ON GLOBAL OPTIMIZATION

Isaac Siwale

Technical Report RD-23-2014
Apex Research Limited
London
England

e-mail: ike_siwale@hotmail.com

Abstract

This paper presents a relatively "unfettered" practical method for finding global optima to constrained nonlinear programs. The method reformulates the given program into a bi-objective mixed-integer program that is then solved for the Nash equilibrium. A numerical example is included to illustrate the efficacy of the method; the solution computed is a benchmark against which other algorithms may be assessed.

Key Words: Global Optimization, Nonlinear Programming, Karush-Kuhn-Tucker Theorem, Mixed-integer Programming, Nash Equilibrium.

1 Introduction

The field of nonlinear programming is said to have started in 1951 with the publication of a theorem by Harold W. Kuhn and Albert W. Tucker [11], although results similar to those comprising the theorem were subsequently shown to have been derived much earlier by William Karush in his 1939 Masters Thesis [10]. The said theorem—now commonly known as the 'Karush-Kuhn-Tucker Theorem' (or KKT Theorem for short)—asserts conditions that characterize the optimal solution to a mathematical program defined on the set $\mathbf{X}_1 \equiv \{\mathbf{x} \mid \mathbf{C}(\mathbf{x}) \geq \mathbf{0}\}$, where $\mathbf{C}(\mathbf{x}) : \mathbf{R}^n \to \mathbf{R}^m$ is a vector-valued mapping whose components are (generally) nonlinear numeric functions $c_i(\mathbf{x}) : \mathbf{R}^n \to \mathbf{R}$, $i \in \{1, 2, \ldots, m\}$, viz.:

MP_1: $\quad \underset{\mathbf{x}}{\mathrm{Min}} \left\{ f(\mathbf{x}) \mid \mathbf{x} \in \mathbf{X}_1 \subset \mathbf{R}^n \right\}$

Given MP_1, the KKT theorem is stated thus: let $\boldsymbol{\lambda} \in \mathbf{R}^m$ be a vector of non-negative Lagrange multipliers; then at the solution $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ of MP_1, the following conditions are necessary:

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

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

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

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

And if all the functions involved in MP_1 are convex then, subject to a qualification on the constraints, the conditions (1) are also sufficient.

But convexity is not always guaranteed: and although the notion of 'invexity'—originally due to Hanson [7]—substantially expands the range of functions for which the KKT conditions are necessary and sufficient, the theorem remains rather restrictive in its application since the model has to conform to the smoothness as well as the constraint qualification requirements dictated therein. From a utilitarian and practical point of view therefore, a more general 'solution-characterizing' theorem is warranted.

The purpose of this paper is to present a computational scheme anchored on a saddle-value theorem that sufficiently characterizes the *global* solution to MP_1; the said theorem does not require any convexity or smoothness assumptions, and no constraint qualifications are necessary either. The paper is organised as follows: §2 presents the saddle-value sufficiency theorem; §3 explains the proposed method for computing the global solution to MP_1; a numerical example is in §4; §5 summarises and concludes the presentation; last but not least, the legal framework governing this document is set forth in §6.

2 Characterizing the Global Solution via a Saddle-point Theorem

In the derivation of their theorem, Kuhn and Tucker used an auxiliary saddle-value problem based on the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}) \equiv f(\mathbf{x}) - \langle \boldsymbol{\lambda}, \mathbf{C}(\mathbf{x}) \rangle$, viz.:[1]

- **THE LAGRANGIAN SADDLE-VALUE PROBLEM:** Given MP_1 in which $\mathbf{C}(\mathbf{x}) : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is a vector of (generally) non-linear functions and $\boldsymbol{\lambda} \in \mathbf{R}^m$ is a vector of non-negative multipliers, 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*:

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

The computational method presented in §3 is based directly on the Lagrangian saddle-value problem; it is anchored on a sufficiency theorem that connects the global solution to MP_1 to the solution of the saddle-value problem. Similar theorems may be found in Takayama [18, p.74] and Mangasarian [12, p. 74], but the exposition presented below is different and hopefully much simpler—the proof merely draws on properties of the inner product function and some well-known results in optimization which are restated as the following 'non-negative variable lemma'.

- **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: $\omega \geq 0$; $\partial f / \partial \omega \geq 0$; $\langle \omega, \partial f / \partial \omega \rangle = 0$ (3a)

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

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

- **PROOF** of **Case I**: There are three cases to consider: the first is the well-known first order optimality criteria for a local minimum located in the interior of the feasible set, viz.:

$$\omega > 0 \quad \text{and} \quad \partial f / \partial \omega = 0 \tag{4a}$$

For a solution located on the border of the feasible region, i.e. at $\omega = 0$, the optimality conditions are:

$$\omega = 0 \quad \text{and} \quad \partial f / \partial \omega = 0 \tag{4b}$$

But there is another way in which a 'solution on the border' may occur. Consider a first-order approximation (via Taylor's theorem) of the variation in f following a variation $\delta\omega$ from the point $\omega = 0$; this is given by:

$$\delta f(\omega) \cong \langle \delta\omega, \partial f / \partial \omega \rangle \tag{4c}$$

The normal definition of a variation is 'a small increase', hence $\delta\omega > 0$; and if $\partial f / \partial \omega > 0$, then the right-hand side of (4c) is non-negative since it is the inner product of two vectors in the same orthant. In other words, as one moves from $\omega = 0$ to $\omega = 0 + \delta\omega$, there is an *increase* in the value of f—i.e. $f(0)$ is a 'minimum' relative to $f(\delta\omega)$; and of course f would continue increasing in value on successive moves as long as the gradient of f remains strictly positive. Thus, the minimum point in this case is characterized by:

$$\omega = 0 \quad \text{and} \quad \partial f / \partial \omega > 0 \tag{4d}$$

[1] "Why did Kuhn and Tucker choose the saddle point formulation, and why were they looking for equivalence between [MP_1] and the saddle value problem? [Apparently] Kuhn and Tucker actually were searching for a way to extend the duality theorem for linear programming [that had just been developed by George B. Danzig] to more general cases, [and] it seems perfectly natural to take the saddle point for the Lagrange function as the starting point." [Paraphrased from 10, pp.344-345]

The inequality $\omega \geq 0$ describes all values of ω that satisfy (4a), (4b) and (4d); the ine quality $\partial f / \partial \omega \geq 0$ does the same for the gradient; and the inter-relations between the two—i.e. the conjunctions indicated by the word 'and' in (4a) and (4b)—is subsumed by the equation $\langle \omega, \partial f / \partial \omega \rangle = 0$. Thus, a complete characterization of an inflection or minimum point of $f(\omega)$ is totally summarised by the triple: $\omega \geq 0$, $\partial f / \partial \omega \geq 0$ and $\langle \omega, \partial f / \partial \omega \rangle = 0$.

Note that the criterion $\langle \omega, \partial f / \partial \omega \rangle = 0$—which is alternately known as the 'complementary slackness' or 'complementarity' or 'orthogonality' condition—only arises because the variable ω is restricted to the half-line $[0, \infty)$, hence the descriptor 'Non-negative Variable lemma'; 'strict complementarity' means that one and only one of the operands of the inner product is zero at any instance; but in this case, the complementarity is assumed to be non-strict in order to accommodate (4b) ∎

PROOF of **Case II**: The same argument applies but with the order relations involving the gradient reversed ∎

□ **THEOREM 1**: [2] If the pair $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a solution to the Lagrangian saddle value problem of (2a), then $\mathbf{x}^*$ is a *global* solution to MP_1.

□ **PROOF**: The proof draws on two facts: (i) that the Lagrangian of MP_1 is a continuously differentiable function of a non-negative variable $\boldsymbol{\lambda}$, and so **Case I** of **LEMMA 1** applies—the gradient term being $\partial L / \partial \boldsymbol{\lambda} = \mathbf{C}(\mathbf{x})$; (ii) that the inner product function is non-negative whenever its operands are in the same orthant, and that if one of the vectors is in the orthant's interior and the other is non-zero, then the inner product is strictly positive[3]

Let $\boldsymbol{\lambda}^*$ denote a critical point of the Lagrangian, and consider the left-hand inequality in the saddle-value problem of (2a); that inequality implies a maximization process with respect to $\boldsymbol{\lambda}$ for a fixed $\mathbf{x}^*$ (whose character is yet to be determined), and the inequality may be simplified as follows:

$$L(\mathbf{x}^*, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \tag{5a}$$

$$f(\mathbf{x}^*) - \langle \boldsymbol{\lambda}, \mathbf{C}(\mathbf{x}^*) \rangle \leq f(\mathbf{x}^*) - \langle \boldsymbol{\lambda}^*, \mathbf{C}(\mathbf{x}^*) \rangle \tag{5b}$$

$$- \langle \boldsymbol{\lambda}, \mathbf{C}(\mathbf{x}^*) \rangle \leq - \langle \boldsymbol{\lambda}^*, \mathbf{C}(\mathbf{x}^*) \rangle \tag{5c}$$

$$\langle \boldsymbol{\lambda}, \mathbf{C}(\mathbf{x}^*) \rangle \geq 0 \tag{5d}$$

The right-hand side of (5d) follows directly from the orthogonality condition in (3a) of **LEMMA 1**, since $\boldsymbol{\lambda}^*$ is a critical point of the Lagrangian by assumption. If the inequality in (5d) is to hold, then the inner product on the left-hand side must be non-negative; and since $\boldsymbol{\lambda} \geq \mathbf{0}$ by design, we therefore require conditions that ensure that $\mathbf{C}(\mathbf{x}^*) \geq \mathbf{0}$—the said conditions obtain if and only if $\mathbf{x}^*$ is an element of the feasible set. Furthermore, since the term $\langle \boldsymbol{\lambda}, \mathbf{C}(\mathbf{x}^*) \rangle$ contributes negatively to the value of the Lagrangian, it follows that the assumed critical point $\boldsymbol{\lambda}^*$ is actually a maximizer because it is when $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ that the said term subtracts nothing from $L(\mathbf{x}^*, \boldsymbol{\lambda})$.

Now consider the right-hand inequality in (2a); that inequality implies a minimization process over the set $\mathbf{X}_1$ for a fixed $\boldsymbol{\lambda}^*$, and the relation may be simplified as follows:

$$L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*) \tag{6a}$$

$$f(\mathbf{x}^*) - \langle \boldsymbol{\lambda}^*, \mathbf{C}(\mathbf{x}^*) \rangle \leq f(\mathbf{x}) - \langle \boldsymbol{\lambda}^*, \mathbf{C}(\mathbf{x}) \rangle \tag{6b}$$

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \langle \boldsymbol{\lambda}^*, \mathbf{C}(\mathbf{x}) \rangle \tag{6c}$$

But for all feasible $\mathbf{x}$, $\mathbf{C}(\mathbf{x}) \geq \mathbf{0}$ and therefore $\langle \boldsymbol{\lambda}^*, \mathbf{C}(\mathbf{x}) \rangle \geq 0$ because $\boldsymbol{\lambda}^* \geq \mathbf{0}$ by definition. The right-hand side of (6c) is therefore non-negative *for all* $\mathbf{x} \in \mathbf{X}_1$; and consequently, for the inequality to be valid, $f(\mathbf{x}^*)$ must always equal or evaluate onto the left side of $f(\mathbf{x})$ on the real line for all $\mathbf{x} \in \mathbf{X}_1$, and this criterion is fulfilled if and only if $\mathbf{x}^*$ is the *global* minimizer of f on $\mathbf{X}_1$ ∎

[2] cf. Takayama [18, p.74] and Mangasarian [12, p. 74].

[3] In expanded form, the inner product function is simply a sum of corresponding elements of its vector operands, viz.: $\langle \mathbf{a}, \mathbf{b} \rangle \equiv \sum a_i b_i$. If **a** and **b**, are in the same orthant, then corresponding non-zero elements have the same sign and their product is therefore non-negative; and the sum of non-negative products is obviously non-negative. By the same argument, the assertion regarding strict positivity is also apparent.

3 Computation: The Decentralization-cum-Coordination Method

THEOREM 1 affords an "unfettered" method for computing the global solution to MP_1 in the sense that the given program needn't conform to the smoothness and the constraint qualification requirements dictated by the KKT theorem—the method's strategy may be summarised thus:

> **Solution Strategy**. Rather than an algorithm that attempts to solve or is based on the complete KKT system in (1), one may compute the saddle-point of the Lagrangian instead; and to that end, one possible computational scheme is self-evident—it is a decentralized implementation of the optimization processes implicit in the Lagrangian saddle-value problem itself, perhaps combined with coordinating mechanisms whose purpose is to control the disaggregated solution process.

The method advocated by this paper begins with the following "meta-algorithm":

Step 1: Given MP_1, "unpack" the appropriate saddle-value problem into its two constituent inequalities, viz.:

$$L(\mathbf{x}^*, \boldsymbol{\lambda}) \le L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \le L(\mathbf{x}, \boldsymbol{\lambda}^*) \Leftrightarrow \{L(\mathbf{x}^*, \boldsymbol{\lambda}) \le L(\mathbf{x}^*, \boldsymbol{\lambda}^*)\} \wedge \{L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \le L(\mathbf{x}, \boldsymbol{\lambda}^*)\} \tag{7}$$

Step 2: Restate the two constituent inequalities in (7) as optimization problems, viz.:

$$L(\mathbf{x}^*, \boldsymbol{\lambda}) \le L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \Leftrightarrow \text{Max } \{L(\mathbf{x}^*, \boldsymbol{\lambda})\} \text{ with respect to } \boldsymbol{\lambda} \tag{8a}$$

$$L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \le L(\mathbf{x}, \boldsymbol{\lambda}^*) \Leftrightarrow \text{Min } \{L(\mathbf{x}, \boldsymbol{\lambda}^*)\} \text{ with respect to } \mathbf{x} \tag{8b}$$

Step 3: Introduce constraints on the decision variables in (8) in accordance with MP_1, **LEMMA 1** and **THEOREM 1**, viz.:

$$\underset{\boldsymbol{\lambda}}{\text{Max}} \left\{ L(\mathbf{x}^*, \boldsymbol{\lambda}) \,\middle|\, \boldsymbol{\lambda} \in \Lambda \subseteq \mathbf{R}_+^m ; \langle \boldsymbol{\lambda}, \mathrm{C}(\mathbf{x}^*) \rangle = 0 \right\} \tag{9a}$$

$$\underset{\mathbf{x}}{\text{Min}} \left\{ L(\mathbf{x}, \boldsymbol{\lambda}^*) \,\middle|\, \mathbf{x} \in \mathbf{X}_1 \subset \mathbf{R}^n ; \langle \boldsymbol{\lambda}^*, \mathrm{C}(\mathbf{x}) \rangle = 0 \right\} \tag{9b}$$

Step 4: Solve the bi-objective problem in (9) using an appropriate solution concept and numerical method.

The various types of solution that one may define for a multi-objective problem are discussed in [17]; the same paper advocates for single-point solutions based on the equilibrium or compromise rationale; and in this particular case, the Nash equilibrium solution is the more appropriate solution.[4]

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 (9). However numerical experience indicates that unless extra information is used to tightly delineate the feasible set for the Lagrange multipliers, i.e. $\boldsymbol{\Lambda}$, the solution algorithm is likely to be inefficient because the coordinating mechanism inherent in the Nash equilibrium concept is rather weak over large search spaces, and this often leads to rather erratic algorithmic performance.

[4] The multi-objective optimization exposition in [17] is couched in terms of game theory, and a game-theoretic perspective of the decentralized scheme suggested above is as follows. The Lagrangian saddle-value problem may be viewed as a statement of a two-person, zero-sum, *infinite* game. To see this, first, the decision vector is notionally expanded by appending the multiplier vector $\boldsymbol{\lambda}$ to yield a new decision vector $\mathbf{y}' = (\mathbf{x}' \mid \boldsymbol{\lambda}')$; secondly, the two parts of the "new" decision vector $\mathbf{y}$ are assumed to be controlled by two distinct agents in a decentralized scheme in which each parochially seeks to optimize the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda})$— the "utility" accruing to the maximizing agent being $L(\mathbf{x}, \boldsymbol{\lambda})$ and that accruing to the minimizer being $-L(\mathbf{x}, \boldsymbol{\lambda})$.

In two-person, zero-sum, *finite* games, it is well known that the solution is a saddle-point of the game matrix that may be computed (assuming it exists) by a technique known as 'iterated elimination of strictly dominated strategies'. And according to Moulin [12, p.1339]: "If a game is dominant solvable, the set of solutions obtained by the iterated elimination of strictly dominated strategies is a subset of the set of its Nash equilibri[a]."

But unlike in the two-person, zero-sum, finite game where the existence of the saddle-point is not assured, a saddle-point in the two-person zero-sum game comprising the saddle-value problem will always exists by virtue of the manner in which the pay-off functions are defined, viz.: $L(\mathbf{x}, \boldsymbol{\lambda})$ and $-L(\mathbf{x}, \boldsymbol{\lambda})$. And provided the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda})$ is finite and possesses a unique global maximum, the assumed game is dominant solvable—each iteration of the solution algorithm is in essence an 'elimination of strictly dominated strategies' because each point that is not visited by the solution algorithm may rightly be viewed as being part of a dominated strategy for each agent. Therefore, an algorithm that is capable of converging to Nash equilibria should successfully compute the Lagrangian saddle point.

An auxiliary coordinating mechanism is therefore necessary, and in this regard what is required is an extra term augmenting the Lagrangian, but one that vanishes at the optimal solution ($\mathbf{x}^*, \boldsymbol{\lambda}^*$). The Fischer-Burmeister function ('FB-function') affords such a coordinator—it is defined as follows [5]:

$$\phi(a, b) = \sqrt{a^2 + b^2} - (a + b); \quad a \geq 0; \quad b \geq 0.$$

The FB-function is sub-additive, positive homogeneous and Lipschitz-continuous; it is non-positive on $\mathbf{R}^2_+$ and it only attains its maximum value of '0' when either $a = 0$ or $b = 0$ or both, which in all cases implies that $ab = 0$; such a scenario is formally encapsulated by the following relations:

$$(a > 0) \wedge (b = 0) \Leftrightarrow \phi(a, b) = 0 \tag{10a}$$

$$(a = 0) \wedge (b > 0) \Leftrightarrow \phi(a, b) = 0 \tag{10b}$$

$$(a = 0) \wedge (b = 0) \Leftrightarrow \phi(a, b) = 0 \tag{10c}$$

Note that the conjunctions on left are exactly those in **Case I** of **LEMMA 1** when 'a' replaced by an element of the m-vector $\mathbf{C}(\mathbf{x})$ and 'b' replaced by the corresponding element of the Lagrange multiplier vector $\boldsymbol{\lambda}$. One way of implementing the coordination mechanism proper is to evaluate (at iteration k) each candidate solution ($\mathbf{x}^k, \boldsymbol{\lambda}^k$) using the relations in (10) and according to the criterion: 'the closer $\phi(\mathbf{x}^k, \boldsymbol{\lambda}^k)$ is to zero, the better'.

The presence of a coordinator does not necessarily yield efficient solution processes in all cases. But from the KKT theorem stated in (1), one may rightly deduce that the only information regarding $\boldsymbol{\lambda}^*$ that matters in this saddle-point approach where differentiability and hence (1a) are irrelevant is whether each λ_i is positive or zero, and so restricting the λ_i's to the binary set {0, 1} should equally suffice. The efficiency limitations due to the size of $\mathbf{R}^m$ may therefore be alleviated by defining $\boldsymbol{\lambda}$ as a vector in the much smaller space $\{0, 1\}^m$, thus creating a mixed-binary, bi-objective optimization problem,[5] viz.:

$$\text{MP}_2: \quad \min_{\mathbf{x}} \left\{ L(\mathbf{x}, \boldsymbol{\lambda}^*) \,\middle|\, \mathbf{x} \in \mathbf{X}_1 \subset \mathbf{R}^n; \langle \boldsymbol{\lambda}^*, \mathbf{C}(\mathbf{x}) \rangle = 0 \right\} \qquad \max_{\boldsymbol{\lambda}} \left\{ L(\mathbf{x}^*, \boldsymbol{\lambda}) \,\middle|\, \boldsymbol{\lambda} \in \{0, 1\}^m; \langle \boldsymbol{\lambda}, \mathbf{C}(\mathbf{x}^*) \rangle = 0 \right\}$$

Further algorithmic efficiency may be realized by converting the binary maximization sub-program of MP_2 into an integer program as follows:

1. First, a different perspective is adopted: the decision vector is notionally expanded by appending the multiplier vector $\boldsymbol{\lambda}$ to yield a new decision vector $\mathbf{y}' = (\mathbf{x}' \mid \boldsymbol{\lambda}')$;[6] the binary sub-vector $\boldsymbol{\lambda}'$ is viewed as a binary number with the least significant bit on the right-hand side; so, as far as the binary part is concerned, the search space for say a 4-dimensional binary sub-vector extends from '`0 0 0 0`' to '`1 1 1 1`';
2. Second, the binary number $\boldsymbol{\lambda}'$ is notionally converted into its base-10 equivalent, say z, and this scalar integer replaces the binary sub-vector $\boldsymbol{\lambda}'$ in $\mathbf{y}'$—this reduces the dimension of the decision vector by '`dim(λ) - 1`';
3. Third, the lower and upper bounds for the integer variable 'z' are calculated from their binary counter parts using the well-known binary-to-integer conversion formula:

$$z = \sum_{n=0}^{\dim(\boldsymbol{\lambda})-1} b_n \cdot 2^n$$

where: b_n — is the n-th binary bit; b_0 being the least significant bit;

$\dim(\boldsymbol{\lambda})$ — is the dimension of the binary vector $\boldsymbol{\lambda}$.

Obviously, the lower bound is trivially zero since all the b_n coefficients are zero; for the upper bound, all the b_n's are '1', and so for the 4-dimensional example cited above, we have: $2^3 + 2^2 + 2^1 + 2^0 = 15$.

[5] The mixed-binary optimization problem may be described in game-theoretic terms as a 'Two-person, zero-sum, semi-infinite game'

[6] See footnote 4

The final formulation is a bi-objective mixed-integer optimization problem in which one simultaneously seeks to minimise the Lagrangian $L(\mathbf{x}, \mathbf{z})$ over a continuous decision space and to maximise $L(\mathbf{x}, \mathbf{z})$ over a discrete decision space; it may formally be stated thus:

MP_3: Given $\mathbf{Z}_1 = \{0, 1, \ldots, \zeta\}$ where $\zeta = \sum_{n=0}^{\dim(\boldsymbol{\lambda})-1} 2^n$; and $\mathbf{X}_1 \equiv \{\mathbf{x} \mid \mathbf{C}(\mathbf{x}) \geq \mathbf{0}\}$:

$$\underset{\mathbf{x}}{\text{Min}} \left\{ L(\mathbf{x}, \mathbf{z}^*) \,\middle|\, \mathbf{x} \in \mathbf{X}_1; \mathbf{z}^* \in \mathbf{Z}_1 \right\} \qquad \underset{\mathbf{z}}{\text{Max}} \left\{ L(\mathbf{x}^*, \mathbf{z}) \,\middle|\, \mathbf{z} \in \mathbf{Z}_1; \mathbf{x}^* \in \mathbf{X}_1 \right\}$$

The rationale underlying MP_3 as a vehicle for seeking global optima to MP_1 is clear; but whether one can successfully compute a global solution much depends on the numerical algorithm involved. The solver GENO—whose description beyond this footnote[7] is unnecessary for the purposes of this paper—is well equipped for the task; the following numerical example serves to demonstrate its efficacy.[8]

4 Numerical Example

MP_4: $\underset{\mathbf{x}}{\text{Min}}\, f(\mathbf{x}) = 0.6224 x_1 x_3 x_4 + 1.7781 x_2 x_3^2 + 3.1661 x_4 x_1^2 + 19.84 x_3 x_1^2$

Subject to:
$$-x_1 + 0.0193 x_3 \leq 0$$
$$-x_2 + 0.00954 x_3 \leq 0$$
$$-\pi x_3^2 x_4 - (4/3)\pi x_3^3 + 1{,}296{,}000 \leq 0$$
$$x_i \in [0.0625, 99] \cap \{\xi_i : \xi_i = 0.0625\,n,\ n \in \mathbf{Z}\},\ \ i = 1, 2$$
$$x_i \in [10.0, 200],\ i = 3, 4$$

The program MP_4 was first re-cast into MP_3 mould and coded accordingly; the solution generated by GENO is as follows.

I. GENO Output

Generation Number	Time[9] (sec)	Objective 1	Objective 2
0	0.0000	-694079.15705000	694079.15705000
20	2.7650	-6071.74042904	6071.74042904
40	2.7140	-6060.57134714	6060.57134714
60	2.7350	-6059.71607451	6059.71607451
80	2.7180	-6059.71433505	6059.71433505
100	2.7660	-6059.71433505	6059.71433505
120	2.7500	-6059.71433505	6059.71433505
140	2.7500	-6059.71433505	6059.71433505
160	2.7810	-6059.71433505	6059.71433505
180	2.7660	-6059.71433505	6059.71433505
200	2.8120	-6059.71433505	6059.71433505

Optimal Decision Vector:	$\mathbf{x} = (0.81250000, 0.43750000, 42.09844560, 176.63659584)^T$
Optimal Lagrange Vector:	$\boldsymbol{\lambda} = (1.00000000, 0.00000000, 0.00000000)^T$
Objective Function Value:	$f(\mathbf{x}) = 6059.71433505$

[7] GENO is an acronym for ***G***eneral ***E***volutionary ***N***umerical ***O***ptimizer. GENO is a real-coded evolutionary algorithm that can be used to solve uni- or multi-objective optimization problems; the problems presented may be static or dynamic in character; they may be unconstrained or constrained by functional equality or inequality constraints, coupled with set constraints on the variables; the variables themselves may assume real or discrete values in any combination. For a more detailed description and performance evaluation of GENO, see [16].

[8] **Source**: Coello Coello [3], but Coello Coello's problem statement includes the inequality $x_4 \leq 240$ which is clearly redundant given the set-constraint $x_4 \in [10, 200]$. This example is not particularly special—it has been chosen merely to illustrate use of the MP_3 formulation; it can be (and has been) solved directly and more efficiently by GENO (**Example 3.20** in [16]) as well as by other specialist GAMS solvers [1].

[9] 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 used was of size 30.

II. General Remarks

This is a real-life practical engineering design problem concerning a pressure vessel; the problem has previously been tackled by Deb [4] using Genetic Adaptive Search (GeneAS); by Kannan & Kramer [9] using an augmented Lagrangian multiplier method; by Sandgren [14] using a branch and bound technique; and by Coello Coello [3] using a genetic algorithm. The table below is an extract from [3] to which has been appended the result by GENO.

	Coello Coello [3]	Deb [4]	Kannan, et al. [9]	Sandgren [14]	GENO
Best Function Value	6069.3267	6410.3811	7198.0428	8129.1036	6059.71433505

But, as pointed out earlier, this example can be (and has been) solved directly and more efficiently by GENO as well as some specialist deterministic solvers in the GAMS suite [1]; its use here is more to validate (or refute) the 'globally optimal' claim of the computational scheme described in §**2** and §**3**. The GENO solution coincides with the best known thus far; this assertion is supported by Yang, *et al* [18] who not only provide an extensive survey of published results, but also prove, via analysis, that the said solution *is* globally optimal; some specialist global solvers in the GAMS suite also provide numerical proof of this fact [1]. Other researchers, e.g. Cagnina, *et al* [2], have since found the same solution using various methods; García-Palomares, *et al* [6] have recently used the GENO solution as a benchmark in evaluating the efficacy of their algorithm.

Hedar & Fukushima [8] report a seemingly better solution valued 5868.764836; however, their method imposes a different set of constraints on x_3 and x_4 (namely $x_i \in [10, 240]$, $i = 3, 4$) and ignores the discreteness restriction on the same variables; but in any case, their solution is still sub-optimal because GENO actually finds a better one—assuming $x_i \in [10, 240]$, $i = 3, 4$—valued 5850.38306310 and located at $\mathbf{x} = (0.75, 0.375, 38.86010361, 221.36547166)^T$.

5 Summary and Conclusions

This paper has presents a relatively "unfettered" method for finding global optima to constrained nonlinear programs—the "unfettered" claim being in the sense that the given program needn't conform to the smoothness and the constraint qualification requirements dictated by the Karush-Kuhn-Tucker theorem—the founding theorem of nonlinear programming. The method is anchored on a sufficiency theorem derived from the Lagrangian saddle-value problem associated with the given nonlinear program; it involves a reformulation the given program into a bi-objective mixed-integer program that is then solved for the Nash equilibrium. A real-life practical engineering problem concerning the design of a pressure vessel has been presented to illustrate the efficacy of the method; its numerical solution by the solver GENO is globally optimal; it provides a benchmark against which other algorithms may be assessed.

6 Legalities

I. Licence and Trademarks

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

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

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.

References

1. Bussieck, M. R.: Private communication (2014).

2. Cagnina, L. C., Esquivel, S. C., Coello, Coello A. C.: Solving Engineering Optimization Problems with the Simple Constrained Particle Swarm Optimizer. *Informatica (Slovenia)*, **32**, 319-326 (2008)

3. Coello Coello, C. A.: Constraint-handling using an evolutionary multi-objective optimization technique. *Civil Engineering and Environmental Systems*, **17**, 319-346 (2000).

4. Deb, K.: GeneAS: A robust optimal design technique for mechanical component design. In: Dasgupta, D., Michalewicz, Z. (eds.). *Evolutionary Algorithms in Engineering Applications*. Springer-Verlag, Berlin (1997).

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

6. García-Palomares, U. M.,·Costa-Montenegro, E.·Asorey-Cacheda, R., González-Castaño, F. J.: Adapting derivative free optimization methods to engineering models with discrete variables. *Engineering Optimization*, **13**, 579-594 (2012).

7. Hanson, M. A.: Invexity and the Kuhn-Tucker theorem. *Journal of Mathematical Analysis and Applications*, **236**, 594-604 (1999)

8. Hedar, A., Fukushima, M.: Derivative-free filter simulated annealing method for constrained continuous global optimization. In: Di Pillo, G., Giannessi, F. (eds.). *Nonlinear Optimization and Applications 2*. Kluwer, (2005).

9. Kannan, B. K., Kramer, S. N.: An augmented Lagrangian multiplier based method for mixed integer discrete continuous optimization and its applications to mechanical design. *Journal of Mechanical Design. Transactions of the ASME*, **116**, 318-320 (1994).

10. Kjeldsen, T. H.: A contextualized historical analysis of the Kuhn-Tucker theorem in nonlinear programming: The impact of World War II. *Historia Mathematica*, **27**, 331-361 (2000).

11. 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).

12. Mangasarian, O. L.: *Nonlinear Programming*. SIAM, Philadelphia (1994).

13. Moulin, H.: Dominant solvable voting schemes. *Econometrica*, **47**, 1337-1351 (1979).

14. Sandgren, E.: Nonlinear integer and discrete programming in mechanical design. *Proceedings of ASME Design Technology Conference*, Kissimine, Florida, pp.95-105 (1988).

15. Siwale, I.: GENO™ 1.0: Supplement to User's Manual Part I: Static and Dynamic Programs. *Tech. Rep. RD-4-2005*, Apex Research Ltd, London (2005). [Online] Available at: http://www.aptech.com/resources

16. Siwale, I.: GENO™ 2.0: The GAUSS User's Manual. *Tech. Rep. RD-13-2013*, Apex Research Ltd, London (2013). Available with a trial version of GENO—contact Aptech Systems Inc.

17. Siwale, I.: Practical multi-objective programming. *Tech. Rep. RD-14-2013*, Apex Research Ltd, London (2013). [Online] Available at: http://www.researchgate.net or at: http://www.optimization-online.org

18. Takayama, A.: *Mathematical Economics*. Cambridge University Press, Cambridge (1985).

19. Yang, X-S, Huyck, C., Karamanoglu, M., Nawaz Khan, N.: True global optimality of the pressure vessel design problem: A benchmark for bio-inspired optimisation algorithms. *International Journal of Bio-Inspired Computation*, **6**, 329-335 (2013).