Department of Industrial Engineering and Economics

Working Paper

No. 2019-3

M-convex Function Minimization Under L1-Distance Constraint and Its Application to Dock Re-allocation in Bike Sharing System

Akiyoshi Shioura

February, 2019

Tokyo Institute of Technology

2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, JAPAN
http://educ.titech.ac.jp/iee/

M-convex Function Minimization Under L1-Distance Constraint and Its Application to Dock Re-allocation in Bike Sharing System*

Akiyoshi Shioura†

February 4, 2019

Abstract

In this paper we consider a new problem of minimizing an M-convex function under L1-distance constraint (MML1); the constraint is given by an upper bound for L1-distance between a feasible solution and a given "center." This is motivated by a nonlinear integer programming problem for re-allocation of dock capacity in a bike sharing system discussed by Freund et al. (2017). The main aim of this paper is to better understand the combinatorial structure of the dock re-allocation problem through the connection with M-convexity, and show its polynomial-time solvability using this connection. For this, we first show that the dock re-allocation problem can be reformulated in the form of (MML1). We then present a pseudo-polynomial-time algorithm for (MML1) based on steepest descent approach. We also propose two polynomial-time algorithms for (MML1) by replacing the L1-distance constraint with a simple linear constraint. Finally, we apply the results for (MML1) to the dock re-allocation problem to obtain a pseudo-polynomial-time steepest descent algorithm and also polynomial-time algorithms for this problem. The proposed algorithm is based on a proximity-scaling algorithm for a relaxation of the dock re-allocation problem, which is of interest in its own right.

1 Introduction

The concepts of M-convexity and M$^\natural$-convexity for functions in integer variables play a primary role in the theory of discrete convex analysis [11]. M-convex function, introduced by Murota [9, 10], is defined by a certain exchange axiom (see Section 2 for a precise definition), and enjoys various nice properties as "discrete convexity" such as a local characterization for global minimality, extensibility to ordinary convex functions, conjugacy, duality, etc. M$^\natural$-convex function is introduced by Murota and Shioura [14] as a variant of M-convex function. While the class of M$^\natural$-convex functions properly contains that of M-convex functions, the concept of M$^\natural$-convexity is essentially equivalent to M-convexity in the sense that an M$^\natural$-convex function can be obtained by the projection of some M-convex function (see, e.g., [11]). Minimization of an M-convex function is the most fundamental optimization problem concerning M-convex functions, and a common generalization of the separable convex resource allocation problem under a submodular constraint and some classes of nonseparable convex function minimization on integer lattice points. M-convex function minimization can be solved by a steepest descent algorithm (or greedy algorithm) that runs in pseudo-polynomial time [11, 12], and various polynomial-time algorithms have been proposed [8, 15, 16, 18].

*This work was supported by JSPS KAKENHI Grant Numbers 15K00030, 18K11177.

†Department of Industrial Engineering and Economics, Tokyo Institute of Technology, Tokyo 152-8550, Japan shioura.a.aa@m.titech.ac.jp

In this paper, we consider a new problem of minimizing an M-convex function under the L1-distance constraint, which is formulated as follows:

$$\text{(MML1)}\quad \left|\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & \sum_{i=1}^{n} x(i) = \theta, \\ & \|x - x_{\mathrm{c}}\|_1 \leq 2\gamma, \\ & x \in \mathrm{dom}\, f, \end{array}\right.$$

where n, θ, γ are integers with $n > 0$ and $\gamma \geq 0$, $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ is an M-convex function such that $\sum_{i=1}^{n} x(i) = \theta$ holds for every $x \in \mathbb{Z}^n$ with $f(x) < +\infty$, and x_{c} is a vector (called the "center") with $f(x_{\mathrm{c}}) < +\infty$ and $\sum_{i=1}^{n} x_{\mathrm{c}}(i) = \theta$. This problem is motivated by a nonlinear integer programming problem for re-allocation of dock-capacity in a bike sharing system [1].

In a bike sharing system, many bike stations are located around a city so that users can rent and return bikes there. Each bike station has several docks and bikes; some docks are equipped with bikes, and the other docks are kept open so that users can return bikes at the station. The numbers of docks with bike and of open docks change as time passes, and it is possible that some users cannot rent or return a bike at a station due to the shortage of bikes or open docks, and in such situation users feel dissatisfied. To reduce users' dissatisfaction, operators of a bike sharing system need to re-allocate docks (and bikes) among bike stations appropriately. Change to a new allocation, however, requires the movement of docks and bikes, which yields some amount of cost. Therefore, it is desirable that a new allocation is not so different from the current allocation. Hence, the task of operators in a bike sharing system is to minimize users' dissatisfaction by changing the allocation of docks, while bounding the number of docks to be moved in the re-allocation.

This problem, which we refer to as the *dock re-allocation problem*, is discussed by Freund, Henderson, and Shmoys [1] and formulated as follows[1]:

$$\text{(DR)}\quad \left|\begin{array}{ll} \text{Minimize} & \sum_{i=1}^{n} c_i(d(i), b(i)) \\ \text{subject to} & \sum_{i=1}^{n} (d(i) + b(i)) = D + B, \\ & \sum_{i=1}^{n} b(i) \leq B, \\ & \sum_{i=1}^{n} |(d(i) + b(i)) - (\bar{d}(i) + \bar{b}(i))| \leq 2\gamma, \\ & \ell(i) \leq d(i) + b(i) \leq u(i) \quad (i \in N), \\ & d(i), b(i) \in \mathbb{Z}_+ \quad (i \in N). \end{array}\right.$$

Here, $N = \{1, 2, \ldots, n\}$ denotes the set of bike stations. For a station $i \in N$, we denote by $b(i), d(i) \in \mathbb{Z}_+$, respectively, the decision variables representing the numbers of docks with bike and of open docks allocated at the station. The expected number of dissatisfied users at the station i is represented by a function $c_i : \mathbb{Z}_+^2 \to \mathbb{R}$ in variables $d(i)$ and $b(i)$, and shown to have the property of *multimodularity* (see Section 2 for the definition).

The first constraint in (DR) means that the total number of docks (i.e., docks with bike and open docks) is equal to a fixed constant $D + B$. The second constraint gives an upper bound for the total number of docks with bike. The third constraint, given in the form of L1-distance constraint, means that the difference between the current and the new allocations of docks should be small, where $\bar{d}(i)$ and $\bar{b}(i)$ denote, respectively, the numbers of docks with bike and of open docks at the station i in the current allocation. In addition, the number of docks $d(i)+b(i)$ at each station i should be between lower and upper bounds $[\ell(i), u(i)]$, as represented by the fourth constraint.

[1]While the first constraint is given as an inequality $\sum_{i=1}^{n}(d_i + b_i) \leq D + B$ in [1], it is implicitly assumed in [1] that the inequality holds with equality. Indeed, the algorithm in [1] applies only to the problem with the equality constraint.

For the problem (DR), Freund et al. [1] propose a steepest descent (or greedy) algorithm that repeatedly update a constant number of variables by ± 1, and prove by using the multimodularity of the objective function that the algorithm finds an optimal solution of (DR) in at most γ iterations. Hence, the problem (DR) can be solved in pseudo-polynomial time.

Our Contribution The main aim of this paper is to better understand the combinatorial structure of the problem (DR) through the connection with M-convexity, and to provide polynomial-time algorithms for (DR) by using the connection.

We first show that the dock re-allocation problem (DR) can be reformulated in the form of the minimization of an M-convex function under the L1-distance constraint (MML1), where we regard $d(i) + b(i)$ as a single variable (see Section 3 for details).

We then consider the problem (MML1) and present a steepest descent algorithm that runs in pseudo-polynomial time. While unconstrained M-convex function minimization (i.e., the problem (MML1) without the L1-distance constraint) can be solved by a certain steepest descent algorithm (see Section 2.2; see also [11, 12] for details), a naive application of the algorithm does not work for the problem (MML1), due to the L1-distance constraint. Nevertheless, we prove in Section 4 that if the center $x_{\rm c}$ is used as an initial solution of the algorithm, then the steepest descent algorithm finds an optimal solution in γ iterations. Moreover, we prove a stronger statement that for each $k = 0, 1, 2, \ldots$, the vector generated in the k-th iteration of the steepest descent algorithm is an optimal solution of the M-convex function minimization under the constraint $\|x - x_{\rm c}\|_1 = 2k$ (i.e., the problem (MML1) with γ replaced with k). As a byproduct of this result, we obtain new properties of the steepest descent algorithm for unconstrained M-convex function minimization. In particular, we provide a nontrivial tight bound on the number of iterations required by the algorithm, and show that the trajectory of the solutions generated by the algorithm is a geodesic (i.e, a "shortest" path) to the nearest optimal solution from the initial solution.

While the problem (MML1) can be solved by a steepest descent algorithm, its running time is pseudo-polynomial time. To obtain faster algorithms, we present in Section 5 two approaches to solve (MML1) in polynomial time. For this, we show that by using a minimizer of the M-convex objective function, the L1-distance constraint in (MML1) can be replaced with a simple linear constraint; the two approaches proposed in this section solve the M-convex function minimization under the simple linear constraint instead of the original problem. The first approach is to reduce the problem to the minimization of the sum of two M-convex functions, for which polynomial-time algorithms are available. The second approach is based on the reduction to the minimization of another M-convex function with smaller number of variables, and the resulting algorithm is faster than the first approach.

Finally, in Section 6 we apply the algorithms for (MML1) presented in Sections 4 and 5 to the dock re-allocation problem (DR), which can be regarded as a special case of (MML1). We aim at obtaining fast algorithms by making use of the special structure of (DR).

In Section 6.1, we discuss an application of the steepest descent algorithm in Section 4 to (DR). A naive application of the algorithm takes $\mathrm{O}(n^3 \log(B/n))$ time in each iteration since it requires $\mathrm{O}(n \log(B/n))$ time for the evaluation of the M-convex function f used in the reformulation of (DR). To reduce the time complexity, we present a useful property of the M-convex function f that the update of function value $f(x)$ can be done quickly in $\mathrm{O}(\log n)$ time if the vector x is updated to a vector in a neighborhood. Furthermore, we make full use of this property to implement the steepest descent algorithm so that the algorithm works for the original formulation and each iteration requires $\mathrm{O}(\log n)$ time only. We also discuss the connection with the steepest descent algorithm in [1], and show that the fast implementation of our algorithm is nothing but the steepest descent algorithm in [1].

Section 6.2 is devoted to polynomial-time algorithms for (DR). While the polynomial-time solvability of (DR) follows from the results in Section 5, a naive application of an algorithm in Section 5 leads to a polynomial-time but rather slow algorithm for (DR); a faster implementation is difficult this time since the algorithms in Section 4 are more involved. Instead, we use an idea in Section 5 and the structure of (DR) to obtain a faster polynomial-time algorithm. For this, we replace the L1-distance constraint in (DR) with a simple linear constraint, as in Section 5. This new formulation, together with the use of a new problem parameter, makes it possible to decompose the problem (DR) into two independent subproblems, both of which can be reduced to M-convex function minimization and therefore can be solved efficiently. We show that an algorithm based on this approach runs in $\mathrm{O}(n \log n \log((D+B)/n) \log B)$ time. To obtain this time bound, we prove a proximity theorem for a relaxation of the problem (DR) and devise a proximity-scaling algorithm for the relaxation in Section 6.3; the proximity theorem and the algorithm are of interest in their own right.

We finally point out that quite recently, Freund et al. [2] also propose a wealy-polynomial-time algorithm for (DR). Their approach is based on proximity-scaling and different from ours.

2 Preliminaries on M-convexity

Throughout the paper, let n be a positive integer with $n \geq 2$ and $N = \{1, 2, \ldots, n\}$. We denote by $\mathbb{R}$ the set of real numbers, and by $\mathbb{Z}$ (resp., by $\mathbb{Z}_+$) the sets of integers (resp., nonnegative integers); $\mathbb{Z}_{++}$ denotes the set of positive integers.

Let $x = (x(1), x(2), \ldots, x(n)) \in \mathbb{R}^n$ be a vector. We denote $\mathrm{supp}^+(x) = \{i \in N \mid x(i) > 0\}$ and $\mathrm{supp}^-(x) = \{i \in N \mid x(i) < 0\}$. For a subset $Y \subseteq N$, we denote $x(Y) = \sum_{i \in Y} x(i)$. We define $\|x\|_1 = \sum_{i \in N} |x(i)|$ and $\|x\|_\infty = \max_{i \in N} |x(i)|$.

We define $\mathbf{0} = (0, 0, \ldots, 0) \in \mathbb{Z}^n$. For $Y \subseteq N$, we denote by $\chi_Y \in \{0,1\}^n$ the characteristic vector of Y, i.e., $\chi_Y(i) = 1$ if $i \in Y$ and $\chi_Y(i) = 0$ otherwise. In particular, we denote $\chi_i = \chi_{\{i\}}$ for every $i \in N$. We also denote $\chi_0 = \mathbf{0}$. Inequality $x \leq y$ for vectors $x, y \in \mathbb{R}^n$ means component-wise inequality $x(i) \leq y(i)$ for all $i \in N$. For two vectors $x, y \in \mathbb{Z}^n$ with $x \leq y$, we denote $[x, y] = \{z \in \mathbb{Z}^n \mid x \leq z \leq y\}$.

2.1 M-convex and Multimodular Functions

Let $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ be a function. The *effective domain* of f is defined by $\mathrm{dom}\, f = \{x \in \mathbb{Z}^n \mid f(x) < +\infty\}$, and the set of minimizers of f is denoted by $\arg\min f$. Function f is said to be *M$^\natural$-convex* if it satisfies the following exchange property:

(M$^\natural$-EXC) $\forall x, y \in \mathrm{dom}\, f$, $\forall i \in \mathrm{supp}^+(x-y)$, $\exists j \in \mathrm{supp}^-(x-y) \cup \{0\}$:

$$f(x) + f(y) \geq f(x - \chi_i + \chi_j) + f(y + \chi_i - \chi_j).$$

For an M$^\natural$-convex function f, if $\mathrm{dom}\, f$ is contained in a hyperplane $\{x \in \mathbb{Z}^n \mid x(N) = \theta\}$ for some $\theta \in \mathbb{Z}$, then f is called an *M-convex function*, in particular. It is known (see, e.g., [11]) that a function f is M-convex if and only if it satisfies the following exchange property:

(M-EXC) $\forall x, y \in \mathrm{dom}\, f$, $\forall i \in \mathrm{supp}^+(x-y)$, $\exists j \in \mathrm{supp}^-(x-y)$:

$$f(x) + f(y) \geq f(x - \chi_i + \chi_j) + f(y + \chi_i - \chi_j).$$

M-convex functions can be characterized by a seemingly weaker exchange property.

Theorem 2.1 ([11, Theorem 6.4])**.** *A function $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ is M-convex if and only if it satisfies the following condition:*
$\forall x, y \in \mathrm{dom}\, f$ *with* $x \neq y$, $\exists i \in \mathrm{supp}^+(x - y)$, $\exists j \in \mathrm{supp}^-(x - y)$:

$$f(x) + f(y) \geq f(x - \chi_i + \chi_j) + f(y + \chi_i - \chi_j).$$

M$^\natural$-convexity of a function implies the following exchange properties.

Theorem 2.2 ([14])**.** *Let $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ be an M$^\natural$-convex function and $x, y \in \mathrm{dom}\, f$.*
(i) *If $x(N) \leq y(N)$, then for every $i \in \mathrm{supp}^+(x - y)$ there exists some $j \in \mathrm{supp}^-(x - y)$ such that*

$$f(x) + f(y) \geq f(x - \chi_i + \chi_j) + f(y + \chi_i - \chi_j).$$

(ii) *If $x(N) < y(N)$, then there exists some $j \in \mathrm{supp}^-(x - y)$ such that*

$$f(x) + f(y) \geq f(x + \chi_j) + f(y - \chi_j).$$

We then explain the concept of multimodularity and its connection with M$^\natural$-convexity. A function $\varphi : \mathbb{Z}^2_+ \to \mathbb{R}$ in two variables is called *multimodular* if it satisfies the following conditions:

$$\begin{aligned}
\varphi(\eta + 1, \zeta + 1) - \varphi(\eta + 1, \zeta) &\geq \varphi(\eta, \zeta + 1) - \varphi(\eta, \zeta) \qquad (\forall \eta, \zeta \in \mathbb{Z}_+),\\
\varphi(\eta - 1, \zeta + 1) - \varphi(\eta - 1, \zeta) &\geq \varphi(\eta, \zeta) - \varphi(\eta, \zeta - 1) \qquad (\forall \eta, \zeta \in \mathbb{Z}_{++}),\\
\varphi(\eta + 1, \zeta - 1) - \varphi(\eta, \zeta - 1) &\geq \varphi(\eta, \zeta) - \varphi(\eta - 1, \zeta) \qquad (\forall \eta, \zeta \in \mathbb{Z}_{++}).
\end{aligned}$$

For functions in two variables, multimodularity and M$^\natural$-convexity are essentially equivalent.

Proposition 2.3 (cf. [7])**.** *A function $\varphi : \mathbb{Z}^2_+ \to \mathbb{R}$ in two variables is multimodular if and only if the function $f : \mathbb{Z}^2 \to \mathbb{R} \cup \{+\infty\}$ given by*

$$\mathrm{dom}\, f = \mathbb{Z}^2_+, \qquad f(\alpha, \beta) = \varphi(\alpha, \beta) \quad ((\alpha, \beta) \in \mathrm{dom}\, f) \tag{2.1}$$

is M$^\natural$-convex. □

This relationship and Theorem 2.2 immediately imply the following property of multimodular functions.

Proposition 2.4. *Let $\varphi : \mathbb{Z}^2 \to \mathbb{R} \cup \{+\infty\}$ be an M$^\natural$-convex function, and $\eta, \zeta, \eta', \zeta' \in \mathbb{Z}$.*
(i) *If $\eta > \eta'$ and $\zeta < \zeta'$, then it holds that*

$$\varphi(\eta, \zeta) + \varphi(\eta', \zeta') \geq \varphi(\eta - 1, \zeta + 1) + \varphi(\eta' + 1, \zeta' - 1). \tag{2.2}$$

(ii) *If $\eta > \eta'$ and $\eta + \zeta > \eta' + \zeta'$, then it holds that*

$$\varphi(\eta, \zeta) + \varphi(\eta', \zeta') \geq \varphi(\eta - 1, \zeta) + \varphi(\eta' + 1, \zeta'). \tag{2.3}$$

Proof. We first prove the claim (i). If $\eta + \zeta \leq \eta' + \zeta'$ Theorem 2.2 (i) immediately implies the inequality (2.2). If $\eta + \zeta \geq \eta' + \zeta'$, then we can also obtain the inequality (2.2) from Theorem 2.2 (i) by interchanging the roles of (η, ζ) and (η', ζ').

We then prove the claim (ii). If $\zeta < \zeta'$, then the inequality (2.3) follows immediately from Theorem 2.2 (ii). If $\zeta \geq \zeta'$, then the inequality (2.3) follows immediately from (M$^\natural$-EXC). □

2.2 Minimization of an M-convex Function

We review some known results for the minimization of an M-convex function. A minimizer of an M-convex function can be characterized by a local optimality condition.

Theorem 2.5 (cf. [11, Theorem 6.26]). *For an M-convex function* $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$*, a vector* $x^* \in \operatorname{dom} f$ *is a minimizer of* f *if and only if* $f(x^* - \chi_i + \chi_j) \geq f(x^*)$ $(\forall i, j \in N)$*.*

This theorem immediately implies that the minimization of an M-convex function can be solved by the following steepest descent algorithm (see, e.g., [11, Section 10.1.1]):

Algorithm SteepestDescent
Step 0: Let $x_0 \in \operatorname{dom} f$ be an appropriately chosen initial vector. Set $k := 1$.
Step 1: If $f(x_{k-1} + \chi_i - \chi_j) \geq f(x_{k-1})$ for every $i, j \in N$, then output x_{k-1} and stop.
Step 2: Find $i_k, j_k \in N$ that minimize $f(x_{k-1} + \chi_{i_k} - \chi_{j_k})$.
Step 3: Set $x_k := x_{k-1} + \chi_{i_k} - \chi_{j_k}$, $k := k + 1$, and go to Step 1.

Theorem 2.6 (cf. [11, Section 10.1.1]). *Let* $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ *be an M-convex function* $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ *that has a minimizer, i.e..* $\arg\min f \neq \emptyset$*. Then, the algorithm* Steepest-Descent *outputs a minimizer of* f *after a finite number of iterations.*

Polynomial-time algorithms based on proximity-scaling approach are proposed for M-convex function minimization [8, 15, 16, 18], and the current best time complexity bound is given as follows. For a set $S \subseteq \mathbb{Z}^n$, we define the L_∞*-diameter* of S by

$$L = \max\{\|x - y\|_\infty \mid x, y \in S\}. \tag{2.4}$$

Theorem 2.7 ([16, 18]). *Minimization of an M-convex function* $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ *can be done in* $\mathrm{O}(n^3 \log(L/n)F)$ *time, where* L *is the* L_∞*-diameter of* $\operatorname{dom} f$ *and* F *denotes the time to evaluate the function value of* f*.*

3 Reformulation of Dock Re-allocation Problem as (MML1)

We consider the dock re-allocation problem (DR) explained in Section 1. Using vector notation, the problem (DR) can be simply rewritten as follows:

$$\text{(DR)}\ \left|\ \begin{array}{ll} \text{Minimize} & c(d, b) \\ \text{subject to} & d(N) + b(N) = D + B, \\ & b(N) \leq B, \\ & \|(d + b) - (\bar{d} + \bar{b})\|_1 \leq 2\gamma, \\ & \ell \leq d + b \leq u, \\ & d, b \in \mathbb{Z}^n_+, \end{array}\right.$$

where $c : \mathbb{Z}^n_+ \times \mathbb{Z}^n_+ \to \mathbb{R}$ is a function given by $c(d, b) = \sum_{i=1}^n c_i(d(i), b(i))$ $((d, b) \in \mathbb{Z}^n_+ \times \mathbb{Z}^n_+)$. In this section, we show that (DR) can be reformulated as the problem (MML1).

For the reformulation, we define a function $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ by

$$\begin{aligned} &\operatorname{dom} f = \{x \in \mathbb{Z}^n \mid x(N) = D + B,\ \ell \leq x \leq u\}, \\ &f(x) = \min\{c(d, b) \mid d, b \in \mathbb{Z}^n_+,\ d + b = x,\ b(N) \leq B\} \quad (x \in \operatorname{dom} f). \end{aligned} \tag{3.1}$$

Theorem 3.1. *Function* f *in* (3.1) *is M-convex.*

Proof of Theorem 3.1 is given at the end of this section. With this function f, the problem (DR) can be reformulated as

$$\left|\begin{array}{ll}\text{Minimize} & f(x) \\ \text{subject to} & x(N) = D + B, \\ & \|x - (\bar{d} + \bar{b})\|_1 \leq 2\gamma, \\ & x \in \operatorname{dom} f.\end{array}\right.$$

Hence, (DR) is reformulated as (MML1).

We now give a proof of Theorem 3.1.

Proof of Theorem 3.1. By Theorem 2.1 (i), it suffices to show that the following condition holds for every distinct vectors $x', x'' \in \operatorname{dom} f$:

$$\begin{aligned}&\exists i \in \operatorname{supp}^+(x' - x''), \exists j \in \operatorname{supp}^-(x' - x'') \text{ such that} \\ &\qquad f(x') + f(x'') \geq f(x' - \chi_i + \chi_j) + f(x'' + \chi_i - \chi_j).\end{aligned} \tag{3.2}$$

For $x \in \operatorname{dom} f$, we denote

$$S(x) = \{(d, b) \in \mathbb{Z}^n_+ \times \mathbb{Z}^n_+ \mid y + z = x,\ b(N) \leq B\}.$$

Let $(d', b') \in S(x')$ (resp., $(d'', b'') \in S(x'')$) be a pair of vectors such that $f(x') = c(d', b')$ (resp., $f(x'') = c(d'', b'')$). We denote

$$N^+ = \operatorname{supp}^+(x' - x''), \quad N^- = \operatorname{supp}^-(x' - x''), \quad N^0 = N \setminus (N^+ \cup N^-).$$

We have $N^+ \neq \emptyset$ and $N^- \neq \emptyset$ since x' and x'' are distinct vectors with $x'(N) = x''(N)$. In the following, we consider only the case with $b'(N) = b''(N) = B$ since the remaining case can be proved similarly and more easily. Note that this assumption and the equation $x'(N) = x''(N)$ implies $d'(N) = d''(N)$.

We first show by using Proposition 2.4 that the condition (3.2) holds if at least one of the following four conditions holds:

(C1) $N^+ \cap \operatorname{supp}^+(d' - d'') \neq \emptyset,\ N^- \cap \operatorname{supp}^-(d' - d'') \neq \emptyset$,
(C2) $N^+ \cap \operatorname{supp}^+(b' - b'') \neq \emptyset,\ N^- \cap \operatorname{supp}^-(b' - b'') \neq \emptyset$,
(C3) $N^+ \cap \operatorname{supp}^+(d' - d'') \neq \emptyset,\ N^- \cap \operatorname{supp}^-(b' - b'') \neq \emptyset,\ N^0 \cap \operatorname{supp}^-(d' - d'') \neq \emptyset$,
(C4) $N^+ \cap \operatorname{supp}^+(b' - b'') \neq \emptyset,\ N^- \cap \operatorname{supp}^-(d' - d'') \neq \emptyset,\ N^0 \cap \operatorname{supp}^-(b' - b'') \neq \emptyset$.

In the following, we give a proof for only the case with (C3); the proof for other cases are similar and omitted. Let $i, j, s \in N$ be distinct elements such that

$$i \in N^+ \cap \operatorname{supp}^+(d' - d''), \quad j \in N^- \cap \operatorname{supp}^-(b' - b''), \quad s \in N^0 \cap \operatorname{supp}^-(d' - d'').$$

Since $N^0 \cap \operatorname{supp}^-(d' - d'') = N^0 \cap \operatorname{supp}^+(b' - b'')$, we have $s \in \operatorname{supp}^+(b' - b'')$. We define vectors $\tilde{d}', \tilde{d}'', \tilde{b}', \tilde{b}'', \tilde{x}', \tilde{x}'' \in \mathbb{Z}^n$ by

$$\begin{aligned}\tilde{d}' &= d' - \chi_i + \chi_s, & \tilde{d}'' &= d'' + \chi_i - \chi_s, \\ \tilde{b}' &= b' + \chi_j - \chi_s, & \tilde{b}'' &= b'' - \chi_j + \chi_s, \\ \tilde{x}' &= \tilde{d}' + \tilde{b}'\ (= x' - \chi_i + \chi_j), & \tilde{x}'' &= \tilde{d}'' + \tilde{b}''(= x'' + \chi_i - \chi_j).\end{aligned}$$

Since

$$\begin{aligned}
&\tilde{x}'(i) = x'(i) - 1 \geq x''(i), && \tilde{x}'(j) = x'(j) + 1 \leq x''(j),\\
&\tilde{x}''(i) = x''(i) + 1 \leq x'(i), && \tilde{x}''(j) = x''(j) - 1 \geq x'(j),\\
&\tilde{x}'(N) = x'(N) = D + B, && \tilde{x}''(N) = x''(N) = D + B,\\
&\tilde{b}'(N) = b'(N) = B, && \tilde{b}''(N) = b''(N) = B,
\end{aligned}$$

it holds that $\tilde{x}', \tilde{x}'' \in \operatorname{dom} f$, $(\tilde{d}', \tilde{b}') \in S(\tilde{x}')$, and $(\tilde{d}'', \tilde{b}'') \in S(\tilde{x}'')$. Hence, we have

$$f(\tilde{x}') \leq c(\tilde{d}', \tilde{b}'), \qquad f(\tilde{x}'') \leq c(\tilde{d}'', \tilde{b}''). \tag{3.3}$$

By the choice of $i, j, s \in N$, the following inequalities follow from Proposition 2.4:

$$\begin{aligned}
&c_i(d'(i), b'(i)) + c_i(d''(i), b''(i))\\
&\geq c_i(d'(i) - 1, b'(i)) + c_i(d''(i) + 1, b''(i)) = c_i(\tilde{d}'(i), \tilde{b}'(i)) + c_i(\tilde{d}''(i), \tilde{b}''(i)),\\
&c_j(d'(j), b'(j)) + c_j(d''(j), b''(j))\\
&\geq c_j(d'(j), b'(j) + 1) + c_j(d''(j), b''(j) - 1) = c_j(\tilde{d}'(j), \tilde{b}'(j)) + c_j(\tilde{d}''(j), \tilde{b}''(j)),\\
&c_s(d'(s), b'(s)) + c_s(d''(s), b''(s))\\
&\geq c_s(d'(s) + 1, b'(s) - 1) + c_s(d''(s) - 1, b''(s) + 1) = c_s(\tilde{d}'(s), \tilde{b}'(s)) + c_s(\tilde{d}''(s), \tilde{b}''(s)).
\end{aligned}$$

From these inequalities and (3.3), the inequality (3.2) can be obtained as follows:

$$\begin{aligned}
f(x') + f(x'') &= c(d', b') + c(d'', b'')\\
&\geq c(\tilde{d}', \tilde{b}') + c(\tilde{d}'', \tilde{b}'')\\
&\geq f(\tilde{x}') + f(\tilde{x}'') = f(x' - \chi_i + \chi_j) + f(x'' + \chi_i - \chi_j).
\end{aligned}$$

To conclude the proof, we show that at least one of the four conditions (C1)–(C4) holds. It follows from the equations $x'(h) = d'(h) + b'(h)$ and $x''(h) = d''(h) + b''(h)$ for $h \in N$ that

$$N^+ \subseteq \operatorname{supp}^+(d' - d'') \cup \operatorname{supp}^+(b' - b''), \tag{3.4}$$
$$N^- \subseteq \operatorname{supp}^-(d' - d'') \cup \operatorname{supp}^-(b' - b''). \tag{3.5}$$

Sine N^+ and N^- are non-empty sets, (3.4) and (3.5) imply that

$$N^+ \cap \operatorname{supp}^+(d' - d'') \neq \emptyset \quad \text{or} \quad N^+ \cap \operatorname{supp}^+(b' - b'') \neq \emptyset, \tag{3.6}$$
$$N^- \cap \operatorname{supp}^-(d' - d'') \neq \emptyset \quad \text{or} \quad N^- \cap \operatorname{supp}^-(b' - b'') \neq \emptyset. \tag{3.7}$$

Assume that neither of conditions (C1) and (C2) holds. Then, we have

$$\begin{aligned}
&N^+ \cap \operatorname{supp}^+(d' - d'') = \emptyset \quad \text{or} \quad N^- \cap \operatorname{supp}^-(d' - d'') = \emptyset,\\
&N^+ \cap \operatorname{supp}^+(b' - b'') = \emptyset \quad \text{or} \quad N^- \cap \operatorname{supp}^-(b' - b'') = \emptyset,
\end{aligned}$$

which, together with (3.6) and (3.7), imply that the following two cases are possible:

(a) $N^+ \cap \operatorname{supp}^+(d' - d'') \neq \emptyset$, $N^- \cap \operatorname{supp}^-(d' - d'') = \emptyset$, $N^+ \cap \operatorname{supp}^+(b' - b'') = \emptyset$, and $N^- \cap \operatorname{supp}^-(b' - b'') \neq \emptyset$,
(b) $N^+ \cap \operatorname{supp}^+(d' - d'') = \emptyset$, $N^- \cap \operatorname{supp}^-(d' - d'') \neq \emptyset$, $N^+ \cap \operatorname{supp}^+(b' - b'') \neq \emptyset$, and $N^- \cap \operatorname{supp}^-(b' - b'') = \emptyset$.

We will show that (a) implies (C3), and (b) implies (C4); below we prove the former implication only since the latter can be proven in a similar way.

Suppose that the condition (a) holds. It suffices to show that $N^0 \cap \mathrm{supp}^-(d' - d'') \neq \emptyset$. Since $\emptyset \neq N^+ \subseteq \mathrm{supp}^+(d' - d'')$ and $N^- \subseteq N \setminus \mathrm{supp}^-(d' - d'')$, we have $d'(N^+) > d''(N^+)$ and $d'(N^-) \geq d''(N^-)$, which, together with $d'(N) = d''(N)$, implies

$$d'(N^0) = d'(N) - d'(N^+) - d'(N^-) < d''(N) - d''(N^+) - d''(N^-) = d''(N^0).$$

Hence, $N^0 \cap \mathrm{supp}^-(d' - d'') \neq \emptyset$ follows. $\square$

4 Steepest Descent Algorithm for (MML1)

4.1 Algorithm

In this section, we show that an optimal solution of the problem (MML1) can be obtained by using a variant of the steepest descent algorithm SteepestDescent in Section 2.2 for unconstrained M-convex function minimization. While we are mainly interested in the case where the center x_c is a feasible solution to (MML1), we also consider the case with infeasible x_c. Since we consider only vectors x with $\|x - x_c\|_1 \leq 2\gamma$ in (MML1), we may assume, without loss of generality, that the effective domain $\mathrm{dom}\, f$ of the function f is bounded; this assumption implies that $\arg\min f \neq \emptyset$, in particular.

Let $\sigma \in \mathbb{Z}_+$ be the half of L1-distance between x_c and a nearest vector in $\mathrm{dom}\, f$, and $\tau \in \mathbb{Z}_+$ the half of L1-distance between x_c and a nearest minimizer of f, i.e.,

$$\sigma = (1/2)\min\{\|x - x_c\|_1 \mid x \in \mathrm{dom}\, f\}, \qquad \tau = (1/2)\min\{\|x - x_c\|_1 \mid x \in \arg\min f\}. \tag{4.1}$$

We have $\sigma = 0$ if x_c is a feasible solution.

For every integer k with $k \geq \sigma$, we denote by (MML1(k)) the problem (MML1) with the constant γ in the L1-distance constraint is replaced with the parameter k. That is, (MML1(k)) is given as follows:

$$\text{(MML1}(k)\text{)} \left| \begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & x(N) = \theta, \\ & \|x - x_c\|_1 \leq 2k, \\ & x \in \mathrm{dom}\, f. \end{array} \right.$$

We first present some properties of optimal solutions for the problem (MML1(k)). For every integer k with $k \geq \sigma$, we denote by $M_k \subseteq \mathbb{Z}^n$ and by $\mu_k \in \mathbb{R}$, respectively, the set of optimal solutions and the optimal value of the problem (MML1(k)). Note that for every $k \geq \tau$, we have

$$M_k = \{x \in \arg\min f \mid \|x - x_c\|_1 \leq 2k\}, \qquad \mu_k = \min f.$$

We also have $M_0 = \{x_c\}$ and $\mu_0 = f(x_c)$ if x_c is a feasible solution of (MML1);

Theorem 4.1.
(i) *It holds that $\mu_\sigma > \mu_{\sigma+1} > \cdots > \mu_\tau$ and $M_k \subseteq \{x \in \mathbb{Z}^n \mid \|x - x_c\|_1 = 2k\}$ for $k \in [\sigma, \tau]$.*
(ii) *For every $k \in [\sigma, \tau - 1]$ and $y \in M_k$, there exist some $i \in N \setminus \mathrm{supp}^-(y - x_c)$ and $j \in N \setminus \mathrm{supp}^+(y - x_c)$ such that $y + \chi_i - \chi_j \in M_{k+1}$.*
(iii) *For every $k \in [\sigma, \tau - 1]$ and $y \in M_{k+1}$, there exist some $i \in \mathrm{supp}^+(y - x_c)$ and $j \in \mathrm{supp}^-(y - x_c)$ such that $y - \chi_i + \chi_j \in M_k$.*

Proof is given in Section 4.3.

Theorem 4.1 (ii) implies that a variant of the steepest descent algorithm for unconstrained M-convex function minimization finds an optimal solution of (MML1).

Algorithm SteepestDescentMML1
Step 0: Compute σ in (4.1) and $x^\circ \in M_\sigma$. Set $x_\sigma := x^\circ$ and $k := \sigma + 1$.
Step 1: If $k - 1 = \gamma$, then output x_{k-1} and stop.
Step 2: Find $i_k, j_k \in N$ that minimize $f(x_{k-1} + \chi_{i_k} - \chi_{j_k})$.
Step 3: If $f(x_{k-1} + \chi_{i_k} - \chi_{j_k}) \geq f(x_{k-1})$, then output x_{k-1} and stop.
Otherwise, set $x_k := x_{k-1} + \chi_{i_k} - \chi_{j_k}$, $k := k + 1$, and go to Step 1.

Theorem 4.2. *The algorithm* SteepestDescentMML1 *outputs an optimal solution of* (MML1) *in at most $\gamma - \sigma + 1$ iterations. Moreover, the vector x_k generated in each iteration of the algorithm satisfies $x_k \in M_k$.*

Proof. Suppose that the algorithm terminates when $k = k'$. We first prove by induction that $k \leq \tau$ and $x_k \in M_k$ hold for $k = \sigma, \sigma+1, \ldots, k'-1$. Assume that $k-1 \leq \tau$ and $x_{k-1} \in M_{k-1}$ hold for some k with $\sigma \leq k - 1 \leq k' - 2$. Since the algorithm does not stop in Step 3 of the previous iteration, we have $f(x_k) < f(x_{k-1})$, which implies that x_{k-1} is not a minimizer of f, i.e., $k \leq \tau$ holds. Since $f(x_k) < f(x_{k-1})$, we have $i_k \in N \setminus \mathrm{supp}^-(x_{k-1} - x_\mathrm{c})$, $j_k \in N \setminus \mathrm{supp}^+(x_{k-1} - x_\mathrm{c})$, and $\|x_k - x_\mathrm{c}\|_1 = k$. Hence, it follows from Theorem 4.1 (ii) that $x_k \in M_k$.

We then prove that the output of the algorithm is an optimal solution of (MML1). We have either $k' - 1 = \gamma$ or $f(x_{k'-1} + \chi_{i_{k'}} - \chi_{j_{k'}}) \geq f(x_{k'-1})$ (or both). If $k' - 1 = \gamma$, then the output is the vector x_γ and therefore an optimal solution of (MML1) since $x_\gamma \in M_\gamma$. If $f(x_{k'-1} + \chi_{i_{k'}} - \chi_{j_{k'}}) \geq f(x_{k'-1})$, then $x_{k'-1}$ is a minimizer of f by Theorem 2.5, which is an optimal solution of (MML1). $\square$

The running time of the algorithm SteepestDescentMML1, except for Step 0, is $\mathrm{O}(n^2(\gamma - \sigma))$, provided that the evaluation of function value of f can be done in constant time. Computation of σ and x° in Step 0 can be done by finding a minimizer x° of a function $f(x) + \Upsilon\|x - x_\mathrm{c}\|_1$ with a sufficiently large real number Υ with $\Upsilon > \max\{f(x) \mid x \in \mathrm{dom}\, f\}$ and then setting $\sigma = \|x^\circ - x_\mathrm{c}\|_1$. Since the sum of an M-convex function and a separable-convex function is M-convex [11, Theorem 6.13], function $f(x) + \Upsilon\|x - x_\mathrm{c}\|_1$ is also M-convex, and therefore its minimization can be done by any algorithm for unconstrained M-convex function minimization, even if the value Υ is not given specifically.

Theorem 4.1 (iii) suggests another variant of steepest descent algorithm that starts from a nearest minimizer $x^\bullet$ of f and greedily approaches x_c. This algorithm is faster than SteepestDescentMML1 if $\tau - \gamma$ is smaller than $\gamma - \sigma$.

Algorithm ReverseSteepestDescentMML1
Step 0: Compute the value τ in (4.1) and a minimizer $x^\bullet$ of f with $\|x^\bullet - x_\mathrm{c}\|_1 = 2\tau$.
Set $x_\tau := x^\bullet$, and $k := \tau - 1$.
Step 1: If $k + 1 \leq \gamma$, then output x_{k+1} and stop.
Step 2: Find $i_k \in \mathrm{supp}^+(x_{k+1} - x_\mathrm{c})$ and $j_k \in \mathrm{supp}^-(x_{k+1} - x_\mathrm{c})$ that minimize $f(x_{k+1} - \chi_{i_k} + \chi_{j_k})$.
Step 3: Set $x_k := x_{k+1} - \chi_{i_k} + \chi_{j_k}$, $k := k - 1$, and go to Step 1.

Theorem 4.3. *The algorithm* ReverseSteepestDescentMML1 *outputs an optimal solution of* (MML1) *in at most* $\max\{\tau - \gamma, 0\} + 1$ *iterations. Moreover, the vector x_k generated in each iteration of the algorithm satisfies $x_k \in M_k$.*

Proof. If $\tau \leq \gamma$, then the minimizer $x^\bullet$ of f found in Step 0 is a feasible solution of the problem (MML1). Otherwise (i.e., $\tau > \gamma$), then we can show that $x_k \in M_k$ holds for $k = \tau, \tau - 1, \ldots, \gamma$, in a similar way as in the proof of Theorem 4.2. Hence, the output of the algorithm is an optimal solution of (MML1). □

A minimizer $x^\bullet$ of f with $\|x^\bullet - x_{\mathrm{c}}\|_1 = 2\tau$ is a minimizer of a function $f(x) + \varepsilon\|x - x_{\mathrm{c}}\|_1$ with a sufficiently small positive ε. Since $f(x) + \varepsilon\|x - x_{\mathrm{c}}\|_1$ is M-convex in x, a minimizer of $f(x) + \varepsilon\|x - x_{\mathrm{c}}\|_1$ can be obtained by any algorithm for unconstrained M-convex function minimization.

Remark 4.4. The sequence of optimal values μ_k for (MML1(k)) is a convex sequence, i.e., for every integer $k \in [\sigma + 1, \tau - 1]$, it holds that $\mu_{k-1} + \mu_{k+1} \geq 2\mu_k$. This fact can be shown as follows.

For $k \in [\sigma + 1, \tau - 1]$, let $x_{k-1} \in M_{k-1}$ and $x_{k+1} \in M_{k+1}$ be vectors such that

$$x_{k+1} = x_{k-1} - \chi_i - \chi_{i'} + \chi_j + \chi_{j'}$$

for some $i, i', j, j' \in N$ with $\{i, i'\} \cap \{j, j'\} = \emptyset$, $i, i' \in \mathrm{supp}^-(x_{k+1} - x_{\mathrm{c}})$, and $j, j' \in \mathrm{supp}^+(x_{k+1} - x_{\mathrm{c}})$; the existence of such x_{k-1} and x_{k+1} follows from the claim (ii) (or (iii)) of Theorem 4.1. By (M-EXC) applied to x_{k-1} and x_{k+1}, we have $f(x_{k-1}) + f(x_{k+1}) \geq f(y) + f(z)$ with $(y, z) = (x_{k-1} - \chi_i + \chi_j, x_{k-1} - \chi_{i'} + \chi_{j'})$ or $(y, z) = (x_{k-1} - \chi_i + \chi_{j'}, x_{k-1} - \chi_{i'} + \chi_j)$. In either case we have $\|y - x_{\mathrm{c}}\|_1 = \|z - x_{\mathrm{c}}\|_1 = 2k$, and therefore it follows that

$$\mu_{k-1} + \mu_{k+1} = f(x_{k-1}) + f(x_{k+1}) \geq f(y) + f(z) \geq 2\mu_k.$$

□

4.2 Implication to Unconstrained M-convex Function Minimization

Theorem 4.2 has an important implication to the algorithm SteepestDescent for unconstrained M-convex function minimization.

It is easy to see that the behavior of the algorithm SteepestDescent coincides with that of the algorithm SteepestDescentMML1 with x_{c} given by the initial vector x_0 of SteepestDescent. Hence, following properties of the algorithm SteepestDescent can be obtained as an immediate corollary of Theorem 4.2 for SteepestDescentMML1. In particular, the exact bound on the number of iterations required by the algorithm SteepestDescent is obtained.

Corollary 4.5. *The algorithm* SteepestDescent *outputs a minimizer of an M-convex function $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$, and the number of updates of the vector x is exactly equal to*

$$(1/2) \min\{\|x^* - x_0\|_1 \mid x^* \in \arg\min f\}. \tag{4.2}$$

Moreover, the vector x_k generated in the k-th iteration of the algorithm satisfies

$$x_k \in \arg\min\{f(x) \mid \|x - x_0\|_1 \leq 2k\}.$$

If we update exactly two components of a vector x_k in each iteration, as in the algorithm SteepestDescent, then the number of iterations is at least the bound in (4.2). Hence, Corollary 4.5 shows that SteepestDescent achieves the best-possible bound in this sense, and the trajectory of the solutions generated by the algorithm is a geodesic to the nearest optimal solution from the initial solution.

Remark 4.6. In Corollary 4.5 we obtained the exact bound (4.2) on the number of iterations required by the algorithm SteepestDescent. While the bound (4.2) is obtained for some special case of M-convex functions and for some variants of the algorithm, it is not proven so far for the "naive" steepest descent algorithm (i.e., SteepestDescent).

The bound (4.2) is obtained by [12] for the special case where an M-convex function has a unique minimizer. Based on this fact, the bound (4.2) for *general* M-convex functions is obtained in [12], by using a variant of SteepestDescent with certain tie-breaking rules in the choice of i_k and j_k in Step 1.

The bound (4.2) can be also obtained in [16] by using another variant of SteepestDescent, where a region containing an optimal solution is explicitly maintained by lower and upper bound vectors. Corollary 4.5 shows that no modification of the algorithm SteepestDescent is necessary to obtain the same exact bound. □

4.3 Proof of Theorem 4.1

We give a proof of Theorem 4.1. For this, we show some technical lemmas.

Lemma 4.7. *Let $y, \tilde{y} \in \mathbb{Z}^n$ be distinct vectors satisfying $y(N) = \tilde{y}(N)$. If $\|y - x_{\mathrm{c}}\|_1 \le \|\tilde{y} - x_{\mathrm{c}}\|_1$, then we have $\tilde{y}(i) > x_{\mathrm{c}}(i)$ for some $i \in \mathrm{supp}^+(\tilde{y} - y)$ or $\tilde{y}(j) < x_{\mathrm{c}}(j)$ for some $j \in \mathrm{supp}^-(\tilde{y} - y)$ (or both).*

Proof. We prove the statement by contradiction. Assume, to the contrary, that $\tilde{y}(i) \le x_{\mathrm{c}}(i)$ for all $i \in \mathrm{supp}^+(\tilde{y} - y)$ and $\tilde{y}(j) \ge x_{\mathrm{c}}(j)$ for all $j \in \mathrm{supp}^-(\tilde{y} - y)$. Then, it holds that

$$\begin{aligned}
&\|\tilde{y} - x_{\mathrm{c}}\|_1 - \|y - x_{\mathrm{c}}\|_1 \\
&= \sum_{i \in \mathrm{supp}^+(\tilde{y} - y)} (|\tilde{y}(i) - x_{\mathrm{c}}(i)| - |y(i) - x_{\mathrm{c}}(i)|) + \sum_{j \in \mathrm{supp}^-(\tilde{y} - y)} (|\tilde{y}(j) - x_{\mathrm{c}}(j)| - |y(j) - x_{\mathrm{c}}(j)|) \\
&= \sum_{i \in \mathrm{supp}^+(\tilde{y} - y)} ((x_{\mathrm{c}}(i) - \tilde{y}(i)) - (x_{\mathrm{c}}(i) - y(i))) + \sum_{j \in \mathrm{supp}^-(\tilde{y} - y)} ((\tilde{y}(j) - x_{\mathrm{c}}(j)) - (y(j) - x_{\mathrm{c}}(j))) \\
&= \sum_{i \in \mathrm{supp}^+(\tilde{y} - y)} (-\tilde{y}(i) + y(i)) + \sum_{j \in \mathrm{supp}^-(\tilde{y} - y)} (\tilde{y}(j) - y(j)) < 0,
\end{aligned}$$

a contradiction to the inequality $\|y - x_{\mathrm{c}}\|_1 \le \|\tilde{y} - x_{\mathrm{c}}\|_1$. □

Lemma 4.8. *Let $x, y, z \in \mathbb{Z}^n$, $i \in \mathrm{supp}^+(x - y)$, and $j \in \mathrm{supp}^-(x - y)$. Then, we have*

$$\|x - z\|_1 + \|y - z\|_1 \ge \|(x - \chi_i + \chi_j) - z\|_1 + \|(y + \chi_i - \chi_j) - z\|_1.$$

Proof. For a univariate convex function $\varphi : \mathbb{R} \to \mathbb{R}$ and integers η, ζ with $\eta < \zeta$, it holds that

$$\varphi(\eta) + \varphi(\zeta) \ge \varphi(\eta + 1) + \varphi(\zeta - 1).$$

We have $\|x - z\|_1 = \sum_{i=1}^n |x(i) - z(i)|$ and each term $|x(i) - z(i)|$ is a univariate convex function in $x(i)$. Hence, the claim follows. □

We say that a sequence $y_0, y_1, \ldots, y_h \in \mathrm{dom}\, f$ of vectors is *monotone* if $\|y_k - y_0\|_1 = 2k$ holds for $k = 0, 1, \ldots, h$; this condition can be rewritten as follows: for $k = 0, 1, \ldots, h-1$, there exist $i \in \mathrm{supp}^+(y_k - y_h)$ and $j \in \mathrm{supp}^-(y_k - y_h)$ such that $y_{k+1} = y_k - \chi_i + \chi_j$. Recall that by the definition of τ, every optimal solution of the problem (MML1(τ)) is a minimizer of f.

Lemma 4.9. *Let $y \in \operatorname{dom} f$ be a vector with $\|y - x_{\mathrm{c}}\|_1 < 2\tau$, and $x^\bullet \in M_\tau$ be a vector minimizing the value $\|x^\bullet - y\|_1$. Then, there exists a monotone sequence $y_0, y_1, \dots, y_h \in \operatorname{dom} f$ with $h = (1/2)\|y - x^\bullet\|_1$ such that $y_0 = y$, $y_h = x^\bullet$, and $f(y_0) > f(y_1) > \cdots > f(y_h)$.*

Proof. We prove the claim by induction on h. It suffices to show that there exist some $i \in \operatorname{supp}^+(y - x^\bullet)$ and $j \in \operatorname{supp}^-(y - x^\bullet)$ such that $f(y - \chi_i + \chi_j) < f(y)$ since $(1/2)\|(y - \chi_i + \chi_j) - x^\bullet\|_1 = h - 1$.

Since $\|x^\bullet - x_{\mathrm{c}}\|_1 = 2\tau > \|y - x_{\mathrm{c}}\|_1$, it follows from Lemma 4.7 that $x^\bullet(i) > x_{\mathrm{c}}(i)$ for some $i \in \operatorname{supp}^+(x^\bullet - y)$ or $x^\bullet(j) < x_{\mathrm{c}}(j)$ for some $j \in \operatorname{supp}^-(x^\bullet - y)$ (or both); we assume, without loss of generality, that the former holds. Then, the exchange property (M-EXC) of M-convex function f applied to $x^\bullet, y$, and i implies that there exists some $j \in \operatorname{supp}^-(x^\bullet - y)$ such that

$$f(x^\bullet) + f(y) \geq f(x^\bullet - \chi_i + \chi_j) + f(y + \chi_i - \chi_j). \tag{4.3}$$

Hence, if we have $f(x^\bullet) < f(x^\bullet - \chi_i + \chi_j)$, then (4.3) implies the desired inequality $f(y - \chi_i + \chi_j) < f(y)$. In the following, we prove $f(x^\bullet) < f(x^\bullet - \chi_i + \chi_j)$.

By the choice of i, we have $\|(x^\bullet - \chi_i + \chi_j) - x_{\mathrm{c}}\|_1 - \|x^\bullet - x_{\mathrm{c}}\|_1 \in \{0, -2\}$. If $\|(x^\bullet - \chi_i + \chi_j) - x_{\mathrm{c}}\|_1 - \|x^\bullet - x_{\mathrm{c}}\|_1 = 0$ then we have $f(x^\bullet) < f(x^\bullet - \chi_i + \chi_j)$ by the choice of $x^\bullet$ since $\|(x^\bullet - \chi_i + \chi_j) - y\|_1 < \|x^\bullet - y\|_1$. If $\|(x^\bullet - \chi_i + \chi_j) - x_{\mathrm{c}}\|_1 - \|x^\bullet - x_{\mathrm{c}}\|_1 = -2$ then we have $\|(x^\bullet - \chi_i + \chi_j) - x_{\mathrm{c}}\|_1 < 2\tau$ and therefore $f(x^\bullet) < f(x^\bullet - \chi_i + \chi_j)$ holds by the definition of τ. Hence, we have $f(x^\bullet) < f(x^\bullet - \chi_i + \chi_j)$ in either case. □

We now prove the claims (i), (ii), and (iii) of Theorem 4.1 in turn.

Proof of Theorem 4.1 (i). We have $M_\sigma \subseteq \{x \in \mathbb{Z}^n \mid \|x - x_{\mathrm{c}}\|_1 = 2\sigma\}$ by the definition of σ. In the following, we prove by induction on k that $\mu_k > \mu_{k+1}$ and $M_{k+1} \subseteq \{x \in \mathbb{Z}^n \mid \|x - x_{\mathrm{c}}\|_1 = 2(k+1)\}$ for each integer $k \in [\sigma, \tau - 1]$.

Let $y \in M_k$, and $x^\bullet \in M_\tau$ be a vector that minimizes the value $\|x^\bullet - y\|_1$. By the induction hypothesis we have $\|y - x_{\mathrm{c}}\|_1 = 2k$. By Lemma 4.9, there exists a monotone sequence $y_0, y_1, \dots, y_h \in \operatorname{dom} f$ with $h = \|x^\bullet - y\|_1$ such that $y_0 = y$, $y_h = x^\bullet$, and $\mu_k = f(y_0) > f(y_1) > \cdots > f(y_h)$. Since $\|y_{t+1} - x_{\mathrm{c}}\|_1 - \|y_t - x_{\mathrm{c}}\|_1 \in \{-2, 0, +2\}$ for every integer $t \in [0, h-1]$ and $\|y_h - x_{\mathrm{c}}\|_1 = 2\tau > 2k \geq \|y_0 - x_{\mathrm{c}}\|_1$, there exists some integer $s \in [1, h]$ such that $\|y_s - x_{\mathrm{c}}\|_1 = 2(k+1)$; such s satisfies $\mu_{k+1} \leq f(y_s) < f(y_0) = \mu_k$.

The inclusion $M_{k+1} \subseteq \{x \in \mathbb{Z}^n \mid \|x - x_{\mathrm{c}}\|_1 = 2(k+1)\}$ follows from the inequality $\mu_{k+1} < \mu_k$ since $f(x) \geq \mu_k > \mu_{k+1}$ holds for every $x \in \operatorname{dom} f$ with $\|x - x_{\mathrm{c}}\|_1 < 2(k+1)$. □

Proof of Theorem 4.1 (ii). We fix $y \in M_k$, and let $\tilde{y}$ be a vector in M_{k+1} that minimizes $\|\tilde{y} - y\|_1$. By Lemma 4.7, it suffices to consider the following two cases:

Case 1: $\operatorname{supp}^+(\tilde{y} - y) \cap \operatorname{supp}^+(\tilde{y} - x_{\mathrm{c}}) \neq \emptyset$,
Case 2: $\operatorname{supp}^-(\tilde{y} - y) \cap \operatorname{supp}^-(\tilde{y} - x_{\mathrm{c}}) \neq \emptyset$.

In the following we give a proof for Case 1 only since Case 2 can be proven in a similar way.

Suppose that there exists some $i \in \operatorname{supp}^+(\tilde{y} - y) \cap \operatorname{supp}^+(\tilde{y} - x_{\mathrm{c}})$. By (M-EXC) applied to $\tilde{y}$ and y, there exists some $j \in \operatorname{supp}^-(\tilde{y} - y)$ such that

$$f(\tilde{y}) + f(y) \geq f(\tilde{y} - \chi_i + \chi_j) + f(y + \chi_i - \chi_j). \tag{4.4}$$

Put $\tilde{z} = \tilde{y} - \chi_i + \chi_j$, $z = y + \chi_i - \chi_j$, and

$$\alpha = \|\tilde{z} - x_{\mathrm{c}}\|_1 - \|\tilde{y} - x_{\mathrm{c}}\|_1, \quad \beta = \|z - x_{\mathrm{c}}\|_1 - \|y - x_{\mathrm{c}}\|_1.$$

Then, we have $\beta \in \{-2, 0, +2\}$ and $\alpha \in \{-2, 0\}$ since $\tilde{y}(i) > x_{\rm c}(i)$.

Assume first that $\alpha = 0$ holds. By Lemma 4.8, we have

$$\alpha + \beta = \|\tilde{z} - x_{\rm c}\|_1 + \|z - x_{\rm c}\|_1 - \|\tilde{y} - x_{\rm c}\|_1 - \|y - x_{\rm c}\|_1 \leq 0,$$

which, together with $\alpha = 0$, implies $\beta \leq 0$. Hence, it holds that $\|z - x_{\rm c}\|_1 \leq \|y - x_{\rm c}\|_1 = 2k$, implying $f(z) \geq \mu_k$. Since $\|\tilde{z} - x_{\rm c}\|_1 = \|\tilde{y} - x_{\rm c}\|_1 = 2(k+1)$, we have $f(\tilde{z}) \geq \mu_{k+1}$. Combining these inequalities with (4.4), we have

$$\mu_{k+1} + \mu_k = f(\tilde{y}) + f(y) \geq f(\tilde{z}) + f(z) \geq \mu_{k+1} + \mu_k,$$

from which follows that $f(\tilde{z}) = \mu_{k+1}$, a contradiction to the choice of $\tilde{y}$ since $\|\tilde{z} - y\|_1 = \|\tilde{y} - y\|_1 - 2$. This shows that $\alpha = 0$ cannot occur. Hence, we have $\alpha = -2$.

Since $\alpha = -2$, we have $\|\tilde{z} - x_{\rm c}\|_1 = \|\tilde{y} - x_{\rm c}\|_1 - 2 = 2k$, from which follows that $f(\tilde{z}) \geq \mu_k$. We also have $\|z - x_{\rm c}\|_1 \leq \|y - x_{\rm c}\|_1 + 2 = 2(k+1)$, and therefore $f(z) \geq \mu_{k+1}$. Combining these inequalities with (4.4), we have

$$\mu_{k+1} + \mu_k = f(\tilde{y}) + f(y) \geq f(\tilde{z}) + f(z) \geq \mu_k + \mu_{k+1},$$

from which follows that $f(z) = \mu_{k+1}$. This implies $\|z - x_{\rm c}\|_1 = 2(k+1)$ by Theorem 4.1 (i). Hence, we have $z = y + \chi_i - \chi_j \in M_{k+1}$ with $i \in N \setminus \mathrm{supp}^-(y - x_{\rm c})$ and $j \in N \setminus \mathrm{supp}^+(y - x_{\rm c})$. □

Proof of Theorem 4.1 (iii). The proof below is quite similar to that for Theorem 4.1 (ii) and omitted. □

5 Polynomial-Time Algorithms for (MML1)

In this section we show that the problem (MML1) can be solved in polynomial time by presenting two algorithms. The first one is based on the reduction to the minimization of the sum of two M-convex functions and its running time is dependent on the function value of the objective function f, while the second one is based on the reduction to the minimization of a single M-convex function and its running time is independent of the function value of f.

We may assume, without loss of generality, that the value τ in (4.1) satisfies $\tau > \gamma$ since otherwise an optimal solution of (MML1) can be obtained easily by solving unconstrained M-convex function minimization problem. Let $x^\bullet \in \mathrm{dom}\, f$ be a minimizer of f with $\|x^\bullet - x_{\rm c}\|_1 = 2\tau$, which is fixed throughout this section.

5.1 Reduction to Problem with Linear Constraints

We first show that the L1-distance constraint $\|x - x_{\rm c}\|_1 \leq 2\gamma$ in (MML1) can be replaced with a system of linear constraints. Let us consider the following problem:

$$\text{(MM-L)} \left| \begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & x(N) = \theta, \\ & x(P) = x_{\rm c}(P) + \gamma, \\ & \hat{\ell} \leq x \leq \hat{u}, \\ & x \in \mathrm{dom}\, f, \end{array} \right.$$

where $P = \mathrm{supp}^+(x^\bullet - x_{\rm c})$, and $\hat{\ell}, \hat{u} \in \mathbb{Z}^n$ are vectors given by

$$\hat{\ell}(i) = \begin{cases} x_{\rm c}(i) & (i \in P), \\ \max\{x^\bullet(i), x_{\rm c}(i) - \gamma\} & (i \in N \setminus P), \end{cases} \qquad \hat{u}(i) = \begin{cases} \min\{x^\bullet(i), x_{\rm c}(i) + \gamma\} & (i \in P), \\ x_{\rm c}(i) & (i \in N \setminus P). \end{cases}$$

Lemma 5.1. *Every optimal solution of* (MM-L) *is also optimal for* (MML1).

Proof. We first show that every feasible solution of (MM-L) is a feasible solution of (MML1). For this, it suffices to prove that every feasible solution x of (MM-L) satisfies the L1-distance constraint $\|x - x_{\mathrm{c}}\|_1 \leq 2\gamma$. Under the condition $\hat{\ell} \leq x \leq \hat{u}$ we have

$$\|x - x_{\mathrm{c}}\|_1 = (x(P) - x_{\mathrm{c}}(P)) + (x_{\mathrm{c}}(N \setminus P) - x(N \setminus P)),$$

and the equation $x(N) = \theta = x_{\mathrm{c}}(N)$ implies that $x(P) - x_{\mathrm{c}}(P) = x_{\mathrm{c}}(N \setminus P) - x(N \setminus P)$. Since $x(P) = x_{\mathrm{c}}(P) + \gamma$, the L1-distance $\|x - x_{\mathrm{c}}\|_1$ is bounded by 2γ.

To conclude the proof, we show that there exists an optimal solution x^* of (MML1) that is a feasible solution of (MM-L), i.e., x^* satisfies $x^*(P) = x_{\mathrm{c}}(P) + \gamma$ and $\hat{\ell} \leq x^* \leq \hat{u}$. Repeated application of Theorem 4.1 (iii) implies that there exists an optimal solution $x^* \in \mathrm{dom}\, f$ of (MML1) such that $\|x^* - x_{\mathrm{c}}\|_1 = 2\gamma$, $x^*(P) = x_{\mathrm{c}}(P) + \gamma$, $x^*(N \setminus P) = x_{\mathrm{c}}(N \setminus P) - \gamma$, and

$$x_{\mathrm{c}}(i) \leq x^*(i) \leq x^\bullet(i) \ (i \in P), \quad x^\bullet(i) \leq x^*(i) \leq x_{\mathrm{c}}(i) \ (i \in N \setminus P). \tag{5.1}$$

By the equation $x^*(P) = x_{\mathrm{c}}(P) + \gamma$ and the former inequality in (5.1), we have

$$x^*(i) = x_{\mathrm{c}}(P) + \gamma - x^*(P \setminus \{i\}) \leq x_{\mathrm{c}}(i) + \gamma \qquad (\forall i \in P).$$

Similarly, for $i \in N \setminus P$ we have $x^*(i) \geq x_{\mathrm{c}}(i) - \gamma$. Hence, $\hat{\ell} \leq x^* \leq \hat{u}$ holds. □

While the problem (MM-L) does not fit into the framework of M-convex function minimization problem, due to the constraint $x(P) = x_{\mathrm{c}}(P) + \gamma$, it can be seen as the minimization of the sum of two M-convex functions. Indeed, (MM-L) is equivalent to the minimization of the sum of functions $f_1, f_2 : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ given by

$$f_1(x) = \begin{cases} f(x) & (\text{if } x(N) = \theta,\ \hat{\ell} \leq x \leq \hat{u}), \\ +\infty & (\text{otherwise}), \end{cases}$$

$$f_2(x) = \begin{cases} 0 & (\text{if } x(N) = \theta,\ x(P) = x_{\mathrm{c}}(P) + \gamma,\ \hat{\ell} \leq x \leq \hat{u}), \\ +\infty & (\text{otherwise}). \end{cases}$$

It is not difficult to show that f_1 and f_2 satisfy (M-EXC), i.e., the two functions f_1 and f_2 are M-convex.

It is known that minimization of the sum of two M$^\natural$-convex functions $f_1, f_2 : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ can be solved in polynomial time (see, e.g., [11]), and the fastest algorithm runs in $\mathrm{O}(n^6 (\log L)^2 \log(nK))$ time [5], where L is the maximum of the L_∞-diameter of $\mathrm{dom}\, f_1$ and $\mathrm{dom}\, f_2$ (see (2.4) for the definition of L_∞-diameter) and

$$K = \max_{h=1,2} \max\{|f_h(x) - f_h(y)| \mid x, y \in \mathrm{dom}\, f_h\}.$$

For the functions f_1 and f_2 defined above, the L_∞-diameter of $\mathrm{dom}\, f_1$ and $\mathrm{dom}\, f_2$ is bounded by $\max_{i \in N}\{\hat{u}(i) - \hat{\ell}(i)\} \leq \gamma$. Hence, we obtain the following result.

Theorem 5.2. *The problem* (MML1) *can be solved in* $\mathrm{O}(n^6 (\log \gamma)^2 \log(nK_f))$ *time with* $K_f = \max\{|f(x) - f(y)| \mid x, y \in \mathrm{dom}\, f\}$.

Note that this time bound contains the parameter K_f and is dependent on the function value of f.

5.2 Reduction to M-convex Function Minimization

We now explain an alternative approach to solve the problem (MM-L) by the reduction to the minimization of an M-convex function.

For a vector $y \in \mathbb{Z}^{N\setminus P}$, define a set $T(y) \subseteq \mathbb{Z}^n$ by

$$T(y) = \{x \in \mathrm{dom}\, f \mid x(N) = \theta,\ x(i) = y(i)\ (i \in N \setminus P),\ \hat{\ell}(i) \le x(i) \le \hat{u}(i)\ (i \in P)\}.$$

Then, define a function $g : \mathbb{Z}^{N\setminus P} \to \mathbb{R} \cup \{+\infty\}$ as follows:

$$g(y) = \begin{cases} \min\{f(x) \mid x \in T(y)\} & (\text{if } y(N \setminus P) = \theta - (x_{\mathrm{c}}(P) + \gamma) \\ & \quad \text{and } \hat{\ell}(i) \le y(i) \le \hat{u}(i)\ (\forall i \in N \setminus P)), \\ +\infty & (\text{otherwise}). \end{cases} \tag{5.2}$$

By definition, $x \in \mathbb{Z}^n$ is a feasible solution of (MM-L) if and only if the vector $y \in \mathbb{Z}^{N\setminus P}$ given by $y(i) = x(i)$ $(i \in N \setminus P)$ satisfies $y \in \mathrm{dom}\, g$ and $x \in T(y)$. Therefore, the problem (MM-L) can be reduced to the minimization of function g; for a minimizer $y^* \in \mathbb{Z}^{N\setminus P}$ of g, the vector $x^* \in T(y^*)$ with $g(y^*) = f(x^*)$ is an optimal solution of (MM-L).

We show that function g is an M-convex function, which implies that existing algorithms can be applied to the minimization of g.

Proposition 5.3. *Function g is M-convex.*

Proof. Define a function $f' : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ by

$$f'(x) = \begin{cases} f(x) & (\text{if } x(N) = \theta \text{ and } \hat{\ell}(i) \le x(i) \le \hat{u}(i)\ (\forall i \in P)), \\ +\infty & (\text{otherwise}). \end{cases} \tag{5.3}$$

Since f is M-convex, the function f' is also an M-convex function [11, Theorem 6.13 (5)]. The function $g' : \mathbb{Z}^{N\setminus P} \to \mathbb{R} \cup \{+\infty\}$ given by

$$g'(y) = \min\{f'(x) \mid x(i) = y(i)\ (i \in N \setminus P)\} \qquad (y \in \mathbb{Z}^{N\setminus P})$$

is an M$^\natural$-convex function since g' is a projection of f' [11, Theorem 6.15 (2)]. Finally, function g in (5.2) is given as

$$g(y) = \begin{cases} g'(y) & (\text{if } y(N \setminus P) = \theta - (x_{\mathrm{c}}(P) + \gamma),\ \hat{\ell}(i) \le x(i) \le \hat{u}(i)\ (i \in N \setminus P)), \\ +\infty & (\text{otherwise}), \end{cases}$$

and therefore g is M-convex (cf. [11, Theorem 6.13]). $\square$

We analyze the running time of the algorithm. By Theorem 2.7, the minimization of g can be done in $\mathrm{O}(n^3 \log(\gamma/n) F_g)$ time, where F_g denotes the time to evaluate the function value of g. The evaluation of the function value $g(y)$ can be seen as the minimization of the function f' in (5.3) under the constraint $x(i) = y(i)$ $(i \in N \setminus P)$. Since the L_∞-diameter of $\mathrm{dom}\, f'$ is bounded by γ, the evaluation of g can be done in $\mathrm{O}(n^3 \log(\gamma/n))$ time by Theorem 2.7, provided that the function evaluation of f can be done in constant time. Hence, we obtain the following time bound:

Theorem 5.4. *The problem* (MML1) *can be solved in* $\mathrm{O}(n^6 (\log(\gamma/n))^2)$ *time.*

6 Application to Dock Re-allocation Problem

As observed in Section 3, the dock re-allocation problem (DR) can be seen as a special case of the problem (MML1). In this section, we apply the results obtained in Sections 4 and 5 for (MML1) to obtain algorithms for (DR). In particular, we show that the problem (DR) can be solved in polynomial time.

6.1 Steepest Descent Algorithm

We first propose a steepest descent algorithm for (DR) by applying the algorithm in Section 4 to (DR). We also show that a fast implementation of the steepest descent algorithm coincides with the greedy algorithm by Freund et al. [1].

Recall that (DR) can be reformulated in the form of (MML1) as

$$\left|\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & x(N) = D + B, \\ & \|x - (\bar{d} + \bar{b})\|_1 \leq 2\gamma, \\ & x \in \operatorname{dom} f, \end{array}\right.$$

where the M-convex function $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ is given by

$$\begin{aligned} &\operatorname{dom} f = \{x \in \mathbb{Z}^n \mid x(N) = D + B,\ \ell \leq x \leq u\}, \\ &f(x) = \min\{c(d, b)\ \mid d, b \in \mathbb{Z}^n_+,\ d + b = x,\ b(N) \leq B\} \quad (x \in \operatorname{dom} f). \end{aligned} \tag{6.1}$$

By definition, the function value $f(x)$ for a given $x \in \operatorname{dom} f$ can be computed by solving the following problem:

$$(\mathrm{SRA}(x)) \left|\begin{array}{ll} \text{Minimize} & c(x - b, b) \equiv \sum_{i=1}^n c_i(x(i) - b(i), b(i)) \\ \text{subject to} & b(N) \leq B, \\ & \mathbf{0} \leq b \leq x, \\ & b \in \mathbb{Z}^n_+. \end{array}\right.$$

For each $i \in N$, $c_i(x(i) - b(i), b(i))$ is a convex function in variable $b(i)$ since c_i is a multimodular (or M$^\natural$-convex) function. Hence, the problem (SRA(x)) can be seen as a simple resource allocation problem and therefore the evaluation of the function value of f can be done efficiently.

Proposition 6.1 ([3, 4])**.** *The problem* (SRA(x)) *can be solved in* $\mathrm{O}(n \log(B/n))$ *time and in* $\mathrm{O}(n + B \log n)$ *time. Moreover, if a feasible solution* $b \in \mathbb{Z}^n_+$ *of* (SRA(x)) *is available, then the problem can be solved in* $\mathrm{O}(n + B' \log n)$ *time with*

$$B' = \min\{\|b' - b\|_1 \mid b' \textit{ is an optimal solution of } (\mathrm{SRA}(x))\}.$$

We say that a feasible solution (d, b) of the problem (DR) is *bike-optimal* if the vector b is an optimal solution of the problem (SRA($d + b$)). Throughout Section 6, we assume that $(\bar{d}, \bar{b})$ is a bike-optimal solution. If $(\bar{d}, \bar{b})$ is not bike-optimal, then we compute an optimal solution b^* of the problem (SRA($\bar{d}+\bar{b}$)) and replace $(\bar{d}, \bar{b})$ with $(\bar{d}+\bar{b}-b^*, b^*)$; note that with this replacement the sum $\bar{d}+\bar{b}$ remains the same. This assumption requires an extra time for solving (SRA($\bar{d}+\bar{b}$)), which does not affect the total running time of the algorithms proposed in this section.

The algorithm SteepestDescentMML1 is rewritten in terms of the problem (DR) as follows. Since $(\bar{d}, \bar{b})$ is a feasible solution of the problem (DR), the vector $\bar{x} = \bar{d} + \bar{b}$ can be used as the initial solution of the steepest descent algorithm.

Algorithm STEEPESTDESCENTDR
Step 0: Set $x_0 := \bar{d} + \bar{b}$ and $k := 1$.
Step 1: If $k - 1 = \gamma$, then output the solution $(x_{k-1} - b_{k-1}, b_{k-1})$ and stop.
Step 2: For every distinct $i, j \in N$, compute the value $f(x_{k-1} + \chi_i - \chi_j)$ by solving (SRA($x_{k-1} + \chi_i - \chi_j$)), and find $i_k, j_k \in N$ minimizing $f(x_{k-1} + \chi_{i_k} - \chi_{j_k})$.
Step 3: If $f(x_{k-1} + \chi_{i_k} - \chi_{j_k}) \geq f(x_{k-1})$, output the solution $(x_{k-1} - b_{k-1}, b_{k-1})$ and stop. Otherwise, compute an optimal solution b_k of (SRA($x_{k-1} + \chi_{i_k} - \chi_{j_k}$)), set $x_k := x_{k-1} + \chi_{i_k} - \chi_{j_k}$, $k := k + 1$, and go to Step 1.

Since the evaluation of the function value $f(x)$ requires $\mathrm{O}(n \log(B/n))$ time, each iteration requires $\mathrm{O}(n^3 \log(B/n))$ time, and the total running time of the algorithm is $\mathrm{O}(\gamma n^3 \log(B/n))$.

The next lemma shows that the evaluation of the value $f(x)$ can be done faster by maintaining an optimal solution of the problem (SRA(x_k)) for each k. This lemma is essentially equivalent to Lemma 6 in Freund et al. [1], while the statement of the lemma is described in our notation. For completeness, the proof of the lemma is given in Appendix.

Lemma 6.2 ([1, Lemma 6])**.** *Let $x \in \operatorname{dom} f$, and $b \in \mathbb{Z}^n$ be an optimal solution of the problem* (SRA(x))*. Also, let $i, j \in N$ be distinct elements such that $x + \chi_i - \chi_j \in \operatorname{dom} f$. Then, there exists an optimal solution $\hat{b} \in \mathbb{Z}^n$ of the problem* (SRA($x + \chi_i - \chi_j$)) *such that*

$$\begin{aligned}\hat{b} \in \{&b, b + \chi_i, b - \chi_j, b + \chi_i - \chi_j\} \\ &\cup \{b + \chi_i - \chi_t \mid t \in N \setminus \{i, j\}\} \cup \{b + \chi_s - \chi_j \mid s \in N \setminus \{i, j\}\}.\end{aligned} \tag{6.2}$$

It follows from Lemma 6.2 that for each $i, j \in N$, an optimal solution of the problem (SRA($x + \chi_i - \chi_j$)) can be found in $\mathrm{O}(n)$ time, provided that an optimal solution of the problem (SRA(x)) is available. Therefore, the function value $f(x)$ can be evaluated in $\mathrm{O}(n)$ time, and the running time of the algorithm STEEPESTDESCENTDR is reduced to $\mathrm{O}(\gamma\, n^3)$.

Moreover, Lemma 6.2 implies that the running time $\mathrm{O}(n^3)$ in each iteration can be further reduced to $\mathrm{O}(\log n)$ by computing elements $i_k, j_k \in N$ minimizing the value $f(x_{k-1} + \chi_{i_k} - \chi_{j_k})$ and an optimal solution of the problem (SRA($x + \chi_{i_k} - \chi_{j_k}$)) simultaneously. We denote

$$R = \{(d, b) \in \mathbb{Z}^n \times \mathbb{Z}^n \mid d(N) + b(N) = D + B,\ b(N) \leq B,\ \ell \leq d + b \leq u,\ d \geq \mathbf{0},\ b \geq \mathbf{0}\},$$

i.e., R is the set of vectors $(d, b) \in \mathbb{Z}^n \times \mathbb{Z}^n$ satisfying the constraints of the problem (DR), except for the L1-distance constraint $\|(\bar{d} + \bar{b}) - (d + b)\|_1 \leq 2\gamma$. We also define

$$\begin{aligned}
N(d, b) &= N_1(d, b) \cup N_2(d, b) \cup \cdots \cup N_6(d, b), && (6.3)\\
N_1(d, b) &= \{(d + \chi_i - \chi_j, b) \in \mathbb{Z}^n \times \mathbb{Z}^n \mid i, j \in N,\ i \neq j\},\\
N_2(d, b) &= \{(d - \chi_j, b + \chi_i) \in \mathbb{Z}^n \times \mathbb{Z}^n \mid i, j \in N,\ i \neq j\},\\
N_3(d, b) &= \{(d + \chi_i, b - \chi_j) \in \mathbb{Z}^n \times \mathbb{Z}^n \mid i, j \in N,\ i \neq j\},\\
N_4(d, b) &= \{(d, b + \chi_i - \chi_j) \in \mathbb{Z}^n \times \mathbb{Z}^n \mid i, j \in N,\ i \neq j\},\\
N_5(d, b) &= \{(d - \chi_j + \chi_t, b + \chi_i - \chi_t) \in \mathbb{Z}^n \times \mathbb{Z}^n \mid i, j \in N,\ i \neq j,\ t \in N \setminus \{i, j\}\},\\
N_6(d, b) &= \{(d - \chi_s + \chi_i, b + \chi_s - \chi_j) \in \mathbb{Z}^n \times \mathbb{Z}^n \mid i, j \in N,\ i \neq j,\ s \in N \setminus \{i, j\}\}.
\end{aligned}$$

The following property follows immediately from Lemma 6.2.

Lemma 6.3. *Let $x \in \operatorname{dom} f$, and $b \in \mathbb{Z}^n$ be an optimal solution of* (SRA(x))*. Then, we have*

$$\min\{f(x + \chi_i - \chi_j) \mid i, j \in N,\ \ell \leq x + \chi_i - \chi_j \leq u\} = \min\{c(d', b') \mid (d', b') \in N(d, b) \cap R\}.$$

By Lemma 6.3, the algorithm SteepestDescentDR can be rewritten in terms of original variables (d, b) as follows, which is nothing but the greedy algorithm by Freund et al. [1].

Algorithm SteepestDescentDR$'$
Step 0: Set $d_0 := \bar{d}$, $b_0 := \bar{b}$, and $k := 1$.
Step 1: If $k - 1 = \gamma$, then output the solution (d_{k-1}, b_{k-1}) and stop.
Step 2: Find $(d', b') \in N(d_{k-1}, b_{k-1}) \cap R$ that minimizes $c(d', b')$.
Step 3: If $c(d', b') \geq c(d_{k-1}, b_{k-1})$, then output the solution (d_{k-1}, b_{k-1}) and stop.
Otherwise, set $(d_k, b_k) := (d', b')$, $k := k + 1$, and go to Step 1.

For $h = 1, 2, \ldots, 6$, the value $\min\{c(d', b') \mid (d', b') \in N_h(d_{k-1}, b_{k-1}) \cap R\}$ can be computed in $\mathrm{O}(\log n)$ time by using six binary heaps that maintain the following six sets of numbers, as in [1, Section 3.1]:

$$
\begin{aligned}
&\{c_i(d_{k-1}(i) + 1, b(i)) - c_i(d_{k-1}(i), b(i)) \mid i \in N\},\\
&\{c_i(d_{k-1}(i) - 1, b(i)) - c_i(d_{k-1}(i), b(i)) \mid i \in N\},\\
&\{c_i(d(i), b(i) + 1) - c_i(d(i), b(i)) \mid i \in N\},\\
&\{c_i(d(i), b(i) - 1) - c_i(d(i), b(i)) \mid i \in N\},\\
&\{c_i(d(i) + 1, b(i) - 1) - c_i(d(i), b(i)) \mid i \in N\},\\
&\{c_i(d(i) - 1, b(i) + 1) - c_i(d(i), b(i)) \mid i \in N\}.
\end{aligned}
$$

Hence, each iteration of the algorithm can be done in $\mathrm{O}(\log n)$ time. Since the initialization of the heaps requires $\mathrm{O}(n)$ time, we obtain the following result:

Theorem 6.4 ([1]). *Suppose that $(\bar{d}, \bar{b})$ is a bike-optimal solution of the problem* (DR). *Then, the algorithm* SteepestDescentDR$'$ *finds an optimal solution of* (DR) *in* $\mathrm{O}(n + \gamma \log n)$ *time.*

6.2 Polynomial-Time Solvability of (DR)

The running time of the algorithm SteepestDescentDR is proportional to the problem parameter γ and therefore is not polynomial. We show that the problem (DR) can be solved in polynomial time.

Since the problem (DR) can be reformulated as (MML1) with the M-convex objective function $f : \mathbb{Z}^n \to \mathbb{R} \cup \{+\infty\}$ given by (6.1), we obtain the following result from Theorem 5.4. Recall that the evaluation of function value of f can be done in $\mathrm{O}(n \log(B/n))$ time by Proposition 6.1.

Proposition 6.5. *The problem* (DR) *can be solved in* $\mathrm{O}(n^7 (\log(\gamma/n))^2 \log(B/n))$ *time.*

To obtain a better time bound for (DR), we consider a different approach. In this approach, we consider the following problem, denoted as (DA), obtained from (DR) by relaxing the L1-distance constraint $\|(d + b) - (\bar{d} + \bar{b})\|_1 \leq 2\gamma$.

$$
\text{(DA)}\ \left|\ \begin{array}{ll}
\text{Minimize} & c(d, b)\\
\text{subject to} & d(N) + b(N) = D + B,\\
& b(N) \leq B,\\
& \ell \leq d + b \leq u,\\
& d, b \in \mathbb{Z}^n_+.
\end{array}\right.
$$

We will show in Section 6.3 that (DA) can be solved in $\mathrm{O}(n \log n \log((D + B)/n))$ time by using scaling approach, which is of interest in its own right.

Theorem 6.6. *A proximity-scaling algorithm finds an optimal solution of the problem* (DA) *in* $\mathrm{O}(n \log n \log((D+B)/n))$ *time.*

The discussion in Section 5 shows that using an optimal solution $(d^\bullet, b^\bullet)$ of (DA), the problem (DR) can be reformulated as the following problem without L1-distance constraint:

$$
\text{(DR-L)} \quad \left| \begin{array}{ll} \text{Minimize} & c(d, b) \\ \text{subject to} & d(N) + b(N) = D + B, \\ & b(N) \leq B, \\ & d(P) + b(P) = \bar{d}(P) + \bar{b}(P) + \gamma, \\ & d(N \setminus P) + b(N \setminus P) = \bar{d}(N \setminus P) + \bar{b}(N \setminus P) - \gamma, \\ & \ell \leq d + b \leq u, \\ & d, b \in \mathbb{Z}^n_+. \end{array} \right.
$$

where the set $P \subseteq N$ is given as

$$
P = \mathrm{supp}^+((d^\bullet + b^\bullet) - (\bar{d} + \bar{b})).
$$

Note that the first equality constraint $d(N) + b(N) = D + B$ is redundant and follows from other equality constraints.

To solve the problem (DR-L) efficiently, we consider the two subproblems (DR-L-A(α)) and (DR-L-B(α)) with parameter $\alpha \in [0, B]$:

$$
\text{(DR-L-A}(\alpha)) \quad \left| \begin{array}{ll} \text{Minimize} & \sum_{i \in P} c_i(d(i), b(i)) \\ \text{subject to} & b(P) \leq \alpha, \\ & d(P) + b(P) = \bar{d}(P) + \bar{b}(P) + \gamma, \\ & \ell(i) \leq d(i) + b(i) \leq u(i), \\ & d(i), b(i) \in \mathbb{Z}_+ \ (i \in P); \end{array} \right.
$$

(DR-L-B(α)) is defined similarly to (DR-L-A(α)), where P is replaced with $N \setminus P$ and the first constraint $b(P) \leq \alpha$ is replaced with $b(N \setminus P) \leq B - \alpha$. The two subproblems have (almost) the same structure as the problem (DA), and therefore can be solved in $\mathrm{O}(n \log n \log((D+B)/n))$ time by Theorem 6.6.

We denote by $\psi_{\mathrm{A}}(\alpha)$ (resp., $\psi_{\mathrm{B}}(\alpha)$) the optimal value of the problem (DR-L-A(α)) (resp., (DR-L-B(α))). Then, the optimal value of the problem (DR-L) is equal to

$$
\min\{\psi_{\mathrm{A}}(\alpha) + \psi_{\mathrm{B}}(\alpha) \mid \alpha \in [0, B] \cap \mathbb{Z}\}.
$$

The next property shows that the minimum value of $\psi_{\mathrm{A}}(\alpha) + \psi_{\mathrm{B}}(\alpha)$ can be computed by binary search with respect to α.

Proposition 6.7. *The values* $\psi_{\mathrm{A}}(\alpha)$ *and* $\psi_{\mathrm{B}}(\alpha)$ *are convex functions in* $\alpha \in [0, B] \cap \mathbb{Z}$.

Proof is given at the end of this section. Since the binary search terminates in $\mathrm{O}(\log B)$ iterations and each iteration requires $\mathrm{O}(n \log n \log((D+B)/n))$ time by Theorem 6.6, we obtain the following time bound.

Theorem 6.8. *The problem* (DR) *can be solved in* $\mathrm{O}(n \log n \log((D+B)/n) \log B)$ *time.*

We now give a proof of Proposition 6.7. Consider a variant of the problem (DA), where the constant B is replaced with a parameter α:

$$
\text{(DA}[\alpha]) \quad \left| \begin{array}{ll} \text{Minimize} & c(d, b) \\ \text{subject to} & d(N) + b(N) = D + B, \\ & b(N) \leq \alpha, \\ & \ell \leq d + b \leq u, \\ & d, b \in \mathbb{Z}^n_+. \end{array} \right.
$$

We denote by $\psi(\alpha)$ the optimal value of (DA[α]). To prove Proposition 6.7, it suffices to show the following property of $\psi(\alpha)$.

Lemma 6.9. *The value $\psi(\alpha)$ is a convex function in $\alpha \in [0, B] \cap \mathbb{Z}$.*

Proof. We show that $\psi(\alpha) + \psi(\alpha + 2) \geq 2\psi(\alpha + 1)$ holds for $\alpha \in [0, B - 2] \cap \mathbb{Z}$.

Let $(d, b) \in \mathbb{Z}^n \times \mathbb{Z}^n$ be an optimal solution of (DA[α]). Also, let $(\hat{d}, \hat{b}) \in \mathbb{Z}^n \times \mathbb{Z}^n$ be an optimal solution of (DA[$\alpha + 2$]) that has the minimum value of $\|\hat{d} - d\|_1 + \|\hat{b} - b\|_1$. Note that $c(d, b) = \psi(\alpha)$ and $c(\hat{d}, \hat{b}) = \psi(\alpha + 2)$ hold. Since (d, b) is a feasible solution of (DA[$\alpha + 1$]), we have $\psi(\alpha) \geq \psi(\alpha + 1)$. Hence, if $(\hat{d}, \hat{b})$ is a feasible solution of (DA[$\alpha + 1$]) (i.e., $\hat{b}(N) \leq \alpha + 1$), then we have $\psi(\alpha + 1) \leq \psi(\alpha + 2)$ and therefore the inequality $\psi(\alpha) + \psi(\alpha + 2) \geq 2\psi(\alpha + 1)$ follows. Therefore, we may assume $\hat{b}(N) = \alpha + 2$ in the following.

Since $\hat{b}(N) = \alpha+2 > \alpha = b(N)$ and $\hat{d}(N)+\hat{b}(N) = d(N)+b(N)$, it holds that $\mathrm{supp}^+(\hat{b}-b) \neq \emptyset$ and $\mathrm{supp}^-(\hat{d} - d) \neq \emptyset$. We first consider the case where there exists some $i \in \mathrm{supp}^+(\hat{b} - b) \cap \mathrm{supp}^-(\hat{d} - d)$. Then, Proposition 2.4 (i) implies that

$$c_i(\hat{d}(i), \hat{b}(i)) + c_i(d(i), b(i)) \geq c_i(\hat{d}(i) + 1, \hat{b}(i) - 1) + c_i(d(i) - 1, b(i) + 1),$$

from which follows that the vectors $\hat{d}' = \hat{d} + \chi_i$, $\hat{b}' = \hat{b} - \chi_i$, $d' = d - \chi_i$, and $b' = b + \chi_i$ satisfy the inequality

$$\psi(\alpha + 2) + \psi(\alpha) = c(\hat{d}, \hat{b}) + d(d, b) \geq c(\hat{d}', \hat{b}') + c(d', b') \geq 2\psi(\alpha + 1),$$

where the last inequality is by the fact that $(\hat{d}', \hat{b}')$ and (d', b') are feasible solutions of (DA[α+1]).

We next consider the case with $\mathrm{supp}^+(\hat{b} - b) \cap \mathrm{supp}^-(\hat{d} - d) = \emptyset$. We have

$$\begin{aligned} \hat{d}(i) &\geq d(i) \qquad (\forall i \in \mathrm{supp}^+(\hat{b} - b)), \\ \hat{b}(j) &\leq b(j) \qquad (\forall j \in \mathrm{supp}^-(\hat{d} - d)), \end{aligned}$$

from which follows that

$$\begin{aligned} \hat{d}(i) + \hat{b}(i) &> d(i) + b(i) \qquad (\forall i \in \mathrm{supp}^+(\hat{b} - b)), \\ \hat{d}(j) + \hat{b}(j) &< d(j) + b(j) \qquad (\forall j \in \mathrm{supp}^-(\hat{d} - d)). \end{aligned}$$

Proposition 2.4 (ii) implies that for arbitrarily chosen $i \in \mathrm{supp}^+(\hat{b} - b)$ and $j \in \mathrm{supp}^-(\hat{d} - d)$, it holds that

$$\begin{aligned} c_i(\hat{d}(i), \hat{b}(i)) + c_i(d(i), b(i)) &\geq c_i(\hat{d}(i), \hat{b}(i) - 1) + c_i(d(i), b(i) + 1), \\ c_j(\hat{d}(j), \hat{b}(j)) + c_j(d(j), b(j)) &\geq c_j(\hat{d}(j) + 1, \hat{b}(j)) + c_j(d(j) - 1, b(j)), \end{aligned}$$

from which follows that the vectors $\hat{d}'' = \hat{d} + \chi_j$, $\hat{b}'' = \hat{b} - \chi_i$, $d'' = d - \chi_j$, and $b'' = b + \chi_i$ satisfy the inequality

$$\psi(\alpha + 2) + \psi(\alpha) = c(\hat{d}, \hat{b}) + d(d, b) \geq c(\hat{d}'', \hat{b}'') + c(d'', b'') \geq 2\psi(\alpha + 1),$$

where the last inequality is by the fact that $(\hat{d}'', \hat{b}'')$ and (d'', b'') are feasible solutions of (DA[$\alpha +$ 1]). $\square$

6.3 Algorithms for (DA)

In this section, we give a proof of Theorem 6.6, stating that the problem (DA) can be solved in $\mathrm{O}(n \log n \log((D+B)/n))$ time. For this, we first present a steepest descent algorithm for (DA) and analyze its running time in Section 6.3.1. We then propose a polynomial-time algorithm that is based on scaling technique in Section 6.3.2. To apply scaling technique to some discrete optimization problem, we need a property called "proximity theorem," which states that in an appropriately chosen neighborhood of an "approximate" solution, there exists an optimal solution. We give a proof of the proximity theorem for the problem (DA) in Section 6.3.3.

6.3.1 Pseudo-Polynomial-Time Algorithm

We show that the problem (DA) can be solved by a steepest descent algorithm for (DA). The problem (DA) can be seen as a special case of (DR), where $(\bar{d}, \bar{b}) = (d_0, b_0)$ is an arbitrarily chosen bike-optimal solution of (DA) and $\gamma = \tau$ with τ given by

$$\tau = \min\{\|(d+b) - (d_0 + b_0)\|_1 \mid (d,b) \text{ is an optimal solution of (DA)}\}. \tag{6.4}$$

This implies that the algorithm SteepestDescentDR$'$ can be specialized to (DA) as follows.

Algorithm SteepestDescentDA
Step 0: Let (d_0, b_0) be an initial solution. Set $k := 1$.
Step 1: Find $(d', b') \in N(d_{k-1}, b_{k-1}) \cap R$ that minimizes $c(d', b')$.
Step 2: If $c(d', b') \geq c(d_{k-1}, b_{k-1})$, then output the solution (d_{k-1}, b_{k-1}) and stop.
Otherwise, set $(d_k, b_k) := (d', b')$, $k := k+1$, and go to Step 1.

The following time complexity result for SteepestDescentDA follows immediately from Theorem 6.4.

Proposition 6.10. *Suppose that the initial solution (d_0, b_0) is a bike-optimal solution of the problem* (DA). *Then, the algorithm* SteepestDescentDA *finds an optimal solution of* (DA) *in $\mathrm{O}(n + \tau \log n)$ time with τ given by* (6.4).

6.3.2 Polynomial-Time Algorithm

We then propose a polynomial-time proximity-scaling algorithm for (DA). In the following, we fix an arbitrarily chosen feasible solution $(\tilde{d}, \tilde{b})$ of (DA).

Let ν be a positive integer. We consider the following problem, a scaled variant of the problem (DA):

$$\text{(DA}(\nu)) \left| \begin{array}{ll} \text{Minimize} & c(d,b) \\ \text{subject to} & d(N) + b(N) = D + B, \\ & b(N) \leq B, \\ & \ell \leq d + b \leq u, \\ & d, b \in \mathbb{Z}^n_+, \\ & d(i) \equiv \tilde{d}(i) \bmod \nu,\ b(i) \equiv \tilde{b}(i) \bmod \nu\ (i \in N). \end{array} \right.$$

A feasible solution (resp., optimal solution) of (DA(ν)) is called a *ν-feasible solution* (resp., *ν-optimal solution*) of (DA). We also consider a scaled variant of the problem (SRA(x)) with

$x \in \mathbb{Z}_+^n$:

$$\text{(SRA}(x,\nu)) \left| \begin{array}{ll} \text{Minimize} & c(x-b,b) \equiv \sum_{i=1}^n c_i(x(i)-b(i),b(i)) \\ \text{subject to} & b(N) \leq B, \\ & \mathbf{0} \leq b \leq x, \\ & b \in \mathbb{Z}_+^n, \\ & b(i) \equiv \tilde{b}(i) \bmod \nu \ (i \in N). \end{array} \right.$$

A feasible solution (d,b) of the problem $(\mathrm{DA}(\nu))$ is called *ν-bike optimal* if b is an optimal solution of the problem $(\mathrm{SRA}(d+b,\nu))$.

Our proximity-scaling algorithm is based on the following proximity theorem for (DA). This implies, in particular, that a ν-optimal solution of (DA) is close to a 2ν-optimal solution.

Theorem 6.11. *Let ν be a positive integer with $\nu \geq 2$, and $(d,b) \in \mathbb{Z}^n \times \mathbb{Z}^n$ be a 2ν-optimal solution of* (DA). *Then, there exists some ν-optimal solution $(d^*,b^*) \in \mathbb{Z}^n \times \mathbb{Z}^n$ of* (DA) *such that*

$$\|(d^*+b^*)-(d+b)\|_1 < 16\nu n.$$

Proof will be given in Section 6.3.3.

We observe that the problem $(\mathrm{DA}(\nu))$ has the same combinatorial structure as (DA). Indeed, for $i \in N$ the function

$$c_i^\nu(\eta,\zeta) = c_i(\nu\eta + \tilde{d}(i), \nu\zeta + \tilde{b}(i))$$

is also a multimodular (or $\mathrm{M}^\natural$-convex) function in (η,ζ). Therefore, any algorithm for (DA) can be applied to $(\mathrm{DA}(\nu))$. In particular, a variant of the algorithm SteepestDescentDA, where step length is set to ν instead of unit step length, is applicable to find its optimal solution of $(\mathrm{DA}(\nu))$. We denote this variant of the algorithm by SteepestDescentDA(ν). The following property is an immediate consequence of Proposition 6.10 for the algorithm SteepestDescentDA.

Proposition 6.12. *Suppose that the initial solution (d_0,b_0) is a ν-bike-optimal solution of the problem* $(\mathrm{DA}(\nu))$. *Then, the algorithm* SteepestDescentDA(ν) *finds an optimal solution of* $(\mathrm{DA}(\nu))$ *in* $\mathrm{O}(n+\tau_\nu \log n)$ *time with τ_ν given by*

$$\tau_\nu = \min\{\|(d+b)-(d_0+b_0)\|_1/\nu \mid (d,b) \textit{ is an optimal solution of } (\mathrm{DA}(\nu))\}.$$

Our proximity-scaling algorithm consists of several scaling phases. Each scaling phase is associated with a scaling parameter ν that is a power of 2. In each scaling phase, a ν-optimal solution is found by the algorithm SteepestDescentDA(ν). To reduce the number of iterations in SteepestDescentDA(ν), we use a 2ν-optimal solution (d,b) computed in the previous scaling phase as an initial solution. The solution (d,b), however, is not a ν-bike-optimal solution in general, and cannot be used as an initial solution of SteepestDescentDA(ν) as it is. Therefore, we need to modify the solution (d,b) to a new solution (d',b') with $d'+b' = d+b$ that is ν-bike optimal, and use the new solution as an initial solution of SteepestDescentDA(ν). In other words, we need to compute an optimal solution of the problem $(\mathrm{SRA}(d+b,\nu))$. The next property shows that this computation can be done efficiently by using the solution (d,b); note that (d,b) is a 2ν-optimal solution of (DA) and therefore 2ν-bike optimal.

Proposition 6.13. *Let $x \in \mathbb{Z}_+^n$, ν be a positive integer with $\nu \geq 2$, and $b \in \mathbb{Z}_+^n$ be an optimal solution of* $(\mathrm{SRA}(x,2\nu))$. *Then, there exists some optimal solution $b^* \in \mathbb{Z}_+^n$ of* $(\mathrm{SRA}(x,\nu))$ *such that*

$$\|b^*-b\|_1 < 2\nu n.$$

Proof. It suffices to consider the case with $\nu = 1$. Let $b^* \in \mathbb{Z}^n_+$ be an optimal solution of (SRA$(x, 1)$) with the minimum value of $\|b^* - b\|_1$.

We first show that

$$\{i \in N \mid b^*(i) - b(i) \geq 2\} = \emptyset \text{ or } \{j \in N \mid b^*(j) - b(j) \leq -2\} = \emptyset \text{ (or both).} \tag{6.5}$$

Assume, to the contrary, that there exist $i, j \in N$ with $b^*(i) - b(i) \geq 2$ and $b^*(j) - b(j) \leq -2$. Since $c(x - b', b')$ is separable-convex in b', we have

$$c(x-b, b)+c(x-b^*, b^*) \geq c(x-(b+2\chi_i-2\chi_j), b+2\chi_i-2\chi_j)+c(x-(b^*-2\chi_i+2\chi_j), b^*-2\chi_i+2\chi_j). \tag{6.6}$$

Since b is an optimal solution of (SRA$(x, 2)$), we have

$$c(x - b, b) \leq c(x - (b + 2\chi_i - 2\chi_j), b + 2\chi_i - 2\chi_j). \tag{6.7}$$

Since $\|b - (b^* - 2\chi_i + 2\chi_j)\|_1 < \|b - b^*\|_1$, it follows from the choice of b^* that

$$c(x - b^*, b^*) < c(x - (b^* - 2\chi_i + 2\chi_j), b^* - 2\chi_i + 2\chi_j),$$

which, together with (6.7), implies

$$c(x-b, b)+c(x-b^*, b^*) < c(x-(b+2\chi_i-2\chi_j), b+2\chi_i-2\chi_j)+c(x-(b^*-2\chi_i+2\chi_j), b^*-2\chi_i+2\chi_j),$$

a contradiction to (6.6).

In a similar way, we can show the following:

$$\text{if } b^*(N) - b(N) \geq 2, \text{ then, } \{i \in N \mid b^*(i) - b(i) \geq 2\} = \emptyset, \tag{6.8}$$

$$\text{if } b^*(N) - b(N) \leq -2, \text{ then, } \{j \in N \mid b^*(j) - b(j) \leq -2\} = \emptyset. \tag{6.9}$$

We now prove the inequality $\|b^* - b\|_1 < 2n$. We first consider the case where $\{i \in N \mid b^*(i) - b(i) \geq 2\} = \emptyset$ and $\{j \in N \mid b^*(j) - b(j) \leq -2\} = \emptyset$ hold. Then, we have $|b^*(i) - b(i)| \leq 1$ for every $i \in N$, implying that

$$\|b^* - b\|_1 = \sum_{i \in N} |b^*(i) - b(i)| \leq n.$$

We then consider the case where $\{i \in N \mid b^*(i) - b(i) \geq 2\} \neq \emptyset$ or $\{j \in N \mid b^*(j) - b(j) \leq -2\} \neq \emptyset$ holds; we assume the latter holds without loss of generality.

Since $\{j \in N \mid b^*(j) - b(j) \leq -2\} \neq \emptyset$, it follows from (6.5) and (6.9) that $\{i \in N \mid b^*(i) - b(i) \geq 2\} = \emptyset$ and

$$b^*(N) - b(N) \geq -1; \tag{6.10}$$

the former condition implies that $b^*(i) - b(i) \leq 1$ for every $i \in N$, from which follows that the set $N^+ = \mathrm{supp}^+(b^* - b)$ satisfies

$$b^*(N^+) - b(N^+) \leq |N^+|. \tag{6.11}$$

Hence, if $N^+ = N$ then we have

$$\|b^* - b\|_1 = \sum_{i \in N^+} (b^*(i) - b(i)) = b^*(N^+) - b(N^+) \leq |N^+| \leq n.$$

If $N^+ \neq N$ then we have

$$\begin{aligned}
\|b^* - b\|_1 &= \sum_{i \in N^+} (b^*(i) - b(i)) + \sum_{j \in N \setminus N^+} (b(j) - b^*(j)) \\
&= \Big(b^*(N^+) - b(N^+)\Big) + \Big(b(N \setminus N^+) - b^*(N \setminus N^+)\Big) \\
&\leq 2\Big(b^*(N^+) - b(N^+)\Big) + 1 \\
&\leq 2|N^+| + 1 \leq 2(n-1) + 1 < 2n,
\end{aligned}$$

where the first inequality is by (6.10) and the second by (6.11). □

Our proximity-scaling algorithm is described as follows.

Algorithm ProximityScalingDA
Step 0: Let (d_0, b_0) be an arbitrarily chosen feasible solution of (DA).
Set $x_0 := d_0 + b_0$, $\nu := 2^{\lceil \log_2((D+B)/n) \rceil}$, and $p := 1$.
Step 1: Compute an optimal solution $b'_{p-1} \in \mathbb{Z}^n$ of the problem (SRA(x_{p-1}, ν)) by using the vector b_{p-1}. Set $d'_{p-1} := x_{p-1} - b'_{p-1}$.
Step 2: Apply the algorithm SteepestDescentDA(ν) to (DA(ν)) with the initial solution (d'_{p-1}, b'_{p-1}) to find a ν-optimal solution (d_p, b_p).
Step 3: If $\nu = 1$, then output (d_p, b_p) and stop.
Otherwise, set $x_p := d_p + b_p$, $\nu := \nu/2$, $p := p + 1$, and go to Step 1.

Theorem 6.14. *The algorithm* ProximityScalingDA *finds an optimal solution of the problem* (DA) *in* $\mathrm{O}(n \log n \log((D+B)/n))$ *time.*

Proof. In the last iteration of the algorithm, the algorithm SteepestDescentDA(1) is used to obtain an output of the algorithm ProximityScalingDA. Hence, the output of ProximityScalingDA is an optimal solution of (DA) by Proposition 6.10.

We analyze the time complexity of the algorithm ProximityScalingDA. The number of iterations is $\mathrm{O}(\log((D+B)/n))$. We will show that each iteration of the algorithm can be done in $\mathrm{O}(n \log n)$ time.

The definition of the initial ν in Step 0 implies that there exists a ν-optimal solution (d, b) with

$$\|(d+b) - x_0\|_1 \leq \|d+b\|_1 + \|x_0\|_1 \leq 2(D+B) \leq 2\nu n.$$

Also, in the p-th iterations with $p \geq 2$, Theorem 6.11 implies that there exists a ν-optimal solution (d, b) with $\|(d+b) - x_{p-1}\|_1 < 16\nu n$. Hence, it follows from Proposition 6.12 that each iteration, except for Step 1, can be done in $\mathrm{O}(n \log n)$ time.

We then show that Step 1 can be done in $\mathrm{O}(n \log n)$ time. We observe that the problem (SRA(x, ν)) has the same combinatorial structure as (SRA(x)), and therefore any algorithm for the latter can be applied to the former. Hence, Proposition 6.1 can be used to derive the time bound for Step 1. In the first iteration, we have $B \leq D + B \leq \nu n$, implying that (SRA(x_0, ν)) can be solved in $\mathrm{O}(n \log n)$ time by Proposition 6.1. In the p-th iteration with $p \geq 2$, we have

$$\min\{\|b' - b_{p-1}\|_1 \mid b' \text{ is an optimal solution of } (\mathrm{SRA}(x_{p-1}, \nu))\} < 2\nu n$$

by Proposition 6.13 since b_{p-1} is an optimal solution of (SRA$(x_{p-1}, 2\nu)$). Hence, the latter statement in Proposition 6.1 implies that (SRA(x_{p-1}, ν)) can be solved in $\mathrm{O}(n \log n)$ time. □

6.3.3 Proof of Theorem 6.11

Instead of proving Theorem 6.11, we prove the following stronger statement; Theorem 6.11 corresponds to the case with $\lambda = 2$.

Theorem 6.15. *Let λ and ν be positive integers with $\lambda \geq 2$, and $(d, b) \in \mathbb{Z}^n \times \mathbb{Z}^n$ be a $\lambda\nu$-optimal solution of* (DA)*. Then, there exists some ν-optimal solution $(d^*, b^*) \in \mathbb{Z}^n \times \mathbb{Z}^n$ of* (DA) *such that*

$$\|(d^* + b^*) - (d + b)\|_1 < 8\lambda\nu n.$$

Since the structure of the problem (DA(ν)) is essentially the same as (DA), it suffices to consider the case with $\nu = 1$, i.e., it suffices to prove the following statement:

> Let λ be a positive integer with $\lambda \geq 2$, and $(d, b) \in \mathbb{Z}^n \times \mathbb{Z}^n$ be a λ-optimal solution of (DA). Then, there exists some optimal solution $(d^*, b^*) \in \mathbb{Z}^n \times \mathbb{Z}^n$ of (DA) such that
> $$\|(d^* + b^*) - (d + b)\|_1 < 8\lambda n. \tag{6.12}$$

Let (d^*, b^*) be an optimal solution of (DA) that minimizes the value $\|d^* - d\|_1 + \|b^* - b\|_1$. Also, let $x = d + b$ and $x^* = d^* + b^*$. We prove that (d^*, b^*) satisfies the inequality $\|x^* - x\|_1 < 8\lambda n$.

In the proof we consider the following six sets.

$$I_1 = \{i \in N \mid d(i) - d^*(i) \geq \lambda,\ b(i) - b^*(i) \leq -\lambda\}, \tag{6.13}$$
$$I_2 = \{j \in N \mid d(j) - d^*(j) \leq -\lambda,\ b(j) - b^*(j) \geq \lambda\}, \tag{6.14}$$
$$I_3 = \{i \in N \mid x(i) - x^*(i) \geq \lambda,\ d(i) - d^*(i) \geq \lambda\}, \tag{6.15}$$
$$I_4 = \{j \in N \mid x(j) - x^*(j) \leq -\lambda,\ d(j) - d^*(j) \leq -\lambda\}, \tag{6.16}$$
$$I_5 = \{i \in N \mid x(i) - x^*(i) \geq \lambda,\ b(i) - b^*(i) \geq \lambda\}, \tag{6.17}$$
$$I_6 = \{j \in N \mid x(j) - x^*(j) \leq -\lambda,\ b(j) - b^*(j) \leq -\lambda\}. \tag{6.18}$$

We first show that at least one of the following four cases occurs.

(Case 1) $I_4 = I_6 = \emptyset$,
(Case 2) $I_3 = I_5 = \emptyset$,
(Case 3) $I_2 = I_4 = I_5 = \emptyset$ and $b(N) - b^*(N) > -\lambda$,
(Case 4) $I_1 = I_3 = I_6 = \emptyset$ and $b(N) - b^*(N) < \lambda$.

Lemma 6.16.
(i) *At least one of $I_3 = \emptyset$ and $I_4 = \emptyset$ holds.* (ii) *At least one of $I_5 = \emptyset$ and $I_6 = \emptyset$ holds.*

Proof. We prove (i) only since (ii) can be proven similarly.

Assume, to the contrary, that both of $I_3 \neq \emptyset$ and $I_4 \neq \emptyset$ hold. Then, there exist distinct $i, j \in N$ such that

$$x(i) - x^*(i) \geq \lambda,\ d(i) - d^*(i) \geq \lambda,$$
$$x(j) - x^*(j) \leq -\lambda,\ d(j) - d^*(j) \leq -\lambda.$$

We consider the pair of vectors $(d - \lambda\chi_i + \lambda\chi_j, b)$, which is a feasible solution of (DA(λ)). As shown below, we have

$$c(d, b) > c(d - \chi_i + \chi_j, b) > c(d - 2\chi_i + 2\chi_j, b) > \cdots > c(d - \lambda\chi_i + \lambda\chi_j, b).$$

This, however, is a contradiction to the λ-optimality of (d, b).

For an integer λ' with $0 \leq \lambda' < \lambda$, put

$$d' = d - \lambda'\chi_i + \lambda'\chi_j, \qquad x' = d' + b.$$

Then, (d', b) is also a feasible solution of (DA). We will show that $c(d', b) > c(d' - \chi_i + \chi_j, b)$ holds.

Since $i \in \mathrm{supp}^+(d' - d^*) \cap \mathrm{supp}^+(x' - x^*)$ and $j \in \mathrm{supp}^-(d' - d^*) \cap \mathrm{supp}^-(x' - x^*)$, Proposition 2.4 (ii) implies that

$$\begin{aligned} c_i(d'(i), b'(i)) + c_i(d^*(i), b^*(i)) &\geq c_i(d'(i) - 1, b'(i)) + c_i(d^*(i) + 1, b^*(i)), \\ c_j(d'(j), b'(j)) + c_j(d^*(j), b^*(j)) &\geq c_j(d'(j) + 1, b'(j)) + c_j(d^*(j) - 1, b^*(j)). \end{aligned}$$

Hence, we have

$$c(d', b') + c(d^*, b^*) \geq c(d' - \chi_i + \chi_j, b') + c(d^* + \chi_i - \chi_j, b^*). \tag{6.19}$$

Note that $(d^* + \chi_i - \chi_j, b^*)$ is also a feasible solution of (DA). Since

$$\|(d^* + \chi_i - \chi_j) - d\|_1 + \|b^* - b\|_1 < \|d^* - d\|_1 + \|b^* - b\|_1,$$

we have $c(d^*, b^*) < c(d^* + \chi_i - \chi_j, b^*)$, which, together with (6.19), implies $c(d', b') > c(d' - \chi_i + \chi_j, b')$. □

Lemma 6.17.
(i) *At least one of I_1, I_4, and I_5 is an empty set.*
(ii) *If $b(N) - b^*(N) \geq \lambda$, then at least one of I_4 and I_5 is an empty set.*
(iii) *At least one of I_2, I_3, and I_6 is an empty set.*
(iv) *If $b(N) - b^*(N) \leq -\lambda$, then at least one of I_3 and I_6 is an empty set.*

Proof. We prove (i) and (ii) only since (iii) and (iv) can be proven similarly.

[Proof of (i)] Assume, to the contrary, that all of the sets I_1, I_4, and I_5 are nonempty, and let $s \in I_1$, $i \in I_4$, and $j \in I_5$. Then, elements s, i, j are distinct since sets I_1, I_4, and I_5 are mutually disjoint. We consider the pair of vectors $(d - \lambda\chi_s + \lambda\chi_i, b + \lambda\chi_s - \lambda\chi_j)$, which is a feasible solution of (DA(λ)) since

$$\begin{aligned} &(d - \lambda\chi_s + \lambda\chi_i)(N) + (b + \lambda\chi_s - \lambda\chi_j)(N) = d(N) + b(N) = D + B, \\ &(b + \lambda\chi_i - \lambda\chi_j)(N) = b(N) \leq B. \end{aligned}$$

As shown below, we have

$$\begin{aligned} c(d, b) > c(d - \chi_s + \chi_i, b + \chi_s - \chi_j) &> c(d - 2\chi_s + 2\chi_i, b + 2\chi_s - 2\chi_j) \\ &> \cdots > c(d - \lambda\chi_s + \lambda\chi_i, b + \lambda\chi_s - \lambda\chi_j). \end{aligned}$$

This, however, is a contradiction to the λ-optimality of (d, b).

For an integer λ' with $0 \leq \lambda' < \lambda$, put

$$d' = d - \lambda'\chi_s + \lambda'\chi_i, \quad b' = b + \lambda'\chi_s - \lambda'\chi_j.$$

Then, (d', b') is also a feasible solution of (DA). We will show that $c(d', b') > c(d' - \chi_s + \chi_i, b' + \chi_s - \chi_j)$ holds.

Since $i \in \mathrm{supp}^-(x'-x^*) \cap \mathrm{supp}^-(d'-d^*)$ and $j \in \mathrm{supp}^+(x'-x^*) \cap \mathrm{supp}^+(b'-b^*)$, Proposition 2.4 (ii) implies that

$$\begin{aligned} c_i(d'(i), b'(i)) + c_i(d^*(i), b^*(i)) &\geq c_i(d'(i)+1, b'(i)) + c_i(d^*(i)-1, b^*(i)), \\ c_j(d'(j), b'(j)) + c_j(d^*(j), b^*(j)) &\geq c_j(d'(j), b'(j)-1) + c_j(d^*(j), b^*(j)+1). \end{aligned}$$

Since $s \in \mathrm{supp}^+(d'-d^*) \cap \mathrm{supp}^-(b'-b^*)$, Proposition 2.4 (i) implies that

$$c_s(d'(s), b'(s)) + c_s(d^*(s), b^*(s)) \geq c_s(d'(s)-1, b'(s)+1) + c_s(d^*(s)+1, b^*(s)-1).$$

Hence, we have

$$c(d', b') + c(d^*, b^*) \geq c(d' - \chi_s + \chi_i, b' + \chi_s - \chi_j) + c(d^* + \chi_s - \chi_i, b^* - \chi_s + \chi_j). \tag{6.20}$$

Note that $(d^{**}, b^{**}) \equiv (d^* + \chi_s - \chi_i, b^* - \chi_s + \chi_j)$ is also a feasible solution of (DA) since $d^{**}(N) = d^*(N)$ and $b^{**}(N) = b^*(N)$. Since (d^{**}, b^{**}) satisfies

$$\|d^{**} - d\|_1 + \|b^{**} - b\|_1 < \|d^* - d\|_1 + \|b^* - b\|_1,$$

we have $c(d^*, b^*) < c(d^{**}, b^{**})$, which, together with (6.20), implies $c(d', b') > c(d' - \chi_s + \chi_i, b' + \chi_s - \chi_j)$.

[Proof of (ii)] Proof is similar to that for (i). Assume, to the contrary, that $b(N) - b^*(N) \geq \lambda$, $I_4 \neq \emptyset$, and $I_5 \neq \emptyset$. Let $i \in I_4$ and $j \in I_5$. Then, elements i, j are distinct since sets I_4 and I_5 are disjoint. We consider the pair of vectors $(d + \lambda\chi_i, b - \lambda\chi_j)$, which is a feasible solution of (DA(λ)) since

$$\begin{aligned} &(d + \lambda\chi_i)(N) + (b - \lambda\chi_j)(N) = d(N) + b(N) = D + B, \\ &(b - \lambda\chi_j)(N) = b(N) - \lambda \leq B. \end{aligned}$$

As shown below, we have

$$c(d, b) > c(d + \chi_i, b - \chi_j) > c(d + 2\chi_i, b - 2\chi_j) > \cdots > c(d + \lambda\chi_i, b - \lambda\chi_j).$$

This, however, is a contradiction to the λ-optimality of (d, b).

For an integer λ' with $0 \leq \lambda' < \lambda$, put

$$d' = d + \lambda'\chi_i, \quad b' = b - \lambda'\chi_j.$$

Then, (d', b') is also a feasible solution of (DA). We will show that $c(d', b') > c(d' + \chi_i, b' - \chi_j)$ holds.

Since $i \in \mathrm{supp}^-(x'-x^*) \cap \mathrm{supp}^-(d'-d^*)$ and $j \in \mathrm{supp}^+(x'-x^*) \cap \mathrm{supp}^+(b'-b^*)$, Proposition 2.4 (ii) implies that

$$\begin{aligned} c_i(d'(i), b'(i)) + c_i(d^*(i), b^*(i)) &\geq c_i(d'(i)+1, b'(i)) + c_i(d^*(i)-1, b^*(i)), \\ c_j(d'(j), b'(j)) + c_j(d^*(j), b^*(j)) &\geq c_j(d'(j), b'(j)-1) + c_j(d^*(j), b^*(j)+1). \end{aligned}$$

Hence, we have

$$c(d', b') + c(d^*, b^*) \geq c(d' + \chi_i, b' - \chi_j) + c(d^* - \chi_i, b^* + \chi_j). \tag{6.21}$$

Note that $(d^{**}, b^{**}) \equiv (d^* - \chi_i, b^* + \chi_j)$ is also a feasible solution of (DA) since $d^{**}(N) + b^{**}(N) = d^*(N) + b^*(N)$ and $b^{**}(N) = b^*(N) + 1 \leq b(N) \leq B$. Since (d^{**}, b^{**}) satisfies

$$\|d^{**} - d\|_1 + \|b^{**} - b\|_1 < \|d^* - d\|_1 + \|b^* - b\|_1,$$

we have $c(d^*, b^*) < c(d^{**}, b^{**})$, which, together with (6.21), implies $c(d', b') > c(d' + \chi_i, b' - \chi_j)$. $\square$

Lemma 6.18. *At least one of Cases 1, 2, 3, and 4 occurs.*

Proof. Suppose that neither of Cases 1 and 2 occurs. By Lemma 6.16, we have either

$$I_4 = I_5 = \emptyset,\ I_3 \neq \emptyset,\ I_6 \neq \emptyset, \qquad \text{or} \qquad I_3 = I_6 = \emptyset,\ I_4 \neq \emptyset,\ I_5 \neq \emptyset.$$

In the former case, we have $I_2 = \emptyset$ by Lemma 6.16 (iii) and $b(N) - b^*(N) > -\lambda$ by Lemma 6.16 (iv), i.e., Case 3 occurs. Similarly, the latter case implies Case 4. □

We prove the inequality (6.12) in Cases 1 and 2.

Lemma 6.19. *We have $\|x - x^*\|_1 < 4\lambda n$ if Case 1 or 2 occurs.*

Proof. We consider Case 1 only; the proof for Case 2 is similar. Suppose that $I_4 = I_6 = \emptyset$ holds. Then, we have $x(i) - x^*(i) > -2\lambda$ for every $i \in N$; indeed, if $x(i) - x^*(i) \leq -\lambda$, then $I_4 = I_6 = \emptyset$ implies $x(i) - x^*(i) > -2\lambda$.

Let $N^- = \mathrm{supp}^-(x - x^*)$. Then, we have

$$x^*(N^-) - x(N^-) = \sum_{j \in N^-} (x^*(i) - x(i)) < 2\lambda |N^-|,$$

which, together with $x(N) = D + B = x^*(N)$, implies that

$$\begin{aligned} \|x - x^*\|_1 &= \sum_{i \in N \setminus N^-} (x(i) - x^*(i)) + \sum_{j \in N^-} (x^*(i) - x(i)) \\ &= \Big(x(N \setminus N^-) - x^*(N \setminus N^-)\Big) + \Big(x^*(N^-) - x(N^-)\Big) \\ &= 2(x^*(N^-) - x(N^-)) < 4\lambda |N^-| \leq 4\lambda n. \end{aligned}$$

□

We then prove the inequality (6.12) in Cases 3 and 4.

Lemma 6.20. *We have $\|x - x^*\|_1 < 8\lambda n$ if Case 3 or 4 occurs.*

Proof. We consider Case 3 only since the proof in Case 4 is similar. Since

$$\|x - x^*\|_1 \leq \|d - d^*\|_1 + \|b - b^*\|_1,$$

it suffices to show that $\|d - d^*\|_1 < 4\lambda n$ and $\|b - b^*\|_1 < 4\lambda n$ hold. Below we prove $\|d - d^*\|_1 < 4\lambda n$ only since the inequality $\|b - b^*\|_1 < 4\lambda n$ can be proven in a similar way.

Since $I_2 = I_4 = \emptyset$, it holds that

$$d(i) - d^*(i) > -2\lambda \qquad (i \in N). \tag{6.22}$$

Indeed, for $i \in N$, if $d(i) - d^*(i) > -\lambda$, then we are done; if $d(i) - d^*(i) \leq -\lambda$, then $I_2 = I_4 = \emptyset$ implies that

$$b(i) - b^*(i) < \lambda, \qquad x(i) - x^*(i) > -\lambda,$$

from which follows that

$$d(i) - d^*(i) = (x(i) - x^*(i)) - (b(i) - b^*(i)) > -2\lambda.$$

Since $b(N) - b^*(N) > -\lambda$ and $x(N) - x^*(N) = 0$, we also have

$$d(N) - d^*(N) < \lambda. \tag{6.23}$$

To prove the inequality $\|d - d^*\|_1 < 4\lambda n$, let $N^- = \mathrm{supp}^-(d - d^*)$. By (6.22), it holds that

$$d^*(N^-) - d(N^-) < 2\lambda|N^-|. \tag{6.24}$$

Hence, if $N^- = N$ then we have

$$\|d - d^*\|_1 = \sum_{j \in N}(d^*(j) - d(j)) = d^*(N^-) - d(N^-) < 2\lambda|N^-| \leq 2\lambda n.$$

If $N^- \neq N$ then we have

$$\begin{aligned}
\|d - d^*\|_1 &= \sum_{i \in N \setminus N^-}(d(i) - d^*(i)) + \sum_{j \in N^-}(d^*(j) - d(j)) \\
&= \Big(d(N \setminus N^-) - d^*(N \setminus N^-)\Big) + \Big(d^*(N^-) - d(N^-)\Big) \\
&= (d(N) - d^*(N)) + 2(d^*(N^-) - d(N^-)) \\
&< \lambda + 4\lambda|N^-| \leq \lambda + 4\lambda(n-1) \leq 4\lambda n,
\end{aligned}$$

where the first inequality is by (6.23) and (6.24). □

References

[1] D. Freund, S.G. Henderson, and D.B. Shmoys. Minimizing multimodular functions and allocating capacity in bike-sharing systems. In *Proc. 19th IPCO*, LNCS 10328, pages 186–198. Springer, Berlin, 2017.

[2] D. Freund, S.G. Henderson, and D.B. Shmoys, Minimizing multimodular functions and allocating capacity in bike-sharing systems. preprint, arXiv:1611.09304v3, October 2018.

[3] D. S. Hochbaum, Lower and upper bounds for the allocation problem and other nonlinear optimization problems. *Math. Oper. Res.*, 19:390–409, 1994.

[4] T. Ibaraki and N. Katoh: *Resource Allocation Problems: Algorithmic Approaches.* MIT Press, Cambridge, MA, 1988.

[5] S. Iwata, S. Moriguchi, and K. Murota. A capacity scaling algorithm for M-convex submodular flow. *Math. Program.*, 103:181-202, 2005.

[6] S. Iwata and M. Shigeno. Conjugate scaling algorithm for Fenchel-type duality in discrete convex optimization. *SIAM J. Optim.*, 13:204–211, 2003.

[7] S. Moriguchi and K. Murota. On fundamental operations for multimodular functions. preprint, arXiv:1805.04245, 2018.

[8] S. Moriguchi, K. Murota, and A. Shioura. Scaling algorithms for M-convex function minimization. *IEICE Transactions on Fundamentals*, E85-A: 922–929, 2002.

[9] K. Murota. Convexity and Steinitz's exchange property. *Adv. Math.*, 124:272–311, 1996.

[10] K. Murota. Discrete convex analysis. *Math. Program.*, 83:313–371, 1998.

[11] K. Murota. *Discrete Convex Analysis.* SIAM, Philadelphia, 2003.

[12] K. Murota. On steepest descent algorithms for discrete convex functions. *SIAM J. Optim.*, 14:699–707, 2003.

[13] K. Murota. Note on multimodularity and L-convexity. *Math. Oper. Res.*, 30:658-661, 2005.

[14] K. Murota and A. Shioura. M-convex function on generalized polymatroid. Math. Oper. Res., 24:95–105, 1999.

[15] A. Shioura. Minimization of an M-convex function. *Discrete Appl. Math.*, 84:215–220, 1998.

[16] A. Shioura. Fast scaling algorithms for M-convex function minimization with application to the resource allocation problem. *Discrete Appl. Math.*, 134:303–316, 2004.

[17] A. Shioura. M-convex function minimization under L1-distance constraint. preprint, arXiv:1809.03126v3, September 2018.

[18] A. Tamura. Coordinatewise domain scaling algorithm for M-convex function minimization. *Math. Program.*, 102:339–354, 2005.

A Appendix

A.1 Proof of Lemma 6.2

We denote $\hat{x} = x + \chi_i - \chi_j$, and let $\hat{b} \in \mathbb{Z}^n$ be an optimal solution of the problem (SRA($\hat{x}$)) that minimizes the value $\|\hat{b} - b\|_1$. We show that the vector $\hat{b}$ satisfies the condition (6.2).

Claim 1: (i) Suppose that $s \in N$ satisfies the following condition:

$$\text{either} \quad s \in \mathrm{supp}^+(\hat{b} - b) \setminus \{i\} \quad \text{or} \quad s = i \text{ and } \hat{b}(i) - b(i) \geq 2. \tag{A.1}$$

Then,

$$\begin{aligned} &c_s(\hat{x}(s) - \hat{b}(s), \hat{b}(s)) + c_s(x(s) - b(s), b(s)) \\ &\geq c_s(\hat{x}(s) - \hat{b}(s) + 1, \hat{b}(s) - 1) + c_s(x(s) - b(s) - 1, b(s) + 1). \end{aligned} \tag{A.2}$$

(ii) Suppose that $t \in N$ satisfies the following condition:

$$\text{either} \quad t \in \mathrm{supp}^-(\hat{b} - b) \setminus \{j\} \quad \text{or} \quad t = j \text{ and } \hat{b}(j) - b(j) \leq -2. \tag{A.3}$$

Then,

$$\begin{aligned} &c_t(\hat{x}(t) - \hat{b}(t), \hat{b}(t)) + c_t(x(t) - b(t), b(t)) \\ &\geq c_t(\hat{x}(t) - \hat{b}(t) - 1, \hat{b}(t) + 1) + c_t(x(t) - b(t) + 1, b(t) - 1). \end{aligned} \tag{A.4}$$

[Proof of Claim] We prove (i) only since (ii) can be shown similarly. If $s \in \mathrm{supp}^+(\hat{b} - b) \setminus \{i\}$ then we have $\hat{b}(s) > b(s)$ and $\hat{x}(s) - \hat{b}(s) < x(s) - b(s)$. If $s = i$ and $\hat{b}(i) - b(i) \geq 2$, then we have $\hat{b}(s) > b(s)$ and $\hat{x}(i) - \hat{b}(i) \leq (x(i) + 1) - (b(i) + 2) < x(i) - b(i)$. In either case, (A.2) follows from Proposition 2.4 (i). [End of Claim]

To prove the lemma, we consider the following two conditions:

(a) $\mathrm{supp}^+(\hat{b} - b) \subseteq \{i\}$ and $\hat{b}(i) - b(i) \leq 1$,

(b) $\mathrm{supp}^-(\hat{b} - b) \subseteq \{j\}$ and $\hat{b}(j) - b(j) \geq -1$.

Note that the condition (a) holds if and only if there exists no $s \in N$ satisfying (A.1); similarly, (b) holds if and only if there exists no $t \in N$ satisfying (A.3).

Claim 2: At least one of the conditions (a) and (b) holds. Moreover, the condition (a) holds if $b(N) < B$, and the condition (b) holds if $\hat{b} < B$.
[Proof of Claim] To prove the former statement, assume, to the contrary, that there exist some $s, t \in N$ satisfying the conditions (A.1) and (A.3). Then, it holds that

$$\begin{aligned}
&c(\hat{x} - \hat{b}, \hat{b}) + c(x - b, b) \\
&\quad - c(\hat{x} - (\hat{b} - \chi_s + \chi_t), \hat{b} - \chi_s + \chi_t) - c(x - (b + \chi_s - \chi_t), b + \chi_s - \chi_t) \\
&= \Big[c_s(\hat{x}(s) - \hat{b}(s), \hat{b}(s)) + c_s(x(s) - b(s), b(s)) \\
&\qquad - c_s(\hat{x}(s) - \hat{b}(s) + 1, \hat{b}(s) - 1) - c_s(x(s) - b(s) - 1, b(s) + 1) \Big] \\
&\quad + \Big[c_t(\hat{x}(t) - \hat{b}(t), \hat{b}(t)) + c_t(x(t) - b(t), b(t)) \\
&\qquad - c_t(\hat{x}(t) - \hat{b}(t) - 1, \hat{b}(t) + 1) - c_t(x(t) - b(t) + 1, b(t) - 1) \Big] \\
&\geq 0,
\end{aligned} \tag{A.5}$$

where the inequality is by (A.2) and (A.4) in Claim 1. Note that $b + \chi_s - \chi_t$ is a feasible solution of (SRA(x)) since $(b + \chi_s - \chi_t)(N) = b(N) \leq B$, $b(s) < \hat{b}(s) \leq x(s)$, and $b(t) > \hat{b}(t) \geq 0$. Therefore, we have

$$c(x - b, b) \leq c(x - (b + \chi_s - \chi_t), b + \chi_s - \chi_t),$$

which, together with (A.5), implies

$$c(\hat{x} - \hat{b}, \hat{b}) \geq c(\hat{x} - (\hat{b} - \chi_s + \chi_t), \hat{b} - \chi_s + \chi_t).$$

Since $\hat{b}$ is an optimal solution of (SRA($x + \chi_i - \chi_j$)) and $\hat{b} - \chi_s + \chi_t$ is a feasible solution of (SRA($x + \chi_i - \chi_j$)), the vector $\hat{b} - \chi_s + \chi_t$ is also an optimal solution of (SRA($x + \chi_i - \chi_j$)), a contradiction to the choice of $\hat{b}$ since $\|(\hat{b} - \chi_s + \chi_t) - b\|_1 = \|\hat{b} - b\|_1 - 2$. This concludes the proof of the former statement.

To prove the latter statement, we assume $b(N) < B$; proof for the case $\hat{b}(N) < B$ is similar and omitted. Assume, to the contrary, that there exists some $s \in N$ satisfying the condition (A.1). Then, we have

$$\begin{aligned}
&c(\hat{x} - \hat{b}, \hat{b}) + c(x - b, b) - c(\hat{x} - (\hat{b} - \chi_s), \hat{b} - \chi_s) - c(x - (b + \chi_s), b + \chi_s) \\
&= c_s(\hat{x}(s) - \hat{b}(s), \hat{b}(s)) + c_s(x(s) - b(s), b(s)) \\
&\qquad - c_s(\hat{x}(s) - \hat{b}(s) + 1, \hat{b}(s) - 1) - c_s(x(s) - b(s) - 1, b(s) + 1) \\
&\geq 0,
\end{aligned} \tag{A.6}$$

where the inequality is by (A.2) in Claim 1. The vector $b + \chi_s$ is a feasible solution of (SRA(x)) since $b(N) < B$ and $b(s) < \hat{b}(s) \leq x(s)$. Hence, we have

$$c(x - b, b) \leq c(x - (b + \chi_s), b + \chi_s),$$

which, together with (A.6), implies

$$c(\hat{x} - \hat{b}, \hat{b}) \geq c(\hat{x} - (\hat{b} - \chi_s), \hat{b} - \chi_s). \tag{A.7}$$

Since $\hat{b} - \chi_s$ is a feasible solution of (SRA($x + \chi_i - \chi_j$)), optimality of $\hat{b}$ and the inequality (A.7) imply that $\hat{b} - \chi_s$ is also an optimal solution of (SRA($x + \chi_i - \chi_j$)), a contradiction to the choice of $\hat{b}$ since $\|(\hat{b} - \chi_s) - b\|_1 = \|\hat{b} - b\|_1 - 1$. Hence, the condition (a) holds. [End of Claim]

We now prove the lemma. It is easy to see from Claim 2 that the following properties hold:

- if $b(N) = \hat{b}(N) < B$, then $\hat{b} \in \{b, b + \chi_i - \chi_j\}$,
- if $b(N) < \hat{b}(N) \leq B$, then $\hat{b} = b + \chi_i$,
- if $B \geq b(N) > \hat{b}(N)$, then $\hat{b} = b - \chi_j$.

In either case, the condition (6.2) holds.

To conclude the proof of the lemma, consider the remaining case with $b(N) = \hat{b}(N) = B$. Then, at least one of (a) and (b) holds by Claim 2. Suppose that (a) holds. If $\mathrm{supp}^+(\hat{b} - b) = \emptyset$, then $\hat{b} = b$ follows from $b(N) = \hat{b}(N)$. If $\mathrm{supp}^+(\hat{b} - b) \neq \emptyset$, then we have $\mathrm{supp}^+(\hat{b} - b) = \{i\}$ and $\hat{b}(i) = b(i) + 1$. Since $b(N) = \hat{b}(N)$, there exists a unique element t in $\mathrm{supp}^-(\hat{b} - b)$, which satisfies $t \neq i$ and $\hat{b}(t) = b(t) - 1$. Hence, we have $\hat{b} = b$ or $\hat{b} = b + \chi_i - \chi_t$ for some $t \in N \setminus \{i\}$. If the condition (b) holds, then we can show in a similar way that $\hat{b} = b$ or $\hat{b} = b + \chi_s - \chi_j$ for some $s \in N \setminus \{j\}$. Hence, the condition (6.2) holds.