GLOBAL OPTIMIZATION VIA SLACK VARIABLES

Isaac Siwale

Technical Report RD-16-2013
Apex Research Limited
London
England

e-mail: ike_siwale@hotmail.com

Abstract

This paper presents a method for finding global optima to constrained nonlinear programs via slack variables. The method only applies if all functions involved are of class C^1 but without any further qualification on the types of constraints allowed; it proceeds by reformulating the given program into a bi-objective program that is then solved for the Nash equilibrium. A numerical example is included to demonstrate its efficacy.

Key Words: Global Optimization, Nonlinear Programming, Slack variables, Multi-objective programming, Nash Equilibrium.

1 Introduction

A companion paper [14] presented a relatively "unfettered" method for seeking the global solution to a generic program defined on the set $\mathbf{X}_1 \equiv \{\mathbf{x} \mid \mathbf{C}(\mathbf{x}) \geq 0\}$, where the vector $\mathbf{C}(\mathbf{x})$ is, in general, a nonlinear mapping from $\mathbf{R}^n$ to $\mathbf{R}^m$, viz.:

MP₁: $$\operatorname*{Min}_{\mathbf{x}} \left\{ f(\mathbf{x}) \mid \mathbf{x} \in \mathbf{X}_1 \subset \mathbf{R}^n \right\}$$

The purpose of this paper is to contribute yet another global optimization scheme; the method—which is based on slack variables—is less general than that in [14] in the sense that it only applies if all the functions involved are of differentiability class C^1 (at least); however, no constraint qualifications checks are necessary. The paper is organised as follows: §**2** presents some preliminaries required later in the exposition; §**3** presents the conceptual foundation of the proposed method and §**4** outlines the requisite computational algorithm; a numerical example is in §**5**; §**6** summarises and concludes the presentation; last but not least, the legal framework governing this publication is set forth in §**7**.

2 Preliminaries

Let $\boldsymbol{\lambda} \in \mathbf{R}^m$ be a vector of non-negative multipliers introduced into MP_1 to create the Lagrange function $L(\mathbf{x}, \boldsymbol{\lambda}) \equiv f(\mathbf{x}) - \langle \boldsymbol{\lambda}, \mathbf{C}(\mathbf{x}) \rangle$; at a critical point $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ of the Lagrangian in which $\mathbf{x}^*$ is solution to MP_1, the Karush-Kuhn-Tucker (KKT) theorem asserts the following conditions as necessary [5]:

$$\nabla f(\mathbf{x}^*) - \nabla \langle \boldsymbol{\lambda}^*, \mathbf{C}(\mathbf{x}^*) \rangle = \mathbf{0} \tag{1a}$$

$$\langle \boldsymbol{\lambda}^*, \mathbf{C}(\mathbf{x}^*) \rangle = 0 \tag{1b}$$

$$\mathbf{C}(\mathbf{x}^*) \geq \mathbf{0} \tag{1c}$$

$$\boldsymbol{\lambda}^* \geq \mathbf{0} \tag{1d}$$

And if all the functions in MP_1 are convex on $\mathbf{X}_1$, then, subject to a qualification on the constraints, the conditions (1) are also sufficient. In deriving the optimality conditions above, Kuhn and Tucker used an auxiliary saddle-value problem based on the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda})$—this is restated below for ease of reference later; the scheme in [14] for determining the global solution to MP_1 proceeds by solving the saddle-value problem directly, using a saddle-point theorem whose proof employs a 'non-negative variable lemma'—this is also restated below; and the method proffered by this paper employs the 'variable endogenization' technique that is described in [13, 16]—its use shall be illustrated by example.

- **THE SADDLE-VALUE PROBLEM:** Given MP1 in which $\mathbf{C}(\mathbf{x}) : \mathbf{R}^n \to \mathbf{R}^m$ is a vector of (generally) non-linear functions and $\boldsymbol{\lambda} \in \mathbf{R}^m$ is a vector of non-negative multipliers, the original Kuhn-Tucker saddle-value auxiliary problem is to find the pair $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ that results in a saddle value of the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda})$, i.e. the point $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ which is such that, for all $\mathbf{x} \in \mathbf{X}_1$ and $\boldsymbol{\lambda} \geq \mathbf{0}$, the following holds in the case of *minimization*:

$$L(\mathbf{x}^*, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*) \tag{2a}$$

Or the following holds in the case of *maximization*:

$$L(\mathbf{x}^*, \boldsymbol{\lambda}) \geq L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \geq L(\mathbf{x}, \boldsymbol{\lambda}^*) \tag{2b}$$

- **LEMMA 1** [Non-negative Variables]: Given a continuously differentiable function $f(\omega)$ of a *non-negative* variable ω, the following relations are true at all points where there is an inflection or *minimum* point of $f(\omega)$:

Case I: $\omega \geq 0\,;$ $\quad \partial f/\partial\omega \geq 0\,;$ $\quad \langle \omega, \partial f/\partial\omega \rangle = 0$ (3a)

And the following relations are true at all points where there is an inflection or *maximum* point of $f(\omega)$:

Case II: $\omega \geq 0\,;$ $\quad \partial f/\partial\omega \leq 0\,;$ $\quad \langle \omega, \partial f/\partial\omega \rangle = 0$ (3b)

- **PROOF:** See [16] ▮

3 Global Optimization: The Slack Variable Approach

Let $\boldsymbol{\mu} \in \mathrm{M} \subset \mathbf{R}^p_+ + \{\mathbf{0}\}$ denote a p-vector of non-negative slack variables that are introduced into p of the m inequality constraints in MP1 and for the sake of generality, assume $p < m$. The introduction of $\boldsymbol{\mu}$ normally entails an expansion of the decision vector by p and yields an augmented, equality-constrained program defined on the set $\mathbf{X}_2 \equiv \{(\mathbf{x}, \boldsymbol{\mu}) \mid \mathbf{C}(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{0}\}$ in which the constraints vector $\mathbf{C}$ now maps $\mathbf{R}^{n+p}$ into $\mathbf{R}^m$, viz.:

MP2: $\quad \underset{\mathbf{x},\boldsymbol{\mu}}{\mathrm{Min}} \left\{ f(\mathbf{x}) \mid (\mathbf{x}, \boldsymbol{\mu}) \in \mathbf{X}_2 \subset \mathbf{R}^{n+p} \right\}$

But if $\boldsymbol{\mu}$ is added using the variable endogenization technique described in [13, 16] (see also §5 below), then the dimension of the decision vector remains the same; the endogenization process merely changes the composition of the decision vector—each element x_i that is "internalized" by the process is replaced by a new element μ_i of the slack vector. The said process also introduces a new endogenous variable $\mathbf{z}$ that is defined by a map $\mathbf{D}$, and it effectively induces a "new" feasible set $\mathbf{X}_3 \subset \mathbf{R}^n$ that is circumscribed by a subset of size $(m - p)$ of the original m constraints in MP1. If we let $\mathbf{v}_{-p}$ denote a vector $\mathbf{v}$ whose dimension has decreased by p, then $\mathbf{z}$ and $\mathbf{X}_3$ may formally be stated thus:

$$\boldsymbol{\mu} \in \mathrm{M} \subset \mathbf{R}^p_+ \cup \{\mathbf{0}\} \tag{4a}$$

$$\mathbf{z} \equiv \mathbf{D}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \in \mathbf{Z} \subset \mathbf{R}^p \tag{4b}$$

$$\mathbf{X}_3 \equiv \{(\mathbf{x}_{-p}, \boldsymbol{\mu}) \mid (\mathbf{C}_{-p}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \geq \mathbf{0})\} \tag{4c}$$

And the effective search space for MP2 is thus:

$$\mathbf{X}_4 \equiv \{(\mathbf{x}_{-p}, \boldsymbol{\mu}) \mid (\mathbf{C}_{-p}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \geq \mathbf{0}) \wedge (\mathbf{z} \equiv \mathbf{D}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \in \mathbf{Z}) \wedge (\boldsymbol{\mu} \in \mathrm{M} \subset \mathbf{R}^p_+ \cup \{\mathbf{0}\})\} \subset \mathbf{R}^n \tag{5}$$

Following the introduction of $\boldsymbol{\mu}$—even in cases where this is done without 'variable endogenization'—the objective function f effectively becomes an implicit function of the slack variables. One may appreciate this by considering the variation in f arising from a variation in the k-th slack variable; by Taylor's theorem and the chain rule of differentiation, a first-order approximation of the said variation is:

$$\delta f = \sum_i \left(\frac{\partial f}{\partial x_i} \cdot \frac{\partial x_i}{\partial \mu_k} \right) \cdot \delta\mu_k + \sum_m \sum_n \sum_k \left(\frac{\partial f}{\partial x_m} \cdot \frac{\partial x_m}{\partial x_n} \cdot \frac{\partial x_n}{\partial \mu_k} \right) \cdot \delta\mu_k \tag{6}$$

The variation in (6) has two components: (i) a "direct gradient" effect (represented by the first term) that is transmitted directly via variables (indexed by *i*) that are common to *f* and the *k*-th constraint; and (ii) an "indirect gradient" effect that transmits changes in the *k*-th constraint to *f* via the other variables and equations indexed by *m* and *n* in the constraint system. The method presented below assumes problems in which at least one of these transmission mechanisms is present.

Accordingly, the search space for MP_2 needs to be circumscribed even further because the optimal solution on $\mathbf{X}_4$ is required to meet the optimality conditions asserted by **Lemma 1**, purely by virtue of $\boldsymbol{\mu}$ being non-negative. To that end, a practical way forward is to create a conjunction of $\mathbf{X}_4$ and part of the criteria in **Lemma 1**; if we let $\mathbf{G} : \mathbf{R}^n \rightarrow \mathbf{R}^p$ denote the gradient of *f* with respect to $\boldsymbol{\mu}$, then a more appropriate feasible set for MP_2 is as follows:[1]

$$\mathbf{X}_5 \equiv \{(\mathbf{x}_{-p}, \boldsymbol{\mu}) \mid (\mathbf{C}_{-p}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \geq \mathbf{0}) \wedge (\mathbf{z} \equiv \mathbf{D}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \in \mathbf{Z} \subset \mathbf{R}^p) \wedge (\mathbf{G}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \geq \mathbf{0}) \wedge (\boldsymbol{\mu} \in \mathrm{M} \subset \mathbf{R}^p_+ \cup \{\mathbf{0}\})\} \subset \mathbf{R}^n \tag{7}$$

Thus, a final reformulation of the original M-program in terms of the slack vector $\boldsymbol{\mu}$, the endogenous vector $\mathbf{z}$ and the reduced decision vector $\mathbf{x}_{-p}$ may simply be stated as:

MP_3: $$\underset{\mathbf{x},\boldsymbol{\mu}}{\mathrm{Min}} \left\{ f(\mathbf{x}_{-p}, \boldsymbol{\mu}) \mid (\mathbf{x}_{-p}, \boldsymbol{\mu}) \in \mathbf{X}_5 \subset \mathbf{R}^n \right\}$$

The decentralized-cum-coordination scheme for solving MP_1 presented in [14] may also be applied to MP_3. This is so because of an analogy that one may draw between the KKT criteria for MP_1, and the optimality criteria comprising **Lemma 1**; the said analogy allows one to formulate, as in [14], an auxiliary saddle-value problem and prove a sufficiency theorem upon which to base the decentralized algorithm—the supporting argument proceeds as follows.

- Assume a slack *p*-vector has been introduced by endogenization and consider a "Lagrangian" of the form:

  $$S(\mathbf{x}_{-p}, \boldsymbol{\mu}) \equiv f(\mathbf{x}_{-p}, \boldsymbol{\mu}) - \langle \boldsymbol{\mu}, \mathbf{G}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \rangle$$

  The pseudo-Lagrangian $S(\mathbf{x}_{-p}, \boldsymbol{\mu})$ shall hereafter be called the 'Slack Saddle Function' (SSF), for want of a better term. The inspiration for proposing the SSF comes from an analogy that one may draw between the KKT optimality conditions pertaining to MP_1, and those pertaining to the partial function $f(\cdot, \boldsymbol{\mu})$, given that $\boldsymbol{\mu}$ is non-negative and therefore **Lemma 1** applies, viz.:

  $$\boldsymbol{\lambda} \geq 0 \quad \Leftrightarrow \quad \boldsymbol{\mu} \geq 0 \tag{8a}$$

  $$\mathbf{C}(\mathbf{x}) \geq 0 \quad \Leftrightarrow \quad \mathbf{G}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \geq 0 \tag{8b}$$

  $$\langle \boldsymbol{\lambda}, \mathbf{C}(\mathbf{x}) \rangle = 0 \quad \Leftrightarrow \quad \langle \boldsymbol{\mu}, \mathbf{G}(\mathbf{x}_{-p}, \boldsymbol{\mu}) \rangle = 0 \tag{8c}$$

- By appending $\boldsymbol{\mu}$ to the reduced vector $\mathbf{x}_{-p}$, one may define a new decision vector $\boldsymbol{\omega}$, viz.:

  $$\boldsymbol{\omega} \in \boldsymbol{\Omega} \subset \mathbf{R}^n \text{ with } \boldsymbol{\omega}' \equiv (\mathbf{x}_{-p}' \mid \boldsymbol{\mu}')$$

  And at this point, a useful comprehension aid is to notionally separate the $\boldsymbol{\mu}$ that appears inside the functions *f* and $\mathbf{G}$ from the one that is outside—the "copy" of $\boldsymbol{\mu}$ outside of *f* and $\mathbf{G}$ may be viewed as playing a role similar to the normal Lagrange multiplier $\boldsymbol{\lambda}$, and the "copy" inside *f* and $\mathbf{G}$ may be regarded simply as a component of the decision vector $\boldsymbol{\omega}$. With this devise and $\boldsymbol{\omega}$ thus defined, the relations on the right-hand side in (8) assume an identical structure to the standard KKT conditions on the left-hand side, and naturally one is lead to pose the following saddle point problem and propose a sufficiency theorem similar to that in [14] which is based on the normal Lagrangian, viz.:

[1] The orthogonality criterion $\langle \boldsymbol{\mu}, \mathbf{G} \rangle = 0$ has been left out for now with the implicit assumption that this will be satisfied at the solution since it comprises the coordinating mechanism (see §4).

- **THE SSF SADDLE-VALUE PROBLEM:** Consider the program MP3. Given the gradient $\mathbf{G}(\omega)$ of f with respect to μ, and assuming ω is feasible,[2] find the pair (ω^*, μ^*) that results in a saddle-value for $S(\omega, \mu)$ for all $\omega \in \mathbf{\Omega}$, and $\mu \in \mathrm{M}$, i.e.:

$$S(\omega^*, \mu) \;\le\; S(\omega^*, \mu^*) \;\le\; S(\omega, \mu^*) \tag{9}$$

- **THEOREM 1**: If the pair (ω^*, μ^*) is a solution to the saddle-value problem in (9), then ω^* is the global solution of MP3.

- **PROOF**: The style of proof is similar to that in the true Lagrangian case in [14]—it also merely relies on the properties of the inner product function and proceeds as follows.

  Let μ^* denote a critical point of the SSF and consider the left-hand inequality in the saddle-value problem of (9); that inequality implies a maximization process of $S(\omega^*, \mu)$ with respect to $\mu \in \mathrm{M}$ for a quasi-constant ω^* (whose character is yet to be determined), and the said relation may be simplified as follows:

$$\begin{aligned} S(\omega^*, \mu) &\le S(\omega^*, \mu^*) && \text{(10a)}\\ f(\omega^*) - \langle \mu, \mathbf{G}(\omega^*)\rangle &\le f(\omega^*) - \langle \mu^*, \mathbf{G}(\omega^*)\rangle && \text{(10b)}\\ -\langle \mu, \mathbf{G}(\omega^*)\rangle &\le -\langle \mu^*, \mathbf{G}(\omega^*)\rangle && \text{(10c)}\\ \langle \mu, \mathbf{G}(\omega^*)\rangle &\ge 0 && \text{(10d)} \end{aligned}$$

  The right-hand side of (10d) follows directly from the complementary slackness condition in (8c), since μ^* is a critical point by assumption. But if (10d) is to hold for all μ, then the inner product on the left-hand side has to be non-negative; and since $\mu \ge \mathbf{0}$ by design, we therefore require conditions on ω^* that ensure that $\mathbf{G}(\omega^*) \ge \mathbf{0}$; the said conditions obtain if and only if ω^* is an element of the collection of points that satisfy LEMMA 1, and this of course is consistent with the quest for a solution to the SSF saddle-value problem. Furthermore, since the term $\langle \mu, \mathbf{G}(\omega^*)\rangle$ contributes negatively to the value of $S(\omega^*, \mu)$ for all ω^* such that $\mathbf{G}(\omega^*) \ge \mathbf{0}$, it follows that the assumed critical point μ^* is actually the maximizer of the SSF because, by (8c), it is only at $\mu = \mu^*$ when the said term subtracts nothing from $S(\omega^*, \mu)$.

  Now consider the right-hand inequality in (9); that inequality implies a minimization process of the saddle function $S(\omega^*, \mu)$ with respect to $\omega \in \mathbf{\Omega}$ for a fixed μ^* and simplifies as follows:

$$\begin{aligned} S(\omega^*, \mu^*) &\le S(\omega, \mu^*) && \text{(11a)}\\ f(\omega^*) - \langle \mu^*, \mathbf{G}(\omega^*)\rangle &\le f(\omega) - \langle \mu^*, \mathbf{G}(\omega)\rangle && \text{(11b)}\\ f(\omega^*) - f(\omega) &\le -\langle \mu^*, \mathbf{G}(\omega)\rangle && \text{(11c)}\\ f(\omega) - f(\omega^*) &\ge \langle \mu^*, \mathbf{G}(\omega)\rangle && \text{(11d)} \end{aligned}$$

  But for all vectors ω that potentially satisfy LEMMA 1, we have that $\mathbf{G}(\omega) \ge \mathbf{0}$ and therefore $\langle \mu^*, \mathbf{G}(\omega)\rangle \ge 0$, since $\mu^* \ge \mathbf{0}$ by definition. Consequently, if the inequality in (11d) is to hold for all $\omega \in \mathbf{\Omega}$, then $f(\omega^*)$ must evaluate onto the left side of $f(\omega)$ on the real line, and this criterion is fulfilled if and only if ω^* is the *global* minimizer of f on $\mathbf{\Omega}$ ▮

Closing Remarks. It's worth reiterating that the method described here is less general than its "true-Lagrangian" counterpart described in [14], in the sense that it only applies when all functions involved are continuously differentiable; in addition, it is also not always easy to implement in the absence of an automatic method for evaluating the gradient $\mathbf{G}(\omega)$. However, no constraint qualification tests are necessary, and the same decentralisation-cum-coordination numerical technique presented in [14] to solve MP1 to global optimality may also be used for the SSF case as well. This is explained next.

[2] In terms of items defined previously, feasibility implies $(\mathbf{C}_{-p}(\omega) \ge \mathbf{0}) \wedge (\mathbf{D}(\omega) \in \mathbf{Z})$

4 Computation: The Decentralization-cum-Coordination Method

Theorem 1 affords a method for computing the global solution to MP1 via MP3. Although the method only applies when all functions involved are of differentiability class C^1, there is no need for a qualification on the type of constraints allowed. For simplicity of notation, assume the vector **x** is already in reduced form; then the computational method proposed may be summarised thus:

Step 1: Given MP1, introduce slack variables, preferably by endogenization (if at all possible), to produce MP3

Step 2: Evaluate the gradient $\mathbf{G}(\boldsymbol{\omega})$ and the 'Slack Saddle Function' $S(\boldsymbol{\omega}, \boldsymbol{\mu}^*)$; state the requisite saddle-value problem in accordance with **Lemma 1** and "unpack" it into its two constituent inequalities, viz.:

$$S(\boldsymbol{\omega}^*, \boldsymbol{\mu}) \le S(\boldsymbol{\omega}^*, \boldsymbol{\mu}^*) \le S(\boldsymbol{\omega}, \boldsymbol{\mu}^*) \Leftrightarrow \{S(\boldsymbol{\omega}^*, \boldsymbol{\mu}) \le S(\boldsymbol{\omega}^*, \boldsymbol{\mu}^*)\} \wedge \{S(\boldsymbol{\omega}^*, \boldsymbol{\mu}^*) \le S(\boldsymbol{\omega}, \boldsymbol{\mu}^*)\} \tag{12}$$

Step 3: Restate the two constituent inequalities in (12) as optimization problems, viz.:

$$S(\boldsymbol{\omega}^*, \boldsymbol{\mu}) \le S(\boldsymbol{\omega}^*, \boldsymbol{\mu}^*) \Leftrightarrow \text{Max } \{S(\boldsymbol{\omega}^*, \boldsymbol{\mu})\} \text{ with respect to } \boldsymbol{\mu} \tag{13a}$$

$$S(\boldsymbol{\omega}^*, \boldsymbol{\mu}^*) \le S(\boldsymbol{\omega}, \boldsymbol{\mu}^*) \Leftrightarrow \text{Min } \{S(\boldsymbol{\omega}, \boldsymbol{\mu}^*)\} \text{ with respect to } \boldsymbol{\omega} \tag{13b}$$

Step 4: Introduce constraints on the decision variables in (13) in accordance with MP3 and **Lemma 1**, viz.:

$$\underset{\boldsymbol{\mu}}{\text{Max}} \left\{ S(\boldsymbol{\omega}^*, \boldsymbol{\mu}) \,\middle|\, \boldsymbol{\mu} \in \mathrm{M} \subset \mathbf{R}^p_+ \cup \{\mathbf{0}\}; \langle \boldsymbol{\mu}, \mathrm{G}(\boldsymbol{\omega}^*) \rangle = 0 \right\} \tag{14a}$$

$$\underset{\boldsymbol{\omega}}{\text{Min}} \left\{ S(\boldsymbol{\omega}, \boldsymbol{\mu}^*) \,\middle|\, \boldsymbol{\omega} \in \boldsymbol{\Omega} \subset \mathbf{R}^n; \langle \boldsymbol{\mu}^*, \mathrm{G}(\boldsymbol{\omega}) \rangle = 0 \right\} \tag{14b}$$

Step 5: Solve the bi-objective problem in (14) using an appropriate solution concept and numerical method. Note that unlike the method of [14], the sub-program (14a) cannot be "reduced" to a search over a subset of the integers.

Multi-objective solution concepts are explained fully in [15]. And as argued in the companion paper [14, footnote **4**]: (i) the 'Nash equilibrium' is the more appropriate solution in this case; (ii) to compute the Nash equilibrium, one may use a decentralization-cum-coordination approach that is evidently suggested by the saddle-value problem itself and its optimization reformulation in (14). But numerical experience suggests that the search for the equilibrium point is likely to be erratic unless extra coordinating mechanisms are added to supplement that inherent in the Nash solution concept itself.

The Fischer-Burmeister function [3] affords one such a coordinator—it is defined by equation (15) and characterized by the relations in (16):

$$\phi(\mathrm{a}, \mathrm{b}) = \sqrt{\mathrm{a}^2 + \mathrm{b}^2} - (\mathrm{a} + \mathrm{b})\,;\ \mathrm{a} \ge 0;\ \mathrm{b} \ge 0 \tag{15}$$

$$(\mathrm{a} > 0) \wedge (\mathrm{b} = 0) \Leftrightarrow \phi(\mathrm{a}, \mathrm{b}) = 0 \tag{16a}$$

$$(\mathrm{a} = 0) \wedge (\mathrm{b} > 0) \Leftrightarrow \phi(\mathrm{a}, \mathrm{b}) = 0 \tag{16b}$$

$$(\mathrm{a} = 0) \wedge (\mathrm{b} = 0) \Leftrightarrow \phi(\mathrm{a}, \mathrm{b}) = 0 \tag{16c}$$

Note that the conjunctions on left are exactly those in **Case I** of **Lemma 1** when 'a' replaced by an element of the gradient vector **G** and 'b' replaced by the corresponding element of the slack vector $\boldsymbol{\mu}$. One way of implementing the coordination mechanism proper is to evaluate (at iteration k) each candidate solution $(\boldsymbol{\omega}^k, \boldsymbol{\mu}^k)$ using the relations in (15) and according to the criterion: 'the closer $\phi(\boldsymbol{\omega}^k, \boldsymbol{\mu}^k)$ is to zero, the better'

A second technique that helps prevent erratic algorithmic behaviour is to restrict the slack variables to closed intervals with finite endpoints—as opposed to the half-line $[0, \infty)$—using interval arithmetic methods [9]. This is relatively easy to implement when endogenous z-variables are present: for example, given the endogenous equation $z_1 = 400 - x_1 - \mu_1$ with $x_1 \in [10, 1000]$, $z_1 \in [10, 1000]$ and $\mu_1 \in [0, \infty)$, the net natural interval extension pertaining to the slack variable μ_1 is (see Step 8 and/or Step 9 in §**5**):

$$\mu_1 \in [-1600, 380] \cap [0, \infty) = [0, 380]$$

5 Numerical Example

MP4: $$\min_{\mathbf{x}} f(\mathbf{x}) = x_1 + x_2 + x_3$$

Subject to:

$$1 - 0.0025(x_4 + x_6) \geq 0 \quad \text{(i)}$$

$$1 - 0.0025(x_5 + x_7 - x_4) \geq 0 \quad \text{(ii)}$$

$$1 - 0.01(x_8 - x_5) \geq 0 \quad \text{(iii)}$$

$$x_1 x_6 - 833.33252 x_4 - 100 x_1 + 83333.333 \geq 0 \quad \text{(iv)}$$

$$x_2 x_7 - 1250 x_5 - x_2 x_4 + 1250 x_4 \geq 0 \quad \text{(v)}$$

$$x_3 x_8 - 1250000 - x_3 x_5 + 2500 x_5 \geq 0 \quad \text{(vi)}$$

$$x_1 \in [100, 10000]$$

$$x_i \in [1000, 10000], \quad i = 2, 3;$$

$$x_i \in [10, 1000], \quad i = 4, \ldots, 8$$

I. Pre-Processing: Introducing Slacks via Variable Endogenization

Step 1: Create a ‘connexion matrix’—previously known as the ‘incidence matrix’—that indicates the presence or absence of each variable per constraint, viz.:

Table 1: Original connexion matrix of the constraints

CONSTRAINT NO	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
i				✓		✓		
ii				✓	✓		✓	
iii					✓			✓
iv	✓			✓		✓		
v		✓		✓	✓		✓	
vi			✓		✓			✓

Step 2: Manipulate the connexion matrix by swapping columns and / or rows so as to form an echelon of rows.[3] For the variables that constitute the slopping edge of the echelon (indicated by the red tick ✓ in Table 2 below), it is imperative that the equations that follow from the introduction of slacks are easily manipulated to make the “edge variables” the subjects of the said equations. Introduce sequentially numbered z-variables into the matrix—which will form the subjects of the endogenous equations—in such a way as to avoid circular definitions, viz.:

Table 2: Connexion matrix in ‘row echelon’ form

CONSTRAINT NO	x_4	x_5	x_6	x_7	x_8	x_1	x_2	x_3	
i	✓		✓						z_1
ii	✓	✓		✓					z_2
iii		✓			✓				z_3
iv	✓		✓			✓			z_4
v	✓	✓		✓			✓		z_5
vi		✓			✓			✓	z_6

[3] Note that such manipulations of the connexion matrix do not alter the search space at all: swapping rows corresponds to changing the order in which the constraints are presented; and swapping columns corresponds to changing the sequence of terms in each affected function.

Step 3: Into the "edge constraints" (which in this particular case just happens to be all the original constraints), introduce the slack variables μ_1, μ_2, μ_3, μ_4, μ_5 and μ_6 respectively and define endogenous z-equations using the "edge variables", viz.:

$$z_1 \equiv x_6 = 400 - x_4 - \mu_1 \quad (17a)$$

$$z_2 \equiv x_7 = 400 + x_4 - x_5 - \mu_2 \quad (17b)$$

$$z_3 \equiv x_8 = 100 + x_5 - \mu_3 \quad (17c)$$

$$z_4 \equiv x_1 = [\mu_4 + 833.33252 x_4 - 83333.333] / [z_1 - 100] \quad (17d)$$

$$z_5 \equiv x_2 = [\mu_5 + 1250 x_5 - 1250 x_4] / [z_2 - x_4] \quad (17e)$$

$$z_6 \equiv x_3 = [\mu_6 + 1250000 - 2500 x_5] / [z_3 - x_5] \quad (17f)$$

Step 4: For convenience, rename the variables as follows:

Table 3: Variable replacement table

OLD NAME	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Intermediate	z_4	z_5	z_6	-	-	z_1	z_2	z_3
NEW NAME	z_4	z_5	z_6	x_1	x_2	z_1	z_2	z_3

Step 5: Restate the original M-program using the "new" variables from Step 4, viz.:

Given:

$$z_1 = 400 - x_1 - \mu_1 \quad (18a)$$

$$z_2 = 400 + x_1 - x_2 - \mu_2 \quad (18b)$$

$$z_3 = 100 + x_2 - \mu_3 \quad (18c)$$

$$z_4 = [\mu_4 + 833.33252 x_1 - 83333.333] / [z_1 - 100] \quad (18d)$$

$$z_5 = [\mu_5 + 1250 x_2 - 1250 x_1] / [z_2 - x_1] \quad (18e)$$

$$z_6 = [\mu_6 + 1250000 - 2500 x_2] / [z_3 - x_2] \quad (18f)$$

$$\min_{\mathbf{x}} f(\mathbf{x}) = z_4 + z_5 + z_6$$

Subject to:

$$x_i \in [10, 1000], \ i = 1, 2;$$

$$z_i \in [10, 1000], \ i = 1, 2, 3;$$

$$z_4 \in [100, 10000];$$

$$z_i \in [1000, 10000], \ i = 5, 6.$$

Step 6: In accordance with **LEMMA 1**, the gradient vector $\partial f / \partial \mu \equiv g(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu})$ may be used to place further constraints on the evolution of the slack variable vector $\boldsymbol{\mu}$, and in this case the gradient vector is easy to derive, viz.:

$$g_1 \equiv \partial f / \partial \mu_1 = (\mu_4 + 833.33252 x_1 - 83333.333) / (z_1 - 100)^2 \quad (19a)$$

$$g_2 \equiv \partial f / \partial \mu_2 = (\mu_5 + 1250 x_2 - 1250 x_1) / (z_2 - x_1)^2 \quad (19b)$$

$$g_3 \equiv \partial f / \partial \mu_3 = (\mu_6 + 1250000 - 2500 x_2) / (z_3 - x_2)^2 \quad (19c)$$

$$g_4 \equiv \partial f / \partial \mu_4 = 1 / (z_1 - 100) \quad (19a)$$

$$g_5 \equiv \partial f / \partial \mu_5 = 1 / (z_2 - x_1) \quad (19b)$$

$$g_6 \equiv \partial f / \partial \mu_6 = 1 / (z_3 - x_2) \quad (19c)$$

Step 7: Since the original program is a *minimization* problem, the gradient is required to be non-negative in accordance with **Case I** of **Lemma 1**; and ignoring the complementary slackness condition for now (see footnote 1), the final program may be stated thus:

Given:

$$z_1 = 400 - x_1 - \mu_1$$

$$z_2 = 400 + x_1 - x_2 - \mu_2$$

$$z_3 = 100 + x_2 - \mu_3$$

$$z_4 = [\mu_4 + 833.33252 x_1 - 83333.333]/[z_1 - 100]$$

$$z_5 = [\mu_5 + 1250 x_2 - 1250 x_1]/[z_2 - x_1]$$

$$z_6 = [\mu_6 + 1250000 - 2500 x_2]/[z_3 - x_2]$$

$$\underset{\mathbf{x}}{\text{Min}}\ f(\mathbf{x}) = z_4 + z_5 + z_6$$

Subject to: $g(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}) \geq 0$;

$$x_i \in [10, 1000], \quad i = 1, 2;$$

$$z_i \in [10, 1000],\ i = 1, 2, 3;$$

$$z_4 \in [100, 10000];$$

$$z_i \in [1000, 10000],\ i = 5, 6$$

$$\mu \in \mathbf{R}_+$$

Step 8: Append the vector $\boldsymbol{\mu}$ to the remaining (i.e. "un-internalized") components of the original decision vector $\mathbf{x}$ to create a new vector, (say) $\boldsymbol{\omega}$:

$$\boldsymbol{\omega}' = (x_1, x_2, \mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6)$$

Use interval arithmetic methods to compute proper set-bounds for at least some of the slack variables using the endogenous z-equations. For example, from equation (18a), we have that:

$$\mu_1 = 400 - (x_1 + z_1) \tag{20a}$$

$$x_1 \in [10, 1000] \tag{20b}$$

$$z_1 \in [10, 1000] \tag{20c}$$

$$\mu_1 \in [0, \infty) \tag{20d}$$

By elementary interval arithmetic [9], it can shown that the natural interval extension of the right-hand side of (20a) evaluates to the set [–1600, 380]. Therefore, if the equality in (20a) is to hold under all circumstances, then we must have that:

$$\mu_1 \in [-1600, 380] \cap [0, \infty) = [0, 380] \tag{20e}$$

An alternative or additional approach—which is only possible where the solution algorithm is such that components of $\boldsymbol{\omega}$ are determined sequentially—is to code the bounds in a dynamic fashion as follows.

Step 9: Let [v] denote the interval pertaining to the variable, 'v'; then from equation (18a), we have that:

$$[a, b] = 400 - (x_1 + [z_1]) \tag{21a}$$

$$x_1 = \text{a "known" scalar} \tag{21b}$$

$$[z_1] = [10, 1000] \tag{21c}$$

$$\mu_1 \in [a, b] \cap [0, \infty) \tag{21d}$$

Alternatively, since (20e) accounts for all possible values of x_1, instead of (21d) one could also write:

$$\mu_1 \in [a, b] \cap [0, 380] \tag{21e}$$

II. GENO Output

Generation Number	Time[4] (sec)	Objective 1	Objective 2
0	0.0000	-3138.60174223	3138.60174223
10	1.3590	-7049.15048902	7049.15048902
20	1.3600	-7049.24802053	7049.24802053
30	1.3430	-7049.24802053	7049.24802053
40	1.3440	-7049.24802053	7049.24802053
50	1.3440	-7049.24802053	7049.24802053
60	1.3280	-7049.24802053	7049.24802053
70	1.3750	-7049.24802053	7049.24802053
80	1.3590	-7049.24802053	7049.24802053
90	1.3750	-7049.24802053	7049.24802053
100	1.3130	-7049.24802053	7049.24802053

Optimal (Reduced) Decision Vector: $\mathbf{x}^* = (182.01769976, 295.60117370)^T$

Optimal Slack Variable Vector: $\boldsymbol{\mu}^* = (0.00000, 0.00000, 0.00000, 0.00000, 0.00000, 0.00000)^T$

Optimal Endogenous Vector: $\mathbf{z}^* = (217.9823005, 286.4165263, 395.6011731, 579.306683, 1359.9706661, 5109.9706714)^T$

Optimal Objective Function Value: $f(\boldsymbol{\omega}^*) = 7049.24802053$

Average execution time per 10 generations: 1.3500 seconds

Overall execution time on 100 generations: 13.5000 seconds

Approximate time to first optimum (8th decimal place accuracy): 2.7000 seconds

III. General Remarks

This is a real engineering problem whose aim is to optimize the performance of a heat exchanger. The problem was first tackled using the geometric programming method by Avriel & Williams [1], and it has been subjected to various other optimization techniques ever since—two of the latest efforts being Azad & Fernandes [2], and Pinter [10]. It features in the comparative study in [7], as well as in [7] and [11], and the best known solution up until then (i.e., *circa*, 2001) remained that reported by Hock & Schittkowski [4], namely,

$f(\mathbf{x}^*) = 7049.330923$, which is located at:

$\mathbf{x}^* = (579.3167, 1359.943, 5110.071, 182.0174, 295.5985, 217.9799, 286.4162, 395.5979)^T$

But this has since been superseded: at $f(\boldsymbol{\omega}^*) = 7049.24802053$, the GENO solution reported above—which was first found by an earlier version of the solver using a different solution method [12]—is better than the Hock-Schittkowski solution; all six constraints are active at the solution (as evidenced by the value of $\boldsymbol{\mu}^*$). Azad & Fernandes [2] report a solution of similar quality, at least up to the third decimal place; but the 'Lipschitz Global Optimizer' [10] returns a solution of lower quality valued 7049.720652.

6 Summary and Conclusions

This paper has presented a new method for finding global optima to constrained nonlinear programs; the method is akin to one reported earlier but less general in the sense that it only applies when all the functions involved are of class C^1 (at least). The method is anchored on a sufficiency theorem derived from a pseudo-Lagrangian saddle-value problem associated with the given nonlinear program; it involves a reformulation the given program into a bi-objective program that is then solved for the Nash equilibrium.

A real practical engineering problem concerning the optimal operating conditions of a heat exchanger has been presented to illustrate the efficacy of the method; its numerical solution by the solver GENO is the best known so far—it provides a benchmark against which other algorithms may be assessed.

[4] The execution times pertain to a C++ version of GENO running under Windows 8.1 on a Laptop machine with the following hardware specs: AMD A4-5000 APU Processor, 1.5GHz, 4GB RAM. The mating population was of size 30.

7 Legalities

I. Licence and Trademarks

II. Disclaimer

This document contains proprietary material created by Apex Research Ltd which is subject to further verification and change without notice; however, Apex Research Ltd is under no obligation to provide an updated version. Furthermore, Apex Research Ltd does not make any expressed or implied warranty— including the warranties of merchantability and fitness for a particular purpose—as to the accuracy or completeness of the methods described herein; accordingly, Apex Research Ltd accepts no liability for any damages that may occur from use.

[5] http://creativecommons.org/licenses/by-nc-sa/4.0/

[6] http://creativecommons.org/licenses/by-nc-sa/4.0/legalcode

References

1. Avriel, M., Williams, A. C.: An Extension of Geometric Programming with applications in engineering optimization. *Journal of Engineering Mathematics*, **5**, pp. 187-194 (1971).

2. Azad, Md. K. A., Fernandes, M. G. P.: Modified constrained differential evolution for solving nonlinear global optimization problems. In: *Computational Intelligence*, pp. 85-100. Springer-Verlag, Berlin (2013).

3. Fischer, A.: A special Newton-type optimization method. *Optimization*, **24**, 269-284 (1992).

4. Hock, W., Schittkowski, K.: *Test Examples for Non-linear Programming Codes*, Lecture Notes in Economics and Mathematical Systems, **187**, Spring-Verlag, Berlin (1981).

5. Kuhn, H. W., Tucker, A. W.: Nonlinear programming. In: Neyman, J. (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 481–492. University of California Press, Berkeley (1951).

6. Michalewicz, Z., Attia, N.: Evolutionary optimization of constrained Problems. In: Sebald, A.V., Fogel, L.J. (eds.). *Proceedings of the 3rd Annual Conference on Evolutionary Programming*. World Scientific, River Edge (1994).

7. Michalewicz, Z., Fogel, D. B.: *How to Solve It. Modern Heuristics*. Springer-Verlag, Berlin (2000).

8. Michalewicz, Z.: Genetic algorithms, numerical optimization and constraints. In: *Proceedings of the 6th International Conference on Genetic Algorithms*, pp. 151-158. Pittsburg, (1995).

9. Moore, R. E., Kearfott, R. B., Cloud, M. J.: *Introduction to Interval Analysis*. SIAM, Philadelphia, (2009).

10. Pinter, J.: How difficult is nonlinear optimization? A practical solver tuning approach with illustrative results. (2014). [Online] Available at: http://www.optimization-online.org

11. Runarsson, T. P., Yao, X.: Stochastic ranking for constrained evolutionary optimization. *IEEE Transactions on Evolutionary Computation*, **4**, 284–294 (2000).

12. Siwale, I.: GENO™ 1.0: Supplement to User's Manual Part I: Static and Dynamic Programs. *Technical Report No. RD-4-2005*, Apex Research Ltd, London (2005). [Online] Available at: http://www.researchgate.net

13. Siwale, I.: GENO™ 2.0: The GAUSS User's Manual. *Technical Report No. RD-13-2013*, Apex Research Ltd, London (2013). Available with a trial version of GENO—contact Aptech Systems Inc.

14. Siwale, I.: On Global Optimization. *Technical Report No. RD-23-2014*, Apex Research Ltd, London (2014). [Online] Available at: http://www.researchgate.net or at: http://www.optimization-online.org

15. Siwale, I.: Practical multi-objective programming. *Technical Report No. RD-14-2013*, Apex Research Ltd, London (2013). [Online] Available at: http://www.researchgate.net or at: http://www.optimization-online.org

16. Siwale, I.: Solution of nonlinear equation systems via optimization. *Technical Report No. RD-15-2013*, Apex Research Ltd, London (2013). [Online] Available at: http://www.researchgate.net or at: http://www.optimization-online.org