The Theory of Discrete Lagrange Multipliers for Nonlinear Discrete Optimization*

Benjamin W. Wah and Zhe Wu

Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois, Urbana-Champaign
Urbana, IL 61801, USA
{wah, zhewu}@manip.crhc.uiuc.edu
http://www.manip.crhc.uiuc.edu

Abstract. In this paper we present a Lagrange-multiplier formulation of discrete constrained optimization problems, the associated discrete-space first-order necessary and sufficient conditions for saddle points, and an efficient first-order search procedure that looks for saddle points in discrete space. Our new theory provides a strong mathematical foundation for solving general nonlinear discrete optimization problems. Specifically, we propose a new vector-based definition of descent directions in discrete space and show that the new definition does not obey the rules of calculus in continuous space. Starting from the concept of saddle points and using only vector calculus, we then prove the discrete-space first-order necessary and sufficient conditions for saddle points. Using well-defined transformations on the constraint functions, we further prove that the set of discrete-space saddle points is the same as the set of constrained local minima, leading to the first-order necessary and sufficient conditions for constrained local minima. Based on the first-order conditions, we propose a local-search method to look for saddle points that satisfy the first-order conditions.

1 Introduction

Many applications in engineering, decision science and operations research can be formulated as nonlinear discrete optimization problems, whose variables are restricted to discrete values and whose objective and constraint functions are nonlinear. The general formulation of a *nonlinear, nonconvex, discrete constrained minimization problem* is as follows:

$$\begin{aligned} &\text{minimize} && f(x) && \\ &\text{subject to} && g(x) \le 0 && x = (x_1, x_2, \ldots, x_n) \text{ is a vector} \\ & && h(x) = 0 && \text{of discrete variables} \end{aligned} \tag{1}$$

where $f(x)$ is the objective function, $g(x) = [g_1(x), \ldots, g_k(x)]^T$ is a k-component vector of inequality constraints, and $h(x) = [h_1(x), \ldots, h_m(x)]^T$ is an m-component vector of equality constraints, $f(x)$, $g(x)$, and $h(x)$ can be either

* Research supported by National Science Foundation Grant NSF MIP 96-32316.

convex or non-c
alytic (in closed
The possible
To formally cha
neighborhoods

Definition 1.
set of points {x
Neighborhoods
reachable from

For example
Hamming dista
which a variabl
points in which

Definition 2.
the following t

– *x is a feasi*
– *For all x′ ∈*

Note that w
only feasible po
minimum. Furt
of $\mathcal{N}(x)$ but ma
In general, t
however, does
neighborhood i
to include the
its original mea
points that are
guidance for do
search to explo

Example 1. Co
strained minim
between −2 an

minimize
subject to

where

and x is an int
Applying t
can be defined

convex or non-convex, linear or nonlinear, continuous or discontinuous, and analytic (in closed-form formulae) or procedural.

The possible solutions to (1) are local minima that satisfy all the constraints. To formally characterize the solutions to be found, we introduce the concepts of neighborhoods and constrained local minima in discrete space.

Definition 1. *$\mathcal{N}(x)$, the* <u>*neighborhood*</u> *of point x in space X, is a user-defined set of points $\{x' \in X\}$ such that $x \notin \mathcal{N}(x)$ and that $x' \in \mathcal{N}(x) \iff x \in \mathcal{N}(x')$. Neighborhoods must be defined such that any point in the finite search space is reachable from any other point through traversals of neighboring points.*

For example, in $\{0,1\}^n$ space, y (the neighborhood of x) can be points whose Hamming distance between x and y is less than 2. In modulo-integer space in which a variable is an integer element in $\{0, 1, \ldots, q-1\}$, y can be the set of points in which $\text{mod}(y_1 - x_1, k) + \ldots + \text{mod}(y_n - x_n, k)$ is less than 2.

Definition 2. *A point x is a* <u>*discrete constrained local minimum*</u> *if it satisfies the following two properties:*

- *x is a feasible point, implying that x satisfies all the constraints;*
- *For all $x' \in \mathcal{N}(x)$ such that x' is feasible, $f(x') \geq f(x)$ holds true.*

Note that when all neighboring points of x are infeasible and that x is the only feasible point surrounded by infeasible points, x is still a constrained local minimum. Further, note that point x may be a local minimum to one definition of $\mathcal{N}(x)$ but may not be for another definition of $\mathcal{N}'(x)$.

In general, the choice of neighborhoods is application-dependent. The choice, however, does not affect the validity of a search as long as one definition of neighborhood is used consistently throughout. Normally, one may choose $\mathcal{N}(x)$ to include the nearest discrete points to x so that neighborhood still carries its original meaning. However one can also choose the neighborhood to contain points that are "far away." Intuitively, neighboring points close to x can provide guidance for doing local descents, while neighboring points far away allow the search to explore larger regions in the search space.

Example 1. Consider the following one-dimensional nonlinear discrete constrained minimization problem whose constraints are satisfied at integer points between -2 and 3.

$$\begin{aligned} &\text{minimize } f(x) = 2 - 0.4x - 2.0x^2 + 0.75x^3 + 0.4x^4 - 0.15x^5 + \sin(5x) \\ &\text{subject to } h(x) = 0 \\ &\quad \text{where} \quad h(x) = \begin{cases} \sin(\pi x) & \text{if } -2 \leq x \leq 3 \\ 1 & \text{otherwise} \end{cases} \end{aligned} \tag{2}$$

and x is an integer variable. Figure 1 plots the curve of the objective function.

Applying the first two definitions on the example, the neighborhood of x can be defined as $\{x-1, x+1\}$. Given that $x = -3$ and $x = 4$ are infeasible,

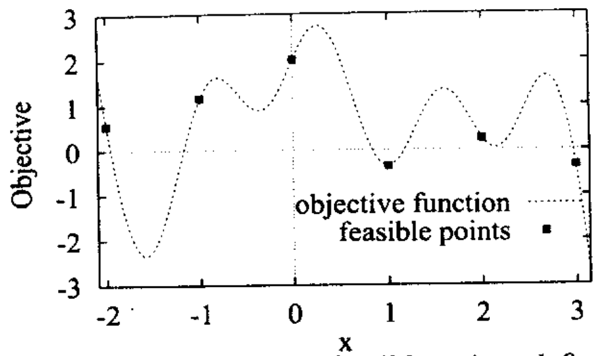

Fig. 1. The objective function and feasible points defined in (2)

$f(-1) > f(-2)$, $f(0) > f(1)$, $f(2) > f(1)$, and $f(2) > f(3)$, $x = -2, 1, 3$ are all the constrained local minima in (2), Out of the six points in the feasible region, the global minimum is at $x = 1$ where $f(1) = -0.359$. ■

Our goal in this paper is to show the equivalence between the set of discrete saddle points and the set of constrained local minima in the discrete Lagrangian space of a discrete nonlinear constrained optimization problem. That is, the condition for a saddle point is necessary and sufficient for a point to be a discrete constrained local minimum. The condition, therefore, provides a formal basis for the development of efficient algorithms for solving problems in this class. Our work here extends substantially our previous results in this area [23] that only proved the condition to be sufficient.

It is worth emphasizing that the theories and methods proposed here address discrete constrained *non-convex* optimization problems. We do not require the objective or constraint functions to be convex, as some of the existing continuous optimization methods do, because convexity is too restricted in general applications. Further, the original concept of convexity [20] is only applicable to continuous space, although some recent research [19] tries to extend the theory of convex analysis using matroids [5] to integer-valued functions.

The paper is organized as follows. Section 2 surveys existing work on discrete nonlinear constrained optimization and the basic concepts, theories and methods of continuous Lagrange multipliers. In Section 3, we prove the discrete-space first-order necessary and sufficient conditions, after introducing the concept of direction of maximum potential drop. These conditions form the mathematical foundation that leads to the discrete space first-order local-search method proposed in Section 4. Finally, Section 5 concludes the paper.

2 Existing Work

We summarize in this section related work in the area of discrete optimization and that in Lagrange multipliers for continuous optimization.

2.1 Transformation-Based Methods for Solving Discrete Problems

In the past, there has been extensive work on efficient techniques to solve (1). One major approach is to rewrite a discrete optimization problem as a constrained nonlinear 0-1 programming problem before solving it. Existing nonlinear 0-1 integer programming algorithms can be classified into four categories [12]. First, a nonlinear problem can be linearized by replacing each distinct product of variables by a new 0-1 variable and by adding some new constraints [30, 10]. This method only works for simple nonlinear problems. Second, algebraic methods [21] express the objective function as a polynomial function of the variables and their complements. These only work for the case in which all the constraints can be removed. Third, enumerative methods [29] utilize branch-and-bound algorithms to find lower bounds using linearized constraints. Lower bounds found this way are inaccurate when constraints are highly nonlinear. For a similar reason, branch-and-bound methods do not work well on general nonlinear integer programming problems. Last, cutting-plane methods [11] reduce a constrained nonlinear 0-1 problem into a generalized covering problem. However, they are limited because not all nonlinear 0-1 problems can be transformed this way.

A second approach to solving (1) is to transform it into an unconstrained problem before solving it using incomplete methods that may find a feasible solution in finite time if one exists but are not able to prove infeasibility. In this approach, an unconstrained problem is generally formulated as a weighted sum of the objective and the constraints. Examples of methods in this class include simulated annealing (SA) [16], genetic algorithms (GA) [13], tabu search [9], gradient descent [1], Hopfield networks [14], and penalty methods. These methods generally have difficulties in finding feasible solutions when their weights are not chosen properly. A class of penalty methods adjust the penalties (or weights) dynamically according to the amount of constraint violation in order to force all the constraints into satisfaction; examples of which include heuristic repair methods [4] and break-out strategies [18]. Although they work well in practice, there is no theoretical foundation on why they work well and under what conditions they will converge.

2.2 Lagrangian Methods for Solving Continuous Problems

Traditionally, Lagrangian methods were developed for solving continuous optimization problems. In this section, we present briefly their basic definitions and concepts, and examine the relationships among their solution spaces. These relationships are important because they show that some constrained local minima may not be found by existing continuous Lagrangian methods. This issue can be overcome in Lagrangian methods that work in discrete space (see Section 3).

Basic Definitions. A general ***continuous equality-constrained minimization problem*** is formulated as follows:

$$\begin{array}{lll} \text{minimize} \quad f(x) & x = (x_1, x_2, \ldots, x_n) \text{ is a vector} & \\ \text{subject to } h(x) = 0 & \quad \text{of continuous variables} & \end{array} \qquad (3)$$

where $h(x) = [h_1(x), \ldots, h_m(x)]^T$ is an m-component vector. Both $f(x)$ and $h(x)$ may be linear or nonlinear continuous functions.

In continuous space, the solutions to (3) are also called constrained local minima. However, their definition is slightly different from that in discrete space defined in Definition 2.

Definition 3. *A point x in continuous space is a* constrained local minimum *[17] iff there exists a small $\varepsilon > 0$ such that for all x' that satisfy $|x' - x| < \varepsilon$ and that x' is also a feasible point, $f(x') \geq f(x)$ holds true.*

The following two definitions define the Lagrangian function as a weighted sum of the objective and the constraints, and the saddle point as a point in the Lagrangian space where the Lagrange multipliers are at their local maxima and the objective function is at its local minimum.

Definition 4. *The Lagrangian function of (3) is $L(x, \lambda) = f(x) + \lambda^T h(x)$, where λ is a vector of Lagrange multipliers.*

Definition 5. *A saddle point (x^*, λ^*) of function L is defined as one that satisfies $L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$ for all (x^*, λ) and all (x, λ^*) sufficiently close to (x^*, λ^*).*

In general, there is no efficient method to find saddle points in continuous space.

First-Order Necessary Conditions for Continuous Problems. There are various continuous Lagrangian methods that can be used to locate constrained local minima. They are all based on two first-order necessary conditions.

Theorem 1. *First-order necessary conditions for continuous problems [17]. Let x be a local extremum point of $f(x)$ subject to $h(x) = 0$. Further, assume that $x = (x_1, \ldots, x_n)$ is a regular point (see [17]) of these constraints. Then there exists $\lambda \in E^m$ such that $\nabla_x f(x) + \lambda^T \nabla_x h(x) = 0$. Based on the definition of Lagrangian function, the necessary conditions can be expressed as:*

$$\nabla_x L(x, \lambda) = 0; \qquad \nabla_\lambda L(x, \lambda) = 0. \tag{4}$$

First-Order Methods for Solving Continuous Problems. Based on the first-order necessary conditions in continuous space, there are a number of methods for solving constrained minimization problems. These include the first-order method, Newton's method, modified Newton's methods, quasi-Newton methods [17], and sequential quadratic programming [15]. Among these methods, a popular one is the first-order method represented as an iterative process:

$$x^{k+1} = x^k - \alpha_k \nabla L_x(x^k, \lambda^k)^T \tag{5}$$

$$\lambda^{k+1} = \lambda^k + \alpha_k h(x^k) \tag{6}$$

where α_k is a step-size parameter to be determined.

Intuitively, these two equations represent two counter-acting forces to resolve constraints and find high-quality solutions. When any of the constraints is violated, its degree of violation is used in (6) to increase the corresponding Lagrange multiplier in order to increase the penalty on the unsatisfied constraint and to force it into satisfaction. When the constraint is satisfied, its Lagrange multiplier stops to grow. Hence, (6) performs ascents in the Lagrange-multiplier space until each Lagrange multiplier is at its maximum. In contrast, (5) performs descents in the objective space when all the constraints are satisfied and stops at an extremum in that space. By using a combination of simultaneous ascents and descents, an equilibrium is eventually reached in which all the constraints are satisfied and the objective is at a local extremum.

To guarantee that the equilibrium point is a local minimum, there are second order sufficient conditions to make the solution a strict constrained local minimum. They are omitted since they are irrelevant to solving discrete problems.

Relationships among solution spaces in continuous problems. We present without proof [32] three lemmas and one theorem to show the relationships among the three solution spaces: the set of constrained local minima (**A**), the set of solutions satisfying the first-order necessary and second-order sufficient conditions (**B**), and the set of saddle points (**C**).

Lemma 1. *The set of saddle points is a proper subset of the set of constrained local minima;* i.e., $x \in \mathbf{C} \Rightarrow x \in \mathbf{A}$, *and* $x \in \mathbf{A} \not\Rightarrow x \in \mathbf{C}$.

Lemma 2. *The set of solutions satisfying the first-order necessary and second-order sufficient conditions is a proper subset of the set of constrained local minima;* i.e., $x \in \mathbf{B} \Rightarrow x \in \mathbf{A}$, *and* $x \in \mathbf{A} \not\Rightarrow x \in \mathbf{B}$.

Lemma 3. *The set of solutions* (x^*, λ^*) *to the first-order necessary and second-order sufficient conditions may not be equal to the set of saddle points;* i.e., $x \in \mathbf{B} \not\Rightarrow x \in \mathbf{C}$, *and* $x \in \mathbf{C} \not\Rightarrow x \in \mathbf{B}$.

Theorem 2. *The relationships among the solutions sets* **A**, **B**, *and* **C** *is as follows: (a)* $\mathbf{B} \subseteq \mathbf{A}$, *(b)* $\mathbf{C} \subseteq \mathbf{A}$, *and* $\mathbf{B} \neq \mathbf{C}$. *Figure 2 depicts their relationships.*

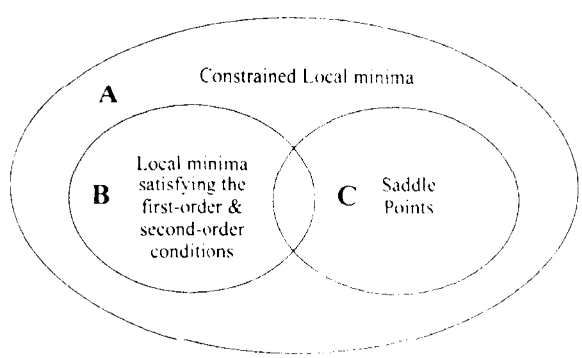

Fig. 2. Relationship among solution sets of Lagrangian methods for solving continuous problems.

2.3 Lagrangian Relaxation

There is a class of algorithms called *Lagrangian relaxation* [7,8,6,24,3] proposed in the literature that should not be confused with the Lagrange-multiplier methods proposed in this paper. Lagrangian relaxation reformulates a *linear* integer minimization problem:

$$\begin{aligned} z = \text{minimize}_x\ & Cx \\ \text{subject to } & Gx \leq b \quad \text{where } x \text{ is an integer vector of variables} \\ & x \geq 0 \quad\quad \text{and } C \text{ and } G \text{ are constant matrices} \end{aligned} \tag{7}$$

into the following form:

$$\begin{aligned} L(\lambda) = \text{minimize}_x \quad & (Cx + \lambda(b - Gx)) \\ \text{subject to} \quad & x \geq 0. \end{aligned} \tag{8}$$

Obviously, the new relaxed problem is simple and can be solved efficiently for any given vector λ. The method is based on Lagrangian Duality theory [25] upon which a general relationship between the solution to the original minimization problem and the solution to the relaxed problem can be deduced. There was some research [2] addressing nonlinear optimization problems. However, as pointed out in [25], Lagrangian relaxation aims to find an optimal primal solution given an optimal dual solution, or vice versa. This approach is simple in the case of linear functions but not for nonlinear functions. It has been proved [7] that $L(\lambda)$ can be used as an upper bound on z in (7), although better bounds can be computed in some special cases [6]. The bounds computed can be applied in a branch-and-bound search to solve linear integer programming problems.

In contrast, our Lagrange-multiplier formulation addresses *nonlinear* discrete constrained minimization problems defined in (1) that cannot be solved by Lagrangian relaxation. The mathematical foundation of our proposed formulation is based on two new discrete-space first-order necessary and sufficient conditions that define the necessary and sufficient conditions when the search finds a constrained local minimum.

3 Discrete Space Lagrangian Formulations/Methods

In this section we derive a new theory of Lagrange multipliers to work in discrete space.

3.1 Basic Definitions

We first consider a special case of (1) with equality constraints, and will defer till Section 3.3 to discuss ways to handle inequality constraints.

$$\begin{aligned} \text{minimize} \quad & f(x) \qquad x = (x_1, x_2, \ldots, x_n) \text{ is a vector} \\ \text{subject to } & h(x) = 0 \qquad \text{of discrete variables} \end{aligned} \tag{9}$$

Similar to the continuous case [17], the <u>*discrete Lagrangian function*</u> [23] of (9) is defined to be:

$$L_d(x, \lambda) = f(x) + \lambda^T H(h(x)). \tag{10}$$

where H is a continuous transformation function that satisfies $H(x) = 0 \Leftrightarrow x = 0$. There are various transformation functions that can be used. Note that function H was not used in our previous paper [23] and is introduced here to prove the necessary and sufficient condition. Its exact form is discussed in the next subsection.

We cannot use L_d to derive first-order necessary conditions similar to those in continuous space [17] because there are no gradients or differentiation in discrete space. Without these concepts, none of the calculus in continuous space is applicable in discrete space.

An understanding of gradients in continuous space shows that they define directions in a small neighborhood in which function values decrease. To this end, we define in discrete space a <u>*direction of maximum potential drop*</u> (*DMPD*) for $L_d(x, \lambda)$ at point x for fixed λ as a vector[1] that points from x to a neighbor of $x \in \mathcal{N}(x)$ with the minimum L_d:

$$\Delta_x L_d(x, \lambda) = \boldsymbol{\nu}_x = y \ominus x = (y_1 - x_1, \ldots, y_n - x_n) \tag{11}$$
$$\text{where} \quad y \in \mathcal{N}(x) \cup \{x\} \text{ and } L_d(y, \lambda) = \min_{\substack{x' \in \mathcal{N}(x) \\ \cup \{x\}}} L_d(x', \lambda).$$

Here, $\ominus$ is the vector-subtraction operator for changing x in discrete space to one of its "user-defined" neighborhood points $\mathcal{N}(x)$. Intuitively, $\boldsymbol{\nu}_x$ is a vector pointing from x to y, the point with the minimum L_d value among all neighboring points of x, including x itself. That is, if x itself has the minimum L_d, then $\boldsymbol{\nu}_x = \mathbf{0}$. It is important to emphasize that, with this definition of discrete descent directions, DMPDs cannot be added/subtracted in discrete space [32]. Consequently, the proof procedure of first-order conditions in continuous space is not applicable here. This is shown in the following lemma.

Lemma 4. *There is no addition operation for* DMPD *defined in (11).*

Proof. In general, *DMPD* of $[f_1(x) + f_2(x)]$ is not the same as the summation of *DMPDs* of $f_1(x)$ and of $f_2(x)$. The following example illustrates the point.

$$f_1(0,0) = 0,\ f_1(0,1) = -3,\ f_1(1,0) = 0, \quad f_1(-1,0) = -2,\ f_1(0,-1) = 0,$$
$$f_2(0,0) = 0,\ f_2(0,1) = 0, \quad f_2(1,0) = -3,\ f_2(-1,0) = -2,\ f_2(0,-1) = 0.$$

From the definition of *DMPD*, we know that $\Delta_x f_1(0,0) = (0,1)$, $\Delta_x f_2(0,0) = (1,0)$ and $\Delta_x(f_1 + f_2)(0,0) = (-1,0)$. Hence, $\Delta_x(f_1 + f_2)(0,0) \neq \Delta_x f_1(0,0) \oplus \Delta_x f_2(0,0)$. ■

[1] To simplify our symbols, we represent points in x space without the explicit vector notation.

The result of this lemma implies that the addition of two *DMPDs* cannot be carried out, rendering it impossible to prove the first-order conditions similar to those proved in Theorem 1 for continuous problems.

Based on DMPD, we define the concept of discrete saddle points [23,32] in discrete space similar to those in continuous space [17].

Definition 6. *A point (x^*, λ^*) is a discrete saddle point when:*

$$L_d(x^*, \lambda) \leq L_d(x^*, \lambda^*) \leq L_d(x, \lambda^*), \tag{12}$$

for all $x \in \mathcal{N}(x^*)$ and all possible λ.

Note that the first inequality in (12) only holds when all the constraints are satisfied, which implies that it must be true for all λ.

3.2 Characteristics of Discrete Saddle Points

The concept of saddle points is of great importance to discrete problems because, starting from saddle points, we can derive first-order necessary and sufficient conditions for discrete problems and develop efficient first-order procedures for finding constrained local minima. Although these conditions are similar to those for continuous problems, they were derived from the concept of saddle points rather than from regular points [17].

Similar to the continuous case, we denote **A** to be the set of all constrained local minima, **B** to be the set of all solutions that satisfy the discrete-space first-order necessary and sufficient conditions (13) and (14), and **C** to be the set of all discrete saddle points satisfying (12).

Lemma 5. *First-order necessary and sufficient conditions for discrete saddle points. In discrete space, the set of all saddle points is equal to the set of all solutions that satisfy the following two equations:*

$$\Delta_x L_d(x, \lambda) = \Delta_x[f(x) + \lambda^T H(h(x))] = 0 \tag{13}$$

$$h(x) = 0 \tag{14}$$

Note that the Δ operator in (13) is for discrete space.

Proof. The proof is done in two parts:

"$\Rightarrow$" part: Given a saddle point (x^*, λ^*), we like to prove it to be a solution to (13) and (14). Eq. (13) is true because L_d cannot be improved among $\mathcal{N}(x^*)$ from the definition of saddle points. Hence, $\Delta_x L_d(x^*, \lambda^*) = 0$ must be true from the definition of *DMPD*. Eq. (14) is true because $h(x) = 0$ must be satisfied at any solution point.

"$\Leftarrow$" part: Given a solution (x^*, λ^*) to (13) and (14), we like to prove it to be a discrete saddle point. The first condition $L_d(x^*, \lambda^*) \leq L_d(x, \lambda^*)$ holds for all $x \in \mathcal{N}(x^*)$ because $\Delta_x L_d(x^*, \lambda^*) = 0$. Hence, no improvement of L_d can be found in the neighborhood of x^*. The second condition $L_d(x^*, \lambda) \leq L_d(x^*, \lambda^*)$ is true for all λ because $h(x^*) = 0$ according to (14). Thus, (x^*, λ^*) is a saddle point in discrete space. ■

For the discrete Lagrangian definition in (10), we found that, if $H(h(x))$ is always non-negative (or non-positive), then the set of constrained local minima is the same as the set of saddle points. Examples of transformation H are the absolute function and the square function.

Lemma 6. *Sufficient conditions for constrained local minimum to be a saddle point. In the discrete Lagrangian function defined in (10), if $H(x)$ is a continuous function satisfying $H(x) = 0 \Leftrightarrow x = 0$ and is non-negative (or non-positive), then* $\mathbf{A} = \mathbf{C}$ *holds.*

Proof. We only prove the case when $H(x)$ is non-negative, and the other case can be proved similarly. To prove this lemma, we construct λ^* for every constrained local minimum x^* in order to make (x^*, λ^*) a saddle point. This λ^* must be bounded and be found in finite time in order for the procedure to be useful.

(a): Constructing λ^*. Given x^*, consider $x \in \mathcal{N}(x^*)$. Let $h(x) = (h_1(x), \dots, h_m(x))$ be an m-element vector, and the initial $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*) = (0, \dots, 0)$. For every x such that $H(h(x)) > 0$, there is at least one constraint that is not satisfied, say $H(h_i(x)) > 0$. For this constraint, we set:

$$\lambda_i^* \to \max\left(\lambda_i^*, \frac{f(x^*) - f(x)}{H(h_i(x))}\right) \tag{15}$$

The update defined in (15) is repeated for every unsatisfied constraint of x and every $x \in \mathcal{N}(x^*)$ until no further update is possible. Since $\mathcal{N}(x^*)$ has finite number of elements in discrete space, (15) will terminate in finite time and result in finite λ^* values.

(b) Proving that (x^*, λ^*) is a saddle point. To prove that (x^*, λ^*) is a saddle point, we need to prove that, for all $x \in \mathcal{N}(x^*)$, $L_d(x^*, \lambda) \le L_d(x^*, \lambda^*) \le L_d(x, \lambda^*)$. The first inequality is trivial because $L_d(x^*, \lambda) = f(x^*) = L_d(x^*, \lambda^*)$. In the second inequality, for all $x \in \mathcal{N}(x^*)$ such that $h(x) = 0$, it is obvious that $L_d(x^*, \lambda^*) = f(x^*) \le f(x) = L_d(x, \lambda^*)$ since x^* is a constrained local minimum. For all $x \in \mathcal{N}(x^*)$ such that $h(x) \ne 0$, there must be at least one constraint that is not satisfied, say $H(h_i(x)) > 0$. Moreover, from the construction method, we know that $\lambda_i^* \ge \frac{f(x^*) - f(x)}{H(h_i(x))}$. Therefore, $L_d(x^*, \lambda^*) = f(x^*) \le f(x) + \lambda_i^* H(h_i(x))$ holds. Further, since $\sum_{j=1, j \ne i}^{m} \lambda_j^* H(h_j(x))$ is non-negative (assuming all constraints are transformed by H into non-negative functions), it is obvious true that

$$L_d(x^*, \lambda^*) = f(x^*) \le f(x) + \sum_{j=1}^{m} \lambda_j^* H(h_j(x)) = L_d(x, \lambda^*).$$

Hence, (x^*, λ^*) is a saddle point. ■

The above lemma shows that transformation function H should be non-negative or non-positive, but not both. We illustrate in the following example that, given the discrete Lagrangian function $L_d(x,\lambda) = f(x) + \lambda^T h(x)$, a constrained local minimum x^* may not be a saddle point when $h(x)$ can have both positive and negative values. We construct a counter example to demonstrate that it is not always possible to find λ^* to make (x^*, λ^*) a saddle point even when x^* is a constrained local minimum in discrete space.

Consider a two-dimensional discrete equality-constrained problem with objective $f(x)$ and constraint $h(x) = 0$, where

$$\begin{aligned} f(0,0) = 0,\ f(0,1) = 1,\ f(0,-1) = 0,\ f(1,0) = 0,\ f(-1,0) = -1, \\ h(0,0) = 0,\ h(0,1) = 0,\ h(0,-1) = 1,\ h(1,0) = 0,\ h(-1,0) = -1. \end{aligned}$$

Obviously, $(x^*, y^*) = (0,0)$ is a constrained local minimum. Furthermore, from the definition of Lagrangian function, we know that $L_d((0,0),\lambda) = 0$ holds true for any λ because $h(0,0) = f(0,0) = 0$.

To draw a contradiction, assume that $(0,0)$ is a saddle point. Hence, there exists λ^* such that $L_d((0,0),\lambda^*) \leq L_d((-1,0),\lambda^*)$, and $L_d((0,0),\lambda^*) \leq L_d((0,-1),\lambda^*)$. After substitution, we get the following equations:

$$\begin{aligned} 0 &\leq f(-1,0) + \lambda^* \times h(-1,0) = -1 + \lambda^* \times (-1), \\ 0 &\leq f(0,-1) + \lambda^* \times h(0,-1) = 0 + \lambda^* \times 1. \end{aligned} \tag{16}$$

Since there is no solution for λ^* in (16), the example shows that $(0,0)$ is a constrained local minimum but not a saddle point.

Finally, we show that finding any $\lambda \geq \lambda^*$ suffices to find a saddle point.

Corollary 1. *Given λ^* defined in Lemma 6, (x^*, λ') is a saddle point for any $\lambda' \geq \lambda^*$, where $\lambda' \geq \lambda^*$ means that every element of λ' is not less than the corresponding element of λ^*.*

Proof. The proof of the corollary is similar to that of Lemma 6 and will not be shown here. The corollary is important because, in practice, a search algorithm may only be able to find $\lambda \geq \lambda^*$ but not exactly λ^*. ∎

3.3 Handling Inequality Constraints

The results discussed so far apply only to discrete optimization problems with equality constraints. To handle (1) with inequality constraints, we need to transform them into equality constraints.

One general method of transforming inequality constraint $g_j(x) \leq 0$ is to apply a maximum function to convert the constraint into $\max(g_j(x), 0) = 0$. Obviously, the new constraint is satisfied iff $g_j(x) \leq 0$. The discrete Lagrangian function using the maximum transformation for inequality constraints is:

$$L_d(x,\lambda,\mu) = f(x) + \lambda^T H(h(x)) + \sum_{i=1}^{k} \mu_i H(\max(0, g_i(x))) \tag{17}$$

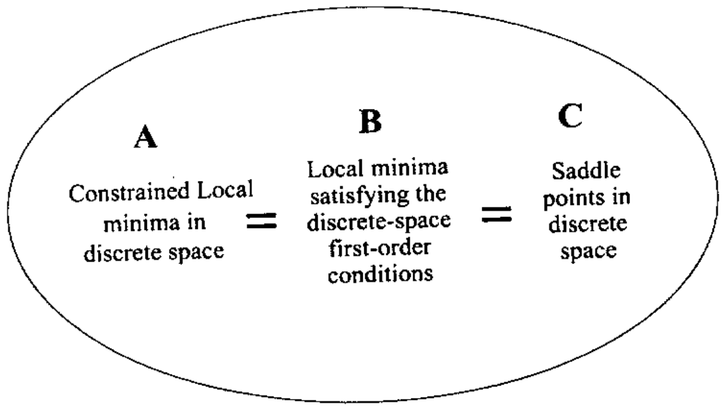

Fig. 3. Relationship among solution sets of Lagrangian methods for solving discrete problems when H satisfies the conditions in Theorem 3.

3.4 Discrete Space First-Order Necessary & Sufficient Conditions

This section summarizes the main theorem on discrete Lagrange multipliers.

Theorem 3. *First-order necessary and sufficient conditions for discrete constrained local minima. In discrete space, if $H(x)$ in the discrete Lagrangian definition in (10) is a continuous function satisfying $H(x) = 0 \Leftrightarrow x = 0$ and is non-negative (or non-positive), then the three solutions sets are equal, namely,* $\mathbf{A} = \mathbf{B} = \mathbf{C}$.

Proof. The proof follows from Lemmas 5 and 6. Figure 3 illustrates the relationships proved by this theorem. ■

Theorem 3 is of great importance in the following sense:

- Any search strategy aiming to find saddle points is sufficient because it is equivalent to searching for constrained local minima.
- A global optimization strategy looking for the saddle point with the minimum objective value will result in the constrained global minimum because the set of global minima is a subset of the set of saddle points.

In contrast, a continuous-space global-minimization method based on continuous first-order necessary and second-order sufficient conditions is not guaranteed to find the global minimum because the global solution may be outside the set of solutions that satisfy the first-order and second-order conditions. This fact is illustrated in Figure 2.

4 Discrete Space First-Order Local-Search Methods

The first-order necessary and sufficient conditions provide a basis for an efficient local-search method. In a way similar to that in continuous space, we propose an iterative *discrete first-order method* that looks for discrete saddle points.

$$x^{k+1} = x^k \oplus \Delta_x L_d(x^k, \lambda^k) \tag{18}$$

$$\lambda^{k+1} = \lambda^k + c_1 H(h(x^k)) \tag{19}$$

where $\oplus$ is the vector-addition operator ($x \oplus y = (x_1 + y_1, \ldots x_n + y_n)$), and c_1 is a positive step-wise real number controlling how fast the Lagrange multipliers change. This method is more general than the one used earlier [23] because it is not restricted to a special neighborhood function of Hamming distance one and arithmetic additions.

It should be obvious that the necessary condition for (18) and (19) to converge is when $h(x) = 0$, implying that x is a feasible solution to the original problem. If any of the constraints in $h(x)$ is not satisfied, then λ will continue to evolve to suppress the unsatisfied constraints. Further, as in continuous Lagrangian methods, the time for (18) and (19) to find a saddle point may be unbounded and can only be determined empirically, even when there exist feasible solutions.

The first-order search method in (18) and (19) has been applied to design multiplierless filter banks [26,28], solve satisfiability problems [23,22,31], and evaluate nonlinear discrete optimization benchmarks [32].

The following theorem guarantees that (18) and (19) can be used to locate saddle points [23].

Theorem 4. *Fixed-point theorem for discrete problems. A saddle point (x^*, λ^*) of (10) is reached iff (18) and (19) terminates [23].*

Proof. The proof consists of two parts.

"$\Rightarrow$" part: We prove that at a saddle point (18) and (19) will stop. This means that if (x^*, λ^*) is a saddle point, then it is also a fixed point of the iterative process. Given a saddle point (x^*, λ^*), we know from its definition that $L_d(x^*, \lambda^*) \leq L_d(x, \lambda^*)$ holds true for all $x \in \mathcal{N}(x^*)$. Thus, from the definition of *DMPD*, we conclude that $x_{k+1} = x_k$. In addition, since x^* is feasible, we conclude that $H(h(x_k)) = 0 = h(x_k)$. Since $x_{k+1} = x_k$ and $\lambda_{k+1} = \lambda_k$, (18) and (19) will stop at (x^*, λ^*).

"$\Leftarrow$" part: We prove the point that (18) and (19) stop at must be a saddle point of $L_d(x, \lambda) = f(x) + \lambda^T H(h(x))$. This also means that the method will not stop at points other than saddle points. Since the method stops at (x_k, λ_k), it implies that $x_{k+1} = x_k$ and $\lambda_{k+1} = \lambda_k$. These conditions imply that $\Delta_x L_d(x_k, \lambda_k) = \mathbf{0}$, meaning that $L_d(x_k, \lambda_k) \leq L_d(x'_k, \lambda_k)$ for any $x'_k \in \mathcal{N}(x_k)$. Moreover, $h(x_k) = 0$ is true because $\lambda_{k+1} = \lambda_k$. Hence, x_k is a feasible point, and (x_k, λ_k) is a saddle point. ■

5 Conclusions

In this paper, we have extended the theory of Lagrange multipliers to discrete optimization problems and have shown two important results that form the mathematical foundation in this area:

- For general discrete constrained optimization problems, we have shown the first-order necessary and sufficient conditions for saddle points in the solution space.
- After transforming general constraints into non-negative or non-positive functions, we have further shown the equivalence between the set of saddle points and the set of constrained local minima. Hence, the same first-order necessary and sufficient conditions derived for saddle points become the necessary and sufficient conditions for constrained local minima.

The last result is particularly significant because it implies that finding the saddle point with the best solution value amounts to global optimization of a discrete constrained optimization problem. A global optimization procedure with asymptotic convergence is presented in another paper [27].

References

1. K. J. Arrow and L. Hurwicz. Gradient method for concave programming, I: Local results. In K. J. Arrow, L. Hurwica, and H. Uzawa, editors, *Studies in Linear and Nonlinear Programming*. Stanford University Press, Stanford, CA, 1958.
2. E. Balas. *Minimax and Duality for Linear and Nonlinear Mixed-Integer Programming*. North-Holland, Amsterdam, Netherlands, 1970.
3. M. S. Bazaraa and J. J. Goode. A survey of various tactics for generating Lagrangian multipliers in the context of Lagrangian duality. *European Journal of Operational Research*, 3:322–338, 1979.
4. K. F. M. Choi, J. H. M. Lee, and P. J. Stuckey. A Lagrangian reconstruction of a class of local search methods. In *Proc. 10th Int'l Conf. on Artificial Intelligence Tools*. IEEE Computer Society, 1998.
5. A. Frank. An algorithm for submodular functions on graphs. *Ann. Discrete Math.*, 16:97–120, 1982.
6. B. Gavish. On obtaining the 'best' multilpliers for a lagrangean relaxation for integer programming. *Comput. & Ops. Res.*, 5:55–71, 1978.
7. A. M. Geoffrion. Lagrangean relaxation for integer programming. *Mathematical Programming Study*, 2:82–114, 1974.
8. F. R. Giles and W. R. Pulleyblank. *Total Dual Integrality and Integer Polyhedra*, volume 25. Elsevier North Holland, Inc., 1979.
9. F. Glover. Tabu search — Part I. *ORSA J. Computing*, 1(3):190–206, 1989.
10. F. Glover and E. Woolsey. Converting the 0-1 polynomial programming problem to a 0-1 linear program. *Operations Research*, 22:180–182, 1975.
11. D. Granot, F. Granot, and W. Vaessen. An accelerated covering relaxation algorithm for solving positive 0-1 polynomial programs. *Mathematical Programming*, 22:350–357, 1982.
12. P. Hansen, B. Jaumard, and V. Mathon. Constrained nonlinear 0-1 programming. *ORSA Journal on Computing*, 5(2):97–119, 1993.
13. J. H. Holland. *Adaption in Natural and Adaptive Systems*. University of Michigan Press, Ann Arbor, 1975.
14. J. J. Hopfield and D. W. Tank. Neural computation by concentrating information in time. In *Proc. National Academy of Sciences*, volume 84, pages 1896–1900, Washington, D.C., 1987. National Academy of Sciences.

15. M. E. Hribar. Large scale constrained optimization. *Ph.D. Disertation, Northeasten University*, 1996.
16. S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983.
17. D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley Publishing Company, 1984.
18. P. Morris. The breakout method for escaping from local minima. In *Proc. of the 11th National Conf. on Artificial Intelligence*, pages 40–45, Washington, DC, 1993.
19. K. Murota. Discrete convex analysis. *Mathematical Programming*, 83(3):313 – 371, 1998.
20. R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.
21. I. Rosenberg. Minimization of pseudo-boolean functions by binary development. *Discrete Mathematics*, 7:151–165, 1974.
22. Y. Shang and B. W. Wah. Discrete Lagrangian-based search for solving MAX-SAT problems. In *Proc. Int'l Joint Conf. on Artificial Intelligence*, pages 378–383. IJCAI, August 1997.
23. Y. Shang and B. W. Wah. A discrete Lagrangian based global search method for solving satisfiability problems. *J. Global Optimization*, 12(1):61–99, January 1998.
24. J. F. Shapiro. Generalized Lagrange multipliers in integer programming. *Operations Research*, 19:68–76, 1971.
25. J. Tind and L. A. Wolsey. An elementary survey of general duality theory in mathematical programming. *Mathematical Programming*, pages 241–261, 1981.
26. B. W. Wah, Y. Shang, and Z. Wu. Discrete Lagrangian method for optimizing the design of multiplierless QMF filter banks. *IEEE Transactions on Circuits and Systems, Part II*, (accepted to appear) 1999.
27. B. W. Wah and T. Wang. Simulated annealing with asymptotic convergence for nonlinear constrained global optimization. *Principles and Practice of Constraint Programming*, (accepted to appear) October 1999.
28. B. W. Wah and Z. Wu. Discrete Lagrangian method for designing multiplierless two-channel PR-LP filter banks. *VLSI Signal Processing*, 21(2):131–150, June 1999.
29. X. D. Wang. An algorithm for nonlinear 0-1 programming and its application in structural optimization. *Journal of Numerical Method and Computational Applications*, 1(9):22–31, 1988.
30. L. J. Watters. Reduction of integer polynomial programming to zero-one linear programming problems. *Operations Research*, 15:1171–1174, 1967.
31. Z. Wu and B. W. Wah. Solving hard satisfiability problems using the discrete Lagrange-multiplier method. In *Proc. 1999 National Conference on Artificial Intelligence*, pages 673–678. AAAI, July 1999.
32. Zhe Wu. *Discrete Lagrangian Methods for Solving Nonlinear Discrete Constrained Optimization Problems.* M.Sc. Thesis, Dept. of Computer Science, Univ. of Illinois, Urbana, IL, May 1998.