

In what follows in this section, we consider an optimization problem involving equality constraints and a multiplier algorithm for its solution involving partial elimination of constraints. A convergence and rate-of-convergence result for this algorithm will be given. This result yields in a straightforward manner convergence and rate-of-convergence results for algorithms such as those considered in the previous section.

Consider the problem

$$\begin{aligned}&\text{minimize } q(x),\\ &\text{subject to } h_i(x)=0, \quad l_j(x)=0, \quad i=1,\dots,m, \quad j=1,\dots,r.\end{aligned} \tag{25}$$

Let $\bar{x}$ be a local minimum for this problem. We assume that the functions

$$q\colon R^n \to R, \qquad h_i\colon R^n \to R, \qquad l_j\colon R^n \to R$$

have Hessian matrices

$$\nabla^2 q(x), \qquad \nabla^2 h_i(x), \qquad \nabla^2 l_j(x)$$

which are Lipschitz continuous in a neighborhood of $\bar{x}$. Furthermore, we assume that the gradients

$$\nabla h_i(\bar{x}), \qquad \nabla l_j(\bar{x}), \qquad i=1,\dots,m, \qquad j=1,\dots,r,$$

are linearly independent. As a consequence, we obtain that there exist unique Lagrange multipliers

$$\bar{\lambda}=(\bar{\lambda}^1,\dots,\bar{\lambda}^m)', \qquad \bar{\mu}=(\bar{\mu}^1,\dots,\bar{\mu}^r),$$

such that

$$\nabla q(\bar{x})+\sum_{i=1}^m \bar{\lambda}^i \nabla h_i(\bar{x})+\sum_{j=1}^r \bar{\mu}^j \nabla l_j(\bar{x})=0. \tag{26}$$

Consider now the following multiplier iteration for solving problem (25):

$$\lambda^i_{k+1}=\lambda^i_k+c_k h_i(x_k), \qquad i=1,\dots,m, \tag{27}$$

where x_k solves, within a neighborhood of $\bar{x}$, the problem

$$\begin{aligned}&\text{minimize } q(x)+\sum_{i=1}^m \{\lambda^i_k h_i(x)+\tfrac{1}{2}c_k[h_i(x)]^2\},\\ &\text{subject to } l_j(x)=0, \qquad j=1,\dots,r.\end{aligned} \tag{28}$$

In the above relations, $\{c_k\}$ is as earlier a sequence of positive penalty parameters, and

$$\lambda_0 = (\lambda_0^1, \ldots, \lambda_0^m)'$$

is a given initial multiplier vector.

In what follows in this section, we shall investigate the convergence properties of the multiplier method specified by (27)–(28). This method may be cast into the usual form of the method of multipliers where all constraints are eliminated by means of a penalty function by using the constraints

$$l_j(x) = 0$$

to reduce the number of the variables of the problem in a manner which is familiar from constructions used in the reduced gradient method (see, e.g., Ref. 13).

Indeed, consider the $n \times r$ matrix

$$\nabla l(\bar{x}) = [\nabla l_1(\bar{x}), \nabla l_2(\bar{x}), \ldots, \nabla l_r(\bar{x})]$$

having as columns the gradients

$$\nabla l_j(\bar{x}), \qquad j = 1, \ldots, r.$$

Since these gradients are linearly independent by assumption, one may find r rows of $\nabla l(\bar{x})$ which are linearly independent. Assume, without loss of generality, that the first r rows of $\nabla l(\bar{x})$ are linearly independent and partition the vector x into

$$x = \begin{bmatrix} x_B \\ x_R \end{bmatrix},$$

where

$$x_B \in R^r, \qquad x_R \in R^{n-r}.$$

Then, the system of equations

$$l_j(x) = l_j(x_B, x_R) = 0, \qquad j = 1, \ldots, r, \tag{29}$$

may be solved for x_B in terms of x_R in a neighborhood of

$$\bar{x} = (\bar{x}_B', \bar{x}_R')'$$

by using the implicit function theorem and the linear independence of the first r rows of $\nabla l(\bar{x})$ (i.e., the invertibility of the Jacobian matrix corresponding to the coordinates of x_B). More specifically, there exists an $\epsilon > 0$ and a function

$$\varphi: S(\bar{x}_R; \epsilon) \to S(\bar{x}_B; \epsilon),$$

where

$$S(\bar{x}_R;\epsilon), \qquad S(\bar{x}_B;\epsilon)$$

denote the open spheres of radius ϵ centered at $\bar{x}_R$ and $\bar{x}_B$, respectively, with the following property:

$$l_i[\varphi(x_R), x_R]=0, \qquad \forall_i=1,\ldots,r, \qquad x_R\in S(\bar{x}_R;\epsilon).$$

Furthermore, $(\varphi(x_R)', x_R')'$ is the unique solution of the system of equations (29) within

$$S(\bar{x}_B;\epsilon)\times S(\bar{x}_R;\epsilon).$$

As a result,

$$\bar{x}_B=\varphi(\bar{x}_R).$$

In addition, all first and second derivatives of φ exist and are Lipschitz continuous within $S(\bar{x}_R;\epsilon)$.

Now, in view of the above construction, the algorithm specified by (27)–(28) is, within a neighborhood of

$$\bar{x}=(\varphi(\bar{x}_R)', \bar{x}_R')',$$

equivalent to the algorithm

$$\lambda_{k+1}^i=\lambda_k^i+c_k\tilde{h}_i(z_k), \qquad i=1,\ldots,m, \tag{30}$$

where

$$z_k\in R^{n-r}$$

solves, within a neighborhood of

$$\bar{z}=\bar{x}_R,$$

the problem

$$\text{minimize } \tilde{q}(z)+\sum_{i=1}^m \{\lambda_k^i\tilde{h}_i(z)+\tfrac{1}{2}c_k[\tilde{h}_i(z)]^2\}, \tag{31}$$

and $\tilde{q}$ and $\tilde{h}_i$ are real-valued functions defined on $S(\bar{x}_R;\epsilon)$ by

$$\tilde{q}(z)=q[\varphi(z), z], \tag{32}$$

$$\tilde{h}_i(z)=h_i[\varphi(z), z], \qquad i=1,\ldots,m. \tag{33}$$

Notice that the algorithm specified by (30)–(31) is the ordinary method of multipliers for the problem

$$\begin{aligned}&\text{minimize } \tilde{q}(z),\\ &\text{subject to } \tilde{h}_i(z)=0, \qquad i=1,\ldots,m,\end{aligned} \tag{34}$$

which has

$$\bar{z}=\bar{x}_R$$

as a local minimum.

Now, based on the above construction, and utilizing a known result for the method of multipliers with full elimination of constraints as applied to problem (34), we can prove the following proposition which constitutes both a convergence and a rate-of-convergence result for the algorithm specified by (27)–(28). The assumptions of the proposition constitute second-order sufficiency conditions for problem (25).

Proposition 3.1. In addition to linear independence of the gradients

$$\nabla h_i(\bar{x}), \quad \nabla l_j(\bar{x}), \quad i = 1, \ldots, m, \quad j = 1, \ldots, r,$$

and Lipschitz continuity of the Hessian matrices

$$\nabla^2 q(x), \quad \nabla^2 h_i(x), \quad \nabla^2 l_j(x), \quad i = 1, \ldots, m, \quad j = 1, \ldots, r,$$

within a neighborhood of $\bar{x}$, assume that

$$w'\{\nabla^2 q(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}^i \nabla^2 h_i(\bar{x}) + \sum_{j=1}^{r} \bar{\mu}^j \nabla^2 l_j(\bar{x})\} w > 0, \tag{35}$$

for all $w \in R^n$ such that $w \neq 0$ and

$$w'\nabla h_i(\bar{x}) = 0, \quad w'\nabla l_j(\bar{x}) = 0 \quad \text{for all } i = 1, \ldots, m, \quad j = 1, \ldots, r.$$

Then, for any subset $\Lambda \subset R^m$, there exists a scalar $c^* > 0$ (depending on Λ) such that, for all

$$c \geq c^* \quad \text{and} \quad \lambda = (\lambda^1, \ldots, \lambda^m)' \in \Lambda,$$

the problem

$$\begin{aligned} &\text{minimize } q(x) + \sum_{i=1}^{m} \{\lambda^i h_i(x) + \tfrac{1}{2} c [h_i(x)]^2\}, \\ &\text{subject to } l_j(x) = 0, \quad j = 1, \ldots, r, \end{aligned} \tag{36}$$

has a unique solution within some neighborhood of $\bar{x}$, denoted by $x(\lambda, c)$. Furthermore, there exists a scalar $M > 0$ such that, for all $c \geq c^*$ and $\lambda \in \Lambda$,

$$\|x(\lambda, c) - \bar{x}\| \leq M \|\lambda - \bar{\lambda}\| / c, \tag{37}$$

$$\|\tilde{\lambda}(\lambda, c) - \bar{\lambda}\| \leq M \|\lambda - \bar{\lambda}\| / c, \tag{38}$$

where $\bar{\lambda}$ is the Lagrange multiplier vector for problem (25) [see (26)] and the vector $\tilde{\lambda}(\lambda, c)$ has coordinates given by

$$\tilde{\lambda}^i(\lambda, c) = \lambda^i + c h_i[x(\lambda, c)], \quad i = 1, \ldots, m. \tag{39}$$

Prior to proving Proposition 3.1, we show the following lemma.

Lemma 3.1. Under the assumptions of Proposition 3.1, the gradients

$$\nabla \tilde{h}_i(\bar{x}_R), \qquad i = 1, \ldots, m,$$

of the functions $\tilde{h}_i$ of (33) are linearly independent, and we have

$$\nabla \tilde{q}(\bar{x}_R) + \sum_{i=1}^{m} \bar{\lambda}^i \nabla \tilde{h}_i(\bar{x}_R) = 0, \tag{40}$$

where $\tilde{q}$ is given by (32) [i.e., $\bar{\lambda}$ is a Lagrange multiplier vector for problem (34) as well as for problem (25)]. Furthermore, the Hessian matrices $\nabla^2 \tilde{q}(z)$, $\nabla^2 \tilde{h}_i(z)$ exist and are Lipschitz continuous in a neighborhood of the vector

$$\bar{z} = \bar{x}_R.$$

In addition, we have

$$v'\{\nabla^2 \tilde{q}(\bar{x}_R) + \sum_{i=1}^{m} \bar{\lambda}^i \nabla^2 \tilde{h}_i(\bar{x}_R)\} v > 0, \tag{41}$$

for all $v \in R^{n-r}$ such that $v \neq 0$ and

$$v' \nabla \tilde{h}_i(\bar{x}_R) = 0 \qquad \text{for all } i = 1, \ldots, m.$$

Proof. Consider the matrices

$$\nabla h(x) = [\nabla h_1(x), \nabla h_2(x), \ldots, \nabla h_m(x)],$$

$$\nabla l(x) = [\nabla l_1(x), \nabla l_2(x), \ldots, \nabla l_r(x)],$$

having as columns the gradients $\nabla h_i(x)$, $\nabla l_j(x)$, and partition them as follows

$$\nabla h(x) = \begin{bmatrix} \nabla_B h(x) \\ \nabla_R h(x) \end{bmatrix}, \qquad \nabla l(x) = \begin{bmatrix} \nabla_B l(x) \\ \nabla_R l(x) \end{bmatrix}.$$

In the above relations, $\nabla_B h(x)$, $\nabla_B l(x)$ are $r \times m$ and $r \times r$ matrices, respectively, consisting of the partial derivatives of h_i, l_j with respect to the first r components of x, i.e., the r-tuple of variables x_B. From the implicit function theorem and the definition of the implicit function φ, we have, for all $x_R \in S(\bar{x}_R; \epsilon)$,

$$\nabla \varphi(x_R) = -[\nabla_B l[\varphi(x_R), x_R]']^{-1} \nabla_R l[\varphi(x_R), x_R]'; \tag{42}$$

and, in particular,

$$\nabla \varphi(\bar{x}_R) = -[\nabla_B l(\bar{x}_B, \bar{x}_R)']^{-1} \nabla_R l(\bar{x}_B, \bar{x}_R)', \tag{43}$$

where $\nabla\varphi$ is the $r\times(n-r)$ Jacobian matrix of φ. We also have, from (32)–(33),

$$\nabla\tilde{q}(\bar{x}_R)=\nabla\varphi(\bar{x}_R)'\nabla_B q(\bar{x}_B,\bar{x}_R)+\nabla_R q(\bar{x}_B,\bar{x}_R), \tag{44}$$

$$\nabla\tilde{h}_i(\bar{x}_R)=\nabla\varphi(\bar{x}_R)'\nabla_B h_i(\bar{x}_B,\bar{x}_R)+\nabla_R h_i(\bar{x}_B,\bar{x}_R), \tag{45}$$

where

$$\nabla_B q, \nabla_B h_i \in R^r \qquad \text{and} \qquad \nabla_R q, \nabla_R h_i \in R^{n-r}$$

are the vectors of partial derivatives of q and h_i with respect to the coordinates of x_B and x_R, respectively.

Now, (26) can be written as

$$\nabla_B q(\bar{x})+\nabla_B h(\bar{x})\bar{\lambda}+\nabla_B l(\bar{x})\bar{\mu}=0, \tag{46}$$

$$\nabla_R q(\bar{x})+\nabla_R h(\bar{x})\bar{\lambda}+\nabla_R l(\bar{x})\bar{\mu}=0. \tag{47}$$

From (46), we obtain

$$\bar{\mu}=-[\nabla_B l(\bar{x})]^{-1}[\nabla_B q(\bar{x})+\nabla_B h(\bar{x})\bar{\lambda}]; \tag{48}$$

and substitution in (47) yields

$$\nabla_R q(\bar{x})-\nabla_R l(\bar{x})[\nabla_B l(\bar{x})]^{-1}\nabla_B q(\bar{x})+[\nabla_R h(\bar{x})-\nabla_R l(\bar{x})[\nabla_B l(\bar{x})]^{-1}\nabla_B h(\bar{x})]\bar{\lambda}=0.$$

Using (42), (44), (45) in the above relation, we obtain

$$\nabla\tilde{q}(\bar{x}_R)+\sum_{i=1}^{m}\bar{\lambda}^i\nabla\tilde{h}_i(\bar{x}_R)=0,$$

which is identical to (40), which was to be proved.

To show linear independence of the gradients

$$\nabla\tilde{h}_i(\bar{x}_R), \qquad i=1,\ldots,m,$$

assume that there exists another vector

$$\tilde{\lambda}=(\tilde{\lambda}^1,\ldots,\tilde{\lambda}^m)'\neq\bar{\lambda},$$

such that

$$\nabla\tilde{q}(\bar{x}_R)+\sum_{i=1}^{m}\tilde{\lambda}^i\nabla\tilde{h}_i(\bar{x}_R)=0,$$

and define

$$\tilde{\mu}=-[\nabla_B l(\bar{x})]^{-1}[\nabla_B g(\bar{x})+\nabla_B h(\bar{x})\tilde{\lambda}].$$

Then, by reversing the argument given above, we obtain

$$\nabla_B q(\bar{x})+\nabla_B h(\bar{x})\tilde{\lambda}+\nabla_B l(\bar{x})\tilde{\mu}=0, \qquad \nabla_R q(\bar{x})+\nabla_R h(\bar{x})\tilde{\lambda}+\nabla_R l(\bar{x})\tilde{\mu}=0.$$

Comparing the above relations with (46)–(47), we obtain

$$\nabla h(\bar{x})(\bar{\lambda}-\tilde{\lambda})+\nabla l(\bar{x})(\bar{\mu}-\tilde{\mu})=0.$$

Since the gradients $\nabla h_i(\bar{x})$, $\nabla l_j(\bar{x})$ are linearly independent by assumption, it follows that

$$\tilde{\lambda}=\bar{\lambda},$$

a contradiction. Hence,

$$\nabla \tilde{h}_i(\bar{x}_R), \qquad i=1,\ldots,m,$$

are linearly independent vectors.

To prove (41), we first express the matrix within braces in terms of q, h_i, l_j, $\bar{\lambda}$, $\bar{\mu}$. We have, by a straightforward calculation based on analysis given in Ref. 13, pp. 268–269, and on the relations verified above

$$\nabla^2\tilde{q}(\bar{x}_R)+\sum_{i=1}^{m}\bar{\lambda}^i\nabla^2\tilde{h}_i(\bar{x}_R)$$
$$=[\nabla\varphi(\bar{x}_R)', I_{n-r}]\{\nabla^2 q(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}^i\nabla^2 h_i(\bar{x})+\sum_{j=1}^{r}\bar{\mu}^j\nabla^2 l_j(\bar{x})\}\begin{bmatrix}\nabla\varphi(\bar{x}_R)\\ I_{n-r}\end{bmatrix},$$

where I_{n-r} is the $(n-r)\times(n-r)$ identity matrix. Now, based on the above relation and the expressions (43) and (45) for $\nabla\varphi(\bar{x}_R)$ and $\nabla\tilde{h}_i(\bar{x}_R)$, it can be seen that (41), which is to be proved, is equivalent to the assumption (35). □

Proof of Proposition 3.1. Lemma 3.1 guarantees that, under the assumptions of Proposition 3.1, one may apply the result of Proposition 1 of Ref. 14 (also, Proposition 1 in Refs. 2, 15) to problem (34). This result yields that, given any subset $\Lambda\subset R^m$, there exists a scalar $c^*>0$ (depending on Λ) such that, for all $c\geq c^*$ and $\lambda\in\Lambda$, the problem

$$\text{minimize } \tilde{q}(z)+\sum_{i=1}^{m}\{\lambda^i\tilde{h}_i(z)+\tfrac{1}{2}c[\tilde{h}_i(z)]^2\} \tag{49}$$

has a unique solution $z(\lambda, c)$ within some neighborhood of $\bar{z}=\bar{x}_R$. Furthermore, there exists a scalar $\tilde{M}>0$ such that, for all $c\geq c^*$ and $\lambda\in\Lambda$,

$$\|z(\lambda,c)-\bar{z}\|\leq\tilde{M}\|\lambda-\bar{\lambda}\|/c, \tag{50}$$

$$\|\lambda+c\tilde{h}[z(\lambda,c)]-\bar{\lambda}\|\leq\tilde{M}\|\lambda-\bar{\lambda}\|/c. \tag{51}$$

Denote

$$x(\lambda,c)=\begin{bmatrix}\varphi[z(\lambda,c)]\\ z(\lambda,c)\end{bmatrix}. \tag{52}$$

Since problem (49) is equivalent to problem (36) in a neighborhood of $\bar{x}$, the first part of Proposition 3.1 follows. Since φ is Lipschitz continuous in a neighborhood of $\bar{z} = \bar{x}_R$, we have, for some $L > 0$,

$$\|\varphi[z(\lambda, c)] - \varphi(\bar{z})\| \leq L\|z(\lambda, c) - \bar{z}\| \leq L\tilde{M}\|\lambda - \bar{\lambda}\|/c. \tag{53}$$

Combining (50)–(53) and (39), we obtain, for some $M > 0$ and all $c \geq c^*$, and $\lambda \in \Lambda$,

$$\|x(\lambda, c) - \bar{x}\| \leq M\|\lambda - \bar{\lambda}\|/c, \qquad \|\tilde{\lambda}(\lambda, c) - \bar{\lambda}\| \leq M\|\lambda - \bar{\lambda}\|/c,$$

which were to be proved. □

References

1. BERTSEKAS, D. P., *Combined Primal–Dual and Penalty Methods for Constrained Minimization*, SIAM Journal on Control, Vol. 13, pp. 521–544, 1974.
2. BERTSEKAS, D. P., *Multiplier Methods: A Survey*, Automatica, Vol. 12, pp. 133–145, 1976.
3. ROCKAFELLAR, R. T., *Solving a Nonlinear Programming Problem by Way of a Dual Problem*, Symposia Matematica, 1976.
4. ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
5. HESTENES, M. R., *Multiplier and Gradient Methods*, Journal of Optimization Theory and Applications, Vol. 4, pp. 303–320, 1969.
6. ROCKAFELLAR, R. T., *The Multiplier Method of Hestenes and Powell Applied to Convex Programming*, Journal of Optimization Theory and Applications, Vol. 12, pp. 555–562, 1973.
7. BERTSEKAS, D. P., *Nondifferentiable Optimization via Approximation*, Mathematical Programming Study 3, Nondifferentiable Optimization, Edited by M. Balinski and P. Wolfe, North-Holland Publishing Company, Amsterdam, Holland, 1975.
8. KORT, B. W., and BERTSEKAS, D. P., *A New Penalty Function Algorithm for Constrained Minimization*, Proceedings of the 1972 IEEE Conference on Decision and Control, New Orleans, Louisiana, 1972.
9. BERTSEKAS, D. P., *A New Algorithm for Solution of Nonlinear Resistive Networks Involving Diodes*, IEEE Transactions on Circuit Theory, Vol. CAS-23, pp. 599–608, 1976.
10. GABAY, D., and MERCIER, B., *A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation*, IRIA Laboratory, Research Report No. 126, 1975.
11. BERTSEKAS, D. P., *Minimax Methods Based on Approximation*, Proceedings of the 1976 Johns Hopkins Conference on Information Sciences and Systems, Baltimore, Maryland, 1976.

12. PAPAVASSILOPOULOS, G., *Algorithms for a Class of Nondifferentiable Problems*, University of Illinois, Urbana, Illinois, Department of Electrical Engineering, MS Thesis, 1977.
13. LUENBERGER, D. G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1973.
14. BERTSEKAS, D. P., *On Penalty and Multiplier Methods for Constrained Minimization*, SIAM Journal on Control and Optimization, Vol. 14, pp. 216–235, 1976.
15. BERTSEKAS, D. P., *Convergence Rate of Penalty and Multiplier Methods*, Proceedings of the 1973 IEEE Conference on Decision and Control, San Diego, California, pp. 260–264, 1973.