Mathematical Programming 37 (1987) 19-40
North-Holland

AN EXACT PENALTY FUNCTION FOR SEMI-INFINITE PROGRAMMING

Andrew R. CONN

Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Nicholas I.M. GOULD

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Received 8 November 1984
Revised manuscript received 8 September 1986

This paper introduces a global approach to the semi-infinite programming problem that is based upon a generalisation of the ℓ_1 exact penalty function. The advantages are that the ensuing penalty function is exact and the penalties include all violations. The merit function requires integrals for the penalties, which provides a consistent model for the algorithm. The discretization is a result of the approximate quadrature rather than an a priori aspect of the model.

Key words: Semi-infinite programming, exact ℓ_1 penalty functions, global algorithms.

Introduction

Recently there has been considerable interest in so-called semi-infinite programming problems—the optimization of an objective function in finitely many variables over a feasible region defined by an infinite number of constraints. To date, much of the interest has been confined to theoretical results with, sometimes, suggestions of implementable algorithms (see, for example, the conference proceedings edited by Hettich (1979) and Fiacco and Kortanek (1983)). The majority of proposed algorithms have been local—that is, convergence to a local solution of the semi-infinite programming problem can be guaranteed provided a "sufficiently" good initial estimate of the solution is given.

To the best of our knowledge, the only global algorithms for the problem—those algorithms which guarantee convergence to a stationary point of the problem from an arbitrary initial estimate—have been those proposed by Coope and Watson (1985), Gfrerer et al. (1983), and Watson (1981, 1983).

An essential ingredient in the construction of global algorithms for nonlinear programming problems is the use of a merit function against which progress towards a solution may be measured. Such merit functions have a twofold purpose; they ensure that any sequence of iterates which decrease the merit function sufficiently

This research was partially supported by Natural Sciences and Engineering Research Council of Canada grants A-8639 and A-8442. This paper was typeset using software developed at Bell Laboratories and the University of California at Berkeley.

will converge to a stationary point, and they offer guidance as to how such successive iterates should be chosen.

In this paper we describe an exact penalty function for semi-infinite programming. This function is a generalisation of the ℓ_1 exact penalty function for nonlinear programming (see, e.g. Conn and Pietrzykowski (1977)) and may be used as a merit function for semi-infinite programming methods. The only other exact penalty function suggested to date, that of Watson (1981), may also be considered as such a generalization but, in our opinion, it is more closely related to the ℓ_∞ exact penalty function (see, e.g. Bertsekas (1982)).

In Section 2, we show that our proposed penalty function is exact under rather strong (convexity) assumptions. In Section 3, by restricting our attention to a certain class of commonly occurring semi-infinite programming problems, we are able to weaken considerably the assumptions of Section 2. Section 4 contains our conclusions and future research.

1. The problem and the penalty function

We consider the following problem: Let $T_i \subset \mathbb{R}^{p_i}$ be a compact set and let $\phi_i(x, t)$ be a function whose domain is $\mathbb{R}^n \times T_i$ and whose range is $\mathbb{R}$. Furthermore let $f:\mathbb{R}^n \to \mathbb{R}$ be a given objective function. Finally let f and ϕ_i be a continuously differentiable throughout their domains of definition. Then we shall be interested in the following semi-infinite programming problem.

$$\text{SIP: } \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) \qquad \text{subject to} \quad \phi_i(x, t) \geq 0, \ \forall t \in T_i, \ i = 1, \dots, m.$$

We make the following definitions and assumptions. Let x^* be a local minimizer of SIP and let $I_i = \{t \in T_i \mid \phi_i(x^*, t) = 0\}$.

Assumption 1. The gradients $\nabla_x \phi_i(x^*, t)$ for all $t \in I_i$ and all $1 \leq i \leq m$ are linearly independent.

Under Assumption 1, the sets I_i are necessarily finite. Hence we may write $I_i = \{t_{ik}^* \in T_i \mid \phi_i(x^*, t_{ik}^*) = 0, 1 \leq k \leq k_i\}$. It then follows that necessary conditions for x^* to be a local minimizer of SIP (see e.g. Borwein (1983)) are that there exist finite Langrange multipliers $\lambda_{ik} \geq 0$ such that

$$\nabla_x f(x^*) = \sum_{i=1}^{m} \sum_{k=1}^{k_i} \lambda_{ik} \nabla_x \phi_i(x^*, t_{ik}^*). \tag{1.1}$$

Assumption 2. For any x, there is a (possibly empty) finite set of sets $\Omega_{ij}(x)$ such that

(i) $\Omega_{ij}(x) \subseteq T_i$, $1 \leq j \leq s_i = s_i(x) < \infty$,

(ii) $\phi_i(x, t) \leq 0$, $\forall t \in \Omega_{ij}$ and $\phi_i(x, t) > 0$, $\forall t \in T_i \backslash \bigcup_{j=1}^{s_i} \Omega_{ij}(x)$,

(iii) $\Omega_{ij}(x) \cap \Omega_{ik}(x) = \{\phi\}$ if $j \neq k$, and

(iv) $\Omega_{ij}(x)$ is connected and non-trivial, i.e., $\int_{\Omega_{ij}(x)} \mathrm{d}t > 0$.

We note that almost all functions $\phi_i(x, t)$ will satisfy this assumption.

Assumption 3. For any x, and any index i, there is no *open* region U_i *strictly* contained in T_i such that $\phi_i(x, t) = 0$ for all $t \in U_i$.

The purpose of Assumption 3 is to guarantee that the penalty function which we shall construct is everywhere continuous. We note that any analytic function satisfies Assumption 3.

The aim of the penalty function approach to any nonlinear programming problem is to construct a function, the penalty function, which has the following (penalty function) property:

PFP: any local solution to the nonlinear programming problem (in our case SIP) is a local minimizer of the penalty function.

The idea is then to minimize the "easy" penalty function rather than solve the "hard" nonlinear programming problem.

An early attempt to define a penalty function for semi-infinite programming is that of Pietrzykowski (1970) (see also Germeyer (1969) and Eremin and Mazurov (1967)). Pietrzykowski defines the function

$$\rho_1(x, \mu) = \mu f(x) - \sum_{i=1}^{m} \sum_{j=1}^{s_i} \int_{\Omega_{ij}(x)} \phi_i(x, t)\, \mathrm{d}t \tag{1.2}$$

where μ is a positive scalar and shows that $\rho_1(x, \mu)$ satisfies PFP in the limit as $\mu \to 0$. Unfortunately simple examples may be constructed to show that $\rho_1(x, \mu)$ is *not* an exact penalty function. That is, it is necessary that $\mu \to 0$ for the PFP to hold. It is well known that having to let the penalty parameter $\mu \to 0$ may be undesirable for any practical method for solving a nonlinear program based upon penalty function minimization (see for example Gill, Murray and Wright (1981)).

The trouble with Pietrzykowski's penalty function appears to be that the penalty for infeasibility is too weak. This leads us to consider the following penalty function:

$$\rho_2(x, \mu) = \mu f(x) - \sum_{i=1}^{m} \left(\sum_{j=1}^{s_i} \left[\int_{\Omega_{ij}(x)} \phi_i(x, t)\, \mathrm{d}t \Big/ \int_{\Omega_{ij}(x)} \mathrm{d}t \right] \right)$$

where μ is a positive scalar.

It is possible to show that this function is an *exact* penalty function. That is, there is a threshold value $\mu_0 > 0$ such that PFP holds for all $0 < \mu \leq \mu_0$. However, this penalty function has the unfortunate drawback of being discontinuous—this difficulty can be overcome by suitably redefining the SIP but this leads to implementational difficulties we prefer to avoid.

In this paper we consider the following alternative to (1.2);

$$\rho(x, \mu) = \mu f(x) - \sum_{i=1}^{m} \left(\sum_{j=1}^{s_i} \left(\int_{\Omega_{ij}(x)} \phi_i(x, t)\, \mathrm{d}t \right) \Big/ \sum_{j=1}^{s_i} \left(\int_{\Omega_{ij}(x)} \mathrm{d}t \right) \right), \tag{1.3}$$

where μ is a positive scalar.

Such a function is easy to motivate as it is just the limit of an ℓ_1 penalty function for nonlinear programming as the number of constraints increases to infinity. Furthermore, under Assumption 3, it is clearly continuous and thus from Pietrzykowski's result it satisfies PFP in the limit as μ tends to zero.

We now intend to show that (1.3) is actually an exact penalty function. We shall find it convenient to define

$$\Delta_{ij}(x) = \int_{\Omega_{ij}(x)} \mathrm{d}t, \tag{1.4a}$$

$$\Phi_{ij}(x) = \int_{\Omega_{ij}(x)} \phi_i(x, t)\, \mathrm{d}t \tag{1.4b}$$

and thus we may write (1.3) as

$$\rho(x, \mu) = \mu f(x) - \sum_{i=1}^{m} \left(\sum_{j=1}^{s_i} \Phi_{ij}(x) \Big/ \sum_{j=1}^{s_i} \Delta_{ij}(x) \right).$$

2. The convex–concave case

We start by showing that under certain assumptions any solution to SIP is also a minimizer of $\rho(x, \mu)$. In this section, we assume

Assumption 4. $f(x)$ is convex and $\phi_i(x, t)$ is concave in x for $1 \leq i \leq m$.

Assumption 5. For all $1 \leq i \leq m$ and $1 \leq j \leq s_i$, there is a constant $\beta > 0$ such that

$$\sum_{j=1}^{s_i} \Phi_{ij}(x) \leq \beta \phi_i(x, t) \sum_{j=1}^{s_i} \Delta_{ij}(x) \tag{2.1}$$

for any $t \in \bigcup_{j=1}^{s_i} \Omega_{ij}(x)$ and for all $x \in \mathbb{R}^n$.

We shall subsequently show that Assumption 5 is automatically satisfied if T_i is convex and $\phi_i(x, t)$ is convex in t over T_i for $1 \leq i \leq m$. We note that, under Assumption 4, any local solution to SIP is a global solution. We now prove

Theorem 2.1. *Suppose Assumptions* 1-5 *hold. Then x^* is a global minimizer of $\rho(x, \mu)$ for all μ such that $0 \leq \mu \leq \mu^*$ for some $\mu^* > 0$.*

Proof. Let x be any feasible point for SIP. Then $\rho(x, \mu) = \mu f(x) \geq \mu f(x^*) = \rho(x^*, \mu)$. Thus x^* is a global minimizer of $\rho(x, \mu)$ over all feasible points x.

Conversely, let x be any infeasible point for SIP. Then, Assumption 4, elementary properties of differentiable convex functions (see for example Rockafellar (1970)) and (1.1) give

$$f(x)-f(x^*) \geq \nabla_x f(x^*)^T(x-x^*) = \sum_{i=1}^{m}\sum_{k=1}^{k_i} \lambda_{ik}\nabla_x\phi_i(x^*, t_{ik}^*)^T(x-x^*)$$

$$\geq \sum_{i=1}^{m}\sum_{k=1}^{k_i} \lambda_{ik}(\phi_i(x, t_{ik}^*) - \phi_i(x^*, t_{ik}^*))$$

$$= \sum_{i=1}^{m}\sum_{k=1}^{k_i} \lambda_{ik}\phi_i(x, t_{ik}^*), \quad \text{where } \lambda_{ik} \geq 0.$$

Consider t_{ik}^*. Either $\phi_i(x, t_{ik}^*) \leq 0$, in which case $t_{ik}^* \in \Omega_{ij}(x)$ for some index j, or $\phi_i(x, t_{ik}^*) > 0$. Hence

$$f(x)-f(x^*) \geq \sum_{i=1}^{m} \sum_{\substack{k=1\\ \phi_i(x,t_{ik}) \leq 0}}^{k_i} \lambda_{ik}\phi_i(x, t_{ik}^*) = \sum_{i=1}^{m}\sum_{j=1}^{s_i} \sum_{t_{ik}^* \in \Omega_{ij}(x)} \lambda_{ik}\phi_i(x, t_{ik}^*). \tag{2.2}$$

Let n_i be the number of t_{ik}^* contained in $\bigcup_{j=1}^{s_i} \Omega_{ij}(x)$.

If $n_i = 0$ there is no contribution from $\sum_{j=1}^{s_i} (\sum_{t_{ik}^* \in \Omega_{ij}(x)} \lambda_{ik}\phi_i(x, t_{ik}^*))$—thus, in what follows, there is no loss of generality in assuming $n_i \geq 1$. From Assumption 1, $n_i \leq n$. If $t_{ik}^* \in \bigcup_{j=1}^{s_i} \Omega_{ij}(x)$, (2.1) gives

$$\sum_{j=1}^{s_i} \Phi_{ij}(x) \leq \beta\phi_i(x, t_{ik}^*) \sum_{j=1}^{s_i} \Delta_{ij}(x), \quad \text{for any } t_{ik}^* \in \bigcup_{j=1}^{s_i} \Omega_{ij}(x).$$

Hence

$$\sum_{j=1}^{s_i} \Phi_{ij}(x) \leq \frac{\beta}{n_i} \sum_{l=1}^{s_i} \sum_{t_{ik}^* \in \Omega_{il}(x)} \phi_i(x, t_{ik}^*) \sum_{j=1}^{s_i} \Delta_{ij}(x). \tag{2.3}$$

Combining (2.2) and (2.3),

$$\rho(x,\mu)-\rho(x^*,\mu) = \mu(f(x)-f(x^*)) - \sum_{i=1}^{m}\left(\sum_{j=1}^{s_i} \Phi_{ij}(x) \Big/ \sum_{j=1}^{s_i} \Delta_{ij}(x)\right)$$

$$\geq \sum_{i=1}^{m}\sum_{j=1}^{s_i} \sum_{t_{ik}^* \in \Omega_{ij}(x)} \mu\lambda_{ik}\phi_i(x, t_{ik}^*) - \sum_{i=1}^{m}\sum_{j=1}^{s_i} \sum_{t_{ik}^* \in \Omega_{ij}(x)} \frac{\beta}{n_i}\phi_i(x, t_{ik}^*)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{s_i} \sum_{t_{ik}^* \in \Omega_{ij}(x)} \left(\mu\lambda_{ik} - \frac{\beta}{n_i}\right)\phi_i(x, t_{ik}^*)$$

Hence, provided

$$\mu\lambda_{ik} - \frac{\beta}{n_i} \leq 0 \tag{2.4}$$

for all indices i such that $n_i \geq 1$,

$$\rho(x,\mu) \geq \rho(x^*,\mu).$$

If $\lambda_{ik} = 0$, (2.4) is trivially satisfied. Otherwise, if

$$\mu \leq \beta/[(\max_i n_i)(\max_{ik} \lambda_{ik})], \tag{2.5}$$

(2.4) is satisfied. Specifically, if $\mu^* = \beta/[n \max(\lambda_{ik})]$, (2.4) is satisfied for all $0 \leq \mu \leq \mu^*$ and $\mu^* > 0$. Thus $\rho(x, \mu) \geq \rho(x^*, \mu)$ for all x provided (2.5) is satisfied which proves the theorem. □

As we have mentioned, Assumption 5 is satisfied if T_i is convex, and if $\phi_i(x, t)$ is convex for all $t \in T_i$ for $1 \leq i \leq m$. To show this we need

Lemma 2.2 *Suppose Ω is a closed bounded convex non-trivial subset of $\mathbb{R}^p$ and that $h(t)$ is a non-negative concave function in Ω. Then*

$$\int_\Omega h(t)\,\mathrm{d}t \geq \beta_p \|h\|_\infty \int_\Omega \mathrm{d}t \tag{2.6}$$

where $\|h\|_\infty = \max_{t \in \Omega} h(t)$ and $\beta_p = p^p/(p+1)^{p+1}$.

Proof. See Appendix 1. □

Now identify $h(t)$ with $-\phi_i(x, t)$. As $\phi_i(x, t)$ is convex in t, $h(t)$ is concave. Moreover, on identifying Ω with $\Omega_{ij}(x)$, $\Omega_{ij}(x)$ is clearly closed, bounded and convex as $\phi_i(x, t)$ is convex and $\Omega_{ij}(x) \subseteq T_i \subseteq \mathbb{R}^{p_i}$ with T_i convex and compact. From the lemma, we thus obtain

$$\int_{\Omega_{ij}} \phi_i(x, t)\,\mathrm{d}t \leq \beta\{\min_{t \in \Omega_{ij}(x)} \phi_i(x, t)\} \int_{\Omega_{ij}} \mathrm{d}t \leq \beta\phi_i(x, t) \int_{\Omega_{ij}} \mathrm{d}t$$

for any $t \in \Omega_{ij}(x)$, where $\beta = p^p/(p+1)^{p+1} > 0$ and $p = \min_{1 \leq i \leq m} p_i$.

Furthermore, a simplification occurs when the $\phi_i(x, t)$ are convex in t as then $s_i = 1$, $0 \leq n_i \leq 1$, and the penalty function becomes

$$\rho(x, \mu) = \mu f(x) - \sum_{i=1}^{m} \left\{ \int_{\Omega_i(x)} \phi_i(x, t)\,\mathrm{d}t \Big/ \int_{\Omega_i(x)} \mathrm{d}t \right\} \tag{2.7}$$

where $\Omega_i(x) = \{t \in T_i \mid \phi_i(x, t) \leq 0\}$—in other words j is fixed at one.

We thus have

Corollary 2.3. *Suppose Assumptions* 1-4 *are satisfied and $\phi_i(x, t)$ is convex in t over the convex region T_i, $1 \leq i \leq m$. Then x^* is a global minimizer of* (2.7) *for all μ such that*

$$0 \leq \mu \leq \mu^* = [p^p/(p+1)^{p+1}][1/(\max_{1 \leq i \leq m} \max_{1 \leq k \leq k_i} (\lambda_{ik}))],$$

where $p = \min_{1 \leq i \leq m} p_i$.

Theorem 2.1 shows that, under the stated assumptions, any solution to SIP is also a global minimizer of the penalty function $\rho(x, \mu)$. We next give a partial converse to this result.

Theorem 2.4 (partial converse to Theorem 2.1). *Suppose Assumptions* 1-5 *hold and furthermore that $x(\mu)$ is the global minimizer of $\rho(x, \mu)$. Then, if μ is sufficiently small, $x(\mu) = x^*$.*

Proof. From Assumption 5,

$$-\sum_{j=1}^{s_i} \Phi_{ij}(x) \Big/ \sum_{j=1}^{s_i} \Delta_{ij}(x) \geqslant -\beta\phi_i(x, t) \geqslant 0 \quad \text{for all } t \in \bigcup_{j=1}^{s_i} \Omega_{ij}(x).$$

$$\text{As } \min(0, \phi_i(x, t_{ik}^*)) = \begin{cases} \phi_i(x, t_{ik}^*) & \text{if } t_{ik}^* \in \bigcup_{j=1}^{s_i} \Omega_{ij}(x), \\ 0 & \text{otherwise,} \end{cases}$$

$$-\sum_{j=1}^{s_i} \Phi_{ij}(x) \Big/ \sum_{j=1}^{s_i} \Delta_{ij}(x) \geqslant -\beta \min(0, \phi_i(x, t_{ik}^*)) \quad \text{for } 1 \leqslant k \leqslant k_i.$$

Hence, summing over k,

$$\begin{aligned} -\sum_{j=1}^{s_i} \Phi_{ij}(x) \Big/ \sum_{j=1}^{s_i} \Delta_{ij}(x) &\geqslant -\frac{\beta}{k_i} \sum_{k=1}^{k_i} \min(0, \phi_i(x, t_{ik}^*)) \\ &\geqslant -\frac{\beta}{n} \sum_{k=1}^{k_i} \min(0, \phi_i(x, t_{ik}^*)) \\ &\geqslant -\sum_{k=1}^{k_i} \min\left(0, \frac{\beta}{n}\phi_i(x, t_{ik}^*)\right). \end{aligned} \tag{2.8}$$

By definition and (2.8),

$$\begin{aligned} \rho(x, \mu) &= \mu f(x) - \sum_{i=1}^{m} \left(\sum_{j=1}^{s_i} \Phi_{ij}(x) \Big/ \sum_{j=1}^{s_i} \Delta_{ij}(x) \right) \\ &\geqslant \mu f(x) - \sum_{i=1}^{m} \sum_{k=1}^{k_i} \min\left(0, \frac{\beta}{n}\phi_i(x, t_{ik}^*)\right) \triangleq \rho_L(x, \mu). \end{aligned}$$

Thus $\rho_L(x, \mu)$ is the l_1 exact penalty function associated with the nonlinear programming problem

$$\underset{x \in \mathbb{R}}{\text{minimize}}\ f(x) \qquad \text{subject to} \quad \frac{\beta}{n}\phi_i(x, t_{ik}^*) \geqslant 0, \ i = 1, \ldots, m, \ k = 1, \ldots, k_i.$$

This problem has the global solution x^* and from Pietrzykowski's Theorem 2 (1969), x^* is also a global solution of $\rho_L(x, \mu)$ for μ sufficiently small. Hence

$$\mu f(x^*) = \rho_L(x^*, \mu) \leqslant \rho_L(x, \mu) \leqslant \rho(x, \mu)$$

for all x and for μ sufficiently small. Hence, in particular,

$$\mu f(x^*) \leqslant \rho(x(\mu), \mu). \tag{2.9}$$

But $x(\mu)$ is the global minimizer of $\rho(x, \mu)$ and thus

$$\rho(x(\mu), \mu) \leqslant \rho(x^*, \mu) = \mu f(x^*). \tag{2.10}$$

Combining (2.9) and (2.10) we obtain

$$\rho(x^*, \mu) = \mu f(x^*) = \rho(x(\mu), \mu), \quad \text{for } \mu \text{ sufficiently small.}$$

As $x(\mu)$ is the global minimizer of $\rho(x, \mu)$, $x(\mu) = x^*$. □

The missing ingredient to a full converse to Theorem 3.1 is the need to assume that $x(\mu)$ is the *global* minimizer of $\rho(x, \mu)$. Ideally we should just like to assume $x(\mu)$ is a local minimizer of $\rho(x, \mu)$ and hope that the conditions on f and ϕ_i are sufficient to imply that any local minimizer of $\rho(x, \mu)$ is global. Indeed, if $\rho(x, \mu)$ were convex, the result would be immediate. However, to date, we have been unable to demonstrate the convexity of $\rho(x, \mu)$ or produce a counterexample.

Although there is considerable theoretical interest in convex-concave problems, we are primarily interested in solving more general problems. Below we consider how this may be achieved.

3. The general case

We now dispense with the strong Assumptions 4 and 5. We have already remarked that Assumption 2 is quite weak. Assumption 1 is essentially the condition that makes the semi-infinite programming problem tractable since it implies that one is able to replace the infinite number of constraints by a finite number of significant constraints. As one would expect, results concerning global minimizers in Section 2 are now replaced by local minimizers.

Before proving the main theorem of this section we require three additional assumptions and a lemma.

Assumption 6. Recall that $T_i \subseteq \mathbb{R}^{p_i}$. We assume that T_i is described by a finite number of continuously differentiable constraints.

Assumption 7. There is a neighbourhood $S(x^*)$ of x^* such that there are differentiable functions $t_{ik}(x) \in T_i$, $1 \leq i \leq m$, $1 \leq k \leq k_i$ with the following properties:

(i) $t_{ik}(x)$ are strong local minimizers of $\phi_i(x, t)$ on T_i, for any given $x \in S(x^*)$, that satisfy the usual second-order sufficiency and strict complementary slackness conditions (see for example Gill, Murray and Wright (1981, p. 82)),

(ii) $t_{ik}(x^*) = t^*_{ik}$

(iii) If t^*_{ik} lies on a certain (possibly null) set of the constraints defining the boundary of T_i, $t_{ik}(x)$ lies on the same set for all $x \in S(x^*)$,

(iv) There is a positive number ε such that any other stationary point $t(x)$ of $\phi_i(x, t)$ satisfies $|\phi_i(x, t(x))| > \varepsilon$ for all x in $S(x^*)$.

Note: It is possible to relax part (iii) of this assumption. However the presentation of the following results is significantly complicated by such a relaxation.

This assumption is similar to those made by Coope and Watson (1985) and Hettich and Van Honstede (1979). It is relatively weak in that it will be satisfied by almost all constraint functions. Moreover, the assumption is entirely local in character.

Assumption 8. We shall assume that the Lagrange multipliers λ_{ik} at any local solution of SIP are strictly positive.

Remark. This assumption is commonly made in nonlinear programming, although its motivation appears to be practical rather than theoretical, since in active set strategies it is assumed that there exists some neighbourhood of a local solution for which the multiplier's sign can be used to indicate inequality constraint activity.

Under the conditions given in Assumption 7, we define functions $\psi_{ik}(x)$, $1 \leq i \leq m$, $1 \leq k \leq k_i$, such that

$$\psi_{ik}(x) = \phi_i(x, t_{ik}(x)).$$

Before proving theorem 3.3, we need a result concerning the derivative of the $\psi_{ik}(x)$.

Lemma 3.1. *Suppose $t(x) \in T \subseteq \mathbb{R}^{p_i}$ is a local minimizer of $\phi(x, t)$ for fixed x. Then, provided assumptions 6 and 7 hold,*

$$\sum_{k=1}^{p_i} \frac{\partial \phi(x, t(x))}{\partial t_k} \nabla_x t_k(x) = 0.$$

Proof. Suppose T is described by the constraints $c_j(t) \geq 0$ and that $t(x)$ lies on the first l of these curves, i.e., $c_j(t(x)) = 0$ for $1 \leq j \leq l$, where we allow the possibility l is zero.

Now Kuhn–Tucker theory implies the existence of non-negative numbers $\lambda_j(x)$ such that

$$\frac{\partial \phi(x, t(x))}{\partial t_k} = \sum_{j=1}^{l} \lambda_j(x) \frac{\partial c_j(t(x))}{\partial t_k}, \quad k = 1, 2, \ldots, p_i, \tag{3.1}$$

(or zero, in the case where l is zero). As we have the identity $c_j(t(x)) = 0$, we may differentiate to obtain

$$\sum_{k=1}^{p_i} \frac{\partial c_j(t(x))}{\partial t_k} \nabla t_k(x) = 0. \tag{3.2}$$

Multiplying (3.1) by $\nabla_x t_k(x)$ and summing over $1 \leq k \leq p_i$, we obtain

$$\begin{aligned} \sum_{k=1}^{p_i} \frac{\partial \phi(x, t(x))}{\partial t_k} \nabla_x t_k(x) &= \sum_{k=1}^{p_i} \nabla_x t_k(x) \sum_{j=1}^{l} \lambda_j(x) \frac{\partial c_j(t(x))}{\partial t_k} \\ &= \sum_{j=1}^{l} \lambda_j(x) \sum_{k=1}^{p_i} \frac{\partial c_j(t(x))}{\partial t_k} \nabla_x t_k(x) = 0, \end{aligned}$$

using (3.2). □

Corollary 3.2. *Recalling* $\psi_{ik}(x) = \phi_i(x, t_{ik}(x))$, *we have that* $\nabla_x \psi_{ik}(x) = \nabla_x \phi_i(x, t)|_{t=t_{ik}(x)}$.

Proof

$$\begin{aligned}\nabla_x \psi_{ik}(x) &= \nabla_x \phi_i(x, t)|_{t=t_{ik}(x)} + \sum_{l=1}^{p_i} \frac{\partial \phi_i(x, t_{ik}(x))}{\partial t_l} \nabla_x (t_{ik}(x))_l \\ &= \nabla_x \phi_i(x, t)|_{t=t_{ik}(x)},\end{aligned}$$

from Lemma 3.1. □

Theorem 3.3. *Under Assumptions* 1, 2, 6, 7, 8, *and the additional assumption that* x^* *is a strong local minimizer of SIP with* f *and* ϕ_i*'s twice continuously differentiable, there exists* $\mu^* > 0$ *such that for all* $0 < \mu \leq \mu^*$, $\rho(x, \mu)$ *has a local minimizer at* x^*.

Proof. Let us suppose the converse, namely that for each arbitrarily small positive μ, there exists an $x(\mu) \neq x^*$, where $x(\mu)$ indicates a local minimizer of $\rho(x, \mu)$ such that $\lim_{\mu \to 0^+} x(\mu) = x^*$. The existence of a sequence $x(\mu)$ such that $x(\mu)$ is a local minimum of $\rho(x, \mu)$ and $\lim_{\mu \to 0^+} x(\mu) = x^*$ is guaranteed by Pietrzykowski's result (Pietrzykowksi (1970)).

Suppose in addition, $x(\mu)$ is feasible for the semi-infinite programming problem. Then we easily arrive at a contradiction as follows. Since $x(\mu)$ is a local minimizer of $\rho(x, \mu)$,

$$\rho(x(\mu), \mu) \leq \rho(x^*, \mu),$$

for μ sufficiently small. But, by the feasibility assumption, this is equivalent to

$$f(x(\mu)) \leq f(x^*),$$

which, for μ sufficiently small, contradicts the hypothesis that x^* is a strong local minimizer of f.

It remains to consider the case where $x(\mu)$ is infeasible.

Let $\bar{x}$ be any infeasible point within the neighbourhood $S(x^*)$ defined in Assumption 7. Now consider the functions $\psi_{ik}(x) = \phi_i(x, t_{ik}(x))$, defined for $1 \leq i \leq m$, $1 \leq k \leq n_i$. There are three possibilities for each such function, namely (i) $\psi_{ik}(\bar{x}) < 0$, (ii) $\psi_{ik}(\bar{x}) = 0$, and (iii) $\psi_{ik}(\bar{x}) > 0$. Let the index set $V_i(x)$ for any point $x \in S(x^*)$ be given as $V_i(x) = \{k: \psi_{ik}(x) < 0\}$. We note that, as $\bar{x}$ is infeasible, there is at least one nonempty set $V_i(\bar{x})$. Without loss of generality we may assume that $V_i(\bar{x}) = \{k: 1 \leq k \leq m_i\}$. Then, it is straightforward to show that, as $\bar{x} \in S(x^*)$, each index pair ik with $1 \leq k \leq m_i$ gives rise to non-zero functions $\Phi_{ik}(x)$ and $\Delta_{ik}(x)$ (defined by (1.4)) in the sense that t_{ik} is contained in $\Omega_{ik}(x)$ and is the only $t_{ik}(x) \in \Omega_{ik}(x)$. Finally it is clear that $\Phi_{ik}(x)$ and $\Delta_{ik}(x)$ are differentiable in some neighbourhood of $\bar{x}$.

Now define the function

$$\bar{\rho}(x, \mu) = \mu f(x) - \sum_{i=1}^{m} \left(\sum_{k=1}^{m_i} \Phi_{ik}(x) \Big/ \sum_{k=1}^{m_i} \Delta_{ik}(x) \right).$$

Notice that $\bar{\rho}(\bar{x}, \mu) = \rho(\bar{x}, \mu)$ and that $\bar{\rho}(x, \mu)$ is differentiable, in a neighbourhood of $\bar{x}$.

Our intention is to show that $\bar{x}$ cannot be a local minimizer of $\rho(x, \mu)$ by constructing a non-zero vector $\bar{h}$ so that $\rho(\bar{x} + \bar{h}, \mu) < \rho(\bar{x}, \mu)$. We shall achieve this by finding a suitable vector h and a positive scalar τ such that

$$\rho(\bar{x} + \tau h, \mu) \leq \bar{\rho}(\bar{x} + \tau h, \mu) + M\tau^2 \tag{3.3a}$$

and

$$\bar{\rho}(\bar{x} + \tau h, \mu) < \bar{\rho}(\bar{x}, \mu) - m\tau \tag{3.3b}$$

for some positive scalars M and m. It then follows that

$$\rho(\bar{x} + \tau h, \mu) < \bar{\rho}(\bar{x}, \mu) - m\tau + M\tau^2 = \rho(\bar{x}, \mu) - m\tau + M\tau^2 \leq \rho(\bar{x}, \mu)$$

for τ sufficiently small. The vector $\bar{h}$ can then be set to τh for small τ and $\rho(\bar{x} + \bar{h}, \mu) < \rho(\bar{x}, \mu)$.

Observe that significant (i.e., $O(\tau h)$) differences between $\rho(\bar{x} + \tau h, \mu)$ and $\bar{\rho}(\bar{x} + \tau h, \mu)$ can only occur if, for any i, $V_i(\bar{x} + \tau h) \neq V_i(\bar{x})$ and this can only happen if one or more of the functions $\psi_{ik}(x)$, for which $\psi_{ik}(\bar{x}) = 0$, attains a significantly negative value at $\bar{x} + \tau h$. (We may assume that any ψ_{ik} which is strictly positive or strictly negative at $\bar{x}$ will remain so for small perturbations $\bar{x} + \tau h$.) In order to prevent this, we choose h so that

$$\psi_{ik}(\bar{x} + \tau h) = \psi_{ik}(\bar{x}) + O(\tau^2) \quad \text{for all indices } ik \text{ for which } \psi_{ik}(\bar{x}) = 0. \tag{3.4}$$

Without losing generality, we suppose that $\psi_{ik}(\bar{x}) = 0$ for $m_i + 1 \leq k \leq n_i$.

To see that this has the desired effect, we note that (see for example Apostol (1974))

$$-\Phi_{ik}(x)/\Delta_{ik}(x) \leq -\psi_{ik}(x)$$

for any index k, follows from the mean value theorem for multiple integrals. Hence

$$\begin{aligned}\rho(\bar{x} + \tau h, \mu) &= \mu f(\bar{x} + \tau h) - \sum_{i=1}^{m} \left(\sum_{k=1}^{n_i} \Phi_{ik}(\bar{x} + \tau h) \Big/ \sum_{k=1}^{n_i} \Delta_{ik}(\bar{x} + \tau h) \right) \\ &\leq \bar{\rho}(\bar{x} + \tau h, \mu) - \sum_{i=1}^{m} \left(\sum_{k=m_i+1}^{n_i} \Phi_{ik}(\bar{x} + \tau h) \Big/ \sum_{k=m_i+1}^{n_i} \Delta_{ik}(\bar{x} + \tau h) \right) \\ &\leq \bar{\rho}(\bar{x} + \tau h, \mu) - \sum_{i=1}^{m} \left(\sum_{k=m_i+1}^{n_i} \Phi_{ik}(\bar{x} + \tau h)/\Delta_{ik}(\bar{x} + \tau h) \right)\end{aligned}$$

using the inequality $(a+b)/(c+d) \leq a/c + b/d$, if $a, b, c, d > 0$,

$$\begin{aligned}&\leq \bar{\rho}(\bar{x} + \tau h, \mu) - \sum_{i=1}^{m} \left(\sum_{k=m_i+1}^{n_i} \psi_{ik}(\bar{x} + \tau h) \right) \\ &\leq \bar{\rho}(\bar{x} + \tau h, \mu) + M\tau^2,\end{aligned}$$

for some $M \geq 0$, using (3.4).

Thus (3.3a) is established.

We may ensure that $\psi_{ik}(\bar{x}+\tau h)=O(\tau^2)$ by picking h so that

$$\nabla_x \psi_{ik}(\bar{x})^T h = 0 \quad m_i+1 \leqslant k \leqslant n_i, \ 1 \leqslant i \leqslant m. \tag{3.5}$$

The existence of a non-trivial solution to (3.5) is guaranteed by virtue of the fact that the number of indices ik with $m_i+1 \leqslant k \leqslant n_i$ is at most $n-1$ (since the number of indices ik with $1 \leqslant k \leqslant k_i$ is at most n, using Assumption 1, and at least one index lies in $V_i(\bar{x})$ for some i) and hence the system of equations (3.5) has a null-space of dimension one or greater. Thus it is possible to find h for which (3.3a) is satisfied.

In order to satisfy (3.3b) we use the remaining degrees of freedom given to h. Thus we choose h to be the projection of the steepest descent direction for $\bar{\rho}(x, \mu)$ at $\bar{x}$ into the subspace defined by (3.5). In fact, all we need to show is that such an h is a descent direction for $\bar{\rho}(x, \mu)$ at $\bar{x}$, as (3.3b) then follows from Taylor's theorem.

We now give a formal definition of h. We first note that

$$\nabla_x \psi_{ik}(x) = \nabla_x \phi_i(x, t)|_{t=t_{ik}(x)},$$

from Corollary 3.2.

Moreover, the $\nabla_x \psi_{ik}(x^*)$, $1 \leqslant i \leqslant m$, $1 \leqslant k \leqslant n_i$, are linearly independent and, provided x is within a suitable neighbourhood of x^*, $\nabla_x \psi_{ik}(x)$, $1 \leqslant i \leqslant m$, $1 \leqslant k \leqslant k_i$ are therefore linearly independent. The set $\nabla_x \psi_{ik}(x)$, $1 \leqslant i \leqslant m$, $m_i+1 \leqslant k \leqslant m_i$, is thus linearly independent. Now, let the columns of the matrix $\bar{Z}(x)$ represent a Lipshitz continuous basis (see for example Coleman and Sorensen (1984)) for the null-space of the vector space spanned by $\{\nabla_x \psi_{ik}(x): 1 \leqslant i \leqslant m, m_i+1 \leqslant k \leqslant n_i\}$.

We define the vectors $h(x) = -\bar{Z}(x)\bar{Z}(x)^T \nabla_x \bar{\rho}(x, \mu)$ and $h = h(\bar{x})$. Clearly such an h satisfies (3.5). It remains to show that $h^T \nabla_x \bar{\rho}(\bar{x}, \mu) < 0$; i.e. we require

$$\bar{Z}(\bar{x})^T \nabla_x \bar{\rho}(\bar{x}, \mu) \neq 0 \tag{3.6}$$

Let $Y(x) = \{x \in S(x^*) \mid \psi_{ik}(x) \leqslant 0, 1 \leqslant i \leqslant m, 1 \leqslant k \leqslant n_i\}$. Y has a non-empty interior, as follows directly from the linear independence Assumption 1. We shall show that

$$\lim_{\substack{x \to x \\ x \in Y(x)}} \bar{Z}(x)^T \nabla_x \bar{\rho}(x, \mu) \neq 0.$$

It then follows that there is a neighbourhood of x^* contained in $Y(x)$ for which $\bar{Z}(\bar{x})^T \nabla_x \bar{\rho}(\bar{x}, \mu) \neq 0$ for all $\bar{x}$ in this neighbourhood.

Consider

$$\lim_{\substack{x \to x^* \\ x \in Y(x)}} \nabla_x \bar{\rho}(x, \mu) = \mu \nabla_x f(x^*) - \sum_{i=1}^{m} \lim_{x \to x^*} \left(\nabla_x \left(\sum_{k=1}^{m_i} \Phi_{ik}(x) \Big/ \sum_{k=1}^{m_i} \Delta_{ik}(x) \right) \right).$$

We show in Appendix 2 that

$$\lim_{x \to x^*} \sum_k \Phi_{ik}(x) \Big/ \sum_k \Delta_{ik}(x) \sim \sum_k (\psi_{ik}(x) C_{ik} \Theta_{ik}(x)) \Big/ \sum_k \Theta_{ik}(x),$$

where the C_{ik} are constants, the $\Theta_{ik}(x)$ are differentiable as $x \to x^*$ and satisfy $\Theta_{ik}(x^*) = 0$. Hence, it can be seen that

$$\lim_{x \to x^*} \nabla_x \left(\sum_k \Phi_{ik}(x) \Big/ \sum_k \Delta_{ik}(x) \right) = \Theta(x^*),$$

where $\Theta(x^*)$ lies in the span of $\nabla_x \psi_{ik}(x^*)$, $1 \leq k \leq m_i$, and where $\Theta(x^*)$ is *independent* of μ. Thus we may write

$$\lim_{x \to x^*} \nabla_x \bar{\rho}(x, \mu) = \mu \nabla_x f(x^*) - \sum_{i=1}^{m} \sum_{k=1}^{m_i} \omega_{ik} \nabla_x \psi_{ik}(x^*)$$

for some coefficients ω_{ik}, *independent* of μ.

Now suppose

$$\lim_{x \to x^*} \bar{Z}(x)^{\mathrm{T}} \nabla_x \bar{\rho}(x, \mu) = 0.$$

Then

$$\mu \nabla_x f(x^*) - \sum_{i=1}^{m} \sum_{k=1}^{m_i} \omega_{ik} \nabla_x \psi_{ik}(x^*) = \sum_{i=1}^{m} \sum_{k=m_i+1}^{n_i} v_{ik} \nabla_x \psi_{ik}(x^*)$$

for some coefficients v_{ik}. But, from (1.1),

$$\mu \nabla_x f(x^*) = \sum_{i=1}^{m} \sum_{k=1}^{m_i} \lambda_{ik} \nabla_x \psi_{ik}(x^*)$$

and from the linear independence of $\nabla_x \psi_{ik}(x^*)$,

$$\mu \lambda_{ik} = \omega_{ik} \quad \text{for } 1 \leq i \leq m, \ 1 \leq k \leq m_i.$$

As not all the m_i are zero, the positivity of the λ_{ik} (Assumption 8) contradicts the non-dependence of ω_{ik} upon μ.

Thus $\lim_{x \to x^*} \bar{Z}(x)^{\mathrm{T}} \nabla_x \bar{\rho}(x, \mu) \neq 0$, (3.4) is true for all x sufficiently close to x^* for which $V_i(x) = V_i(\bar{x})$. As there are only a finite number of different possibilities for $V_i(x)$, (3.4) is therefore true for all x sufficiently close to x^*, (3.3b) is true and therefore there is a neighbourhood of x^* for which an infeasible point $\bar{x}$ cannot be a local minimizer of $\rho(x, \mu)$. As $x(\mu)$ can be made as close to x^* as we please, $x(\mu)$ cannot be infeasible for sufficiently small μ. □

4. Conclusions and future research

In this paper we have demonstrated the existence of a new exact penalty function for the semi-infinite programming problem. The function proposed is a generalisation of the exact l_1 penalty function of nonlinear programming. In fact, the theoretical results are based essentially upon the results given in the case of the l_1 penalty function by Pietrzykowski (1969), complicated by the presence of an infinite number of constraints. As such, the proofs are constructive and indeed, the authors are currently developing a globally convergent, second-order algorithm for semi-infinite programming based on these ideas.

However, in order to indicate that the method is viable numerically, the following problems from Coope and Watson (1985) were all solved numerically using a prototype second order algorithm. The results were comparable to those of Coope and Watson.

Example 1

$$\text{Minimize} \quad f(x, y) = \tfrac{1}{3}x^2 + \tfrac{1}{2}x + y^2 - y$$

$$\text{subject to} \quad \sin t - x^2 - 2xyt \geqslant 0, \quad t \in [0, 2],$$

with starting point $[x^0, y^0] = [1, 2]$, and optimal solution $[x^*, y^*] = [0, 0.5]$.

Example 2

$$\text{Minimize} \quad f(x, y) = \tfrac{1}{3}x^2 + y^2 + \tfrac{1}{2}x$$

$$\text{subject to} \quad xt^2 + y^2 - (1 - x^2t^2)^2 - y \geqslant 0, \quad t \in [0, 1]$$

with $[x^0, y^0] = [1, 2]$ and $[x^*, y^*] = [-0.75, -0.618034]$.

Example 3

$$\text{Minimize} \quad f(x, y, z) = x^2 + y^2 + z^2$$

$$\text{subject to} \quad 2 \sin 4t - x - y\,e^{zt} - e^{2t} \geqslant 0, \quad t \in [0, 1],$$

with $[x^0, y^0, z^0] = [1, 1, 1]$ and $[x^*, y^*, z^*] = [-0.213313, -1.361450, 1.853547]$.

Example 4

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad f(x) = \sum_{i=1}^{n} x_i / i$$

$$\text{subject to} \sum_{i=1}^{n} x_i t^{i-1} - \tan t \geqslant 0, \quad t \in [0, 1],$$

for $n = 3$, 6 and 8 with starting point $x^0 = 0$.

$x_3^* = [0.089096, 0.423052, 1.045260]$,

$x_6^* = [0.0, 1.023260, -0.240604, 1.221679, -1.388257, 0.941331]$,

$x_8^* = [0.0, 1.002913, -0.053486, 0.709802, -1.299414, 2.499344, -2.205328, 0.903576]$.

Example 5

$$\text{Minimize} \quad f(x_1, x_2, x_3) = \sum_{i=1}^{3} e^{x_i}$$

$$\text{subject to} \quad x_1 + x_2 t + x_3 t^2 - \frac{1}{1+t^2} \geq 0, \quad t \in [0, 1],$$

$x^0 = [1, .5, 0]$, $x^* = [1.006607, -0.126904, 0.379703]$.

Example 6

$$\text{Minimize} \quad f(x, y) = (x - 2y + 5y^2 - y^3 - 13)^2 + (x - 14y + y^2 + y^3 - 29)^2$$

$$\text{subject to} \quad e^t - x^2 - 2yt^2 - e^{x+y} \geq 0, \quad t \in [0, 1],$$

$(x^0, y^0) = (1, 2)$ and $(x^*, y^*) = (0.719961, -1.450487)$.

Example 7

$$\text{Minimize} \quad f(x, y) = 1.21e^x + e^y,$$

$$\text{subject to} \quad e^{x+y} - t \geq 0, \quad t \in [0, 1],$$

$[x^0, y^0] = [0.8, 0.9]$, $[x^*, y^*] = (-\ln 1.1, \ln 1.1)$.

The proofs above explicitly determine a first order descent direction for the penalty function. Future research entails refining the algorithm, the details of the global convergence results and consideration of both convergence rates and numerical implementation.

As was already mentioned in Section 1, the problem of generalising the l_1 penalty function of nonlinear programming to the semi-infinite case is not entirely straightforward. In particular, as is true in the nonlinear programming case, the penalty function may introduce undesirable local minima and some of the assumptions required for the theoretical results may be unnecessarily restrictive.

Finally, we also wish to investigate an approach based upon the penalty function

$$p_\infty(x, \mu) = \mu f(x) - \min_{1 \leq i \leq m} \{ \min_{t \in \Omega_i(x)} \phi_i(x, t)\}.$$

Appendix 1. Proof of Lemma 2.2

As Ω is closed and bounded, there is a point $z \in \Omega$ at which $h(z) = \|h\|_\infty$. The result is trivial if $\|h\|_\infty = 0$, so assume otherwise. For any non-zero vector p, there is a unique largest scalar $\alpha \geq 0$ such that $z + \alpha p \in \partial\Omega$, the boundary of Ω (by convexity of Ω). Let $q = \alpha p$ be called a boundary-pointing vector.

Now define the region

$$\Omega'(\beta) = \{t: t = z + \gamma q \ \forall 0 \leq \gamma \leq \beta < 1 \text{ and all boundary pointing vectors } q\}.$$

As $h(z) = \|h\|_\infty$ and $h(z+q) \geq 0$ for all q,

$$h(t) = h(z + \gamma q) = h(1-\gamma)z + \gamma(z+q) \geq (1-\gamma)h(z) + \gamma h(z+q)$$
$$\geq (1-\gamma)h(z) \geq (1-\beta)h(z) = (1-\beta)\|h\|_\infty.$$

Thus, for all $t \in \Omega'(\beta)$,

$$h(t) \geq (1-\beta)\|h\|_\infty.$$

Hence

$$\int_\Omega h(t)\,\mathrm{d}t = \int_{\Omega'(\beta)} h(t)\,\mathrm{d}t + \int_{\Omega-\Omega'(\beta)} h(t)\,\mathrm{d}t \geq \int_{\Omega'(\beta)} h(t)\,\mathrm{d}t \quad (\text{as } h \geq 0)$$
$$\geq (1-\beta)\|h\|_\infty \int_{\Omega'(\beta)} \mathrm{d}t. \tag{A1.1}$$

We now claim that

$$\int_{\Omega'(\beta)} \mathrm{d}t = \beta^p \int_\Omega \mathrm{d}t. \tag{A1.2}$$

For if we transform our co-ordinate axes so that z becomes the origin and then consider any point in $\Omega'(\beta)$ in terms of spherical polar co-ordinates (see, e.g., Edwards (1922, p. 47))

$$t_1 = r\cos\Theta_1, \ldots, \quad t_j = r\sin\Theta_1 \cdots \sin\Theta_{j-1}\cos\Theta_j, \ldots,$$
$$t_{p-1} = r\sin\Theta_1 \cdots \sin\Theta_{p-2}\cos\Theta_{p-1},$$
$$t_p = r\sin\Theta_1, \ldots, \ \sin\Theta_{p-2}\sin\Theta_{p-1}.$$

Then any point on the boundary of Ω is at $r(\Theta_1, \ldots, \Theta_{p-1})$ and any point on the boundary of $\Omega'(\beta)$ is at $\beta r(\Theta_1, \ldots, \Theta_{p-1})$. Hence

$$\int_{\Omega'(\beta)} \mathrm{d}t = \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\beta r(\Theta_1,\ldots,\Theta_{p-1})} r^{p-1} g(\Theta_1, \ldots, \Theta_{p-1})\,\mathrm{d}\Theta_1 \cdots \mathrm{d}\Theta_{p-1}\,\mathrm{d}r$$
$$= \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} \frac{\beta r(\Theta_1, \ldots, \Theta_{p-1}))^p}{p} g(\Theta_1, \ldots, \Theta_{p-1})\,\mathrm{d}\theta_1 \cdots \mathrm{d}\Theta_{p-1}$$
$$= \beta^p \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} r(\Theta_1, \ldots, \Theta_{p-1})^p g(\Theta_1, \ldots, \Theta_{p-1})\,\mathrm{d}\Theta_1 \cdots \mathrm{d}\Theta_{p-1}$$
$$= \beta^p \int_\Omega \mathrm{d}t,$$

where $g(\Theta_1, \ldots, \Theta_{p-1}) = \sin^{p-2}\Theta_1 \cdots \sin\Theta_{p-2}$.

Thus, combining (A1.1) and (A1.2),

$$\int_\Omega h(t)\,\mathrm{d}t \geq \beta^p(1-\beta)\|h\|_\infty \int_\Omega \mathrm{d}t \quad \text{for all } 0 \leq \beta \leq 1.$$

Therefore,

$$\int_\Omega h(t)\,\mathrm{d}t \geq \max_{0\leq\beta\leq 1} \beta^p(1-\beta)\|h\|_\infty \int_\Omega \mathrm{d}t = \frac{p^p}{(p+1)}\|h\|_\infty \int_\Omega \mathrm{d}t$$

which proves the lemma. □

Appendix 2

We now justify the asymptotic formulae for Φ_{ik} and Δ_{ik} needed in the proof of Theorem 3.3.

Lemma A2. *Suppose Assumptions* 6 *and* 7 *hold and that* $t(x)$ $(=t_{ik}(x)$ *for some indices* $ik) \in T_i \subset \mathbb{R}^{p_i}$. *Furthermore, suppose* $t(x)$ *lies on* m_{ik} *constraints* $c_j(t)$ *for* $1 \leq j \leq m_{ik}$, *where we allow the possibility that* $m_{ik}=0$. *Then, as* $x \to x^*$.

$$\Phi_{ik}(x) \sim (2/(p_i+m_{ik}+2))\Psi_{ik}(x)\Theta_{ik}(x) \quad \textit{and} \quad \Delta_{ik}(x) \sim \Theta_{ik}(x),$$

where, $\Theta_{ik}(x)$ *is differentiable while* $\phi(x, t_{ik}^*) \leq 0$ *and* $\Theta_{ik}(x^*)=0$.

Proof. For simplicity, in what follows we will drop the subscripts i and ik. We wish to evaluate

$$\Phi(x) = \int_{\substack{\phi(x,t)\leq 0\\ c_i(t)\geq 0,\ i=1,\ldots,m}} \phi(x,t)\,\mathrm{d}t \quad \text{and} \quad \Delta(x) = \int_{\substack{\phi(x,t)\leq 0\\ c_i(t)\geq 0,\ i=1,\ldots,m}} \mathrm{d}t,$$

where we know that $\phi(x, t(x)) < 0$, $|\phi(x,t)|$ is small for all $t \in \Omega(x)$ and $c_i(t(x))=0$, $i=1,\ldots,m$.

Without loss of generality we may assume that $t^*=0$. As $\phi(x,t)$ is assumed to be small for all t in the appropriate region we have

$$\phi(x,t) \simeq \phi(x,0) + t^{\mathrm{T}}\nabla_t\phi(x,0) + \tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}\phi(x,0)t,$$

$$c_i(t) \simeq t^{\mathrm{T}}\nabla_t c_i(0) + \tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}c_i(0)t.$$

Hence $\Phi(x)$ and $\Delta(x)$ will be approximated by

$$\Phi(x) \simeq \int_{J(t)} (\phi(x,0) + t^{\mathrm{T}}\nabla_t\phi(x,0) + \tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}\phi(x,0)t)\,\mathrm{d}t$$

and

$$\Delta(x) \simeq \int_{J(t)} \mathrm{d}t,$$

where

$$J(t)=\{t\colon \phi(x,0)+t^{\mathrm{T}}\nabla_t\phi(x,0)+\tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}\phi(x,0)t\leqslant 0,\ t^{\mathrm{T}}\nabla tc_i(0)$$
$$+\tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}c_i(0)t\geqslant 0,\ i=1,\dots,m\}.$$

Now transform co-ordinates as follows. Define

$$s_i=t^{\mathrm{T}}\nabla_t c_i(0)+\tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}c_i(0)t,\quad i=1,\dots,m. \tag{A2.1}$$

Let the $m\times n$ and $n\times\overline{n-m}$ matrices $A(t)$and $Z(t)$ be given by

$$A^{\mathrm{T}}(t)=(\nabla_t c_1(t),\dots,\nabla_t c_m(t)),$$

with $Z(t)$ satisying $A(t)Z(t)=0$, and $Z(t)^{\mathrm{T}}Z(t)=I_{n-m}$. Further, for any vector λ with ith component λ_i, define

$$M(x,\lambda)=\nabla_{tt}\phi(x,0)-\sum_{i=1}^{m}\lambda_i\nabla_{tt}c_i(0).$$

Let $s_{m+i}=(Z(0)^{\mathrm{T}}t)_{i\text{th entry}}$ and let s be the vector whose components are the s_i viz.

$$s=\begin{pmatrix} c(t)\\ Z(0)^{\mathrm{T}}t\end{pmatrix}\simeq\begin{pmatrix} A(0)\\ \hline Z^{\mathrm{T}}(0)\end{pmatrix}t\quad\text{for small perturbations about } t=0.$$

Note, by assumption, $A(0)$ is of full rank and hence the transformation is well-defined and continuous in some neighbourhood of $t=t(x)=0$.

We may now write

$$t\simeq(B\,|\,Z(0))\begin{pmatrix} s_1\\ \hline s_2\end{pmatrix}, \tag{A2.2}$$

where

$$\begin{pmatrix} A(0)\\ \hline Z^{T}(0)\end{pmatrix}(B\,|\,Z(0))=I_n$$

and s is partitioned into s_1 and s_2 with s_1 an m-vector.

As $t=t(x)=0$ is a strong local minimizer of $\phi(x,t)$ in T, by Assumption 7, there are Lagrange multipliers λ_i such that

$$\nabla_t\phi(x,0)=\sum_{i=1}^{m}\lambda_i(x)\nabla_t c_i(0), \tag{A2.3}$$

$$\lambda_i(x)>0,\ i=1,\dots,m \tag{A2.4}$$

and

$$Z(0)^{\mathrm{T}}M(x,\lambda(x))Z(0)\text{ is positive definite.} \tag{A2.5}$$

Thus,

$$\begin{aligned}
&\phi(x,0)+t^{\mathrm{T}}\nabla_t\phi(x,0)+\tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}\phi(x,0)t\\
&\quad=\phi(x,0)+\sum_{i=1}^{m}\lambda_i(x)\nabla_t c_i(0)^{\mathrm{T}}t+\tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}\phi(x,0)t,\quad\text{using (A2.3)}\\
&\quad=\phi(x,0)+\sum_{i=1}^{m}(\lambda_i(x)s_i-\lambda_i(x)\tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}c_i(0)t)+\tfrac{1}{2}t^{\mathrm{T}}\nabla_{tt}\phi(x,0)t\\
&\quad=\phi(x,0)+s_1^{\mathrm{T}}\lambda(x)+\tfrac{1}{2}t^{\mathrm{T}}M(x,\lambda(x))t\\
&\quad\simeq\phi(x,0)+s_1^{\mathrm{T}}\lambda(x)+\tfrac{1}{2}s_1^{\mathrm{T}}B^{\mathrm{T}}M(x,\lambda(x))Bs_1+\tfrac{1}{2}s_2^{\mathrm{T}}Z^{\mathrm{T}}(0)M(x,\lambda(x))Z(0)s_2\\
&\qquad+s_1^{\mathrm{T}}B^{\mathrm{T}}M(x,\lambda(x))Z(0)s_2\\
&\quad\simeq\phi(x,0)+s_1^{\mathrm{T}}\lambda(x)+\tfrac{1}{2}s_2^{\mathrm{T}}Z^{\mathrm{T}}(0)M(x,\lambda(x))Z(0)s_2\quad\text{for small } s_1,\ s_2,
\end{aligned}$$

since the terms $s_1^{\mathrm{T}}B^{\mathrm{T}}M(x,\lambda(x))Bs_1$ and $s_1^{\mathrm{T}}B^{\mathrm{T}}M(x,\lambda(x))Z(0)s_2$ are dominated by $s_1^{\mathrm{T}}\lambda(x)$ if s_1 and s_2 are small.

Thus, under our assumptions,

$$\begin{aligned}
\Phi(x)&\simeq\int_{K(s)}(\phi(x,0)+s_1^{\mathrm{T}}\lambda(x)\\
&\qquad+\tfrac{1}{2}s_2^{\mathrm{T}}Z^{\mathrm{T}}(0)M(x,\lambda(x))Z(0)s_2)\det(B\,|\,Z(0))\,\mathrm{d}s_1\,\mathrm{d}s_2,\\
\Delta(x)&\simeq\int_{K(s)}\det(B\,|\,Z(0))\,\mathrm{d}s_1\,\mathrm{d}s_2,
\end{aligned}$$

where $K(s)=\{s\colon s_1\geqslant 0;\ s_1^{\mathrm{T}}\lambda(x)+\tfrac{1}{2}s_2^{\mathrm{T}}Z^{\mathrm{T}}(0)M(x,\lambda(x))Z(0)s_2\leqslant-\phi(x,0)\}$.

As $Z^{\mathrm{T}}(0)M(x,\lambda(x))Z(0)$ is positive definite from (A2.5), we may transform the s_2 variables so that the new variables $s_3=\sqrt{Z^{\mathrm{T}}(0)M(x,\lambda(x))Z(0)}s_2$ are defined for some appropriate square root. This then gives

$$\begin{aligned}
\Phi(x)&\simeq\det(B\,|\,Z(0))\sqrt{\det(Z^{\mathrm{T}}(0)M(x,\lambda(x))Z(0))}(\phi(x,0)I_2(x)+I(x)),\\
\Delta(x)&\simeq\det(B\,|\,Z(0))\sqrt{\det(Z^{\mathrm{T}}(0)M(x,\lambda(x))Z(0))}I_2(x),
\end{aligned}\tag{A.6}$$

where

$$I(x)=\int_{\bar K(s)}(s_1^{\mathrm{T}}\lambda(x)+\tfrac{1}{2}s_3^{\mathrm{T}}s_3)\,\mathrm{d}s_1\,\mathrm{d}s_3,\quad I_2(x)=\int_{\bar K(s)}\mathrm{d}s_1\,\mathrm{d}s_3$$

and

$$\bar K(s)\triangleq\left\{s=\begin{pmatrix}s_1\\ s_3\end{pmatrix}:\ s_1\geqslant 0;\ s_1^{\mathrm{T}}\lambda(x)+\tfrac{1}{2}s_3^{\mathrm{T}}s_3\leqslant-\phi(x,0)\right\}$$

Writing

$$I(x)=\int_{\substack{s_1\geqslant 0\\ \lambda(x)^{\mathrm{T}}s_1\leqslant-\phi(x,0)}}\left(\int_{\frac{1}{2}s_3^{\mathrm{T}}s_3\leqslant-(\phi(x,0)+\lambda(x)^{\mathrm{T}}s_1)}(\lambda(x)^{\mathrm{T}}s_1+\tfrac{1}{2}s_3^{\mathrm{T}}s_3)\,\mathrm{d}s_3\right)\mathrm{d}s_1,$$

we first evaluate

$$I_1(s_1)=\int_{\frac{1}{2}s_3^{\mathrm{T}}s_3\leq-(\phi(x,0)+\lambda(x)^{\mathrm{T}}s_1)}(\lambda(x)^{\mathrm{T}}s_1+\tfrac{1}{2}s_3^{\mathrm{T}}s_3)\,\mathrm{d}s_3$$

$$=\frac{\pi^{q/2}}{\Gamma(\frac{1}{2}q+1)}\frac{[-2(\phi(x,0)+\lambda(x)^{\mathrm{T}}s_1)]^{q/2}}{(q+2)}(2\lambda(x)^{\mathrm{T}}s_1-q\phi(x,0)),$$

where $q=p-m$ and $\Gamma(u)$ is the gamma function, using a variation of Apostol (1974, p. 431).

By assumption (A2.4), the matrix $\Lambda=\operatorname{diag}(\lambda(x)_i)$ is non-singular. Using the change of variables $s_4=\Lambda s_1$, we may write

$$I(x)=\int_{s_4\geq0,\ e^{\mathrm{T}}s_4\leq-\phi(x,0)}\bar{I}(s_4)\,\mathrm{d}s_4,$$

where e is a vector of ones and

$$\bar{I}(s_4)=\frac{\pi^{q/2}}{\Gamma(\frac{1}{2}q+1)}\frac{[-2(\phi(x,0)+e^{\mathrm{T}}s_4)]^{q/2}}{(q+2)\prod_{i=1}^{m}\lambda(x)_i}[2e^{\mathrm{T}}s_4-q\phi(x,0)]$$

$$=\frac{-\pi^{q/2}}{\Gamma(\frac{1}{2}q+1)(q+2)\prod_{i=1}^{m}\lambda(x)_i}([-2(\phi(x,0)+e^{\mathrm{T}}s_4)]^{q/2+1}$$

$$+(q+2)\phi(x,0)[-2(\phi(x,0)+e^{\mathrm{T}}s_4)]^{q/2}).$$

An elementary exercise in integral calculus then yields

$$I(x)=\frac{\pi^{q/2}2^{q/2}(q/2+m)(-\phi(x,0))^{m+q/2+1}}{\prod_{i=1}^{m}\lambda(x)_i\Gamma(\frac{1}{2}q+m+1)(q/2+m+1)}.$$

Similarly,

$$I_2(x)=\int_{\lambda(x)^{\mathrm{T}}s_1+\frac{1}{2}s_3^{\mathrm{T}}s_3\leq-\phi(x,0),\ s_1\geq0}\mathrm{d}s_1\,\mathrm{d}s_3$$

$$=\frac{\pi^{q/2}}{\Gamma(\frac{1}{2}q+1)}\frac{1}{\prod_{i=1}^{m}\lambda(x)_i}\int_{s_4\geq0,\ e^{\mathrm{T}}s_4\leq-\phi(x,0)}[-2(\phi(x,0)+\mathrm{e}^{\mathrm{T}}s_4)]^{q/2}\,\mathrm{d}s_4$$

$$=\frac{\pi^{q/2}2^{q/2}(-\phi(x,0))^{m+q/2}}{\prod_{i=1}^{m}\lambda_i(x)\Gamma(\frac{1}{2}q+m+1)}.$$

Thus we may rewrite

$$I(x)=\frac{(q/2+m)}{(q/2+m+1)}(-\phi(x,0))I_2(x).$$

Hence on reintroducing $t(x)=0$, (A2.6) gives

$$\Phi(x)\simeq2/(p+m+2)\phi(x,t(x))\Theta(x),\quad \Delta(x)=\Theta(x)$$

where

$$\Theta(x) = \det(B \mid Z(t(x)))\sqrt{\det(z^{\mathrm{T}}(t(x))M(x, \lambda(x))Z(t(x)))}$$
$$\times \frac{\pi^{(p-m)/2} 2^{(p-m)/2} (-\phi(x, t(x)))^{(p+m)/2}}{\prod_{i=1}^{m} \Gamma(\frac{1}{2}(p+m)+1)}.$$

Finally, it is easy to see that $\Theta(x)$ satisfies the conclusions of the theorem. □

Acknowledgements

We would like to thank Jonathan Goodman and M.J.D. Powell for their useful comments and Vera Kovacevic-Vujcic, who, at Oberwolfach, introduced us to semi-infinite programming.

References

T.M. Apostol, *Mathematical Analysis, 2nd edition* (Addison-Wesley, Reading, MA., 1974).

D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press, New York, 1982).

J.M. Borwein, "Semi-infinite programming duality: how special is it?," in: A.V. Fiacco and K.O. Kortanek, eds., *Semi-Infinite Programming and Applications*, Lecture notes in economic and mathematical systems 215 (Springer-Verlag, New York, 1983) pp. 10-36.

T.F. Coleman and D.C. Sorensen, "A note on the computation of an orthonormal basis for the null space of a matrix," *Mathematical Programming* 29 (1984) 234-242.

A.R. Conn and T. Pietrzykowski, "A penalty function method converging directly to a constrained optimum," *SIAM Journal on Numerical Analysis* 14 (1977) 348-375.

I.D. Coope and G.A. Watson, "A projected Lagrangian algorithm for semi-infinite programming," *Mathematical Programming* 32 (1985) 337-356.

J. Edwards, *A Treatise on the Integral Calculus, with Applications, Examples and Problems, Volume 2* (Macmillan, London, 1922).

I.I. Eremin and Vl.D. Mazurov, "Iteration method for solving problems of convex programming," *Soviet Physics-Doklady* 9 (1967) 757-759.

A.V. Fiacco and K.O. Kortanek, eds., *Semi-Infinite Programming and Applications*, Lecture notes in economics and mathematical systems 215 (Springer-Verlag, New York, 1983).

Yu. B. Germeyer, "Approximate reduction of the problem of determining a maximum by means of penalty functions," *USSR Computational Mathematics and Mathematical Physics* 9 (1969) 325-328.

H. Gfrerer, J. Guddat, Hj. Wacker and W. Zulehner, "Globalization of locally convergent algorithms for nonlinear optimization problems with constraints," in: A.V. Fiacco and K.O. Kortanek, eds., Lecture notes in economic and mathematical systems 215 (Springer-Verlag, New York, 1983) pp. 128-137.

P.E. Gill, W. Murray and M.H. Wright, *Practical Optimization* (Academic Press, New York, 1981).

S.-A. Gustafson, "A three-phase algorithm for semi-infinite programs", in: A.V. Fiacco and K.O. Kortanek, eds., *Semi-Infinite Programming and Applications*, Lecture notes in economics and mathematical systems 215 (Springer-Verlag, New York, 1983) pp. 138-157.

R. Hettich, ed., *Semi-infinite Programming*, Lecture notes in control and information sciences 15 (Springer-Verlag, New York, 1979).

R. Hettich and W. Van Honstede, "On quadratically convergent methods for semi-infinite programming," in: R. Hettich, ed., *Semi-infinite Programming*, Lecture notes in control and information science 15 (Springer-Verlag, New York, 1979) pp. 97-111.

T. Pietrzykowski, "An exact potential method for constrained maxima," *SIAM Journal of Numerical Analysis* 6 (1969) 299-304.

T. Pietrzykowski, "The potential method for conditional maxima in locally compact metric spaces," *Numerische Mathematik* 14 (1970) 325-329.

R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, N.J., 1970).

G.A. Watson, "Globally convergent methods for semi-infinite programming," *Nordisk Tidskrift for Informationsbehandling (BIT)* 21 (1981) 36-373.

G.A. Watson, "Numerical experiments with globally convergent methods for semi-infinite programming problems," in: A.V. Fiacco and K.O. Kortanek, eds., *Semi-infinite Programming and Applications*, Lecture notes in economics and mathematical systems 215 (Springer-Verlag, New York, 1983) pp. 193-205.