A short convex-hull proof for the all-different system with the inclusion property

Marco Di Summa

Abstract

An all-different constraint on some discrete variables imposes the condition that no two variables take the same value. A linear-inequality description of the convex hull of solutions to a system of all-different constraints is known under the so-called *inclusion property*: the convex hull is the intersection of the convex hulls of each of the all-different constraints of the system. We give a short proof of this result, which in addition shows the total dual integrality of the linear system.

Keywords: all-different constraint, convex hull, integral polyhedron, total dual integrality
2000 MSC: 90C10, 90C27

1. Introduction

In many combinatorial optimization problems one needs to impose one or more *all-different constraints*, i.e., conditions of the following type: for a given finite (sub)family of discrete variables, no two variables can be assigned the same value. All-different constraints arise, for instance, in problems related to timetabling, scheduling, manufacturing, and in several variants of the assignment problem (see, e.g., [6, 9] and the references therein).

Though all-different constraints are mainly studied in the context of Constraint Programming (see, e.g., [8]), when dealing with a problem that can be modeled as an integer linear program it is useful to have information on the polyhedral structure of the feasible solutions to a system of all-different constraints. For this reason, several authors studied linear-inequality formulations for the convex hull of solutions to a single all-different constraint or a system of all-different constraints [3, 5, 6, 9]. We remark that in some cases these descriptions are extended formulations, i.e., they make use of additional variables; however, here we are only interested in the description of the convex hull in the original space of variables.

If n variables $x_1, \dots, x_n$ can take values in a finite domain $D \subseteq \mathbb{R}$ and an all-different constraint is imposed on them, we will write (following the notation in [6])

$$\{x_1, \dots, x_n\}_{\neq} \tag{1}$$

$$x_1, \dots, x_n \in D. \tag{2}$$

Williams and Yan [9] proved that if $D = \{1, \dots, d\}$ for some positive integer d, then the convex hull of the vectors that satisfy (1)–(2) is described by the linear system

$$\sum_{j \in S} x_j \geq f(S), \quad S \subseteq [n], \tag{3}$$

$$\sum_{j \in S} x_j \leq g(S), \quad S \subseteq [n], \tag{4}$$

where $[n] = \{1, \dots, n\}$ and, for $S \subseteq [n]$,

$$f(S) = \frac{|S|(|S|+1)}{2}, \quad g(S) = |S|(d+1) - f(S). \tag{5}$$

Note that $f(S)$ is the sum of the $|S|$ smallest positive integers, while $g(S)$ is the sum of the $|S|$ largest integers that do not exceed d, therefore inequalities (3)–(4) are certainly valid for every vector x satisfying (1)–(2). This result extends to an arbitrary finite domain $D \subseteq \mathbb{R}$ (with $|D| \geq n$) by defining $f(S)$ (resp., $g(S)$) as the sum of the $|S|$ smallest (resp., largest) elements in D, for every $S \subseteq [n]$. Note however that in the following we assume $D = \{1, \dots, d\}$ for some positive integer d, while we will consider the case of a generic finite domain D in a final remark.

Williams and Yan [9] showed that if $d > n$ then all inequalities (3)–(4) are facet-defining, thus the convex hull of (1)–(2) needs an exponential number of inequalities to be described in the original space of variables $x_1, \dots, x_n$. However, they also gave polynomial-size extended formulations for the convex hull of (1)–(2).

When $d = n$, (1)–(2) is the set of permutations of the elements in $[n]$, and its convex hull is called permutahedron. In this case, the whole family of inequalities (4) can be dropped and replaced by the equation $\sum_{j \in [n]} x_j = f([n])$. The permutahedron admits an extended formulation with $O(n \log n)$ constraints and variables [2].

System (3)–(4) not only defines an integral polyhedron, but it also has the stronger property of being *totally dual integral*. We recall that a linear system of inequalities $Ax \leq b$ is said to be totally dual integral if for every integer vector c such that the linear program $\max\{cx : Ax \leq b\}$ has finite optimum, the dual linear program has an optimal solution with integer components. It is known that if $Ax \leq b$ is totally dual integral and b is an integer vector, then the polyhedron defined by $Ax \leq b$ is integral (see, e.g., [7, Theorem 5.22]). The total dual integrality of system (3)–(4) follows immediately from the fact that f

(resp., g) is a supermodular (resp., submodular) function, along with a classical result on polymatroid intersection [1] (see also [7, Theorem 46.2]). An explicit proof of the total dual integrality of (3)–(4) is given in [5].

In a more general setting, we might have $m \geq 1$ all-different constraints, each enforced on a different subset of variables $N_i \subseteq [n], i \in [m]$. In this case, we have the system of conditions

$$\{x_j : j \in N_i\}_{\neq}, \quad i \in [m], \tag{6}$$
$$x_1, \ldots, x_n \in D. \tag{7}$$

The following inequalities are of course valid for the convex hull of solutions to (6)–(7):

$$\sum_{j \in S} x_j \geq f(S), \quad S \subseteq N_i,\, i \in [m], \tag{8}$$
$$\sum_{j \in S} x_j \leq g(S), \quad S \subseteq N_i,\, i \in [m]. \tag{9}$$

However, the above constraints do not give, in general, the convex hull of the vectors that satisfy (6)–(7). Furthermore, there are even examples in which some integer solutions to system (8)–(9) do not lie in the convex hull of the points satisfying (6)–(7). Therefore it is natural to ask which conditions ensure that the above system provides the convex hull of the vectors satisfying (6)–(7).

A special case, studied in [6], in which constraints (8)–(9) do yield the convex hull of solutions to (6)–(7) is now described. Define $N = [n]$ and assume that $N = T \cup U$, where T and U are disjoint nonempty subsets of N. Define $T_i = N_i \cap T$ and $U_i = N_i \cap U$ for $i \in [m]$. If the T_i's form a monotone family of subsets ($T_1 \supseteq T_2 \supseteq \cdots \supseteq T_m$) and the U_i's are pairwise disjoint, then Magos et al. [6] say that the *inclusion property* holds. They showed that in this case inequalities (8)–(9) provide the convex hull of solutions to (6)–(7). We remark that (to the best of the author's knowledge) this is the only nontrivial case in which formulation (8)–(9) is known to define the convex hull of the all-different system.

The proof of Magos et al. [6] is rather lengthy and involved (overall, it consists of about 25 pages). The purpose of this note is to give a shorter proof of their result. Indeed, we show something more: we prove that, under the inclusion property, system (8)–(9) is totally dual integral. Our proof is an extension of the classical approach to show the total dual integrality of polymatroids (see, e.g., [7, Chapter 44]). Specifically, in Section 2 we describe a greedy algorithm that solves linear optimization over (8)–(9), under the inclusion property. The correctness of the algorithm is shown in Section 3 by completing the feasible solution returned by the algorithm with a dual solution such that the complementary slackness conditions are satisfied. The result of Section 3 also implies the total dual integrality of system (8)–(9), as the dual solution is integer whenever the primal objective function coefficients are all integers. We conclude in Section 4 with an extension of the result, which in particular can be used to deal with a generic finite domain D.

2. Primal algorithm

Assume that the inclusion property holds for an all-different system (6)–(7). Recall that:

- $N = [n] = T \cup U$, with T and U disjoint and nonempty;
- $T_i = N_i \cap T$ and $U_i = N_i \cap U$ for $i \in [m]$;
- $T_1 \supseteq \cdots \supseteq T_m$;
- $U_i \cap U_j = \varnothing$ for all distinct $i, j \in [m]$.

Wlog, $N = N_1 \cup \cdots \cup N_m$ and $T = T_1 = [t]$ for some positive integer t. Also, recall that $D = [d]$. We assume that $d \geq \max_{i \in [m]} |N_i|$, otherwise both (6)–(7) and (8)–(9) are infeasible. We use the notation $t_i = |T_i|$ and $u_i = |U_i|$ for $i \in [m]$. Furthermore, we will sometimes identify an index $j \in N$ with the corresponding variable x_j; e.g., we will indifferently say "the indices in T" or "the variables in T".

Consider the problem of minimizing a linear objective function cx over the polytope defined by (8)–(9), where c is a row-vector in $\mathbb{R}^n$. If we define $\mathcal{S} = \bigcup_{i \in [m]} \{S : S \subseteq N_i\}$, the problem can be written as follows:

$$\begin{aligned} \min \quad & cx & \quad (10)\\ \text{s.t.} \quad & \sum_{j \in S} x_j \geq f(S), \qquad S \in \mathcal{S}, & \quad (11)\\ & -\sum_{j \in S} x_j \geq -g(S), \quad S \in \mathcal{S}. & \quad (12) \end{aligned}$$

We give a greedy algorithm that solves the above linear program for an arbitrary $c \in \mathbb{R}^n$. Since the polyhedron defined by (11)–(12) contains all vectors satisfying (6)–(7), and since we will show that the solution returned by the algorithm satisfies (6)–(7), this will prove that system (11)–(12) (i.e., system (8)–(9)) defines the convex hull of (6)–(7). The algorithm that we present can be seen as an extension of the greedy algorithm for polymatroids (see, e.g., [7, Chapter 44]), and also as an extension of the algorithm given in [5] for the case $m = 1$.

The procedure is shown in Algorithm 1 and is now illustrated. Throughout the algorithm, we maintain d *clusters* of variables $V_1, \ldots, V_d$, i.e., d (possibly empty) disjoint subsets of N gathering those variables that will be assigned the same value at the end of the algorithm. At the beginning (lines 2–3) we have t nonempty clusters $V_1, \ldots, V_t$, where $V_j = \{j\}$ for $j \in [t]$, while the other clusters $V_{t+1}, \ldots, V_d$ are empty. Thus every variable in $T = [t]$ is assigned to a different cluster (as these variables are not allowed to take the same value because of the first all-different constraint), while the variables in U are not assigned to any cluster. During the execution of the algorithm, each variable in U will be assigned to a cluster, and no variable will be ever moved from one cluster to another.

Notation $r(j)$ indicates the index of the cluster to which variable x_j is assigned. With each cluster $V_j, j \in [d]$, we associate a *pseudo-cost* γ_j, which is the sum of the costs of all variables in the cluster. The pseudo-cost of an empty cluster is zero.

For $i = 1, \dots, m$, at the ith iteration of the algorithm we assign each variable in U_i to a different cluster (lines 4–11), as we now explain. Because of the ith all-different constraint, a variable in U_i cannot be assigned to a cluster containing a variable in T_i. Note that for $j \in T_i$, the cluster containing j is V_j. Therefore only the clusters V_j with $j \in [d] \setminus T_i$ are *feasible* for the variables in U_i. Lines 5–6 order the feasible clusters (including the empty ones) and the variables in U_i according to their pseudo-costs and costs, respectively (with ties broken arbitrarily). This is needed to assign the variables in U_i to the feasible clusters in a greedy fashion (lines 8–9): among the variables with nonnegative cost, the one with the highest cost is assigned to the feasible cluster with the highest pseudo-cost (independently of the sign of the pseudo-cost), then the variable with the second highest cost is assigned to the feasible cluster with the second highest pseudo-cost, and so on; on the other hand, among the variables with negative cost, the one with the smallest cost is assigned to the feasible cluster with the smallest pseudo-cost, then the variable with the second smallest cost is assigned to the feasible cluster with the second smallest pseudo-cost, and so on. Lines 10 and 11 consequently update the clusters and the pseudo-costs.

At the end of the above procedure, we simply assign value 1 to the variables in the cluster with the highest pseudo-cost, value 2 to those in the cluster with second highest pseudo-cost, and so forth (lines 12–13).

Algorithm 1: Greedy algorithm for linear optimization over an all-different system with the inclusion property.

1 **begin**
2   **for each** $j \in [t]$ **do** $V_j := \{j\}$, $\gamma_j := c_j$, $r(j) := j$;
3   **for each** $j \in [d] \setminus [t]$ **do** $V_j := \varnothing$, $\gamma_j := 0$;
4   **for** $i = 1, \dots, m$ **do**
5   define a bijection $\sigma : [d - t_i] \to [d] \setminus T_i$ such that $\gamma_{\sigma(1)} \geq \cdots \geq \gamma_{\sigma(d-t_i)}$;
6   define a bijection $\pi : [u_i] \to U_i$ such that $c_{\pi(1)} \geq \cdots \geq c_{\pi(u_i)}$;
7   **for each** $j \in [u_i]$ **do**
8   **if** $c_{\pi(j)} \geq 0$ **then** $r(\pi(j)) := \sigma(j)$;
9   **else** $r(\pi(j)) := \sigma((d - t_i) - (u_i - j))$;
10       $V_{r(\pi(j))} := V_{r(\pi(j))} \cup \{\pi(j)\}$;
11       $\gamma_{r(\pi(j))} := \gamma_{r(\pi(j))} + c_{\pi(j)}$;
12   define a bijection $\sigma : [d] \to [d]$ such that $\gamma_{\sigma(1)} \geq \cdots \geq \gamma_{\sigma(d)}$;
13   **for each** $j \in N$ **do** $\bar{x}_j := \sigma^{-1}(r(j))$;
14   **return** $\bar{x}$

Note that if two variables belong to the same set N_i for some $i \in [m]$, then they are assigned to different clusters, and therefore they receive different values. This implies that the solution returned by the algorithm satisfies the given all-different system (6)–(7), and thus also (11)–(12). The optimality of the solution will follow from the existence of a dual solution satisfying the complementary slackness conditions, as proven in the next section.

3. Dual solution and total dual integrality

In this section we prove the correctness of Algorithm 1 and, at the same time, we show that inequalities (8)–(9) define the convex hull of the vectors satisfying (6)–(7), under the inclusion property. Before proceeding, it is useful to note that if, for a given instance of the problem, Algorithm 1 is run with different tie-breaking choices at lines 5, 6, and 12, then the solutions returned have the same cost. To see this, define at every stage of the algorithm the *pseudo-cost pattern* as the d-dimensional vector containing the pseudo-costs of all clusters (with repetition and including empty clusters), listed in non-increasing order. At the end of every iteration of loop 4–11, the pseudo-cost pattern does not depend on the tie-breaking rule used at lines 5 and 6. Also note that at every iteration the elements contained in a specific cluster are not relevant: what matters is only the pseudo-cost pattern. Finally, given the pseudo-cost pattern, lines 12–13 produce a solution whose cost does not depend on the tie-breaking rule.

Therefore, for every instance of the linear program (10)–(12), it is sufficient to prove the correctness of Algorithm 1 under a specific tie breaking rule. Furthermore, as observed above, the final pseudo-cost pattern does not depend on the rule chosen.

Theorem 1. *Under the inclusion property, inequalities* (8)–(9) *define the convex hull of the vectors satisfying* (6)–(7).

Proof. It is sufficient to show that every instance of the linear program (10)–(12) has an optimal solution that satisfies (6)–(7). For this purpose, given an instance of (10)–(12), let p be the number of distinct nonzero costs of the variables in U_m, and q be the number of distinct pseudo-costs at the end of Algorithm 1:

$$p = |\{c_j : c_j \neq 0,\, j \in U_m\}|, \qquad q = |\{\gamma_j : j \in [d]\}|.$$

(As observed above, the value of q does not depend on the tie-breaking rule.) Assume by contradiction that for some instances the algorithm returns a solution that is not optimal. Among all instances for which the algorithm does not return an optimal solution, we choose an instance I such that the vector $(m, p + q)$ is lexicographically minimum. (Note that $m \geq 2$, as for $m = 1$ Algorithm 1 reduces to the algorithm given in [5] and thus returns an optimal solution.) From now on, I will be a minimal instance according to the above definition, and $\bar{x}$ will be the solution returned by Algorithm 1 (with any fixed tie-breaking rule). Recall that $\bar{x}$ is a feasible solution to (10)–(12).

Consider the dual problem of (10)–(12):

$$\begin{aligned} \max \ & \sum_{S \in \mathcal{S}} \big(f(S) y_S - g(S) z_S\big) \\ \text{s.t.} \ & \sum_{S \in \mathcal{S}: j \in S} (y_S - z_S) = c_j, \qquad j \in N, \\ & y_S, z_S \geq 0, \qquad S \in \mathcal{S}. \end{aligned}$$

Since $\bar{x}$ is feasible but not optimal for I, there is no dual feasible solution $(\bar{y}, \bar{z})$ such that $\bar{x}$ and $(\bar{y}, \bar{z})$ satisfy the complementary slackness conditions, which read as follows:

(a′) for every $S \in \mathcal{S}$, if $\bar{y}_S > 0$ then $\sum_{j \in S} \bar{x}_j = f(S)$;

(b′) for every $S \in \mathcal{S}$, if $\bar{z}_S > 0$ then $\sum_{j \in S} \bar{x}_j = g(S)$.

Note that since $\bar{x}$ satisfies (6)–(7), $\sum_{j \in S} \bar{x}_j = f(S)$ if and only if $\{\bar{x}_j : j \in S\} = \{1, \dots, |S|\}$, and $\sum_{j \in S} \bar{x}_j = g(S)$ if and only if $\{\bar{x}_j : j \in S\} = \{d - |S| + 1, \dots, d\}$. Then we can rewrite conditions (a′) and (b′) as follows:

(a) for every $S \in \mathcal{S}$, if $\bar{y}_S > 0$ then $\{\bar{x}_j : j \in S\} = \{1, \dots, |S|\}$;

(b) for every $S \in \mathcal{S}$, if $\bar{z}_S > 0$ then $\{\bar{x}_j : j \in S\} = \{d - |S| + 1, \dots, d\}$.

Therefore there is no dual feasible solution $(\bar{y}, \bar{z})$ such that (a) and (b) are fulfilled.

Case 1. Suppose that $c_j = 0$ for all $j \in U_m$. If we remove the mth all-different constraint and variables x_j for $j \in U_m$, we obtain a new instance I' with $m-1$ constraints (note that $m - 1 \geq 1$). Clearly, I' has the inclusion property. If we apply Algorithm 1 to I' (by making the same tie-breaking choices as we did for I), we execute exactly the same operations as in the first $m-1$ iterations of the algorithm applied to instance I. Then, since $c_j = 0$ for all $j \in U_m$, the final pseudo-costs are the same for I and I'. Thus we obtain a solution $\bar{x}'$ for I' which is identical to $\bar{x}$, except that $\bar{x}'$ does not have the entries with index $j \in U_m$. By the minimality of I, instance I' admits a dual solution $(\bar{y}, \bar{z})$ that satisfies (a) and (b). If this dual solution is extended by setting $\bar{y}_S = \bar{z}_S = 0$ for every $S \subseteq N_m$ such that $S \cap U_m \neq \varnothing$, we obtain a dual feasible solution for I, and conditions (a) and (b) are still satisfied. This is a contradiction.

Case 2. Suppose that $c_j > 0$ for some $j \in U_m$. Define $c^* = \max\{c_j : j \in U_m\} > 0$ and $C = \{j \in U_m : c_j = c^*\}$. Recall that, for $j \in N$, $r(j)$ denotes the index of the cluster containing j. We extend this notation to subsets by defining $r(J) = \{r(j) : j \in J\}$ for $J \subseteq N$. Let $A = r(C)$ and $\gamma_0 = \min\{\gamma_j : j \in A\}$.

Claim. If $\gamma_j \geq \gamma_0$ for some $j \notin A$, then $j \in T_m$.

Proof of claim. Assume by contradiction that there is an index $j \notin A \cup T_m$ such that $\gamma_j \geq \gamma_0$. Since $j \notin A$, V_j was not assigned any variable in C; and since $j \notin T_m$, cluster V_j was feasible at the mth iteration of the algorithm. This implies that before the execution of the mth iteration, the pseudo-cost γ_j was at most as large as γ_k for every $k \in A$. But then the final pseudo-cost γ_j would be smaller than the final pseudo-cost γ_k for $k \in A$: this is because if V_j was assigned some variable at the mth iteration, the cost of this variable is smaller than c^* (as V_j was not assigned any variable in C), while V_k was assigned a variable of cost c^*, as $k \in A$. If we choose k to be an index in A such that $\gamma_k = \gamma_0$, we obtain a contradiction, as we assumed $\gamma_j \geq \gamma_0 = \gamma_k$. ◇

Define $B = \{j \in T_m : \gamma_j \geq \gamma_0\}$. Note that $r(B) = B$, as $B \subseteq T$. By the claim, $\gamma_j \geq \gamma_0$ if and only if $j \in A \cup B$. Then, since $r(B \cup C) = A \cup B$, independently of the tie-breaking rules we have

$$\{\bar{x}_j : j \in B \cup C\} = \{1, \dots, |B \cup C|\}. \tag{13}$$

Let $\hat{c} = \max\{c_j : c_j < c^*, j \in U_m\}$, with $\hat{c} = -\infty$ if $c_j = c^*$ for all $j \in U_m$, and $\hat{\gamma} = \max\{\gamma_j : \gamma_j < \gamma_0\}$, with $\hat{\gamma} = -\infty$ if γ_0 is the minimum of all pseudo-costs. Define $\delta = \min\{c^*, c^* - \hat{c}, \gamma_0 - \hat{\gamma}\} > 0$.

Construct a new instance I' that is identical to I, except that the costs now are

$$c'_j = \begin{cases} c_j - \delta, & j \in B \cup C, \\ c_j, & j \notin B \cup C. \end{cases}$$

We claim that by applying the algorithm to I', we obtain the same solution $\bar{x}' = \bar{x}$. To see this, note that the first $m-1$ iterations of the algorithm are identical for I and I' (if every tie is broken with the same criterion as that adopted when solving I). Indeed, we only changed the costs of some variables in N_m, and since $T_m \subseteq T_i$ for all $i \in [m-1]$, this does not affect the pseudo-costs of the feasible clusters during the first $m-1$ iterations. At the mth iteration, at line 6 we can choose the same ordering of the variables in U_m as we did when solving instance I: this is because the only variables in U_m whose cost has changed are the x_j's with $j \in C$, and $c'_j = c^* - \delta \geq \hat{c}$ for $j \in C$. Since, after the $(m-1)$th iteration, the pseudo-costs of the clusters V_j with $j \notin T_m$ are the same as they were for instance I, we can assign the elements in U_m to the feasible clusters exactly as we did for I. It follows that after the mth iteration the pseudo-costs for I' are

$$\gamma'_j = \begin{cases} \gamma_j - \delta, & j \in A \cup B, \\ \gamma_j, & j \notin A \cup B. \end{cases}$$

Since $\gamma'_j = \gamma_j - \delta \geq \gamma_0 - \delta \geq \hat{\gamma}$ for all $j \in A \cup B$, at line 12 we can choose the same ordering as we did for I. We then obtain the same solution as for instance I, as claimed.

By the choice of δ, for I' either the number of distinct nonzero costs of the variables in U_m is $p-1$ (this happens if

$\delta \in \{c^*, c^* - \hat{c}\}$), or the number of distinct pseudo-costs is $q-1$ (this happens if $\delta = \gamma_0 - \hat{\gamma}$). Then, by the minimality of instance I, there is a dual solution $(\bar{y}, \bar{z})$ that satisfies conditions (a) and (b) for I'. By increasing $\bar{y}_{B\cup C}$ by δ, we obtain a dual solution for I, with conditions (a) and (b) still satisfied because of (13). This is a contradiction.

Case 3. If $c_j < 0$ for all $j \in U_m$, the proof is similar to that of Case 2. Specifically, one defines $c^* = \min\{c_j : j \in U_m\} < 0$, $C = \{j \in U_m : c_j = c^*\}$, $A = r(C)$, and $\gamma_0 = \max\{\gamma_j : j \in A\}$. The claim now states that if $\gamma_j \leq \gamma_0$ for some $j \notin A$, then $j \in T_m$. One then defines $B = \{j \in T_m : \gamma_j \leq \gamma_0\}$ and checks that $\{\bar{x}_j : j \in B \cup C\} = \{d - |B \cup C| + 1, \ldots, d\}$. After defining $\hat{c} = \min\{c_j : c_j > c^*, j \in U_m\}$, $\hat{\gamma} = \min\{\gamma_j : \gamma_j > \gamma_0\}$, and $\delta = \min\{-c^*, \hat{c} - c^*, \hat{\gamma} - \gamma_0\} > 0$, the costs of the new instance I' are given by

$$c'_j = \begin{cases} c_j + \delta, & j \in B \cup C, \\ c_j, & j \notin B \cup C. \end{cases}$$

Finally, the dual solution $(\bar{y}, \bar{z})$ of I' is modified by increasing $\bar{z}_{B\cup C}$ by δ. ☐

Corollary 2. *Under the inclusion property, system* (8)–(9) *is totally dual integral.*

Proof. The above proof shows that if c is an integer vector then there is an optimal dual solution with integer components. (The existence of such a solution when $m = 1$, which is needed in the base step of the proof, was shown in [5].) ☐

The proof of Theorem 1 can be straightforwardly converted into a recursive algorithm that, given the output of Algorithm 1, constructs an optimal dual solution.

3.1. A remark

One might wonder whether Theorem 1 and Corollary 2 can be proved more directly via the theory of submodular functions. We recall that a set function $\psi : 2^N \to \mathbb{R}$ is *submodular* if

$$\psi(S_1) + \psi(S_2) \geq \psi(S_1 \cup S_2) + \psi(S_1 \cap S_2)$$

for every $S_1, S_2 \subseteq N$, while a set function $\varphi : 2^N \to \mathbb{R}$ is *supermodular* if

$$\varphi(S_1) + \varphi(S_2) \leq \varphi(S_1 \cup S_2) + \varphi(S_1 \cap S_2) \tag{14}$$

for every $S_1, S_2 \subseteq N$. It is known that if φ (resp., ψ) is a supermodular (resp., submodular) function defined on 2^N, then the polyhedron described by the inequalities

$$\sum_{j\in S} x_j \geq \varphi(S), \quad S \subseteq N, \tag{15}$$

$$\sum_{j\in S} x_j \leq \psi(S), \quad S \subseteq N, \tag{16}$$

is totally dual integral: this is a classical result on polymatroids [1] (see also [7, Theorem 46.2]).

In our system (8)–(9), we do not have constraints for every $S \subseteq N$, but only for $S \in \mathcal{S} = \bigcup_{i=1}^m \{S : S \subseteq N_i\}$. Thus the definitions of f and g given in (5) only apply to $S \in \mathcal{S}$, while the value of f and g on the subsets in $2^N \setminus \mathcal{S}$ might be defined in a different way, if needed. Assume that, under the inclusion property, f (resp., g) can be extended to a supermodular function φ (resp., submodular function ψ) defined on 2^N in such a way that the integer solutions to (15)–(16) are precisely the integer vectors in the convex hull of (6)–(7). Then Theorem 1 and Corollary 2 would follow immediately. However, we now show that in general such an extension does not exist.

Consider the all-different system with $N = \{1, 2, 3\}$, $m = 2$, $N_1 = \{1, 2\}$, $N_2 = \{2, 3\}$. The inclusion property is clearly satisfied. We show that if $\varphi, \psi : 2^N \to \mathbb{R}$ are extensions of f, g such that the integer solutions to (15)–(16) are precisely the integer vectors in the convex hull of (6)–(7), then φ violates inequality (14) for $S_1 = N_1$ and $S_2 = N_2$. First, note that $\varphi(N_1) = f(N_1) = 3$, $\varphi(N_2) = f(N_2) = 3$, and $\varphi(N_1 \cap N_2) = f(N_1 \cap N_2) = 1$. On the other hand, the value of $\varphi(N_1 \cup N_2) = \varphi(N)$ cannot be determined a priori, as $N \notin \mathcal{S}$. However, since the vector $(\bar{x}_1, \bar{x}_2, \bar{x}_3) = (1, 2, 1)$ must be a feasible solution, (15) holds only if $\varphi(N_1 \cup N_2) \leq \bar{x}_1 + \bar{x}_2 + \bar{x}_3 = 4$. Then $\varphi(N_1) + \varphi(N_2) = 6$ and $\varphi(N_1 \cup N_2) + \varphi(N_1 \cap N_2) \leq 5$, and therefore φ violates inequality (14).

4. An extension

We finally present an extension of Theorem 1 and Corollary 2, which can be used, for instance, to deal with the case of an all-different system with the inclusion property and an arbitrary finite domain $D \subseteq \mathbb{R}$. In what follows, we say that a function $\phi : \mathbb{N} \to \mathbb{R}$ is convex (resp., concave) if the piecewise linear interpolation of ϕ is convex (resp., concave).

Consider a system of the form

$$\sum_{j\in S} x_j \geq f(S), \qquad S \subseteq N_i,\ i \in [m], \tag{17}$$

$$-\sum_{j\in S} x_j \geq -g(S), \quad S \subseteq N_i,\ i \in [m], \tag{18}$$

where $f(S) = \alpha(|S|)$ for some convex function $\alpha : \mathbb{N} \to \mathbb{R}$, $g(S) = \beta(|S|)$ for some concave function $\beta : \mathbb{N} \to \mathbb{R}$, and $\alpha(0) = \beta(0) = 0$. We assume that $\alpha(k) \leq \beta(k)$ for all $k \in \mathbb{N}$, otherwise the system is infeasible. Note that system (8)–(9) and the functions defined in (5) are of this form.

When $m = 1$, system (17)–(18) is totally dual integral: this follows from the fact that f is a supermodular function and g is a submodular function (see [4, Proposition 5.1]), along with the result on polymatroids mentioned in Section 3.1. However, in general the above system is not totally dual integral for $m > 1$. We now observe that we have total dual integrality if the inclusion property holds.

Theorem 3. *Assume that $f(S) = \alpha(|S|)$ for some convex function $\alpha : \mathbb{N} \to \mathbb{R}$, and $g(S) = \beta(|S|)$ for some concave function $\beta : \mathbb{N} \to \mathbb{R}$, with $\alpha(k) \le \beta(k)$ for all $k \in \mathbb{N}$ and $\alpha(0) = \beta(0) = 0$. Then, under the inclusion property, system (17)–(18) is totally dual integral. Thus, if α and β are integer-valued, the polyhedron defined by inequalities (17)–(18) is integral.*

Proof. We extend Algorithm 1 so that it solves linear optimization over (17)–(18). For easiness of notation, here we write r_j instead of $r(j)$ to denote the index of the cluster containing x_j.

The only modification of the algorithm is at line 13, where we now set

$$\bar{x}_j = \begin{cases} \alpha\left(\sigma^{-1}(r_j)\right) - \alpha\left(\sigma^{-1}(r_j) - 1\right) & \text{if } \gamma_{r_j} \ge 0, \\ \beta\left(\sigma^{-1}(r_j)\right) - \beta\left(\sigma^{-1}(r_j) - 1\right) & \text{if } \gamma_{r_j} < 0. \end{cases}$$

Note that when f and g are the functions defined in (5), this assignment coincides with that in line 13 of Algorithm 1.

In the following we prove that the solution returned by the modified algorithm is feasible for (17)–(18), and then we observe that it can be completed with a dual solution satisfying the complementary slackness conditions.

We show that $\bar{x}$ satisfies (17) for every $S \subseteq N_i$, $i \in [m]$. First we observe that since α is a convex function,

$$\alpha(k) - \alpha(k-1) \le \alpha(h) - \alpha(h-1) \quad \text{for } h \ge k \ge 1. \tag{19}$$

Now fix $S \subseteq N_i$ for some $i \in [m]$. If we define $S^+ = \{j \in S : \gamma_{r_j} \ge 0\}$ and $S^- = \{j \in S : \gamma_{r_j} < 0\}$, then

$$\begin{aligned} \sum_{j \in S} \bar{x}_j &= \sum_{j \in S^+} \left(\alpha\left(\sigma^{-1}(r_j)\right) - \alpha\left(\sigma^{-1}(r_j) - 1\right)\right) \\ &\quad + \sum_{j \in S^-} \left(\beta\left(\sigma^{-1}(r_j)\right) - \beta\left(\sigma^{-1}(r_j) - 1\right)\right) \\ &\ge \sum_{j \in S} \left(\alpha\left(\sigma^{-1}(r_j)\right) - \alpha\left(\sigma^{-1}(r_j) - 1\right)\right) \\ &\ge \sum_{k=1}^{|S|} \left(\alpha(k) - \alpha(k-1)\right) = \alpha(|S|) = f(S), \end{aligned} \tag{20}$$

where the first inequality holds because $\alpha(k) \le \beta(k)$ for all $k \in \mathbb{N}$, and the second inequality follows from (19) along with the fact that the indices $\sigma^{-1}(r_j)$ for $j \in S$ are pairwise distinct. This shows that $\bar{x}$ satisfies (17); for inequalities (18), the proof is similar.

The rest of the proof is the same as the proof of Theorem 1, except that conditions (a) and (b) need to be adapted to this more general context. Note that the complementary slackness conditions take again the form (a′)–(b′) of the proof of Theorem 1. By (20), it follows that $\sum_{j \in S} \bar{x}_j = f(S)$ if and only if $\{\bar{x}_j : j \in S\} = \{\alpha(k) - \alpha(k-1) : k = 1, \ldots, |S|\}$. Similarly, one proves that $\sum_{j \in S} \bar{x}_j = g(S)$ if and only if $\{\bar{x}_j : j \in S\} = \{\beta(k) - \beta(k-1) : k = 1, \ldots, |S|\}$. Therefore, the complementary slackness conditions can be written in the following form:

(a) for every $S \in \mathcal{S}$, if $\bar{y}_S > 0$ then $\{\bar{x}_j : j \in S\} = \{\alpha(k) - \alpha(k-1) : k = 1, \ldots, |S|\}$;

(b) for every $S \in \mathcal{S}$, if $\bar{z}_S > 0$ then $\{\bar{x}_j : j \in S\} = \{\beta(k) - \beta(k-1) : k = 1, \ldots, |S|\}$.

The proof now proceeds similarly to that of Theorem 1. □

Remark 4. *The above result implies in particular that Theorem 1 and Corollary 2 also hold if D is an arbitrary finite subset of $\mathbb{R}$ (with $|D| \ge n$), provided that $f(S)$ (resp., $g(S)$) is defined as the sum of the $|S|$ smallest (resp., largest) elements in D, for every $S \subseteq N$. (These functions f and g are easily checked to satisfy the conditions of Theorem 3.)*

Acknowledgements

The author would like to thank Michele Conforti and Laurence A. Wolsey for commenting on a previous version of the paper.

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

[2] M. X. Goemans, *Smallest compact formulation for the permutahedron.* `http://math.mit.edu/~goemans/PAPERS/permutahedron.pdf`, 2009.

[3] J. Lee, *All-different polytope*, J. Comb. Optim. 6 (2002) 335–352.

[4] L. Lovász, *Submodular functions and convexity*, in Mathematical Programming: The State of the Art, Springer, 1983, pp. 235–257.

[5] D. Magos, *The constraint of difference and total dual integrality*, in Proceeding of PCI'13, ACM, New York, 2013, pp. 188–194.

[6] D. Magos, I. Mourtos, and G. Appa, *A polyhedral approach to the alldifferent system*, Math. Program. 132 (2012) 209–260.

[7] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*, Springer, 2003.

[8] W. J. van Hoeve, *The alldifferent constraint: A survey.* arXiv:cs/0105015, 2001.

[9] H. P. Williams and H. Yan, *Representations of the all_different predicate of constraint satisfaction in integer programming*, INFORMS J. Comput. 13 (2001) 96–103.