

INVESTIGATING SEMI-INFINITE PROGRAMS USING PENALTY FUNCTIONS AND LAGRANGIAN METHODS

SVEN-ÅKE GUSTAFSON¹

(Received 29 July 1985; revised 15 November 1985)

Abstract

In this paper the relations between semi-infinite programs and optimisation problems with finitely many variables and constraints are reviewed. Two classes of convex semi-infinite programs are defined, one based on the fact that a convex set may be represented as the intersection of closed halfspaces, while the other class is defined using the representation of the elements of a convex set as convex combinations of points and directions. Extension to nonconvex problems is given. A common technique of solving a semi-infinite program computationally is to derive necessary conditions for optimality in the form of a nonlinear system of equations with finitely many equations and unknowns. In the three-phase algorithm, this system is constructed from the optimal solution of a discretised version of the given semi-infinite program, i.e. a problem with finitely many variables and constraints. The system is solved numerically, often by means of some linearisation method. One option is to use a direct analog of the familiar SOLVER method.

1. Introduction

Linear semi-infinite programs were originally defined by generalising linear programs allowing *either* the number of constraints *or* the number of variables (but not both) to become infinite. Consider namely:

PROGRAM(LP). *Minimise*

$$c^T y, \tag{1a}$$

¹Department of Numerical Analysis and Computing Science, The Royal Institute of Technology, S-10044 Stockholm 70, Sweden.

© Copyright Australian Mathematical Society 1986, Serial-fee code 0334-2700/86

over all $y \in R^n$, subject to the constraints

$$A^T y \geq b. \quad (1b)$$

Here $c \in R^n$ and $b \in R^N$ are fixed vectors and A a given n by N matrix.

For all values of N , (1b) defines a convex subset of R^N (which could possibly be empty) and the generalisation to infinitely many constraints (1b) appears straight-forward. (See e.g. Glashoff-Gustafson, [4]). We consider now the dual of the problem (1). (In the sequel we refer to a *group* of formulae whose members are distinguished by letters, by giving the common element of the formula labels, i.e. the number.)

PROGRAM(LD). Let A, b, c be as in (1). Maximise

$$b^T x, \quad (2a)$$

over all vectors $x \in R^N$, subject to the constraints

$$Ax = c, \quad (2b)$$

$$x \geq 0. \quad (2c)$$

If Program(LD) has an optimal solution, then it has an optimal solution x with at most n positive components x_{j_i} , $i = 1, \dots, q$. Therefore we may reformulate Program(LD) thus:

PROGRAM(LD), SECOND FORMULATION. Let A, b, c, n and N be as in Program(LP). Determine an integer q , a subset $\{j_1, \dots, j_q\} \subset \{1, \dots, N\}$ and reals x_{j_1}, \dots, x_{j_q} such that the expression

$$\sum_{i=1}^q x_{j_i} b_{j_i}, \quad (3a)$$

is rendered a maximum, subject to the constraints

$$\sum_{i=1}^q x_{j_i} a_{j_i} = c, \quad (3b)$$

$$x_{j_i} \geq 0, \quad i = 1, \dots, q. \quad (3c)$$

Here the columns of A are denoted a_1, \dots, a_N .

We note that Program(LD) is feasible, if $c \in R^n$ can be written as a nonnegative linear combination of the columns of A . If such a representation exists, then it is easy to establish that $q \leq n$ columns are needed in this representation, independently of the value of N , the total number of columns. This observation paves the way for considering generalisations of Program(LD) to the case when $\{1, \dots, N\}$ is replaced by an infinite set.

Thus we get two classes of semi-infinite programs by generalising Program(LP) and (LD). In the first class, the set of feasible vectors $y \in R^n$ is represented as the intersection of a collection of closed halfspaces, while in the second class the given vector $c \in R^n$ shall be written as a nonnegative linear combination of a given set of vectors. Thus we need to establish that c is an element of a certain convex set whose members are, in the sense of Rockafellar, [11], represented as convex combinations of a given collection of points and directions. Among all possible representations of c , one seeks the ones rendering the preference function a maximum. We illustrate the procedure with

EXAMPLE 1. *Minimise*

$$y_1 + \frac{1}{4}y_2, \tag{4a}$$

over all y_1, y_2 subject to the constraint

$$y_1 + y_2s \geq \sqrt{1 + s}, s \in [0, 1]. \tag{4b}$$

The dual of this problem is (See Glashoff-Gustafson [4])

Maximise

$$\sum_{i=1}^q x_i \sqrt{1 + s_i}, \tag{5a}$$

subject to the constraints:

$$\sum_{i=1}^q x_i = 1, \tag{5b}$$

$$\sum_{i=1}^q x_i s_i = 1/4, \tag{5c}$$

$$s_i \in [0, 1], x_i \geq 0, i = 1, \dots, q. \tag{5d}$$

If we replace the interval $[0, 1]$ with a finite subset $T = \{s_1, \dots, s_N\}$ in (4b) and (5d) then (4) and (5) define a dual pair of linear programs. In the semi-infinite case we need to write $(1, \frac{1}{4})^T$ as a nonnegative linear combination of the vectors $(1, s)^T$ and to determine an optimal representation. We find the optimal solutions specified by $q = 1, x_1 = 1, s_1 = \frac{1}{4}, y_1 = 0.45\sqrt{5}, y_2 = 0.2\sqrt{5}$. We note in passing that if the function f has two continuous derivatives in $[0, 1]$ then the relation holds:

$$\int_0^1 f(s) \ln 1/s ds = f(1/4) + \frac{7}{288} f''(\xi), \quad 0 < \xi < 1.$$

Thus the optimal solution of (5) defines a one-point generalised Gauss rule corresponding to the weighting function $\ln(1/s)$ on the interval $[0, 1]$, since the mechanical quadrature rule

$$\int_0^1 f(s) \ln 1/s ds \approx f(1/4),$$

is exact if f is a polynomial of degree less than 2, i.e. a linear function.

Provided certain regularity assumptions are satisfied when generalising Programs(LP) and (LD), the resulting dual pair of semi-infinite programs preserves many important properties of linear programs, such as complementary slackness.

The procedure outlined above may be applied to *nonlinear convex* programs. Then the Wolfe dual plays the role of Program(LD). This is discussed in Section 2. However, the convexity assumption is not valid for many important tasks like nonlinear approximation. Therefore, generalisation to nonconvex problems is called for. Hence one arrives at nonconvex semi-infinite programs which have been treated numerically with success. Thus problems in a fairly large and diverse class can be referred to as “semi-infinite” programs. However, all of these problems may be treated by means of computational schemes based on the principles of the *three-phase algorithm*:

Phase 1: Determine an approximate solution by solving a discretised version of the given problem.

Phase 2. Determine necessary conditions in the form of a (nonlinear) system of equations.

Phase 3. Solve the nonlinear system of equations (numerically).

2. General semi-infinite programs

We start by considering the following *convex* optimisation problem:

PROGRAM(P_{CT}). Let S be a compact subset of R^k , $k < \infty$, f a convex differentiable function on R^n . Let further $g(y, s)$ be a function which is defined and continuously differentiable for all $s \in S$, $y \in R^n$. For each fixed $s \in S$, f is convex as a function of y . Finally let

$$T = \{s_1, \dots, s_N\} \subset S,$$

be a given subset.

Determine

$$\inf f(y), \tag{6a}$$

when y varies over R^n subject to the constraints

$$g(y, s) \leq 0, s \in T. \tag{6b}$$

We formulate the Wolfe dual of Program(P_{CT}) as follows. (See Fletcher, [3], p 69.)

PROGRAM(D_{CT}). *Let the notations and assumptions of Program(P_{CT}) prevail!*
Determine

$$\max f(y) + \sum_{j=1}^N x_j g(y, s_j), \tag{7a}$$

when y varies over R^n and x_j are real numbers subject to the constraints:

$$f_y(y) + \sum_{j=1}^N x_j g_y(y, s_j) = 0, \tag{7b}$$

$$x_j \geq 0, \quad j = 1, \dots, N. \tag{7c}$$

We next study the semi-infinite case and define:

PROGRAM(P_{CS}). Replace T in Program(P_{CT}) with S .

PROGRAM(D_{CS}). Let the assumptions and notations of Program(D_{CT}) prevail! Determine an integer q , a subset $\{s_1, \dots, s_q\} \subset S$ and reals x_1, \dots, x_q such that the expression

$$f(y) + \sum_{i=1}^q x_i g(y, s_i), \tag{8a}$$

is rendered a maximum when y varies over R^n and $\{s_1, \dots, s_q\}$, x_1, \dots, x_q are subject to the constraints

$$f_y(y) + \sum_{i=1}^q x_i g_y(y, s_i) = 0, \tag{8b}$$

$$x_i \geq 0, \quad i = 1, \dots, q. \tag{8c}$$

The following result is well-known. (See e.g. Theorem 2 on p 145 in Gustafson [7].)

THEOREM 1. Let Program(P_{CS}) have an optimal solution y_0 and assume also that there is a vector $h \in R^n$ such that

$$g(y, s) + h^T g_y(y, s) < 0, \text{ all } s \in S \text{ for } y = y_0. \tag{9}$$

If y is an optimal solution, then there are positive numbers x_i and elements $s_i \in S$ such that

$$f_y(y) + \sum_{i=1}^q x_i g_y(y, s_i) = 0, \tag{10a}$$

where the vectors $g_y(y, s_i)$, $i = 1, \dots, q$ are linearly independent,

$$g(y, s_i) = 0, \quad i = 1, \dots, q, \tag{10b}$$

the function

$$\Delta(s) = g(y, s) \text{ has a local maximum at } s_i. \tag{10c}$$

Finally, Programs(P_{CS}) and (D_{CS}) have a joint optimal value.

The relations (10) are used to construct a nonlinear system of equations from whose solutions y, q, s_1, \dots, s_q and x_1, \dots, x_q one may determine optimal solutions to Programs(P_{CS}) and (D_{CS}).

EXAMPLE 2. Minimise

$$y_1^2 + y_2^2,$$

subject to the constraints

$$e^s - y_1 - y_2s \leq 0, s \in [0, 1]. \tag{11}$$

The dual problem may be written:

Maximise

$$y_1^2 + y_2^2 + \sum_{i=1}^q x_i(e^{s_i} - y_1 - y_2s_i),$$

subject to the constraints

$$\sum_{i=1}^q x_i = 2y_1, \tag{12a}$$

$$\sum_{i=1}^q x_i s_i = 2y_2. \tag{12b}$$

We observe that (11) is equivalent to the simpler conditions

$$e^s - y_1 - y_2s \leq 0 \text{ for } s = 0 \text{ and } s = 1.$$

Therefore the primal is replaced with an equivalent optimisation problem involving just 2 linear constraints. We find the optimal solution

$$y_1 = y_2 = e/2.$$

Entering this into (12) we arrive at

$$q = 1, \quad s_1 = 1, \quad x_1 = e.$$

We next discuss the special case of linear semi-infinite programs. Specialising Program(P_{CS}) we get

PROGRAM(P_{LS}). Use the notations of Program(P_{LS}) but put

$$f(y) = c^T y,$$

$$g(y, s) = b(s) - a(s)^T y.$$

Here *b* is a real-valued function on *S*, *a* a vector-valued function on the same set and *c* ∈ *R*^{*n*} a fixed vector.

The generalisation to nonconvex semi-infinite programs is straight-forward. We introduce

PROGRAM(P_S). Remove the convexity requirements in Program(P_{CS}).

Program(P_S) may have several (or none) local extrema. No duality equality can be established but if condition (9) is satisfied, then (10) delivers necessary conditions for optimality. Program(D_{CS}) may be substantially generalised to give the task:

PROGRAM(Π_S). Let *S* ∈ *R*^{*k*}, *k* < ∞ be a compact set. Determine an integer *q*, a subset {*s*₁, ..., *s*_{*q*}} ⊂ *S*, reals *x*₁, ..., *x*_{*q*} and a vector *y* ∈ *R*^{*n*} in order to maximise the expression

$$F(y) + \sum_{i=1}^q x_i b(y, s_i), \tag{13a}$$

subject to the constraints

$$\sum_{i=1}^q x_i a(y, s_i) = c(y), \tag{13b}$$

$$x_i \geq 0, \quad i = 1, \dots, q. \tag{13c}$$

Here F, a, b, c are given continuously differentiable functions: F and c are defined on R^n , a and b on $S \times R^n$. Finally, F and b are real-valued while c and a take values in R^n .

REMARK. Consider the special case when a, b, c, F all are independent of y . Then (13b) and (13c) define a convex subset of R^n whose elements are represented as convex combinations of points and directions in Rockafellar's terminology.

3. Penalty functions. Lagrange multipliers

Consider Program(P_S). Put

$$m(y) = \max_{s \in S} g(y, s).$$

Then Program(P_S) takes the form

Minimise $f(y)$

over all $y \in R^n$ subject to the constraint

$$m(y) \leq 0.$$

Therefore a semi-infinite program may be cast into the form of a non-smooth optimisation problem with a single constraint.

One may also introduce penalty functions. We discuss the two cases:

Minimise over all $y \in R^n$

$$f(y) + \int_S \max(0, g(y, s)) w(s) ds,$$

w continuous and positive on S and

$$f(y) + M \cdot \max(0, m(y)). \tag{14}$$

We illustrate the two choices with

EXAMPLE 3. Minimise

$$y_1 + y_2/2,$$

subject to the constraint

$$-s^2 - y_1 - y_2 s \leq 0, \quad 0 \leq s \leq 1.$$

The problem

Minimise

$$y_1 + y_2/2 + M \cdot \int_0^1 \max(0, -s^2 - y_1 - y_2 s) ds,$$

has the solution

$$y_1 = \frac{1}{4} - \frac{1}{16M^2}, \quad y_2 = -1.$$

But the problem

Minimise

$$y_1 + y_2/2 + M \max\left(0, \max_{0 \leq s \leq 1} (-s^2 - y_1 - y_2s)\right),$$

has for all $M > 1$ the solution

$$y_1 = 1/4, \quad y_2 = -1.$$

These results are verified by straight-forward calculations, observing that the straight line $y_1 + y_2s$ intersects the curve $-s^2$ at most twice.

THEOREM 2. *Consider the convex problem Program(P_{CT}). If M is chosen large enough the penalty function given in (14) delivers an exact optimal solution, provided the Wolfe dual, Program (D_{CT}) has an optimal solution.*

PROOF. (14) may be written

$$\min_{y, y_0} f(y) + M \cdot y_0,$$

subject to the constraints

$$-y_0 \leq 0, \quad g(y, s) - y_0 \leq 0, \quad s \in S.$$

Its Wolfe dual reads

$$\max f(y) + My_0 - \lambda y_0 + \sum_{i=1}^q x_i (g(y, s_i) - y_0), \quad (15)$$

subject to

$$\begin{aligned} \sum_{i=1}^q x_i g_y(y, s_i) + f_y(y) &= 0, \\ -\lambda - \sum_{i=1}^q x_i + M &= 0, \\ x &\geq 0, \quad i = 1, \dots, q, \quad \lambda \geq 0. \end{aligned} \quad (16)$$

After simplification (15) becomes

$$f(y) + \sum_{i=1}^q x_i g(y, s_i).$$

For M large enough, the constraint (16) will not be binding, giving the desired conclusion. With the general problem Program(P_S) we may associate the

Lagrangian function

$$L = f(y) + \sum_{i=1}^q x_i g(y, s_i).$$

When we seek its stationary values, we are led to the same nonlinear system as before. Since the Lagrange multipliers occur in the solution of the dual program they can sometimes be interpreted as weights of quadrature rules. This is illustrated by Example 1, where only one multiplier occurred.

4. Discretisation. Stability

It is an essential prerequisite for successful numerical treatment that the problem to be solved has a well-defined solution which can be represented in the computer. We therefore introduce the so-called *regularised* problem:

PROGRAM(P_R). Use the notations of Program(P_S). Let y₀ be a real variable, M_B and M_F given positive constants.

Determine

$$\min f(y) + y_0 M_F, \tag{17a}$$

subject to the constraints

$$g(y, s) - y_0 \leq 0, \quad s \in S, \tag{17b}$$

$$-y_0 \leq 0, \tag{17c}$$

$$|y_r| \leq M_B, \quad r = 1, \dots, n. \tag{17d}$$

REMARK. Let *f* and *g* in (17) also satisfy the convexity requirements of Program(P_{CS}). Then Program(P_R) has an optimal solution with y₀ = 0, if Program(P_{CS}) and its dual have optimal solutions satisfying

$$\|y\|_\infty < M_B, \quad \sum_{i=1}^q x_i < M_F.$$

We note that the constraints in (17) are consistent also in a more general context, since we may select an arbitrary $\bar{y} \in R^n$ satisfying (17d) and then put

$$y_0 = \max_{s \in S} g(\bar{y}, s).$$

For the numerical treatment, we approximate Program(P_R) with Program(P_{RT}), the task we get when *S* in (17b) is replaced by a finite subset *T*. The resulting problem may be solved, e.g. by the SOLVER method described in Fletcher [2].

Using the fact that the solutions of Program(P_{RT}) are constrained to a compact subset of *R*^{*n*} we may establish that the solutions of the discretised problem can be made arbitrarily close to those of Program(P_R), provided that the grid *T* is made sufficiently “fine”, as explained in Gustafson [6], [7]. The numerical stability

questions associated with linear semi-infinite programs are examined in Gustafson [5]. We discuss here

EXAMPLE 4. Use the notations of Example 1 but replace in (4b) the interval $[0, 1]$ by the discrete set with $N + 1$ points:

$$T = \{0, 1/N, 2/N, \dots, 1\}.$$

The resulting discretised problem may be solved by means of linear programming. To describe the result, we distinguish between the two cases (a) and (b) below:

Case (a). $\frac{1}{4}$ belongs to the grid T . Then the dual has the optimal solution

$$q = 1, \quad x_1 = 1, \quad s_1 = 1/4.$$

(This result is independent of N as long as $\frac{1}{4}$ belongs to the grid.) Let now t_1, t_2 be the two gridpoints adjacent to $\frac{1}{4}$ and such that $t_1 < \frac{1}{4} < t_2$. Then y_1, y_2 define an optimal solution, if and only if

$$y_1 + y_2/4 = \sqrt{5}/2, \tag{18a}$$

$$y_1 + y_2 t_i \geq \sqrt{1 + t_i}, \quad i = 1, 2. \tag{18b}$$

Thus the optimal solution is not unique.

Case (b). $\frac{1}{4}$ does not belong to the grid T . Let t_1, t_2 be the two adjacent gridpoints satisfying $t_1 < \frac{1}{4} < t_2$. We find that the dual has the unique solution

$$\begin{aligned} q &= 2, \quad s_1 = t_1, \quad s_2 = t_2, \\ x_1 &= \frac{t_2 - \frac{1}{4}}{t_2 - t_1}, \quad x_2 = \frac{\frac{1}{4} - t_1}{t_2 - t_1}, \end{aligned}$$

and the unique solution of the primal is defined by

$$y_1 + y_2 t_i = \sqrt{1 + t_i}, \quad i = 1, 2. \tag{19}$$

We observe here the relation

$$x_1 + x_2 = 1.$$

Further, (19) gives

$$y_1 = \sqrt{1 + t_2} - \frac{t_2}{\sqrt{1 + t_1} + \sqrt{1 + t_2}}, \tag{20a}$$

$$y_2 = \frac{1}{\sqrt{1 + t_1} + \sqrt{1 + t_2}}. \tag{20b}$$

To be able to compare (20) with (18), we observe that (18) is equivalent to the three relations

$$\begin{aligned} y_1 + \frac{y_2}{4} &= \sqrt{5}/2, \\ y_2 \cdot (t_2 - 1/4) &\geq \sqrt{1 + t_2} - \sqrt{5}/2, \\ y_2 \cdot (1/4 - t_1) &\geq \sqrt{5}/2 - \sqrt{1 + t_1}. \end{aligned}$$

Hence we may conclude

$$\frac{1}{\sqrt{5}/2 + \sqrt{1+t_2}} \leq y_2 \leq \frac{1}{\sqrt{5}/2 + \sqrt{1+t_1}}.$$

Thus the solutions of the discretised *primal* problems converge towards that of the corresponding continuous problem when the grid T is made finer, i. e. when t_1 and t_2 approach $\frac{1}{4}$. In the *dual* problem, the integer q depends on N and large values of N do not imply the “correct” value $q = 1$. However,

$$x_1 + x_2 = 1 \quad \frac{x_1 s_1 + x_2 s_2}{x_1 + x_2} = \frac{1}{4},$$

and these relations hold for all $n > 2$, illustrating the clustering procedure given on page 141 in Glashoff-Gustafson [4]. We discuss now the case when S is a real interval and s_1, \dots, s_q are all in its interior. Then an optimal solution of (17) satisfies, provided $y_0 = 0$ and that (17d) is not binding:

$$g(y, s_i) = 0, \quad i = 1, \dots, q, \tag{21a}$$

$$g_s(y, s_i) = 0, \quad i = 1, \dots, q, \tag{21b}$$

$$\sum_{i=1}^q x_i g_y(y, s_i) + f_y(y) = 0. \tag{21c}$$

The optimal solution of the discretised version satisfies under the same assumptions:

$$g(y, \sigma_i) = 0, \quad i = 1, \dots, 2q, \tag{22a}$$

$$\sum_{i=1}^{2q} \xi_i g_y(y, \sigma_i) + f_y(y) = 0. \tag{22b}$$

Here $\sigma_i, i = 1, \dots, 2q$ belong to the subset T . We rewrite (22), putting $\sigma_{2i} - \sigma_{2i-1} = h_i$,

$$g(y, \sigma_{2i-1}) = 0, \tag{23a}$$

$$\frac{g(y, \sigma_{2i}) - g(y, \sigma_{2i-1})}{h_i} = 0, \tag{23b}$$

$$\sum_{i=1}^q \left((\xi_{2i-1} + \xi_{2i}) g(y, \sigma_{2i-1}) + h_i \xi_{2i} \frac{g(y, \sigma_{2i}) - g(y, \sigma_{2i-1})}{h_i} \right) + f_y(y) = 0. \tag{23c}$$

If now the grid T is fine, then the numbers h_i become small and the system (23) approaches (22), if we identify x_i with $\xi_{2i-1} + \xi_{2i}$. Therefore the solution of the discretised problem can be used for constructing an approximate solution of the continuous program.

We have not treated the general problem $\text{Program}(\Pi_S)$. A discretisation would imply that the infinite set S is replaced by a finite subset T resulting in a problem with finitely many variables and constraints. A regularisation could be imposed by constraining y to a bounded subset of R^n .

References

- [1] A. V. Fiacco and K. O. Kortanek (eds), "Semi-infinite programming and applications", *Lecture notes in economics and mathematical systems No 215* (Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, 1983).
- [2] R. Fletcher, *Practical methods of optimization, Volume 1. Unconstrained optimization* (John Wiley & Sons, Chichester-New York-Brisbane-Toronto, 1980).
- [3] R. Fletcher, *Practical methods of optimization, Volume 2. Constrained optimization* (John Wiley & Sons, Chichester-New York-Brisbane-Toronto, 1981).
- [4] K. Glashoff and S.-Å. Gustafson, *Linear optimization and approximation* (Springer-Verlag, New York-Heidelberg-Berlin, 1983).
- [5] S.-Å. Gustafson, "Stability aspects on the numerical treatment of linear semi-infinite programs", TRITA-NA-7604, Dept. of Numerical Analysis and Computing Science, The Royal Institute of Technology, S-10044 Stockholm 70, Sweden.
- [6] S.-Å. Gustafson, "A general three-phase algorithm for nonlinear semi-infinite programs", in *Operations Research '81* (ed. J. P. Brans) (North-Holland, Amsterdam-New York-Oxford, 1981), 343–356.
- [7] S.-Å. Gustafson, "A three-phase algorithm for semi-infinite programs", pp 138–157 in [1].
- [8] S.-Å. Gustafson and K. O. Kortanek, "Semi-infinite programming and applications", in *Mathematical programming. The state of the art. Bonn 1982* (eds. A. Bachem, M. Grötschel and K. Korte) (Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, 1983), 132–157.
- [9] R. Hettich (ed.), "Semi-infinite programming", *Lecture notes in control and information sciences No 15* (Springer-Verlag, Berlin-Heidelberg-New York, 1979).
- [10] D. Luenberger, *Optimization by vector space methods* (John Wiley & Sons, New York-London-Sydney-Toronto, 1969).
- [11] R. T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, 1970).
- [12] G. A. Watson, "Globally convergent methods for semi-infinite programming", *BIT* 21 (1981), 362–373.