DESIGNING CODES FOR STORAGE ALLOCATION

SASCHA KURZ

ABSTRACT. Service rates for storage allocation were considered in [11]. In this notes we consider the design of good or optimal codes with respect to this metric. The cases of two files is completely and the case of three files is partially resolved, see also [8] where a subset of these results are presented in a more compact way.

Keywords: distributed storage; linear codes; service rates of codes

1. PRELIMINARIES

Suppose there are k files $f_1, \ldots, f_k$ with request rates $\lambda_1, \ldots, \lambda_k$. The service rate region $\mathcal{S}(\mathbf{G}, \mu) \subseteq \mathbb{R}^k_{\ge 0}$ is defined as the set of all request vectors λ that can be served by a coded storage system with generator matrix $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ and service rate μ. In the following we assume $\mu = \mathbf{1}$, i.e., $\mu_i = 1$ for all $i \in [n]$, where $[n] = \{1, \ldots, n\}$ for each integer n, and abbreviate $\mathcal{S}(\mathbf{G}, \mathbf{1})$ as $\mathcal{S}(\mathbf{G})$. In order to be more precise we need to introduce more notation. A linear code $\mathcal{C}$ of dimension k over $\mathbb{F}_q$ can be described by a generator matrix $\mathbf{G} \in \mathbb{F}_q^{k\times n}$. Note that there are usually generator matrices that span the same linear code, i.e., whenever the row span of two matrices $\mathbf{G}$ and $\mathbf{G}'$ coincides, the span the same code. Another representation of a linear code $\mathcal{C}$ over $\mathbb{F}_q$ is a multiset $\mathcal{G}$ of points in $\mathrm{PG}(k-1, q)$, where a point is a 1-dimensional subspace of $\mathbb{F}_q^k$. In what follows, we restrict ourselves to the binary field $\mathbb{F}_2$, which allows us to simplify the notation a bit. First we associate the points of $\mathrm{PG}(k-1, 2)$ with the non-zero vectors in $\mathbb{F}_2^k$, then we interpret each such vector v as the binary expansion of a corresponding integer $1 \le i \le l := 2^k - 1$. We denote the vector corresponding to the integer $i \in [l]$ by v_i. As examples, the vector $v_4 = (1, 0, 0)$ corresponds to the integer 4 and the vector $v_3 = (0, 1, 1)$ corresponds to the integer 3. In order to uniquely characterize a multiset of points $\mathcal{G}$ in $\mathrm{PG}(k-1, 2)$ we use multiplicities $n_i \in \mathbb{N}$, where $i \in [l]$, counting the number of occurrences of the vector v_i in $\mathbb{F}_2^k\backslash\{0\}$, where $i \in [l]$, in the generator matrix $\mathbf{G}$. So, we have $\sum_{i\in[l]} n_i = n$. The notion of a multiset of points $\mathcal{G}$ factors out the symmetry of column permutations of corresponding generator matrices $\mathbf{G}$. Due to the correspondence between a generator matrix $\mathbf{G}$ and a multiset of points $\mathcal{G}$ we also write $\mathcal{S}(\mathcal{G})$ instead of $\mathcal{S}(\mathbf{G})$ for the service rate region and remark that we will directly define $\mathcal{S}(\mathcal{G})$ later on.

A recovery set Y for file f_i, where $i \in [k]$, is a subset of $S \subseteq [l]$ such that the span $\langle\{v_j \mid j \in S\}\rangle$ contains the ith unit vector $\mathbf{e}_i$. We call a recovery set S reduced for i there does not exists a proper subset $S' \subsetneq S$ with $\mathbf{e}_i \in \langle\{v_j \mid j \in S'\}\rangle$. For $q = 2$ and a reduced recovery set S there is no need to specify the index i of the file that is recovered since $\sum_{j\in S} v_j = \mathbf{e}_i$. However, in $\mathbb{F}_3$ the set $\{\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 + 2\mathbf{e}_2\}$ spans a 2-dimensional subspace containing both $\mathbf{e}_1$ and $\mathbf{e}_2$, while none of these two unit vectors is contained in the span of a proper subset. Since we assume $q = 2$, we will mostly speak just of a recovery set without specifying i. By $\mathcal{Y}^i$ we denote the set of all reduced recovery sets for file f_i, where $i \in [k]$. For $k = 3$ and $i = 2$ we have

$$\mathcal{Y}^2 = \Big\{\{2\}, \{4, 6\}, \{1, 3\}, \{5, 7\}, \{1, 4, 7\}\Big\},$$

which corresponds to

$$\Big\{\{(0,1,0)\}, \{(1,0,0), (1,1,0)\}, \{(0,0,1), (0,1,1)\}, \{(1,0,1), (1,1,1)\}, \{(0,0,1), (1,0,0), (1,0,1)\}\Big\}.$$

We remark that the maximum cardinality of a reduced recovery set is k, which can indeed be attained.

Given a multiset of points $\mathcal{G}$ in $\mathrm{PG}(k-1,2)$, described by the multiplicities n_j, the service rate region $\mathcal{S}(\mathcal{G})$ is the set of all vectors $\lambda \in \mathbb{R}_{\ge 0}^k$ for which there exists $\alpha_{i,Y}$, satisfying the following constraints:

$$\sum_{Y\in\mathcal{Y}^i} \alpha_{i,Y} = \lambda_i, \text{ for all } i \in [k], \tag{1a}$$

$$\sum_{i=1}^{k} \sum_{Y\in\mathcal{Y}^i\,:\,j\in Y} \alpha_{i,Y} \le n_j, \text{ for all } j \in [l], \tag{1b}$$

$$\alpha_{i,Y} \in \mathbb{R}_{\ge 0}, \text{ for all } i \in [k],\ Y \in \mathcal{Y}^i. \tag{1c}$$

The constraints (1a) guarantee that the demands for all files are served, and constraints (1b) ensure that no node receives requests at a rate in excess of its service rate.

As noted before, for $q = 2$, each reduced recovery set uniquely characterizes the file it recovers. In other words the $\mathcal{Y}^i$ are pairwise disjoint and form a partition of $\mathcal{Y} := \cup_{i\in[k]}\mathcal{Y}^i$. With this we can simplify the above characterization, i.e., the service rate region $\mathcal{S}(\mathcal{G})$ is the set of all vectors $\lambda \in \mathbb{R}_{\ge 0}^k$ for which there exists α_Y, satisfying the following constraints:

$$\sum_{Y\in\mathcal{Y}^i} \alpha_Y \ge \lambda_i, \text{ for all } i \in [k], \tag{2a}$$

$$\sum_{Y\in\mathcal{Y}\,:\,j\in Y} \alpha_Y \le n_j, \text{ for all } j \in [l], \tag{2b}$$

$$\alpha_Y \in \mathbb{R}_{\ge 0}, \text{ for all } Y \in \mathcal{Y}. \tag{2c}$$

Note that constraint (2a) looks like a relaxation of constraint (1a), while it does not matter for the definition of $\mathcal{S}(\mathcal{G})$ if we use "$=$" or "$\ge$".

After these preparations we can come to the main questions of this paper. For each (bounded) subset $\mathcal{R} \subset \mathbb{R}_{\ge 0}^k$ we can ask for the minimum number $n(\mathcal{R})$ of servers such that there exists a generator matrix $\mathbf{G} \in \mathbb{F}_2^{k\times n}$ with $\mathcal{R} \subseteq \mathcal{S}(\mathbf{G})$ (or alternatively, such that there exists a multiset of points $\mathcal{G}$ in $\mathrm{PG}(k-1,2)$ with $\mathcal{R} \subseteq \mathcal{S}(\mathcal{G})$). So, we ask for lower bounds for $n(\mathcal{R})$ and constructive upper bounds for $n(\mathcal{R})$, i.e., the construction of good codes. Note that we can have $\mathcal{S}(\mathbf{G}) \ne \mathcal{S}(\mathbf{G}')$ or $\mathcal{S}(\mathcal{G}) \ne \mathcal{S}(\mathcal{G}')$ even if $\mathbf{G}, \mathbf{G}'$ or $\mathcal{G}, \mathcal{G}'$ generate the same linear code $\mathcal{C}$, so that we have to speak of the construction of good generator matrices or good multisets of points, in order to be more precise.

Before we give integer linear programming (ILP) formulations for the determination of $n(\mathcal{R})$ we first study a few structural properties.

Lemma 1.1. *We have $n(\mathcal{R}) = n(\mathrm{conv}(\mathcal{R}))$, where $\mathrm{conv}(\mathcal{R})$ is the convex hull of $\mathcal{R}$.*

Proof. It suffices to observe that the service rate region $\mathcal{S}(\mathbf{G})$ of every generator matrix $\mathbf{G} \in \mathbb{F}_2^{k\times n}$ is convex. □

The relation $\mathbf{x} \le \mathbf{y}$, i.e., $x_i \le y_i$ for all $1 \le i \le k$, forms a poset in $\mathbb{R}_{\ge 0}^k$ with the unique minimal element $\mathbf{0}$. In that context, the lower set $S{\downarrow}$ of a subset $S \subseteq \mathbb{R}_{\ge 0}^k$ is defined via $S{\downarrow} := \left\{x \in \mathbb{R}_{\ge 0}^k \mid \exists y \in S : x \le y\right\}$. As an example we consider the set $S = \mathrm{conv}(\{(0,0),(1,2),(2,1)\}) \subset \mathbb{R}_{\ge 0}^2$, which is a triangle with area $\frac{3}{2}$. Here, the corresponding lower set

$$S{\downarrow} = \mathrm{conv}(\{(0,0),(0,2),(2,0),(1,2),(2,1)\})$$

is a pentagon with area $\frac{7}{2}$.

Lemma 1.2. *We have $n(\mathcal{R}) = n(\mathcal{R}{\downarrow})$, where $\mathcal{R}{\downarrow}$ is the lower set of $\mathcal{R}$.*

Proof. It suffices to observe that the service rate region $\mathcal{S}(\mathbf{G})$ of every generator matrix $\mathbf{G} \in \mathbb{F}_2^{k\times n}$ is its own lower set, i.e., $\mathcal{S}(\mathbf{G}) = \mathcal{S}(\mathbf{G})\downarrow$. □

Taken the above two observations into account, we want to parameterize a large class of reasonable subsets $\mathcal{R} \subset \mathbb{R}^k_{\geq 0}$ by a function $T\colon 2^{\{1,\dots,k\}} \to \mathbb{N}$ that maps the subsets of $\{1,\dots,k\}$ to integers, where $T(\emptyset) = 0$.

Definition 1.3. Let $T\colon 2^{\{1,\dots,k\}} \to \mathbb{N}$ with $T(\emptyset) = 0$. With this, we set

$$\mathcal{R}(T) := \left\{ \lambda \in \mathbb{R}^k_{\geq 0} \mid \sum_{i\in S} \lambda_i \leq T(S)\, \forall \emptyset \neq S \subseteq \{1,\dots,k\} \right\}$$

and abbreviate $n(\mathcal{R}(T)))$ as $n(T)$.

By construction $\mathcal{R}(T)$ is a polytope, i.e., a bounded polyhedron, which especially is convex, see e.g. [7] for more details. Moreover, $\mathcal{R}(T)\downarrow = \mathcal{R}(T)$, i.e., $\mathcal{R}(T)$ is its own lower set. In some cases we can modify values of the function T without changing $\mathcal{R}(T)$.

Lemma 1.4. *Let $T\colon 2^{\{1,\dots,k\}} \to \mathbb{N}$ with $T(\emptyset) = 0$ and let T' be given by the following algorithm:*

```
for each S ⊆ {1,...,k} do
  T'(S) ← T(S)
end for
changed ← true
while changed = true do
  changed ← false
  for each S ⊆ {1,...,k} do
    for each ∅ ≠ U ⊊ S do
      if T'(S) > T'(U) + T'(S\U) then
        T'(S) ← T'(U) + T'(S\U)
        changed ← true
      end if
    end for
    for each S ⊊ V ⊆ {1,...,k} do
      if T'(S) > T'(V) then
        T'(S) ← T'(V)
        changed ← true
      end if
    end for
  end for
end while
```

Then, we have $\mathcal{R}(T) = \mathcal{R}(T')$. Moreover, if we apply the algorithm again on T' and obtain T'', then $T' = T''$.

Proof. After the first initializing loop we obviously have $\mathcal{R}(T) = \mathcal{R}(T')$. Now we consider a single step where $T'(S)$ is replaced by either $T'(U)+T'(S\backslash U)$ or $T'(V)$. Inductively we know that each $\lambda \in \mathcal{R}(T')$ satisfies $\sum_{i\in S'} \lambda_i \leq T'(S')$ for all $S' \subseteq \{1,\dots,k\}$. Since this especially holds for $S' = U$, $S' = S\backslash U$, and $S' = V$ we also have

$$\sum_{i\in S} \lambda_i \leq T'(U) + T'(S\backslash U)$$

and

$$\sum_{i\in S}\lambda_i \overset{\lambda\geq 0}{\leq} \sum_{i\in V}\lambda_i \leq T'(V).$$

So, after each replacement we still have $\mathcal{R}(T)=\mathcal{R}(T')$.

In order to show that the algorithm terminates let

$$\varepsilon=\min\{T(U)-T(V)\mid \emptyset\subseteq U,V\subseteq\{1,\dots,k\},T(U)-T(V)\}.$$

By induction over the number of replacements we can easily show that $\varepsilon \leq \min\{T'(U)-T'(V) \mid \emptyset\subseteq U,V\subseteq\{1,\dots,k\},T'(U)-T'(V)\}$ at each time after the initialization loop. Thus, every replacement reduces the value of $\sum_{S\subseteq\{1,\dots,\}}T'(S)$ by at least ε, so that the algorithm terminates after at least $\left(\sum_{S\subseteq\{1,\dots,\}}T(S)\right)/\varepsilon+1$ iterations of the while loop.

Since in the last iteration of the while loop non of the if-conditions were true, this is also the case if we apply the algorithm again. □

We remark that the function T' constructed by the algorithm of Lemma 1.4 is subadditive, i.e., we have $T'(U)+T'(V)\geq T'(U\cup V)$ (since T' is non-negative it is no necessary to restrict to the cases where $U\cap V=\emptyset$), and monotone, i.e., we have $T'(U)\leq T'(V)$ for all $\emptyset\subseteq U\subseteq V\subseteq\{1,\dots,k\}$. Indeed, the proof of the following characterization is easy:

Lemma 1.5. *A function $T\colon 2^{\{1,\dots,k\}}\to\mathbb{N}$, with $T(\emptyset)=0$, satisfies $T'=T$, where T' is the result of the algorithm of Lemma 1.4 applied to T, iff T is monotone and subadditive.*

As an example we remark that a function for $k=1$ each function $T\colon 2^{\{1,\dots,k\}}\to\mathbb{N}$ is monotone and subadditive, while for $k=2$ the conditions can be summarized to

$$\max\{T(\{1\}),T(\{2\})\} \quad\leq\quad T(\{1,2\}) \quad\leq\quad T(\{1\})+T(\{2\}). \tag{3}$$

Definition 1.6. Let $\mathcal{R}\subseteq\mathbb{R}^k_{\geq 0}$ be a subset that cannot be enlarged by building the lower set, i.e. $\mathcal{R}{\downarrow}=\mathcal{R}$. Then, we say that a finite set $S\subseteq\mathbb{R}^k_{\geq 0}$ is a generating set of $\mathcal{R}$ if $\operatorname{conv}(S){\downarrow}=\mathcal{R}$. Moreover, we call S minimal if no proper subset of S is a generating set of $\mathcal{R}$.

As an example we consider the function $T\colon 2^{\{1,2\}}\to\mathbb{N}$ given by $T(\emptyset)=0$, $T(\{1\})=T(\{2\})=2$, and $T(\{1,2\})=3$. Here, a generating set of $\mathcal{R}(T)$ is given by $\{(1,2),(1,2)\}$. Actually, the generating set of $\mathcal{R}(T)$ is always unique, since $\mathcal{R}(T)$ is a polytope that can be written as $\mathcal{R}(T)=\operatorname{conv}(V)$, where V is the set of vertices of the polytope, which is the unique minimal set with $\mathcal{R}(T)=\operatorname{conv}(V)$. We obtain a generating set of $\mathcal{R}$ from V be removing all $v\in V$ such there is a different $v'\in V$ with $v\leq v'$.

Before we study bounds for $n(\mathcal{R}(T))$, we give ILP formulations for the determination of $n(\mathcal{R})$.

Proposition 1.7. *Let $\left\{\lambda^{(1)},\dots,\lambda^{(m)}\right\}$ be a generating set of $\mathcal{R}$, i.e., we assume that $\mathcal{R}{\downarrow}=\mathcal{R}$. Then, $n(\mathcal{R})$ coincides with the optimal target value of*

$$\begin{aligned}
\min \quad & \sum_{j\in[l]} n_j \\
\sum_{Y\in\mathcal{Y}\,:\,j\in Y}\alpha^i_Y \leq n_j \quad & \forall j\in[l],\forall i\in[m] \\
\sum_{Y\in\mathcal{Y}^j}\alpha^i_Y \geq \lambda^{(i)}_j \quad & \forall i\in[m],\, j\in[k] \\
n_j\in\mathbb{N} \quad & \forall j\in[l] \\
\alpha^i_Y\in\mathbb{R}_{\geq 0} \quad & \forall i\in[m],\forall Y\in\mathcal{Y},
\end{aligned}$$

Proof. Let the multiset of points $\mathcal{G}$ be uniquely characterized by the integer multiplicities n_j, $j \in [l]$. The stated ILP formulation minimizes the code size $n = \sum_{j\in[l]} n_j$ and ensures that $\lambda^[(i)] \in \mathcal{S}(\mathcal{G})$ by using the characterization (2a)–(2c) for each $i \in [m]$. □

The drawback of the ILP formulation of Proposition 1.7 is that $\#\mathcal{Y}$ grows doubly exponential, i.e., $\#\mathcal{Y}$ gets quite large, even for moderate values of k.

Example 1.8. For $q = 2$, $k = 2$ consider the desired service rate region

$$\mathcal{R} = \left\{(\lambda_1, \lambda_2) \in \mathbb{R}^2_{\geq 0} \,:\, \lambda_1 \leq 2, \lambda_2 \leq 2, \lambda_1 + \lambda_2 \leq 3\right\}.$$

and generating set $\left\{\lambda^{(1)}, \lambda^{(2)}\right\}$ of cardinality $m = 2$, where $\lambda^{(1)} = (2, 1)$ and $\lambda^{(2)} = (1, 2)$. The possible columns of a generator matrix $\mathbf{G}$, i.e., the non-zero vectors in $\mathbb{F}_2^2$ are

$$v_1 = (0, 1),\;\; v_2 = (1, 0), \text{and } v_3 = (1, 1).$$

The recovery sets are given by

$$\mathcal{Y}^1 = \Big\{\{2\}, \{1, 3\}\Big\}.$$

and

$$\mathcal{Y}^2 = \Big\{\{1\}, \{2, 3\}\Big\}$$

With this, the ILP of Proposition 1.7 for the determination of $n(\mathcal{R})$ is:

$$\begin{aligned}
\min n_1 + n_2 + n_3 & \\
\alpha^1_{\{1\}} + \alpha^1_{\{1,3\}} \leq n_1 & \\
\alpha^1_{\{2,3\}} + \alpha^1_{\{2\}} \leq n_2 & \\
\alpha^1_{\{2,3\}} + \alpha^1_{\{1,3\}} \leq n_3 & \\
\alpha^2_{\{1\}} + \alpha^2_{\{1,3\}} \leq n_1 & \\
\alpha^2_{\{2,3\}} + \alpha^2_{\{2\}} \leq n_2 & \\
\alpha^2_{\{2,3\}} + \alpha^2_{\{1,3\}} \leq n_3 & \\
\alpha^1_{\{2\}} + \alpha^1_{\{1,3\}} \geq 2 & \\
\alpha^1_{\{1\}} + \alpha^1_{\{2,3\}} \geq 1 & \\
\alpha^2_{\{2\}} + \alpha^2_{\{1,3\}} \geq 1 & \\
\alpha^2_{\{1\}} + \alpha^2_{\{2,3\}} \geq 2 & \\
n_1, n_2, n_3 \in \mathbb{N} & \\
\alpha^i_{\{1\}}, \alpha^i_{\{2,3\}}, \alpha^i_{\{2\}}, \alpha^i_{\{1,3\}} \in \mathbb{R}_{\geq 0} & \qquad \forall i \in [2]
\end{aligned}$$

An optimal solution is given by $n_1 = 2$, $n_2 = 2$, and $n_3 = 0$, i.e., a code of length $n = 4$ with generator matrix

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Optimal multipliers for the recovery sets are given by $\alpha^1_{\{1\}} = 2$, $\alpha^1_{\{2,3\}} = 0$, $\alpha^1_{\{2\}} = 1$, $\alpha^1_{\{1,3\}} = 0$ and $\alpha^2_{\{1\}} = 1$, $\alpha^2_{\{2,3\}} = 0$, $\alpha^2_{\{2\}} = 2$, $\alpha^2_{\{1,3\}} = 0$. For the optimal multiset of points there are two further possibilities: $(n_1, n_2, n_3) = (2, 1, 1)$ and $(n_1, n_2, n_3) = (1, 2, 1)$.

If we only want to obtain an easier to computer lower bound for $n(\mathcal{R})$, then we can consider the LP relaxation of the ILP of Proposition 1.7:

$$\begin{aligned} \min \quad & \sum_{j\in[l]} n_j \\ n_j - \sum_{Y\in\mathcal{Y}\,:\,j\in Y} \alpha_Y^i \ge 0 \quad & \forall j\in[l], \forall i\in[m] \\ \sum_{Y\in\mathcal{Y}^j} \alpha_Y^i \ge \lambda_j^{(i)} \quad & \forall i\in[m],\, j\in[k] \\ n_j \in \mathbb{R}_{\ge 0} \quad & \forall j\in[l] \\ \alpha_Y^i \in \mathbb{R}_{\ge 0} \quad & \forall i\in[m], \forall Y\in\mathcal{Y}, \end{aligned}$$

The LP relaxation of the ILP in Example 1.8 is given by

$$\begin{aligned} \min n_1 + n_2 + n_3 & \\ n_1 - \alpha_{\{1\}}^1 - \alpha_{\{1,3\}}^1 \ge 0 & \\ n_2 - \alpha_{\{2,3\}}^1 - \alpha_{\{2\}}^1 \ge 0 & \\ n_3 - \alpha_{\{2,3\}}^1 - \alpha_{\{1,3\}}^1 \ge 0 & \\ n_1 - \alpha_{\{1\}}^2 - \alpha_{\{1,3\}}^2 \ge 0 & \\ n_2 - \alpha_{\{2,3\}}^2 - \alpha_{\{2\}}^2 \ge 0 & \\ n_3 - \alpha_{\{2,3\}}^2 - \alpha_{\{1,3\}}^2 \ge 0 & \\ \alpha_{\{2\}}^1 + \alpha_{\{1,3\}}^1 \ge 2 & \\ \alpha_{\{1\}}^1 + \alpha_{\{2,3\}}^1 \ge 1 & \\ \alpha_{\{2\}}^2 + \alpha_{\{1,3\}}^2 \ge 1 & \\ \alpha_{\{1\}}^2 + \alpha_{\{2,3\}}^2 \ge 2 & \\ n_1, n_2, n_3 \in \mathbb{R}_{\ge 0} & \\ \alpha_{\{1\}}^i, \alpha_{\{2,3\}}^i, \alpha_{\{2\}}^i, \alpha_{\{1,3\}}^i \in \mathbb{R}_{\ge 0} & \qquad \forall i\in[2] \end{aligned}$$

and has the unique optimal solution $n_1 = \frac{3}{2}$, $n_2 = \frac{3}{2}$, and $n_3 = \frac{1}{2}$ with optimal multipliers $\alpha_{\{1\}}^1 = \frac{3}{2}$, $\alpha_{\{2,3\}}^1 = \frac{1}{2}$, $\alpha_{\{2\}}^1 = 1$, $\alpha_{\{1,3\}}^1 = 0$ and $\alpha_{\{1\}}^2 = 1$, $\alpha_{\{2,3\}}^2 = 0$, $\alpha_{\{2\}}^2 = \frac{3}{2}$, $\alpha_{\{1,3\}}^2 = \frac{1}{2}$ for the recovery sets. The optimal target value $n = n1 + n_2 + n_3 = \frac{7}{2}$ can be rounded to 4 taking into account that the length of the code has to be an integer.

As mentioned before, the ILP formulation of Proposition 1.7 underlies a massive combinatorial explosion. To be more precise, the number of variables grows exponentially and the number of constraints grows doubly exponentially.

Lemma 1.9. *Let $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ be the generator matrix of an $[n,k]_q$ code $\mathcal{C}$ and $\mathcal{G}$ be the corresponding multiset of points of cardinality n described by point multiplicities n_j. If $\left\{\lambda^{(1)}, \ldots \lambda^{(m)}\right\}$ is a generating set of $\mathcal{R}$, then we have*

$$\sum_{v_j\in\mathrm{PG}(k-1,2)\setminus\mathcal{H}} n_j \ge \max\left\{ \sum_{s\in\mathcal{E}(\mathcal{H})} \lambda_s^{(i)} \mid 1\le i\le m \right\}, \tag{4}$$

where $\mathcal{H}$ is a hyperplane of $\mathrm{PG}(k-1,2)$ and

$$\mathcal{E}(\mathcal{H}) = \{h\in[k] \mid \mathbf{e}_h \notin \langle\{\mathbf{v} \mid \mathbf{v}\in\mathcal{H}\}\rangle\}$$

is the set of indices h such that the hyperplane $\mathcal{H}$ does not contain the unit vector $\mathbf{e}_h$, i.e., $\mathbf{e}_h$ lies in $\operatorname{PG}(k-1,2)\setminus\mathcal{H}$ *.*

Proof. Let $1 \le i \le m$ be an arbitrary index. From the ILP of Proposition 1.7 we conclude

$$\sum_{Y\in\mathcal{Y}^s} \alpha_Y^i \ge \lambda_s^{(i)} \tag{5}$$

for each $s \in \mathcal{E}(\mathcal{H})$ and

$$n_j \ge \sum_{Y\in\mathcal{Y}\,:\,j\in Y} \alpha_Y^i \overset{\alpha_Y^i\ge 0}{\ge} \sum_{s'\in\mathcal{E}(\mathcal{H})} \sum_{Y\in\mathcal{Y}^{s'}\,:\,j\in Y} \alpha_Y^i \tag{6}$$

for each $j \in [l]$ with $v_j \in \operatorname{PG}(k-1,2)\setminus\mathcal{H}$. Thus, we have

$$\sum_{v_j\in\operatorname{PG}(k-1,2)\setminus\mathcal{H}} n_j \ge \sum_{v_j\in\operatorname{PG}(k-1,2)\setminus\mathcal{H}} \sum_{s\in\mathcal{E}(\mathcal{H})} \sum_{Y\in\mathcal{Y}^s\,:\,j\in Y} \alpha_Y^i = \sum_{s\in\mathcal{E}(\mathcal{H})} \sum_{v_j\in\operatorname{PG}(k-1,2)\setminus\mathcal{H}} \sum_{Y\in\mathcal{Y}^s\,:\,j\in Y} \alpha_Y^i.$$

The unit vectors $\mathbf{e}_s$ with index s in $\mathcal{E}(\mathcal{H})$ are not contained in the chosen hyperplane $\mathcal{H}$, so that for each $Y \in \mathcal{Y}^s$ with $s \in \mathcal{E}(\mathcal{H})$ there exists an index $j \in [l]$ with $j \in Y$ and $v_j \in \operatorname{PG}(k-1,2)\setminus\mathcal{H}$. Thus, we conclude

$$\sum_{v_j\in\operatorname{PG}(k-1,2)\setminus\mathcal{H}} n_j \ge \sum_{s\in\mathcal{E}(\mathcal{H})} \sum_{Y\in\mathcal{Y}^s} \alpha_Y^i \ge \sum_{s\in\mathcal{E}(\mathcal{H})} \lambda_s^{(i)}$$

from Inequality (5). □

Corollary 1.10. *If $\left\{\lambda^{(1)},\dots,\lambda^{(m)}\right\}$ is a generating set of $\mathcal{R}$, then $n(\mathcal{R})$ is lower bounded by the optimal target value of*

$$\begin{aligned} \min \quad & \sum_{j\in[l]} n_j \\ \sum_{v_j\in\operatorname{PG}(k-1,2)\setminus\mathcal{H}} n_j \ge \max\left\{\sum_{s\in\mathcal{E}(\mathcal{H})} \lambda_s^{(i)} \mid 1\le i\le m\right\} \quad & \forall \text{ hyperplanes } \mathcal{H} \text{ of } \operatorname{PG}(k-1,2) \\ n_j \in \mathbb{N} \quad & \forall j \in [l]. \end{aligned}$$

Note that the ILP of Corollary 1.10 contains exactly $2^k - 1$ constraints and (integer) variables. So we have obtained a, with respect to Proposition 1.7, smaller formulation for the determination of $n(\mathcal{R})$. However, we only obtained a lower bound on $n(\mathcal{R})$. Indeed, from the context of private information retrieval (PIR) codes, see [9], we know that the optimal target value of the ILP of Corollary 1.10 can differ from $n(\mathcal{R})$, i.e., it can be strictly smaller.

Example 1.11. For $\lambda = (3, 4, 5)$ the ILP of Corollary 1.10 reads

$$\begin{aligned} n_4 + n_5 + n_6 + n_7 &\ge 3, \\ n_2 + n_3 + n_6 + n_7 &\ge 4, \\ n_1 + n_3 + n_5 + n_7 &\ge 5, \\ n_2 + n_3 + n_4 + n_5 &\ge 7, \\ n_1 + n_3 + n_4 + n_6 &\ge 8, \\ n_1 + n_2 + n_5 + n_6 &\ge 9, \\ n_1 + n_2 + n_4 + n_7 &\ge 12. \end{aligned}$$

An integral solution is e.g. given by $n_5 = 7$, $n_6 = 8$, $n_7 = 12$, and $n_i = 0$ for the remaining $i \in \{1, 2, 3, 4\}$. If $\mathcal{G}$ is the multiset of points in $\operatorname{PG}(3-1,2)$ that is uniquely described by the n_i, then we have $\lambda \in \mathcal{S}(\mathcal{G})$. We have $v_5 = (1, 0, 1)$, $v_6 = (1, 1, 0)$, and $v_7 = (1, 1, 1)$, so that the only usable recovery set for $\mathbf{e}_1$ is given by $\{5,6,7\}$, for $\mathbf{e}_2$ we only can use $\{5,7\}$, and for $\mathbf{e}_3$ the only possibility is $\{6,7\}$. Taking these recovery sets with multiplicities 3, 4, and 5 uses all available servers and is indeed the unique solution of the ILP of Proposition 1.7.

Lemma 1.12. *Let $\{\lambda\}$ be a generating set of $\mathcal{R} \subseteq \mathbb{R}^2_{\ge 0}$ and $\mathbf{n}$ be an integral solution of the ILP of Corollary 1.10. If $\lambda \in \mathbb{R}^2_{\ge 0}$ and $\mathcal{G}$ is the multiset corresponding to $\mathbf{n}$, then $\lambda \in \mathcal{S}(\mathcal{G})$, i.e., there exists a feasible choice of α_Y satisfying (2a)-(2c).*

Proof. The constraints of the ILP of Corollary 1.10 read

$$\begin{aligned} n_2 + n_3 &\ge \lambda_1, \\ n_1 + n_3 &\ge \lambda_2, \\ n_1 + n_2 &\ge \lambda_1 + \lambda_2 \end{aligned}$$

and the recovery sets are given by

$$\begin{aligned} \mathcal{Y}^1 &= \Big\{\{2\},\{1,3\}\Big\}, \\ \mathcal{Y}^2 &= \Big\{\{1\},\{2,3\}\Big\}. \end{aligned}$$

Setting

$$\begin{aligned} \alpha_{\{2\}} &= \min\{n_2, \lambda_1\}, \\ \alpha_{\{1\}} &= \min\{n_1, \lambda_2\}, \\ \alpha_{\{1,3\}} &= \max\{0, \lambda_1 - n_2\}, \\ \alpha_{\{2,3\}} &= \max\{0, \lambda_2 - n_1\} \end{aligned}$$

we have

$$\begin{aligned} \alpha_{\{2\}} + \alpha_{\{1,3\}} &= \lambda_1, \\ \alpha_{\{1\}} + \alpha_{\{2,3\}} &= \lambda_2, \\ \alpha_{\{1\}} + \alpha_{\{1,3\}} &\le n_1, \\ \alpha_{\{2\}} + \alpha_{\{2,3\}} &\le n_2, \text{ and} \\ \alpha_{\{1,3\}} + \alpha_{\{2,3\}} &\le n_3. \end{aligned}$$

Only the latter inequality needs a short case analysis. If $n_2 \ge \lambda_1$ and $n_1 \ge \lambda_2$, then $\alpha_{\{1,3\}} + \alpha_{\{2,3\}} = 0 \le n_3$. Since $n_1 + n_2 \ge \lambda_1 + \lambda_2$ we cannot have $n_2 < \lambda_1$ and $n_1 < \lambda_2$. So, let us assume $n_2 < \lambda_1$ and $n_1 \ge \lambda_2$. Then, $\alpha_{\{2,3\}} = 0$, $\alpha_{\{1\}} = \lambda_2$, $\alpha_{\{2\}} = n_2$, $\alpha_{\{1,3\}} = \lambda_1 - n_2$, and $\alpha_{\{1,3\}} + \alpha_{\{2,3\}} = \lambda_1 - n_2$, which is at most n_3 due to $n_2 + n_3 \ge \lambda_1$. The other case $n_2 \ge \lambda_1$ and $n_1 < \lambda_2$ follows analogously. □

In order to apply Corollary 1.10 to Example 1.8 we write $\overline{\mathcal{H}} = \{j \in [k] \mid v_j \in \mathcal{H}\}$ for each hyperplane $\mathcal{H}$ and obtain:

$$\begin{aligned} \mathcal{H}_1 = \{\mathbf{e}_2\}, \ \overline{\mathcal{H}}_1 = \{1\} \quad &\Rightarrow \quad n_2 + n_3 \ge 2, \\ \mathcal{H}_2 = \{\mathbf{e}_1\}, \ \overline{\mathcal{H}}_2 = \{2\} \quad &\Rightarrow \quad n_1 + n_3 \ge 2, \\ \mathcal{H}_3 = \{\mathbf{e}_1 + \mathbf{e}_2\}, \ \overline{\mathcal{H}}_1 = \{3\} \quad &\Rightarrow \quad n_1 + n_2 \ge 3. \end{aligned}$$

Summing up all three inequalities and dividing by two yields $n = n_1 + n_2 + n_3 \ge \frac{7}{2}$, sot that $n \ge 4$. As 3.5 is the optimal target value of the LP relaxation of the ILP from Proposition 1.7, it also has to be the optimal target value of the LP relaxation of the ILP from Corollary 1.10. Again the optimal ILP solutions are given by

$$(n_1, n_2, n_3) \in \Big\{(2,2,0), (2,1,1), (1,2,1)\Big\},$$

where all corresponding generator matrices $\mathbf{G}$ indeed achieve a service rate region $\mathcal{S}(\mathbf{G}) \supseteq \mathcal{R}$.

Let us consider another example in order to illustrate that solving the LP relaxation and uprounding the target value can yield a weaker bound than solving the corresponding ILP.

Example 1.13. For $q = 2$, $k = 3$ consider the desired service rate region $\mathcal{R} = \mathcal{R}(T)$, where $T(\emptyset) = 0$ and $T(S) = \#S + 1$ for $\emptyset \neq S \subseteq [3]$, i.e.,

$$\mathcal{R} = \left\{(\lambda_1, \lambda_2, \lambda_2) \in \mathbb{R}^3_{\geq 0} \,:\, \lambda_1 \leq 2, \lambda_2 \leq 2, \lambda_3 \leq 2, \lambda_1 + \lambda_2 \leq 3, \lambda_1 + \lambda_3 \leq 3, \lambda_2 + \lambda_3 \leq 3, \lambda_1 + \lambda_2 + \lambda_3 \leq 4\right\}.$$

A generating set $\left\{\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}\right\}$ or $\mathcal{R}$ of cardinality $m = 3$ is given by $\lambda^{(1)} = (2, 1, 1)$, $\lambda^{(2)} = (1, 2, 1)$, and $\lambda^{(3)} = (1, 1, 2)$. The possible columns of a generator matrix $\mathbf{G}$, i.e., the non-zero vectors in $\mathbb{F}_2^3$ are

$$v_1 = (0, 0, 1),\; v_2 = (0, 1, 0), v_3 = (0, 1, 1), v_4 = (1, 0, 0), v_5 = (1, 0, 1), v_6 = (1, 1, 0), \text{ and } v_7 = (1, 1, 1).$$

In order to write down the inequalities from Lemma 1.9 we describe a hyperplane $\mathcal{H}$ as a set of vectors $(x_1, x_2, x_3) \in \mathbb{F}_2^3 \setminus \{\mathbf{0}\}$ satisfying a certain constraint $\sum_{i=1}^{3} c_i x_i$, where $(c_1, c_2, c_3) \in \mathbb{F}_2^3 \setminus \{\mathbf{0}\}$:

$$\mathcal{H}_1 :\; x_1 = 0 \;\Rightarrow\; \mathbf{e}_1 \notin \mathcal{H}_1 \;\Rightarrow\; n_4 + n_5 + n_6 + n_7 \geq 2 = \max(\lambda_1^{(1)}, \lambda_1^{(2)}, \lambda_1^{(3)}) \tag{7}$$

$$\mathcal{H}_2 :\; x_2 = 0 \;\Rightarrow\; \mathbf{e}_2 \notin \mathcal{H}_2 \;\Rightarrow\; n_2 + n_3 + n_6 + n_7 \geq 2 = \max(\lambda_2^{(1)}, \lambda_2^{(2)}, \lambda_2^{(3)}) \tag{8}$$

$$\mathcal{H}_3 :\; x_3 = 0 \;\Rightarrow\; \mathbf{e}_3 \notin \mathcal{H}_3 \;\Rightarrow\; n_1 + n_3 + n_5 + n_7 \geq 2 = \max(\lambda_3^{(1)}, \lambda_3^{(2)}, \lambda_3^{(3)}) \tag{9}$$

$$\mathcal{H}_4 :\; x_1 + x_2 = 0 \;\Rightarrow\; \mathbf{e}_1, \mathbf{e}_2 \notin \mathcal{H}_4 \;\Rightarrow\; n_2 + n_3 + n_4 + n_5 \geq 3 = \max\Big(\sum_{j=1,2} \lambda_j^{(i)} : i \in [3]\Big) \tag{10}$$

$$\mathcal{H}_5 :\; x_1 + x_3 = 0 \;\Rightarrow\; \mathbf{e}_1, \mathbf{e}_3 \notin \mathcal{H}_5 \;\Rightarrow\; n_1 + n_3 + n_4 + n_6 \geq 3 = \max\Big(\sum_{j=1,3} \lambda_j^{(i)} : i \in [3]\Big) \tag{11}$$

$$\mathcal{H}_6 :\; x_2 + x_3 = 0 \;\Rightarrow\; \mathbf{e}_2, \mathbf{e}_3 \notin \mathcal{H}_6 \;\Rightarrow\; n_1 + n_2 + n_5 + n_6 \geq 3 = \max\Big(\sum_{j=2,3} \lambda_j^{(i)} : i \in [3]\Big) \tag{12}$$

$$\mathcal{H}_7 : x_1 + x_2 + x_3 = 0 \;\Rightarrow\; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \notin \mathcal{H}_7 \;\Rightarrow\; n_1 + n_2 + n_4 + n_7 \geq 4 = \max\Big(\sum_{j=[3]} \lambda_j^{(i)} : i \in [3]\Big) \tag{13}$$

Summing up inequalities (7)-(13) and dividing by four gives $n \geq \left\lceil \frac{19}{4} \right\rceil = 5$. Indeed, the LP relaxation of the ILP from Corollary 1.10 has an optimal solution $n_1 = n_2 = n_4 = \frac{5}{4}$, $n_3 = n_5 = n_6 = n_7 = \frac{1}{4}$ with target value $\frac{19}{4}$. Next we will show $n \geq 6$ for the ILP and assume that there exists an integral solution with $n = 5$. Summing the inequalities over all hyperplanes $\mathcal{H}_i$ containing $v_1 = \mathbf{e}_3$, i.e., (7), (8), and (10), and dividing by two gives $\sum_{j \in [l] \setminus \{1\}} n_j \geq 3.5$, so that $n_1 \leq 1$. By symmetry, we also conclude $n_2, n_4 \leq 1$. Summing the inequalities over all hyperplanes $\mathcal{H}_i$ not containing $v_1 = \mathbf{e}_3$, i.e., (9), (11), (12), and (13), and dividing by two gives $2n_1 + \sum_{j \in [l] \setminus \{1\}} n_j \geq 6$, so that $n_1 \geq 1$. Thus, $n_1 = 1$ and, by symmetry, also $n_2 = n_4 = 1$. Summing inequalities (10)-(12), plugging in the known values, and dividing by two gives $n_3 + n_5 + n_6 \geq 1.5$, so that $n_7 \leq 0.5$, i.e., $n_7 = 0$. However, this contradicts Inequality (13).

An integral solution for $n = 6$ can indeed be attained by $n_1 = n_2 = n_4 = 2, n_3 = n_5 = n_6 = n_7 = 0$. It can be easily checked that the corresponding generator matrix

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

satisfies $\mathcal{S}(\mathbf{G}) \supseteq \mathcal{R}$.

In Proposition 2.14 we will give a general result that directly yields $n(\mathcal{R}) \geq 6$ for Example 1.13.

2. Bounds for $n(T)$

Let $T\colon 2^{[2]} \to \mathbb{N}$ be monotone, subadditive, and satisfy $T(\emptyset) = 0$, i.e., by Inequality (3)

$$\max\{T(\{1\}), T(\{2\})\} \quad \le \quad T(\{1,2\}) \quad \le \quad T(\{1\}) + T(\{2\}).$$

The corresponding generating sets of $\mathcal{R}(T)$ can be easily described:

Lemma 2.1. *If $T\colon 2^{[2]} \to \mathbb{N}$ is monotone, subadditive, and satisfies $T(\emptyset) = 0$, then a generating set of $\mathcal{R}(T)$ is given by*

$$S = \Big\{ \Big(T(\{1\}), T(\{1,2\}) - T(\{1\})\Big), \Big(T(\{1,2\}) - T(\{2\}), T(\{2\})\Big) \Big\}.$$

Proof. First, we check that each $\lambda \in S$ satisfies the constraints $\lambda_1 \le T(\{1\})$, $\lambda_2 \le T(\{2\})$, and $\lambda_1 + \lambda_2 \le T(\{1,2\})$.

For the other direction let $\lambda \in \mathbb{R}^2_{\ge 0}$ satisfying the constraints $\lambda_1 \le T(\{1\})$, $\lambda_2 \le T(\{2\})$, and $\lambda_1 + \lambda_2 \le T(\{1,2\})$. W.l.o.g. we assume that at least one of these three inequalities is satisfied with equality, since we could increase λ otherwise. If $\lambda_1 + \lambda_2 = T(\{1,2\})$, then $\lambda \in \operatorname{conv}(S)$ since $\lambda_1 \le T(\{1\})$ and $\lambda_2 \le T(\{2\})$. So let us now consider the case $\lambda_1 = T(\{1\})$. If $\lambda_2 < T(\{2\})$ and $\lambda_1 + \lambda_2 < T(\{1,2\})$ then we could increase λ, so that we can assume $\lambda_2 < T(\{2\})$ and conclude $\lambda_1 + \lambda_2 = T(\{1,2\})$ from the subadditivity of T. The case $\lambda_2 = T(\{2\})$ can be treated analogously. □

We remark that the generating set of Lemma 2.1 has cardinality 2 or 1, where the latter happens if the two vectors coincide, which happens iff $T(\{1\}) = T(\{2\})$ and $T(\{1,2\}) = 2T(\{1\})$.

Definition 2.2. For a set $\emptyset \ne S \subset \mathbb{N}_{>0}$ of positive integers we denote by $\operatorname{Simpl}(S)$ the set of non-zero vectors in $\langle \{\mathbf{e}_i \mid i \in S\} \rangle$ over $\mathbb{F}_2$.

We remark that Definition 2.2 defines binary simplex codes, which can be easily generalized to arbitrary finite fields $\mathbb{F}_q$.

The following is well-known:

Lemma 2.3. *For each $\emptyset \ne S \subseteq [k] \subset \mathbb{N}_{>0}$ we have $\#\operatorname{Simpl}(S) = 2^s - 1$ and $\mathcal{S}(\operatorname{Simpl}(S)) = \mathcal{R}(T)$, where $s = \#S$ and $T\colon 2^{[k]} \to \mathbb{N}$ is given by $T(U) = 2^{s-1}$ if $U \cap S \ne \emptyset$ and $T(U) = 0$ otherwise (for all $U \subseteq [k]$).*

As an abbreviate we write $\frac{1}{2} \cdot \operatorname{Simpl}(\{i,j\}) = \{\mathbf{e}_i, \mathbf{e}_j\}$ for two different positive integers i and j. Note that the cardinality is $\left\lceil \frac{1}{2} \cdot \#\operatorname{Simpl}(\{i,j\}) \right\rceil = 2$ and the service rate region contains the service rate region of $\operatorname{Simpl}(\{i,j\})$ scaled by a factor of $\frac{1}{2}$, i.e.,

$$\mathcal{S}(\{\mathbf{e}_i, \mathbf{e}_j\}) = \left\{ \lambda \in \mathbb{R}^k_{\ge 0} \mid \lambda_i \le 1, \lambda_j \le 1 \right\} \supsetneq \left\{ \lambda \in \mathbb{R}^k_{\ge 0} \mid \lambda_i \le 1, \lambda_j \le 1, \lambda_i + \lambda_j \le 1 \right\}.$$

Theorem 2.4. *For the service rate region*

$$\mathcal{R} = \Big\{ \lambda \in \mathbb{R}^2_{\ge 0} \,:\, \lambda_1 \le X, \lambda_2 \le Y, \lambda_1 + \lambda_2 \le \Sigma \Big\},$$

where X, Y, Σ are non-negative integers with $\max\{X, Y\} \le \Sigma \le X + Y$, we have $n(\mathcal{R}) = \left\lceil \frac{X+Y+\Sigma}{2} \right\rceil$.

Proof. Note that the condition of Inequality (3) is satisfied, so that we can apply Lemma 2.1. The inequalities from Lemma 1.9 read

$$\begin{aligned} n_1 + n_3 &\ge X = \max\{X, \Sigma - Y\}, \\ n_2 + n_3 &\ge Y = \max\{Y, \Sigma - X\}, \\ n_1 + n_2 &\ge \Sigma = \max\{\Sigma, \Sigma\}, \end{aligned}$$

so that summing up and dividing by two gives

$$n = n_1 + n_2 + n_3 \ge \frac{X + Y + \Sigma}{2}.$$

Since n is an integer, we obtain $n(\mathcal{R}) \ge \left\lceil \frac{X+Y+\Sigma}{2} \right\rceil$.

For the upper bound on $n(\mathcal{R})$, i.e., the constructive part, we let $\mathcal{G}^{(1)}$ consist of $\Sigma - Y$ copies of $\operatorname{Simpl}(\{1\})$, $\mathcal{G}^{(2)}$ consist of $\Sigma - X$ copies of $\operatorname{Simpl}(\{2\})$, and $\mathcal{G}^{(3)}$ consist of $\frac{X+Y-\Sigma}{2}$ copies of $\operatorname{Simpl}(\{1,2\})$. Now, we set $\mathcal{G} = \cup_{i\in[3]}\mathcal{G}^{(i)}$, which is a multiset of point in $\operatorname{PG}(2-1,2)$ of cardinality

$$(\Sigma - Y) + (\Sigma - X) + \left\lceil \frac{3(X+Y-\Sigma)}{2} \right\rceil = \left\lceil \frac{X+Y+\Sigma}{2} \right\rceil .$$

By construction we have $\mathcal{R}(\mathcal{G}) \supseteq \mathcal{R}(T)$ for $T(\emptyset) = 0$, $T(\{1\}) = X$, $T(\{2\}) = Y$, and $T(\{1,2\}) = \Sigma$. □

Of course, we might give a more direct proof of Theorem 2.4. Instead of basing the constructive part on the "subcodes" introduced in Definition 2.2 we can directly write down the multiplicities for the columns $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_1 + \mathbf{e}_2$.

Conjecture 2.5. *$n(\mathcal{R}(T))$ can be attained by a union of* $\operatorname{Simpl}\big(S^{(i)}\big)$.

Lemma 2.6. *If $T\colon 2^{[3]} \to \mathbb{N}$ is monotone, subadditive, satisfies $T(\emptyset) = 0$ and $T([3]) + T(\{i\}) \le \sum_{j\in[3]\setminus\{i\}} T(\{i,j\})$, then a generating set of $\mathcal{R}(T)$ is given by*

$$S = \Big\{ \Big(T(\pi_{\le 1}) - T(\pi_{<1}), T(\pi_{\le 2}) - T(\pi_{<2}), T(\pi_{\le 1}) - T(\pi_{<3})\Big) \mid \pi \text{ is a bijection } [3] \to [3] \Big\},$$

where $\pi_{\le i} = \{j \in [k] \mid \pi(j) \le \pi(i)\}$ and $\pi_{<i} = \{j \in [k] \mid \pi(j) < \pi(i)\}$.

Proof. First, we need to check that each $\lambda \in S$ satisfies the constraints $\sum_{i\in U} \lambda_i \le T(U)$ for all $U \subseteq [3]$. In explicit form the elements of S in Lemma 2.6 are given by

$$\begin{aligned}
&\Big(T(\{1\}), T(\{1,2\}) - T(\{1\}), T(\{1,2,3\}) - T(\{1,2\})\Big),\\
&\Big(T(\{1\}), T(\{1,2,3\}) - T(\{1,3\}), T(\{1,3\}) - T(\{1\})\Big),\\
&\Big(T(\{1,2\}) - T(\{2\}), T(\{2\}), T(\{1,2,3\}) - T(\{1,2\})\Big),\\
&\Big(T(\{1,2,3\}) - T(\{2,3\}), T(\{2\}), T(\{2,3\} - T(\{2\}))\Big),\\
&\Big(T(\{1,3\}) - T(\{3\}), T(\{1,2,3\}) - T(\{1,3\}), T(\{3\})\Big), \text{ and}\\
&\Big(T(\{1,2,3\}) - T(\{2,3\}), T(\{2,3\}) - T(\{3\}), T(\{3\})\Big).
\end{aligned}$$

It can be easily verified that under the conditions of Lemma 2.6, each $\lambda \in S$ satisfies the constraints $\sum_{i\in U} \lambda_i \le T(U)$ for all $U \subseteq [3]$. For the other direction, we need to show that each $\lambda \in \mathbb{R}^3_{\ge 0}$ satisfying the constraints $\sum_{i\in U} \lambda_i \le T(U)$ for all $U \subseteq [3]$, is in $\operatorname{conv}(S)\!\downarrow$. The proof is similar to the proof of Lemma 2.1. W.l.o.g. we assume that at least one of these seven inequalities is satisfied with equality, since we could increase λ otherwise. If $\lambda_1 + \lambda_2 + \lambda_3 = T(\{1,2,3\})$, then $\lambda \in \operatorname{conv}(S)$ since all other six inequalities are satisfied. So now let us consider the case $\lambda_1 = T(\{1\})$. If $\lambda_2 < T(\{2\})$, $\lambda_3 < T(\{3\})$, $\lambda_1 + \lambda_2 < T(\{1,2\})$, $\lambda_1 + \lambda_3 < T(\{1,3\})$, $\lambda_2 + \lambda_3 < T(\{2,3\})$, and $\lambda_1 + \lambda_2 + \lambda_2 < T(\{1,2,3\})$, then we could increase λ, so that $\lambda_1 + \lambda_2 = T(\{1,2\})$ and $\lambda_1 + \lambda_2 + \lambda_3 = T(\{1,2,3\})$ from the subadditivity and $T([3]) + T(\{i\}) \le \sum_{j\in[3]\setminus\{i\}} T(\{i,j\})$ properties of T. All other cases can be treated analogously. □

The notion of Lemma 2.6 can also be used to characterize the generating set in Lemma 2.1. In explicit form the elements of S in Lemma 2.6 are given by

$$\begin{aligned}
&\Big(T(\{1\}),T(\{1,2\})-T(\{1\}),T(\{1,2,3\})-T(\{1,2\})\Big),\\
&\Big(T(\{1\}),T(\{1,2,3\})-T(\{1,3\}),T(\{1,3\})-T(\{1\})\Big),\\
&\Big(T(\{1,2\})-T(\{2\}),T(\{2\}),T(\{1,2,3\})-T(\{1,2\})\Big),\\
&\Big(T(\{1,2,3\})-T(\{2,3\}),T(\{2\}),T(\{2,3\}-T(\{2\}))\Big),\\
&\Big(T(\{1,3\})-T(\{1\}),T(\{1,2,3\})-T(\{1,3\}),T(\{3\})\Big),\text{ and}\\
&\Big(T(\{1,2,3\})-T(\{2,3\}),T(\{2,3\})-T(\{3\}),T(\{3\})\Big).
\end{aligned}$$

Under the conditions of Lemma 2.6 none of the constraints $\sum_{i\in U}\lambda_i\le T(U)$, for $\emptyset\neq U\subseteq[3]$, is strictly redundant, i.e., each inequality can be attained with equality by some $\lambda\in\mathbb{R}^3_{\ge 0}$ without violating one of the other constraints.

Example 2.7. For $k=3$ the function $T\colon 2^{[k]}\to\mathbb{N}$, given by

$$T(U)=\left\{\begin{array}{rcl} 0 & : & \#U=0,\\ 4 & : & \#U=1,\\ 5 & : & \#U=2,\\ 7 & : & \#U=3\end{array}\right. ,$$

is monotone and subadditive but does not satisfy the last condition of Lemma 2.6. A generating set of $\mathcal{R}(T)$ is given by

$$\Big\{(4,1,1),(1,4,1),(1,1,4),(3,2,2),(2,3,2),(2,2,3)\Big\}.$$

We have $n(\mathcal{R}(T))=9$.

Lemma 2.6 can be generalized in the sense that we can characterize some elements of $\mathcal{R}(T)$ at the very least.

Lemma 2.8. *If $T\colon 2^{[k]}\to\mathbb{N}$ is monotone and satisfies $T(\emptyset)=0$ for some positive integer k, then $\mathcal{R}(T)$ contains the vector x^π for every bijection π of $[k]$, where the components of x^π can be computed recursively in the ordering of π:*

$$x_i^\pi=\min\left\{T(U\cup\{i\})-\sum_{j\in U}x_j\mid U\subseteq\pi_{<i}\right\},$$

where $\pi_{<i}=\{j\in[k]\mid\pi(j)<\pi(i)\}$.

Proof. Directly from the definition of the x_i^π and the ordering π of the evaluation we conclude that the x_i^π are uniquely defined. Next we want to show $x^\pi\ge\mathbf{0}$. So, assume to the contrary that i is the with respect to π earliest index in $[k]$ with $x_i^\pi<0$. Now let $U\subseteq\pi_{<0}$ a subset with $x_i^\pi=T(U\cup\{i\})-\sum_{j\in U}x_j$. By construction we have $\sum_{j\in U}x_j\le T(U)$, so that monotonicity of T, i.e., $T(U\cup\{i\})\ge T(U)$, yields $x_i^\pi\ge 0$, which is a contradiction. Finally we show that $\sum_{j\in U}x_j\le T(U)$ for all $\emptyset\neq U\subseteq[k]$. So, let such a subset U be given and let i be the with respect to π latest element in U. By construction we have

$$x_i^\pi\le T(U'\cup\{i\})-\sum_{j\in U'}x_j,$$

where $U'=U\setminus\{i\}$, so that $\sum_{j\in U}x_j\le T(U)$. □

If $T\colon 2^{[3]}\to\mathbb{N}$ is monotone, subadditive and satisfies $T(\emptyset)=0$, then the formula for x^π of Lemma 2.8 can be simplified to

$$\begin{aligned} x^\pi_{\pi(1)} &= T(\{\pi(1)\}),\\ x^\pi_{\pi(2)} &= T(\{\pi(1),\pi(2)\})-T(\{\pi(1)\}),\ \text{and}\\ x^\pi_{\pi(3)} &= \min\left\{T([3])-T(\{\pi(1),\pi(2)\}),T(\{\pi(1),\pi(3)\})-T(\{\pi(1)\})\right\}. \end{aligned}$$

Example 2.9. Let $k=3$ and $T\colon 2^{[k]}\to\mathbb{N}$ by defined by $T(\emptyset)=0$, $T(\{1\})=13$, $T(\{2\})=14$, $T(\{5\})=15$, $T(\{1,2\})=18$, $T(\{1,3\})=21$, $T(\{2,3\})=22$, and $T(\{1,2,3\})=30$. From Lemma 2.8 we conclude

$$\Big\{(13,5,8),(4,14,8),(6,7,15)\Big\}\subseteq\mathcal{R}(T).$$

We can easily check that also $(9,9,12)\in\mathcal{R}(T)$, while $(9,9,12)\notin\operatorname{conv}(\{(13,5,8),(4,14,8),(6,7,15))\})$ since

$$13a+4b+6c \geq 9 \tag{14}$$
$$5a+14b+7c \geq 9 \tag{15}$$
$$8a+8b+15c \geq 12 \tag{16}$$

with $a,b,c\in\mathbb{R}_{\geq 0}$ and $a+b+c=1$ implies $a+b\geq 1$ (summing the first two inequalities), so that $c=0$, which contradicts the last inequality. A generating set of $\mathcal{R}(T)$ is given by

$$\Big\{(13,5,8),(4,14,8),(6,7,15),(9,9,12),(8,10,12),(8,9,13)\Big\}$$

as we will see in the subsequent lemma. The ILP of Corollary 1.10 has an optimal solution $n_1=10$, $n_2=9$, $n_3=1$, $n_4=8$, $n_5=1$, $n_6=2$, and $n_7=3$, so that $n(\mathcal{R}(T))\geq 34$. We remark that Proposition 2.12 gives $n(\mathcal{R}(T))\geq\lceil 33.25\rceil=34$.

Proposition 2.10. *Let $T\colon 2^{[3]}\to\mathbb{N}$ with $T(\emptyset)=0$ and none of the constraints $\sum_{i\in U}x_i\leq T(U)$ is strictly redundant in $\mathbb{R}^k_{\geq 0}$ for $\emptyset\neq U\subseteq[3]$. Then, the following list of vectors gives a generating set of $\mathcal{R}(T)$:*

- $\Gamma_1=\Big(T(1),T(12)-T(1),\min\big\{T(123)-T(12),T(13)-T(1)\big\}\Big)$;
- $\Gamma_2=\Big(T(1),\min\big\{T(123)-T(13),T(12)-T(1)\big\},T(13)-T(1)\Big)$;
- $\Gamma_3=\Big(T(12)-T(2),T(2),\min\big\{T(123)-T(12),T(23)-T2)\big\}\Big)$;
- $\Gamma_4=\Big(\min\big\{T(123)-T(23),T(12)-T(2)\big\},T(2),T(23)-T(2)\Big)$;
- $\Gamma_5=\Big(T(13)-T(3),\min\big\{T(123)-T(13),T(23)-T(3)\big\},T(3)\Big)$;
- $\Gamma_6=\Big(\min\big\{T(123)-T(23),T(13)-T(3)\big\},T(23)-T(3),T(3)\Big)$;
- $\Gamma_7=\Big(T(12)+T(13)-T(123),T(123)-T(13),T(123)-T(12)\Big)$ *if* $T(12)+T(13)\leq T(123)+T(1)$;
- $\Gamma_8=\Big(T(123)-T(23),T(12)+T(23)-T(123),T(123)-T(12)\Big)$ *if* $T(12)+T(23)\leq T(123)+T(2)$;
- $\Gamma_9=\Big(T(123)-T(23),T(123)-T(13),T(13)+T(23)-T(123)\Big)$ *if* $T(13)+T(23)\leq T(123)+T(3)$.

Proof. Due to our assumption T is monotone and subadditive. The first six vectors of our list are contained in $\mathcal{R}(T)$ due to Lemma 2.8. Given the assumption $T(12)+T(13)\leq T(123)+T(1)$ we have for $(x_1,x_2,x_3)=\Gamma_7$ that $x_1+x_2=T(12)$, $x_1+x_3=T(13)$, and $x_1+x_2+x_3=T(123)$. The condition $x_2 + x_3 \le T(23)$ follows from $2T(123) \le T(12) + T(13) + T(23)$ since $x_2 + x_3 = 2T(123) - T(12) - T(13)$. The conditions $x_2 \le T(2)$ and $x_3 \le T(3)$ follow from the subadditivity of T and the condition $x_1 \le T(1)$ is equivalent to the assumption. Thus, given the assumption, $\Gamma_7 \in \mathcal{R}(T)$. By symmetry, we have the analogues statement for Γ_8 and Γ_9. The polytope $\mathcal{R}(T)$ is described by four types of inequalities:

(i) $x_1 \ge 0, x_2 \ge 0, x_3 \ge 0$;
(ii) $x_1 \le T(1), x_2 \le T(2), x_3 \le T(3)$;
(iii) $x_1 + x_2 \le T(12), x_1 + x_3 \le T(13), x_2 + x_3 \le T(23)$;
(iv) $x_1 + x_2 + x_3 \le T(123)$.

Since we assume that no constraint is strictly redundant the vertices of $\mathcal{R}(T)$ are given by each triple of linear independent inequalities, i.e., the coefficient vectors of the inequalities are linearly independent. Next we will check all possible cases taking the symmetry of the symmetric group on 3 elements into account. Note that $\Gamma_1, \dots, \Gamma_6$ form indeed an orbit under this group action. The vectors $\Gamma_7, \dots, \Gamma_9$ form another orbit under this group action.

Three times type (i) gives the vertex $(0, 0, 0) \le \Gamma_1$. If type (i) occurs two times, then we assume $x_1 = x_2 = 0$ w.l.o.g. Since $\min\{T(3), T(13), T(23), T(123)\} = T(3)$ the corresponding vertex is given by $(0, 0, T(3)) \le \Gamma_5$. If type (i) occurs exactly one time, then we assume $x_3 = 0$ w.l.o.g. Moreover for the other two inequalities we only need to consider those which do not involve x_3. If the other two are $x_1 \le T(1)$ and $x_2 \le T(2)$, then the corresponding vertex is given by $(T(1), T(2), 0) \le \Gamma_1$ since T is subadditive, i.e., $T(12) - T(1) \le T(2)$. If the other two are $x_1 \le T(1)$ and $x_1 + x_2 \le T(12)$, then the corresponding vertex is $(T(1), T(12) - T(1), 0) \le \Gamma_1$. If the other two are $x_2 \le T(2)$ and $x_1 + x_2 \le T(12)$, then the corresponding vertex is $(T(12) - T(2), T(2), 0) \le \Gamma_2$. Thus, in the following we can assume that type (i) does not occur at all.

If type (ii) is attained at least once then we assume $x_1 = T(1)$ w.l.o.g. If either $x_3 \le T(3)$ or $x_1 + x_3 \le T(13)$ occurs then we assume w.l.o.g. that $x_2 \le T(2)$ or $x_1 + x_2 \le T(12)$. Due to subadditivity of T we then have $x_1 + x_2 = T(12)$, i.e., $x_2 = T(12) - T(1)$. For x_3 we then have

$$x_3 = \min\Big\{T(3), T(13) - T(1), T(123) - T(12), T(23) - T(12) + T(1)\Big\},$$

so that the corresponding vertex equals Γ_1. Otherwise, neither $x_2 = T(2)$, $x_3 = T(3)$, $x_1 + x_3 = T(13)$, nor $x_2 + x_3 = T(23)$ are valid. The only two remaining possibilities are $x_2 + x_3 = T(23)$ and $x_1 + x_2 + x_3 = T(123)$, which, however, are linearly dependent.

In the remaining cases types (i) and (ii) do not occur at all. If type (iii) is attained three times, then we have $x_1 = \big(T(12) + T(13) - T(23)\big)/2$, $x_2 = \big(T(12) + T(23) - T(13)\big)/2$, and $x_3 = \big(T(13) + T(23) - T(12)\big)/2$, so that $x_1 + x_2 + x_3 = \big(T(12) + T(13) + T(23)\big)/2$. Since the inequalities $x_1 + x_2 \le T(12)$, $x_1 + x_3 \le T(13)$, and $x_2 + x_3 \le T(23)$ imply $T(123) \le \big(T(12) + T(13) + T(23)\big)/2$, so that $T(123) = \big(T(12) + T(13) + T(23)\big)/2$. In this situation our vector (x_1, x_2, x_3) equals $\Gamma_7 = \Gamma_8 = \Gamma_9$.

Next, we assume $x_1 + x_2 = T(12)$, $x_1 + x_3 = T(13)$, and $x_1 + x_2 + x_3 = T(123)$, i.e., $x_3 = T(123) - T(12)$, $x_2 = T(123) - T(13)$, and $x_1 = T(12) + T(13) - T(123)$. So, the corresponding vertex equals Γ_7. Due to our symmetry assumptions we also have to consider Γ_8 and Γ_9. □

While it might be hard to give explicit formulas for the generating set of $\mathcal{R}(T)$,[1] we can easily generalize the lower bound of Theorem 2.4 if we assume that none of the constraints is (strictly) redundant.

Definition 2.11. Let $P = \big\{x \in \mathbb{R}^k \mid Ax \le b, x \ge 0\big\}$ be a polyhedron in $\mathbb{R}^k$ with description (A, b). We say that a constraint $a^{(i)}x \le b_i$ is redundant, where $a^{(i)}$ denotes the ith row of A, if $P = \big\{x \in \mathbb{R}^k \mid A'x \le b', x \ge 0\big\}$,

[1] In general a polytope $\{x \in \mathbb{R}^n : Ax \ge b\}$ described by m "$\ge$"-inequalities has at most $\binom{m - \lfloor \frac{n+1}{2} \rfloor}{m-n} + \binom{m - \lceil \frac{n+1}{2} \rceil}{m-n}$ extreme points [10], which are vertices in the case of a polytope. This upper bound is attained by the so-called cyclic polytopes, see [4]. If the entries of A are all contained in $\{0, 1\}$, then there is an upper of $n!$, see [3].

where A' and b' arise from A and b by removing the ith row, respectively. We say that a constraint $a^{(i)}x \le b_i$ is strictly redundant if there does not exist $\bar{x} \in P$ with $a^{(i)}\bar{x} = b_i$.

As an example consider $T\colon 2^{[2]} \to \mathbb{N}$ defined via $T(\emptyset) = 0$ and $T(\{1\}) = T(\{2\}) = T(\{1,2\}) = 1$. With this we consider the polyhedron in $\mathbb{R}^2_{\ge 0}$ defined by the inequalities $\sum_{i\in U} \lambda_i \le T(U)$ (choosing λ as variable) for all $\emptyset \ne U \subseteq [2]$. Since the vectors $(1,0)$ and $(0,1)$ are contained in the polyhedron, no inequality is strictly redundant. The inqualities $\lambda_1 \le T(\{1\})$ and $\lambda_2 \le T(\{2\})$ are redundant, while the inequality $\lambda_1 + \lambda_2 \le T(\{1,2\})$ is not redundant since e.g. $(1,1)$ is not contained in the polyhedron.

Proposition 2.12. *We have*

$$n(\mathcal{R}(T)) \ge \left\lceil \frac{\sum_{\emptyset \ne U \subseteq [k]} T(U)}{2^{k-1}} \right\rceil,$$

where $T\colon 2^{[k]} \to \mathbb{N}$ for some positive integer k and none of the constraints $\sum_{i\in U} \lambda_i \le T(U)$ is strictly redundant in $\mathbb{R}^k_{\ge 0}$.

Proof. We want to apply the ILP formulation of Corollary 1.10. First we observe that each hyperplane $\mathcal{H}$ in $\operatorname{PG}(k-1,2)$ can be uniquely characterized by a set $\emptyset \ne S \subseteq [k]$ such that $\{i \in [k] \mid \mathbf{e}_i \notin \mathcal{H}\} = S$. So, we will write $S(\mathcal{H})$ for this set S in the following. Due to our assumption that no constraint (of the ILP formulation of Corollary 1.10) is strictly redundant, we can choose $T(S)$ as right hand side, i.e.,

$$\sum_{j\in[l]\,:\,v_j\notin\mathcal{H}} n_j \ge T(S(\mathcal{H})),$$

where $S(\mathcal{H}) = \{i \in [k] : \mathbf{e}_i \notin \mathcal{H}\}$, as described above. For each $j \in [l]$ we have $v_j \notin \mathcal{H}$ for exactly 2^{k-1} hyperplanes $\mathcal{H}$. Thus, summing all of the above $2^k - 1$ inequalities and dividing by 2^{k-1} yields

$$n = \sum_{j\in[l]} n_j \ge \frac{\sum_{\emptyset\ne U\subseteq[k]} T(U)}{2^{k-1}}.$$

Finally, we observe that n has to be an integer. □

We remark that the lower bound of Proposition 2.12 is indeed tight (if $k = 2$ and T is monotone and subadditive), see Theorem 2.4. However, it is not tight in general, e.g., in Example 1.13 $n(\mathcal{R}(T))$ is one larger than the corresponding lower bound of Proposition 2.12, while none of the constraints strictly redundant.

Corollary 2.13. *We have*

$$n(\mathcal{R}(T)) \ge \left\lceil \frac{X \cdot \left(2^k - 1\right)}{2^{k-1}} \right\rceil,$$

where $T\colon 2^{[k]} \to \mathbb{N}$ for some positive integer k, $X \in \mathbb{N}$, $T(\emptyset) = 0$, and $T(U) = X$ for all $\emptyset \ne U \subseteq [k]$. redundant in $\mathbb{R}^k_{\ge 0}$. Moreover, if $X = t \cdot 2^{k-1}$ for some integer t, then the lower bound is tight.

Proof. We can easily check that none of the constraints is strictly redundant, so that we can apply Proposition 2.12. Indeed, a generating set of $\mathcal{R}(T)$ is given by $\{X \cdot \mathbf{e}_i \mid i \in [k]\}$. A t-fold k-dimensional binary simplex code $\operatorname{Simpl}([k])$ achieves the desired service rate region. □

We remark that the situation of $T(U) = X \in \mathbb{N}$ for all $\emptyset \ne U \subseteq [k]$ is equivalent to the situation of PIR codes, see e.g. [9] for some recent lower bounds.

In the light of Example 1.13 we want to give further general lower bounds similar to the bound of Proposition 2.12.

Proposition 2.14. *For some positive integer $k \ge 2$ let $T\colon 2^{[k]} \to \mathbb{N}$ be a function such that none of the constraints $\sum_{i\in U} \lambda_i \le T(U)$ is strictly redundant in $\mathbb{R}^k_{\ge 0}$. For each $i \in [k]$ we have $n(\mathcal{R}(T)) \ge \left\lceil \frac{\alpha_i+\beta_i}{2} \right\rceil$ where*

$$\alpha_i = \left\lceil \frac{\sum_{\emptyset\neq U\subseteq [k]\setminus\{i\}} T(U)}{2^{k-2}} \right\rceil$$

and

$$\beta_i = \left\lceil \frac{\sum_{\{i\}\subseteq U\subseteq [k]} T(U)}{2^{k-2}} \right\rceil.$$

Proof. We proceed similar as in the proof of Proposition 2.12 and utilize

$$\sum_{j\in[l]\,:\,v_j\notin\mathcal{H}} n_j \ge T(S(\mathcal{H})), \tag{17}$$

where $S(\mathcal{H}) = \{i \in [k] \,:\, \mathbf{e}_i \notin \mathcal{H}\}$. We can also parameterize those constraints by subsets $\emptyset \neq U \subseteq [k]$ by uniquely characterizing $\mathcal{H}$ by $S(\mathcal{H}) = U$. Now let $i \in [k]$ be arbitrary but fix and $\bar{i} = 2^{k-i}$, so that $v_{\bar{i}} = \mathbf{e}_i$. Summing Inequality (17) for all $\emptyset \neq U \subseteq [k] \setminus \{i\}$ gives

$$2^{k-2} \cdot \sum_{j\in[l]\setminus\{\bar{i}\}} n_j \ge \sum_{\emptyset\neq U\subseteq[k]\setminus\{i\}} T(U).$$

Summing Inequality (17) for all $\{i\} \subseteq U \subseteq [k]$ gives

$$2^{k-1} \cdot n_{\bar{i}} + 2^{k-2} \cdot \sum_{j\in[l]\setminus\{\bar{i}\}} n_j \ge \sum_{\{i\}\subseteq U\subseteq[k]} T(U).$$

Since the n_j are integers we have $\sum_{j\in[l]\setminus\{\bar{i}\}} n_j \ge \alpha_i$ and $2n_{\bar{i}} + \sum_{j\in[l]\setminus\{\bar{i}\}} n_j \ge \beta_i$. Dividing the sum of these two inequalities by 2 gives

$$n = n_{\bar{i}} + \sum_{j\in[l]\setminus\{\bar{i}\}} n_j \ge \frac{\alpha_i+\beta_i}{2},$$

where we again can upround the right hand side since n is an integer. □

We remark that Proposition 2.14 implies Proposition 2.12 (for $k \ge 2$). However, for Example 1.13 also Proposition 2.14 implies only $n(\mathcal{R}) \ge 5$ since we have $\alpha_i = 4$ and $\beta_i = 6$ for all $i \in [3]$. We remark that Proposition 3.10 gives the tight lower bound $n(\mathcal{R}) \ge 6$.

Proposition 2.15. *For some positive integer $k \ge 2$ let $T\colon 2^{[k]} \to \mathbb{N}$ be a function such that none of the constraints $\sum_{i\in U} \lambda_i \le T(U)$ is strictly redundant in $\mathbb{R}^k_{\ge 0}$. For each $j \in [l]$ we have*

$$n(\mathcal{R}(T)) \ge \left\lceil \frac{\sum_{\emptyset\neq U\subseteq[k]\,:\,\#(U\cap J)\equiv 0 \pmod 2} T(U)}{2^{k-2}} \right\rceil,$$

where $J \subseteq [k]$ such that $v_j = \sum_{h\in J} \mathbf{e}_h$.

Proof. We proceed similar as in the proof of Proposition 2.12 and utilize

$$\sum_{j\in[l]\,:\,v_j\notin\mathcal{H}} n_j \ge T(S(\mathcal{H})), \tag{18}$$

where $S(\mathcal{H}) = \{i \in [k] \,:\, \mathbf{e}_i \notin \mathcal{H}\}$. We can also parameterize those constraints by subsets $\emptyset \neq U \subseteq [k]$ by uniquely characterizing $\mathcal{H}$ by $S(\mathcal{H}) = U$. Now let $j \in [l]$ by arbitrary but fix. Our aim is to sum

Inequality (18) over all $2^{k-1}-1$ hyperplanes $\mathcal{H}$ that contain v_j. We claim that $v_j=\sum_{h\in J}\mathbf{e}_h\in\mathcal{H}$ iff $\#(U\cap J)\equiv 0 \pmod 2$, where $U=S(\mathcal{H})$. If $\#U\ge 2$, then for some arbitrary element $x\in U$ the set

$$\{\mathbf{e}_i\mid i\in[k]\backslash U\}\cup\{\mathbf{e}_x+\mathbf{e}_i\mid i\in U\setminus x\}$$

is a basis of $\mathcal{H}$. So, we have $v_j=\sum_{h\in J}\mathbf{e}_h\in\mathcal{H}$ iff $\#(U\cap J)\equiv 0 \pmod 2$. In the remaining cases we have $\#U=1$ and choose $x\in[k]$ such that $U=\{x\}$. A basis of $\mathcal{H}$ is given by $\{\mathbf{e}_h\mid h\in[k]\setminus\{x\}\}$. So, $v_j\in\mathcal{H}$ iff $x\notin J$, i.e., $\#(U\cap J)=0\equiv \pmod 2$. Thus, we obtain

$$2^{k-2}\cdot\sum_{a\in[l]\backslash\{j\}}n_a=\sum_{\text{hyperplane }\mathcal{H}\,:\,v_j\in\mathcal{H}}\ \sum_{a\,:\,v_a\notin\mathcal{H}}n_a\ge\sum_{\emptyset\neq U\subseteq[k]\,:\,\#(U\cap J)\equiv 0 \pmod 2}T(U).$$

Since $\sum_{a\in[l]\backslash\{j\}}n_a$ is not larger than n and an integer, we obtain the stated lower bound. □

Example 2.16. Let x be a positive integer and $T\colon 2^{[3]}\to\mathbb{N}$ be defined by $T(\{1\})=T(\{2\})=T(\{3\})=T(\{1,2\})=T(\{1,3\})=x$ and $T(\{2,3\})=T(\{1,2,3\})=2x$. Proposition 2.12 gives $n(\mathcal{R}(T))\ge\left\lceil\frac{9x}{4}\right\rceil$. For $j=3$ Proposition 2.15 gives $n(\mathcal{R}(T))\ge\left\lceil\frac{5x}{2}\right\rceil$.

So, there for $k=3$ (and indeed for all $k\ge 3$) there is no finite upper bound on the deviation of the lower bound of Proposition 2.12 and the exact value of $n(\mathcal{R}(T))$.

3. Partial results for three files

For $k=3$ files the possible reduced recovery sets are given by

$$\begin{aligned}\mathcal{Y}^1 &= \Big\{\{4\},\{1,5\},\{2,6\},\{3,7\},\{5,6,7\},\{2,3,5\},\{1,3,6\},\{1,2,7\}\Big\},\\ \mathcal{Y}^2 &= \Big\{\{2\},\{4,6\},\{1,3\},\{5,7\},\{3,6,7\},\{1,5,6\},\{3,4,5\},\{1,4,7\}\Big\},\text{ and}\\ \mathcal{Y}^3 &= \Big\{\{1\},\{4,5\},\{2,3\},\{6,7\},\{3,5,7\},\{2,5,6\},\{3,4,6\},\{2,4,7\}\Big\}.\end{aligned}$$

Lemma 3.1. *Let $\{\lambda\}$ be a generating set of $\mathcal{R}$ and $\mathbf{n}$ be an integral solution of the ILP of Corollary 1.10 with $n_1=n_2=n_4=0$. If $\lambda\in\mathbb{R}^3_{\ge 0}$, and $\mathcal{G}$ is the multiset corresponding to $\mathbf{n}$, then $\lambda\in\mathcal{S}(\mathcal{G})$, i.e., there exists a feasible choice of α_Y satisfying (2a)-(2c).*

Proof. Plugging in $n_1=n_2=n_4=0$ the constraints of Corollary 1.10 read

$$\begin{aligned}n_5+n_6+n_7 &\ge \lambda_1,\\ n_3+n_6+n_7 &\ge \lambda_2,\\ n_3+n_5+n_7 &\ge \lambda_3,\\ n_3+n_5 &\ge \lambda_1+\lambda_2,\\ n_3+n_6 &\ge \lambda_1+\lambda_3,\\ n_5+n_6 &\ge \lambda_2+\lambda_3,\text{ and}\\ n_7 &\ge \lambda_1+\lambda_2+\lambda_3.\end{aligned}$$

We choose

$$\begin{aligned}\alpha_{\{3,7\}} &= \min\{\lambda_1,n_3\},\\ \alpha_{\{5,7\}} &= \min\{\lambda_2,n_5\},\\ \alpha_{\{6,7\}} &= \min\{\lambda_3,n_6\},\\ \alpha_{\{5,6,7\}} &= \lambda_1-\alpha_{\{3,7\}},\\ \alpha_{\{3,6,7\}} &= \lambda_2-\alpha_{\{5,7\}},\text{ and}\\ \alpha_{\{3,5,7\}} &= \lambda_3-\alpha_{\{6,7\}},\end{aligned}$$

so that we can clearly recover λ. It remains to be checked that the node capacities of Inequality (2b) are satisfied. Since $\alpha_{\{3,7\}} + \alpha_{\{5,6,7\}} = \lambda_1$, $\alpha_{\{5,7\}} + \alpha_{\{3,6,7\}} = \lambda_2$, $\alpha_{\{6,7\}} + \alpha_{\{3,5,7\}} = \lambda_3$, and $n_7 \ge \lambda_1 + \lambda_2 + \lambda_3$ Inequality (2b) is valid for $j = 7$. Since we do not use the corresponding nodes at all, Inequality (2b) is valid for $j \in \{1, 2, 4\}$. If $\lambda_1 \le n_3$, $\lambda_2 \le n_5$, and $\lambda_3 \le n_6$, then $\alpha_{\{5,6,7\}} = \alpha_{\{3,6,7\}} = \alpha_{\{3,5,7\}}$ and Inequality (2b) is valid for $j \in \{3, 5, 6\}$.

So, let us assume $n_3 < \lambda_1$. Due to $n_3 + n_5 \ge \lambda_1 + \lambda_2$ we have $n_5 > \lambda_2$ and due to $n_3 + n_6 \ge \lambda_1 + \lambda_3$ we have $n_6 > \lambda_3$, so that $\alpha_{\{5,7\}} = \lambda_2$, $\alpha_{\{6,7\}} = \lambda_3$, and $\alpha_{\{3,6,7\}} = \alpha_{\{3,5,7\}} = 0$. Thus, Inequality (2b) is valid for $j = 3$. Since $\alpha_{\{5,6,7\}} = \lambda_1 - n_3$, we can use $n_3 + n_5 \ge \lambda_1 + \lambda_2$ to conclude $n_5 \ge (\lambda_1 - n_3) + \lambda_2 = \alpha_{\{5,6,7\}} + \alpha_{\{5,7\}} + \alpha_{\{3,5,7\}}$, i.e., Inequality (2b) is valid for $j = 5$. Similarly, Since $\alpha_{\{5,6,7\}} = \lambda_1 - n_3$, we can use $n_3 + n_6 \ge \lambda_1 + \lambda_3$ to conclude $n_6 \ge (\lambda_1 - n_3) + \lambda_3 = \alpha_{\{5,6,7\}} + \alpha_{\{6,7\}} + \alpha_{\{3,6,7\}}$, i.e., Inequality (2b) is valid for $j = 6$.

The cases $n_5 < \lambda_2$ and $n_6 < \lambda_3$ can be treated analogously. □

Lemma 3.2. *Let $\{\lambda\}$ be a generating set of $\mathcal{R}$ and $\mathbf{n}$ be an integral solution of the ILP of Corollary 1.10, $\lambda \in \mathbb{R}^3_{\ge 0}$, and $\mathcal{G}$ is the multiset corresponding to $\mathbf{n}$. If one of the following three conditions is satisfied, then $\lambda \in \mathcal{S}(\mathcal{G})$, i.e., there exists a feasible choice of α_Y satisfying (2a)-(2c):*

- $n_1 = 0$, $n_2 = 0$, $\lambda_1 = 0$;
- $n_1 = 0$, $n_4 = 0$, $\lambda_2 = 0$;
- $n_2 = 0$, $n_4 = 0$, $\lambda_3 = 0$.

Proof. We only treat the first case, i.e., $n_1 = 0$, $n_2 = 0$, and $\lambda_1 = 0$. The other two cases can be handled analogously. Plugging in $n_1 = n_2 = 0$ and $\lambda_1 = 0$ the constraints of Corollary 1.10 read

$$\begin{aligned}
n_4 + n_5 + n_6 + n_7 &\ge 0,\\
n_3 + n_6 + n_7 &\ge \lambda_2,\\
n_3 + n_5 + n_7 &\ge \lambda_3,\\
n_3 + n_4 + n_5 &\ge \lambda_2,\\
n_3 + n_4 + n_6 &\ge \lambda_3,\\
n_5 + n_6 &\ge \lambda_2 + \lambda_3, \text{ and}\\
n_4 + n_7 &\ge \lambda_2 + \lambda_3.
\end{aligned}$$

We assume $n_4 > 0$ since we can otherwise apply Lemma 3.1.

If $n_7 = 0$, then we have $n_4 \ge \lambda_2 + \lambda_3$, $n_3 + n_6 \ge \lambda_2$, $n_3 + n_5 \ge \lambda_3$, and $n_5 + n_6 \ge \lambda_2 + \lambda_3$. Due to $n_5 + n_6 \ge \lambda_2 + \lambda_3$ we have $n_5 \ge \lambda_3$ or $n_6 \ge \lambda_2$. If $n_5 \ge \lambda_3$ and $n_6 \ge \lambda_2$, then we choose $\alpha_{\{4,5\}} = \lambda_3$ and $\alpha_{\{4,6\}} = \lambda_2$, which is feasible since $n_4 \ge \lambda_2 + \lambda_3 = \alpha_{\{4,5\}} + \alpha_{\{4,6\}}$, $n_5 \ge \lambda_3 = \alpha_{\{4,5\}}$, and $n_6 \ge \lambda_2 = \alpha_{\{4,6\}}$. If $n_5 \ge \lambda_3$ and $n_6 < \lambda_2$, then we choose $\alpha_{\{4,5\}} = \lambda_3$, $\alpha_{\{4,6\}} = n_6 \le \lambda_2$, and $\alpha_{\{3,4,5\}} = \lambda_2 - n_6$, which is feasible since $n_3 \ge \lambda_2 - n_6 = \alpha_{\{3,4,5\}}$, $n_4 \ge \lambda_2 + \lambda_3 = \alpha_{\{4,5\}} + \alpha_{\{4,6\}} + \alpha_{\{3,4,5\}}$, and $n_5 \ge \lambda_2 + \lambda_3 - n_6 = \alpha_{\{4,5\}} + \alpha_{\{3,4,5\}}$. Similarly, if $n_5 < \lambda_3$ and $n_6 \ge \lambda_2$, then we choose $\alpha_{\{4,6\}} = \lambda_2$, $\alpha_{\{4,5\}} = n_5 \le \lambda_3$, and $\alpha_{\{3,4,6\}} = \lambda_3 - n_5$, which is feasible since $n_3 \ge \lambda_3 - n_5 = \alpha_{\{3,4,6\}}$, $n_4 \ge \lambda_2 + \lambda_3 = \alpha_{\{4,5\}} + \alpha_{\{4,6\}} + \alpha_{\{3,4,6\}}$, and $n_6 \ge \lambda_2 + \lambda_3 - n_5 = \alpha_{\{4,6\}} + \alpha_{\{3,4,6\}}$. Thus, we can assume $n_7 > 0$.

If $n_5 = 0$, then we have $n_6 \ge \lambda_2 + \lambda_3$, $n_3 + n_7 \ge \lambda_3$, $n_3 + n_4 \ge \lambda_2$, and $n_4 + n_7 \ge \lambda_2 + \lambda_3$. Due to $n_4 + n_7 \ge \lambda_2 + \lambda_3$ we have $n_7 \ge \lambda_3$ or $n_4 \ge \lambda_2$. If $n_7 \ge \lambda_3$ and $n_4 \ge \lambda_2$, then we choose $\alpha_{\{6,7\}} = \lambda_3$ and $\alpha_{\{4,6\}} = \lambda_2$, which is feasible since $n_4 \ge \lambda_2 = \alpha_{\{4,6\}}$, $n_6 \ge \lambda_2 + \lambda_3 = \alpha_{\{6,7\}} + \alpha_{\{4,6\}}$, and $n_7 \ge \lambda_3 = \alpha_{\{6,7\}}$. If $n_7 \ge \lambda_3$ and $n_4 < \lambda_2$, then we choose $\alpha_{\{6,7\}} = \lambda_3$, $\alpha_{\{4,6\}} = n_4$, and $\alpha_{\{3,6,7\}} = \lambda_2 - n_4$, which is feasible since $n_3 \ge \lambda_2 - n_4 = \alpha_{\{3,6,7\}} = \lambda_2 - n_4$, $n_4 = \alpha_{\{4,6\}}$, $n_6 \ge \lambda_2 + \lambda_3 = \alpha_{\{6,7\}} + \alpha_{\{4,6\}} + \alpha_{\{3,6,7\}}$, and $n_7 \ge \lambda_2 + \lambda_3 - n_4 = \alpha_{\{6,7\}} + \alpha_{\{3,6,7\}} = \lambda_2 - n_4$. If $n_4 \ge \lambda_2$ and $n_7 < \lambda_3$, then we choose $\alpha_{\{4,6\}} = \lambda_2$, $\alpha_{\{6,7\}} = n_7$, and $\alpha_{\{3,4,6\}} = \lambda_3 - n_7$, which is feasible since $n_3 \ge \lambda_3 - n_7 = \alpha_{\{3,4,6\}} = \lambda_2 - n_4$, $n_4 \ge \lambda_2 = \alpha_{\{4,6\}}$, $n_6 \ge \lambda_2 + \lambda_3 = \alpha_{\{6,7\}} + \alpha_{\{4,6\}} + \alpha_{\{3,4,6\}}$, and $n_7 = \alpha_{\{6,7\}}$. Thus, we can assume $n_5 > 0$.

For the case $n_6 = 0$ we can proceed similarly as for the case $n_5 = 0$, so that we can assume $n_6 > 0$.

As argued above we can assume $n_4, n_5, n_6, n_7 > 0$ w.l.o.g. Now assume that at least of of the equations $n_3 + n_6 + n_7 \ge \lambda_2$, $n_3 + n_5 + n_7 \ge \lambda_3$, $n_3 + n_4 + n_5 \ge \lambda_2$, or $n_3 + n_4 + n_6 \ge \lambda_3$ is tight, i.e., satisfied with equality. Here we consider only the case $n_3 + n_6 + n_7 = \lambda_2$ and remark that the other three cases can be treated analogously. If also $n_3 + n_5 + n_7 = \lambda_3$, then summing yields $2n_3 + n_5 + n_6 + 2n_7 = \lambda_2 + \lambda_3$, so that $n_5 + n_6 \ge \lambda_2 + \lambda_3$ gives $n_7 = 0$ – contradiction. If, alternatively, also $n_3 + n_4 + n_6 = \lambda_3$, then summing yields $2n_3 + n_4 + 2n_6 + n_7 = \lambda_2 + \lambda_3$, so that $n_4 + n_7 \ge \lambda_2 + \lambda_3$ gives $n_6 = 0$ – contradiction. Thus, we have $n_3 + n_5 + n_7 \ge \lambda_3 + 1$ and $n_3 + n_4 + n_6 \ge \lambda_3 + 1$. If $n_3 + n_5 + n_7 = \lambda_3 + 1$, then $n_3 + n_6 + n_7 = \lambda_2$ yields $2n_3 + n_5 + n_6 + 2n_7 = \lambda_2 + \lambda_3 + 1$. Since $n_5 + n_6 \ge \lambda_2 + \lambda_3$, we conclude $n_7 = 0$ – contradiction. Thus, we have $n_3 + n_5 + n_7 \ge \lambda_3 + 2$. Next we choose $\alpha_{\{5,7\}} = 1$ and set $n'_1 = n_1 = 0$, $n'_2 = n_2 = 0$, $n'_3 = n_3$, $n'_4 = n_4$, $n'_5 = n_5 - 1$, $n'_6 = n_6$, $n'_7 = n_7 - 1$, $\lambda'_1 = \lambda_1 = 0$, $\lambda'_2 = \max\{0, \lambda_2 - 1\}$, and $\lambda'_3 = \lambda_3$. With this we have $n' = (n'_1, \ldots, n'_7) \in \mathbb{N}^7$ and $\lambda' = (\lambda'_1, \lambda'_2, \lambda'_3) \in \mathbb{N}^3$ satisfying

$$\begin{aligned}
n'_4 + n'_5 + n'_6 + n'_7 &\ge 0, \\
n'_3 + n'_6 + n'_7 &\ge \lambda'_2, \\
n'_3 + n'_5 + n'_7 &\ge \lambda'_3, \\
n'_3 + n'_4 + n'_5 &\ge \lambda'_2, \\
n'_3 + n'_4 + n'_6 &\ge \lambda'_3, \\
n'_5 + n'_6 &\ge \lambda'_2 + \lambda'_3, \\
n'_4 + n'_7 &\ge \lambda'_2 + \lambda'_3, \\
n'_1 &= 0, \text{ and} \\
n'_2 &= 0.
\end{aligned}$$

If none of the four inequalities $n_3 + n_6 + n_7 \ge \lambda_2$, $n_3 + n_5 + n_7 \ge \lambda_3$, $n_3 + n_4 + n_5 \ge \lambda_2$, or $n_3 + n_4 + n_6 \ge \lambda_3$ is tight, i.e., satisfied with equality, then we can choose $\alpha_{\{4,5\}} = \frac{1}{2}$, $\alpha_{\{4,6\}} = \frac{1}{2}$, $\alpha_{\{5,7\}} = \frac{1}{2}$, $\alpha_{\{6,7\}} = \frac{1}{2}$ and set $n'_1 = n_1 = 0$, $n'_2 = n_2 = 0$, $n'_3 = n_3$, $n'_4 = n_4 - 1$, $n'_5 = n_5 - 1$, $n'_6 = n_6 - 1$, $n'_7 = n_7 - 1$, $\lambda'_1 = \lambda_1 = 0$, $\lambda'_2 = \max\{0, \lambda_2 - 1\}$, and $\lambda'_3 = \max\{0, \lambda_3 - 1\}$. With this we have $n' = (n'_1, \ldots, n'_7) \in \mathbb{N}^7$ and $\lambda' = (\lambda'_1, \lambda'_2, \lambda'_3) \in \mathbb{N}^3$ satisfying

$$\begin{aligned}
n'_4 + n'_5 + n'_6 + n'_7 &\ge 0, \\
n'_3 + n'_6 + n'_7 &\ge \lambda'_2, \\
n'_3 + n'_5 + n'_7 &\ge \lambda'_3, \\
n'_3 + n'_4 + n'_5 &\ge \lambda'_2, \\
n'_3 + n'_4 + n'_6 &\ge \lambda'_3, \\
n'_5 + n'_6 &\ge \lambda'_2 + \lambda'_3, \\
n'_4 + n'_7 &\ge \lambda'_2 + \lambda'_3, \\
n'_1 &= 0, \text{ and} \\
n'_2 &= 0.
\end{aligned}$$

Thus, in all remaining cases we can iteratively decrease the value of $\sum_{i\in[7]} n_i$ by at least one and receive a smaller instance satisfying the same assumptions as the original one. So, the proof if finished by induction on $\sum_{i\in[7]} n_i$. □

We remark that in Lemma 3.2 it is essential, that we allow fractional α_Y, since for $\lambda = (0, 1, 1)$ and $\mathbf{n} = (0, 0, 0, 1, 1, 1, 1)$ no integral solution of the constraints (2a)-(2c) exists.

Lemma 3.3. *Let $\{\lambda\}$ be a generating set of $\mathcal{R}$ and $\mathbf{n}$ be an integral solution of the ILP of Corollary 1.10, $\lambda \in \mathbb{R}^3_{\ge 0}$, and $\mathcal{G}$ is the multiset corresponding to $\mathbf{n}$. If one of the following three conditions is satisfied, then $\lambda \in \mathcal{S}(\mathcal{G})$, i.e., there exists a feasible choice of α_Y satisfying (2a)-(2c):*

- $n_1 = 0$, $n_3 = 0$, $n_5 = 0$, $\min\{n_6, n_7\} = 0$, $\lambda_1 = 0$, $\lambda_2 = 0$;
- $n_2 = 0$, $n_3 = 0$, $n_6 = 0$, $\min\{n_5, n_7\} = 0$, $\lambda_1 = 0$, $\lambda_3 = 0$;
- $n_4 = 0$, $n_5 = 0$, $n_6 = 0$, $\min\{n_3, n_7\} = 0$, $\lambda_2 = 0$, $\lambda_3 = 0$.

Proof. We only treat the first case, i.e., $n_1 = 0$, $n_3 = 0$, $n_5 = 0$, $\min\{n_6, n_7\} = 0$, λ_1, and $\lambda_2 = 0$. The other two cases can be handled analogously. Plugging in these assumptions the constraints of Corollary 1.10 read

$$\begin{aligned} n_4 + n_6 + n_7 &\ge 0, \\ n_2 + n_6 + n_7 &\ge 0, \\ n_7 &\ge \lambda_3, \\ n_2 + n_4 &\ge 0, \\ n_4 + n_6 &\ge \lambda_3, \\ n_2 + n_6 &\ge \lambda_3, \text{ and} \\ n_2 + n_4 + n_7 &\ge \lambda_3. \end{aligned}$$

If $n_7 = 0$, then $\lambda_3 = 0$. In this case we have $\lambda = \mathbf{0}$ for which the statement is satisfied by choosing $\alpha_Y = 0$ for all reduced recovery sets Y. Thus, we have $n_6 = 0$, $n_7 \ge \lambda_3$, $n_4 \ge \lambda_3$, and $n_2 \ge \lambda_3$, so that we can choose $\alpha_{\{2,4,7\}} = \lambda_3$. □

Lemma 3.4. *Let $\{\lambda\}$ be a generating set of $\mathcal{R}$ and $\mathbf{n}$ be an integral solution of the ILP of Corollary 1.10, $\lambda \in \mathbb{R}^3_{\ge 0}$, and $\mathcal{G}$ is the multiset corresponding to $\mathbf{n}$. If one of the following three conditions is satisfied, then $\lambda \in \mathcal{S}(\mathcal{G})$, i.e., there exists a feasible choice of α_Y satisfying (2a)-(2c):*

- $n_1 = 0$, $\lambda_1 = 0$, $\lambda_2 = 0$;
- $n_2 = 0$, $\lambda_1 = 0$, $\lambda_3 = 0$;
- $n_4 = 0$, $\lambda_2 = 0$, $\lambda_3 = 0$.

Proof. We only treat the first case, i.e., $n_1 = 0$, λ_1, and $\lambda_2 = 0$. The other two cases can be handled analogously. Plugging in n_1 and $\lambda_1 = \lambda_2 = 0$ the constraints of Corollary 1.10 read

$$\begin{aligned} n_4 + n_5 + n_6 + n_7 &\ge 0, \\ n_2 + n_3 + n_6 + n_7 &\ge 0, \\ n_3 + n_5 + n_7 &\ge \lambda_3, \\ n_2 + n_3 + n_4 + n_5 &\ge 0, \\ n_3 + n_4 + n_6 &\ge \lambda_3, \\ n_2 + n_5 + n_6 &\ge \lambda_3, \text{ and} \\ n_2 + n_4 + n_7 &\ge \lambda_3. \end{aligned}$$

We choose

$$\begin{aligned} \alpha_{\{2,3\}} &= \min\{n_2, n_3\}, \\ \alpha_{\{4,5\}} &= \min\{n_4, n_5\}, \\ \alpha_{\{6,7\}} &= \min\{n_6, n_7\} \end{aligned}$$

and set

$$\begin{aligned} n'_1 &= n_1, \\ n'_2 &= n_2 - \alpha_{\{2,3\}}, \\ n'_3 &= n_3 - \alpha_{\{2,3\}}, \\ n'_4 &= n_4 - \alpha_{\{4,5\}}, \\ n'_5 &= n_5 - \alpha_{\{4,5\}} \\ n'_6 &= n_6 - \alpha_{\{6,7\}}, \\ n'_7 &= n_7 - \alpha_{\{6,7\}}, \\ \lambda'_1 &= \lambda_1, \\ \lambda'_2 &= \lambda_2, \text{ and} \\ \lambda'_3 &= \max\left\{0, \lambda_3 - \alpha_{\{2,3\}} - \alpha_{\{4,5\}} - \alpha_{\{6,7\}}\right\}. \end{aligned}$$

With this we have $n' = (n'_1, \ldots, n'_7) \in \mathbb{N}^7$ and $\lambda' = (\lambda'_1, \lambda'_2, \lambda'_3) \in \mathbb{N}^3$ satisfying

$$\begin{aligned} n'_4 + n'_5 + n'_6 + n'_7 &\geq \lambda'_1, \\ n'_2 + n'_3 + n'_6 + n'_7 &\geq \lambda'_2, \\ n'_1 + n'_3 + n'_5 + n'_7 &\geq \lambda'_3, \\ n'_2 + n'_3 + n'_4 + n'_5 &\geq \lambda'_1 + \lambda'_2, \\ n'_1 + n'_2 + n'_5 + n'_6 &\geq \lambda'_2 + \lambda'_3, \\ n'_1 + n'_3 + n'_4 + n'_6 &\geq \lambda'_1 + \lambda'_3, \\ n'_1 + n'_2 + n'_4 + n'_7 &\geq \lambda'_1 + \lambda'_2 + \lambda'_3, \\ n'_1 &= 0, \\ \lambda'_1 &= 0, \\ \lambda'_2 &= 0, \\ n'_2 = 0 &\vee n'_3 = 0, \\ n'_4 = 0 &\vee n'_5 = 0, \text{ and} \\ n'_6 = 0 &\vee n'_7 = 0. \end{aligned}$$

If $n'_2 = 0$ or $n'_4 = 0$, then we can apply Lemma 3.2. Otherwise we have $n'_3 = n'_5 = 0$ and can apply Lemma 3.3. □

Theorem 3.5. *Let $\{\lambda\}$ be a generating set of $\mathcal{R}$ and $\mathbf{n}$ be an integral solution of the ILP of Corollary 1.10. If $\lambda \in \mathbb{R}^3_{\geq 0}$, and $\mathcal{G}$ is the multiset corresponding to $\mathbf{n}$, then $\lambda \in \mathcal{S}(\mathcal{G})$, i.e., there exists a feasible choice of α_Y satisfying (2a)-(2c).*

Proof. The constraints of Corollary 1.10 read

$$\begin{aligned} n_4 + n_5 + n_6 + n_7 &\geq \lambda_1, \\ n_2 + n_3 + n_6 + n_7 &\geq \lambda_2, \\ n_1 + n_3 + n_5 + n_7 &\geq \lambda_3, \\ n_2 + n_3 + n_4 + n_5 &\geq \lambda_1 + \lambda_2, \\ n_1 + n_3 + n_4 + n_6 &\geq \lambda_1 + \lambda_3, \\ n_1 + n_2 + n_5 + n_6 &\geq \lambda_2 + \lambda_3, \text{ and} \\ n_1 + n_2 + n_4 + n_7 &\geq \lambda_1 + \lambda_2 + \lambda_3. \end{aligned}$$

We choose

$$\begin{aligned} \alpha_{\{4\}} &= \min\{n_4, \lambda_1\}, \\ \alpha_{\{2\}} &= \min\{n_2, \lambda_2\}, \\ \alpha_{\{1\}} &= \min\{n_1, \lambda_3\} \end{aligned}$$

and set

$$\begin{aligned} n'_1 &= n_1 - \alpha_{\{1\}}, \\ n'_2 &= n_2 - \alpha_{\{2\}}, \\ n'_3 &= n_3, \\ n'_4 &= n_4 - \alpha_{\{4\}}, \\ n'_5 &= n_5, \\ n'_6 &= n_6, \\ n'_7 &= n_7, \\ \lambda'_1 &= \lambda_1 - \alpha_{\{4\}}, \\ \lambda'_2 &= \lambda_2 - \alpha_{\{2\}}, \text{ and} \\ \lambda'_3 &= \lambda_3 - \alpha_{\{1\}}. \end{aligned}$$

With this we have $n' = (n'_1, \dots, n'_7) \in \mathbb{N}^7$ and $\lambda' = (\lambda'_1, \lambda'_2, \lambda'_3) \in \mathbb{N}^3$ satisfying

$$\begin{aligned} n'_4 + n_5 +' n'_6 + n'_7 &\ge \lambda'_1, \\ n'_2 + n'_3 + n'_6 + n'_7 &\ge \lambda'_2, \\ n'_1 + n'_3 + n'_5 + n'_7 &\ge \lambda'_3, \\ n'_2 + n'_3 + n'_4 + n'_5 &\ge \lambda'_1 + \lambda'_2, \\ n'_1 + n'_2 + n'_5 + n'_6 &\ge \lambda'_2 + \lambda'_3, \\ n'_1 + n'_3 + n'_4 + n'_6 &\ge \lambda'_1 + \lambda'_3, \\ n'_1 + n'_2 + n'_4 + n'_7 &\ge \lambda'_1 + \lambda'_2 + \lambda'_3, \\ n'_4 = 0 &\vee \lambda'_1 = 0, \\ n'_2 = 0 &\vee \lambda'_2 = 0, \text{ and} \\ n'_1 = 0 &\vee \lambda'_3 = 0. \end{aligned}$$

If $\lambda' = \mathbf{0}$, then our choice of the α_Y is feasible. If $n'_1 = n'_2 = n'_4 = 0$, then we can apply Lemma 3.1. If exactly one value in $\{n'_1, n'_2, n'_4\}$ is equal to zero, then we can apply Lemma 3.4, if there exactly two zero values, then we can apply Lemma 3.2. □

In the following we assume that $T\colon 2^{[3]} \to \mathbb{N}$ is a function such that that $T(\emptyset) = 0$ and none of the constraints $\sum_{i\in U} \lambda_i \le T(U)$ is strictly redundant in $\mathbb{R}^3_{\ge 0}$. Especially, we have that T is monotone and subadditive. As an abbreviation we write $T(\emptyset) = T(0)$, $T(\{1\}) = T(1)$, $T(\{2\}) = T(2)$, $T(\{3\}) = T(3)$, $T(\{1,2\}) = T(12)$, $T(\{1,3\}) = T(13)$, $T(\{2,3\}) = T(2,3)$, and $T(\{1,2,3\}) = T(123)$. Next, we start to study the LP relaxation of the ILP of Corollary 1.10. Under certain conditions the optimal target value is $\frac{\sum_{\emptyset\neq U\subseteq[3]} T(U)}{4}$, which is used in Proposition 2.12.

Due to our assumption that none constraint is strictly redundant the constraints of the ILP of Corollary 1.10 are given by

$$\begin{aligned}
n_4+n_5+n_6+n_7 &\ge T(1),\\
n_2+n_3+n_6+n_7 &\ge T(2),\\
n_1+n_3+n_5+n_7 &\ge T(3),\\
n_2+n_3+n_4+n_5 &\ge T(12),\\
n_1+n_3+n_4+n_6 &\ge T(13),\\
n_1+n_2+n_5+n_6 &\ge T(23), \text{ and}\\
n_1+n_2+n_4+n_7 &\ge T(123).
\end{aligned}$$

Lemma 3.6. *If*

$$\begin{aligned}
-T(123)+T(12)+T(13)-T(23)+T(2)+T(3)-T(1) &\ge 0,\\
-T(123)+T(12)+T(23)-T(13)+T(1)+T(3)-T(2) &\ge 0,\\
-T(123)+T(13)+T(23)-T(12)+T(1)+T(2)-T(3) &\ge 0, \textit{ and}\\
T(123)-T(12)-T(13)-T(23)+T(1)+T(2)+T(3) &\ge 0,
\end{aligned}$$

then the LP relaxation of the ILP of Corollary 1.10 admits

$$\begin{aligned}
4n_1 &= T(123)+T(13)+T(23)-T(12)-T(1)-T(2)+T(3),\\
4n_2 &= T(123)+T(12)+T(23)-T(13)-T(1)-T(3)+T(2),\\
4n_3 &= -T(123)+T(12)+T(13)-T(23)+T(2)+T(3)-T(1),\\
4n_4 &= T(123)+T(12)+T(13)-T(23)-T(2)-T(3)+T(1),\\
4n_5 &= -T(123)+T(12)+T(23)-T(13)+T(1)+T(3)-T(2),\\
4n_6 &= -T(123)+T(13)+T(23)-T(12)+T(1)+T(2)-T(3),\\
4n_7 &= T(123)-T(12)-T(13)-T(23)+T(1)+T(2)+T(3),
\end{aligned}$$

as an optimal solution with target value

$$\frac{T(123)+T(12)+T(13)+T(23)+T(1)+T(2)+T(3)}{4}.$$

Proof. As mentioned before, we have

$$\sum_{j\in[l]\,:\,v_j\notin\mathcal{H}} n_j \ge T(S(\mathcal{H})),$$

where $S(\mathcal{H})=\{i\in[k]\,:\,\mathbf{e}_i\notin\mathcal{H}\}$. We can easily check that the proposed vector $\mathbf{n}=(n_1,\ldots,n_7)$ satisfies all of these equations with equality so that $\sum_{i\in[7]} n_i$ clearly is minimal and has the stated value. In remains to check $n\ge 0$. From the monotonicity of T we conclude $T(123)\ge T(12)$, $T(13)\ge T(1)$ and $T(23)\ge T(2)$, so that $n_1\ge 0$. Applying the symmetries of the symmetric group on three elements we similarly conclude $n_2\ge 0$ and $n_4\ge 0$. The conditions for $n_3,n_5,n_6,n_7\ge 0$ are equivalent to the assumed constraints. □

We remark that the values of n_3, n_5, and n_6 can indeed be negative as it is the case in Example 2.16. For n_7 such an example is given by:

Example 3.7. Let x be a positive integer and $T\colon 2^{[3]}\to\mathbb{N}$ be defined by $T(\{1\})=T(\{2\})=T(\{3\})=x$ and $T(\{1,2\})=T(\{1,3\})=T(\{2,3\})=T(\{1,2,3\})=2x$. Proposition 2.12 gives $n(\mathcal{R}(T))\ge\left\lceil\frac{11x}{4}\right\rceil$. For $j=7$ Proposition 2.15 gives $n(\mathcal{R}(T))\ge\left\lceil\frac{6x}{2}\right\rceil=3x$. In Lemma 3.6 the value for n_7, i.e., $\big(T(123)-T(12)-T(13)-T(23)+T(1)+T(2)+T(3)\big)/4$, equals $-x/4$, which is negative for $x>0$.

Conjecture 3.8. *Given the assumptions of Lemma 3.6, we have*

$$n(\mathcal{R}(T)) \le \left\lceil \frac{\sum_{\emptyset\neq U\subseteq[3]} T(U)}{4} \right\rceil + 5$$

and a corresponding code is obtained by uprounding the values for the n_i in Lemma 3.6.

Next we want to reduce the complexity of the problem an restrict ourselves to functions $T\colon 2^{[3]} \to \mathbb{N}$, where $T(U)$ only depends on the cardinality of U, i.e., we set $T(0) = 0$, $T(1) = T(2) = T(3) = X$, $T(12) = T(13) = T(23) = Y$, and $T(123) = Z$ for $X, Y, Z \in \mathbb{N}$. Monotonicity of T is equivalent to $X \le Y \le Z$ and subadditivity is equivalent to $Y \le 2X$ and $Z \le X + Y$. The latter constraint can be tightened to $Z \le \left\lfloor \frac{3Y}{2} \right\rfloor$, since $\lambda_1 + \lambda_2 \le Y$, $\lambda_1 + \lambda_3 \le Y$, and $\lambda_2 + \lambda_3 \le Y$ imply $\lambda_1 + \lambda_2 + \lambda_3 \le \frac{3Y}{2}$, which is upper bounded by $X + Y$ due to $Y \le 2X$ (in general: $T(123) \le \left\lfloor \frac{T(12)+T(13)+T(23)}{2} \right\rfloor$).

In our situation the constraints $\sum_{j\in U} \lambda_j \le T^{(i)}(U)$ for all $\emptyset \neq U \subseteq [k]$ read

$$\lambda_1 \le X, \tag{19}$$
$$\lambda_2 \le X, \tag{20}$$
$$\lambda_3 \le X, \tag{21}$$
$$\lambda_1+\lambda_2 \le Y, \tag{22}$$
$$\lambda_1+\lambda_3 \le Y, \tag{23}$$
$$\lambda_2+\lambda_3 \le Y, \text{ and} \tag{24}$$
$$\lambda_1+\lambda_2+\lambda_3 \le Z. \tag{25}$$

Before we study under which conditions non of these is strictly redundant, we state that the list of vectors of Proposition 2.10 that form a generating set "simplifies" to:

- $\Gamma_1 = \Big(X, Y-X, \min\{Z-Y, Y-X\}\Big)$;
- $\Gamma_2 = \Big(X, \min\{Z-Y, Y-X\}, Y-X\Big)$;
- $\Gamma_3 = \Big(Y-X, X, \min\{Z-Y, Y-X\}\Big)$;
- $\Gamma_4 = \Big(\min\{Z-Y, Y-X\}, X, Y-X\Big)$;
- $\Gamma_5 = \Big(Y-X, \min\{Z-Y, Y-X\}, X\Big)$;
- $\Gamma_6 = \Big(\min\{Z-Y, Y-X\}, Y-X, X\Big)$;
- $\Gamma_7 = (2Y-Z, Z-Y, Z-Y)$ if $2Y \le X+Z$;
- $\Gamma_8 = (Z-Y, 2Y-Z, Z-Y)$ if $2Y \le X+Z$;
- $\Gamma_9 = (Z-Y, Z-Y, 2Y-Z)$ if $2Y \le X+Z$.

Lemma 3.9. *Given $(X, Y, Z) \in \mathbb{N}^3$, then none of the constraints (19)-(25) is strictly redundant iff $X \le Y \le 2X$ and $Y \le Z \le \left\lfloor \frac{3Y}{2} \right\rfloor$.*

Proof. As shown above, the conditions $X \le Y \le 2X$ and $Y \le Z \le \left\lfloor \frac{3Y}{2} \right\rfloor$ are necessary. It remains to show that for each of the above seven constraints there there exists a vector $\lambda \in \mathbb{N}^3$ for which this constraint is tight, i.e., satisfied with equality, and also the other six constraints are satisfied.

If $2Y \ge X + Z$ we can easily check that the vectors

- $\Gamma_1 = \Big(X, Y-X, Z-Y\Big)$;
- $\Gamma_2 = \Big(X, Z-Y, Y-X\Big)$;
- $\Gamma_3 = \Big(Y-X, X, Z-Y\Big)$;

- $\Gamma_4 = \Big(Z-Y, X, Y-X\Big)$;
- $\Gamma_5 = \Big(Y-X, Z-Y, X\Big)$;
- $\Gamma_6 = \Big(Z-Y, Y-X, X\Big)$

satisfy all constraints (19)-(25). Moreover, all of them satisfy Inequality (25) with equality and for each other constraint there exists an index $i \in [6]$ such that it is tight for Γ_i.

If $2Y \le X + Z$ we can easily check that the vectors

- $\Gamma_1 = \Big(X, Y-X, Y-X\Big)$;
- $\Gamma_3 = \Big(Y-X, X, Y-X\Big)$;
- $\Gamma_5 = \Big(Y-X, Y-X, X\Big)$;
- $\Gamma_7 = (2Y-Z, Z-Y, Z-Y)$;
- $\Gamma_8 = (Z-Y, 2Y-Z, Z-Y)$;
- $\Gamma_9 = (Z-Y, Z-Y, 2Y-Z)$

satisfy all constraints (19)-(25). Moreover, for $i \in \{7, 8, 9\}$ the vector Γ_i satisfies Inequality (25) with equality and for each other constraint there exists an index $i\{1, 3, 5\}$ such that it is tight for Γ_i. □

So, assuming $X \le Y \le 2X$ and $Y \le Z \le \left\lfloor \frac{3Y}{2} \right\rfloor$ we have: Proposition 2.12 gives

$$n(\mathcal{R}(T)) \ge \left\lceil \frac{3X + 3Y + Z}{4} \right\rceil, \tag{26}$$

Proposition 2.14 gives

$$n(\mathcal{R}(T)) \ge \left\lceil \frac{\left\lceil \frac{2X+Y}{2} \right\rceil + \left\lceil \frac{X+2Y+Z}{2} \right\rceil}{2} \right\rceil, \tag{27}$$

Proposition 2.15 gives

$$n(\mathcal{R}(T)) \ge \max\left\{ \left\lceil \frac{2X+Y}{2} \right\rceil, \left\lceil \frac{X+Y+Z}{2} \right\rceil, \left\lceil \frac{3Y}{2} \right\rceil \right\} = \max\left\{ \left\lceil \frac{X+Y+Z}{2} \right\rceil, \left\lceil \frac{3Y}{2} \right\rceil \right\} \tag{28}$$

Note that the assumption that none of the constraints (19)-(25) is strictly redundant is indeed essential. To this end consider $(X, Y, Z) = (8n, 8n, 40n)$ for some integer n. Here Inequality (26) and Inequality (27) both would give $n(\mathcal{R}(T)) \ge 22n$ while the $3n$-fold 3-dimensional simplex code gives $n(\mathcal{R}(T)) \le 21n$. Note that we also have $\left\lceil \frac{X+Y+Z}{2} \right\rceil = 28n > 21n$. These contradictory results are due to the fact that there are indeed strictly redundant constraints. Similarly, for $(X, Y, Z) = (4n, 16n, 16n)$, for some integer n, we have $\left\lceil \frac{3Y}{2} \right\rceil = 24n$, while the $3n$-fold 3-dimensional simplex code gives $n(\mathcal{R}(T)) \le 21n$.

The constraints of the ILP of Corollary 1.10 simplify to

$$\begin{aligned}
n_4 + n_5 + n_6 + n_7 &\ge X, &(29)\\
n_2 + n_3 + n_6 + n_7 &\ge X, &(30)\\
n_1 + n_3 + n_5 + n_7 &\ge X, &(31)\\
n_2 + n_3 + n_4 + n_5 &\ge Y, &(32)\\
n_1 + n_3 + n_4 + n_6 &\ge Y, &(33)\\
n_1 + n_2 + n_5 + n_6 &\ge Y, \text{ and} &(34)\\
n_1 + n_2 + n_4 + n_7 &\ge Z. &(35)
\end{aligned}$$

For the corresponding LP relaxation we can utilize the group action of the symmetric group $\mathcal{S}_3$ on 3 elements. More precisely, for a permutation $\pi \in \mathcal{S}_3$ and a vector $\mathbf{x} \in \{0, 1\}^3$ let $\mathbf{x}^\pi \in \{0, 1\}^3$ denote the vector arising from $\mathbf{x}$ by permuting the coordinates according to π. Using this, we can write i^π for each

$i \in [7]$ by setting $i^\pi = j \in [7]$, where $v_i^\pi = v_j$, i.e., we use the correspondence used in our indexing. Next, we extend the notation to solution vectors $\mathbf{n} = (n_1, \dots, n_7) \in \mathbb{R}_{\ge 0}^7$ by setting

$$\mathbf{n}^\pi = (n_{1^\pi}, \dots n_{7^\pi}) .$$

Due to the underlying symmetry of the constraints (29)- (35) we have that $\mathbf{n}$ is a feasible solution iff $\mathbf{n}^\pi$ is a feasible solution, where $\pi \in \mathcal{S}_3$ is arbitrary. Since the problem is convex, also

$$\mathbf{n}^\star = \frac{1}{6} \cdot \sum_{\pi \in \mathcal{S}^3} \mathbf{n}^\pi$$

is a feasible solution. (In this last step it is essential that we have removed the integrality conditions on the n_i.) The group action of the $\mathcal{S}_3$ partitions the variables into orbits

$$\Big\{n_1, n_2, n_4\Big\}, \Big\{n_3, n_5, n_6\Big\}, \text{ and } \Big\{n_7\Big\},$$

so that we replace the variables of the first orbit by N_1, those of the second orbit by N_2, and, for consistency, set $n_7 = N_3$. With this the target function $\sum_{i\in[7]} n_i$ translates to $3N_1 + 3N_2 + N_3$. Plugging in these new variables into the constraints (29)- (35) yields several duplicates. (Also set of constraints is partitioned by the group action into orbits and we only have to take one representant from each orbit.) So, in our case the system reduces to

$$N_1 + 2N_2 + N_3 \ge X, \qquad (36)$$

$$2N_1 + 2N_2 \ge Y, \text{ and} \qquad (37)$$

$$3N_1 + N_3 \ge Z, \qquad (38)$$

and $X, Y, Z \ge 0$. We remark that this "symmetrization technique" is well known in the context of semidefinite programming, see e.g. [1, 5]. For an application to symmetric linear programs we refer to e.g. [2].

Proposition 3.10. *If $X, Y, Z \in \mathbb{N}$ satisfy $X \le Y \le 2X$ and $Y \le Z \le \left\lfloor \frac{3Y}{2} \right\rfloor$, then for $T\colon 2^{[3]} \to \mathbb{N}$ depending on X, Y, and Z as described above we have*

$$n(\mathcal{R}(T)) \ge \left\lceil \frac{3 \cdot \left\lceil \frac{3X+Y+Z}{2} \right\rceil + 3 \cdot \left\lceil \frac{2X+3Y}{2} \right\rceil + \left\lceil \frac{3Y+2Z}{2} \right\rceil - 7 \cdot \Omega}{3} \right\rceil \qquad (39)$$

where $\Omega = \left\lceil \frac{3X+3Y+Z}{4} \right\rceil$.

Proof. First we show

$$\left\lceil \frac{3 \cdot \left\lceil \frac{3X+Y+Z}{2} \right\rceil + 3 \cdot \left\lceil \frac{2X+3Y}{2} \right\rceil + \left\lceil \frac{3Y+2Z}{2} \right\rceil - 7 \cdot \Omega}{3} \right\rceil \le \Omega + 1.$$

To this end we observe

$$\left\lceil \frac{3 \cdot \left\lceil \frac{3X+Y+Z}{2} \right\rceil + 3 \cdot \left\lceil \frac{2X+3Y}{2} \right\rceil + \left\lceil \frac{3Y+2Z}{2} \right\rceil - 7 \cdot \Omega}{3} \right\rceil \le \left\lceil \frac{\frac{1}{4} \cdot (3X + 3Y + Z) - 7\Omega + \frac{7}{2}}{3} \right\rceil . \qquad (40)$$

If $\Omega - \frac{3X+3Y+Z}{4} \ge \frac{1}{4}$, then the right hand side of Inequality (40) is at most

$$\left\lceil \frac{\frac{10}{4} \cdot (3X + 3Y + Z) - 7\Omega + \frac{7}{2}}{3} \right\rceil = \left\lceil \frac{3X + 3Y + Z}{4} + \frac{7}{12} \right\rceil \le \Omega + 1.$$

Now, let us assume $\Omega = \frac{3X+3Y+Z}{4}$, so that either two or none of the integers X, Y, and Z is odd. In that case the left hand side of Inequality (40) is at most

$$\left\lceil \frac{\frac{10}{4} \cdot (3X + 3Y + Z) - 7\Omega + \frac{4}{2}}{3} \right\rceil = \left\lceil \frac{3X + 3Y + Z}{4} + \frac{2}{3} \right\rceil \le \Omega + 1.$$

Thus, we can assume

$$n_1+n_2+n_3+n_4+n_5+n_6+n_7 \le \left\lceil \frac{3X+3Y+Z}{4} \right\rceil = \Omega \tag{41}$$

for a solution $(n_1,\dots,n_7)$ for a moment (since for $n_1+n_2+n_3+n_4+n_5+n_6+n_7 \ge \Omega+1$ the proposed inequality is satisfied). Now we combine the inequalities (29)(41):

$$\begin{aligned}
n_4+n_5+n_7 &\ge \left\lceil \frac{3X+Y+Z}{2} \right\rceil - \Omega && (29)+\frac{1}{2}\cdot\Big((31)+(32)+(35)\Big)-(41);\\
n_2+n_6+n_7 &\ge \left\lceil \frac{3X+Y+Z}{2} \right\rceil - \Omega && (30)+\frac{1}{2}\cdot\Big((29)+(34)+(35)\Big)-(41);\\
n_1+n_3+n_7 &\ge \left\lceil \frac{3X+Y+Z}{2} \right\rceil - \Omega && (31)+\frac{1}{2}\cdot\Big((30)+(33)+(35)\Big)-(41);\\
n_3+n_4+n_5 &\ge \left\lceil \frac{2X+3Y}{2} \right\rceil - \Omega && (32)+\frac{1}{2}\cdot\Big((29)+(31)+(33)\Big)-(41);\\
n_1+n_3+n_6 &\ge \left\lceil \frac{2X+3Y}{2} \right\rceil - \Omega && (33)+\frac{1}{2}\cdot\Big((30)+(31)+(34)\Big)-(41);\\
n_2+n_5+n_6 &\ge \left\lceil \frac{2X+3Y}{2} \right\rceil - \Omega && (34)+\frac{1}{2}\cdot\Big((29)+(30)+(32)\Big)-(41);\\
n_1+n_2+n_4 &\ge \left\lceil \frac{3Y+2Z}{2} \right\rceil - \Omega && (35)+\frac{1}{2}\cdot\Big((32)+(33)+(34)\Big)-(41),
\end{aligned}$$

using the fact that the left hand sides are integers, so that summing up and dividing by three yields

$$\sum_{j\in[l]} n_j \ge \left\lceil \frac{3\cdot\left\lceil\frac{3X+Y+Z}{2}\right\rceil + 3\cdot\left\lceil\frac{2X+3Y}{2}\right\rceil + \left\lceil\frac{3Y+2Z}{2}\right\rceil - 7\cdot\Omega}{3} \right\rceil, \tag{42}$$

since the left hand side is an integer. (So either $n(\mathcal{R}(T))$ is at least $\Omega+1\in\mathbb{N}$ or at least as large as the right hand side of Inequality (42), which gives Inequality (39). Technically, if the right hand side of Inequality (42) equals $\Omega+1$, then our assumption $n_1+n_2+n_3+n_4+n_5+n_6+n_7 \le \Omega$ yields a contradiction. So, in any case, Inequality (39) is a valid inequality.) □

Example 3.11. For $(X,Y,Z)=(80,105,120)$ and a corresponding function $T\colon 2^{[3]}\to\mathbb{N}$ Inequality (26) and Inequality (27) both yield $n(\mathcal{R}(T))\ge 169$. Inequality (28) yields $n(\mathcal{R}(T))\ge 158$, while Proposition 3.10 yields $n(\mathcal{R}(T))\ge 170$, which is indeed attained.[2] For $(X,Y,Z)=(80,104,121)$ Proposition 3.10 yields $n(\mathcal{R}(T))\ge 167$ and Inequality (26) yields $n(\mathcal{R}(T))\ge 169$, which is indeed attained.

We remark that

$$\Big\{\{4,5,7\},\{2,6,7\},\{1,3,7\},\{3,4,5\},\{1,3,6\},\{2,5,6\},\{1,2,4\}\Big\}$$

is a $2-(7,3,1)$ design, i.e., isomorphic to the Fano plane.

Proposition 3.12. *If X,Y,Z are integers with $X\le Y\le 2X$, $Y\le Z\le \left\lfloor\frac{3Y}{2}\right\rfloor$, and $3X-3Y+Z\ge 0$ then for $T\colon 2^{[3]}\to\mathbb{N}$ depending on X, Y, and Z as described above we have*

$$n(\mathcal{R}(T)) = \max\left\{\Omega, \left\lceil \frac{3\cdot\left\lceil\frac{3X+Y+Z}{2}\right\rceil + 3\cdot\left\lceil\frac{2X+3Y}{2}\right\rceil + \left\lceil\frac{3Y+2Z}{2}\right\rceil - 7\cdot\Omega}{3} \right\rceil\right\} \le \Omega+1,$$

[2] This instance is an example of a non-IRUP instance, see e.g. [6], i.e., the optimal target value of the ILP is strictly larger than the uprounded optimal target value of the corresponding LP relaxation. In general, this happens frequently for ILPs. The interesting case is when the ILP gap, i.e., the difference between both values, is small. So far, it is unclear what happens in our problem.

where $\Omega = \left\lceil \frac{3X+3Y+Z}{4} \right\rceil$.

Proof. Using $3X - 3Y + Z \ge 0$ we conclude that an optimal solution of the symmetrized LP relaxation of the ILP of Corollary 1.10, i.e., constraint (36)-(38), is given by

$$\begin{aligned} N_1 &= \frac{-X+Y+Z}{4}, \\ N_2 &= \frac{X+Y-Z}{4}, \\ N_3 &= \frac{3X-3Y+Z}{4} \end{aligned}$$

with target value $\frac{3X+3Y+Z}{4}$. (I.e., all constraints (36)-(38) are tight. Since we do not need the optimality but only feasibility, we will only show the latter.) Since $Y \ge X$ we have $N_1 \ge 0$. From $2X - Y \ge 0$ and $3Y - 2Z \ge 0$ we conclude $2N_2 = 2X + 2Y - 2Z \ge 0$. Let $\mathbf{r} = (r_1, r_2, r_3) \in \{0, 1, 2, 3\}^3$ be defined by $r_1 \equiv 4N_1$, $r_2 \equiv 4N_2$, and $r_3 \equiv 4N_3$. Depending on $\mathbf{r}$ we will define a code by $(s_1\, s_2\, s_4)\,(s_3\, s_5\, s_6)\,(s_7)$, where $s_i \in \left\{ <, =, > \right\}$. The translation to $(n_1\, n_2\, n_4)\,(n_3\, n_5\, n_6)\,(n_7)$ is as follows: "$<$" means downrounding, "$>$" means uprounding, and "$=$" means no rounding of N_1, N_2, or N_3 for the first, second or third block, respectively.

(1) $(0,0,0)$: $(===)(===)(=)$
(2) $(0,2,0)$: $(===)(<>>)(=)$
(3) $(1,1,1)$: $(<<<)(<>>)(>)$
(4) $(1,3,1)$: $(<<<)(>>>)(>)$
(5) $(2,0,2)$: $(>>>)(===)(<)$
(6) $(2,2,2)$: $(<<>)(<>>)(>)$
(7) $(3,1,3)$: $(>>>)(<<<)(>)$
(8) $(3,3,3)$: $(<>>)(>>>)(>)$

Note that in the cases (1), (2), (4), (6), (7), and (8) the proposed value of $n(\mathcal{R}(T))$ equals Ω and $n = \sum_{j\in[l]} n_j$ also equals Ω. The lower bound $n(\mathcal{R}(T)) \ge \Omega$ is given by Proposition 2.12, i.e., Inequality (26). In the remaining cases (3) and (5) proposed value of $n(\mathcal{R}(T))$ equals $\Omega+1$ and $n = \sum_{j\in[l]} n_j$ also equals $\Omega + 1$. The lower bound $n(\mathcal{R}(T)) \ge \Omega + 1$ is given by Proposition 3.10, where we have to check that the stated lower bound indeed equals $\Omega + 1$ in the mentioned cases.

Thus, it remains to show that the desired service rate region is attained. Since the chosen $\mathbf{n}$ is integral and satisfies the constraints of Corollary 1.10 in all cases we can apply Theorem 3.5 to every vector λ of the desired service region (or to every vector of a generating set, where it is not necessary to known how those vectors look like). □

(X,Y,Z)	(n_1,n_2,n_4)	(n_3,n_5,n_6)	(n_7)	lb	$n(\mathcal{R}(T))$
$(80,104,120)$	$(36,36,36)$	$(16,16,16)$	(12)	168	168
$(80,104,121)$	$(36,36,37)$	$(15,16,16)$	(13)	169	169
$(80,104,122)$	$(36,36,37)$	$(15,16,16)$	(13)	169	169
$(80,104,123)$	$(37,37,37)$	$(15,15,15)$	(13)	169	169
$(80,105,120)$	$(36,36,36)$	$(16,17,17)$	(12)	169	170
$(80,105,121)$	$(36,37,37)$	$(16,16,16)$	(12)	169	170
$(80,105,122)$	$(36,37,37)$	$(16,16,16)$	(12)	170	170
$(80,105,123)$	$(37,37,37)$	$(15,16,16)$	(12)	170	170
$(80,106,120)$	$(36,36,37)$	$(16,16,17)$	(11)	170	170
$\ldots$					
$(83,107,123)$	$(36,37,37)$	$(17,17,17)$	(13)	174	174

Proposition 3.13. *If X, Y, Z are integers with $X \le Y \le 2X$, $Y \le Z \le \lfloor \frac{3Y}{2} \rfloor$, and $3X - 3Y + Z < 0$, then for $T\colon 2^{[3]} \to \mathbb{N}$ depending on X, Y, and Z as described above we have*

$$n(\mathcal{R}(T)) = \left\lceil \frac{3Y}{2} \right\rceil .$$

Proof. Using $3X - 3Y + Z \le -1$ we conclude that an optimal solution of the symmetrized LP relaxation of the ILP of Corollary 1.10, i.e., constraint (36)-(38), is given by

$$\begin{aligned} N_1 &= Y - X, \\ N_2 &= \frac{2X - Y}{2}, \\ N_3 &= 0 \end{aligned}$$

with target value $\frac{3Y}{2}$. (I.e., the constraints (36), (37), and $N_3 \ge 0$ are tight. Since we do not need the optimality but only feasibility, we will only show the latter.) Since $X \le Y$ we have $N_1 \ge 0$ and since $Y \le 2X$ we have $N_2 \ge 0$.

If Y is even, then we choose the multiset of points $\mathcal{G}$ by setting $n_1 = n_2 = n_4 = N_1 = Y - X \in \mathbb{N}$, $n_3 = n_5 = n_6 = X - \frac{Y}{2} \in \mathbb{N}$, and $n_7 = 0$, so that $\sum_{i \in [7]} n_i = \frac{3Y}{2}$. If Y is odd, then we choose the multiset of points $\mathcal{G}$ by setting $n_1 = n_2 = n_4 = N_1 = Y - X \in \mathbb{N}$, $n_3 = n_5 = X - \frac{Y-1}{2} \in \mathbb{N}$, $n_6 = X - \frac{Y+1}{2} \in \mathbb{N}$, and $n_7 = 0$, so that $\sum_{i \in [7]} n_i = \frac{3Y+1}{2} = \lceil \frac{3Y}{2} \rceil$.

In both cases $\mathbf{n}$ is an integral solution of the ILP of Corollary 1.10 so that we can apply Theorem 3.5 to conclude $n(\mathcal{R}(T)) \le \lceil \frac{3Y}{2} \rceil$. The matching lower bound is given by Inequality (28) based on Proposition 2.15. □

An example where Proposition 3.13 can be applied is given by $(X, Y, Z) = (5, 9, 9)$.

Theorem 3.14. *If X, Y, Z are integers with $X \le Y \le 2X$ and $Y \le Z \le \lfloor \frac{3Y}{2} \rfloor$, and $3X - 3Y + Z < 0$, then for $T\colon 2^{[3]} \to \mathbb{N}$ given by*

$$T(U) = \begin{cases} 0 & : \ \#U = 0, \\ X & : \ \#U = 1, \\ Y & : \ \#U = 2, \\ Z & : \ \#U = 3 \end{cases}$$

we have

$$\begin{aligned} n(\mathcal{R}(T)) &= \max\left\{ \Omega, \left\lceil \frac{3 \cdot \lceil \frac{3X+Y+Z}{2} \rceil + 3 \cdot \lceil \frac{2X+3Y}{2} \rceil + \lceil \frac{3Y+2Z}{2} \rceil - 7 \cdot \Omega}{3} \right\rceil, \left\lceil \frac{3Y}{2} \right\rceil \right\} \\ &\le \max\left\{ \Omega + 1, \left\lceil \tfrac{3Y}{2} \right\rceil \right\}, \end{aligned}$$

where $\Omega = \lceil \frac{3X+3Y+Z}{4} \rceil$.

Proof. We apply Proposition 3.12 and Proposition 3.13 noting that the corresponding lower bounds are valid in both cases. □

We remark that for $(X, Y, Z) = (n, 2n, 2n)$ we have

$$\left\lceil \frac{3Y}{2} \right\rceil = 3n \quad \text{and} \quad \Omega + 1 = \left\lceil \frac{11n}{4} \right\rceil + 1$$

and for $(X, Y, Z) = (2, 2n, 2n)$ we have

$$\left\lceil \frac{3Y}{2} \right\rceil = 3n \quad \text{and} \quad \Omega + 1 = \left\lceil \frac{7n}{2} \right\rceil + 1.$$

REFERENCES

[1] C. Bachoc, D. C. Gijswijt, A. Schrijver, and F. Vallentin. Invariant semidefinite programs. In *Handbook on semidefinite, conic and polynomial optimization*, pages 219–269. Springer, 2012.
[2] R. Bödi, K. Herr, and M. Joswig. Algorithms for highly symmetric linear and integer programs. *Mathematical Programming*, 137(1-2):65–90, 2013.
[3] J. Derks and J. Kuipers. On the number of extreme points of the core of a transferable utility game. In *Chapters in game theory*, pages 83–97. Springer, 2002.
[4] D. Gale. Neighborly and cyclic polytopes. In *Proc. Sympos. Pure Math*, volume 7, pages 225–232, 1963.
[5] K. Gatermann and P. A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *Journal of Pure and Applied Algebra*, 192(1-3):95–128, 2004.
[6] V. M. Kartak, A. V. Ripatti, G. Scheithauer, and S. Kurz. Minimal proper non-IRUP instances of the one-dimensional cutting stock problem. *Discrete Applied Mathematics*, 187:120–129, 2015.
[7] F. Kazemi, S. Kurz, and E. Soljanin. A geometric view of the service rates of codes problem and its application to the service rate of the first order Reed-Muller codes. *arXiv preprint 2001.09121*, 2020.
[8] F. Kazemi, S. Kurz, E. Soljanin, and A. Sprintson. Efficient storage schemes for desired service rate regions. *arXiv preprint arXiv:2010.12614*, 2020.
[9] S. Kurz and E. Yaakobi. PIR codes with short block length. *arXiv preprint 2001.03433*, 2020.
[10] P. McMullen. The maximum numbers of faces of a convex polytope. *Mathematika*, 17(2):179–184, 1970.
[11] M. Noori, E. Soljanin, and M. Ardakani. On storage allocation for maximum service rate in distributed storage systems. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 240–244. IEEE, 2016.

SASCHA KURZ, UNIVERSITY OF BAYREUTH, 95440 BAYREUTH, GERMANY
Email address: sascha.kurz@uni-bayreuth.de