3 Graver Bases & Nonlinear Integer Programming

In this chapter we introduce the Graver basis, an object which is of fundamental importance for linear and nonlinear integer programming. Here S is presented as the set of integer points satisfying a given system of linear inequalities. Without loss of generality we may and will usually assume that S is given in *standard form*

$$S \quad := \quad \{x \in \mathbb{Z}^n \ : \ Ax = b,\ l \le x \le u\} \ , \tag{11}$$

where A is an integer $m \times n$ matrix, $b \in \mathbb{Z}^m$, and $l, u \in \mathbb{Z}^n_\infty$. So, we consider nonlinear discrete optimization problems of the form

$$\min / \max \{f(Wx) \ : \ x \in \mathbb{Z}^n,\ Ax = b,\ l \le x \le u\} \ , \tag{12}$$

with W an integer $d \times n$ matrix (with the useful special case of $W = I_n$ the $n \times n$ identity matrix) and $f : \mathbb{R}^d \to \mathbb{R}$ a function typically given by a comparison oracle.

In §3.1 we define the Graver basis of an integer matrix, show that in can be computed by a finite algorithm, and establish some of its basic properties that will be needed in later sections. In §3.2 we establish results of [19, 41] showing that the Graver basis enables polynomial time integer minimization of linear and separable convex functions. In §3.3 we establish a result of [20] showing that, moreover the Graver basis also enables polynomial time integer maximization of convex functions. In §3.4 we specialize these results to nonlinear optimization over totally unimodular systems and describe applications to vector partitioning and clustering.

3.1 Graver Bases and Circuits

In this subsection we define and discuss some basic properties of *Graver bases* and *circuits* of integer matrices, that will be used in our algorithmic results later on. Throughout this subsection let A be an integer $m \times n$ matrix. The *lattice* of A is the set $\mathcal{L}(A) := \{x \in \mathbb{Z}^n \ : \ Ax = 0\}$ of integer vectors in its kernel. We will use $\mathcal{L}^*(A) := \{x \in \mathbb{Z}^n \ : \ Ax = 0,\ x \ne 0\}$ to denote the set of nonzero elements in $\mathcal{L}(A)$.

Throughout these lecture notes we will make use of a partial order $\sqsubseteq$ on $\mathbb{R}^n$ which extends the usual coordinate-wise partial order $\le$ on the nonnegative orthant $\mathbb{R}^n_+$ and is defined as follows. For two vectors $x, y \in \mathbb{R}^n$ we write $x \sqsubseteq y$ and say that x is *conformal* to y if $x_i y_i \ge 0$ and $|x_i| \le |y_i|$ for $i = 1, \ldots, n$, that is, x and y lie in the same orthant of $\mathbb{R}^n$ and each component of x is bounded by the corresponding component of y in absolute value. A suitable extension of the classical lemma of Gordan [31] implies that every subset of $\mathbb{Z}^n$ has finitely-many $\sqsubseteq$-minimal elements. The following object, of fundamental importance, was introduced in [32].

Definition 3.1 The *Graver basis* of an integer matrix A is defined to be the finite set $\mathcal{G}(A) \subset \mathbb{Z}^n$ of $\sqsubseteq$-minimal elements in $\mathcal{L}^*(A) = \{x \in \mathbb{Z}^n \ : \ Ax = 0,\ x \ne 0\}$.

The Graver basis of a matrix is typically a very large set and is hard to compute. Yet, as we show below, the length of its radius is polynomially bounded in the length of A, which implies a simple finite (though exponential) procedure for computing it. Moreover, in Chapter 4 we will show that for broad and fundamental classes of integer programs we can compute the Graver basis in polynomial time.

A *circuit* of an integer matrix A is an element $c \in \mathcal{L}^*(A)$ whose support $\mathrm{supp}(c)$ is minimal with respect to inclusion and whose entries are relatively prime integers. We denote the set of circuits of A by $\mathcal{C}(A)$. Note that A has at most $2\sum_{k=0}^{m}\binom{n}{k+1}$ circuits, a bound depending only on m and n and independent of the entries of A.

Note also that both $\mathcal{G}(A)$ and $\mathcal{C}(A)$ are centrally symmetric, that is, $g \in \mathcal{G}(A)$ if and only if $-g \in \mathcal{G}(A)$, and $c \in \mathcal{C}(A)$ if and only if $-c \in \mathcal{C}(A)$.

It follows directly from the definitions that the set of circuits of a matrix is contained in its Graver basis, that is $\mathcal{C}(A) \subseteq \mathcal{G}(A)$. The converse is typically false. For instance, the set of circuits and the Graver basis of the matrix $A := [1\ \ 2\ \ 1]$ are

$$\mathcal{C}(A) \ = \ \pm\{(2,-1,0),(0,-1,2),(1,0,-1)\}\ , \quad \mathcal{G}(A) \ = \ \mathcal{C}(A) \cup \pm\{(1,-1,1)\}\ .$$

Unlike the set of circuits, the cardinality of the Graver basis cannot be bounded in terms of m and n only and does depend on the entries of A. Nonetheless, we proceed to show that the Graver basis can be bounded in terms of the set of circuits.

A finite sum $u := \sum_i v_i$ of vectors in $\mathbb{R}^n$ is called a *conformal sum* if all summands lie in the same orthant and hence $v_i \sqsubseteq u$ for all i.

In the following sequence of lemmas we denote by r the rank of the matrix A.

Lemma 3.2 *Any nonzero rational $x \in \mathbb{R}^n$ satisfying $Ax = 0$ is a conformal sum $x = \sum_{i=1}^t \lambda_i c_i$ involving $t \leq n - r$ linearly independent circuits $c_i \in \mathcal{C}(A)$ with nonnegative rational coefficients λ_i and with* $\mathrm{supp}(c_i) \nsubseteq \bigcup_{j>i} \mathrm{supp}(c_j)$ *for each i.*

Proof. First, we show that for any such x there are $c \in \mathcal{C}(A)$ and nonnegative λ with $\lambda c \sqsubseteq x$. Suppose indirectly this is false and let x be a counterexample with minimal support. Then there must exist an $h \in \mathcal{C}(A)$ with $\mathrm{supp}(h) \subsetneq \mathrm{supp}(x)$ and hence there exists a μ with $y := x - \mu h \sqsubset x$ and $\mathrm{supp}(y) \subsetneq \mathrm{supp}(x)$. Since $y \neq 0$ and $Ay = 0$, there are $c \in \mathcal{C}(A)$ and nonnegative λ with $\lambda c \sqsubseteq y \sqsubseteq x$, a contradiction.

We proceed to prove that every such x is a conformal sum involving circuits with $\mathrm{supp}(c_i) \nsubseteq \bigcup_{j>i} \mathrm{supp}(c_j)$ for each i. This in particular implies that the c_i are linearly independent, and, since all circuits lie in the orthogonal complement of the row space of A, that at most $n-r$ circuits are involved. Suppose indirectly this is not the case and let x be a counterexample with minimal support. By what we just proved, there are $c_0 \in \mathcal{C}(A)$ and $\lambda_0 \in \mathbb{R}_+$ with $\lambda_0 c_0 \sqsubseteq x$ and hence $y := x - \lambda_0 c_0 \sqsubseteq x$. Clearly, we can choose λ_0 so that, furthermore, y satisfies $\mathrm{supp}(y) \subsetneq \mathrm{supp}(x)$. Then y is a conformal sum $y = \sum_{i=1}^s \lambda_i c_i$ involving circuits c_i with $\mathrm{supp}(c_i) \nsubseteq \bigcup_{j>i} \mathrm{supp}(c_j)$ for each i. But then $x = \lambda_0 c_0 + \sum_{i=1}^s \lambda_i c_i$ is a conformal sum involving circuits with $\mathrm{supp}(c_i) \nsubseteq \bigcup_{j>i} \mathrm{supp}(c_j)$ for each i, which is a contradiction. $\square$

Let $\Delta(A)$ denote the maximum absolute value of a subdeterminant of A, namely, of the determinant of a square submatrix of A. Then the Hadamard bound implies $\Delta(A) \leq (\sqrt{m}\|A\|_\infty)^m$, where $\|A\|_\infty = \max_{i,j}|A_{i,j}|$, and therefore it follows that $\langle \Delta(A) \rangle = O(m(\log m + \log \|A\|_\infty))$ is polynomially bounded in $\langle A \rangle$. The following useful bound on the size of components of circuits follows from Cramer's rule.

Proposition 3.3 *Any circuit $c \in \mathcal{C}(A)$ satisfies $\|c\|_\infty = \max_i |c_i| \leq \Delta(A)$.*

We have the following analogous bound for the elements of the Graver basis.

Lemma 3.4 *Any Graver basis element $g \in \mathcal{G}(A)$ satisfies $\|g\|_\infty \leq (n-r)\Delta(A)$.*

Proof. Consider any $g \in \mathcal{G}(A)$. By Lemma 3.2 we have that $g = \sum_{i=1}^t \lambda_i c_i$ is a conformal sum involving $t \leq n-r$ circuits $c_i = (c_{i,1}, \ldots, c_{i,n}) \in \mathcal{C}(A)$ with $\lambda_i \in \mathbb{R}_+$. Now each $\lambda_i \leq 1$ else $c_i \sqsubset g$ and $c_i \neq g$, contradicting g being in the Graver basis. Consider any $1 \leq j \leq n$. The sum being conformal implies that $c_{1,j}, \ldots, c_{t,j}$ all have the same sign. Hence we have the following inequality, which completes the proof,

$$|g_j| \ = \ |\sum_{i=1}^t \lambda_i c_{i,j}| \ = \ \sum_{i=1}^t \lambda_i |c_{i,j}| \ \leq \ t \max\{\|c_i\|_\infty : 1 \leq i \leq t\} \ \leq \ (n-r)\Delta(A) \ . \quad \square$$

Lemma 3.4 implies the following simple finite, though exponential, generic algorithm for computing the Graver basis of any matrix A.

Algorithm 3.5 (generic algorithm for computing Graver bases).

1. Construct the following finite subset of $\mathcal{L}(A)$,

$$L \ := \ \{x \in \mathbb{Z}^n \ : \ \|x\|_\infty \leq (n-r)\Delta(A) \, , \ Ax = 0\} \ .$$

2. Distill out the Graver basis $\mathcal{G}(A)$ as the set of $\sqsubseteq$-minimal elements in $L \setminus \{0\}$.

We proceed to provide an integer, Graver basis, analogue of Lemma 3.2, which is an important ingredient in the algorithms developed in the following two sections. We start with the following weaker form of the statement.

Lemma 3.6 *Any $x \in \mathcal{L}^*(A)$ is a conformal sum $x = \sum_i g_i$ of Graver basis elements $g_i \in \mathcal{G}(A)$, with some elements possibly appearing more than once in the sum.*

Proof. We use induction on the well partial order $\sqsubseteq$. Consider any $x \in \mathcal{L}^*(A)$. If it is $\sqsubseteq$-minimal in $\mathcal{L}^*(A)$ then $x \in \mathcal{G}(A)$ by definition of the Graver basis and we are done. Otherwise, there is a $g \in \mathcal{G}(A)$ such that $g \sqsubset x$ and $g \neq x$. Set $y := x - g$. Then $y \in \mathcal{L}^*(A)$, $y \sqsubset x$ and $y \neq x$, so by induction there is a conformal sum $y = \sum_i g_i$ with $g_i \in \mathcal{G}(A)$ for all i. Now $x = g + \sum_i g_i$ is a conformal sum of x. $\square$

We now provide the stronger form of Lemma 3.6, which is the integer analogue of Lemma 3.2, where the Graver basis replaces the set of circuits.

Lemma 3.7 *Any $x \in \mathcal{L}^*(A)$ is a conformal sum $x = \sum_{i=1}^t \lambda_i g_i$ involving $t \leq 2n-2$ Graver basis elements $g_i \in \mathcal{G}(A)$ with nonnegative integer coefficients $\lambda_i \in \mathbb{Z}_+$.*

Proof. We prove the slightly weaker bound $t \leq 2n-1$ from [13]. A proof of the stronger bound can be found in [68]. Consider any $x \in \mathcal{L}^*(A)$ and let $g_1, \ldots, g_s$ be all elements of $\mathcal{G}(A)$ lying in the same orthant as x. Consider the linear program

$$\max\left\{\sum_{i=1}^{s} \lambda_i \ : \ x = \sum_{i=1}^{s} \lambda_i g_i\,, \quad \lambda_i \in \mathbb{R}_+\right\}\ . \tag{13}$$

By Lemma 3.6 the point x is a nonnegative linear combination of the g_i and hence the program (13) is feasible. Since all g_i are nonzero and in the same orthant as x, program (13) is also bounded. As is well known, it then has a *basic* optimal solution, that is, an optimal solution $\lambda_1, \ldots, \lambda_s$ with at most n of the λ_i nonzero. Let

$$y \ := \ \sum(\lambda_i - \lfloor\lambda_i\rfloor)g_i \ = \ x - \sum\lfloor\lambda_i\rfloor g_i\ .$$

If $y = 0$ then $x = \sum\lfloor\lambda_i\rfloor g_i$ is a conformal sum of at most n of the g_i and we are done. Otherwise, $y \in \mathcal{L}^*(A)$ and y lies in the same orthant as x, and hence, by Lemma 3.6 again, $y = \sum_{i=1}^{s} \mu_i g_i$ with all $\mu_i \in \mathbb{Z}_+$. Then $x = \sum(\mu_i + \lfloor\lambda_i\rfloor)g_i$ and hence, since the λ_i form an optimal solution to (13), we have $\sum(\mu_i + \lfloor\lambda_i\rfloor) \leq \sum\lambda_i$. Therefore $\sum\mu_i \leq \sum(\lambda_i - \lfloor\lambda_i\rfloor) < n$ with the last inequality holding since at most n of the λ_i are nonzero. Since the μ_i are integer, at most $n-1$ of them are nonzero. So $x = \sum(\mu_i + \lfloor\lambda_i\rfloor)g_i$ is a conformal sum of x involving at most $2n-1$ of the g_i. $\square$

We remark that the smallest possible bound $t(n)$ that could possibly replace $2n-2$ in the statement of Lemma 3.7, sometimes called the *integer Carathéodory number*, is yet unknown. While it has been conjectured in [13] that $t(n) = n$, it was eventually shown in [12] that in fact $t(n) \geq \lfloor\frac{7}{6}n\rfloor > n$ for any dimension $n \geq 6$.

Another useful property of the Graver basis is that it enables to check if a set S of the form (11) is finite or not, without applying the procedure of Lemma 1.3.

Lemma 3.8 *Suppose x is any point in $S := \{z \in \mathbb{Z}^n : Az = b, l \leq z \leq u\}$. Then S is infinite if and only if there is $g \in \mathcal{G}(A)$ such that $l \leq x + \lambda g \leq u$ for all $\lambda \in \mathbb{Z}_+$. In particular, the finiteness of S can be checked in time polynomial in $\langle\mathcal{G}(A), l, u\rangle$.*

Proof. The existence of such g obviously implies that S is infinite. For the converse, suppose S is infinite. Then $P := \{y \in \mathbb{R}^n : Ay = b, l \leq y \leq u\}$ is an unbounded polyhedron and hence, as is well known, it has a recession vector h, that is, a nonzero vector such that $y + \alpha h \in P$ for all $y \in P$ and $\alpha \geq 0$. Scaling h if necessary we may assume it is integer and hence $h \in \mathcal{L}^*(A)$ and is a conformal sum $h = \sum g_i$ of vectors $g_i \in \mathcal{G}(A)$. Then $l \leq x + \lambda h \leq u$ for all $\lambda \in \mathbb{Z}_+$ and hence also $l \leq x + \lambda g \leq u$ for all $\lambda \in \mathbb{Z}_+$ with $g := g_i$ any of the summands. This proves the first statement.

For the algorithmic statement, just note that $l \leq x + \lambda g \leq u$ for all $\lambda \in \mathbb{Z}_+$ if and only if $g_i \leq 0$ whenever $u_i < \infty$ and $g_i \geq 0$ whenever $l_i > -\infty$. $\square$

Recall from §2.3.1 that an augmentation oracle for $S \subset \mathbb{Z}^n$ is one that, queried on $x \in S$ and $w \in \mathbb{Z}^n$, either asserts that x is optimal for the linear optimization problem $\max\{wx : x \in S\}$, or returns a better $\hat{x} \in S$. We conclude this section by noting that the Graver basis of A allows to efficiently realize an augmentation oracle for a set S of integer points satisfying a linear inequality system defined by A.

Lemma 3.9 *There is an algorithm that, given the Graver basis $\mathcal{G}(A)$ of an integer $m \times n$ matrix A, $l, u \in \mathbb{Z}_\infty^n$, and $x, w \in \mathbb{Z}^n$ with $l \leq x \leq u$, in time polynomial in $\langle \mathcal{G}(A), l, u, x, w \rangle$, either asserts x is optimal for the linear optimization problem $\max\{wz : z \in \mathbb{Z}^n, Az = b, l \leq z \leq u\}$ with $b := Ax$, or returns a better point $\hat{x}$.*

Proof. Suppose x is not optimal. Let x^* be any better feasible point and put $h := x^* - x$. Then $Ah = b - b = 0$ so $h \in \mathcal{L}^*(A)$ and hence, by Lemma 3.6, there is a conformal sum $h = \sum_i g_i$ with $g_i \in \mathcal{G}(A)$ for all i. Now, $l \leq x, x + h = x^* \leq u$ and $g_i \sqsubseteq h$ imply that $l \leq x + g_i \leq u$ for all i. Also, $A(x + g_i) = Ax = b$ for all i. Therefore $x + g_i$ is feasible for all i. Now $0 < wx^* - wx = wh = \sum_i wg_i$ implies that $wg_i > 0$ for some g_i in this sum. Therefore $x + g_i$ is feasible and better.

So the algorithm is very simple: if $wg > 0$ for some $g \in G$ with $l \leq x + g \leq u$ then $\hat{x} := x + g$ is an easily computable better point; otherwise, x is optimal. □

Note, however, that the derivation of linear optimization from augmentation provided by Lemma 2.14 was polynomial only in the *unary* length of the radius $\rho(S)$ of the feasible set. In integer programming, for S of the form (11), this is not satisfactory. In the next two sections we provide much stronger results, showing that the Graver basis enables, in fact, to do linear and nonlinear integer optimization in time which is polynomial moreover in the *binary* length of the radius $\rho(S)$.

3.2 Linear and Separable Convex Integer Minimization

Consider the following nonlinear integer minimization problem in standard form, with A an integer $m \times n$ matrix, W an integer $d \times n$ matrix, $b \in \mathbb{Z}^m$, and $l, u \in \mathbb{Z}_\infty^n$,

$$\min\{f(Wx) \;:\; x \in \mathbb{Z}^n,\; Ax = b,\; l \leq x \leq u\} \;. \tag{14}$$

We now show that Graver basis allows to efficiently solve such minimization problems for an important useful class of convex functions. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *separable convex* if $f(x) = \sum_{j=1}^d f_j(x_j)$ where $f_j : \mathbb{R} \to \mathbb{R}$ is a univariate convex function for all j. Clearly, linear functions are separable convex.

We start with two simple lemmas about univariate convex functions. The first lemma establishes a certain *supermodularity* property of such functions.

Lemma 3.10 *Let $f : \mathbb{R} \to \mathbb{R}$ be a univariate convex function, let r be a real number, and let $s_1, \ldots, s_m$ be real numbers satisfying $s_i s_j \geq 0$ for all i, j. Then we have*

$$f\left(r + \sum_{i=1}^m s_i\right) - f(r) \;\geq\; \sum_{i=1}^m \left(f(r + s_i) - f(r)\right) \;.$$

Proof. We use induction on m. The claim holding trivially for $m = 1$, consider $m > 1$. Since all nonzero s_i have the same sign, $s_m = \lambda \sum_{i=1}^m s_i$ for some $0 \leq \lambda \leq 1$.

Then

$$r + s_m = (1-\lambda)r + \lambda\left(r + \sum_{i=1}^{m} s_i\right), \quad r + \sum_{i=1}^{m-1} s_i = \lambda r + (1-\lambda)\left(r + \sum_{i=1}^{m} s_i\right),$$

and so the convexity of f implies

$$\begin{aligned} f(r+s_m) \quad + \quad & f\left(r + \sum_{i=1}^{m-1} s_i\right) \\ & \le \quad (1-\lambda)f(r) + \lambda f\left(r + \sum_{i=1}^{m} s_i\right) + \lambda f(r) + (1-\lambda) f\left(r + \sum_{i=1}^{m} s_i\right) \\ & = \quad f(r) \ + \ f\left(r + \sum_{i=1}^{m} s_i\right). \end{aligned}$$

Subtracting $2f(r)$ from both sides and applying induction, we obtain, as claimed,

$$\begin{aligned} f\left(r + \sum_{i=1}^{m} s_i\right) \quad - \quad & f(r) \\ & \ge \quad f(r+s_m) - f(r) + f\left(r + \sum_{i=1}^{m-1} s_i\right) - f(r) \\ & \ge \quad \sum_{i=1}^{m} \left(f(r+s_i) - f(r)\right) . \quad \square \end{aligned}$$

The second lemma shows that univariate convex functions can be minimized efficiently over an interval of integers using repeated bisections.

Lemma 3.11 *There is an algorithm that, given any two integer numbers $r \le s$ and any univariate convex function $f : \mathbb{Z} \to \mathbb{R}$ given by a comparison oracle, solves in time polynomial in $\langle r, s\rangle$ the following univariate integer minimization problem,*

$$\min \{ f(\lambda) \ : \ \lambda \in \mathbb{Z}, \quad r \le \lambda \le s \}.$$

Proof. If $r = s$ then $\lambda := r$ is optimal. Assume then $r \le s-1$. Consider the integers

$$r \ \le \ \left\lfloor \frac{r+s}{2} \right\rfloor \ < \ \left\lfloor \frac{r+s}{2} \right\rfloor + 1 \ \le \ s \ .$$

Use the oracle of f to compare $f\left(\lfloor \frac{r+s}{2} \rfloor\right)$ and $f\left(\lfloor \frac{r+s}{2} \rfloor + 1\right)$. By the convexity of f:

$$\begin{array}{lcl} f\left(\lfloor \frac{r+s}{2} \rfloor\right) = f\left(\lfloor \frac{r+s}{2} \rfloor + 1\right) & \Rightarrow & \lambda := \lfloor \frac{r+s}{2} \rfloor \text{ is a minimum of } f; \\ f\left(\lfloor \frac{r+s}{2} \rfloor\right) < f\left(\lfloor \frac{r+s}{2} \rfloor + 1\right) & \Rightarrow & \text{the minimum of } f \text{ is in the interval } [r, \lfloor \frac{r+s}{2} \rfloor]; \\ f\left(\lfloor \frac{r+s}{2} \rfloor\right) > f\left(\lfloor \frac{r+s}{2} \rfloor + 1\right) & \Rightarrow & \text{the minimum of } f \text{ is in the interval } [\lfloor \frac{r+s}{2} \rfloor + 1, s]. \end{array}$$

Thus, we either obtain the optimal point, or bisect the interval $[r, s]$ and repeat. So in $O(\log(s-r)) = O(\langle r, s\rangle)$ bisections we find an optimal solution $\lambda \in \mathbb{Z} \cap [r, s]$. $\square$

The next two lemmas follow easily from Lemmas 3.10 and 3.11 respectively. Recall that for $x, y \in \mathbb{R}^n$ we write $x \sqsubseteq y$ and say that x is conformal to y if $x_i y_i \geq 0$ and $|x_i| \leq |y_i|$ for all i. Recall also that a finite sum $u := \sum_i v_i$ of vectors in $\mathbb{R}^n$ is conformal if $v_i \sqsubseteq u$ for all i. The first lemma shows a suitable supermodularity property of separable convex functions with respect to conformal sums.

Lemma 3.12 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be any separable convex function, let $x \in \mathbb{R}^n$ be any point, and let $\sum g_i$ be any conformal sum in $\mathbb{R}^n$. Then the following inequality holds,*

$$f\left(x + \sum g_i\right) - f(x) \ \geq \ \sum \left(f\left(x + g_i\right) - f(x)\right) \ .$$

Proof. Let $f_j : \mathbb{R} \to \mathbb{R}$ be univariate convex functions such that $f(x) = \sum_{j=1}^n f_j(x_j)$. Consider any $1 \leq j \leq n$. Since $\sum g_i$ is a conformal sum, we have $g_{i,j} g_{k,j} \geq 0$ for all i, k and so, setting $r := x_j$ and $s_i := g_{i,j}$ for all i, Lemma 3.10 applied to f_j implies

$$f_j\left(x_j + \sum_i g_{i,j}\right) - f_j(x_j) \ \geq \ \sum_i \left(f_j\left(x_j + g_{i,j}\right) - f_j(x_j)\right) \ . \tag{15}$$

Summing the equations (15) for $j = 1, \ldots, n$, we obtain the claimed inequality. □

The second lemma concerns finding a best improvement step in a given direction.

Lemma 3.13 *There is an algorithm that, given $l, u \in \mathbb{Z}_\infty^n$, $x \in \mathbb{Z}^n$ with $l \leq x \leq u$, $g \in \mathcal{L}^*(A)$ for some integer $m \times n$ matrix A, and convex function $f : \mathbb{Z}^n \to \mathbb{R}$ presented by a comparison oracle, letting $S := \{z \in \mathbb{Z}^n,\ Az = b,\ l \leq z \leq u\}$ with $b := Ax$, solves in time polynomial in $\langle l, u, x, g\rangle$, the following univariate problem,*

$$\min\{f(x + \lambda g) \ : \ \lambda \in \mathbb{Z}_+ ,\ x + \lambda g \in S\} \ . \tag{16}$$

Proof. First, determine $r := \sup\{s \in \mathbb{Z}_+ \ : \ l \leq x + sg \leq u\}$. If $r = \infty$ then terminate asserting that S is infinite. Suppose r is finite. Then for any integer $0 \leq \lambda \leq r$ we have $l \leq x + \lambda g \leq u$ and since $g \in \mathcal{L}(A)$ we have $A(x + \lambda g) = Ax = b$, and therefore $x + \lambda g \in S$. Now use the algorithm of Lemma 3.11 to determine an integer $\lambda \in \{0, \ldots, r\}$ minimizing the univariate convex function $h(\lambda) := h(x + \lambda g)$. □

We can now prove the important special case of problem (14) with $W = I_n$ the identity matrix and f a separable convex function on $\mathbb{Z}^n$, that is, $f(x) = \sum_{j=1}^n f_j(x_j)$ where $f_j : \mathbb{Z} \to \mathbb{Z}$ is a univariate convex function for all j. Note that this, in turn, includes any linear objective function $wx = \sum_{j=1}^n w_j x_j$ as a more specialized case.

An outline of the algorithm is provided in Figure 6 below.

Theorem 3.14 *There is an algorithm that, given the Graver basis $\mathcal{G}(A)$ of some integer $m \times n$ matrix A, $l, u \in \mathbb{Z}_\infty^n$, $x \in \mathbb{Z}^n$ with $l \leq x \leq u$, and separable convex function $f : \mathbb{Z}^n \to \mathbb{Z}$ presented by a comparison oracle, solves the problem*

$$\min\{f(z) \ : \ z \in \mathbb{Z}^n ,\ Az = b ,\ l \leq z \leq u\} \ , \quad b := Ax \ , \tag{17}$$

in time which polynomial in the binary-encoding length $\langle \mathcal{G}(A), l, u, x, f(x), f(x^*)\rangle$.

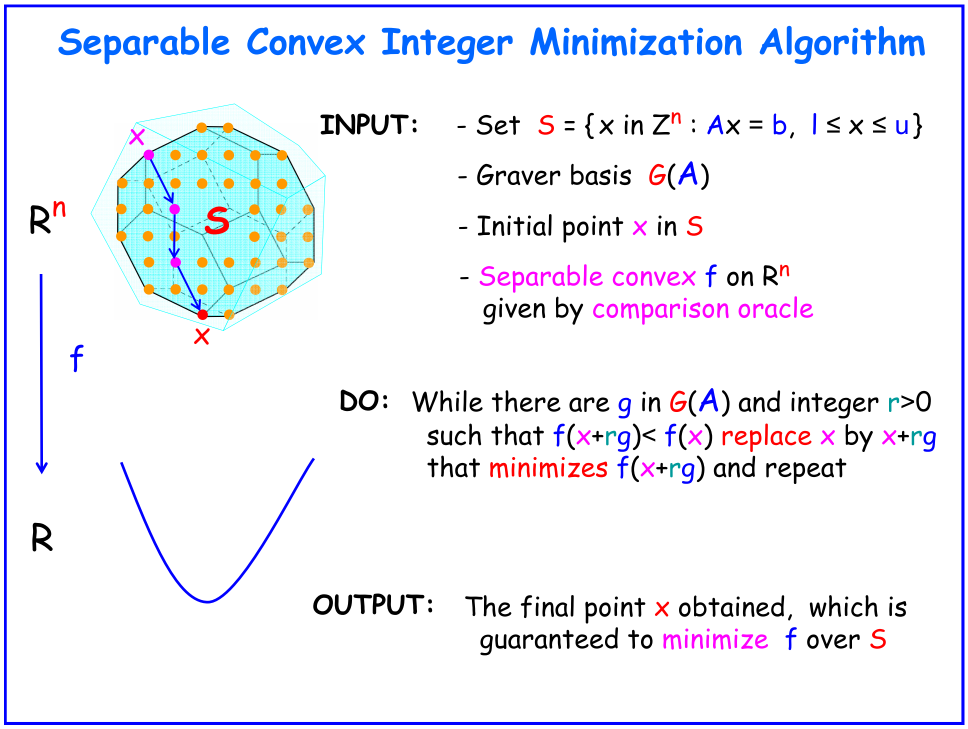

Figure 6: Separable Convex Minimization Using the Graver Basis

Proof. First, apply the algorithm of Lemma 3.8 to $\mathcal{G}(A)$, l, u, and either detect that S is infinite and stop, or conclude that S is finite and continue. Assume S is finite. Produce a sequence of points $x_0, x_1, \ldots, x_s$ in S starting with $x_0 := x$ the given point, as follows. Having determined x_k, solve the univariate minimization problem

$$\min\{f(x_k + \lambda g) \ : \ \lambda \in \mathbb{Z}_+ \, , \ x_k + \lambda g \in S \, , \ g \in \mathcal{G}(A)\,\} \tag{18}$$

by applying the algorithm of Lemma 3.13 for each $g \in \mathcal{G}(A)$. If the minimal value in (18) satisfies $f(x_k + \lambda g) < f(x_k)$ then set $x_{k+1} := x_k + \lambda g$ and repeat, else stop and output the last point x_s in the sequence. Since S is finite and the x_k have decreasing objective values and hence are distinct, the algorithm always terminates.

We now show that the point x_s output by the algorithm is optimal. Let $x^* \in S$ be any optimal solution to (17). Consider any point x_k in the sequence and suppose it is not optimal. We claim that a new point x_{k+1} will be produced and will satisfy

$$f(x_{k+1}) - f\left(x^*\right) \ \leq \ \frac{2n-3}{2n-2}\left(f(x_k) - f(x^*)\right) \ . \tag{19}$$

By Lemma 3.7, we can write the difference $x^* - x_k = \sum_{i=1}^t \lambda_i g_i$ as a conformal sum involving $1 \leq t \leq 2n-2$ elements $g_i \in \mathcal{G}(A)$ with all $\lambda_i \in \mathbb{Z}_+$. By Lemma 3.12,

$$f(x^*) - f\left(x_k\right) \ = \ f\left(x_k + \sum_{i=1}^t \lambda_i g_i\right) - f(x_k) \ \geq \ \sum_{i=1}^t \left(f\left(x_k + \lambda_i g_i\right) - f(x_k)\right) \ .$$

Adding $t\,(f(x_k) - f(x^*))$ on both sides and rearranging terms we obtain

$$\sum_{i=1}^{t} \left(f\left(x_k + \lambda_i g_i\right) - f(x^*)\right) \;\leq\; (t-1)\left(f(x_k) - f(x^*)\right) \;.$$

Therefore there is some summand on the left-hand side satisfying

$$f\left(x_k + \lambda_i g_i\right) - f(x^*) \;\leq\; \frac{t-1}{t}\left(f(x_k) - f(x^*)\right) \;\leq\; \frac{2n-3}{2n-2}\left(f(x_k) - f(x^*)\right) \;.$$

So the point $x_k + \lambda g$ attaining minimum in (18) satisfies

$$f(x_k + \lambda g) - f(x^*) \;\leq\; f\left(x_k + \lambda_i g_i\right) - f(x^*) \;\leq\; \frac{2n-3}{2n-2}\left(f(x_k) - f(x^*)\right) \;,$$

and so indeed $x_{k+1} := x_k + \lambda g$ will be produced and will satisfy (19). This shows that the last point x_s produced and output by the algorithm is indeed optimal.

We proceed to bound the number s of points. Consider any $i < s$ and the intermediate non optimal point x_i in the sequence produced by the algorithm. Then $f(x_i) > f(x^*)$ and, since both values are integer, repeated application of (19) gives

$$\begin{aligned} 1 \leq f(x_i) - f(x^*) \;&=\; \prod_{k=0}^{i-1} \frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)}\left(f(x) - f(x^*)\right) \\ &\leq\; \left(\frac{2n-3}{2n-2}\right)^{i}\left(f(x) - f(x^*)\right) \;, \end{aligned}$$

and therefore

$$i \;\leq\; \left(\log\frac{2n-2}{2n-3}\right)^{-1} \log\left(f(x) - f(x^*)\right) \;.$$

Therefore the number s of points produced by the algorithm is at most one unit larger than this bound, and using a simple bound on the logarithm, we obtain

$$s \;=\; O\left(n \log(f(x) - f(x^*))\right) \;.$$

Thus, the number of points produced and the total running time are polynomial. □

In particular, we can do linear integer programming in polynomial time.

Theorem 3.15 *There is an algorithm that, given an integer $m \times n$ matrix A, its Graver basis $\mathcal{G}(A)$, $l, u \in \mathbb{Z}_\infty^n$, and $x, w \in \mathbb{Z}^n$ with $l \leq x \leq u$, solves in time polynomial in $\langle A, \mathcal{G}(A), l, u, x, w\rangle$, the following linear integer programming problem,*

$$\min\{wz \;:\; z \in \mathbb{Z}^n\,,\; Az = b, l \leq z \leq u\} \;, \quad b := Ax \;.$$

We continue with problem (14) with separable convex f and arbitrary $d \times n$ matrix W. To solve it, we need the Graver basis of an extended, $(m+d) \times (n+d)$ matrix, composed of A, W, the $d \times d$ identity matrix I, and the zero $m \times d$ matrix. Note that, in the next theorem, the number d of rows of W is allowed to be variable.

Theorem 3.16 *There is an algorithm that, given the Graver basis $\mathcal{G}(B)$ of an integer matrix B of the following form, with $m \times n$ block A and $d \times n$ block W,*

$$B \ := \ \begin{pmatrix} A & 0 \\ W & I \end{pmatrix} \ ,$$

$l, u \in \mathbb{Z}_\infty^n$, $x \in \mathbb{Z}^n$ with $l \leq x \leq u$, and separable convex function $f : \mathbb{Z}^d \to \mathbb{Z}$ presented by a comparison oracle, solves the convex integer minimization problem

$$\min\{f(Wy) \ : \ y \in \mathbb{Z}^n \,, \ Ay = b \,, \ l \leq y \leq u\} \ , \quad b := Ax \ , \tag{20}$$

in time which polynomial in the binary-encoding length $\langle \mathcal{G}(B), l, u, x, f(x), f(x^*)\rangle$.

Proof. Define a function $g : \mathbb{Z}^{n+d} \to \mathbb{Z}$ by $g(y, z) := f(-z)$ for all $y \in \mathbb{Z}^n$ and $z \in \mathbb{Z}^d$. Clearly, g is separable convex since f is. Now, problem (20) can be rewritten as

$$\min\{g(y, z) \ : \ (y, z) \in \mathbb{Z}^{n+d} \,, \ \begin{pmatrix} A & 0 \\ W & I \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} \,, \ \ l \leq y \leq u\} \ ,$$

and the statement follows at once by applying Theorem 3.14 to this problem. □

3.3 Convex Integer Maximization

We proceed to discuss the maximization of a convex function over a set given by inequalities in standard form, with A an integer $m \times n$ matrix, $b \in \mathbb{Z}^m$, and $l, u \in \mathbb{Z}_\infty^n$,

$$S \quad := \quad \{x \in \mathbb{Z}^n \ : \ Ax = b, \ l \leq x \leq u\} \ .$$

Consider the polyhedron P of rational points defined by the same system of inequalities and its so-called *integer hull* $P_I := \text{conv}(P \cap \mathbb{Z}^n) = \text{conv}(S)$,

$$P_I \ = \ \text{conv}\{x \in \mathbb{Z}^n \ : \ Ax = b, \ l \leq x \leq u\} \ \subseteq \ \{x \in \mathbb{R}^n \ : \ Ax = b, \ l \leq x \leq u\} \ = \ P \,.$$

We now show that the set of circuits and the Graver basis of A cover the edge-directions of P and P_I, respectively.

Lemma 3.17 *For every integer $m \times n$ matrix A, $l, u \in \mathbb{Z}_\infty^n$, and $b \in \mathbb{Z}^m$, the set of circuits $\mathcal{C}(A)$ of A covers all edge-directions of $P = \{x \in \mathbb{R}^n : Ax = b, \ l \leq x \leq u\}$.*

Proof. Consider any edge e of P. Pick two distinct rational points $x, y \in e$ and set $g := y - x$. Then a suitable multiple of g is in $\mathcal{L}^*(A)$ and hence it follows from Lemma 3.2 that $g = \sum_i \lambda_i c_i$ is a conformal sum for suitable circuits $c_i \in \mathcal{C}(A)$ and $\lambda_i \in \mathbb{R}_+$. We claim that $x + \lambda_i c_i \in P$ for all i. Indeed, c_i being a circuit implies $A(x + \lambda_i c_i) = Ax = b$, and $l \leq x, x + g \leq u$ and $\lambda_i c_i \sqsubseteq g$ imply $l \leq x + \lambda_i c_i \leq u$.

Now let $w \in \mathbb{R}^n$ be uniquely maximized over P at the edge e. Then $w\lambda_i c_i = w(x + \lambda_i c_i) - wx \leq 0$ for all i. But $\sum w\lambda_i c_i = wg = wy - wx = 0$, implying that in fact $w\lambda_i c_i = 0$ and hence $x + \lambda_i c_i \in e$ for all i. This implies that each c_i is a direction of e (in fact, all c_i are the same and g is a multiple of some circuit). □

Lemma 3.18 *For every integer $m \times n$ matrix A, $l, u \in \mathbb{Z}^n_\infty$, and $b \in \mathbb{Z}^m$, the Graver basis $\mathcal{G}(A)$ of A covers all edge-directions of $P_I = \mathrm{conv}\{x \in \mathbb{Z}^n : Ax = b,\ l \leq x \leq u\}$.*

Proof. Consider any edge e of P_I and pick two distinct integer points $x, y \in e$. Then $g := y - x$ is in $\mathcal{L}^*(A)$ and hence Lemma 3.6 implies that $g = \sum_i h_i$ is a conformal sum for suitable $h_i \in \mathcal{G}(A)$. We claim that $x + h_i \in P_I$ for all i. Indeed, $h_i \in \mathcal{G}(A)$ implies $A(x + h_i) = Ax = b$, and $l \leq x, x + g \leq u$ and $h_i \sqsubseteq g$ imply $l \leq x + h_i \leq u$.

Now let $w \in \mathbb{Z}^n$ be uniquely maximized over P_I at the edge e. Then $wh_i = w(x + h_i) - wx \leq 0$ for all i. But $\sum wh_i = wg = wy - wx = 0$, implying that in fact $wh_i = 0$ and hence $x + h_i \in e$ for all i. This implies that h_i is a direction of e (in fact, all h_i are the same and g is a multiple of some Graver basis element). ∎

Using Theorems 2.12 and 3.15 and Lemma 3.18 we obtain the following theorem.

Theorem 3.19 *For every fixed d there is an algorithm that, given an integer $m \times n$ matrix A, its Graver basis $\mathcal{G}(A)$, $l, u \in \mathbb{Z}^n_\infty$, $x \in \mathbb{Z}^n$ with $l \leq x \leq u$, integer $d \times n$ matrix W, and convex function $f : \mathbb{Z}^d \to \mathbb{R}$ presented by a comparison oracle, solves in time polynomial in $\langle A, \mathcal{G}(A), l, u, x, W\rangle$, the convex integer maximization problem*

$$\max \{f(Wy) \ : \ y \in \mathbb{Z}^n,\ Ay = b,\ l \leq y \leq u\} \ , \quad b := Ax \ .$$

Proof. First, apply the algorithm of Lemma 3.8 to $\mathcal{G}(A)$, l, and u, and either detect that S is infinite and stop, or conclude that S is finite and continue.

Now, the algorithm of Theorem 3.15 allows to simulate in polynomial time a linear optimization oracle for S. Moreover, by Lemma 3.18, the given Graver basis covers all edge-directions of $\mathrm{conv}(S)$. Hence the algorithm of Theorem 2.12 can be applied, and enables to solve the convex integer program in polynomial time. ∎

3.4 Totally Unimodular Matrices, Partitioning & Clustering

We now demonstrate some applications of the results on Graver bases developed above to problems of vector partitioning and clustering. Much more involved applications will be given in Chapter 4 where we discuss n-fold systems. We start by discussing a well known important class of matrices which arise in linear integer programming and in particular in network flows. An integer matrix A is *totally unimodular* if every subdeterminant of A is -1, 0, or 1, that is, $\Delta(A) \leq 1$. The following fundamental result of [42], asserting that polyhedra defined by totally unimodular matrices are integer, implies that linear integer programming over totally unimodular systems can be done in polynomial time using linear programming.

Proposition 3.20 *For totally unimodular $m \times n$ matrix A, $b \in \mathbb{Z}^m$, and $l, u \in \mathbb{Z}^n_\infty$,*

$$P_I \ = \ \mathrm{conv}\{x \in \mathbb{Z}^n \ : \ Ax = b,\ l \leq x \leq u\} \ = \ \{x \in \mathbb{R}^n \ : \ Ax = b,\ l \leq x \leq u\} \ = \ P \ .$$

Therefore, there is an algorithm that, given data as above along with $w \in \mathbb{Z}^n$, solves in time polynomial in $\langle b, l, u, w\rangle$ the linear integer programming problem

$$\max\{wx \ : \ x \in \mathbb{Z}^n\,,\ Ax = b,\ l \leq x \leq u\} \ .$$

We proceed to discuss a certain extension of this fundamental result to convex integer programming where the matrix defining the system of inequalities is totally unimodular. As mentioned before, for any matrix $\mathcal{C}(A) \subseteq \mathcal{G}(A)$ holds. For totally unimodular matrices the converse is true as well, providing a useful link between the theory of totally unimodular matrices and the theory of Graver bases.

Lemma 3.21 *The Graver basis and set of circuits of totally unimodular A coincide,*

$$\mathcal{G}(A) \;=\; \mathcal{C}(A) \; .$$

Proof. By Proposition 3.3, any component of any circuit of A satisfies $|c_i| \leq \Delta(A)$. For totally unimodular A this implies $c_i \in \{-1, 0, 1\}$ for each $c \in \mathcal{C}(A)$. Now consider any $g \in \mathcal{G}(A)$. Since $g \in \mathcal{L}^*(A)$, Lemma 3.2 implies that $g = \sum_{i=1}^{t} \lambda_i c_i$ is a conformal sum involving circuits $c_i \in \mathcal{C}(A)$ that satisfy $\mathrm{supp}(c_i) \nsubseteq \bigcup_{j>i} \mathrm{supp}(c_j)$ and $\lambda_i \in \mathbb{R}_+$ for each i. Pick any $k \in \mathrm{supp}(c_1) \setminus \bigcup_{j>1} \mathrm{supp}(c_j)$. Then $g_k = \lambda_1 c_{1,k}$ and $c_{1,k} = \pm 1$ and hence $\lambda_1 \geq 1$, which implies that $c_1 \sqsubseteq g$. Since g is a Graver basis element, it must be that $g = c_1$ and hence g is a circuit. $\square$

The following theorem asserts that if A is totally unimodular and its set of circuits $\mathcal{C}(A)$ is available then we can maximize convex functions over sets of integer points defined by A efficiently. We provide the proof using Lemma 3.21 and Theorem 3.19. Alternatively, it can be proved using Lemma 3.17 and Theorem 2.12.

Theorem 3.22 *For every fixed d there is an algorithm that, given totally unimodular $m \times n$ matrix A, its set of circuits $\mathcal{C}(A)$, $b \in \mathbb{Z}^m$, $l, u \in \mathbb{Z}^n_\infty$, integer $d \times n$ matrix W, and convex function $f : \mathbb{Z}^d \to \mathbb{R}$ presented by a comparison oracle, solves in time polynomial in $\langle \mathcal{C}(A), l, u, b, W \rangle$ the following convex integer maximization program,*

$$\max\left\{f(Wz) \;:\; z \in \mathbb{Z}^n,\ Az = b,\ l \leq z \leq u\right\} \; .$$

Proof. By Proposition 3.20, we have $P_I = P := \{x \in \mathbb{R}^n : Ax = b,\ l \leq x \leq u\}$, so we can check in polynomial time if the integer program is feasible using linear programming over P, and if it is, obtain an integer vertex x of $P_I = P$. By Lemma 3.21 we have the Graver basis $\mathcal{G}(A) = \mathcal{C}(A)$ and so we can apply the algorithm of Theorem 3.19 to the given data augmented with the computed initial feasible point x, and solve the convex maximization problem in polynomial time as claimed. $\square$

Note that if m is fixed then $\mathcal{C}(A)$ can be computed in polynomial time by checking all $2\sum_{k=0}^{m}\binom{n}{k+1} = O(n^{m+1})$ potential supports of circuits. Even if m grows slowly in terms of n, say $m = O(\log n)$, then $\mathcal{C}(A)$ can be computed in subexponential time. Then, using $\mathcal{C}(A)$, the algorithm of Theorem 3.22 can be applied as a good strategy for convex integer maximization over totally unimodular systems.

We proceed to discuss applications of Theorem 3.22 to vector partitioning. This class of problems concerns the partitioning of n items among p players to maximize social utility. A small example of this problem in the special case of identical players is provided in Figure 7 below. Each player i has an integer $q \times n$ utility matrix U^i

whose k-th column $U^{i,k}$ is the utility of item k to player i under q criteria, with entry $U^i_{j,k}$ the utility to player i under criterion j of item k. The utility matrix of an ordered partition $\pi = (\pi_1, \ldots, \pi_p)$ of the set $\{1, \ldots, n\}$ of items is the $q \times p$ matrix

$$U^\pi \ := \ \left(\sum_{k \in \pi_1} U^{1,k}, \ldots, \sum_{k \in \pi_p} U^{p,k} \right) \ ,$$

whose i-th column is the sum $\sum_{k \in \pi_i} U^{i,k}$ of utility vectors of items assigned to player i under π, and is the total utility to player i under π. The social utility of π is the balancing of the player utilities by a given convex function $f : \mathbb{Z}^{q \times p} \cong \mathbb{Z}^{pq} \to \mathbb{R}$,

$$f(U^\pi) \ := \ f\left(\sum_{k \in \pi_1} U^1_{1,k}, \ldots, \sum_{k \in \pi_1} U^1_{q,k}, \ldots, \sum_{k \in \pi_p} U^p_{1,k}, \ldots, \sum_{k \in \pi_p} U^p_{q,k} \right) \ .$$

In the constrained version, the partition must be of a given *shape*, i.e. the number $|\pi_i|$ of items that player i gets has to be a given positive integer λ_i (with $\sum \lambda_i = n$). In the unconstrained version, the number of items per player is unrestricted.

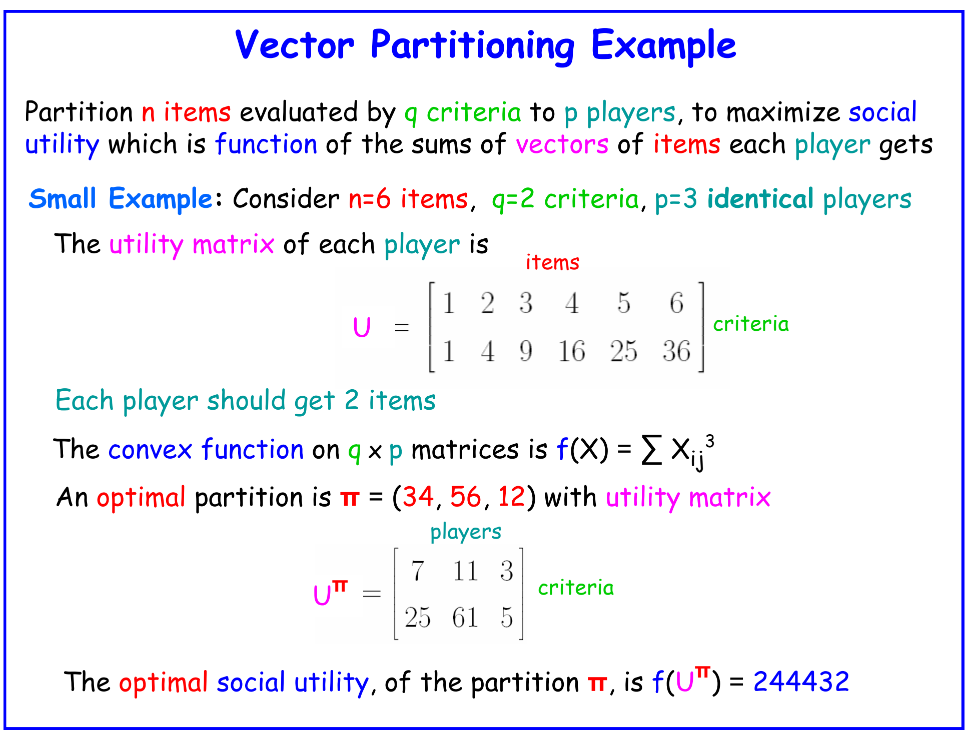

Figure 7: Vector Partitioning Example

If either the number of players p or the number of criteria q is variable, the partition problem is NP-hard. We now show that in contrast, when both p, q are fixed, as a consequence of Theorem 3.22, both the constrained and unconstrained versions of the vector partition problem can be solved in polynomial time.

Corollary 3.23 *For every fixed p and q, there is an algorithm that, given utility matrices $U^1, \ldots, U^p \in \mathbb{Z}^{q\times n}$, $\lambda_1, \ldots, \lambda_p \in \{1, \ldots, n\}$, and convex function $f : \mathbb{Z}^{pq} \to \mathbb{R}$ presented by a comparison oracle, solves the constrained or unconstrained partitioning problem with p players and q criteria, in time polynomial in $\langle U^1, \ldots, U^p\rangle$.*

Proof. We demonstrate only the constrained problem, the unconstrained problem being similar and simpler. There is an obvious one-to-one correspondence between partitions and matrices $x \in \{0,1\}^{n\times p}$ satisfying $\sum_{i=1}^{p} x_{k,i} = 1$ for $k = 1, \ldots, n$, where partition π corresponds to that matrix x with $x_{k,i} = 1$ if $k \in \pi_i$ and $x_{k,i} = 0$ otherwise. Identifying $\mathbb{Z}^{n\times p} \cong \mathbb{Z}^{np}$ using the components order $x = (x_{1,1}, \ldots, x_{1,p}, \ldots, x_{n,1}, \ldots, x_{n,p})$, the set of feasible partitions becomes the set

$$\left\{x \in \mathbb{Z}^{np}_{+} \ : \ \sum_{k=1}^{n} x_{k,i} = \lambda_i\,, \quad i = 1, \ldots, p\,, \qquad \sum_{i=1}^{p} x_{k,i} = 1\,, \quad k = 1, \ldots, n\right\}\,, \qquad (21)$$

of integer points of a suitable transportation problem. Let A denote the $(p+n)\times np$ matrix defining the system of equations in (21). Then A is the adjacency matrix of the complete bipartite graph $K_{p,n}$ and is well known to be totally unimodular. Moreover, $c \in \mathbb{Z}^{np} \cong \mathbb{Z}^{n\times p}$ is a circuit of A if and only if it is supported on the set of edges of a circuit of $K_{p,n}$ with values ± 1 alternating along the circuit. Since p is fixed, the number of circuits of $K_{p,n}$ satisfies $\sum_{i=2}^{p} \binom{p}{i}\binom{n}{i} i!(i-1)! = O(n^p)$, and therefore the set $\mathcal{C}(A)$ of circuits of A can be constructed in polynomial time.

Now define an integer $pq \times np$ matrix W by setting $W_{(i,j),(k,i)} := U^i_{j,k}$ for $i = 1, \ldots, p$, $j = 1, \ldots, q$, $k = 1, \ldots, n$, and setting all other entries to zero. Then for any partition π and its corresponding vector $x \in \{0,1\}^{np}$ we have $\sum_{k\in\pi_i} U^i_{j,k} = (Wx)_{(i,j)}$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. Thus, the vector partitioning problem becomes the following convex integer program with W having a fixed number $d := pq$ of rows,

$$\max\left\{f(Wx) \ : \ x \in \mathbb{Z}^{np}_{+}\,, \ \sum_{k=1}^{n} x_{k,i} = \lambda_i\,, \ i = 1, \ldots, p\,, \ \sum_{i=1}^{p} x_{k,i} = 1\,, \ k = 1, \ldots, n\right\}\,.$$

We can now apply the algorithm of Theorem 3.22 to data consisting of the totally unimodular matrix A defining the system (21), its set of circuits $\mathcal{C}(A)$ computed in polynomial time as explained above, $l := 0$, $u := \mathbf{1}$ and $b := (\lambda_1, \ldots, \lambda_p, 1, \ldots, 1)$, and solve the vector partitioning problem in polynomial time as claimed. □

We conclude with a concrete important example of a vector partitioning problem.

Example 3.24 (minimal variance clustering). This problem has numerous applications in the analysis of statistical data: group n observed points $u^1, \ldots, u^n$ in $\mathbb{R}^q$ into p clusters $\pi_1, \ldots, \pi_p$ so as to minimize the sum of cluster variances

$$\sum_{i=1}^{p} \frac{1}{|\pi_i|} \sum_{k\in\pi_i} \left\| u^k - \left(\frac{1}{|\pi_i|} \sum_{k\in\pi_i} u^k\right)\right\|^2 \,.$$

Consider instances where there are $n = pc$ points and the desired clustering is balanced, that is, the clusters should have equal size c. Assume the observation points are rational, and then suitably scale them to become integer. Suitable manipulation of the sum of variances expression above shows that the problem is equivalent to a constrained vector partitioning problem, with p identical players (the clusters) having a common $q \times n$ utility matrix $U := (u^1, \ldots, u^n)$, with $\lambda_i = c$ for all i, and with $f : \mathbb{Z}^{q \times p} \cong \mathbb{Z}^{pq} \to \mathbb{R}$ (to be maximized) being the Euclidean norm squared,

$$f(z) \ := \ \|z\|^2 \ := \ \sum_{j=1}^{q} \sum_{i=1}^{p} |z_{j,i}|^2 \ .$$

By Corollary 3.23 we can cluster optimally in polynomial time for all fixed p and q.