MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

A Framework of Discrete DC Programming by Discrete Convex Analysis

Takanori MAEHARA and Kazuo MUROTA

METR 2013–28 October 2013

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

A Framework of Discrete DC Programming by Discrete Convex Analysis

Takanori MAEHARA

National Institute of Informatics
JST, ERATO, Kawarabayashi Project
maehara@nii.ac.jp

Kazuo MUROTA

Department of Mathematical Informatics, University of Tokyo
murota@mist.i.u-tokyo.ac.jp

October 2013

Abstract

A theoretical framework of difference of discrete convex functions (discrete DC functions) and optimization problems for discrete DC functions is established. Standard results in continuous DC theory are exported to the discrete DC theory with the use of discrete convex analysis. A discrete DC algorithm, which is a discrete analogue of the continuous DC algorithm (Concave-Convex Procedure in machine learning) is proposed. The algorithm contains the submodular-supermodular procedure as a special case. Exploiting the polyhedral structure discrete convex functions, the algorithms tailored to specific types of discrete DC functions are proposed.

1 Introduction

The theory of DC functions (difference of two convex functions) and DC programming, which treats minimization problems of DC functions (DC programs) is the one of the most successful areas of non-convex optimization [18, 50]. Most of the non-convex optimization problems that arise in practice are indeed DC programming problems [15]. The DC theory is based on a basic non-convex duality theorem, called Toland-Singer duality [44,47]:

$$\inf_{x \in \mathbb{R}^n} \{g(x) - h(x)\} = \inf_{p \in \mathbb{R}^n} \{h^*(p) - g^*(p)\}.$$

A DC program is hard to solve in general, but when the objective function has a nice DC representation, there are some practical algorithms based on the Toland-Singer duality, such as the DC algorithm [45,46,52] and branch-and-bound/cutting-plane type algorithms [49,50]. DC functions are studied in many areas, such as optimization, game theory, variational analysis, spectral theory, and operator theory; see [1] for more details.

The objective of this paper is to establish a discrete analogue of the theory of DC programming. In the conventional (continuous) DC programming, the following properties of convex functions play key roles: (1) biconjugacy: $f^{**} = f$, which is used to prove the Toland-Singer duality, and (2) subdifferentiability: $\partial f(x) \neq \emptyset$ for each $x \in \operatorname{dom} f$, which is used in the DC algorithm. To export these ingredients to discrete functions, we utilize discrete convex analysis developed by Murota and others [8, 33, 35, 36].

In discrete convex analysis, two convexity notions, M$^\natural$-convexity and L$^\natural$-convexity, are distinguished: M$^\natural$-convexity is a generalization of matroid property and L$^\natural$-convexity is a generalization of submodularity on subsets. Conjugacy between M$^\natural$-convex functions and L$^\natural$-convex functions under discrete Legendre-Fenchel transformation is a distinctive feature of discrete convex analysis. Fundamental results in continuous convex analysis, in particular (1) biconjugacy and (2) subdifferentiability, are established in discrete convex analysis in a suitable way. Furthermore, efficient (i.e., polynomial time) algorithms are also available for minimizing discrete convex functions.

We define a discrete DC function as a difference of two discrete convex functions. Since there are two classes of discrete convex functions (M$^\natural$-convex functions and L$^\natural$-convex functions), there are four types of discrete DC functions (an M$^\natural$-convex function minus an M$^\natural$-convex function, an M$^\natural$-convex function minus an L$^\natural$-convex function, and so on). These types of functions contain many functions appearing in practice: a difference of submodular functions [38] is an L$^\natural$−L$^\natural$ DC function, a supermodular function that is restricted to a matroid [3] is an M$^\natural$−L$^\natural$ DC function, and so on. Similarly to the continuous case, many discrete functions that arise in practice can be represented as a difference of two L$^\natural$-convex functions but there exists a function that is not a discrete DC function of other types.

We propose discrete DC programming problems as optimization problems of discrete DC functions:

$$\text{minimize } g(x) - h(x).$$

Since there are two conjugate classes ($M^\natural$ and $L^\natural$) of discrete convex functions, there are four types of discrete DC programs. We prove the discrete version of the Toland-Singer duality for discrete DC programs. The discrete Toland-Singer duality establishes the relation of four types of discrete DC programs, which is a main feature of discrete DC programming.

We also propose algorithms for discrete DC programming. These algorithms are obtained by combining the general discrete DC algorithm, which is a straightforward adaption of the continuous case, and the polyhedral structure of discrete convex functions. The algorithms decrease the function value strictly in each iteration and hence terminate in a finite number of iterations. Furthermore, when the algorithms terminate, the obtained solutions satisfy the local optimality condition. In some special case, the algorithm has a theoretical guarantee for the approximation ratio of the obtained solution.

Related work

There are only a few existing studies of discrete DC theory. Narasimhan and Bilmes [38] considered minimization problems of a difference of two submodular set functions (DS programs) and propose an algorithm, which is named submodular-supermodular procedure. As described later, the DS programming is a special case of our discrete DC programming (since submodular set functions coincide exactly with $L^\natural$-convex functions on $\{0,1\}^n$), and their algorithm is a special case of our general discrete DC algorithm (Section 5). Recently Iyer-Bilmes [20] proposed two algorithms, named supermodular-submodular procedure and modular-modular procedure for DS programs, and compared the performance of these three algorithms in numerical experiments.

Kawahara and Washio [22] recently proposed a prismatic algorithm for DS programming, which applies a branch-and-bound algorithm for continuous DC programs to the Lovász extension of submodular set functions. It is certainly an interesting problem to construct an enumerative algorithm for general discrete DC programs, but we do not persue this direction in this paper.

Kawahara, Nagano, and Okamoto [21] considered a fractional submodular programming problem, minimization of a ratio of two submodular functions. They applied the discrete Newton method to this problem and proved that the problem can be solved exactly by solving a polynomial number of DS programs.

Our contribution

Our contributions are summarized as follows:

- **Theoretical framework.** We define discrete DC functions and discrete DC programs, and develop a theory of discrete DC functions in parallel to the continuous DC theory.

- **DC representability.** We show that every function that has bounded Hessian is an $\mathrm{L}^\natural{-}\mathrm{L}^\natural$ DC function; but there exist discrete functions that are neither $\mathrm{M}^\natural{-}\mathrm{M}^\natural$, $\mathrm{M}^\natural{-}\mathrm{L}^\natural$, nor $\mathrm{L}^\natural{-}\mathrm{L}^\natural$ DC functions.

- **Local optimality.** We discuss local optimality conditions of discrete DC functions. In particular, for some subclasses of $\mathrm{L}^\natural{-}\mathrm{L}^\natural$ DC programs and $\mathrm{M}^\natural{-}\mathrm{L}^\natural$ DC programs, the local optimal solution has an approximation guarantee.

- **Algorithm.** We establish the general framework of discrete DC algorithm that is a direct translation of the continuous DC algorithm. We further propose algorithms tailored to $\mathrm{M}^\natural{-}\mathrm{M}^\natural$, $\mathrm{M}^\natural{-}\mathrm{L}^\natural$ and $\mathrm{L}^\natural{-}\mathrm{L}^\natural$ programs that exploit the polyhedral structure of discrete convex functions of respective types.

Organization of the paper

In Section 2, we review the basics of discrete convex analysis. In Section 3, discrete DC functions are introduced and some representability results are proved. In Section 4, discrete DC programming is introduced. A discrete version of the Toland-Singer duality and local optimality conditions are shown. In Section 5, a generic form of discrete DC algorithm and its ramifications for some types of DC programs exploiting the polyhedral structure of discrete convex functions are proposed.

2 Preliminary: Discrete convex analysis

2.1 Definitions

We introduce some basic notions from discrete convex analysis [35]. Let $\mathbb{Z}$ be the set of integers and let $[n] := \{1, \ldots, n\}$ for positive $n \in \mathbb{Z}$. The positive support of $x \in \mathbb{Z}^n$ is defined as

$$\mathrm{supp}^+(x) := \{i \in [n] : x(i) > 0\}.$$

The characteristic vector χ_S of $S \subseteq [n]$ is defined as

$$\chi_S(i) = \begin{cases} 1, & i \in S, \\ 0, & i \notin S. \end{cases}$$

We identify a set S and its characteristic vector. For simplicity, we write χ_i for $\chi_{\{i\}}$ for $i \in [n]$. Let $\mathbf{1} = (1, \ldots, 1) = \chi_{[n]} \in \{0, 1\}^n$. For a function $f : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$, its effective domain is defined by

$$\mathrm{dom}\, f := \{x \in \mathbb{Z}^n : f(x) < +\infty\}. \tag{2.1}$$

Two classes of discrete convex functions, M$^\natural$-convex functions and L$^\natural$-convex functions[1], are defined as follows. A function $f : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ is called M$^\natural$-convex if it satisfies the exchange axiom:

For all $x, y \in \mathbb{Z}^n$ and $i \in \mathrm{supp}^+(x - y)$,

$$f(x) + f(y) \geq \min \Bigg[f(x - \chi_i) + f(x + \chi_i), \min_{j \in \mathrm{supp}^+(y-x)} \{f(x - \chi_i + \chi_j) + f(y + \chi_i - \chi_j)\} \Bigg].$$

A function $f : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ is called L$^\natural$-convex if it satisfies the discrete midpoint convexity:

For all $x, y \in \mathbb{Z}^n$,

$$f(x) + f(y) \geq f(\left\lceil \frac{x+y}{2} \right\rceil) + f(\left\lfloor \frac{x+y}{2} \right\rfloor),$$

where $\lceil a \rceil$ denotes the smallest integer not less than a, and $\lfloor a \rfloor$ denotes the largest integer not greater than a. In this paper, we refer to M$^\natural$- or L$^\natural$-convex functions as *discrete convex functions*. It is emphasized that we consider integer-valued functions. Some authors (e.g., [6, 27, 39]) studied other discrete convex functions but we do not cover them.

[1] M$^\natural$-convex and L$^\natural$-convex functions are introduced by Fujishige and Murota [9] and Murota and Shioura [37], respectively, as variants of M-convex and L-convex functions; see [35]. M$^\natural$-concave functions on $\{0, 1\}^n$ coincide with valuated matroids of Dress and Wenzel [4]. We note that "M" stands for matroid, and "L" stands for lattice, and the symbol $\natural$ is to read "natural."

Example 2.1. Let $\mathcal{M}$ be a matroid on V and r be the rank function of $\mathcal{M}$. Then the function

$$f(x) = \begin{cases} -r(X), & x = \chi_X \\ +\infty, & \text{otherwise} \end{cases}$$

is an M$^\natural$-convex function with $\operatorname{dom} f = \{0,1\}^V$. In this connection, we also mention that an M$^\natural$-convex function on $\{0,1\}^n$ is essentially the negative of a matroid valuation on independent sets[2]. ■

Example 2.2. Let $\rho : 2^V \to \mathbb{Z}$ be a submodular set function [8], i.e.,

$$\rho(X) + \rho(Y) \geq \rho(X \cup Y) + \rho(X \cap Y), \quad X, Y \subseteq V.$$

Then the function

$$f(x) = \begin{cases} \rho(X), & x = \chi_X \\ +\infty, & \text{otherwise} \end{cases}$$

is an L$^\natural$-convex function with $\operatorname{dom} f = \{0,1\}^V$. Note that every L$^\natural$-convex function f is submodular on $\mathbb{Z}^n$, i.e.,

$$f(x \vee y) + f(x \wedge y) \leq f(x) + f(y).$$

(The converse is not true. Consider a univariate "non-convex" function $f : \mathbb{Z} \to \mathbb{Z}$, e.g., $f(x) = (-1)^{|x|}$. Then it is submodular but not L$^\natural$-convex.) ■

2.2 Optimality and subgradients

For a continuous convex function, a local minimum is also a global minimum. This property is shared by a discrete convex function.

Theorem 2.3 (Optimality condition of M$^\natural$-convex functions (Theorem 6.26 [35]))**.** Let $f : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ be an M$^\natural$-convex function. If $x \in \operatorname{dom} f$ satisfies

$$f(x) \leq \min_{i,j \in [n]} \{f(x - \chi_i), f(x + \chi_j), f(x - \chi_i + \chi_j)\}, \tag{2.2}$$

then x is a global minimum of f. ■

Theorem 2.4 (Optimality condition for L$^\natural$-convex functions (Theorem 7.14 [35]))**.** Let $f : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ be an L$^\natural$-convex function. If $x \in \operatorname{dom} f$ satisfies

$$f(x) \leq \min_{X \subseteq [n]} \{f(x + \chi_X), f(x - \chi_X)\}, \tag{2.3}$$

then x is a global minimum of f. ■

---

[2] The original definition [32] of matroid valuation on independent sets is a monotone M$^\natural$-concave function on $\{0,1\}^n$.

The local optimality conditions above are useful to derive an explicit representation of subgradients and subdifferentials. A vector $p \in \mathbb{Z}^n$ is a subgradient of f at $x \in \mathrm{dom}\, f$ if for all $y \in \mathbb{Z}^n$,

$$\langle p, y - x \rangle \leq f(y) - f(x). \tag{2.4}$$

The subdifferential $\partial f(x)$ of f at x is the set of all subgradients of f at x, i.e.,

$$\partial f(x) := \{p \in \mathbb{Z}^n : \langle p, y - x \rangle \leq f(y) - f(x) \ (\forall y \in \mathbb{Z}^n)\}. \tag{2.5}$$

It is emphasized that we consider integer vectors p as subgradients. Subgradients and subdifferentials are key ingredients in optimization algorithms. By definition, $p \in \partial f(x)$ if and only if $x \in \mathrm{argmin}_y\{f(y) - \langle p, y \rangle\}$, and in particular, $0 \in \partial f(x)$ if and only if x is a global minimum of f.

The subdifferentials of discrete convex functions turn out to be familiar object in combinatorial optimization. Since the sum of an M$^\natural$- (resp. L$^\natural$-) convex function and a linear function is also an M$^\natural$- (resp. L$^\natural$-)convex function, we obtain the following theorems by combining the local optimality conditions of discrete convex functions.

Theorem 2.5 (Subdifferential of M$^\natural$-convex functions (Theorem 6.61 [35])). Let $f : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ be an M$^\natural$-convex function. Then $\partial f(x) \neq \emptyset$ for all $x \in \mathrm{dom}\, f$. We have $p \in \partial f(x)$ if and only if, for all $i, j \in [n]$,

$$\begin{aligned} -p_i &\leq f(x - \chi_i) - f(x), \\ p_j &\leq f(x + \chi_j) - f(x), \\ p_j - p_i &\leq f(x - \chi_i + \chi_j) - f(x). \end{aligned} \tag{2.6}$$

■

The above theorem shows that the subgradients of an M$^\natural$-convex function are integer vectors contained the polytope arising from the dual of the shortest path problem [42]. For an L$^\natural$-convex function f, on the other hand, the subdifferential $\partial f(x)$ at x forms an integral generalized polymatroid [10, 11, 12, 35] as follows.

Theorem 2.6 (Subdifferential of L$^\natural$-convex functions (Theorem 7.43 [35])). Let $f : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ be an L$^\natural$-convex function. Then $\partial f(x) \neq \emptyset$ for all $x \in \mathrm{dom}\, f$. We have $p \in \partial f(x)$ if and only if, for all $X \subseteq [n]$,

$$f(x) - f(x - \chi_X) \leq \langle p, \chi_X \rangle \leq f(x + \chi_X) - f(x).$$

■

2.3 Conjugacy

One of the most important property of discrete convex functions is the Legendre-Fenchel conjugacy between M$^\natural$ and L$^\natural$. The discrete Legendre-Fenchel conjugate f^* of a function $f : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ is defined as follows:

$$f^*(p) := \sup_{x \in \mathbb{Z}^n} \{\langle p, x\rangle - f(x)\}, \quad p \in \mathbb{Z}^n.$$

Note that this defines a function $f^* : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$, provided $\mathrm{dom}\, f \neq \emptyset$. Then the following theorem holds.

Theorem 2.7 (Discrete conjugacy (Theorem 8.12 [35])). If f is M$^\natural$-convex, then f^* is L$^\natural$-convex. Similarly, if f is L$^\natural$-convex, then f^* is M$^\natural$-convex. Furthermore, if f is M$^\natural$- or L$^\natural$-convex, then $f^{**} = f$. ■

Corollary 2.8 (Discrete Young-Fenchel inequality). Let f be an M$^\natural$-convex or L$^\natural$-convex function. Then

$$f(x) + f^*(p) \geq \langle p, x\rangle \quad \text{for all } x, p \in \mathbb{Z}^n.$$

Equality holds if and only if $x \in \partial f^*(p)$. ■

The relation of global optimalities and subgradients with respect to the Legendre-Fenchel conjugacy is summarized as follows [35].

$$\begin{array}{ccc} p \in \underset{q \in \mathbb{Z}^n}{\mathrm{argmin}}\{f^*(q) - \langle q, x\rangle\} & \Longleftrightarrow & p \in \partial f(x) \\ & & \Updownarrow \\ x \in \underset{x \in \mathbb{Z}^n}{\mathrm{argmin}}\{f(y) - \langle p, y\rangle\} & \Longleftrightarrow & x \in \partial f^*(p) \end{array}$$

Figure 2.1: Relation of global optimalities, subgradients, with respect to the Legendre-Fenchel duality

The following theorem, discrete Fenchel duality, is a fundamental theorem in discrete convex analysis. See [35] for details.

Theorem 2.9 (Discrete Fenchel duality (Theorem 8.21 [35])). Let f and g be both M$^\natural$-convex functions or both L$^\natural$-convex functions. Then

$$\inf_{x \in \mathbb{Z}^n} \{f(x) + g(x)\} = - \inf_{p \in \mathbb{Z}^n} \{f^*(p) + g^*(-p)\}. \tag{2.7}$$

■

Remark 2.10. In the continuous DC theory, subdifferentiability: $\partial f(x) \neq \emptyset$ for $x \in \mathrm{dom}\, f$ and biconjugacy: $f^{**} = f$ play crucial roles. Hence, to construct a discrete analogue of the DC theory, we need a class of "convex functions" on $\mathbb{Z}^n$ that have these properties. As mentioned in this section, M$^\natural$- and L$^\natural$-convex functions do have these properties. This is the reason why we employ these classes of functions in developing a discrete DC theory.

We here illustrate that subdifferentiability and biconjugacy are nontrivial or even rare in discrete case by showing a concrete example of a "convex function" f such that $\partial f(x) = \emptyset$ for some $x \in \mathrm{dom}\, f$ and $f^{**} \neq f$. This example is taken from [33].

Let $D = \{(0,0,0), \pm(1,1,0), \pm(0,1,1), \pm(1,0,1)\}$ and $f : \mathbb{Z}^3 \to \mathbb{Z} \cup \{+\infty\}$ be defined by

$$f(x_1, x_2, x_3) := \begin{cases} (x_1 + x_2 + x_3)/2, & x \in D, \\ +\infty, & \text{otherwise.} \end{cases}$$

This function can be naturally extended to a convex function on $\mathrm{conv}(D)$ (the convex hull of D) and D is a "convex set" in the sense of $\mathrm{conv}(D) \cap \mathbb{Z}^n = D$.

We first calculate the subgradient of f at the origin. Suppose $p \in \partial f(0) \subseteq \mathbb{Z}^3$. Since $f(y) - f(0) \geq \langle p, y \rangle$ for all y, we must have

$$\begin{array}{lll} 1 \geq p_1 + p_2, & 1 \geq p_2 + p_3, & 1 \geq p_3 + p_1, \\ -1 \geq -p_1 - p_2, & -1 \geq -p_2 - p_3, & -1 \geq -p_3 - p_1. \end{array}$$

However, this system does not admit an integer solution, although it is satisfied by $(p_1, p_2, p_3) = (1/2, 1/2, 1/2)$. Hence $\partial f(0) = \emptyset$.

Next we calculate the biconjugate of f. The conjugate function of f is

$$f^*(p) = \max\{0, |p_1 + p_2 - 1|, |p_2 + p_3 - 1|, |p_3 + p_1 - 1|\}$$

and the biconjugate is

$$f^{**}(x) = \sup_{p \in \mathbb{Z}^3} \{\langle p, x \rangle - f^*(p)\}.$$

Hence

$$f^{**}(0) = -\inf_{p \in \mathbb{Z}^3} \max\{0, |p_1 + p_2 - 1|, |p_2 + p_3 - 1|, |p_3 + p_1 - 1|\}.$$

Therefore we have $f^{**}(0) = -1 \neq 0 = f(0)$. This shows $f^{**} \neq f$. ■

Remark 2.11. There are some possible candidates for a class of "discrete convex functions," but many of them lack biconjugacy and/or subdifferentiability. In particular, integrally convex functions, proposed by Favati and Tardella [6], are a reasonable candidate, but do not have these properties. The function f demonstrated in Remark 2.10 is, in fact, an integrally convex function.

2.4 Algorithms

We close this section with some algorithmic aspects of discrete convex analysis. Suppose that we can only access to functions by evaluating the value of the functions. This model is called value oracle model (e.g., [51]).

Function minimization Since the local optimality of discrete convex functions can be efficiently checked, a global minimum of discrete convex functions can be efficiently computed by descent type methods with scaling techniques. Note also that we can evaluate the conjugate function $f^*(p)$ by minimizing $f(x) - \langle p, x \rangle$.

Theorem 2.12 (M$^\natural$-convex minimization [35])**.** M$^\natural$-convex minimization can be done with $O(n^3 \log(K_\infty/n))$ function evaluations, where $K_\infty := \max\{\|x-y\|_\infty : x, y \in \mathrm{dom} f\}$. ■

Theorem 2.13 (L$^\natural$-convex minimization [23, 35])**.** L$^\natural$-convex minimization can be done with $O(\sigma(n) \log(\hat{K}_\infty/n))$ function evaluations, where $\sigma(n)$ is the number of function evaluations in submodular set function minimization, and $\hat{K}_\infty := \max\{\|x-y\|_\infty : x, y \in \mathrm{dom} f\}$. ■

The current best complexity $\sigma(n)$ of submodular set function minimization is $O(n^5)$ by Iwata and Orlin [19] or Orlin [40].

Subgradient computation For an M$^\natural$-convex function f, the subdifferential $\partial f(x)$ at x corresponds to a polyhedron (2.6) that appears in the dual of the shortest path problem. Therefore we can obtain a subgradient p by solving a shortest path problem. More concretely, consider a weighted digraph $G = (V, E, w)$, where $V = \{0\} \cup [n]$ and $E = V \times V$, and the edge length w defined as

$$\begin{aligned} w(0, j) &= f(x + \chi_j) - f(x), \\ w(i, 0) &= f(x - \chi_i) - f(x), \\ w(i, j) &= f(x + \chi_j - \chi_i) - f(x). \end{aligned}$$

As a consequence of M$^\natural$-convexity of f, this edge length w satisfies the triangle inequality and hence the graph does not have a negative cycle [35]. Therefore there exists a feasible potential p that satisfies (2.6). We can obtain such potential p explicitly by computing the shortest paths on this graph with $O(n^2)$ function evaluations.

For an L$^\natural$-convex function f, the subdifferential $\partial f(x)$ at x forms an integral generalized polymatroid. Therefore we can obtain a subgradient p by the following simple method [5, 11]. Let π be a permutation of $[n]$ and let $k \in [n] \cup \{0\}$. Then the vector $p_{\pi,k} \in \mathbb{Z}^n$ defined by

$$\begin{aligned} p_{\pi,k}(\pi(j)) &:= f(x + \chi_{\{\pi(1),\ldots,\pi(j)\}}) - f(x + \chi_{\{\pi(1),\ldots,\pi(j-1)\}}), \quad 1 \leq j \leq k, \\ p_{\pi,k}(\pi(j)) &:= f(x - \chi_{\{\pi(j),\ldots,\pi(n)\}}) - f(x - \chi_{\{\pi(j+1),\ldots,\pi(n)\}}), \quad k < j \leq n \end{aligned} \tag{2.8}$$

is a subgradient of f at x, which can be computed with $O(n)$ function evaluations. Every $p \in \partial f(x)$ is obtained in this way.

3 Discrete DC functions

3.1 Definition and examples

Let us say that a function $f : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty, -\infty\}$ is a *discrete DC function* if it can be written as a difference of two discrete convex functions: $f = g - h$ for some discrete ($\mathrm{M}^\natural$- or $\mathrm{L}^\natural$-)convex functions $g, h : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$. We use the convention that $(+\infty) - (+\infty) = (+\infty)$.

We call an expression $f = g - h$ a *discrete DC representation* of f. For a discrete DC function f, its discrete DC representation is not unique, i.e., there are many possibilities of representing $f = g_1 - h_1 = g_2 - h_2 = \cdots$. It is worth noting that a function f is a discrete DC function if and only if there exists a function h such that both $f + h$ and h are discrete convex. We refer to such function h as a *control function* of f.

Since there are two classes of convex functions, $\mathrm{M}^\natural$-convex functions and $\mathrm{L}^\natural$-convex functions, we can distinguish four types of discrete DC functions: $\mathrm{M}^\natural$−$\mathrm{M}^\natural$ DC functions, $\mathrm{M}^\natural$−$\mathrm{L}^\natural$ DC functions, $\mathrm{L}^\natural$−$\mathrm{M}^\natural$ DC functions, and $\mathrm{L}^\natural$−$\mathrm{L}^\natural$ DC functions, depending on the types of g and h in $f = g - h$.

Example 3.1 (Cut function of weighted graph)**.** Let $G = (V, E)$ be a graph and $w : E \to \mathbb{Z}$ be a (not necessarily nonnegative) weight function on the edge set E. A cut S is a nonempty proper subset of V and the value of a cut function $c : 2^V \to \mathbb{Z}$ is the sum of the weights of the edges crossing a cut. If all weights are nonnegative, c is a submodular (i.e., $\mathrm{L}^\natural$-convex) function. Therefore, if we split the edges to negative ones and positive ones, we obtain a representation of a cut function as an $\mathrm{L}^\natural$−$\mathrm{L}^\natural$ DC function. ∎

Example 3.2 (Difference of submodular functions)**.** A function which is the difference of two submodular functions naturally arise in the area of machine learning. One typical example is the mutual information $f(X) := I(X|C) = H(X) - H(X|C)$, which is a difference of two entropies. Since the entropy function H is submodular, the mutual information is a difference of two submodular functions. See [20, 38] for more applications in the area of machine learning. Note that, the class of such functions coincide exactly with $\mathrm{L}^\natural$−$\mathrm{L}^\natural$ DC functions on $\{0, 1\}^n$. ∎

Example 3.3. Let δ be the indicator function of the family of independent sets of a matroid $\mathcal{M}$ and f be a submodular set function. Then $\delta - f$ is an $\mathrm{M}^\natural$−$\mathrm{L}^\natural$ DC function which corresponds to a supermodular function restricted to the independent sets of a matroid.

Example 3.4 (Degree-determinant of polynomial matrix)**.** Let $A = A(s)$ be a polynomial matrix of size $m \times n$ ($m \le n$) in variable s. Let $A[I]$ denote the submatrix induced by a column index set $I \subseteq [n]$. Then $g(I) := \deg(\det A[I])$ is an $\mathrm{M}^\natural$-concave function on $\{I \subseteq [n] : |I| = m\}$; see [34] for more details.

For two polynomial matrices $A(s)$ and $B(s)$ of size $m \times n$ ($m \le n$), $\deg(\det A[I]/\det B[I])$ is an $\mathrm{M}^\natural$−$\mathrm{M}^\natural$ DC function on $\{I \subseteq [n] : |I| = m\}$. ∎

3.2 Representability

We turn to representability of discrete DC functions.

Proposition 3.5. Every lower-bounded discrete DC function $f = g - h$ admits a nonnegative discrete DC representation of the same type.

Proof. The proof is shown in later (Section 4.2). □

In the continuous case, many functions that arise in practice are DC functions [18, 50]. In particular, the following characterization is useful.

Theorem 3.6 (Classical result, e.g., p. 47 in Hiriart-Urruty [15])**.** Every C^2-function with bounded Hessian is a (continuous) DC function[3].

Proof. Let $h(x) := \mu\|x\|^2$ with a sufficiently large $\mu > 0$. Then $f = (f + h) - h$ is a discrete DC representation of f since $f + h$ is convex because the Hessian of $f + h$ is positive definite for a sufficiently large μ. □

In the discrete case, a similar theorem holds for L♮–L♮ DC functions (Theorem 3.8). However, for the other classes of discrete DC functions, this is not the case.

L♮–L♮ DC representability To prove the representability as an L♮–L♮ DC function, we trace the proof of Theorem 3.6 with the aid of discrete convex analysis. The discrete L♮-Hessian matrix $H(x) := [H_{ij}(x)]$ of a function $f : \mathbb{Z}^n \to \mathbb{Z}$ is defined as follows [28, 29]:

$$H_{ij}(x) := f(x + e_i + e_j) - f(x + e_i) - f(x + e_j) + f(x) \quad (i \neq j),$$
$$H_{ii}(x) := f(x) + f(x + \mathbf{1} + e_i) - f(x + \mathbf{1}) - f(x + e_i) - \sum_{j \neq i} H_{ij}(x).$$

Theorem 3.7 (Hessian characterization of L♮-convexity [28])**.** A function $f : \mathbb{Z}^n \to \mathbb{Z}$ is L♮-convex if and only if the L♮-Hessian H of f satisfies, for every $x \in \mathbb{Z}^n$,

$$H_{ij}(x) \leq 0 \ (i \neq j), \quad \sum_{j=1}^{n} H_{ij}(x) \geq 0 \ (\forall i \in [n]). \tag{3.1}$$

■

Using the L♮-Hessian, we can prove L♮–L♮ DC representability of an arbitrary (well-behaved) function.

[3] Usually this theorem is used in combination of Hartman's theorem [13] to prove the statement that "every C^2 function is a continuous DC function." See Hartman [13], Hiriart-Urruty [15] or Tuy [50].

Theorem 3.8. If a function $f : \mathbb{Z}^n \to \mathbb{Z}$ has bounded L$^\natural$-Hessian, then f is an L$^\natural$–L$^\natural$ DC function.

Proof. Let $H(x)$ be the L$^\natural$-Hessian of f at $x \in \mathbb{Z}^n$. Suppose $|H_{ij}(x)| \leq \alpha$ for all $x \in \mathbb{Z}^n$ and $i, j \in [n]$. Then it is easy to verify that the function $h(x) := (n\alpha/2)x^\top Ax$ is a control function of f, where

$$A = \begin{bmatrix} n & -1 & \cdots & -1 \\ -1 & n & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n \end{bmatrix}.$$

Indeed, the L$^\natural$-Hessian of h is equal to $n\alpha A$, which satisfies (3.1), and the L$^\natural$-Hessian of $f + h$ is equal to $H + n\alpha A$, which also satisfies (3.1) by the choice of α. □

The above theorem is a generalization of the classical result that every set function is a difference of submodular set functions. The control function h used in the above proof gives $h(\chi_X) = (n\alpha/2)\,(|X|(n - |X|) + |X|)$ for $X \subseteq [n]$.

Corollary 3.9.

(1) Let $D \subset \mathbb{Z}^n$ be a finite subset. Then for every function $f : D \to \mathbb{Z}$, there exists nonnegative L$^\natural$-convex functions $g, h : \mathbb{Z}^n \to \mathbb{Z}$ such that $f(x) = g(x) - h(x)$ for every $x \in D$.

(2) Every set function $f : \{0, 1\}^n \to \mathbb{Z}$ is a difference of nonnegative monotone submodular functions.

Proof. (1) Let $\bar{f} : \mathbb{Z}^n \to \mathbb{Z}$ be defined by $\bar{f}(x) = f(x)$ $(x \in D)$ and $\bar{f}(x) = 0$ $(x \notin D)$. Since the L$^\natural$-Hessian of $\bar{f}$ is bounded, $\bar{f}$ is an L$^\natural$–L$^\natural$ DC function by Theorem 3.8. Since $\bar{f}$ is lower bounded, $\bar{f}$ admits a nonnegative L$^\natural$–L$^\natural$ representation: $\bar{f} = g - h$ by Proposition 3.5. By restricting the functions onto D, we obtain (1).

(2) By (1) of this corollary, we have nonnegative submodular set functions $g, h : \{0, 1\}^n \to \mathbb{Z}$ such that $f = g - h$. Note that L$^\natural$-convex functions on $\{0, 1\}^n$ are nothing but submodular set functions. We can modify g and h to monotone functions as follows. Let

$$\alpha = \min_{i \in [n]}\{\min\{0, g(\chi_{[n]}) - g(\chi_{[n]\setminus\{i\}}), h(\chi_{[n]}) - h(\chi_{[n]\setminus\{i\}})\}\}.$$

Then both $g(x) - \alpha\langle \mathbf{1}, x\rangle$ and $h(x) - \alpha\langle \mathbf{1}, x\rangle$ are nonnegative monotone and hence $f(x) = (g(x) - \alpha\langle \mathbf{1}, x\rangle) - (h(x) - \alpha\langle \mathbf{1}, x\rangle)$ is a desired representation. Note that $\mathbf{1} = \chi_{[n]}$ and $\langle \mathbf{1}, x\rangle = |\operatorname{supp}^+(x)|$ for $x \in \{0, 1\}^n$. □

We mention that Iyer and Bilmes [20] recently proved a proposition similar to our Corollary 3.9 (2). But our proof technique is different from theirs.

M$^\natural$−M$^\natural$ DC representability We here show the existence of a non-M$^\natural$−M$^\natural$ DC function. To prove this, we use the Hessian characterization of M$^\natural$-convex functions. The discrete M$^\natural$-Hessian matrix $H(x) := [H_{ij}(x)]$ of a function $f : \mathbb{Z}^n \to \mathbb{Z}$ is defined as follows [14, 29]:

$$H_{ij}(x) := f(x + e_i + e_j) - f(x + e_i) - f(x + e_j) + f(x), \quad i, j \in [n].$$

Theorem 3.10 (Hessian characterization of M$^\natural$-convex functions [14, 36])**.** A function $f : \mathbb{Z}^n \to \mathbb{Z}$ is M$^\natural$-convex if and only if the M$^\natural$-Hessian H of f satisfies, for every $x \in \mathbb{Z}^n$,

$$H_{ij}(x) \geq 0 \ (\forall i, j), \quad H_{ij}(x) \geq \min\{H_{ik}(x), H_{jk}(x)\} \ (i \neq k, j \neq k). \quad (3.2)$$

■

Proposition 3.11. There exists a function that is not an M$^\natural$−M$^\natural$ DC function.

Proof. We verify that $f(x, y, z, w) := xz + xw + yz$ is not an M$^\natural$−M$^\natural$ DC function. The M$^\natural$-Hessian of f (at the origin) is given by

$$H_f = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Suppose that h is a control function of f, i.e., both $f+h$ and h are M$^\natural$-convex. Denote the M$^\natural$-Hessian of h by

$$H_h = \begin{bmatrix} * & a & p & q \\ a & * & r & b \\ p & r & * & c \\ q & b & c & * \end{bmatrix}.$$

We show that, for any choice of the diagonal elements of H_h, there is no $\{a, b, c, p, q, r\}$ such that both $H_{f+h} = H_f + H_h$ and H_h satisfy (3.2).

We first show that a, b, c cannot be the minimum in $\{a, b, c, p, q, r\}$. Since $a \geq \min\{p+1, r+1\}$ by (3.2) for H_{f+h}, a cannot be the minimum. Similarly, since $c \geq \min\{p + 1, q + 1\}$ by (3.2) for H_{f+h}, c cannot be the minimum. Suppose that b is the minimum. Then, since $b \geq \min\{a, q\}$ by (3.2) for H_h and a cannot be the minimum, we must have $b = q$. However, since $b \geq \min\{a, q+1\}$ by (3.2) for H_{f+h}, we must have $b = a$. This is a contradiction.

Next, we show that p, q, r cannot be the minimum in $\{a, b, c, p, q, r\}$. Since $q \geq \min\{a, b\}$ and neither a nor b is the minimum, q cannot be the minimum. Similarly, since $r \geq \min\{b, c\}$ and neither b nor c is the minimum, r cannot be the minimum. Then, since $p \geq \min\{a, r\}$ and neither a nor r is the minimum, p cannot be the minimum. □

Remark 3.12. For the function $f(x, y, z, w)$ used in the proof of Proposition 3.11, we have the expression

$$xz + xw + yz = \frac{1}{2}\left((x+z)^2 + (x+w)^2 + (y+z)^2 - 2x^2 - 2z^2 - w^2 - y^2\right),$$

where each term on the right-hand side above is $\mathrm{M}^\natural$-convex or $\mathrm{M}^\natural$-concave. This shows that the class of $\mathrm{M}^\natural$–$\mathrm{M}^\natural$ DC functions is not closed under sum. This is related to the fact that the class of $\mathrm{M}^\natural$-convex functions is not closed under sum. On the other hand, the set of $\mathrm{L}^\natural$–$\mathrm{L}^\natural$ DC functions forms a linear space since a sum of $\mathrm{L}^\natural$-convex functions is $\mathrm{L}^\natural$-convex.

In the continuous case, the set of DC functions is closed under sum, multiplication, scalar-product, min, and max. In other words, continuous DC functions form an algebra and a lattice. ■

$\mathbf{M}^\natural$–$\mathbf{L}^\natural$ and $\mathbf{L}^\natural$–$\mathbf{M}^\natural$ representability Since an $\mathrm{M}^\natural$-convex function $g : \mathbb{Z}^n \to \mathbb{Z}$ is supermodular and an $\mathrm{L}^\natural$-convex function $h : \mathbb{Z}^n \to \mathbb{Z}$ is submodular, an $\mathrm{M}^\natural$–$\mathrm{L}^\natural$ DC function $f = g - h$ with $\mathrm{dom}\, f = \mathbb{Z}^n$ is necessarily supermodular on $\mathbb{Z}^n$. Hence a function that is not supermodular is not an $\mathrm{M}^\natural$–$\mathrm{L}^\natural$ DC function. Similarly, a function that is not submodular is not an $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC function.

We pose the following as an open problem.

Problem 3.13. Give a necessary and sufficient condition for a function to be an $\mathrm{M}^\natural$–$\mathrm{M}^\natural$, $\mathrm{M}^\natural$–$\mathrm{L}^\natural$ or $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC function.

4 Discrete DC programming

4.1 General framework and examples

Discrete DC programming is a framework of minimization of a discrete DC function $f = g - h$:

$$\text{minimize } g(x) - h(x), \quad \text{for } x \in \mathbb{Z}^n. \tag{4.1}$$

We call this problem *discrete DC programming problem.* We distinguish four classes, according to the four types of discrete DC functions: M$^\natural$−M$^\natural$ DC programming, M$^\natural$−L$^\natural$ DC programming, L$^\natural$−M$^\natural$ DC programming, and L$^\natural$−L$^\natural$ DC programming.

Example 4.1 (Minimum cut with some negative weight edges)**.** A minimum cut problem of a graph with possibly negative weight edges is an L$^\natural$−L$^\natural$ DC programming problem. See Example 3.1. ■

Example 4.2 (DS programming)**.** Minimization problem of a difference of submodular set functions is exactly L$^\natural$−L$^\natural$ DC programming problem on $\{0, 1\}^n$. See Example 3.2. ■

Example 4.3 (Matroid constrained submodular maximization)**.** Let g be the indicator function of the family of independent sets of a matroid and h be a submodular set function. Then minimizing $g - h$, which is a matroid constrained submodular maximization [7], is an M$^\natural$−L$^\natural$ DC programming problem. See Example 3.3. ■

Example 4.4 (Minimization of degree of determinant)**.** Let $A = A(s)$ be a polynomial matrix of size $m \times n$ $(m \leq n)$ in variable s and let B be a constant matrix of size $m' \times n$. Then the following problem is an M$^\natural$−M$^\natural$ DC programming problem:

$$\begin{array}{ll} \text{minimize} & \deg(\det(A[I])) \\ \text{subject to} & I \subseteq [n],\ |I| = m, \text{ columns of } B[I] \text{ are linear independent.} \end{array}$$

Note that, the maximization counterpart:

$$\begin{array}{ll} \text{maximize} & \deg(\det(A[I])) \\ \text{subject to} & I \subseteq [n],\ |I| = m, \text{ columns of } B[I] \text{ are linear independent} \end{array}$$

is a problem of maximizing a sum of two M$^\natural$-concave functions, which can be solved efficiently by valuated matroid intersection algorithms [30, 31, 34]. ■

Example 4.5 (Fractional discrete convex programming)**.** Let g and h be positive-valued discrete convex functions. Then

$$\inf_{x \in \mathbb{Z}^n} \frac{g(x)}{h(x)} \geq \alpha \iff \inf_{x \in \mathbb{Z}^n} \{g(x) - \alpha h(x)\} \geq 0.$$

Therefore we can solve the fractional discrete convex programming problem by solving a sequence of discrete DC programming problems. ■

4.2 Toland-Singer duality

The most important theorem in DC programming (in the continuous case) is the Toland-Singer duality [44, 47]:

$$\inf_{x\in\mathbb{R}^n}\{g(x)-h(x)\} = \inf_{p\in\mathbb{R}^n}\{h^*(p)-g^*(p)\}. \tag{4.2}$$

We first establish the discrete version of this theorem. The identity of the Toland-Singer duality should not be confused with the Fenchel duality (2.7) in Theorem 2.9.

Theorem 4.6 (Discrete Toland-Singer duality)**.** Let $f = g - h$ be a discrete DC function of either type ($\mathrm{M}^\natural$–$\mathrm{M}^\natural$, $\mathrm{M}^\natural$–$\mathrm{L}^\natural$, $\mathrm{L}^\natural$–$\mathrm{M}^\natural$, $\mathrm{L}^\natural$–$\mathrm{L}^\natural$). Then

$$\inf_{x\in\mathbb{Z}^n}\{g(x)-h(x)\} = \inf_{p\in\mathbb{Z}^n}\{h^*(p)-g^*(p)\}. \tag{4.3}$$

Proof. By virtue of the conjugacy (Theorem 2.7) in discrete convex analysis, the proof goes in parallel with that for the continuous version of the Toland-Singer duality [44, 47]:

$$\begin{aligned}
\inf_x\{g(x)-h(x)\} &= \inf_x\{g(x)-h^{**}(x)\}\\
&= \inf_x\{g(x)-\sup_p\{\langle p,x\rangle - h^*(p)\}\}\\
&= \inf_x\inf_p\{g(x)-\langle p,x\rangle + h^*(p)\}\\
&= \inf_p\{h^*(p)-\sup_x\{\langle p,x\rangle - g(x)\}\}\\
&= \inf_p\{h^*(p)-g^*(p)\}.
\end{aligned}$$

□

It is emphasized that the key of the above proof is the biconjugacy, $h^{**} = h$, that holds for both $\mathrm{M}^\natural$- and $\mathrm{L}^\natural$-convex functions (Theorem 2.7).

Let us dwell on the issue of $\mathrm{M}^\natural/\mathrm{L}^\natural$ conjugacy in the discrete Toland-Singer duality. Recall from Theorem 2.7 that the Legendre-Fenchel conjugate of an $\mathrm{M}^\natural$-convex function is an $\mathrm{L}^\natural$-convex function and vice versa. For an $\mathrm{M}^\natural$–$\mathrm{M}^\natural$ DC program for $g - h$, for example, the discrete Toland-Singer duality establishes its relation to an $\mathrm{L}^\natural$–$\mathrm{L}^\natural$ DC program for $h^* - g^*$, and vice versa. For an $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC program for $g - h$, in contrast, the dual problem is another $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC program for $h^* - g^*$. Similarly, for an $\mathrm{M}^\natural$–$\mathrm{L}^\natural$ DC program for $g - h$, the dual problem is another $\mathrm{M}^\natural$–$\mathrm{L}^\natural$ DC program for $h^* - g^*$. In this sense, the class of $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC programs and that of $\mathrm{M}^\natural$–$\mathrm{L}^\natural$ DC programs are both self-dual. It is worth mentioning that replacing g and h with their conjugates g^* and h^*, respectively, in (4.3) results in the formula

$$\inf_{x\in\mathbb{Z}^n}\{g^*(x)-h^*(x)\} = \inf_{p\in\mathbb{Z}^n}\{h^{**}(p)-g^{**}(p)\},$$

which is equivalent, by biconjugacy, to (4.3) with g and h interchanged. In this sense, the discrete Toland-Singer duality is self-dual under conjugacy. Such richer duality structure, which we may refer to as "typed duality," is the main feature of discrete DC programming that is not shared by continuous DC programming.

In the rest of the paper, we consider the case that the problem (4.1) has an optimal solution with a finite optimal value, i.e., $\inf_x\{g(x)-h(x)\} > -\infty$. In this case, by the Toland-Singer duality, we also have $\inf_p\{h^*(p)-g^*(p)\} > -\infty$. These finiteness conditions imply, in particular, that

$$\operatorname{dom} g \subseteq \operatorname{dom} h \quad \text{and} \quad \operatorname{dom} h^* \subseteq \operatorname{dom} g^*. \tag{4.4}$$

We assume these inclusions in the rest of the paper.

As an application of the Toland-Singer duality, we give here a proof of Proposition 3.5.

Proof of Proposition 3.5. The proof goes in parallel with that for the continuous version by Hiriart-Urruty [15].

Without loss of generality we assume $f \geq 0$; otherwise we may replace f by $f - \inf f$. Let $g_1(x) := g(x) - (\langle b, x\rangle - h^*(b))$ and $h_1(x) := h(x) - (\langle b, x\rangle - h^*(b))$ with $b \in \operatorname{dom} h^*$. Then $g_1 - h_1 = g - h$, $h_1(x) \geq 0$ by the Young-Fenchel inequality (Corollary 2.8), and $\min g_1(x) = -g^*(b) + h^*(b) \geq \min_p\{h^*(p) - g^*(p)\} = \min_x f(x) \geq 0$ by the Toland-Singer duality. □

4.3 Hardness results

In the continuous case, DC programming is known to be hard. Hence we expect that the discrete DC programming is also hard. Indeed, we can prove the following proposition.

Proposition 4.7.

- M$^\natural$–M$^\natural$, M$^\natural$–L$^\natural$, and L$^\natural$–L$^\natural$ DC programming on $\mathbb{Z}^n$ are NP-hard.
- M$^\natural$–L$^\natural$ and L$^\natural$–L$^\natural$ DC programming on $\{0,1\}^n$ are NP-hard, and L$^\natural$–M$^\natural$ DC programming on $\{0,1\}^n$ is in P.

Proof. M$^\natural$–L$^\natural$ and L$^\natural$–L$^\natural$ DC programs on $\{0,1\}^n$ contain submodular set function maximization, which is NP-hard, and M$^\natural$–M$^\natural$ DC programming is the Toland-Singer dual of L$^\natural$–L$^\natural$ DC programming. L$^\natural$–M$^\natural$ DC programming on $\{0,1\}^n$ is a submodular set function minimization, which is in P. □

Remark 4.8. L$^\natural$–M$^\natural$ DC programming problem is a special case of submodular function minimization problem on a distributive lattice [48]. Hence it can be solved in polynomial time in the length of a maximal chain in the lattice [35, 41]. See Section 5.5. ■

Table 4.1: Complexity of discrete DC programming $\min_x\{g(x)-h(x)\}$

(a) $x \in \mathbb{Z}^n$

$g\backslash h$	M$^\natural$	L$^\natural$
M$^\natural$	NP-hard	NP-hard
L$^\natural$	open	NP-hard

(b) $x \in \{0,1\}^n$

$g\backslash h$	M$^\natural$	L$^\natural$
M$^\natural$	open	NP-hard
L$^\natural$	P	NP-hard

4.4 Optimality criteria

We first state a necessary and sufficient condition for the global optimality of discrete DC function, which is a discrete version of the well-known theorem by Hiriart-Urruty [16].

For $\epsilon \geq 0$ we define the (integral) ϵ-subdifferential [17] of g at x by

$$\partial_\epsilon g(x) := \{p \in \mathbb{Z}^n : \langle p, y-x\rangle \leq g(y) - g(x) + \epsilon \ (\forall y \in \mathbb{Z}^n)\}. \tag{4.5}$$

Note that $p \in \partial_\epsilon g(x)$ if and only if

$$g(x) + g^*(p) \leq \langle p, x\rangle + \epsilon. \tag{4.6}$$

Since g is assumed to be integer-valued and p is an integer vector, we have $\partial_\epsilon g(x) = \partial_{\epsilon'} g(x)$ for $\epsilon' = \lceil \epsilon \rceil$. Therefore, we may assume $\epsilon \in \mathbb{Z}$. If g is a discrete convex function of either type (M$^\natural$ or L$^\natural$), we have

$$\partial_\epsilon g(x) \neq \emptyset \quad \text{for every } \epsilon \geq 0$$

for every $x \in \operatorname{dom} f$ by Theorems 2.5 and 2.6, as well as the inclusion $\partial_\epsilon g(x) \supseteq \partial g(x)$.

Proposition 4.9. $x \in \mathbb{Z}^n$ is a global minimum of a discrete DC function $f = g - h$ if and only if, for every $\epsilon \geq 0$,

$$\partial_\epsilon h(x) \subseteq \partial_\epsilon g(x). \tag{4.7}$$

Proof. Suppose that x is a minimum of $g-h$. Then, by the discrete Toland-Singer duality (Theorem 4.6), for all $p \in \operatorname{dom} h^*$,

$$g(x) - h(x) \leq h^*(p) - g^*(p), \qquad \text{i.e.,} \qquad g(x) + g^*(p) \leq h(x) + h^*(p).$$

Therefore, by (4.6) for g and h, we see that $p \in \partial_\epsilon h(x)$ implies $p \in \partial_\epsilon g(x)$, i.e., $\partial_\epsilon h(x) \subseteq \partial_\epsilon g(x)$.

Conversely, suppose that x is not a global minimum. Then, by the discrete Toland-Singer duality, there exists $p \in \operatorname{dom} h^*$ such that

$$g(x) - h(x) > h^*(p) - g^*(p),$$

i.e.,

$$g(x) + g^*(p) - \langle p, x\rangle > h(x) + h^*(p) - \langle p, x\rangle.$$

By the discrete Young-Fenchel inequality (Corollary 2.8), the right-hand side is nonnegative. Hence we can take $\epsilon = h(x) + h^*(p) - \langle p, x \rangle$ to meet

$$g(x) + g(p) > \langle p, x \rangle + \epsilon \geq h(x) + h^*(p).$$

This shows $p \in \partial_\epsilon h(x)$ but $p \notin \partial_\epsilon g(x)$, i.e., $\partial_\epsilon h(x) \not\subseteq \partial_\epsilon g(x)$. □

Checking the condition $\partial_\epsilon h(x) \subseteq \partial_\epsilon g(x)$ $(\forall \epsilon \geq 0)$ in (4.7) for global optimality is difficult, even in the continuous case [46]. This is not surprising, since (discrete) DC programming problem is a non-convex optimization problem. Hence we may reasonably focus on a local optimal solution.

Instead of requiring the condition $\partial_\epsilon h(x) \subseteq \partial_\epsilon g(x)$ for all $\epsilon \geq 0$, we consider a special case with $\epsilon = 0$, i.e.,

$$\partial h(x) \subseteq \partial g(x). \tag{4.8}$$

This will turn out to be a fruitful compromise; this condition guarantees a certain local optimality (Proposition 4.11), and it is amenable to algorithmic verification for some types of discrete DC programming (Section 5). In this paper, we refer to (4.8) as the *local optimality condition* for $f = g - h$.

We start with a technical lemma.

Lemma 4.10. Let $f = g - h$ be a discrete DC function and $U \subseteq \mathbb{Z}^n$ be a set containing x. If $\partial g(x) \cap \partial h(y) \neq \emptyset$ for every $y \in U$, then x is a minimum of f in U.

Proof. This proof goes in parallel with the proof of Corollary 1 of Tao and Hoai An [46]. Let $y \in U$ and pick $p \in \partial g(x) \cap \partial h(y)$. Then we have

$$\langle p, y - x \rangle \leq g(y) - g(x),$$
$$\langle p, x - y \rangle \leq h(x) - h(y).$$

Adding these two we obtain

$$g(x) - h(x) \leq g(y) - h(y).$$

□

The local optimality condition (4.8) does imply local optimality in a certain neighborhood specified by (4.9) below.

Proposition 4.11 (Local optimality)**.** If $x \in \mathbb{Z}^n$ satisfies (4.8), then x is a minimum of $f = g - h$ in

$$U = \bigcup \{\partial h^*(p) : p \in \partial g(x)\}. \tag{4.9}$$

Proof. For each $y \in U$, $y \in \partial h^*(p)$ for some $p \in \partial g(x)$. Since $y \in \partial h^*(p)$ is equivalent to $p \in \partial h(y)$, we obtain $p \in \partial h(y) \cap \partial g(x) \neq \emptyset$. Then $f(x) \leq f(y)$ by Lemma 4.10. □

Example 4.12. Consider a discrete DC function $f(x) = x^4 - 16(x-1)^2$ $(x \in \mathbb{Z})$, i.e., $f = g - h$ where $g(x) = x^4$ and $h(x) = 16(x-1)^2$. For $x = 2$, we have $\partial h(x) \subseteq \partial g(x)$ since

$$\partial h(x) = \{16, 17, \ldots, 48\}, \quad \partial g(x) = \{15, 16, \ldots, 65\}.$$

Therefore x is minimum in $U = \bigcup_{p \in \partial g(x)} \partial h^*(p) = \{1, 2, 3\}$ by Proposition 4.11. However, $x = 2$ is not a global minimum, since $f(2) = 0 > -175 = f(-3)$. Indeed, for $\epsilon = 17$, the inclusion of (4.7) fails, since $\partial_\epsilon h(x) \not\subseteq \partial_\epsilon g(x)$ with

$$\partial_\epsilon h(x) = \{-1, 0, \ldots, 65\}, \quad \partial_\epsilon g(x) = \{0, 1, \ldots, 82\}.$$

■

The relation of the global optimality and the local optimality is summarized in Figure 4.1.

global optimality	Prop. 4.9 $\Longleftrightarrow$	$\partial_\epsilon h(x) \subseteq \partial_\epsilon g(x)$
		$\Downarrow$
		$\partial h(x) \subseteq \partial g(x)$
		$\Downarrow$ Prop. 4.11
local optimality:		x is minimum in $U = \bigcup_{p \in \partial g(x)} \partial h^*(p)$

Figure 4.1: Relation of global and local optimalities.

Recall the discrete Toland-Singer duality (4.3) in Theorem 4.6. The following proposition states that an optimal solution p of the dual problem can be constructed from an optimal solution x of the primal problem. This is sometimes referred to as optimal solution transportation in the literature of (continuous) DC programming.

Proposition 4.13 (Optimal solution transportation)**.**

(1) If $p \in \partial g(x) \cap \partial h(x)$ then

$$g(x) - h(x) = h^*(p) - g^*(p).$$

(2) If x is a global minimum of $g - h$, then any $p \in \partial g(x) \cap h(x)$ is a global minimum of $h^* - g^*$.

Proof. (1) By the discrete Young-Fenchel inequality (Corollary 2.8), we have

$$g(x) + g^*(p) = \langle p, x\rangle = h(x) + h^*(p).$$

Therefore

$$g(x) - h(x) = h^*(p) - g^*(p).$$

(2) By (1), we have $g(x) - h(x) = h^*(p) - g^*(p)$. Since the left-hand side is a minimum, p is a minimum of $h^* - g^*$, by the discrete Toland-Singer duality. □

It is noted that, in Proposition 4.13 (2), the condition $p \in \partial g(x) \cap \partial h(x)$ with a global minimum x can be simplified to $p \in \partial h(x)$ by (4.8).

4.5 Approximation ratio

In general, there are no theoretical guarantee for the approximation ratio of a local optimal solution.

Example 4.14. Let $a < b$ be positive integers. Let g and h be univariate convex functions defined as

$$g(x) = \begin{cases} 0, & x \le b, \\ +\infty, & x > b, \end{cases}$$
$$h(x) = \begin{cases} 0, & x \le a, \\ x - a, & x > a \end{cases}$$

and consider a discrete DC programming problem to minimize $f(x) = g(x) - h(x)$ for $x \in \mathbb{Z}$. Every $x < a$ satisfies (4.8) and hence x is a local minimum with $f(x) = 0$. However, the global minimum is $x = b$ and $f(b) = a - b$. ■

In some special case, however, we can prove an approximation ratio. In this case, we consider that $f = g - h$ is a "perturbed" convex function, i.e., h is smaller than g.

Theorem 4.15. Let $g : \{0,1\}^n \to \mathbb{Z} \cup \{+\infty\}$ be a function on $\{0,1\}^n$ and $h : \{0,1\}^n \to \mathbb{Z}$ be a monotone nondecreasing L$^\natural$-convex function. If $x \in \{0,1\}^n$ satisfies $\partial h(x) \subseteq \partial g(x)$ in (4.8), then for any $y \in \{0,1\}^n$,

$$g(x) - 2h(x) \le g(y) - h(y). \tag{4.10}$$

Proof. Let $X = \mathrm{supp}^+(x)$ and $Y = \mathrm{supp}^+(y)$, and take a permutation π of $[n]$ as

$$\{\pi(1), \dots, \pi(|X|)\} = X,$$
$$\{\pi(|X \setminus Y| + 1), \dots, \pi(|X \cup Y|)\} = Y.$$

By the construction (2.8) of subgradients of L$^\natural$-convex functions, the vector $p \in \mathbb{Z}^n$ whose $\pi(i)$-th entries are given by

$$p(\pi(i)) := h(\chi_{\{\pi(1),\ldots,\pi(i))\}}) - h(\chi_{\{\pi(1),\ldots,\pi(i-1)\}}), \quad i = 1, \ldots, n,$$

belongs to $\partial h(x)$. Since $\partial h(x) \subseteq \partial g(x)$, we have $p \in \partial g(x)$. Therefore

$$g(x) - \langle p, x \rangle \leq g(y) - \langle p, y \rangle. \tag{4.11}$$

By the definition of p and monotonicity of h,

$$\begin{aligned} \langle p, x \rangle &= h(x), \\ \langle p, y \rangle &= h(\chi_{X \cup Y}) - h(\chi_{X \setminus Y}) \geq h(y) - h(x). \end{aligned}$$

Substituting these into (4.11), we obtain

$$g(x) - 2h(x) \leq g(y) - h(y).$$

□

Corollary 4.16. Let $g : \{0,1\}^n \to \mathbb{Z} \cup \{+\infty\}$ be a function that takes negative values on its effective domain, i.e., $g(x) \leq 0$ for all $x \in \operatorname{dom} g$, and let $h : \{0,1\}^n \to \mathbb{Z}$ be a monotone nondecreasing L$^\natural$-convex function. If x satisfies $\partial h(x) \subseteq \partial g(x)$, then

$$h(x) - g(x) \geq (1/2) \max_{y \in \{0,1\}^n} \{h(y) - g(y)\}.$$

Proof. Since $g(x) \leq 0$, we have

$$h(x) - g(x) \geq (1/2)(2h(x) - g(x)).$$

By (4.10), the right-hand side is bounded from below by $h(y) - g(y)$ for any y. □

As a further corollary, we obtain the following bound for the matroid constraint submodular maximization problem (see Example 4.3).

Corollary 4.17. Let δ be the indicator function of the family $\mathcal{I}$ of independent sets of a matroid $\mathcal{M}$ and h be a monotone nondecreasing submodular set function. If $x = \chi_X$ satisfies $\partial h(x) \subseteq \partial \delta(x)$, then x is a 1/2 approximation solution, i.e.,

$$h(\chi_X) \geq (1/2) \max_{Y \in \mathcal{I}} h(\chi_Y).$$

■

It is known [3] that the matroid constraint monotone submodular maximization problem can be solved in polynomial time with approximation ratio $1 - 1/e$ by the continuous greedy algorithm.

Example 4.18. We cannot drop the monotonicity condition in Theorem 4.15. Consider the following functions g and h on $\{0,1\}^2$:

$$g(0,0) = 0, \quad g(1,0) = 0, \quad g(0,1) = 0, \quad g(1,1) = +\infty,$$
$$h(0,0) = 0, \quad h(1,0) = 1, \quad h(1,0) = 3, \quad h(1,1) = 0.$$

The function h is submodular. Let us compute the subdifferential at $x = (1,0)$. First, $p \in \partial g(x)$ if and only if

$$-p_1 \leq 0, \quad p_2 - p_1 \leq 0.$$

Hence we have $\partial g(x) = \{p \in \mathbb{Z}^2 : p_1 \geq 0, p_2 \leq p_1\}$. Next, $p \in \partial h(x)$ if and only if

$$p_2 \leq -1, \quad -p_1 \leq -1, \quad p_2 - p_1 \leq 2.$$

The third inequality is redundant and we have $\partial h(x) = \{p \in \mathbb{Z}^2 : p_1 \geq 1, p_2 \leq -1\}$. Therefore $\partial h(x) \subseteq \partial g(x)$. Let $y = (0,1)$. Then the inequality in (4.10) reads as

$$g(x) - 2h(x) = -2 \not\leq -3 = g(y) - h(y).$$

Indeed, by increasing $h(0,1)$, the approximation ratio tends to be arbitrarily worse.

Example 4.19. The inequality in (4.10) is tight. Consider

$$g(0,0) = 0, \quad g(1,0) = 0, \quad g(0,1) = 0, \quad g(1,1) = +\infty,$$
$$h(0,0) = 0, \quad h(1,0) = 1, \quad h(1,0) = 2, \quad h(1,1) = 2.$$

The function h is monotone submodular. Let us compute the subdifferential at $x = (1,0)$. First, we have $\partial g(x) = \{p \in \mathbb{Z}^2 : p_1 \geq 0, p_2 \leq p_1\}$ as in Example 4.18. Next, $p \in \partial g(x)$ if and only if

$$-p_1 \leq -1, \quad p_2 \leq 1, \quad p_2 - p_1 \leq 1.$$

The third inequality is redundant and we have $\partial h(x) = \{p \in \mathbb{Z}^2 : p_1 \geq 1, p_2 \leq 1\}$. Therefore $\partial h(x) \subseteq \partial g(x)$. Let $y = (0,1)$. Then the inequality in (4.10) holds with equality:

$$g(x) - 2h(x) = -2 = g(y) - h(y).$$

Problem 4.20. Establish a statement about approximation ratio when h is an M$^\natural$-convex function.

5 Algorithms

As mentioned in Section 1, there are two types of algorithms in continuous DC programming; the convex analysis approach (e.g., the DC algorithm) and the enumerative approach (e.g., branch-and-bound/cutting plane algorithm). We here propose *discrete DC algorithms* to be categorized as "discrete convex analysis approach."

We first propose a general framework of discrete DC algorithm (Algorithm 1), which is a direct translation of the continuous DC algorithm. Our general framework can be applied to all class of discrete DC programming problems. We then propose algorithms tailored to $\mathrm{M}^\natural$−$\mathrm{M}^\natural$, $\mathrm{M}^\natural$−$\mathrm{L}^\natural$, and $\mathrm{L}^\natural$−$\mathrm{L}^\natural$ DC programming problems (Algorithms 2, 4, and 5). These algorithms exploit the polyhedral structure of discrete convex functions.

5.1 Generic discrete DC algorithm

We first describe the general framework of discrete DC algorithm which is a direct translation of the continuous DC algorithm (also known as Concave-Convex Procedure in the area of machine learning). In DC programming, the difficulty in the problem $\min_x\{g(x) - h(x)\}$ comes from the concavity of $-h$. The idea of the DC algorithm is to iteratively approximate h with its subgradient $p \in \partial h(x)$ and solve the convex minimization problem

$$\min_x \{g(x) - \langle p, x\rangle\}.$$

The solution set of this problem is given by $\partial g^*(p)$.

A straightforward adaptation of the above idea to discrete DC programming yields the following algorithm (Algorithm 1). Note that the algorithm has a symmetry between the primal and dual problems in the Toland-Singer duality.

Algorithm 1 Generic form of discrete DC algorithm

Let $x^{(1)}$ be an initial solution
for $k = 1, 2, \ldots$ **do**
  (Primal phase) Pick $p^{(k)} \in \partial h(x^{(k)})$
  (Dual phase) Pick $x^{(k+1)} \in \partial g^*(p^{(k)})$
  if $g(x^{(k)}) - \langle p^{(k)}, x^{(k)}\rangle = g(x^{(k+1)}) - \langle p^{(k)}, x^{(k+1)}\rangle$ **then**
    Return $x^{(k)}$
  end if
end for

Proposition 5.1 (Convergence of the discrete DC algorithm)**.** Let $g, h : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ be discrete ($\mathrm{M}^\natural$- or $\mathrm{L}^\natural$-)convex functions.

(1) $g(x^{(k)}) - h(x^{(k)})$ decreases strictly monotonically. Hence, the algorithm terminates in a finite number of iterations.

(2) When the algorithm terminates, x is a minimum of $f = g - h$ within $\partial g^*(p)$.

Proof. (1) We first note that the termination condition is equivalent to $x^{(k)} \in \partial g^*(p^{(k)})$. Suppose $x^{(k)} \notin \partial g^*(p^{(k)})$. By the Young-Fenchel inequality,

$$g(x^{(k)}) + g^*(p^{(k)}) > \langle p^{(k)}, x^{(k)} \rangle = h(x^{(k)}) + h^*(p^{(k)}).$$

Therefore

$$g(x^{(k)}) - h(x^{(k)}) > h^*(p^{(k)}) - g^*(p^{(k)}). \tag{5.1}$$

Similarly, by the Young-Fenchel inequality,

$$h(x^{(k+1)}) + h^*(p^{(k)}) \geq \langle p^{(k)}, x^{(k+1)} \rangle = g(x^{(k+1)}) + g^*(p^{(k)}).$$

Therefore

$$h^*(p^{(k)}) - g^*(p^{(k)}) \geq g(x^{(k+1)}) - h(x^{(k+1)}). \tag{5.2}$$

By combining (5.1) and (5.2) we obtain

$$g(x^{(k)}) - h(x^{(k)}) > g(x^{(k+1)}) - h(x^{(k+1)}),$$

which shows that $g(x^{(k)}) - h(x^{(k)})$ strictly decreases monotonically. This guarantees the finite termination since $g - h$ is integer-valued and bounded from below.

(2) When the algorithm terminates, we obtain a pair (x, p) such that

$$x \in \partial g^*(p), \quad p \in \partial h(x).$$

For every $y \in \partial g^*(p)$ we have $p^* \in \partial g(y) \cap \partial h(x)$. Then the claim follows from Lemma 4.10. □

To realize the generic discrete DC algorithm, we need to implement the primal phase (subgradient computation for h) and the dual phase (function minimization for g). As mentioned in Section 2, both phases can be carried out efficiently by the existing algorithms in discrete convex analysis. For an $\mathrm{L}^\natural$–$\mathrm{L}^\natural$ DC program where both g and h are $\mathrm{L}^\natural$-convex and $\mathrm{dom}\, g, \mathrm{dom}\, h \subseteq \{0,1\}^n$, our generic algorithm coincides with the submodular-supermodular procedure proposed by Narasimhan and Bilmes [38].

It would be nice if the local optimality condition $\partial h(x) \subseteq \partial g(x)$ in (4.8) is guaranteed at the termination of the discrete DC algorithm (Algorithm 1). As it stands, however, the algorithm does not have this property, mainly because it does not specify which subgradient $p^{(k)} \in \partial g(x^{(k)})$ to pick in the primal phase. By modifying the primal phase to

$$p^{(k)} \in \partial g(x^{(k)}) \setminus \partial h(x^{(k)}), \tag{5.3}$$

we can guarantee the local optimality condition $\partial h(x) \subseteq \partial g(x)$ at the termination of the algorithm.

To implement the modified primal phase (5.3), we take advantage of the "discreteness" and/or "polyhedral structure" of each class of discrete convex functions, which is described below.

5.2 M$^\natural$–M$^\natural$ DC programming

If both g and h are M$^\natural$-convex, we can verify the local optimality condition $\partial h(x) \subseteq \partial g(x)$ in (4.8) in polynomial time by enumerating all faces of these subdifferentials since the number of faces of $\partial h(x)$ is polynomial in n (Proposition 5.4). Therefore we can construct a polynomial time (per iteration) algorithm (Algorithm 2 below) to obtain a local optimal solution that satisfies $\partial h(x) \subseteq \partial g(x)$ in (4.8).

Lemma 5.2. For an M$^\natural$-convex function h,

$$\begin{aligned}
\max_{p \in \partial h(x)} \langle p, -\chi_i \rangle &= h(x - \chi_i) - h(x), \\
\max_{p \in \partial h(x)} \langle p, \chi_j \rangle &= h(x + \chi_j) - h(x), \\
\max_{p \in \partial h(x)} \langle p, \chi_j - \chi_i \rangle &= h(x + \chi_j - \chi_i) - h(x).
\end{aligned}$$

Proof. This lemma can be obtained immediately from Proposition 5.1 of [35] but, to be self-contained, we give a direct proof here. We only prove the third identity. The other two can be proved in the same way.

By the definition of subgradients, we have $\max \langle p, \chi_j - \chi_i \rangle \leq h(x + \chi_j - \chi_i) - h(x)$. To prove the equality, we construct a subgradient $p \in \partial h(x)$ that satisfies the equality. Consider a weighted directed graph $G = (V, E)$ where $V = [n] \cup \{0\}$ and $E = V \times V$ with $w(k, 0) = h(x - \chi_k) - h(x)$, $w(0, l) = h(x + \chi_l) - h(x)$, and $w(k, l) = h(x - \chi_k + \chi_l) - h(x)$ for all $k, l \in [n]$. Note that w satisfies the triangle inequality by the M$^\natural$-convexity of h. Let $d(k)$ $(k \in V)$ be the shortest path distance from i to k and let $p(k) = d(k) - d(i)$ for $k \in [n]$. Then p is a feasible potential (i.e., $p \in \partial h(x)$) and $p(j) - p(i) = d(j) = w(i, j) = h(x - \chi_i + \chi_j) - h(x)$. Therefore we have $\langle p, \chi_j - \chi_i \rangle = h(x - i + j) - h(x)$ for this p. □

Lemma 5.3. Let $g : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ be an M$^\natural$-convex function and $D \subseteq \mathbb{Z}^n$. Then $D \subseteq \partial g(x)$ holds if and only if, for all $i, j \in [n]$,

$$\begin{aligned} \max_{p \in D} \langle p, -\chi_i \rangle &\leq g(x - \chi_i) - g(x), \\ \max_{p \in D} \langle p, \chi_j \rangle &\leq g(x + \chi_j) - g(x), \\ \max_{p \in D} \langle p, \chi_j - \chi_i \rangle &\leq g(x - \chi_i + \chi_j) - g(x). \end{aligned} \tag{5.4}$$

Proof. (if part): Suppose (5.4). Then, for every $p \in D$,

$$\begin{aligned} -p(i) &\leq g(x - \chi_i) - g(x), \\ p(j) &\leq g(x + \chi_j) - g(x), \\ p(j) - p(i) &\leq g(x - \chi_i + \chi_j) - g(x). \end{aligned} \tag{5.5}$$

Therefore $p \in \partial g(x)$ by the explicit formula of M$^\natural$-subgradients (Theorem 2.5).

(only if part): Suppose $D \subseteq \partial g(x)$. Then, for every $a \in \mathbb{R}^n$,

$$\max_{p \in D} \langle p, a \rangle \leq \max_{p \in \partial g(x)} \langle p, a \rangle.$$

By choosing $a = -\chi_i$, χ_j, and $\chi_j - \chi_i$ and using Lemma 5.2, we obtain (5.4). □

Proposition 5.4. Let g and h be M$^\natural$-convex functions. Then $\partial h(x) \subseteq \partial g(x)$ holds if and only if, for all $i, j \in [n]$,

$$\begin{aligned} h(x - \chi_i) - h(x) &\leq g(x - \chi_i) - g(x), \\ h(x + \chi_i) - h(x) &\leq g(x + \chi_j) - g(x), \\ h(x - \chi_i + \chi_j) - h(x) &\leq g(x - \chi_i + \chi_j) - g(x). \end{aligned} \tag{5.6}$$

Proof. Apply Lemma 5.3 with $D = \partial h(x)$ and use Lemma 5.2. □

From Proposition 5.4, we are naturally led to Algorithm 2 below. When the algorithm terminates, x satisfies (5.6) and hence the local optimality condition $\partial h(x) \subseteq \partial g(x)$ in (4.8).

Algorithm 2 M$^\natural$–M$^\natural$ DC algorithm

---

loop
  Find $a \in \bigcup_{i,j \in [n]} \{-\chi_i, \chi_j, \chi_j - \chi_i\}$ such that $h(x + a) - h(x) > g(x + a) - g(x)$
  If no such a exists, return x
  Find $p \in \partial h(x)$ such that $\langle p, x \rangle = h(x + a) - h(x)$ (by the shortest path algorithm)
  $x \leftarrow \mathrm{argmin}\{g(y) - \langle p, y \rangle : y \in \mathbb{Z}^n\}$
end loop

Remark 5.5. Since the optimality condition (5.6) is expressed only in the (primal) function values, we can construct a simple "primal" algorithm below with guaranteed convergence to a local optimal solution.

Algorithm 3 Simplified M$^\natural$–M$^\natural$ DC algorithm

---

repeat
  $x \leftarrow \mathrm{argmin}\{g(y) - h(y) : y \in \bigcup_{i,j \in [n]} \{x - \chi_i, x + \chi_j, x - \chi_i + \chi_j\}\}$
until convergence
Return x

5.3 M$^\natural$–L$^\natural$ DC programming

If g is M$^\natural$-convex and h is L$^\natural$-convex, the local optimality condition $\partial h(x) \subseteq \partial g(x)$ in (4.8) can be verified in polynomial time. The basic idea is the same as in the M$^\natural$–M$^\natural$ case: we enumerate all faces of a subdifferential of an M$^\natural$-convex function.

Proposition 5.6. Let g be an M$^\natural$-convex function and h be an L$^\natural$-convex function. Then $\partial h(x) \subseteq \partial g(x)$ holds if and only if, for all $p \in \partial h(x)$,

$$\begin{aligned} \langle p, -\chi_i \rangle &\leq g(x - \chi_i) - g(x), \quad i \in [n], \\ \langle p, \chi_j \rangle &\leq g(x + \chi_j) - g(x), \quad j \in [n], \\ \langle p, \chi_i - \chi_i \rangle &\leq g(x - \chi_i + \chi_j) - g(x), \quad i, j \in [n]. \end{aligned} \tag{5.7}$$

Proof. Apply Lemma 5.3 with $D = \partial h(x)$. □

The maximization of the left-hand sides of (5.7) can be solved by the greedy algorithm, although no simple expressions are available for those maximum values. Therefore, if $\partial h(x) \not\subseteq \partial g(x)$, we can obtain a certificate p such that $p \in \partial h(x)$ and $p \notin \partial g(x)$ in polynomial time. Combining this procedure with the generic discrete DC algorithm, we obtain Algorithm 4 below. When the algorithm terminates, we have $x \in \mathbb{Z}^n$ that satisfies the local optimality condition $\partial h(x) \subseteq \partial g(x)$ in (4.8).

Algorithm 4 M$^\natural$–L$^\natural$ DC algorithm

---

loop
  Find $p \in \partial h(x)$ that violates (5.7) (by the greedy algorithm)
  If no such p exists, return x
  $x \leftarrow \mathrm{argmin}\{g(y) - \langle p, y \rangle : y \in \mathbb{Z}^n\}$
end loop

5.4 $\mathrm{L}^\natural$–$\mathrm{L}^\natural$ DC programming

For an $\mathrm{L}^\natural$-convex function g, its subdifferential $\partial g(x)$ has exponentially many faces. Therefore we cannot use a similar technique as that for $\mathrm{M}^\natural$–$\mathrm{M}^\natural$ or $\mathrm{M}^\natural$–$\mathrm{L}^\natural$ DC programming. Indeed, if both g and h are $\mathrm{L}^\natural$-convex, the problem of testing for inclusion $\partial h(x) \subseteq \partial g(x)$ is equivalent to the so-called submodular containment problem, which is co-NP complete [25]. However, by the Toland-Singer duality, an $\mathrm{L}^\natural$–$\mathrm{L}^\natural$ DC program is transformed to an $\mathrm{M}^\natural$–$\mathrm{M}^\natural$ DC program and its local optimality condition $\partial g^*(p) \subseteq \partial h^*(p)$ can be checked in polynomial time by the method described above.

Once a dual local minimum solution p is obtained, by optimal solution transportation (Proposition 4.13), we can construct a primal solution x by taking x from $\partial g^*(p)$ that satisfies $g(x) - h(x) = h^*(p) - g^*(p)$. Note that the dual solution p certainly satisfies the dual local optimality condition $\partial g^*(p) \subseteq \partial h^*(p)$, but the constructed primal solution x does not necessarily satisfy the primal local optimality condition $\partial h(x) \subseteq \partial g(x)$.

Algorithm 5 $\mathrm{L}^\natural$–$\mathrm{L}^\natural$ DC algorithm

Compute $p \in \mathbb{Z}^n$ by applying Algorithm 2 or 3 to the dual problem: $\min\{h^*(p) - g^*(p) : p \in \mathbb{Z}^n\}$
Return $x \in \partial g^*(p)$

5.5 $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC programming

Since g is $\mathrm{L}^\natural$-convex, we cannot use a similar technique as that for $\mathrm{M}^\natural$–$\mathrm{M}^\natural$ or $\mathrm{M}^\natural$–$\mathrm{L}^\natural$ DC programming. Furthermore, since the dual of $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC program is another $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC program, we cannot use the technique for $\mathrm{L}^\natural$–$\mathrm{L}^\natural$ DC programming. However, since the problem is a "submodular function minimization on a distributive lattice", it can be solved efficiently (see Remark 5.7 below), independently of our generic form of the discrete DC algorithm. Establishing an algorithm of "discrete DC algorithm" type for $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC programming is an interesting problem.

Remark 5.7. Consider an $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC programming problem $\min_x\{g(x) - h(x)\}$ on $\mathbb{Z}^n$. Since an $\mathrm{L}^\natural$-convex function $g : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ is submodular on $\mathbb{Z}^n$ and an $\mathrm{M}^\natural$-convex function $h : \mathbb{Z}^n \to \mathbb{Z} \cup \{+\infty\}$ is supermodular on $\mathbb{Z}^n$, $f := g - h$ is submodular on $\mathrm{dom}\, g$ (by the convention of $(+\infty) - (+\infty) = (+\infty)$). Furthermore, by the finiteness assumption (4.4), we have $\mathrm{dom}\, f = \mathrm{dom}\, g$. Since $\mathrm{dom}\, f$ forms a distributive lattice (with respect to component-wise max and min), an $\mathrm{L}^\natural$–$\mathrm{M}^\natural$ DC programming problem is a special case of submodular function minimization problem on a distributive lattice [48]. Hence, if $\mathrm{dom}\, f$ is a finite set, this problem can be solved in polynomial time in the length of a maximal chain in the lattice [26, 35, 41]. For example, if $\mathrm{dom}\, f = [0, K]^n$, we can minimize f in $O(\mathrm{poly}(nK))$ time.

For the convenience of readers, we give a brief description of a method for submodular function minimization on a distributive lattice. The following argument is based on Note 10.15 in [35] and Section 4 in [41].

Let $f : L \to \mathbb{Z}$ be a submodular function on a distributive lattice $\mathcal{L} = (L, \vee, \wedge)$. Let $I \subseteq L$ be the set of join-irreducible elements, and define $\phi : L \to 2^I$ by

$$\phi(x) = \{y \in I : x \wedge y = y\}, \quad x \in L.$$

By Birkhoff's representation theorem [2], ϕ is a lattice isomorphism. Note that $\phi^{-1}(X) = \bigvee_{x \in X} x$ for $X \subseteq I$.

Let $F : 2^I \to \mathbb{Z}$ be defined by $F(X) := f(\phi^{-1}(X))$. Then F is a submodular set function on I. Indeed,

$$\begin{aligned} F(X) + F(Y) &= f(\phi^{-1}(X)) + f(\phi^{-1}(Y)) \\ &\geq f(\phi^{-1}(X) \vee \phi^{-1}(Y)) + f(\phi^{-1}(X) \wedge \phi^{-1}(Y)) \\ &= f(\phi^{-1}(X \cup Y)) + f(\phi^{-1}(X \cap Y)) \\ &= F(X \cup Y) + F(X \cap Y). \end{aligned}$$

Furthermore, since $f(x) = F(\phi(x))$ for every $x \in L$, we have

$$\min_{X \subseteq I} F(X) = \min_{x \in L} f(x).$$

Therefore we can solve the submodular function minimization problem on a distributive lattice (right-hand side) through the submodular set function minimization (left-hand side). The complexity is polynomial in $|I|$, which is equal to the length of a maximal chain of $\mathcal{L}$. ■

Acknowledgement

The authors thank Akiyoshi Shioura for helpful discussions, Satoru Iwata for providing the information about reference [25], Tom McCormick and Maurice Queyranne for communicating references [41].

This work is supported by KAKENHI (21360045) and the Aihara Project, the FIRST program from JSPS.

References

[1] M. Bačák and J. M. Borwein (2011): On difference convexity of locally Lipschitz functions. Optimization, vol. 60, pp. 961–978.

[2] G. Birkhoff (1937): Rings of sets. Duke Mathematical Journal, vol 3, pp. 443—454.

[3] G. Calinescu, C. Chekuri, and M. Pál, and J. Vondrák (2011): Maximizing a monotone submodular function subject to a matroid constraint. SIAM Journal on Computing, vol. 40, pp. 1740–1766.

[4] A. W. M. Dress and W. Wenzel (1992): Valuated matroids. Advances in Mathematics, vol. 93, pp. 214–250.

[5] J. Edmonds (1970): Submodular functions, matroids and certain polyhedra. In: R. Guy, H. Hanani, N. Sauer, and J. Schönheim, eds., Combinatorial Structures and Their Applications, Gordon and Breach, New York, pp. 69–87.

[6] P. Favati and F. Tardella (1990): Convexity in nonlinear integer programming. Ricerca Operativa, vol. 53, pp. 3–44.

[7] U. Feige, V. S. Mirrokni, and J. Vondrák (2011): Maximizing non-monotone submodular functions. SIAM Journal on Computing, vol. 40, pp. 1133–1153.

[8] S. Fujishige (2005): Submodular Functions and Optimization. 2nd ed., Annals of Discrete Mathematics, vol. 58, Elsevier, Amsterdam.

[9] S. Fujishige and K. Murota (2000): Notes on L-/M-convex functions and the separation theorems. Mathematical Programming, vol. 88, pp. 129–146.

[10] A. Frank (1984): Generalized polymatroids. In A. Hajnal, L. Lovász, and V. T. Sós, eds., Finite and Infinite Sets, (Proceedings of 6th Hungarian Combinatorial Colloquium, 1981), Colloquia Mathematica Societatis János Bolyai, vol. 37, pp. 285–294, North-Holland.

[11] A. Frank (2011): Connections in Combinatorial Optimization. Oxford Lecture Series in Mathematics and Its Applications, 38, Oxford University Press.

[12] A. Frank and É. Tardos (1988): Generalized polymatroids and submodular flows. Mathematical Programming, vol. 42, pp. 489–563.

[13] P. Hartman (1959): On functions representable as a difference of convex functions. Pacific Journal of Mathematics, vol. 9, pp. 707–713.

[14] H. Hirai and K. Murota (2004): M-convex functions and tree metric. Japan Journal of Industrial and Applied Mathematics, vol. 21, pp. 391–403.

[15] J. B. Hiriart-Urruty (1985): Generalized differentiability, duality and optimization for problems dealing with differences of convex functions. Lecture Notes in Economics and Mathematical Systems, vol. 256, pp. 37–70.

[16] J. B. Hiriart-Urruty (1989): From convex optimization to nonconvex optimization, Part I: Necessary and sufficient conditions for global optimality. In: F. H. Clarke, V. F. Dem'yanov, and F. Giannessi, eds., Nonsmooth Optimization and Related Topics, Ettore Majorana International Sciences, Physical Sciences, Series 43, Plenum Press, New York, pp. 219–239.

[17] J.-B. Hiriart-Urruty and C. Lemaréchal (1993): Convex Analysis and Minimization Algorithms I, II. Springer-Verlag, Berlin.

[18] R. Horst, N. V. Thoai (1999): DC Programming: Overview. Journal of Optimization Theory and Applications, vol. 103, pp. 1–43.

[19] S. Iwata and J. B. Orlin (2009): A simple combinatorial algorithm for submodular function minimization. In Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, pp. 1230–1237.

[20] R. Iyer and J. Bilmes (2012): Algorithms for approximate minimization of the difference between submodular functions, with applications. In Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence, pp. 407–417. Also: `http://arxiv.org/abs/1207.0560`.

[21] Y. Kawahara, K. Nagano, and Y. Okamoto (2011): Submodular fractional programming for balanced clustering. Pattern Recognition Letters, vol. 32, pp. 235–243.

[22] Y. Kawahara and T. Washio (2011): Prismatic algorithm for discrete D.C. programming problem. In Proceedings of the 25th Annual Conference on Neural Information Processing Systems, pp. 2106–2114.

[23] V. Kolmogorov and A. Shioura (2009): New algorithms for convex cost tension problem with application to computer vision. Discrete Optimization, vol. 6, pp. 378–393.

[24] J. Lee, V. S. Mirrokni, V. Nagarajan, and M. Sviridenko (2009): Non-monotone submodular maximization under matroid and knapsack constraints. In Proceedings of the 41st Annual ACM Symposium on Theory of Computing, pp. 323–332.

[25] S. T. McCormick (1996): Submodular containment is hard, even for networks. Operations Research Letters, vol. 19, pp. 95–99.

[26] S. T. McCormick (2006): Submodular function minimization. In: K. Aardal, G. Nemhauser, and R. Weismantel, eds., Discrete Optimization, Handbooks in Operations Research and Management Science, vol. 12, Elsevier Science Publishers, Berlin, Chapter 7, pp. 321–391.

[27] B. L. Miller (1971): On minimizing nonseparable functions defined on the integers with an inventory applications. SIAM Journal on Applied Mathematics, vol. 21, pp. 166–185.

[28] S. Moriguchi and K. Murota (2005): Discrete Hessian matrix for L-convex functions. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E83-A, pp. 1104–1108.

[29] S. Moriguchi and K. Murota (2012): On discrete Hessian matrix and convex extensibility. Journal of Operations Research Society of Japan, vol. 55, pp. 48–62.

[30] K. Murota (1996): Valuated matroid intersection, I: optimality criteria. SIAM Journal on Discrete Mathematics, vol.9, pp. 545–561.

[31] K. Murota (1996): Valuated matroid intersection, II: algorithms. SIAM Journal on Discrete Mathematics, vol.9, pp. 562–576.

[32] K. Murota (1997): Matroid valuation on independent sets. Journal of Combinatorial Theory, Series B, vol. 69, pp. 59–78.

[33] K. Murota (1998): Discrete convex analysis. Mathematical Programming, vol. 83, pp. 313–371.

[34] K. Murota (2000): Matrices and Matroids for Systems Analysis. Springer-Verlag, Berlin.

[35] K. Murota (2003): Discrete Convex Analysis. Society for Industrial and Applied Mathematics, Philadelphia.

[36] K. Murota (2009): Recent developments in discrete convex analysis. In: W. Cook., L. Lovász, J. Vygen, eds., Research Trends in Combinatorial Optimization, Springer, Berlin, Chapter 11, pp. 219–260.

[37] K. Murota and A. Shioura (1999): M-convex function on generalized polymatroid. Mathematics of Operations Research, vol. 24, pp. 95–105.

[38] M. Narasimhan and J. Bilmes (2005): A submodular-supermodular procedure with applications to discriminative structure learning. In Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence, pp. 404–412.

[39] S. Onn (2010): Nonlinear Discrete Optimization: An Algorithmic Theory. European Mathematical Society, Zurich.

[40] J. B. Orlin (2009): A faster strongly polynomial time algorithm for submodular function minimization. Mathematical Programming, Series A, vol. 118, pp. 237–251.

[41] M. Queyranne and F. Tardella (2004): Submodular function minimization in $\mathbb{Z}^n$ and searching in Monge arrays. Unpublished manuscript, Presented by M. Queyranne at CTW04 (Cologne Twente Workshop 2004), Electronic Notes in Discrete Mathematics, vol. 17, p. 5.

[42] A. Schrijver (2003): Combinatorial Optimization: Polyhedra and Efficiency, Springer-Verlag, Berlin.

[43] A. Shioura (2000): Level set characterization of M-convex functions. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E83-A, pp. 586–589.

[44] I. Singer (1979): A Fenchel-Rockafellar type duality theorem for maximization. Bulletin of the Australian Mathematical Society, vol. 20, pp. 193–198.

[45] P. D. Tao and S. El Bernoussi (1987): Duality in D.C. (difference of convex functions) optimization: Subgradient methods. In: K. H. Hoffman, J. Zowe, J. B. Hiriart-Urruty, and C. Lemaréchal, eds., Trends in Mathematical Optimization, International Series of Numerical Mathematics, vol. 84, Birkhäuser, Basel, pp. 277–293.

[46] P. D. Tao and L. T. Hoai An (1997): Convex analysis approach to D.C. programming: Theory, algorithms and applications. Acta Mathematica Vietnamica, vol. 22, pp. 289–355.

[47] J. F. Toland (1979): A duality principle for non-convex optimisation and the calculus of variations. Archive for Rational Mechanics and Analysis, vol. 71, pp. 41–61.

[48] D. M. Topkis (1978): Minimizing a submodular function on a lattice. Operations Research, vol. 26, pp. 305–321.

[49] H. Tuy (1987): Global minimization of a difference of two convex functions. Mathematical Programming Study, vol. 30, pp. 150–182.

[50] H. Tuy (1995): D.C. optimization: Theory, methods and algorithms. In: R. Horst and P. M. Pardalos, eds., Handbook of Global Optimization, Kluwer Academic Publishers, Dordrecht, pp. 149–216.

[51] J. Vondrák (2008): Optimal approximation for the submodular welfare problem in the value oracle model. In Proceedings of the 40th annual ACM Symposium on Theory of Computing, pp. 67–74.

[52] A. L. Yuille and A. Rangarajan (2003): The concave-convex procedure. Neural Computation, vol. 15, pp. 915–936.