Università degli Studi di Padova
Dipartimento di Matematica "Tullio Levi-Civita"
Corso di Laurea Magistrale in Matematica

Markov bases and linear algebra: connecting fibres of integer linear maps

Relatore:
Prof. Marco Andrea Garuti

Laureando: **Giorgio Spadaccini**
Matricola: **1237005**

Anno Accademico 2020/2021

21 Luglio 2021

Contents

Introduction

Background

This thesis explores the mathematical background of the algorithm presented by Diaconis and Sturmfels in their paper [1], the purpose of which was to extract random vectors from a finite, yet very large subset of $\mathbb{N}^r$, according to any given probability distribution π over such set. To better understand the scenario that we want to explore, let us see a simplified version of the example on pages 1 and 2 of [1]: consider three natural numbers $a, b, c \in \mathbb{N}$ and take $d := a + b - c$. Then we can define the set:

$$\mathcal{M}(a,b,c) := \{\begin{pmatrix} x & z \\ y & t \end{pmatrix} \in \mathbb{N}^{2\times 2} : x + z = a, y + t = b, x + y = c, z + t = d\}$$

of all the natural tables with fixed row and column sums. If we stack up the columns into one vector, we can express this as:

$$\mathcal{M}'(a,b,c) := \{\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \in \mathbb{N}^4 : A \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}\} \qquad \text{with } A := \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Now suppose that we needed to extract a vector at random from this set $\mathcal{M}'(a,b,c)$. If a, b, c are small, then so is $|\mathcal{M}'(a,b,c)|$, therefore it would be very reasonable to simply label all its elements with a natural number and then randomly extract a label according to the distribution π. However, memorizing exact integers means that the computer cannot use a floating-point notation to store such numbers: instead, it stores them by simply writing them in binary notation. This means that 64 bits can only represent the integers from $-(2^{63} - 1)$ to $2^{63} - 1$ or, more reasonably in our case, from 0 to $2^{64} - 1 \approx 1.8 \cdot 10^{19}$. Any set with a cardinality bigger than such quantity could therefore not be labelled with the standard word size of 64 bits.
The article [1] introduces the idea of using a Markov chain to perform such extraction. More specifically, it assumes the vector pool to be in the form of what we will call a "fibre".

Definition 1. *Given a matrix $A \in \mathbb{Z}^{s\times r}$ and a vector $u \in \mathbb{N}^r$, we define the fibre of A containing u as:*

$$\mathcal{F}(u) = \{v \in \mathbb{N}^r : Av = Au\} = (u + \mathrm{Ker}_{\mathbb{Z}}(A)) \cap \mathbb{N}^r.$$

Notice that here the notation $\mathrm{Ker}_{\mathbb{Z}}(A)$ denotes the Kernel of the linear map $f_A : \mathbb{Z}^r \to \mathbb{Z}^s$ induced by A. In other words, it is the integer null space of the matrix.
This assumption is not really much of a restriction, as many conditions can be expressed linearly: for instance, the set $\mathcal{M}(a, b, c)$ depicted above is the 2×2 realization of the more general case of matrices whose row and column sums are constant, a scenario which appears for instance in Fisher's exact test.
Furthermore, even if our pool of vectors cannot be exactly expressed in the form $\mathcal{F}(u)$, it is very easily contained in such a fibre. Given the fact that we are setting the problem in a way that allows us to choose the probability π, we can always decide to set $\pi(v) = 0$ if the vector $v \in \mathcal{F}(u)$ is not in the pool of vectors that we were starting with. This way the pool does not necessarily have to be a fibre, it only has to be contained in one.
Other examples of applications of this method can be found in [2].
After fixing A and defining $\mathcal{F}(u)$, Diaconis and Sturmfels use any given set $\mathcal{B} \subseteq \mathrm{Ker}_{\mathbb{Z}}(A)$ to upgrade fibres into graphs by drawing edges according to the following rule:

$$u \sim v \text{ in } \mathcal{F}(u) \iff u - v \in \pm\mathcal{B} \tag{1}$$

where, as one may imagine, $\pm\mathcal{B}$ denotes the set $\mathcal{B} \cup \{-b : b \in \mathcal{B}\}$.
For instance, we may go back to our small example $\mathcal{M}'(a, b, c)$ with $a = b = c = d = 3$ and draw the lines on $\mathcal{M}'(3, 3, 3) = \mathcal{F}(\begin{pmatrix}3 & 3 & 3 & 3\end{pmatrix}^t)$ using the following set $\mathcal{B}$:

$$\mathcal{B} := \{\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}\}$$

In this case, the graph looks like this:

$$\begin{pmatrix}0\\3\\3\\0\end{pmatrix} \text{———} \begin{pmatrix}1\\2\\2\\1\end{pmatrix} \text{———} \begin{pmatrix}2\\1\\1\\2\end{pmatrix} \text{———} \begin{pmatrix}3\\0\\0\\3\end{pmatrix}$$

and it is therefore connected.
If the graph $\mathcal{F}(u)$ is connected, then the Markov chain performing a random walk on this graph is also connected. If we then apply the Metropolis-Hastings algorithm, we can turn this Markov chain into one whose stationary distribution is equal to π. This means that, if we let the random walk X_n run for long enough (i.e. we take n any big enough), then we have, regardless of the initial position X_0:

$$\mathbb{P}(X_n = v) \approx \pi(v).$$

However, it is not easy to make sure a priori that the set $\mathcal{B}$ connects a fibre. The idea by Diaconis and Sturmfels is, once we fix the matrix A, to try and connect all fibres at once through means of what we will call a Markov basis.

Definition 2. *Let $A \in \mathbb{Z}^{s \times r}$ be a matrix. A set $\mathcal{B} \subseteq \mathrm{Ker}_{\mathbb{Z}}(A)$ that connects $\mathcal{F}(u)$ for every $u \in \mathbb{N}^r$ is called a Markov basis.*

Before we go any further, it may be necessary to notice that, if we want to connect all fibres, we need $\mathcal{B}$ to generate $\mathrm{Ker}_{\mathbb{Z}}(A)$ as a group: assume all the fibres to be connected and consider $w \in \mathrm{Ker}_{\mathbb{Z}}(A)$. Take any $v \in \mathbb{N}^r$ big enough for $v + w$ to also be in $\mathbb{N}^r$. Then the fact that $v, v + w \in \mathcal{F}(v)$ are connected implies the existence of a chain $v_0 = v, v_1, \ldots, v_n = v + w$, with $v_i - v_{i-1} \in \pm\mathcal{B}$ for all i. Using a telescoping sum, we can conclude:

$$w = v + w - v = v_n - v_0 = \sum_{i=1}^{n} (v_i - v_{i-1}) \in \underbrace{\pm\mathcal{B} \pm \cdots \pm \mathcal{B}}_{n \text{ times}} \subseteq \langle \mathcal{B} \rangle.$$

However, the set $\mathcal{B}$ generating the integer null space $\mathrm{Ker}_{\mathbb{Z}}(A)$ is definitely not enough to connect every fibre, as one can see from the following example taken from [3].

Example 3. *Consider the setting:*

$$A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{pmatrix}, \qquad \mathcal{B} = \{b_1, b_2\} = \{ \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -2 \\ 1 \end{pmatrix} \}, \qquad u = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, v = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

While $\mathcal{B}$ generates $\mathrm{Ker}_{\mathbb{Z}}(A)$ as a group, not all fibres are connected: for instance, there is no way to connect u and v in the graph built on $\mathcal{F}(u) = \{u, v\}$.

This means that, in order to better understand when a subset $\mathcal{B} \subseteq \mathrm{Ker}_{\mathbb{Z}}(A)$ is a Markov basis, we need to go further in the characterization given by Diaconis and Sturmfels, which involves the following algebraic tool.

Definition 4. *Let R be a ring and consider $A \in \mathbb{Z}^{s \times r}$. Let $\mathcal{B} \subseteq \mathrm{Ker}_{\mathbb{Z}}(A)$ generate the integer null space as a group. Then we define the following ideal of $R[\underline{x}] := R[x_1, \ldots, x_r]$:*

$$\mathcal{I}_{\mathcal{B}}^{(R)} := \langle \{ \prod_{i=1}^{r} x_i^{\max(b_i, 0)} - \prod_{i=1}^{r} x_i^{\max(-b_i, 0)} \}_{b \in \mathcal{B}} \rangle_{R[x_1, \ldots, x_r]}.$$

When the choice of the ring does not matter or is implicit, we will denote the ideal simply by $\mathcal{I}_{\mathcal{B}}$.

To simplify the notation, we will often write the following.

Notation 5. *Given a vector $v \in \mathbb{R}^r$, we define the notation:*

$$v^+ := (\max(v_i, 0))_{i=1,\ldots,r} \qquad\qquad v^- := (\max(-v_i, 0))_{i=1,\ldots,r}.$$

In particular, the ideal $\mathcal{I}_{\mathcal{B}}^{(R)}$ can now be re-written simply as:

$$\mathcal{I}_{\mathcal{B}}^{(R)} := \langle \{ \underline{x}^{b^+} - \underline{x}^{b^-} \}_{b \in \mathcal{B}} \rangle_{R[\underline{x}]}.$$

Diaconis and Sturmfels then proceed to study the saturation of such ideal.

Definition 6. *Let S be a commutative ring and let J be an ideal of S. Given some elements $s_1, \ldots, s_k \in S$, we say that J is saturated in S with respect to $s_1, \ldots, s_k$ if, for any $r \in S$, the following equivalence of statements holds:*

$$r \in J \iff s_1 r \in J \iff \cdots \iff s_k r \in J.$$

In the whole thesis we will always exclusively discuss saturation of ideals in the polynomial ring $R[x_1, \dots, x_r]$ with respect to $x_1, \dots, x_r$, therefore we will often just talk about "saturated ideals of $R[x_1, \dots, x_r]$" and imply that the elements we are referring to are $x_1, \dots, x_r$. Following the notation on [4], we will denote the saturation of an ideal J with respect to such elements by $(J : \underline{x}^\infty)$.
Diaconis and Sturmfels then proceed to formulate the following statement.

Theorem 7. *(Diaconis-Sturmfels, [1, Theorem 3.1]) Let R be a field and consider $A \in \mathbb{Z}^{s \times r}$. Let $\mathcal{B} \subseteq \mathrm{Ker}_{\mathbb{Z}}(A)$ generate the integer null space of A as a group. Then the ideal $\mathcal{I}_{\mathcal{B}}$ is saturated with respect to $x_1, \dots, x_r$ if and only if every fibre $\mathcal{F}(u)$ is connected by $\mathcal{B}$.*

Then they show how, starting from a Gröbner basis of an ideal of the form $\mathcal{I}_{\mathcal{B}}$, one can easily obtain a set $\mathcal{C} \supseteq \mathcal{B}$ such that $\mathcal{I}_{\mathcal{C}}$ is the saturation of $\mathcal{I}_{\mathcal{B}}$. Using Theorem 7, they therefore deduce that the set $\mathcal{C}$ yielded by this algorithm is a Markov basis. For instance, let us see how this applies to the example we have seen above.

Example 8. *Consider the setting:*

$$A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{pmatrix}, \qquad \mathcal{B} := \{b_1, b_2\} := \{ \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -2 \\ 1 \end{pmatrix} \}, \qquad u = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, v = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

The ideal $\mathcal{I}_{\mathcal{B}} = \langle x_1x_3 - x_2^2, x_2x_4 - x_3^2 \rangle$ is not saturated, and its saturation is indeed equal to:

$$(\mathcal{I}_{\mathcal{B}} : \underline{x}^\infty) = \langle x_1x_3 - x_2^2, x_2x_4 - x_3^2, x_1x_4 - x_2x_3 \rangle =: \mathcal{I}_{\mathcal{C}}, \qquad \text{with } \mathcal{C} := \{b_1, b_2, \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} \}.$$

So $\mathcal{C}$ is a Markov basis now, which makes sense since u and v are now connected by the third vector of $\mathcal{C}$, in the graph built on $\mathcal{F}(u) = \{u, v\}$.

The only problem with this method is that computing a Gröbner basis can be computationally challenging for large numbers of generators. This is why Aoki, Hara and Takemura introduce a different algorithm in their paper [5]: instead of extending $\mathcal{B}$ to a Markov basis, they consider $\mathcal{F}(u)$ as a sub-graph of $u + \mathrm{Ker}_{\mathbb{Z}}(A)$, where the latter also has its edges drawn according to the rule defined in Line (1). Notice that $u + \mathrm{Ker}_{\mathbb{Z}}(A)$ is always a connected graph, since $\mathcal{B}$ generates $\mathrm{Ker}_{\mathbb{Z}}(A)$ as a group.
The algorithm defined in [5] then allows the Markov chain to do more than just jumping from a vector u to a vector $v \in \mathcal{F}(u)$ of the form $u \pm b$ for a suitable $b \in \mathcal{B}$; more precisely, given the position $X_t = u$, the position X_{t+1} of the Markov chain is defined by the following process, where $n := |\mathcal{B}|$:

1. Choose $\alpha_1, \dots, \alpha_n \in \mathbb{N}$ at random from a Poisson distribution of a certain parameter λ which we can decide arbitrarily;

2. Choose $\epsilon_1, \dots, \epsilon_n \in \{\pm 1\}$ uniformly at random;

3. Compute $v := u + \sum_{i=1}^{n} \epsilon_i \alpha_i b_i$. This vector is reachable from u in the graph $u + \mathrm{Ker}_{\mathbb{Z}}(A)$ by adding α_i times the vector $\epsilon_i b_i$, for every $i = 1, \ldots, n$. If $v \in \mathbb{N}^r$, i.e. if it is inside $\mathcal{F}(u)$, then define $X_{k+1} := v$, otherwise start over from the first step.

As we have already discussed, the graph $u + \mathrm{Ker}_{\mathbb{Z}}(A)$ is connected and therefore, since this algorithm allows to add any $\epsilon_i b_i$ for as many times as one wants, X_{t+1} can be any vector in the fibre. This means that this Markov chain on $\mathcal{F}(u)$ is always connected, even when the edges defined on Line (1) do not make $\mathcal{F}(u)$ a connected graph.
The authors also present a variation of this algorithm where the first step uses a combination of a geometric and a multinomial distribution to determine the coefficients $\alpha_1, \ldots, \alpha_n$, but the concept of the algorithm remains the same otherwise.
As Aoki, Hara and Takemura show in their paper [5], this algorithm can be very effective, for suitable parameters of the Poisson or the Geometric distribution. The choice, however, is not always easy: bigger parameters allow an easier connection of connected components that can be potentially very far apart. On the other hand, smaller parameters imply a smaller average value of $\sum_{i=1}^{n} \alpha_i$, which we can interpret as the number of average steps "potentially outside $\mathcal{F}(u)$" that connect X_t and X_{t+1}. This means that the Markov chain wastes less time "wandering" far away from the fibre and therefore we have a faster apparent convergence to the stationary distribution.
The latter argument is what motivates Holmes to tweak the algorithm above by adding the extra condition $\sum_{i=1}^{n} \alpha_n \leqslant N$ in step 1, for a big enough constant $N \in \mathbb{N}$. In other words, just like Aoki, Hara and Takemura do, Holmes allows the Markov chain to move from u to a vector $v \in \mathcal{F}(u)$ which is reachable in the bigger graph $u + \mathrm{Ker}_{\mathbb{Z}}(A)$, even if there is no direct edge connecting u and v and even if there is no path from u to v that stays entirely inside $\mathcal{F}(u)$. However, he allows such jump only as long as v is "within N steps outside of $\mathcal{F}(u)$ from u ".
It is therefore natural to try to find the minimum number of steps outside any fibre $\mathcal{F}(u)$ that the Markov chain needs to take in order to be connected. This number is introduced by Holmes in [3] as $||\mathcal{B}||$ and will be properly explored in the second chapter.

The main results

The most prominent result of this thesis is the following Theorem from Chapter 2.

Theorem. *(Corollary 53) Let $A \in \mathbb{Z}^{s \times r}$ be a matrix and take a set $\mathcal{B} = \{b_1, \ldots, b_n\}$ generating its integer null space as a group. Set $\beta := \max_{j=1,\ldots,n} ||b_j||_\infty$ and $k := \min(\mathrm{rk}(B), n-1)$. Denote by B the $r \times n$ matrix whose columns are the vectors of $\mathcal{B}$. Then:*

$$||\mathcal{B}|| \leqslant n \cdot \sqrt{k+1}^{k+1} \cdot \beta^k \leqslant n^{\frac{n}{2}+1} \beta^{n-1}.$$

This result presents an improvement [1] when compared to Theorem 1.3 in [3], which bounds the norm of $\mathcal{B}$ with $n^{n+1}\beta^{n-1}$.
The background to fully understand this result and its consequences is properly explained

[1] The quantity $n^2\sqrt{k+1}^{k+1}\beta^{k+1}$ is always at most $n^{\frac{n}{2}+1}\beta^n$, which is an improvement for any non-trivial case $n \geqslant 2$.

in Chapter 1. However, this is not the only purpose of the first chapter. More specifically, we are firstly going to explore the proof of Theorem 7 from [1] and [4] in more detail and generalize it to every ring. Secondly, we are going to focus on the connection of some classes of fibres, such as the "big fibres" and "small fibres" (sets of the form $\mathcal{F}(u)$ with u having respectively a big or a small p-norm). Lastly, we will discuss how the connectedness of some fibres implies the connectedness of others. An example of interesting result that comes from this kind of arguing will be the following.

Theorem (Theorem 40)**.** *Let $A \in \mathbb{Z}^{s\times r}$ be a matrix and take a set $\mathcal{B} = \{b_1, \dots, b_n\}$ generating its integer null space as a group. Set $\beta := \max_{j=1,\dots,n} ||b_j||_\infty$. Then the set:*

$$\mathcal{B}' := \{v \in \mathrm{Ker}_{\mathbb{Z}}(A) : |v_i| \leqslant ||\mathcal{B}|| \cdot \beta \ \forall i = 1, \dots, r\}$$

is a Markov basis.

Combining this theorem with the aforementioned Corollary 53 allows us to conclude that the set:

$$\mathcal{B}' := \{v \in \mathrm{Ker}_{\mathbb{Z}}(A) : |v_i| \leqslant n\sqrt{k+1}^{k+1}\beta^{k+1} \ \forall i = 1, \dots, r\}$$

is always a Markov Basis.
Unlike Chapter 1, the second chapter is much more oriented towards one specific goal, improving the bound of $||\mathcal{B}||$, which we will do by coming up with an alternative to Siegel's Lemma for solutions of integer systems of equations. As we use this alternative lemma, which only works for our specific case, we also want to compare it with the more general result found by Bombieri and Vaaler in [6].
The approach of the second chapter will also yield some computer algorithms which allow further improvement in bounding $||\mathcal{B}||$. These algorithms can not only be helpful in the general problem of maximizing a determinant, they will also allow a substantial improvement on the bound of $||\mathcal{B}||$ which, as described by Holmes in [3], might potentially make the convergence of the methods described above faster.

Chapter 1

Connectedness of some classes of fibres

1.1 Connecting all fibres

As we have seen in the previous section, we will very often use the following notation.

Notation 9. *Let $A \in \mathbb{Z}^{s\times r}$ be a matrix. With a slight abuse of notation, we will denote by $\mathrm{Ker}_{\mathbb{Z}}(A)$ the null space of A in $\mathbb{Z}^r$. Let $\mathcal{B} = \{b_1, \dots, b_n\} \subseteq \mathrm{Ker}_{\mathbb{Z}}(A)$ generate it as a group. Then we write B for the $r \times n$ matrix whose columns are the vectors $b_1, \dots, b_n$.*

We will omit the $\mathcal{B}$ from the notation where it is clear from context.
In this scenario, we are exploring the connectedness of undirected graphs of the form $(\mathcal{F}(u), E)$ where:

$$\mathcal{F}(u) = \{v \in \mathbb{N}^r : Av = Au\}, \qquad E = \{(u, v) \in \mathcal{F}(u)^2 : u - v \in \pm\mathcal{B}\}.$$

More specifically, we will study connectedness in the following sense.

Definition 10. *Let (V, E) be a graph. We will say two vertices $u, v \in V$ to be connected if $u = v$ or there exists a path $v_0 = u, v_1, \dots, v_k = v$ all inside V such that $(v_j, v_{j-1}) \in E$ for every $j = 1, \dots, k$.*
We will call the graph "connected" if any two vertices of V are connected in (V, E).

With an abuse of notation, we will often refer to $\mathcal{F}(u)$ as a graph by implying the structure we just defined, and we will be referring to the connectedness of two vectors u, v in the same fibre by implying that we are once again referring to the connectedness with respect to the graph $(\mathcal{F}(u), E)$ drawn above. To highlight the fact that the edges of the graph (and therefore its connectedness) depend on the set $\mathcal{B}$, we will often say vectors or fibres to be "connected by $\mathcal{B}$".
We have already seen the key result which is Theorem 7. The proofs given in [4] and in [1] are very similar, except for how the ideal $\mathcal{I}_A$ (see Definition 12) is introduced. While [4] deeply analyzes this ideal from a toric geometry perspective, the paper [1] uses it more as a tool. However, in both cases, $\mathcal{I}_A$ is introduced before the theorem without further

explanation, which may make it a bit unclear as to why such ideal is involved in the property of connectedness of graphs.
We will therefore rely on the same overall argument as [1] and [4], but with some changes to the proof of some implications and by focusing more on the connectedness of pairs of vectors, rather than the whole graph. This will hopefully make the definition of $\mathcal{I}_A$ come naturally and will also lead to more general results, starting with the following lemma.

Lemma 11. *Given a ring R and a matrix $A \in \mathbb{Z}^{s\times r}$, let $\mathcal{B} \subseteq \mathrm{Ker}_{\mathbb{Z}}(A)$ generate the integer null space as a group. Consider two vectors $u, v \in \mathcal{F}(u)$. Then they are connected by $\mathcal{B}$ if and only if $\underline{x}^u - \underline{x}^v \in \mathcal{I}_{\mathcal{B}}$.*

Proof. Assume $\underline{x}^u - \underline{x}^v \in \mathcal{I}_{\mathcal{B}}$. Since $\mathcal{I}_{\mathcal{B}}$ is generated by the elements $\underline{x}^{b^+} - \underline{x}^{b^-}$ this means that, for suitable $M \in \mathbb{N}, \alpha_i \in \mathbb{N}^r, \rho_i \in R, b_i \in \mathcal{B}$, we can write:

$$\underline{x}^u - \underline{x}^v = \sum_{i=1}^{M} \rho_i \underline{x}^{\alpha_i}(\underline{x}^{b_i^+} - \underline{x}^{b_i^-}) = \sum_{i=1}^{M} \rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-}). \tag{1.1}$$

Consider the set: [1]

$$J := \{i \in \{1,\dots,M\} : \alpha_i + b_i^+ \in \mathcal{F}(u),\ \alpha_i + b_i^+ \text{ or } \alpha_i + b_i^- \text{ is connected to } u\}.$$

Notice that the vectors $\alpha_i + b_i^+$ and $\alpha_i + b_i^-$ are connected by b_i, so one is connected to u if and only if the other is. This means that we have the equality:

$$J = \{i \in \{1,\dots,M\} : \alpha_i + b_i^+ \in \mathcal{F}(u),\ \alpha_i + b_i^+ \text{ and } \alpha_i + b_i^- \text{ are connected to } u\}. \tag{1.2}$$

Now write $\underline{x}^u - \underline{x}^v$ as:

$$\underline{x}^u - \underline{x}^v = \sum_{i=1}^{M} \rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-}) = \sum_{i\in J} \rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-}) + \sum_{i\notin J} \rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-}).$$

This means that $\underline{x}^u - \underline{x}^v$ can be expressed as the sum of two polynomials $g(\underline{x})$ and $h(\underline{x})$, where:

$$g(\underline{x}) := \sum_{i\in J} \rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-}), \qquad h(\underline{x}) := \sum_{i\notin J} \rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-}).$$

Notice that all the binomials $\rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-})$ in which $\underline{x}^u$ appears with a non-zero coefficient have $i \in J$ by definition of J (and because of course u is connected to itself). This means that all the summands in Line (1.1) which contribute to $\underline{x}^u$ in $\underline{x}^u - \underline{x}^v$ are also summands of the sum that defines $g(\underline{x})$. Since $\underline{x}^u$ appears with coefficient 1 in $\underline{x}^u - \underline{x}^v$, it follows that 1 is also the coefficient of $\underline{x}^u$ in $g(\underline{x})$.
This argument actually holds for any monomial in $g(\underline{x})$: if $\underline{x}^\beta$ appears in $g(\underline{x})$ with a non-zero coefficient, then it means that $\underline{x}^\beta$ must come from at least one binomial $\rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-})$ with $i \in J$, therefore $\beta = \alpha_i + b_i^{\pm}$ must be connected to u, by the equality on

[1] Since $\alpha_i + b_i^+$ and $\alpha_i + b_i^-$ are in the same fibre, the condition $\alpha_i + b_i^+ \in \mathcal{F}(u)$ in the definition of J is equivalent to $\alpha_i + b_i^+, \alpha_i + b_i^- \in \mathcal{F}(u)$. Also, if the expression on Line (1.1) is minimal, such condition is always satisfied and can therefore be dropped in the definition of J

Line (1.2). The definition of J implies that all the binomials $\rho_j(\underline{x}^{\alpha_j+b_j^+} - \underline{x}^{\alpha_j+b_j^-})$ which contribute to $\underline{x}^\beta$ in Line (1.1) have $j \in J$, therefore they all also appear in the sum that defines $g(\underline{x})$. This means that the coefficient of $\underline{x}^\beta$ in $g(\underline{x})$ must be the same as in $\underline{x}^u - \underline{x}^v$. Since $\underline{x}^u - \underline{x}^v$ only has two monomials and we have already proved $\underline{x}^u$ to have coefficient 1 in $g(\underline{x})$, this means that either $g(\underline{x}) = \underline{x}^u$ or $g(\underline{x}) = \underline{x}^u - \underline{x}^v$. However, pure difference ideals can never contain monomials. An easy way to see this is to notice that all the pure difference binomials $\underline{x}^{b^+} - \underline{x}^{b^-}$ are contained in the kernel of the evaluation morphism $\psi_{\underline{1}} : R[\underline{x}] \to R$ that sends every x_i to 1, therefore $\mathcal{I}_\mathcal{B} \subseteq \mathrm{Ker}(\psi_{\underline{1}})$. However, no monomial can be in the kernel of this map.
This means that $g(\underline{x}) \in \mathcal{I}_\mathcal{B}$ cannot be just $\underline{x}^u$ and it is therefore equal to $\underline{x}^u - \underline{x}^v$, so $\underline{x}^v$ also appears in $g(\underline{x})$ with a non-zero coefficient. Specifically, it must come from at least one term $\rho_t \underline{x}^{\alpha_t}(\underline{x}^{b_t^+} - \underline{x}^{b_t^-})$ with $t \in J$. Using Line (1.2) once more, the fact that $t \in J$ implies that v is connected to u, which concludes the proof of the first implication.
For the converse, assume $u, v \in \mathcal{F}(u) = \mathcal{F}(v)$ to be connected through a chain of vectors $v_0 = v, \dots, v_N = u$. This means that, up to a sign, $v_i - v_{i-1} \in \mathcal{B}$. To simplify the notation, assume such a sign to be a + and set $b_i := v_i - v_{i-1}$ for $i = 1, \dots, N$. Then this means that we can conclude the argument by using a telescoping sum:

$$\underline{x}^u - \underline{x}^v = \sum_{i=1}^N (\underline{x}^{v_i} - \underline{x}^{v_{i-1}}) = \sum_{i=1}^N (\underline{x}^{v_{i-1}+b_i^+-b_i^-} - \underline{x}^{v_{i-1}}) = \sum_{i=1}^N \underline{x}^{v_{i-1}-b_i^-}(\underline{x}^{b_i^+} - \underline{x}^{b_i^-}) \in \mathcal{I}_\mathcal{B}.$$

Notice that the fact that $\underline{x}^{v_{i-1}-b_i^-} \in \mathbb{Z}[\underline{x}]$ is because b_i^- is exactly the quantity that is subtracted when moving from v_{i-1} to v_i. More formally, we have:

$$v_i = v_{i-1} + b_i = v_{i-1} + b_i^+ - b_i^-$$

and therefore $v_{i-1} - b_i^- = v_i - b_i^+$. This means that, for every index $j \in \{1, \dots, r\}$, we have:

$$(v_{i-1} - b_i^-)_j = v_{i-1,j} - (b_i^-)_j = \begin{cases} v_{i-1,j} - 0 = v_{i-1,j} \geqslant 0 & \text{if } (b_i)_j \geqslant 0 \\ v_{i,j} - (b_i^+)_j = v_{i,j} + 0 = v_{i,j} \geqslant 0 & \text{if } (b_i)_j \leqslant 0 \end{cases}.$$

Notice that by the notation $(b_i)_j$ we mean the j-th entry of the vector b_i. If we remember Notation 9, then b_i is the i-th column of the matrix B, so essentially we are using the quantity $(b_i)_j = b_{j,i}$.
So for every i we have $\underline{x}^{v_{i-1}-b_i^-} \in \mathbb{Z}[\underline{x}]$ and therefore $\underline{x}^u - \underline{x}^v \in \mathcal{I}_\mathcal{B}$, which is what we wanted to prove. □

Using this lemma, we now know that asking for all the vectors u, v in the same fibre to be connected means asking for all the binomials $\underline{x}^u - \underline{x}^v$ to be inside the ideal $\mathcal{I}_\mathcal{B}$. It is therefore natural to introduce the following definition.

Definition 12. *Given a ring R and a matrix $A \in \mathbb{Z}^{s\times r}$, define the ideal $\mathcal{I}_A^{(R)}$ of $R[\underline{x}]$ as:*

$$\mathcal{I}_A^{(R)} := \langle \{\underline{x}^u - \underline{x}^v : u, v \in \mathbb{N}^r,\ Au = Av\} \rangle_{R[\underline{x}]}.$$

When the choice of the ring does not matter or is implicit, we will denote the ideal simply by $\mathcal{I}_A$.

Notice that $\mathcal{I}_{\mathcal{B}} \subseteq \mathcal{I}_A$ for any $\mathcal{B} \subseteq \mathrm{Ker}(A)$, since we have the implications:

$$b \in \mathrm{Ker}(A) \Longrightarrow 0 = Ab = A(b^+ - b^-) \Longrightarrow Ab^+ = Ab^- \Longrightarrow \underline{x}^{b^+} - \underline{x}^{b^-} \in \mathcal{I}_A.$$

So all the fibres being connected (i.e. any two vectors in the same fibre being connected) is equivalent to the other inclusion holding as well.
So in particular we have now just proved the following theorem by Sturmfels and Diaconis, which they had previously stated only for fields.

Theorem 13 (Diaconis-Sturmfels)**.** *Let R be a ring and consider $A \in \mathbb{Z}^{s\times r}$. Let $\mathcal{B} \subseteq \mathrm{Ker}_{\mathbb{Z}}(A)$ generate the integer null space as a group. Then the equality $\mathcal{I}_{\mathcal{B}} = \mathcal{I}_A$ holds if and only if every fibre $\mathcal{F}(u)$ is connected.*

Apart from the fact that this version can be used for any ring, the only difference with Theorem 7 is that the latter was asking for $\mathcal{I}_{\mathcal{B}}$ to be saturated. Since Lemma 12.2 from [4] proves that $\mathcal{I}_A$ is the saturation of $\mathcal{I}_{\mathcal{B}}$ (and the argument works for any ring R), we notice that the two formulations actually coincide.
To give a more intuitive perspective on why the saturation is involved, let us see the following lemma.

Lemma 14. *Let R be any ring. An ideal $I \trianglelefteq R[\underline{x}]$ is saturated with respect to $x_1, \dots, x_r$ if and only if it is of the form $I = J \cap R[\underline{x}]$, with $J \trianglelefteq R[\underline{x}^{\pm 1}] := R[x_1, \dots, x_r, x_1^{-1}, \dots, x_r^{-1}]$. Moreover, if $I = \langle \mathcal{G} \rangle_{R[\underline{x}]}$ with $\mathcal{G} \subseteq R[\underline{x}]$, then we can take J as $J = \langle \mathcal{G} \rangle_{R[\underline{x}^{\pm 1}]}$.*

Proof. If $I = J \cap R[\underline{x}]$ with $J \trianglelefteq R[\underline{x}^{\pm 1}]$, then I is saturated: assume $f(\underline{x}) \in R[\underline{x}]$ to be such that $\underline{x}^{\alpha} f(\underline{x})$ lies in I. Since $\underline{x}^{\alpha}$ is invertible in $R[\underline{x}^{\pm 1}]$ and multiplying by an invertible element does not alter the presence in an ideal, the fact that $\underline{x}^{\alpha} f(\underline{x}) \in I \subseteq J$ implies that $f(\underline{x})$ is still in J. Also, we were assuming $f(\underline{x}) \in R[\underline{x}]$, so $f(\underline{x}) \in R[\underline{x}] \cap J = I$. Conversely, assume $I = \langle \mathcal{G} \rangle_{R[\underline{x}]}$ to be saturated and take $J := \langle \mathcal{G} \rangle_{R[\underline{x}^{\pm 1}]}$. Then of course $I \subseteq J$ and therefore $I \subseteq J \cap R[\underline{x}]$. For the other inclusion, take an element $f(\underline{x}) = \sum_{i=1}^{N} \rho_i \underline{x}^{\alpha_i} g_i(\underline{x}) \in R[\underline{x}] \cap J$, with $g_i(\underline{x}) \in \mathcal{G}, \rho_i \in R$ for all $i = 1, \dots, N$. Take a vector $\beta \in \mathbb{N}^r$ such that $\beta + \alpha_i \in \mathbb{N}^r$ for every i (namely, take β with $\beta_j \geqslant \max_{i=1}^{N}(-\alpha_{i,j})$ for all j). Then, for such β, we have that $\underline{x}^{\beta+\alpha_i} \in R[\underline{x}]$ for all i and therefore:

$$\underline{x}^{\beta} f(\underline{x}) = \sum_{i=1}^{N} \rho_i \underline{x}^{\beta+\alpha_i} g_i(\underline{x}) \in \langle \mathcal{G} \rangle_{R[\underline{x}]} = I.$$

So we proved that $\underline{x}^{\beta} f(\underline{x}) \in I$ with $f(\underline{x}) \in R[\underline{x}]$. Since I is saturated, this implies $f(\underline{x}) \in I$, which concludes the proof. □

So what we just proved is that an ideal $I \trianglelefteq R[\underline{x}]$ being saturated means that we can "play with its generators" as much as we want, even multiply by Laurent polynomials: as long as the final result is back inside $R[\underline{x}]$, this element is still going to be in I.
Applying Lemma 14 to the ideal $\mathcal{I}_{\mathcal{B}} = \langle \{\underline{x}^{b^+} - \underline{x}^{b^-}\}_{b \in \mathcal{B}} \rangle$ can therefore help us understand the idea behind the property of saturation. Essentially, the point is that $\mathcal{B}$ already generates the integer null space as a group, so every distance $u - v$ is reachable in a suitable number of jumps (i.e. by adding or subtracting a certain number of vectors in $\mathcal{B}$). While it is not necessarily true that throughout the whole path of jumps we will stay

inside $\mathbb{N}^r$, we know that the starting and ending point of the path are indeed natural vectors. Using the same idea as Lemma 11, this translates into $\underline{x}^u - \underline{x}^v$ being generated by $\{\underline{x}^{b^+} - \underline{x}^{b^-}\}_{b\in\mathcal{B}}$ in $R[\underline{x}^{\pm 1}]$ while also being in $R[\underline{x}]$. So for u and v to be connected, again by Lemma 11, we need this to imply that $\underline{x}^u - \underline{x}^v \in \mathcal{I}_{\mathcal{B}}$, which is exactly what saturation is about.
Before we move on to the next section, let us consider the following lemma which was also stated in the paper [1] for the case $\mathcal{I}_{\mathcal{B}} = \mathcal{I}_A$ and which, thanks to the new approach coming from the first part of Lemma 11, can now be generalized as follows.

Lemma 15. *Let S be a subring of another ring R, and take $A, \mathcal{B}$ as in Notation 9. Then:*

$$\mathcal{I}_{\mathcal{B}}^{(R)} \cap S[\underline{x}] = \mathcal{I}_{\mathcal{B}}^{(S)}.$$

Proof. The inclusion $\mathcal{I}_{\mathcal{B}}^{(R)} \cap S[\underline{x}] \supseteq \mathcal{I}_{\mathcal{B}}^{(S)}$ is trivial.
For the other inclusion, assume by contradiction that there exists a polynomial $f(\underline{x}) \in (\mathcal{I}_{\mathcal{B}}^{(R)} \cap S[\underline{x}])\backslash\mathcal{I}_{\mathcal{B}}^{(S)}$. Take a total order $\prec$ on the set $\mathbb{N}^r$ of all the exponents and take $f(\underline{x})$ with a minimal leading monomial with respect to $\prec$. Since $f(\underline{x})$ is generated by $\{\underline{x}^{b^+} - \underline{x}^{b^-}\}_{b\in\mathcal{B}}$ in $R[\underline{x}]$, it can be written as:

$$f(\underline{x}) = \sum_{i=1}^{M} \rho_i \underline{x}^{\alpha_i}(\underline{x}^{b_i^+} - \underline{x}^{b_i^-}) = \sum_{i=1}^{M} \rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-}), \tag{1.3}$$

for suitable $M \in \mathbb{N}, \alpha_i \in \mathbb{N}^r, \rho_i \in R, b_i \in \mathcal{B}$.
Let $c_u\underline{x}^u$ be the leading monomial of $f(\underline{x})$. As we have already seen in the proof of Lemma 11, we can define:

$$J = \{i \in \{1,\dots,M\} : \alpha_i + b_i^+ \in \mathcal{F}(u),\ \alpha_i + b_i^+ \text{ or } \alpha_i + b_i^- \text{ are connected to } u\},$$

which is actually the same as:

$$J = \{i \in \{1,\dots,M\} : \alpha_i + b_i^+ \in \mathcal{F}(u),\ \alpha_i + b_i^+ \text{ and } \alpha_i + b_i^- \text{ are connected to } u\},$$

and we can split $f(\underline{x})$ as $f(\underline{x}) = g(\underline{x}) + h(\underline{x})$ with:

$$g(\underline{x}) = \sum_{i\in J} \rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-}), \qquad h(\underline{x}) = \sum_{i\notin J} \rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-}).$$

As already discussed in Lemma 11, every monomial of $g(\underline{x})$ with a non-zero coefficient also appears in $f(\underline{x})$ with that same coefficient. Indeed, if $\underline{x}^\beta$ appears in $g(\underline{x})$ with a non-zero coefficient, then it means that $\underline{x}^\beta$ must come from at least one binomial $\rho_i(\underline{x}^{\alpha_i+b_i^+} - \underline{x}^{\alpha_i+b_i^-})$ with $i \in J$, therefore $\beta = \alpha_i + b_i^\pm$ lies in $\mathcal{F}(u)$ and it is connected to u. This implies that all the binomials $\rho_j(\underline{x}^{\alpha_j+b_j^+} - \underline{x}^{\alpha_j+b_j^-})$ which contribute to $\underline{x}^\beta$ in Line (1.3) have $j \in J$, therefore they all also appear in the sum that defines $g(\underline{x})$.
This argument allows us to conclude the following:

- Thanks to the construction of $g(\underline{x})$ and by what we just said, $\underline{x}^u$ appears in $g(\underline{x})$ with its coefficient being precisely c_u;

- As discussed in the proof of Lemma 11, $g(\underline{x})$ is not a monomial;
- Since $c_u\underline{x}^u$ is the leading monomial of $f(\underline{x})$, we have that $\underline{x}^u$ is the monomial of $f(\underline{x})$ with the greatest exponent, with respect to $\prec$. Since all the monomials of $g(\underline{x})$ are taken from $f(\underline{x})$, this means that they all have a smaller exponent than u. Combining this with the previous point, we deduce that $c_u\underline{x}^u$ is also the leading monomial of $g(\underline{x})$.

Putting all these considerations together, we conclude that if we take any monomial $c_v\underline{x}^v$ from $g(\underline{x})$ such that $c_v \neq 0$ and $v \neq u$, then:

- The vector v is strictly smaller than u, with respect to $\prec$;
- The vectors u, v are in the same fibre and they are connected by $\mathcal{B}$.

By Lemma 11, this means that $\underline{x}^u - \underline{x}^v$ lies in $\mathcal{I}_{\mathcal{B}}^{(S)}$, and therefore so does $c_u(\underline{x}^u - \underline{x}^v)$. But now we can take $f(\underline{x}) - c_u(\underline{x}^u - \underline{x}^v)$, which is the sum of a polynomial in $(\mathcal{I}_{\mathcal{B}}^{(R)} \cap S[\underline{x}])\backslash\mathcal{I}_{\mathcal{B}}^{(S)}$ and one in $\mathcal{I}_{\mathcal{B}}^{(S)}$. This means that $f(\underline{x}) - c_u(\underline{x}^u - \underline{x}^v) \in (\mathcal{I}_{\mathcal{B}}^{(R)} \cap S[\underline{x}])\backslash\mathcal{I}_{\mathcal{B}}^{(S)}$. However, the leading monomials in $f(\underline{x})$ and $c_u(\underline{x}^u - \underline{x}^v)$ cancel each other out, hence this new polynomial has a smaller leading monomial, which contradicts the choice of $f(\underline{x})$. □

This lemma will be useful for our next section, as it allows us to assume the pure difference binomials $\underline{x}^u - \underline{x}^v$ to be always generated $\mathbb{Z}[\underline{x}]$-linearly in any ideal $\mathcal{I}_{\mathcal{B}}^{(R)}$.

1.2 Connecting big fibres

The first section studied what happens when we try to connect all the fibres at once. For the rest of the chapter, however, we will focus on the connectedness of some smaller classes of fibres. In this section, for instance, we will be interested in only connecting all the "big" fibres, i.e. fibres of the form $\mathcal{F}(u)$ for a vector u with a high p-norm. A small counterexample can show us right away that just big fibres being connected is not enough to guarantee the connectedness of every fibre (which we know to be equivalent to $\mathcal{I}_{\mathcal{B}}$ being saturated).

Example 16. *Consider the following setting:*

$$A = \begin{pmatrix} 1 & 1 \end{pmatrix}, \qquad \mathcal{B} = \{b_1, b_2\} := \{\begin{pmatrix} 4 \\ -4 \end{pmatrix}, \begin{pmatrix} 5 \\ -5 \end{pmatrix}\}.$$

The set $\mathcal{B}$ generates $\mathrm{Ker}_{\mathbb{Z}}(A)$ as a group: thanks to Bézout's identity, any vector $m\cdot\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ with $m \in \mathbb{Z}$ can be expressed as a $\mathbb{Z}$-linear combination of $b_1 = 4\cdot\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $b_2 = 5\cdot\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Also, in $\mathbb{N}^2$, $Au = u_1 + u_2 = ||u||_1$, so the fibres simply consist of the natural vectors with the same L_1-norm. It is obvious that not all fibres are connected: for instance, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

cannot be connected to $\begin{pmatrix}0\\1\end{pmatrix}$*. However, let us show that* $\mathcal{F}(u)$ *is connected for any* u *such that* $||u||_1 \geqslant 10$*: take two vectors* $u = \begin{pmatrix}u_1\\u_2\end{pmatrix}, v = \begin{pmatrix}v_1\\v_2\end{pmatrix}$ *with the same* L_1*-norm, and therefore in the same fibre. We want to prove their connectedness by induction on* $||u-v||_\infty$*. If* $||u-v||_\infty = 0$*, then* $u = v$ *and therefore the two vectors are connected. Otherwise, for* $||u-v||_\infty \geqslant 1$*, we have* $u \neq v$*. Since* $u_1 + u_2 = v_1 + v_2$ *we can assume, without loss of generality, that* $u_1 > v_1$ *and* $u_2 < v_2$*. Also, since* $u_1 + u_2 = ||u||_1 \geqslant 10$*, we have* $u_i \geqslant 5$ *for at least one index* $i \in \{1, 2\}$*.*
If $u_1 \geqslant 5$*, then we can define* $u'' := u - b_2 + b_1 = \begin{pmatrix}u_1 - 1\\u_2 + 1\end{pmatrix} \in \mathbb{N}^2$*, which is connected to* u *through* $u' := u - b_2 \in \mathbb{N}^2$*. By induction, since* $||u'' - v||_\infty = ||u - v||_\infty - 1$*, we know that* u'' *is connected to* v*. By transitivity, so is* u*.*
Analogously, if $u_2 \geqslant 5$*, then we can define* $u'' := u + b_2 - b_1 = \begin{pmatrix}u_1 + 1\\u_2 - 1\end{pmatrix} \in \mathbb{N}^2$*, which is connected to* u *through* $u' := u + b_2 \in \mathbb{N}^2$*. Once again, we can deduce by induction that* u'' *(and therefore* u*) is connected to* v*.*
So we just proved that all fibres $\mathcal{F}(u)$ *with* $||u||_1 \geqslant 10$ *are connected, even though not all fibres are.*

However we will show that, whenever the ring R is reduced, the connectedness of big fibres makes $\sqrt{\mathcal{I}_\mathcal{B}}$ saturated, rather than $\mathcal{I}_\mathcal{B}$. Before we formally state this, let us make a quick remark.

Remark 17. *Given a reduced ring* R *and* A,$\mathcal{B}$ *as in Notation 9, we have:*

$$(\sqrt{\mathcal{I}_\mathcal{B}} : \underline{x}^\infty) = \mathcal{I}_A.$$

In particular, we have that $\mathcal{I}_A \supseteq \sqrt{\mathcal{I}_\mathcal{B}}$ *and the equality holds if and only if* $\sqrt{\mathcal{I}_\mathcal{B}}$ *is saturated.*

Proof. Firstly notice that the computations on pages 31 and 32 of [4] show that $R[\underline{x}]/\mathcal{I}_A$ is isomorphic to a subring of a polynomial ring $R[t_1, \ldots, t_s]$, which is reduced, hence $\mathcal{I}_A$ is radical. Since it also contains $\mathcal{I}_\mathcal{B}$, this means that $\sqrt{\mathcal{I}_\mathcal{B}} \subseteq \mathcal{I}_A$.
Lastly, to prove the equality in the theorem, notice that, using Lemma 12.2 from [4] and considering the inclusions $\mathcal{I}_\mathcal{B} \subseteq \sqrt{\mathcal{I}_\mathcal{B}} \subseteq \mathcal{I}_A$, we can conclude:

$$\mathcal{I}_A = (\mathcal{I}_\mathcal{B} : \underline{x}^\infty) \subseteq (\sqrt{\mathcal{I}_\mathcal{B}} : \underline{x}^\infty) \subseteq (\mathcal{I}_A : \underline{x}^\infty) = \mathcal{I}_A,$$

which implies the equality we wanted. □

So now let us proceed in the same way we did for the first section, but this time with $\sqrt{\mathcal{I}_\mathcal{B}}$.

Lemma 18. *Let* R *be any ring. Given two vectors* u, v *in the same fibre, suppose that there exists* N *such that* $n \cdot u$ *and* $n \cdot v$ *are connected by* $\mathcal{B}$ *for every* $n \geqslant N$*.*
Then $(\underline{x}^u - \underline{x}^v)^{2N+1} \in \mathcal{I}_\mathcal{B}$*. In particular,* $\underline{x}^u - \underline{x}^v \in \sqrt{\mathcal{I}_\mathcal{B}}$*.*

Proof. By hypothesis we know that Nu is connected to Nv and $(N+1)u$ is connected to $(N+1)v$. By Lemma 11, this can be expressed as:

$$\begin{cases} \underline{x}^{Nu} - \underline{x}^{Nv} \in \mathcal{I}_{\mathcal{B}} \\ \underline{x}^{(N+1)u} - \underline{x}^{(N+1)v} \in \mathcal{I}_{\mathcal{B}} \end{cases} = \begin{cases} (\underline{x}^u)^N \equiv_{\mathcal{I}_{\mathcal{B}}} (\underline{x}^v)^N \\ (\underline{x}^u)^{N+1} \equiv_{\mathcal{I}_{\mathcal{B}}} (\underline{x}^v)^{N+1} \end{cases} .$$

If we substitute $(\underline{x}^u)^N$ from the first equality in $(\underline{x}^u)^{N+1}$ from the second (and analogously with $\underline{x}^v$), we get:

$$\begin{cases} (\underline{x}^v)^{N+1} \equiv_{\mathcal{I}_{\mathcal{B}}} (\underline{x}^u)^{N+1} \equiv_{\mathcal{I}_{\mathcal{B}}} \underline{x}^u(\underline{x}^v)^N \\ (\underline{x}^u)^{N+1} \equiv_{\mathcal{I}_{\mathcal{B}}} (\underline{x}^v)^{N+1} \equiv_{\mathcal{I}_{\mathcal{B}}} \underline{x}^v(\underline{x}^u)^N \end{cases} = \begin{cases} (\underline{x}^v)^N(\underline{x}^u - \underline{x}^v) \equiv_{\mathcal{I}_{\mathcal{B}}} 0 \\ (\underline{x}^u)^N(\underline{x}^u - \underline{x}^v) \equiv_{\mathcal{I}_{\mathcal{B}}} 0 \end{cases} .$$

This allows us to conclude that:

$$(\underline{x}^u - \underline{x}^v)^{2N+1} = (\underline{x}^u - \underline{x}^v) \sum_{j=0}^{2N} (-1)^{2N-j} \binom{2N}{j} (\underline{x}^u)^j (\underline{x}^v)^{2N-j} \in \mathcal{I}_{\mathcal{B}}.$$

Notice that this is true thanks to the fact that, in the summation above, at least one between j and $2N-j$ is always at least N, hence all the summands in the summation above are in $\mathcal{I}_{\mathcal{B}}$. □

One may wonder whether it was necessary to actually ask for nu and nv to be connected for all the integers $n \geqslant N$, instead of just for two consecutive integers $n, n+1$. The point is that it actually makes no difference, as one can see from the following lemma.

Lemma 19. *Given two vectors $u, v \in \mathbb{N}^r$ lying in the same fibre $\mathcal{F}(u)$, the following are equivalent:*

1. *There exists $N \in \mathbb{N}$ such that $n \cdot u$ is connected to $n \cdot v$ for all $n \geqslant N$;*
2. *There exists $t \in \mathbb{N}$ such that $t \cdot u$ is connected to $t \cdot v$ and $(t+1) \cdot u$ is connected to $(t+1) \cdot v$;*
3. *There exist two coprime numbers $t, h \in \mathbb{N}$ such that $t \cdot u$ is connected to $t \cdot v$ and $h \cdot u$ is connected to $h \cdot v$.*

Before we proceed with the proof, we need a technical lemma.

Lemma 20. *Consider two coprime natural numbers h, t. Then every $n \geqslant ht$ can be written as $n = \alpha h + \beta t$ with both α, β non-negative integers.*

Proof. Given h and t, take $N := ht$ and $n \geqslant N$.
Since $G.C.D.(h,t) = 1$, we know that there is an integer solution $(\overline{\alpha}, \overline{\beta})$ to the equation above. Since $(\overline{\alpha} + kt, \overline{\beta} - kh)$ is still a solution for every integer k, up to taking $k > 0$ big enough, we can assume $\overline{\alpha} \geqslant 0$. If we also have $\overline{\beta} \geqslant 0$, then we are done, as we found our solution. Otherwise, consider the following set:

$$\mathcal{A} := \{(\alpha, \beta) \in \mathbb{N} \times \mathbb{Z}_{<0} | n = \alpha h + \beta t\} \ni (\overline{\alpha}, \overline{\beta}),$$

which is therefore non-empty.
Take an element $(a, b) \in \mathcal{A}$ such that b is maximal (i.e. its absolute value is minimal). The fact that this element is in $\mathcal{A}$ implies that $ah + bt = n \geqslant N$ and therefore:

$$a \geqslant \frac{N - bt}{h} \geqslant \frac{N}{h} = t,$$

which means that $a - t \in \mathbb{N}$. Now notice that the element $(a - t, b + h)$ is still a solution to $n = \alpha h + \beta t$. If $b + h$ were still negative, then $(a - t, b + h)$ would be in $\mathcal{A}$ and $b + h$ would have a smaller absolute value than b, so it would contradict the minimality of (a, b). Therefore we conclude that $b + h \geqslant 0$ and hence we can exhibit the natural solution:

$$n = \alpha h + \beta t \qquad \text{with } \alpha := a - t \in \mathbb{N}, \;\; \beta := b + h \in \mathbb{N},$$

which concludes our proof. □

Now we proceed with the proof of Lemma 19.

Proof(Lemma 19). It is clear that (1) implies (2) and, since two consecutive integers are always coprime, that (2) implies (3). So we only need to prove that (3) implies (1).
Assume hu to be connected to hv (so $\underline{x}^{hu} - \underline{x}^{hv} \in \mathcal{I}_{\mathcal{B}}$ thanks to Lemma 11) and tu to be connected to tv (so $\underline{x}^{tu} - \underline{x}^{tv} \in \mathcal{I}_{\mathcal{B}}$).
Again by Lemma 11, we only need to prove that $\underline{x}^{nu} - \underline{x}^{nv} \in \mathcal{I}_{\mathcal{B}}$ for any $n \geqslant ht$. For any such n, which by the previous lemma can be expressed as $n = \alpha h + \beta t$ with $\alpha, \beta \geqslant 0$, we can write:

$$\underline{x}^{nu} - \underline{x}^{nv} = \underline{x}^{nu} - \underline{x}^{\alpha hu + \beta tv} + \underline{x}^{\alpha hu + \beta tv} - \underline{x}^{nv} = \underline{x}^{\alpha hu}(\underline{x}^{\beta tu} - \underline{x}^{\beta tv}) + \underline{x}^{\beta tv}(\underline{x}^{\alpha hu} - \underline{x}^{\alpha hv}) =$$

$$= \underline{x}^{\alpha hu}(\underline{x}^{tu} - \underline{x}^{tv})p(\underline{x}) + \underline{x}^{\beta tv}(\underline{x}^{hu} - \underline{x}^{hv})q(\underline{x}) \in \mathcal{I}_{\mathcal{B}}$$

for some suitable polynomials $p(x)$ and $q(x)$, where the last step uses the fact that $f(\underline{x}) - g(\underline{x})$ always divides $f^k(\underline{x}) - g^k(\underline{x})$ for any $k, f(\underline{x}), g(\underline{x})$. □

So now that we have seen that Lemma 18 is stated in the strongest possible way that could come from the proof we gave, it is reasonable to ask ourselves whether the converse also holds. If we consider the case $\mathrm{char}(R) = 0$, then the answer is yes.

Lemma 21. *Consider any ring R with $\mathrm{char}(R) = 0$ and take $u, v \in \mathcal{F}(u)$. If $\underline{x}^u - \underline{x}^v \in \sqrt{\mathcal{I}_{\mathcal{B}}}$, then there exists N such that nu is connected to nv for every $n \geqslant N$.*

Proof. Since by hypothesis we have $\underline{x}^u - \underline{x}^v \in \sqrt{\mathcal{I}_{\mathcal{B}}}$, we know that $(\underline{x}^u - \underline{x}^v)^n \in \mathcal{I}_{\mathcal{B}}$ for any integer n bigger than a suitable $N \in \mathbb{N}$. Since $(\underline{x}^u - \underline{x}^v)^n \in \mathbb{Z}[\underline{x}]$ then, by Lemma 15, this means that all these polynomials are generated $\mathbb{Z}[\underline{x}]$-linearly by $\{\underline{x}^{b^+} - \underline{x}^{b^-}\}_{b \in \mathcal{B}}$. Take any prime number $p \geqslant N$. Then this means that:

$$(\underline{x}^u - \underline{x}^v)^p = \sum_{i=1}^{T} \epsilon_i \underline{x}^{\alpha_i}(\underline{x}^{b_i^+} - \underline{x}^{b_i^-}) \qquad \text{with } \epsilon_i \in \{\pm 1\}, b_i \in \mathcal{B}, \alpha_i \in \mathbb{N}^r, T \in \mathbb{N}.$$

Take this equality modulo p (where the p-th power becomes a ring endomorphism):

$$\underline{x}^{pu} - \underline{x}^{pv} = \sum_{i=1}^{T} \epsilon_i \underline{x}^{\alpha_i}(\underline{x}^{b_i^+} - \underline{x}^{b_i^-}) \in \mathcal{I}_{\mathcal{B}}^{(\mathbb{F}_p)}.$$

By Lemma 11 (this time applied with the ring $R = \mathbb{F}_p$), this implies that pu and pv are connected by $\mathcal{B}$.
Since this works for any prime number bigger than N, one can take two such primes p_1, p_2 and use them in Lemma 19 to conclude that nu and nv are connected for any $n \geqslant p_1 p_2$. □

Unfortunately this type of argument does not work on rings with positive characteristic, as we have no way of reducing modulo two different primes.
This is actually very reasonable, since a positive characteristic makes some powers behave differently. For instance, let us tweak Example 8 so that Lemma 21 does not hold for $\mathrm{char}(R) = p > 0$.

Example 22. *For any prime p, consider the setting:*

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 3 & 2 & 1 & 0 \end{pmatrix}, \qquad \mathcal{B} = \{b_1, b_2, b_3\} := \{ \begin{pmatrix} -1 \\ 1 \\ 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ p \\ -p \\ -p \\ p \end{pmatrix} \},$$

$$u = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, v = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

Since adding $\pm b_1$ or $\pm b_2$ to a multiple of u would produce a vector outside of $\mathbb{N}^r$, b_3 is the only vector that can connect multiples of u and v, therefore $n \cdot u$ is connected to $n \cdot v$ by $\mathcal{B}$ if and only if $p \mid n$. In particular, it is false that $n \cdot u$ and $n \cdot v$ are connected for all large enough values of n.
However, if we take $R := \mathbb{F}_p$, then:

$$\mathcal{I}_{\mathcal{B}} = \langle x_2x_3x_5 - x_1x_4^2, x_1x_4x_6 - x_2x_5^2, x_3^p x_6^p - x_4^p x_5^p \rangle_{R[\underline{x}]} =$$

$$= \langle x_2x_3x_5 - x_1x_4^2, x_1x_4x_6 - x_2x_5^2, (x_3x_6 - x_4x_5)^p \rangle_{R[\underline{x}]}.$$

The element $x_3x_6 - x_4x_5 = \underline{x}^u - \underline{x}^v$ lies therefore in $\sqrt{\mathcal{I}_{\mathcal{B}}}$, even though there exist infinitely many integers n such that $n \cdot u$ and $n \cdot v$ are not connected. [2]

However, for these rings, we get a weaker version; it is not a coincidence that, in the example above, we got $n \cdot u$ being connected to $n \cdot v$ if and only if $p \mid n$:

Lemma 23. *Consider any ring R with prime characteristic $\mathrm{char}(R) = p$ and take $u, v \in \mathcal{F}(u)$. If $\underline{x}^u - \underline{x}^v \in \sqrt{\mathcal{I}_{\mathcal{B}}}$, then there exists T such that nu is connected to nv for every multiple n of p^T.*

[2] Since $\mathcal{B}' := \{b_1, b_2, \frac{1}{p} b_3\}$ is a Markov basis, we actually have the equality $\sqrt{\mathcal{I}_{\mathcal{B}}} = \mathcal{I}_{\mathcal{B}'} = \mathcal{I}_A$, so $\sqrt{\mathcal{I}_{\mathcal{B}}}$ actually contains the binomial $\underline{x}^u - \underline{x}^v$ for all $u, v \in \mathcal{F}(u)$.

Proof. Analogously to the proof of Lemma 21, we are assuming that there exists a natural number N such that $(\underline{x}^u - \underline{x}^v)^N \in \mathcal{I}_{\mathcal{B}}$. Take the smallest integer T such that $p^T \geqslant N$. Then we have:

$$\underline{x}^{p^T \cdot u} - \underline{x}^{p^T \cdot v} = (\underline{x}^u - \underline{x}^v)^{p^T} \in \mathcal{I}_{\mathcal{B}}$$

Which means that $p^T \cdot u$ and $p^T \cdot v$ are connected, and therefore the same holds for any multiple of the two. □

To sum up, we proved the following proposition.

Proposition 24. *Let $A, \mathcal{B}$ be as in Notation 9 and take $u, v \in \mathcal{F}(u)$. Then the following are equivalent:*

1. *There exists $N \in \mathbb{N}$ such that nu and nv are connected for every $n \geqslant N$;*
2. *The ideal $\sqrt{\mathcal{I}_{\mathcal{B}}^{(R)}}$ contains $\underline{x}^u - \underline{x}^v$ for any ring R with $\mathrm{char}(R) = 0$;*
3. *Both ideals $\sqrt{\mathcal{I}_{\mathcal{B}}^{(R_1)}}$ and $\sqrt{\mathcal{I}_{\mathcal{B}}^{(R_2)}}$ contain $\underline{x}^u - \underline{x}^v$ for any two rings R_1, R_2 with different prime characteristic.*

Proof. Thanks to Lemma 18 and Lemma 21, we already know that (1) and (2) are equivalent and that (1) implies (3). For the remaining implication, if (3) holds, then apply Lemma 23 once for each of the two rings: this yields T_1, T_2, p_1, p_2 such that $p_1^{T_1} \cdot u$ is connected to $p_1^{T_1} \cdot v$ and $p_2^{T_2} \cdot u$ is connected to $p_2^{T_2} \cdot v$. Using Lemma 19, we conclude that $n \cdot u$ and $n \cdot v$ are connected for any $n \geqslant p_1^{T_1} p_2^{T_2}$. □

If we ask the condition in the previous proposition to hold for any pair of vectors in the same fibre, we immediately get the following result.

Corollary 25. *Let $A, \mathcal{B}$ be as in Notation 9. Then the following are equivalent:*

1. *For every $u, v \in \mathcal{F}(u)$, there exists $N \in \mathbb{N}$ such that nu and nv are connected for every $n \geqslant N$;*
2. *The equality $\sqrt{\mathcal{I}_{\mathcal{B}}^{(R)}} = \sqrt{\mathcal{I}_A^{(R)}}$ holds for any ring R with $\mathrm{char}(R) = 0$;*
3. *Both equalities $\sqrt{\mathcal{I}_{\mathcal{B}}^{(R_1)}} = \sqrt{\mathcal{I}_A^{(R_1)}}$ and $\sqrt{\mathcal{I}_{\mathcal{B}}^{(R_2)}} = \sqrt{\mathcal{I}_A^{(R_2)}}$ hold for any two rings R_1, R_2 with different prime characteristic.*

Proof. Notice that the inclusion $\sqrt{\mathcal{I}_{\mathcal{B}}} \subseteq \sqrt{\mathcal{I}_A}$ always holds, since $\mathcal{I}_{\mathcal{B}} \subseteq \mathcal{I}_A$.
On the other hand, we have the following equivalences:

$$\underline{x}^u - \underline{x}^v \in \sqrt{\mathcal{I}_{\mathcal{B}}} \ \forall u, v \text{ s.t. } Au = Av \iff \mathcal{I}_A \subseteq \sqrt{\mathcal{I}_{\mathcal{B}}} \xLeftrightarrow{\sqrt{\mathcal{I}_{\mathcal{B}}} \text{ radical}} \sqrt{\mathcal{I}_A} \subseteq \sqrt{\mathcal{I}_{\mathcal{B}}}. \quad (1.4)$$

Together with the other inclusion, this means that the conditions on Line (1.4) are all equivalent to $\sqrt{\mathcal{I}_A} = \sqrt{\mathcal{I}_{\mathcal{B}}}$. Putting this together with the Proposition 24 concludes the proof. □

In particular, thanks to Remark 17, if R is reduced, then we have the previous corollary with the special case of $\mathcal{I}_A$ being radical, and the condition $\sqrt{\mathcal{I}_\mathcal{B}} = \sqrt{\mathcal{I}_A} = \mathcal{I}_A$ becomes equivalent to $\mathcal{I}_\mathcal{B}$ being saturated.
If we go back to the initial question of big fibres being connected, this means the following: if R is reduced and $\mathcal{F}(u)$ is connected for any $u \in \mathbb{N}^r$ such that $||u||_p \geqslant D$, then we can obviously satisfy Statement 1 from Corollary 25 with $N = D$, therefore $\sqrt{\mathcal{I}_\mathcal{B}}$ is saturated and it is equal to $\mathcal{I}_A$.

1.3 Deducing fibre connectedness from other fibres

In this section we want to show how an affine transformation on $\mathcal{F}(u)$ can possibly simplify the problem of connecting the fibres.
Before we go on, let us expand Notation 9.

Notation 26. *Given A and $\mathcal{B}$ as in Notation 9, let $\mathcal{C} \subseteq \mathcal{B}$ be a $\mathbb{Q}$-basis for $\mathrm{Ker}_\mathbb{Q}(A)$ as a $\mathbb{Q}$-vector space. Define $C \in \mathbb{Z}^{r\times t}$ as the matrix whose columns are the vectors of $\mathcal{C}$ (so $t := |\mathcal{C}|$).*

Notice that $\mathcal{B}$ always contains a subset $\mathcal{C}$ as in Notation 26: since $\mathcal{B}$ generates $\mathrm{Ker}_\mathbb{Z}(A)$ as a group, it generates $\mathrm{Ker}_\mathbb{Q}(A) \cong \mathrm{Ker}_\mathbb{Z}(A) \otimes \mathbb{Q}$ as a $\mathbb{Q}$-vector space, so we can extract a basis $\mathcal{C}$ from it. We will omit the $\mathcal{C}$ from the notation where it is clear from context.
Given this set $\mathcal{C}$ and any vector $u \in \mathbb{N}^r$, every element v inside $\mathcal{F}(u)$ (which is just $(u + \mathrm{Ker}_\mathbb{Q}(A)) \cap \mathbb{N}^r$) can now be expressed only in terms of the coordinates of $v - u$ with respect to the basis $\mathcal{C}$. In other words, we are defining the set:

$$\mathcal{G}(u) := \{x \in \mathbb{Q}^t : u + Cx \in \mathcal{F}(u)\}$$

and the bijections:

$$\psi_u : \begin{array}{c} \mathcal{G}(u) \longrightarrow \mathcal{F}(u) \\ x \longmapsto u + Cx \end{array}, \qquad \phi_u : \begin{array}{c} \mathcal{F}(u) \longrightarrow \mathcal{G}(u) \\ v \longmapsto D(v-u) \end{array},$$

where $D \in \mathbb{Q}^{t\times r}$ is a left inverse [3] of C. Notice that these are indeed bijections (and the inverse to one another), as ψ_u is surjective by definition of $\mathcal{G}(u)$ (plus the fact that $\mathcal{C}$ is a basis of the null space) and:

$$(\phi_u \circ \psi_u)(x) = \phi_u(u + Cx) = DCx = I_{t\times t}x = x \qquad \forall x \in \mathcal{G}(u).$$

But now we need to translate the concept of connectedness from $\mathcal{F}(u)$ to our new set $\mathcal{G}(u)$. In order for us to do that, let us introduce the following graph structure on $\mathcal{G}(u)$.

Notation 27. *Given $A, \mathcal{B}, \mathcal{C}$ as in Notation 26 and $u \in \mathbb{N}^r$, set $n := |\mathcal{B}|$ and $\mathcal{Z} := \{z_1 \ldots, z_n\}$ with $z_i := Db_i$. From now on, $\mathcal{G}(u)$ will implicitly have the graph structure $(\mathcal{G}(u), L)$ given by:*

$$L := \{(x, y) \in \mathcal{G}(u)^2 : x - y \in \pm\mathcal{Z}\}.$$

Analogously to the case of $\mathcal{F}(u)$, connectedness according to this kind of graph will be referred to as "connectedness by $\mathcal{Z}$".

[3] It does not matter which inverse D we pick, as all left inverses induce the same linear map over the column space of C (i.e. $\mathrm{Ker}_\mathbb{Q}(A)$)

What we introduced is indeed the corresponding graph to the one on $\mathcal{F}(u)$, as one can see from the following lemma.

Lemma 28. *Given $v, w \in \mathcal{F}(u)$, v is connected to w by $\mathcal{B}$ if and only if $\phi_u(v)$ is connected to $\phi_u(w)$ by $\mathcal{Z}$ in $\mathcal{G}(u)$.*

Proof. It suffices to show that, given $v, w \in \mathcal{F}(u)$, we have $w - v \in \pm\mathcal{B}$ if and only if $\phi_u(w) - \phi_u(v) \in \pm\mathcal{Z}$.
Since D is a left inverse of C, the linear map induced by D is a bijection from the column space of C to $\mathbb{Q}^t$. Since $\pm\mathcal{B} \subseteq \mathrm{Ker}_{\mathbb{Q}}(A)$ and since the columns of C span $\mathrm{Ker}_{\mathbb{Q}}(A)$, we have:

$$w - v \in \pm\mathcal{B} \iff \pm\mathcal{Z} = \pm D\mathcal{B} \ni D(w - v) = D(w - u) - D(v - u) = \phi_u(w) - \phi_u(v),$$

which concludes the proof. □

In other words, this lemma states that our problem did not change much: using $\mathcal{G}(u)$ instead of $\mathcal{F}(u)$ means that we are jumping using different jumps ($z_1, \dots, z_n$ instead of $b_1, \dots, b_n$) on a different set. So why is this in any way helpful?
Firstly, it may be interesting to notice that, in the lemma above, when $b_i \in \mathcal{C}$ we actually get z_i being a vector e_j of the standard basis of $\mathbb{Q}^t$. This means that, for instance, given $v, w \in \mathcal{F}(u)$, if $\phi_u(v), \phi_u(w) \in \mathbb{Z}^r$ then we immediately know that reaching v from u takes at most $||\phi_u(v) - \phi_u(w)||_1$ steps inside $u + \mathrm{Ker}_{\mathbb{Z}}(A)$. The concept of "minimal number of steps" to connect two points in $\mathcal{F}(u)$ will come back later on in the thesis, from the discussion on page 18 and the subsequent Definition 36.
Most importantly though, this change of coordinates allows us to determine interesting relationships between different graphs $\mathcal{G}(u)$. For instance, we are going to prove the two following results, which can potentially say a lot about the connectedness of $\mathcal{G}$ and therefore $\mathcal{F}$.

Lemma 29. *Let $A, \mathcal{B}, \mathcal{C}$ be as in Notation 26 and consider any two vectors $u, v \in \mathbb{N}^r$. Then:*

$$\mathcal{G}(u) \cap \mathcal{G}(v) = \mathcal{G}(w), \qquad \text{where } w_i := \min(u_i, v_i)\ \forall i \in \{1, \dots, r\}.$$

Theorem 30. *Let $A, \mathcal{B}, \mathcal{C}$ be as in Notation 26 and let $u \in \mathbb{N}^r$ be such that $\mathcal{F}(u - e_j)$ is connected for every j for which $u - e_j \in \mathbb{N}^r$.*
Then $\mathcal{F}(u) \cap \{v \in \mathcal{F}(u) : u_i v_i = 0 \forall i\}^c$ is connected.

Before we state the proof to these results, we may have to re-write $\mathcal{G}(u)$ in more explicit terms. Remember that, by its definition, $\mathcal{G}(u)$ contains a rational vector x if and only if $u + Cx$ is in $\mathcal{F}(u)$, i.e. if $u + Cx \in \mathbb{N}^r$. [4] In other words, the elements in $\mathcal{G}(u)$ are the vectors $(x_1, \dots, x_t) \in \mathbb{Q}^t$ such that:

$$\begin{cases} u_1 + \sum_{i=1}^t c_{1,i} x_i \geqslant 0,\ \sum_{i=1}^t c_{1,i} x_i \in \mathbb{Z} \\ \vdots \\ u_r + \sum_{i=1}^t c_{r,i} x_i \geqslant 0,\ \sum_{i=1}^t c_{r,i} x_i \in \mathbb{Z} \end{cases} . \tag{1.5}$$

[4] Since the column space of C is $\mathrm{Ker}_{\mathbb{Q}}(A)$, it is automatic that $u + Cx$ will still have the same image under A as u.

Notice that here the notation $c_{j,i}$ is referring to an entry of the matrix C, whose columns are the vectors $c_1, \ldots, c_t$. In other words, $c_{j,i} = (c_i)_j$ is the j-th entry of the vector c_i. Line (1.5) can be re-expressed as:

$$\begin{cases} u_1 + \sum_{i=1}^{t} c_{1,i} x_i = k_1 \in \mathbb{N} \\ \vdots \\ u_r + \sum_{i=1}^{t} c_{r,i} x_i = k_r \in \mathbb{N} \end{cases}.$$

If we use the intersection of solution sets instead of the system of equations and if we allow every k_i to be any natural number, we get:

$$\mathcal{G}(u) = \bigcap_{j=1}^{r} \Big(\bigcup_{k_j \in \mathbb{N}} \{\underline{x} \in \mathbb{Q}^t : u_j + \sum_{i=1}^{t} c_{j,i} x_i = k_j\} \Big).$$

If we introduce the notation $H_j(\alpha)$ to denote the hyperplane in $\mathbb{Q}^t$ defined by $\sum_{i=1}^{t} c_{j,i} x_i = \alpha$, then this can be written as:

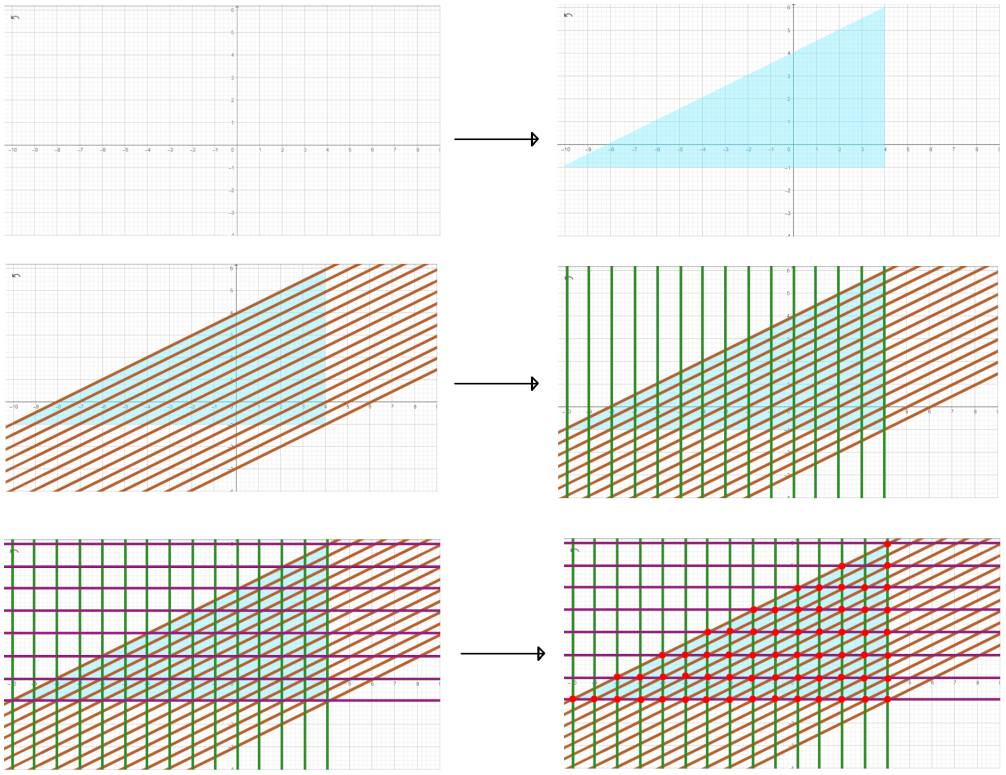
$$\mathcal{G}(u) = \bigcap_{j=1}^{r} \Big(\bigcup_{k_j \in \mathbb{N}} H_j(k_j - u_j) \Big) \subseteq \mathbb{Q}^t. \tag{1.6}$$

Les us see an example to better visualize the set $\mathcal{G}$.

Example 31. *Consider the following setting:*

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \end{pmatrix}, \qquad \mathcal{B} = \mathcal{C} = \{b_1, b_2\} := \{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \}, \qquad u := \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix}.$$

Then the set $\mathcal{G}(u)$ is depicted below. The initial light blue triangle is the rational polyhedron defined by only the inequalities in Line (1.5) and it contains $\mathcal{G}(u)$. The parallel lines are the hyperplanes $H_1(k_1), H_2(k_2)$ and $H_3(k_3)$ for all the relevant values of k_1, k_2, k_3, in order of appearance. The red dots are all the possible intersections of three such hyperplanes, i.e. the points of $\mathcal{G}(u)$.



Now we prove the statements above.

Proof(Lemma 29). Using Line (1.5), we know that $\mathcal{G}(u) \cap \mathcal{G}(v)$ is defined by the following systems of equations:

$$\begin{cases} \sum_{i=1}^{t} c_{1,i} x_i \in \mathbb{Z} \\ \vdots \\ \sum_{i=1}^{t} c_{r,i} x_i \in \mathbb{Z} \end{cases}, \quad \begin{cases} \sum_{i=1}^{t} c_{1,i} x_i \geqslant -u_1 \\ \vdots \\ \sum_{i=1}^{t} c_{r,i} x_i \geqslant -u_r \end{cases}, \quad \begin{cases} \sum_{i=1}^{t} c_{1,i} x_i \geqslant -v_1 \\ \vdots \\ \sum_{i=1}^{t} c_{r,i} x_i \geqslant -v_r \end{cases}.$$

The last two systems can be joined into:

$$\begin{cases} \sum_{i=1}^{t} c_{1,i} x_i \geqslant \max(-u_1, -v_1) = -\min(u_1, v_1) = -w_1 \\ \vdots \\ \sum_{i=1}^{t} c_{r,i} x_i \geqslant \max(-u_r, -v_r) = -\min(u_r, v_r) = -w_r \end{cases},$$

which means that $\mathcal{G}(u) \cap \mathcal{G}(v)$ is defined by:

$$\begin{cases} \sum_{i=1}^{t} c_{1,i} x_i \in \mathbb{Z} \\ \vdots \\ \sum_{i=1}^{t} c_{r,i} x_i \in \mathbb{Z} \end{cases}, \quad \begin{cases} w_1 + \sum_{i=1}^{t} c_{1,i} x_i \geqslant 0 \\ \vdots \\ w_r + \sum_{i=1}^{t} c_{r,i} x_i \geqslant 0 \end{cases}.$$

Which, again by Line (1.5), is exactly how we define $\mathcal{G}(w)$. $\square$

Using this lemma, we now proceed with the proof of the theorem.

Proof(Theorem 30). Let J be the set of indices j such that $u-e_j \in \mathbb{N}^r$. We are assuming $\mathcal{B}$ to connect all the graphs $\mathcal{F}(u-e_j)$, and therefore $\mathcal{Z}$ to connect all the graphs $\mathcal{G}(u-e_j)$. Notice that, by Lemma 29, these graphs are not disjointed:

$$\bigcap_{j\in J} \mathcal{G}(u-e_j) = \mathcal{G}(w), \qquad \text{where } w_i = \min_{j\in J}(u_i - \delta_{i,j}) \in \mathbb{N}.$$

Since $w \in \mathbb{N}^r$ is always contained in $\mathcal{F}(w)$, this means that $0 \in \mathcal{G}(w)$, so in particular the intersection above is non-empty. So if we take the union of these graphs, namely $\bigcup_{j\in J} \mathcal{G}(u-e_j)$, this subset of $\mathcal{G}(u)$ is going to be a non-disjoint union of connected graphs and therefore it will also be connected. By Lemma 28, this means that the subgraph $\psi_u(\bigcup_{j\in J} \mathcal{G}(u-e_j))$ is also going to be connected. We only need to prove the latter to be equal to $\mathcal{F}(u) \cap \{v \in \mathcal{F}(u) : u_i v_i = 0 \forall i\}^c$.
Firstly, let us use the form obtained in Line (1.6) to notice that:

$$\mathcal{G}(u-e_j) = \bigcap_{i\neq j}^{r} \Big(\bigcup_{k_i\in\mathbb{N}} H_i(k_i-u_i)\Big) \cap \big(\bigcup_{k_j\in\mathbb{N}} H_j(k_j-u_j+1)\big) =$$

$$= \bigcap_{i\neq j}^{r} \Big(\bigcup_{k_i\in\mathbb{N}} H_i(k_i-u_i)\Big) \cap \Big(\big(\bigcup_{k_j\in\mathbb{N}} H_j(k_j-u_j)\big) \cap H_j(-u_j)^c\Big) =$$

$$\bigcap_{i=1}^{r} \Big(\bigcup_{k_i\in\mathbb{N}} H_i(k_i-u_i)\Big) \cap H_j(-u_j)^c = \mathcal{G}(u) \cap H_j(-u_j)^c,$$

which now means that

$$\bigcup_{j\in J} \mathcal{G}(u-e_j) = \bigcup_{j\in J} (\mathcal{G}(u) \cap H_j(-u_j)^c) = \mathcal{G}(u) \cap \big(\bigcup_{j\in J} H_j(-u_j)^c\big) = \mathcal{G}(u) \cap \Big(\bigcap_{j\in J} H_j(-u_j)\Big)^c.$$

This allows us to conclude that the connected set is:

$$\psi_u\big(\bigcup_{j\in J} \mathcal{G}(u-e_j)\big) = \psi_u\big(\mathcal{G}(u) \cap \big(\bigcap_{j\in J} H_j(-u_j)\big)^c\big) =$$

$$= \psi_u(\mathcal{G}(u)) \cap \psi_u\big(\big(\bigcap_{j\in J} H_j(-u_j)\big)^c\big) = \mathcal{F}(u) \cap \psi_u\Big(\bigcap_{j\in J} H_j(-u_j)\Big)^c.$$

But now notice that, for every $j \in J$, by very definition of $H_j(-u_j)$, we have that $x \in H_j(-u_j)$ if and only if $\sum_{i=1}^{t} c_{j,i}x_i = -u_j$, which is equivalent to:

$$0 = u_j + \sum_{i=1}^{t} c_{j,i}x_i = (u+Cx)_j = (\psi_u(x))_j.$$

This implies that an element x is in $\bigcap_{j\in J} H_j(-u_j)$ if and only if the equalities $(\psi_u(\underline{x}))_j = 0$ hold for all $j \in J$, so we have that:

$$\psi_u\big(\bigcup_{j\in J} \mathcal{G}(u-e_j)\big) = \mathcal{F}(u) \cap \{v \in \mathcal{F}(u) : v_j = 0 \,\forall j \in J\}^c.$$

Now remember that J was the set of all indices such that $u - e_j \in \mathbb{N}^r$, i.e. $u_j \geqslant 1$, i.e. $u_j \neq 0$. So this means that, in the set $\{v \in \mathcal{F}(u) : v_j = 0 \, \forall j \in J\}$, a component v_i can be any natural number if $u_i = 0$, but it has to be 0 if $u_i \neq 0$. This means that the set above can be re-written as:

$$\psi_u(\bigcup_{j \in J} \mathcal{G}(u - e_j)) = \mathcal{F}(u) \cap \{v \in \mathcal{F}(u) : u_i v_i = 0 \, \forall i = 1, \dots, r\}^c.$$

We are saying that this subgraph is connected, which is what we wanted to prove. □

An immediate application of this theorem is the following result.

Corollary 32. *Let $u \in \mathbb{N}^r$ be a vector such that:*

- *$u_i \geqslant 1$ for every $i = 1, \dots, r$;*
- *$u \notin \mathrm{Ker}_{\mathbb{Z}}(A)$;*
- *$\mathcal{F}(u - e_i)$ is connected for every $i \in \{1, \dots, r\}$.*

Then $\mathcal{F}(u)$ is connected.

Proof. Using Theorem 30, in this case with $J = \{1, \dots, r\}$, we know that the only possible element in $\mathcal{F}(u)$ that is not connected to u would be $\underline{0} \in \mathbb{N}^r$. However, since $u \notin \mathrm{Ker}_{\mathbb{Z}}(A)$, we conclude that $\underline{0} \notin \mathcal{F}(u)$ and therefore $\mathcal{F}(u)$ is entirely connected. □

This corollary, in particular, says that we only need the connectedness of fibres of elements with at least a zero entry, as the remaining fibres automatically become connected.
Theorem 30 and Corollary 32 may not entirely come as a surprise since, after all, up to dividing by $\mathrm{G.C.D.}(\underline{x}^u, \underline{x}^v)$, the vectors u and v appearing in $\underline{x}^u - \underline{x}^v$ are always such that $u_i v_i = 0$ for all i. However, being able to divide by a power of $\underline{x}$ while being sure not to be altering the presence in the ideal is only possible with saturated ideals, so this argument only makes sense if we are looking for $\mathcal{I}_{\mathcal{B}}$ to be equal to $\mathcal{I}_A$. This means that the theorem and corollary above were not only helpful in properly formalizing this concept, but they also cover more general cases, where we ask for only some of the fibres to be connected.
Also, thanks to Theorem 30, even the remaining case of connecting u and v with $u_i v_i = 0$ for all i is simplified: we can still look for a "bridge" element w such that $w_j u_j \neq 0$ and $w_i v_i \neq 0$ for some suitable i, j. In this case, by applying Theorem 30 twice, u would be connected to v by a path that goes through w.
Towards this direction, it may be interesting to notice that if $u \in \mathcal{F}(u)$ admits a "bridge" element for a certain vector v, then so does $u + w \in \mathcal{F}(u + w)$ for the vector $v + w$, for whatever vector $w \in \mathbb{N}^r$, so in a sense one only needs to look up "small" graphs $\mathcal{G}(u)$.

1.4 Checking the connectedness of a finite number of fibres

As the title suggests, in this section we continue the previous work of trying to understand which fibres are "key" to the connectedness of other fibres.

In particular, we are interested in what we will often call "small fibres", namely the fibres of the form $\mathcal{F}(u)$ for a small vector u.
Firstly, let us notice that it is indeed possible to only check small fibres, as suggested by the following remark.

Remark 33. *Given A and $\mathcal{B}$ as in Notation 9, fix a norm $||\cdot||$ on $\mathbb{Q}^r$. Then there exists a positive number $T \in \mathbb{R}$ such that the condition:*

$$\mathcal{F}(u) \text{ is connected for every } u \text{ such that } ||u|| \leqslant T$$

implies $\mathcal{F}(u)$ being connected for any $u \in \mathbb{N}^r$.

Proof. Choose the ring R as a field. By Corollary 4.4 from [4], we know that there is a finite set of pure binomials $\{\underline{x}^{d^+} - \underline{x}^{d^-}\}_{d\in\mathcal{D}}$ that generates $\mathcal{I}_A$, with $\mathcal{D} \subseteq \mathrm{Ker}_{\mathbb{Z}}(A)$. By Lemma 11, the element $\underline{x}^{d^+} - \underline{x}^{d^-}$ belonging to $\mathcal{I}_{\mathcal{B}}$ is equivalent to d^+ being connected to d^- by the set $\mathcal{B}$, for any $d \in \mathcal{D}$. In particular, if all fibres $\mathcal{F}(d^+)$ are connected for any $d \in \mathcal{D}$, then $\mathcal{I}_{\mathcal{B}} = \mathcal{I}_A$ and therefore every fibre is connected.
In other words, if we set $T := \max_{d\in\mathcal{D}} ||d^+||$, then it suffices to check that $\mathcal{F}(u)$ is connected for any vector u such that $||u|| \leqslant T$. □

Unfortunately this proof does not give us an estimation unless we explicitly have $\mathcal{D}$, at which point we would already have a Markov basis to connect every fibre and any further discussion would be pointless anyway.
In the rest of this section we are going to introduce and then prove a bound for such quantity. However, we are going to need to generalize a previous definition first.

Definition 34. *Let A and $\mathcal{B}$ be as in Notation 9, take a natural number N. Given $u \in \mathbb{N}^r$, we denote by $\mathcal{F}_N(u)$ the undirected graph whose vertices are the same as $\mathcal{F}(u)$, but whose set of edges is:*

$$E_N := \{(v,w) \in \mathcal{F}(u)^2 : \exists m \leqslant N, b_1, \dots, b_m \in \mathcal{B},\ \epsilon_1, \dots, \epsilon_m \in \{\pm 1\} \text{ s.t. } v - w = \sum_{i=1}^{m} \epsilon_i b_i\}.$$

So, essentially, we are saying that we can take some steps outside $\mathcal{F}(u)$ to connect points, now.
Of course, this can translate analogously to $\mathcal{G}$.

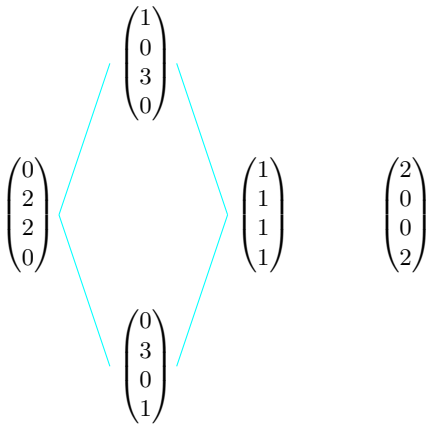
Definition 35. *Let $A, \mathcal{B}, \mathcal{C}$ be as in Notation 26. Take $\mathcal{Z}$ as in Notation 27 and consider a natural number N. Given $u \in \mathbb{N}^r$, we denote by $\mathcal{G}_N(u)$ the undirected graph whose vertices are the same as $\mathcal{G}(u)$, but whose set of edges is:*

$$L_N := \{(v,w) \in \mathcal{G}(u)^2 : \exists m \leqslant N, z_1, \dots, z_m \in \pm\mathcal{Z},\ \epsilon_1, \dots, \epsilon_m \in \{\pm 1\} \text{ s.t. } v - w = \sum_{i=1}^{m} \epsilon_i z_i\}.$$

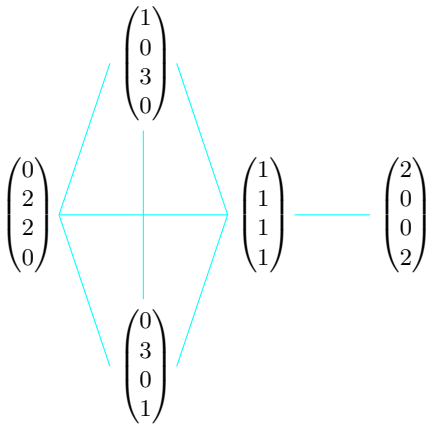
If we compare $\mathcal{F}(u)$ with $\mathcal{F}_2(u)$ and $\mathcal{G}(u)$ with $\mathcal{G}_2(u)$, for the following setting:

$$A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{pmatrix}, \qquad \mathcal{B} := \mathcal{C} := \{\begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -2 \\ 1 \end{pmatrix}\}, \qquad u = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

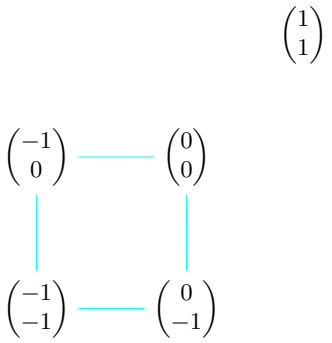
then we get the following pictures:



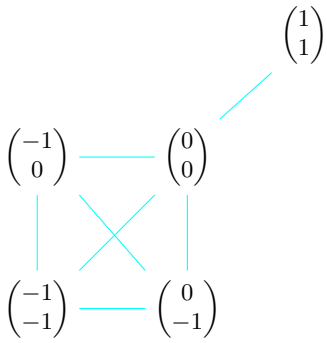
The graph $\mathcal{F}(u)$.



The graph $\mathcal{F}_2(u)$.



The graph $\mathcal{G}(u)$.



The graph $\mathcal{G}_2(u)$.

There are two interesting remarks that we want to make at this point:

- When $N = 1$, we are just taking $\mathcal{F}_N(u) = \mathcal{F}(u)$ as in the previous definition;
- If we set $\mathcal{D} := \{\sum_{i=1}^{m} \epsilon_i b_i : m \leqslant N, b_i \in \mathcal{B}, \epsilon_i = \pm 1\}$, then taking $\mathcal{F}_N(u)$ with respect to $\mathcal{B}$ is the same as taking $\mathcal{F}(u)$ with respect to $\mathcal{D}$, so all the considerations made for $\mathcal{F}(u)$ actually hold for $\mathcal{F}_N(u)$ as well. For instance, Corollary 32 says that if N steps are sufficient to connect every $\mathcal{F}(u - e_i)$ then they are also sufficient to connect $\mathcal{F}(u)$. This is also why, from now on, a lot of the arguments will only be made for $\mathcal{F}(u)$, even though they all also hold for $\mathcal{F}_N(u)$.

At this point one may want to weaken the initial task of looking for a set $\mathcal{B}$ that connects $\mathcal{F}(u)$ and, instead, look for a set $\mathcal{B}$ that connects $\mathcal{F}_N(u)$ for a small N, i.e. in "only a few steps outside the graph". In this case, we need to properly define the number of steps needed, so we introduce the following definition.

Definition 36. *Given a matrix $A \in \mathbb{N}^{s\times r}$ and a generating set $\mathcal{B}$ of $\mathrm{Ker}_{\mathbb{Z}}(A)$ as a group, we define $\nu(\mathcal{B})$ as:*

$$\nu(\mathcal{B}) := \min(\{N \in \mathbb{N} : \mathcal{F}_N(u) \text{ is connected } \forall u \in \mathbb{N}^r\}).$$

We notice immediately that we are taking the maximum of a non empty set thanks to Proposition 1.4 of the paper [3], which also implies the inequality $\nu(\mathcal{B}) \leqslant ||\mathcal{B}||$ (here the notation of "norm of $\mathcal{B}$" is the same as in [3] and we give its definition again in Definition 43).
Now, we are finally ready to introduce our theorem.

Theorem 37. *Let $A, \mathcal{B}$ be as in Notation 9 and take $p \in \mathbb{R}_{\geqslant 1} \cup \{\infty\}$. Assume $\mathcal{F}(w)$ to be connected by $\mathcal{B}$ for every vector $w \in \mathbb{N}^r$ that has norm:*

$$||w||_p \leqslant \frac{\nu(\mathcal{B})M}{2^{1/p}} \qquad \text{where } M = \max_{b\in\mathcal{B}} ||b||_p,$$

with $\frac{1}{\infty}$ being 0, as usual. Then $\mathcal{F}(w)$ is connected for every vector $w \in \mathbb{N}^r$.

Before we prove the theorem, we need a couple of useful lemmas.
When trying to connect two vectors u and v, the idea behind Theorem 37 is to see u as the sum of a big component that remains the same and a smaller one that can be connected because it has small norm (and we are covering the smaller cases).
However, this only makes sense if connectedness behaves well with the sum.

Lemma 38. *Assume $u, v \in \mathcal{F}(u)$ to be connected and assume the same for $t, w \in \mathcal{F}(t)$. Then the vectors $u + t, v + w \in \mathcal{F}(u + t)$ are also connected.*

Proof. There are two interesting, easy ways to prove this: the first one is constructive; assume the vectors to be connected through the chains:

$$u = u_1 \to u_2 \to \cdots \to u_a = v, \qquad t = t_1 \to t_2 \to \cdots \to t_b = w.$$

Then $u + t$ is connected to $v + w$ through the chain:

$$u + t \to u_2 + t \to \cdots \to v + t \to v + t_2 \to \cdots \to v + w.$$

Another easy argument can be made by using Lemma 11 to conclude that:

$$\begin{cases} u, v \text{ connected} \\ t, w \text{ connected} \end{cases} \iff \begin{cases} \underline{x}^u - \underline{x}^v \in \mathcal{I}_{\mathcal{B}} \\ \underline{x}^t - \underline{x}^w \in \mathcal{I}_{\mathcal{B}} \end{cases}.$$

The right-hand side implies that:

$$\underline{x}^{u+t} - \underline{x}^{v+w} = \underline{x}^u(\underline{x}^t - \underline{x}^w) + \underline{x}^w(\underline{x}^u - \underline{x}^v) \in \mathcal{I}_{\mathcal{B}}$$

which, again by Lemma 11, means that $u + t, v + w$ are connected. □

Now we need to better formalize the concept of splitting u into the sum of two components, which is what we will do thanks to the following corollary.

Corollary 39. *Take $u, v \in \mathcal{F}(u)$. Since $u - v \in \mathrm{Ker}_{\mathbb{Z}}(A)$, we already know that $(u - v)^+, (u-v)^- \in \mathcal{F}((u-v)^+)$. Assume $(u-v)^+$ and $(u-v)^-$ to be connected. Then u and v are connected.*

Proof. Write u as:

$$u = u - v + v = (u-v)^+ - (u-v)^- + v = (u-v)^+ + \big(v - (u-v)^-\big) = (u-v)^+ + a,$$

where $a := \big(v-(u-v)^-\big)$. We want to apply Lemma 38, so we need to prove that $a \in \mathbb{N}^r$. This is true because:

$$a = v - (u-v)^- = (v_i - \max(v_i - u_i, 0))_{i=1,\dots,r} = (v_i + \min(u_i - v_i, 0))_{i=1,\dots,r} =$$

$$= (\min(u_i, v_i))_{i=1,\dots,r} \in \mathbb{N}^r.$$

So now we have that u is the sum of $(u-v)^+$ and a, both in $\mathbb{N}^r$. The former is connected to $(u-v)^-$ by hypothesis, while the latter is obviously connected to itself, therefore by Lemma 38 u is connected to:

$$(u-v)^- + a = (u-v)^- + \Big(v - (u-v)^-\Big) = v,$$

which is what we wanted to prove. □

So this means that, instead of having to connect two vectors u and v, we can try connecting $(u-v)^+$ and $(u-v)^-$, which are smaller.
So now we are finally ready to prove our theorem.

Proof(Theorem 37). Given any pair of vectors $u, v \in \mathcal{F}(w)$, we know that by definition of $\nu(\mathcal{B}) =: N$ there must be a path that connects u and v in $\mathcal{F}_N(u)$, so there exists a chain:

$$u_0 = u \to u_1 \to \cdots \to u_k = v \qquad \text{with } u_i - u_{i-1} = \sum_{j=1}^{N_i} \epsilon_j^{(i)} b_j^{(i)},\ N_i \leqslant N \quad \forall i = 1, \dots, k$$

where $\epsilon_j^{(i)} \in \{\pm 1\}, b_j^{(i)} \in \mathcal{B}$ for any i, j. For every i we have therefore that:

$$||u_i - u_{i-1}||_p = || \sum_{j=1}^{N_i} \epsilon_j^{(i)} b_j^{(i)}||_p \leqslant \sum_{j=1}^{N_i} ||b_j^{(i)}||_p \leqslant \sum_{j=1}^{N_i} \max_{b\in\mathcal{B}} ||b||_p = N_i \max_{b\in\mathcal{B}} ||b||_p \leqslant NM.$$

Now let us separate the proof into two different cases:

- If $p = +\infty$, then notice that $u_i - u_{i-1}$ has all the non-zero components of the vector $(u_i - u_{i-1})^+$, plus some extra non-zero components coming from its negative part $(u_i - u_{i-1})^-$. So this means that:

  $$||(u_i - u_{i-1})^+||_\infty \leqslant ||u_i - u_{i-1}||_\infty \leqslant N \max_{b\in\mathcal{B}} ||b||_\infty = \frac{\max_{b\in\mathcal{B}} ||b||_\infty}{2^0} = \frac{NM}{2^{1/p}},$$

  which means that by hypothesis $\mathcal{F}\big((u_i - u_{i-1})^+\big)$ is connected.

- If $p < +\infty$ then notice that when we write $u_i - u_{i-1}$ as $u_i - u_{i-1} = (u_i - u_{i-1})^+ - (u_i - u_{i-1})^-$, we are actually writing it as a sum of two vectors with disjoint supports, so:

$$||(u_i - u_{i-1})^+||_p^p + ||(u_i - u_{i-1})^-||_p^p =$$

$$= \sum_{\substack{j=1,\dots,r \\ (u_i - u_{i-1})_j > 0}} |(u_i - u_{i-1})_j^+|^p + \sum_{\substack{j=1,\dots,r \\ (u_i - u_{i-1})_j < 0}} |(u_i - u_{i-1})_j^-|^p =$$

$$= \sum_{\substack{j=1,\dots,r \\ (u_i - u_{i-1})_j \neq 0}} |(u_i - u_{i-1})_j|^p = ||u_i - u_{i-1}||_p^p \leqslant N^p M^p.$$

  This inequality has the sum of two non-negative numbers on the left-hand side, therefore at least one of them is at most $\frac{N^p M^p}{2}$. Without loss of generality, we can assume $||(u_i - u_{i-1})^+||_p^p \leqslant \frac{N^p M^p}{2}$, which implies:

$$||(u_i - u_{i-1})^+||_p \leqslant \frac{NM}{2^{1/p}}.$$

  So this also means that by hypothesis $\mathcal{F}\big((u_i - u_{i-1})^+\big)$ is connected.

So in both cases we end up with $\mathcal{F}\big((u_i - u_{i-1})^+\big)$ being connected, which means that $(u_i - u_{i-1})^+$ is connected to $(u_i - u_{i-1})^-$ and therefore, by Corollary 39, u_i is connected to u_{i-1} in $\mathcal{F}(u)$. This holds for every i, therefore u is connected to u_1, which is connected to u_2 and so on up until v. By transitivity, this means that u and v are connected.
Since this holds for any pair u, v in any fibre $\mathcal{F}(w)$, it means that the fibres are all connected. □

We have to make two remarks about the proof of this theorem:

- If, for some index i, we have $u_i - u_{i-1} = N_i b^{(i)}$, then we do not need the reasoning above to conclude that u_i and u_{i-1} are connected, since in this special case they would be connected by the chain:

$$u_{i-1} \longrightarrow u_{i-1} + b^{(i)} \longrightarrow \cdots \longrightarrow u_{i-1} + N_i b^{(i)} = u_i.$$

  So this means that we can focus on the case where, once we fix i, the vectors $b_j^{(i)}$ are not all equal. This might be relevant if only one element has the highest norm of $\mathcal{B}$, which we called M. Indeed, instead of using NM, we can use $(N-1)M + M'$ with M' being the second highest norm of the vectors of $\mathcal{B}$ (which is therefore still equal to M if more than one vector in $\mathcal{B}$ has norm M).

- Since the proof of Theorem 37 is essentially based on the connectedness of fibres of the form $\mathcal{F}(b^+)$ or $\mathcal{F}(b^-)$ and since it is impossible for both b^+ and b^- not to have zero entries, it suffices to only check the connectedness of the fibres $\mathcal{F}(u)$ with u satisfying the conditions of Theorem 37 *and* having at least one zero entry. This also makes sense when compared to Corollary 32.

Using an analogous proof, one can also prove the following result.

Theorem 40. *Let $A, \mathcal{B}$ be as in Notation 9 and take $p \in \mathbb{R}_{\geqslant 1} \cup \{+\infty\}$. Then the set:*

$$\mathcal{B}' := \{v \in \mathrm{Ker}_{\mathbb{Z}}(A) : ||v||_p \leqslant \nu(\mathcal{B}) \cdot M\}, \qquad \text{where } M = \max_{b \in \mathcal{B}} ||b||_p,$$

is a Markov basis.

Proof. As proved in the proof of Theorem 37, we already know that, for any two vectors $u, v \in \mathcal{F}(u)$, there is a chain of vectors $u_0 = u, \ldots, u_k = v$ such that:

$$||u_i - u_{i-1}||_p \leqslant \nu(\mathcal{B}) \cdot M \qquad \forall i = 1, \ldots, k$$

Since of course $u_i - u_{i-1} \in \mathrm{Ker}_{\mathbb{Z}}(A)$, this implies that $u_i - u_{i-1} \in \mathcal{B}'$ for every i, and therefore u and v are connected by $\mathcal{B}'$. □

Using Theorem 1.3 from [3], we know that $\nu(\mathcal{B}) \leqslant ||\mathcal{B}|| \leqslant n^{n+1}\beta^{n-1}$ and therefore, if we apply the two last theorems with $p = +\infty$, we have the following corollaries.

Corollary 41. *Let $A, B, \mathcal{B}$ be as in Notation 9 and set $\beta := \max_{\substack{i=1,\ldots,r \\ j=1,\ldots,n}} |b_{i,j}|$. Assume $\mathcal{F}(w)$ to be connected by $\mathcal{B}$ for every vector $w \in \mathbb{N}^r$ that has norm:*

$$||w||_\infty \leqslant n^{n+1}\beta^n.$$

Then $\mathcal{F}(w)$ is connected for every vector $w \in \mathbb{N}^r$.

Corollary 42. *Let $A, B, \mathcal{B}$ be as in Notation 9 and set $\beta := \max_{\substack{i=1,\ldots,r \\ j=1,\ldots,n}} |b_{i,j}|$. Then the set:*

$$\mathcal{B}' := \{v \in \mathrm{Ker}_{\mathbb{Z}}(A) : |v_i| \leqslant n^{n+1}\beta^n \ \forall i = 1, \ldots, r\}$$

is a Markov basis.

These very interesting results push us to try better understand (and bound) the quantity $\nu(\mathcal{B})$, which appears in these results.

Chapter 2

Improving the bound on the norm

2.1 A direct bound

As already noticed right after Definition 36, we already have a bound for $\nu(\mathcal{B})$ to start from, which is given by $||\mathcal{B}||$ as defined in [3]. Let us see its definition.

Definition 43. *Given $R = \mathbb{Z}$, consider $A, \mathcal{B}$ as in Notation 9. We then define the norm of $\mathcal{B}$ as the smallest integer N for which there exists a set $G \subseteq \mathcal{I}_A$ such that:*

1. *The ideal $\mathcal{I}_A$ is generated by G;*
2. *Every element of G is a pure difference binomial, i.e. of the form $\underline{x}^u - \underline{x}^v$ for $u, v \in \mathbb{N}^r$;*
3. *Every $g \in G$ can be written in the form:*

$$g = \sum_{i=1}^{N} \epsilon_i m_i (\underline{x}^{b_i^+} - \underline{x}^{b_i^-})$$

where the $\epsilon_i \in \{-1, 0, 1\}$, the m_i are Laurent monomials, and the b_i are elements of $\mathcal{B}$.

Notice that the definition of this quantity is being given on the ring $\mathbb{Z}$. We have already seen in Theorem 13 that a set being a Markov basis (and therefore generating the ideal $\mathcal{I}_A$) does not depend on the choice of the ring, so the set G from Definition 43 always generates $\mathcal{I}_A$, even if we change ring. However, if we changed ring R, we would not be able to establish a priori whether the number N would remain the same, so we can not be sure whether the definition above would depend on the choice of the ring R. The answer is that the choice of the ring is irrelevant, since we actually have the following remark, whose proof does not depend on the choice of the ring.

Remark 44. *Given $A, \mathcal{B}$ as in Notation 9, we have $\nu(\mathcal{B}) = ||\mathcal{B}||$.*

Proof. We already know from Proposition 1.4 in [3] that $\nu(\mathcal{B}) \leqslant ||\mathcal{B}||$. Thanks to Lemma 15, we know that its proof does not depend on the choice of the ring R.
For the other inequality, we need to find a set G such that the conditions stated in Definition 43 hold for $N := \nu(\mathcal{B})$. In order to exhibit such a set, let us take a Markov basis $\mathcal{H} = \{h_1, \dots, h_l\}$, so $\mathcal{I}_A = \mathcal{I}_{\mathcal{H}}$.
By definition of $\nu(\mathcal{B})$, every h_i^+ is connected to h_i^- in $\mathcal{F}_N(h_i^+)$, therefore for every i we have a path:

$$h_i^+ =: a_0^{(i)} \to a_1^{(i)} \to \cdots \to a_{t_i}^{(i)} := h_i^- \qquad \text{with } a_k^{(i)} - a_{k-1}^{(i)} = \sum_{j=1}^{N_{i,k}} \epsilon_j^{(i,k)} b_j^{(i,k)} \ \forall k = 1, \dots, t_i.$$

Let us now prove that the set:

$$G := \left\{ \underline{x}^{a_k^{(i)}} - \underline{x}^{a_{k-1}^{(i)}} : i \in \{1, \dots, l\}, k \in \{1, \dots, t_i\} \right\}$$

satisfies the conditions stated above:

- G generates $\mathcal{I}_A$ since, for every i, it contains:

$$\sum_{k=1}^{t_i} \underline{x}^{a_k^{(i)}} - \underline{x}^{a_{k-1}^{(i)}} = \underline{x}^{a_{t_i}^{(i)}} - \underline{x}^{a_0^{(i)}} = \underline{x}^{h_i^-} - \underline{x}^{h_i^+}$$

  and since $\mathcal{H}$ is a Markov basis;

- Every element of G is a pure difference binomial by definition;

- Exactly as in the proof of Lemma 11, but this time with Laurent polynomials, every element $\underline{x}^{a_k^{(i)}} - \underline{x}^{a_{k-1}^{(i)}}$ of G can be re-written by using a telescoping sum. Fix i and k, then set $\alpha_m := a_{k-1}^{(i)} + \sum_{j=1}^{m} \epsilon_j^{(i,k)} b_j^{(i,k)}$. Then we have:

$$\underline{x}^{a_k^{(i)}} - \underline{x}^{a_{k-1}^{(i)}} = \underline{x}^{\alpha_{N_{i,k}}} - \underline{x}^{\alpha_0} = \sum_{m=1}^{N_{i,k}} \underline{x}^{\alpha_m} - \underline{x}^{\alpha_{m-1}} = \sum_{m=1}^{N_{i,k}} \underline{x}^{\alpha_{m-1} + \epsilon_m^{(i,k)} b_m^{(i,k)}} - \underline{x}^{\alpha_{m-1}} =$$

$$= \sum_{m=1}^{N_{i,k}} \underline{x}^{\alpha_{m-1} - (\epsilon_m^{(i,k)} b_m^{(i,k)})^-} \cdot \left(\underline{x}^{(\epsilon_m^{(i,k)} b_m^{(i,k)})^+} - \underline{x}^{(\epsilon_m^{(i,k)} b_m^{(i,k)})^-} \right) =$$

$$= \sum_{m=1}^{N_{i,k}} \underline{x}^{\alpha_{m-1} - (\epsilon_m^{(i,k)} b_m^{(i,k)})^-} \cdot \epsilon_m^{(i,k)} \left(\underline{x}^{(b_m^{(i,k)})^+} - \underline{x}^{(b_m^{(i,k)})^-} \right),$$

  which is indeed the form that we wanted any element in G to be written in.

So $||\mathcal{B}|| \leqslant \nu(\mathcal{B})$ and therefore the two quantities are equal. □

So this means that the quantity that we need to analyze is not just bounded by $||\mathcal{B}||$, it actually coincides with it.
This leads us to further looking into the pre-existing work of the paper [3] where this quantity was defined. In particular, what emerges from section 4 of such article is that we can obtain a bound starting from the following rational polyhedral cones in $\mathbb{N}^n$.

Definition 45. *Given $A, B, \mathcal{B}$ as in Notation 9 and $\epsilon \in \{\pm 1\}^n$ and $\delta \in \{\pm 1\}^r$, define the following rational polyhedral cone in $\mathbb{N}^n$:*

$$T_{\epsilon,\delta} := \{t \in \mathbb{N}^n : \delta_i \cdot \Big(\sum_{j=1}^{n} \epsilon_j b_{i,j} t_j \Big) \geqslant 0 \, \forall i = 1, \dots, r\} \subseteq \mathbb{N}^n.$$

More specifically, Holmes proves the following theorem.

Theorem 46 (Holmes, [3])**.** *In the situation described above, assume to have a bound N such that, for every $\epsilon \in \{\pm 1\}^n$ and $\delta \in \{\pm 1\}^r$, there exists a set of generators $V_{\epsilon,\delta}$ of $T_{\epsilon,\delta}$ with the property $||v||_1 \leqslant N$ for any $v \in V_{\epsilon,\delta}$. Then:*

$$||\mathcal{B}|| \leqslant N.$$

He then concludes that a possible set $V_{\epsilon,\delta}$ can be:

$$V_{\epsilon,\delta} = \{v_1, \dots, v_d\} \cup \big(\{ \sum_{i=1}^{d} \lambda_i v_i : \lambda_i \in [0,1) \, \forall i = 1, \dots, t\} \cap \mathbb{N}^n \big) \tag{2.1}$$

where $v_1, \dots, v_d$ are the smallest natural vectors in the extremal rays of $T_{\epsilon,\delta}$.

The extremal rays are all the possible solution sets in $\mathbb{N}^n$ that arise from sets of $n-1$ linearly independent equations taken from the following collection:

$$\begin{cases} \sum_{j=1}^{n} \epsilon_j b_{1,j} t_j = 0 \\ \vdots \\ \sum_{j=1}^{n} \epsilon_j b_{r,j} t_j = 0 \\ t_1 = 0 \\ \vdots \\ t_n = 0 \end{cases} \iff \left(\frac{\epsilon_1 b_1 | \cdots | \epsilon_n b_n}{I_{n \times n}} \right) \underline{t} =: \left(\frac{B_\epsilon}{I_{n \times n}} \right) \underline{t} = 0.$$

Here the matrix B_ϵ is defined according to the following notation.

Notation 47. *Given a sign vector $\epsilon \in \{\pm 1\}^b$ and a matrix $M \in \mathbb{Z}^{a \times b}$ whose columns are $c_1, \dots, c_b$, we denote by M_ϵ the matrix whose columns are $\epsilon_1 c_1, \dots, \epsilon_b c_b$.*

Holmes uses Siegel's Lemma to conclude that $v_1, \dots, v_d$ have norm at most $n \cdot (n\beta)^{n-1} = n^n \beta^{n-1}$, where $\beta := \max_{i,j} |b_{i,j}|$. This means that Theorem 46 can be applied with the quantity $N := n^{n+1} \beta^{n-1}$ and therefore yields the following result.

Theorem 48. *(Holmes, [3, Theorem 1.3]) Take A, B as in Notation 9. Set $\beta := \max_{i,j} |b_{i,j}|$. Then we have:*

$$||\mathcal{B}|| \leqslant n^{n+1} \beta^{n-1}.$$

The purpose of this section is that of exploring some alternatives to Siegel's lemma. Before we start, it may be interesting to notice that the case $n = 2$ is trivial, since the extremal rays are defined by only one linear equation. Such equation can be either of the form $t_i = 0$, in which case the smallest natural solution vector comes from the standard

basis of $\mathbb{Q}^2$ and its L_1-norm is 1, or of the form $\epsilon_1 b_{i,1} t_1 + \epsilon_2 b_{i,2} t_2 = 0$, in which case the smallest natural solution vector is the following, up to a sign:

$$\underline{t} = D^{-1} \begin{pmatrix} \epsilon_2 b_{i,2} \\ -\epsilon_1 b_{i,1} \end{pmatrix} \qquad \text{with } D := \text{G. C. D.}(b_{i,2}, b_{i,1}).$$

This means that, for the case $n = 2$, we can apply Theorem 46 and have that:

$$||\mathcal{B}|| \leqslant 2 \cdot \max\Big(\frac{|b_{1,1}| + |b_{1,2}|}{\text{G. C. D.}(b_{1,1}, b_{1,2})}, \frac{|b_{1,1}| + |b_{1,2}|}{\text{G. C. D.}(b_{2,1}, b_{2,2})}\Big).$$

For $n \geqslant 3$, the first option we may want to introduce is the following result from [6].

Theorem 49. *(Bombieri-Vaaler, [6, Theorem 1]) Given a full-rank matrix $M \in \mathbb{Z}^{m \times n}$ with $m < n$, there is a non-zero vector $w \in \text{Ker}_{\mathbb{Z}}(M)$ such that:*

$$\max_{i=1}^{n} |w_i| \leqslant (D^{-1} \sqrt{|\det(MM^t)|})^{\frac{1}{n-m}},$$

where D is the greatest common divisor of the determinants of all the $m \times m$ submatrices of M.

Starting from the bound suggested by this theorem, we could multiply by n to get a bound on the L_1-norm, then multiply by n again to bound the L_1-norm of all the vectors of $V_{\epsilon,\delta}$ and therefore have a bound on $||\mathcal{B}||$, as per Theorem 46. However, there does not seem to be any easy way to compute or approximate the maximum value that can appear in Theorem 49 unless we use very crude approximations, and computing all the determinants for all possible matrices seems computationally challenging.
This is why we need to come up with a different way to estimate such quantity. The lemma mentioned below will be only applicable to our specific case, as it focuses on the following:

1. We are talking about a one-dimensional null space of a matrix M;
2. We are only interested in the L_1-norm.

The idea is to compute the exact value of the norm of the smallest integer vector, a value which will be then approximated by a more computer-friendly one.
The lemma is the following.

Lemma 50. *Let $M \in \mathbb{Z}^{(n-1) \times n}$ be a full rank matrix and take any $v \in \text{Ker}_{\mathbb{Q}}(M)$. Then, for any vector $\epsilon \in \{\pm 1, 0\}^n$ such that:*

$$\epsilon_i = \text{sign}(v_i) \quad \forall i \in \{1, \dots, n\} \text{ s.t. } v_i \neq 0,$$

we know that the smallest L_1-norm in $\text{Ker}_{\mathbb{Z}}(M)$ is:

$$D^{-1} \big| \det\begin{pmatrix} M \\ \epsilon \end{pmatrix} \big|,$$

where D is the greatest common divisor of the determinants of all the $(n-1) \times (n-1)$ submatrices of M.

Proof. The idea behind this lemma is the following: write M as:

$$M = (v_1|\cdots|v_n) = \left(\begin{array}{c} m_1 \\ \hline \vdots \\ \hline m_{n-1} \end{array}\right).$$

If we extend M with any extra row m such that $\left(\frac{M}{m}\right)$ is invertible, then the last column of its inverse is in $\mathrm{Ker}_{\mathbb{Q}}(M)$. This vector would not be in $\mathbb{Z}^r$ a priori, which is the reason why we have to take the last column of the adjoint matrix $\mathrm{adj}\left(\frac{M}{m}\right)$ instead: consider the vector $w \in \mathbb{Z}^r$ whose components are defined as:

$$w_i := (-1)^{n+i}\det(M_i), \qquad \text{where } M_i = (v_1|\cdots|v_{i-1}|v_{i+1}|\cdots|v_n).$$

This vector is indeed in $\mathrm{Ker}_{\mathbb{Z}}(M)$:

$$\langle m_i, w\rangle = \sum_{j=1}^{n} m_{i,j}w_j = \sum_{j=1}^{n}(-1)^{j+n}m_{i,j}\det(M_j) = \det\left(\begin{array}{c} m_1 \\ \hline \vdots \\ \hline m_{n-1} \\ \hline m_i \end{array}\right) = 0,$$

where the penultimate equality holds thanks to the Laplace expansion formula.
Since $\mathrm{Ker}_{\mathbb{Q}}(M) = \langle v\rangle_{\mathbb{Q}}$, up to taking $-w$ instead of w (which does not change the value of the norm), we can assume $\mathrm{sign}(w_i) = \mathrm{sign}(v_i)$ for every i. So this means that:

$$||w||_1 = \sum_{i=1}^{n}|w_i| = \sum_{i=1}^{n}\mathrm{sign}(w_i)w_i = \sum_{i=1}^{n}\mathrm{sign}(v_i)w_i = \sum_{i=1}^{n}\epsilon_i w_i =$$

$$= \sum_{i=1}^{n}(-1)^{n+i}\epsilon_i\det(M_i) = \det\left(\frac{M}{\epsilon}\right),$$

where the last equality holds again thanks to the Laplace expansion formula for the determinants applied to the last row.
The fact that we are choosing w or $-w$ means that we have to introduce the absolute value, so we get $||w||_1 = |\det\left(\frac{M}{\epsilon}\right)|$.
The only extra remark that we need to make now is that w is scalable by exactly the G.C.D. of all its components. In other words, the smallest vector in $\mathrm{Ker}_{\mathbb{Z}}(M)$ (even though technically there are two, up to a sign) is $\frac{1}{\mathrm{G.C.D.}(\{w_i\}_{i=1}^n)}w$, which means that the smallest norm is:

$$\left|\left|\frac{1}{\mathrm{G.\,C.\,D.}(\{w_i\}_{i=1}^n)}w\right|\right|_1 = \frac{||w||_1}{\mathrm{G.\,C.\,D.}(\{w_i\}_{i=1}^n)} = \frac{|\det\left(\frac{M}{\epsilon}\right)|}{\mathrm{G.\,C.\,D.}(\{\det(M_i)\}_{i=1}^n)} = \frac{|\det\left(\frac{M}{\epsilon}\right)|}{D},$$

which is what we wanted to prove. □

In our case, the extremal rays are always inside $\mathbb{Q}^n_{\geqslant 0}$ and therefore we always have $\epsilon = \underline{1}$, even if the practical implementation from chapter 3 will use an extra step described later.
The value introduced by this lemma is the exact value of the smallest possible norm, so it must be no worse than the result that we would obtain from Theorem 49. However, one may want to know how different the two results are. As we will see in instances such as Corollary 52, the easiest way to approximate these quantities will be ignoring the denominator $D \geqslant 1$ and using Hadamard's inequality to estimate the determinant as the product of the norms of its rows. If we apply this to $\det\left(\frac{M}{\epsilon}\right)$, then we obtain $\sqrt{n}\prod_{i=1}^{n-1}||m_i||_2$. On the other hand, if we apply the same inequality to $\sqrt{\det(MM^t)}$, consider that the diagonal entries of MM^t alone are $||m_i||_2^2$, so Hadamard's inequality will work less effectively.
If we want to make a more general comparison instead, then we might have to use a slightly more general formulation of Hadamard's inequality for determinant in order to compare the two, such as the following lemma.

Lemma 51. *Given a matrix* $M = \left(\frac{m_1}{\vdots}\right) \in \mathbb{R}^{n\times n}$*, choose a non-empty set of indices* $J = \{j_1, \dots, j_k\} \subseteq \{1, \dots, n\}$*. Set* $M_J := \left(\frac{m_{j_1}}{\vdots}\right)$*. Then:*

$$|\det(M)| \leqslant \sqrt{|\det(M_J M_J^t)|} \cdot \prod_{i\notin J} ||m_i||_2.$$

Also, the equality holds if and only if the rows of M_{J^c} *form an orthogonal basis of* $\operatorname{Ker}\mathbb{R}(M_J)$*.*

This result seemed new at first but, during the polishing of the thesis, it emerged that this result was already available, since for example it can be found as Lemma 1 in [7]. For convenience of the reader, we also leave an elementary proof:

Proof. Without loss of generality, we may assume $J = \{1, \dots, k\}$ and therefore $J^c = \{k+1, \dots, n\}$. Also, the case $\det(M) = 0$ makes the lemma trivially true, so we can assume all the rows of M to be linearly independent.
First assume $m_{k+1}, \dots, m_n$ to be an orthogonal basis of $\operatorname{Ker}_{\mathbb{R}}(M_J)$. Then the equality holds:

$$|\det(M)| = \sqrt{\det(MM^t)} = \sqrt{\det\left(\left(\frac{M_J}{M_{J^c}}\right)\left(M_J^t \mid M_{J^c}^t\right)\right)} =$$

$$= \sqrt{\det(\left(\begin{array}{c|c} M_J M_J^t & M_J M_{J^c}^t \\ \hline M_{J^c} M_J^t & M_{J^c} M_{J^c}^t \end{array}\right))} =$$

$$= \sqrt{\det\left(\left(\begin{array}{c|cccccc} M_J M_J^t & \multicolumn{5}{c}{\underline{0}} \\ \hline & ||m_{k+1}||_2^2 & 0 & 0 & \cdots & 0 \\ & 0 & ||m_{k+2}||_2^2 & 0 & \cdots & 0 \\ \underline{0} & \vdots & \vdots & \ddots & \vdots & \vdots \\ & 0 & \cdots & \cdots & 0 & ||m_n||_2^2 \end{array}\right)\right)} =$$

$$= \sqrt{\det(M_J M_J^t) \cdot \prod_{i=k+1}^{n} ||m_i||_2^2} = \sqrt{\det(M_J M_J^t)} \cdot \prod_{i=k+1}^{n} ||m_i||_2.$$

To generalize this to any matrix M_{J^c}, write the vectors $m_{k+1}, \ldots, m_n$ as:

$$m_j = v_j + u_j \qquad \text{with } v_j \in \langle m_1, \ldots, m_k \rangle_{\mathbb{R}}^{\perp}, \quad u_j \in \langle m_1, \ldots, m_k \rangle_{\mathbb{R}}.$$

Also, by applying the Gram-Schmidt algorithm on the vectors v_j, we can re-write them as:

$$m_j = v'_j + u_j + \sum_{i=k+1}^{j-1} \alpha_i^{(j)} v'_i \qquad \text{with } \alpha_i^{(j)} \in \mathbb{R} \forall i, j$$

where $v'_{k+1}, \ldots, v'_n$ form an orthogonal basis of $\mathrm{Ker}_{\mathbb{R}}(M_J)$. By the multilinearity of the determinant, we know that:

$$\det(M) = \det\left(\begin{array}{c} M_J \\ \hline v'_{k+1} + u_{k+1} \\ \hline \vdots \\ \hline v'_n + u_n + \sum_{i=k+1}^{n-1} \alpha_i^{(n)} v'_i \end{array}\right) = \det\left(\begin{array}{c} M_J \\ \hline v'_{k+1} \\ \hline \vdots \\ \hline v'_n \end{array}\right).$$

Using the previous base case, this is equal to:

$$\sqrt{|\det(M_J M_J^t)|} \cdot \prod_{i=k+1}^{n} ||v'_i||_2.$$

Now notice that, since we decomposed m_j into the sum of orthogonal vectors, we have, for every j:

$$||v'_j||_2 = \sqrt{||m_i||_2^2 - ||u_j||_2^2 - \sum_{i=k+1}^{j-1} ||\alpha_i^{(j)} v'_i||_2^2} \leqslant ||m_i||_2$$

and the equality holds if and only if $u_j = 0$ for all j and $\alpha_j^{(i)} = 0$ for all i, j. In other words, we have:

$$\det(M) \leqslant \sqrt{|\det(M_J M_J^t)|} \cdot \prod_{i=k+1}^{n} ||m_i||_2$$

and the equality holds if and only if $m_i = v'_i$ for all $i \geqslant k+1$, so if and only if $m_{k+1}, \ldots, m_n$ are already an orthogonal basis of $\mathrm{Ker}_{\mathbb{R}}(M_J)$. □

This means that, if we want to compare the two results, the norm of the smallest vector $w \in \mathrm{Ker}_{\mathbb{Z}}(M)$ would be bounded by:

- If we use Lemma 50 (and Lemma 51):

$$||w||_1 = \frac{\det\begin{pmatrix} M \\ \underline{1} \end{pmatrix}}{D} \leqslant \frac{\sqrt{|\det(MM^t)|} \cdot ||\underline{1}||_2}{D} = \frac{\sqrt{|\det(MM^t)|} \cdot \sqrt{n}}{D}.$$

- If we use Theorem 49:

$$||w||_1 \leqslant n \cdot \max_{i=1}^{r} |w_i| \leqslant n \cdot \frac{\sqrt{|\det(MM^t)|}}{D}.$$

So, even if we do not immediately use Hadamard's inequality on the two matrices and we explicitly compute the determinants, we still have that Lemma 50 works better in this case, by a factor of at least $\sqrt{n}$.
Now that we analyzed this lemma, let us use it in order to get an easily-computable bound of the norm, as a corollary.

Corollary 52. *Consider $A, \mathcal{B}, B$ as in Notation 9, then take* [1] $k := \min(\mathrm{rk}(B), n-1)$ *and let $\rho_{j_1}, \ldots, \rho_{j_k}$ be the rows of the matrix B with the highest euclidean norms. Then:*

$$||\mathcal{B}|| \leqslant n \cdot \sqrt{k+1} \prod_{i=1}^{k} ||\rho_{j_i}||_2.$$

Proof. We have seen that the vectors v_i in Line (2.1) are generators of the integer null space of full-rank $(n-1) \times n$ submatrices M of $\begin{pmatrix} B_\epsilon \\ I_{n\times n} \end{pmatrix} = \begin{pmatrix} \epsilon_1 b_1 | \cdots | \epsilon_n b_n \\ I_{n\times n} \end{pmatrix}$. Since changing the signs of the columns does not alter the euclidean norms of the rows, we may assume M to be a submatrix of $\begin{pmatrix} B \\ I_{n\times n} \end{pmatrix}$ instead.
First assume $k = n-1$. By Lemma 50 we know that the minimal vector in an extremal ray of $T_{\epsilon,\delta}$ has norm:

$$||v_i||_1 = D^{-1} \det\begin{pmatrix} M \\ \underline{1} \end{pmatrix} \leqslant \det\begin{pmatrix} M \\ \underline{1} \end{pmatrix},$$

where M is a full rank $(n-1)\times n$ submatrix of $\begin{pmatrix} B \\ I_{n\times n} \end{pmatrix}$. Applying Hadamard's inequality by rows yields:

$$||v_i||_1 \leqslant \det\begin{pmatrix} M \\ \underline{1} \end{pmatrix} \leqslant \sqrt{n} \cdot \prod_{i=1}^{n-1} ||m_i||_2 \leqslant \sqrt{n} \cdot \prod_{i=1}^{n-1} ||\rho_{j_i}||_2,$$

where $m_1, \ldots, m_{n-1}$ are the rows of M, which have a smaller norm then the rows ρ_{j_i}, by definition.
This means that we can use this bound instead of Siegel's lemma and use Theorem 46 with $N := n^{3/2} \cdot \prod_{i=1}^{n-1} ||\rho_{j_i}||_2$ to get the result we wanted to prove.
For the case $k = \mathrm{rk}(B)$, simply notice that this implies that only k of the total $n-1$ rows

[1] By the Rank–nullity theorem we know that $\mathrm{rk}(B) = r - \mathrm{rk}(A)$.

of M can actually be taken from B. The remaining ones need to be taken from $I_{n\times n}$, so they have norm 1, so we only have the product of the k rows with the highest norm.
Also, since at least $n-1-\mathrm{rk}(B)$ rows of M come from the identity matrix, we can compute the Laplace expansion formula for determinants on such rows. Since these rows are vectors from the standard basis of $\mathbb{Q}^n$, we essentially get to delete such rows together with as many columns, and reduce the matrix $\left(\frac{M}{\underline{1}}\right)$ to a $(k+1)\times(k+1)$ matrix. Since we are just removing columns, the norms of the rows can only decrease, while the norm of $\underline{1}$ becomes $\sqrt{k+1}$ instead of $\sqrt{n}$. This implies that, if $k=\mathrm{rk}(B)$, then we can use Theorem 46 with $N := n\sqrt{k+1}\prod_{i=1}^{k}||\rho_{j_i}||_2$ to prove our claim.

□

While we will see in the next chapter how this bound can be easily computed by a machine, it may be really hard to compute this quantity without one. If we bound the norm of every row by $\sqrt{k+1}\beta$, where β is the highest entry of B in absolute value, then we immediately get the following result.

Corollary 53. *Consider $A, \mathcal{B}, B$ as in Notation 9, and let $\beta := \max_{\substack{i=1,\dots,r\\ j=1,\dots,n}} |b_{i,j}|$ and $k := \min(\mathrm{rk}(B), n-1)$. Then:*

$$||\mathcal{B}|| \leqslant n\cdot\sqrt{k+1}^{k+1}\cdot\beta^k \leqslant n^{\frac{n}{2}+1}\beta^{n-1}.$$

If we compare this with Theorem 48 then, for the non-trivial cases $n\geqslant 3$, we see that this bound is an improvement by a factor of at least $n^{\frac{n}{2}}>1$.
It may be relevant to notice that, unlike the actual value of $||\mathcal{B}||$, both these bounds actually worsen if one keeps expanding the set $\mathcal{B}$, as the rows in B get a higher norm and β potentially increases.
As already discussed in the introduction, the improvement on the bound of $||\mathcal{B}||$ might be relevant in several ways, such as the following improvement of Corollary 42.

Theorem 54. *Let $A\in\mathbb{Z}^{s\times r}$ be a matrix and take a matrix B whose set of columns $\mathcal{B}=\{b_1,\dots,b_n\}$ generates $\mathrm{Ker}_{\mathbb{Z}}(A)$ as a group. Set $\beta := \max_{\substack{i=1,\dots,r\\ j=1,\dots,n}} |b_{i,j}|$ and $k := \min(\mathrm{rk}(B), n-1)$. Then the set:*

$$\mathcal{B}' := \{v\in\mathrm{Ker}_{\mathbb{Z}}(A) : |v_i| \leqslant n\sqrt{k+1}^{k+1}\beta^{k+1}\ \forall i=1,\dots,r\}$$

is a Markov basis.

2.2 Computer algorithms to improve the bound

In this section we will explore how some algorithms can be defined in order to try and compute the bound suggested by Lemma 50 and Corollary 52. In particular, we will see an algorithm that computes Hadamard's bound from Corollary 52 in a finer way and also an algorithm that tries to explicitly bound the highest absolute determinant through means of local maximums, in a sense that we will clarify in Section 2.2.2.

2.2.1 Introductory lemmas and considerations

Before we start introducing some algorithms, we need a couple of preliminary remarks. The first remark will help us deal with the fact that the extremal rays are defined by sub-matrices of $\left(\frac{B_\epsilon}{I_{n\times n}}\right)$, with ϵ being any possible sign vector.

Naively, one would check any possible full-rank $(n-1)\times n$ submatrix M of $\left(\frac{B_\epsilon}{I_{n\times n}}\right)$, for any ϵ, and then only compute the determinant $\det\left(\frac{M}{\underline{1}}\right)$ for the matrices M whose null space is generated by a vector in $\mathbb{N}^r$. However, there is a much more efficient way suggested by some elementary remarks.

Remark 55. *Consider a full-rank matrix $M \in \mathbb{Z}^{(n-1)\times n}$. Then there exists a sign vector $\epsilon \in \{\pm 1\}^n$ such that $\mathrm{Ker}(M_\epsilon)$ is generated by a vector in $\mathbb{N}^r$*

Proof. Take a non-zero vector $v \in \mathrm{Ker}_{\mathbb{Z}}(M)$ and consider the vector $\epsilon \in \{\pm 1\}^n$ defined as $\epsilon_i := \mathrm{sign}(v_i)$ (with $\mathrm{sign}(0) = 1$). Then we have:

$$0 = Mv = M_\epsilon I_\epsilon v \qquad \text{for } v \in \mathbb{Z}^r,$$

where I is the identity matrix $I_{n\times n}$. This means that we can write:

$$M_\epsilon u = 0 \qquad \text{with } u := I_\epsilon v = (\mathrm{sign}(v_i)v_i)_{i=1}^n = (|v_i|)_{i=1}^n \in \mathbb{N}^n,$$

which concludes the proof. □

As the sign vector described above will recur often, let us properly introduce it in our notation.

Definition 56. *Let $M \in \mathbb{Z}^{(n-1)\times n}$ be a full-rank matrix and let $v \in \mathrm{Ker}_{\mathbb{R}}(M)$ be such that its first non-zero component is positive. Then we define the vector $\varepsilon(M) \in \{\pm 1\}^n$ as:*

$$\varepsilon(M)_i := \mathrm{sign}(v_i) \;\; \forall i = 1, \ldots, n,$$

where $\mathrm{sign}(0) := 1$.

We may now proceed to our second remark.

Remark 57. *Consider the rational polyhedral cones $T_{\epsilon,\delta}$ as defined in Definition 45. Then the set:*

$$\mathcal{P} := \bigcup_{\substack{\epsilon\in\{\pm1\}^n \\ \delta\in\{\pm1\}^r}} \{\ell : \ell \text{ is an extremal ray of } T_{\epsilon,\delta}\}$$

actually coincides with:

$$\mathcal{Q} := \{\mathrm{Ker}_{\mathbb{Z}}(M_{\varepsilon(M)}) \cap \mathbb{N}^n : M \text{ is an } (n-1)\times n \text{ full-rank submatrix of } \left(\frac{B}{I_{n\times n}}\right)\}.$$

Proof. As we already noticed on page 26, and thanks to Remark 55, we have the trivial inclusion $\mathcal{Q} \subseteq \mathcal{P}$.
For the converse, let M be a full-rank sub-matrix of $\begin{pmatrix} B_\epsilon \\ \hline I \end{pmatrix} := \begin{pmatrix} B_\epsilon \\ \hline I_{n\times n} \end{pmatrix}$ whose null space is spanned by a natural vector w. Then $M' := M_\epsilon$ is a full-rank sub-matrix of:

$$\begin{pmatrix} B_\epsilon \\ \hline I \end{pmatrix}_\epsilon = \begin{pmatrix} (B_\epsilon)_\epsilon \\ \hline I_\epsilon \end{pmatrix} = \begin{pmatrix} B \\ \hline I_\epsilon \end{pmatrix}$$

and, as seen in the proof of Remark 55, its null space is spanned by the vector $(\epsilon_i w_i)_i$. Since we were assuming $w \in \mathbb{N}^r$, this means that $\varepsilon(M') = \pm\epsilon$, which allows us to conclude that:

$$\mathrm{Ker}(M) = \mathrm{Ker}((M_\epsilon)_\epsilon) = \mathrm{Ker}(M'_\epsilon) = \mathrm{Ker}(M'_{\pm\varepsilon(M')}) = \mathrm{Ker}(M'_{\varepsilon(M')}).$$

Notice that we may easily assume M' to be a submatrix of $\begin{pmatrix} B \\ \hline I \end{pmatrix}$ instead of $\begin{pmatrix} B \\ \hline I_\epsilon \end{pmatrix}$: we are only computing the half-line in $\mathbb{N}^n$ which is orthogonal to the rows of M', and taking I instead of I_ϵ simply means changing the sign of all the rows of M' taken from the identity matrix, which does not affect orthogonality. □

In other words, we just proved that the collection of all the extremal rays of all the rational polyhedral cones $T_{\epsilon,\delta}$ that come from any possible $\epsilon \in \{\pm 1\}^n$ and $\delta \in \{\pm 1\}^r$ is equal to the collection of all the half-lines $\mathrm{Ker}(M_{\varepsilon(M)}) \cap \mathbb{N}^n$ that come from any full-rank $(n-1) \times n$ sub-matrix M of $\begin{pmatrix} B \\ \hline I_{n\times n} \end{pmatrix}$. The latter is a much smaller collection to check. These remarks together mean that one can try the following strategy:

- Take any $(n-1) \times n$ submatrix M from $\begin{pmatrix} B \\ \hline I_{n\times n} \end{pmatrix}$;
- Compute $\mathrm{Ker}_{\mathbb{Q}}(M)$;
- If it has dimension strictly larger than 1, then M is not full-rank, so move forward to the next submatrix (i.e. go back to step 1);
- If it has dimension 1, take any vector $w \in \mathrm{Ker}_{\mathbb{Q}}(M)$ and compute $\epsilon := (\mathrm{sign}(w_i))_i$;
- Compute $|\det \begin{pmatrix} M \\ \hline \epsilon \end{pmatrix}|$, then move forward to the next submatrix (i.e. go back to step 1);
- Take the maximum of all these determinants.

Notice that the formulation of Lemma 50 also covers the case when a computer approximates zero-components of w into "slightly" positive or negative numbers, when computing $\mathrm{sign}(w_i)$.
This strategy, however, has two flaws: one is that we still check $\binom{n+r}{n-1}$ different combinations, a number which increases very quickly as n grows. The other is that, once we compute the null space of M, we might as well just take a generator of $\mathrm{Ker}_{\mathbb{Z}}(M)$ and compute its L_1-norm explicitly.

This pushes us to look for a different strategy, which is that of trying to determine in advance which combinations of rows of $\left(\frac{B}{I_{n\times n}}\right)$ are going to give a higher absolute determinant.

Lemma 58. *Take a full-rank matrix $M \in \mathbb{R}^{a\times n}$ with $a < n$ and a matrix $W \in \mathbb{R}^{(n-a)\times n}$ whose rows are a basis for $\mathrm{Ker}_{\mathbb{R}}(M)$. Then $|\det\left(\frac{M}{V}\right)|$ and $|\det(VW^t)|$ are related by multiplication with a positive constant which does not depend on the matrix $V \in \mathbb{R}^{(n-a)\times n}$.*

Proof. Since W is full-rank, we can take a left inverse $T \in \mathbb{R}^{(n-a)\times n}$ of W^t. Now notice that:

$$\left(\frac{M}{T}\right)\cdot\left(\frac{M}{W}\right)^t = \left(\begin{array}{c|c} MM^t & MW^t \\ \hline TM^t & TW^t \end{array}\right) = \left(\begin{array}{c|c} MM^t & MW^t \\ \hline TM^t & I_{(n-a)\times(n-a)} \end{array}\right).$$

Since M, and therefore MM^t, is full-rank, we know that the block matrix on the right-hand side is full-rank, therefore so is $\left(\frac{M}{T}\right)$. Let us denote by E its inverse, which we write as $(D|W^t)$, where $D \in \mathbb{R}^{n\times a}$. Its last columns are given by W^t exactly because the rows of W are orthogonal to M and T is a left inverse of W^t. Using again the block matrix product, we have that, given any matrix $V \in \mathbb{R}^{(n-a)\times n}$:

$$\det\left(\frac{M}{V}\right)\det(E) = \det\left(\left(\frac{M}{V}\right)\cdot(D|W^t)\right) = \det\left(\begin{array}{c|c} MD & MW^t \\ \hline VD & VW^t \end{array}\right) =$$

$$= \det\left(\begin{array}{c|c} I_{a\times a} & \underline{0} \\ \hline VD & VW^t \end{array}\right).$$

This is a block lower triangular matrix, therefore its determinant is:

$$\det\left(\frac{M}{V}\right)\det(E) = \det(I_{a\times a})\cdot\det(VW^t) = \det(VW^t),$$

so this means that:

$$\left|\det\left(\frac{M}{V}\right)\right| = \left|\frac{\det(VW^t)}{\det(E)}\right| = \frac{1}{|\det(E)|}\cdot|\det(VW^t)|.$$

Since E does not depend on V, this concludes the proof. □

Essentially, what we proved is that if we have a matrix U (which acts as "pool of available rows") and we want to complete the matrix M to a square matrix by adding rows from U, then we can take the set:

$$\mathcal{M} := \left\{\left(\frac{M}{V}\right) \in \mathbb{R}^{n\times n} : V \text{ is a } (n-a)\times n\text{-submatrix of } U\right\}$$

and Lemma 58 says that the matrices in $\mathcal{M}$ with the highest absolute determinant are the ones such that the quantity $|\det(VW^t)|$ is maximal.
For instance, it may be interesting to notice that, once we fix M, $\varepsilon(M)$ is indeed a sign

vector ϵ (taken from the pool of rows $\{\pm 1\}^n$) that maximizes $|\det\begin{pmatrix} M \\ \hline \epsilon \end{pmatrix}|$. Indeed, given a vector $w \in \mathrm{Ker}_{\mathbb{R}}(M)$, we have:

$$|\langle \epsilon, w \rangle| = \Big| \sum_{i=1}^{n} \epsilon_i w_i \Big| \leqslant \sum_{i=1}^{n} |\epsilon_i w_i| = \sum_{i=1}^{n} |w_i| = \sum_{i=1}^{n} \mathrm{sign}(w_i) w_i = |\langle \varepsilon(M), w \rangle|,$$

therefore $\epsilon(M)$ maximizes $|\langle \epsilon, w \rangle|$ and hence, using Lemma 58, it also maximizes $|\det\begin{pmatrix} M \\ \hline \epsilon \end{pmatrix}|$.

2.2.2 The algorithms

Thanks to all the considerations made in Chapter 1 and in the previous section, we may now design two algorithms to compute the bound in a more precise way than Corollary 52. Let us therefore introduce them, while keeping in mind that these algorithms will assume $n > 2$ since the case $n = 2$ is trivial, as discussed on page 26.

Hadamard's inequality computation

This algorithm is definitely the simplest one that we are going to see. All it does is explicitly computing the $n-1$ highest norms of rows in $U := \begin{pmatrix} B \\ \hline I_{n\times n} \end{pmatrix}$, then using Corollary 52 to output a bound on the norm. There are only two remarks to improve this algorithm:

- Order the rows of $U = \begin{pmatrix} B \\ \hline I_{n\times n} \end{pmatrix}$ in a decreasing order with respect to their euclidean norm, and call such ordered rows $\rho_1, \ldots, \rho_{r+n}$. If we take the rows $\rho_1, \ldots, \rho_{n-1}$, which have the highest norm, then, as seen in Corollary 52, one may bound $|\det\begin{pmatrix} M \\ \hline \epsilon \end{pmatrix}|$ with $\sqrt{n}\prod_{i=1}^{n-1} ||\rho_i||_2$. However, there is room for improvement: say that $\mathrm{rk}\begin{pmatrix} \rho_1 \\ \hline \vdots \\ \hline \rho_{n-1} \end{pmatrix} = k < n-1$. Then, since M is always full-rank, only at most k vectors between $\rho_1, \ldots, \rho_{n-1}$ can be rows of M. So it actually makes sense to only use $||\rho_1||_2, \ldots, ||\rho_k||_2, ||\rho_n||_2, \ldots, ||\rho_{2n-2-k}||_2$ in the product above, instead of $||\rho_1||_2, \ldots, ||\rho_{n-1}||_2$.

  Analogously, if $k_1 := \mathrm{rk}\begin{pmatrix} \rho_1 \\ \hline \vdots \\ \hline \rho_k \end{pmatrix} < k$, then out of the k rows that we take from $\rho_1, \ldots, \rho_{n-1}$ there are only at most k_1 rows that can come from $\rho_1, \ldots, \rho_k$, which means that instead of using $||\rho_1||_2, \ldots, ||\rho_k||_2$ one can actually use:

  $$||\rho_1||_2, \ldots, ||\rho_{k_1}||_2, ||\rho_{k+1}||, \ldots, ||\rho_{2k-2-k_1}||_2$$

  This technique can proceed iteratively.

- If $\mathrm{rk}(B) < n-1$ then, as already discussed in the proof of Corollary 52, there are not enough linearly independent rows of B to cover all the rows of the matrix $\begin{pmatrix} M \\ \hline \epsilon \end{pmatrix}$, therefore at least $n-1-\mathrm{rk}(B)$ rows of this matrix come from the identity matrix. Using the Laplace expansion on these $n-1-\mathrm{rk}(B)$ rows, one by one, leads to the deletion of $n-1-\mathrm{rk}(B)$ columns from the matrix $\begin{pmatrix} M \\ \hline \epsilon \end{pmatrix}$, together with these rows from the standard basis of $\mathbb{Q}^n$. After this computation, we are left with the computation of the determinant of a $(\mathrm{rk}(B)+1) \times (\mathrm{rk}(B)+1)$ matrix which is a submatrix of the initial matrix $\begin{pmatrix} M \\ \hline \epsilon \end{pmatrix}$. We cannot know a priori which columns or rows determine such submatrix. However, if we want to use Hadamard's inequality, then we only need to compute the norms of the rows of such submatrix. If we want to make sure to have a bound then, for every row, we can compute the euclidean norm that we obtain with the highest $\mathrm{rk}(B)+1$ coefficients of such row (highest in absolute value).
  Also, the fact that we end up with a $(\mathrm{rk}(B)+1) \times (\mathrm{rk}(B)+1)$ implies that the last row has norm $\sqrt{\mathrm{rk}(B)+1}$ instead of $||\epsilon||_2 = \sqrt{n}$.

The second point is easy to implement. For the first one, we deploy the following technique, starting from the matrix U which is our pool of rows.

Algorithm 1. *(Find the independent rows with the highest euclidean norm)*

1. *Arrange the rows of the matrix U in decreasing order, according to their euclidean norm (or the euclidean norm of the subvector with the $\mathrm{rk}(B)+1$ highest entries, if $\mathrm{rk}(B) < n-1$).*

2. *We start by looking for $n-1$ rows, or at least "as many linearly independent rows as possible", which means that we look for $k := \min(n-1, \mathrm{rk}(B))$ rows.*

3. *If U has exactly k rows, then it is full-rank and you can take all of them and be done. Otherwise, go on.*

4. *Split U in consecutive $k \times n$ submatrices (i.e. without skipping rows, so U_i having rows $\rho_{i \cdot k}, \ldots, \rho_{(i+1) \cdot k-1}$). Of course the last submatrix can have less than k rows, so as to be able to split U even when its number of rows is not a multiple of k.*

5. *Starting from $i = 1$, check $\Delta_i := \mathrm{rk} \begin{pmatrix} U_1 \\ \hline \vdots \\ \hline U_i \end{pmatrix} - \mathrm{rk} \begin{pmatrix} U_1 \\ \hline \vdots \\ \hline U_{i-1} \end{pmatrix}$. This is the number of rows we can add from U_i. The smallest N such that we have the inequality:*

$$\mathrm{rk} \begin{pmatrix} U_1 \\ \hline \vdots \\ \hline U_N \end{pmatrix} = \sum_{i=1}^{N} \Delta_i \geqslant k$$

   is the number of sub-matrices that we need to take our rows from.

6. *To take Δ_i rows from U_i, we repeat the process: go back to step 2, but with U_i instead of U and with Δ_i instead of k. Of course this does not apply if $\Delta_i = 0$, in which case we skip the whole block U_i, as it has no new possible vectors to add.*

Notice that, technically, we are being imprecise in step 5: potentially, we can take more than Δ_i rows from U_i. We can take up to $\mathrm{rk}(U_i) \geqslant \Delta_i$ rows. However, doing so would mean taking such vectors *instead of* other vectors from some previous U_j with $j < i$. Since these previous blocks have rows with a higher norm, the bound holds anyway. In other words, this model does not necessarily look for the combination of vectors with the highest determinants, but rather the one that makes the Hadamard bound the highest. This is made possible by the ordering of the rows from step 1, combined with the structure of the algorithm which prioritizes the rows that come first.
A way to implement this algorithm in Sage can be found in the appendix.

Local maximums computation

This algorithm deploys Lemma 58 in a more extensive way.
To explain this method, let us first take a step back and ignore the fact that we need the last vector of the matrix to be a sign vector ϵ. Let us assume to have the general problem of having to maximize the absolute determinant of a $n \times n$ matrix whose n rows are taken from the same pool of rows U.
The idea is to start looking at "local maximums" by using the following algorithm.

Algorithm 2. *(Find a local maximum-General case)*

1. *Given a matrix U of available rows, take a full-rank submatrix $M \in \mathbb{Z}^{n \times n}$ with rows $m_1, \ldots, m_n$.*
2. *Take any vector $w \in \mathrm{Ker} \begin{pmatrix} m_2 \\ \hline \vdots \\ \hline m_n \end{pmatrix}$ and look for a row r of U that maximizes $|\langle w, r \rangle|$.*
3. *Replace M with $\begin{pmatrix} m_2 \\ \hline \vdots \\ \hline m_n \\ \hline r \end{pmatrix}$.*
4. *Start over from point 2 and keep on going until n consecutive steps do not change the quantity $|\langle w, r \rangle|$ in step 2 (i.e. no row can be substituted to get a higher absolute determinant, by Lemma 58).*

Notice that every time we substitute a row the determinant does not decrease. Combining this with the fact that we start from an invertible matrix in step 1, this means that the output matrix is always invertible. Also, now it is clear that the meaning of the expression "local maximum" is referred to the fact that there are no better matrices up to only changing one row.
However, this does not make such matrix a global maximum. If we write $M_{\max}$ for the

$n \times n$ submatrix of U which *does* have the maximum absolute determinant, then what we can tell is the following:

$$|\det(M_{\max})| = |\det(M_{\max}M^{-1}M)| = |\det(M_{\max}M^{-1})| \cdot |\det(M)|.$$

Since we know the determinant of the local maximum M, it means that we only need to determine the maximum possible determinant of the $n \times n$ submatrix $M_{\max}M^{-1}$ of the matrix UM^{-1}. So now we have two possibilities:

- Iterating the process, by looking for the maximum absolute determinant obtainable from UM^{-1}
- Using Hadamard's inequality to yield an estimation of the determinant

In a way, we will use both ideas: after finding a local maximum M, we can use the previous Hadamard algorithm to estimate the biggest absolute determinant that can come from UM^{-1} and deduce an immediate bound. As we have seen, such algorithm actually outputs the set of independent rows with the highest norms. These would potentially be the rows of the global maximum $M_{\max}$, so it might make sense to try and find a new local maximum, starting from these rows. At this point, to check the validity of this new local maximum, one would repeat the process we just depicted. One can keep on going for as long as the bound decreases. The resulting algorithm is the following.

Algorithm 3. *(Bound the global maximum-General case)*

1. *Use Algorithm 1 to detect the set of linearly independent rows $\rho_{j_1}, \ldots, \rho_{j_n}$ of U which have the highest norm.*
2. *From $\rho_{j_1}, \ldots, \rho_{j_n}$, find a local maximum M through means of Algorithm 2.*
3. *The determinant $\det(M_{\max})$ whose absolute value is the highest possible is equal to $\det(M)$ multiplied by the maximum absolute determinant coming from an $n \times n$ submatrix of $V := UM^{-1}$. Using Hadamard's estimation on V, we get an immediate bound, which consists of $\det(M)$ multiplied by the product of the euclidean norm of n rows of V.*
4. *Denote by $r'_{k_1}, \ldots, r'_{k_n}$ the rows of V taken with the same technique as point 1, which are the same used in point 3 to compute Hadamard's bound on V. Using those same indices, go back to point 2 with $\rho_{k_1}, \ldots, \rho_{k_n}$ instead of $\rho_{j_1}, \ldots, \rho_{j_n}$, still being rows of U (and not of V).*
5. *Stop when the bound does not improve after some consecutive steps of this type.*

When we introduced the idea of recursion and moving from the maximums of U to the maximums of $V = UM^{-1}$, we did not mention what we did in step 4, which is staying in U instead of actually starting over from V. This is because it does not really make any difference. Indeed, once we have the new local maximum:

$$D := \begin{pmatrix} m'_1M^{-1} \\ \hline \vdots \\ \hline m'_nM^{-1} \end{pmatrix} \qquad \text{submatrix of } V = UM^{-1},$$

then the algorithm would proceed by checking the highest determinant from:

$$V \cdot D^{-1} = V \cdot (\begin{pmatrix} m'_1 \\ \hline \vdots \\ \hline m'_n \end{pmatrix} M^{-1})^{-1} = UM^{-1}M \begin{pmatrix} m'_1 \\ \hline \vdots \\ \hline m'_n \end{pmatrix}^{-1} = U \begin{pmatrix} m'_1 \\ \hline \vdots \\ \hline m'_n \end{pmatrix}^{-1}.$$

So, as we said above, we might as well just stay in U and evaluate $\begin{pmatrix} m'_1 \\ \hline \vdots \\ \hline m'_n \end{pmatrix}$ as local maximum.

If, instead of the classic Hadamard's inequality, one wanted to use Lemma 51, this may be done in the following way: group all the rows of UM^{-1} whose norm is strictly larger than one into one submatrix L of UM^{-1}. Then, given any local maximum D submatrix of UM^{-1}, up to a permutation of its rows, write it as $D = \begin{pmatrix} D_J \\ \hline D_{J^c} \end{pmatrix}$ with D_J being the submatrix of all the rows taken from L (i.e. with norm larger than 1). Then we may use Lemma 51:

$$|\det(D)| \leqslant \sqrt{|\det(D_J D_J^t)|} \prod_{j \in J^c} ||d_j||_2 \leqslant \sqrt{|\det(D_J D_J^t)|}.$$

Since D_J is a sub-matrix of L, we have that $D_J D_J^t$ is a principal submatrix of LL^t. This means that an alternative to the classic Hadamard computation might be taking the maximum of all the possible $|\det(P)|$ over all the principal submatrices P of LL^t. This algorithm is based on the idea that many of the vectors in UM^{-1} seem to have the tendency to have a small norm, while typically only at most n rows have a higher norm. This makes such computation much more doable than the general case with $r + n$ rows. The algorithm we obtain is the following.

Algorithm 4. *(Bound the global maximum-General case-v2)*

1. *Use Algorithm 1 to detect the set of linearly independent rows $\rho_{j_1}, \ldots, \rho_{j_n}$ of U which have the highest norm.*
2. *From $\rho_{j_1}, \ldots, \rho_{j_n}$, find a local maximum M through means of Algorithm 2.*
3. *The determinant $\det(M_{\max})$ whose absolute value is the highest possible is equal to $\det(M)$ multiplied by the maximum absolute determinant coming from an $n \times n$ submatrix of $V := UM^{-1}$. Using Hadamard's estimation on V, we get an immediate bound, which consists of $\det(M)$ multiplied by the product of the euclidean norm of n rows of V.*
4. *Another bound for the quantity of point 3 is given by the greatest absolute minor of LL^t (and L is as defined on page 40) that can be obtained with at most $n-1$ rows and columns. Take the minimum between the two bounds.*
5. *Denote by $r'_{k_1}, \ldots, r'_{k_n}$ the rows of V taken with the same technique as point 1, which are the same used in point 3 to compute Hadamard's bound on V. Using those same indices, go back to point 2 with $\rho_{k_1}, \ldots, \rho_{k_n}$ instead of $\rho_{j_1}, \ldots, \rho_{j_n}$, still being rows of U (and not of V).*

6. *Stop when the bound does not improve after some consecutive steps of this type.*

So now how do we do all of this in our setting, which is that of only choosing $n-1$ vectors from U plus one sign vector? One adaptation of Algorithm 2 might be alternating two kinds of steps: one adjusts the sign vector according to the other $n-1$ rows, while the other changes one of the $n-1$ rows to increase the determinant.
More specifically, this leads us to the following algorithm, the code of which can be found in the appendix.

Algorithm 5. *(Find a local maximum-Specific case)*

1. *Given a matrix U of available rows, take a full-rank submatrix $M \in \mathbb{Z}^{(n-1)\times n}$ with rows $m_1, \ldots, m_{n-1}$ and consider the vector $\epsilon := \varepsilon(M)$.*
2. *Take any vector $w \in \operatorname{Ker} \left(\begin{array}{c} m_2 \\ \hline \vdots \\ \hline m_{n-1} \\ \hline \epsilon \end{array}\right)$ and look for the row r of U that maximizes $|\langle w, r\rangle|$.*
3. *Replace M with $\left(\begin{array}{c} m_2 \\ \hline \vdots \\ \hline m_{n-1} \\ \hline r \end{array}\right)$ and ϵ with $\varepsilon\left(\left(\begin{array}{c} m_2 \\ \hline \vdots \\ \hline m_{n-1} \\ \hline r \end{array}\right)\right)$.*
4. *Start over from point 2 and keep on going until $n-1$ consecutive steps do not change the quantity $|\langle w, r\rangle|$ in step 2 (i.e. no row can be substituted to get a higher absolute determinant, by Lemma 58).*

This allows us to implement an adaptation of Algorithm 3 and Algorithm 4:

Algorithm 6. *(Bound the global maximum-Specific case)*

1. *Use Algorithm 1 to detect the set of linearly independent rows $\rho_{j_1}, \ldots, \rho_{j_{n-1}}$ of U which have the highest norm.*
2. *From $\rho_{j_1}, \ldots, \rho_{j_{n-1}}$, find a local maximum $H := \left(\frac{M}{\epsilon}\right) := \left(\begin{array}{c} m_1 \\ \hline \vdots \\ \hline m_{n-1} \\ \hline \epsilon \end{array}\right)$ through means of Algorithm 5.*
3. *The determinant $\det\left(\frac{M_{\max}}{\epsilon_{\max}}\right)$ whose absolute value is the highest possible is equal to $\det H \cdot \det\left(\frac{M'}{\varepsilon(M')H^{-1}}\right)$, where M' is a submatrix of $V := UH^{-1}$ which maximizes $|\det\left(\frac{M'}{\varepsilon(M')}\right)|$. Using Hadamard's estimation on V, we get an immediate bound. It is the product of the euclidean norm of $n-1$ rows of V and the norm $||\varepsilon(M')H^{-1}||_2 \leqslant ||\varepsilon(M')||_2 \cdot ||H^{-1}||_2 = \sqrt{n}||H^{-1}||_2$.*

4. *Denote by* $r'_{k_1}, \ldots, r'_{k_{n-1}}$ *the rows of* V *taken with the same technique as point 1, which are the same used in point 3 to compute Hadamard's bound on* V*. Using those same indices, go back to point 2 with* $\rho_{k_1}, \ldots, \rho_{k_{n-1}}$ *instead of* $\rho_{j_1}, \ldots, \rho_{j_{n-1}}$*, still being rows of* U *(and not of* V*).*

5. *Stop when the bound does not improve after some consecutive steps of this type.*

Algorithm 7. *(Bound the global maximum-Specific case-v2)*

1. *Use Algorithm 1 to detect the set of linearly independent rows* $\rho_{j_1}, \ldots, \rho_{j_{n-1}}$ *of* U *which have the highest norm.*

2. *From* $\rho_{j_1}, \ldots, \rho_{j_{n-1}}$*, find a local maximum* $H := \begin{pmatrix} M \\ \hline \epsilon \end{pmatrix} := \begin{pmatrix} m_1 \\ \hline \vdots \\ \hline m_{n-1} \\ \hline \epsilon \end{pmatrix}$ *through means of Algorithm 5.*

3. *The determinant* $\det \begin{pmatrix} M_{\max} \\ \hline \epsilon_{\max} \end{pmatrix}$ *whose absolute value is the highest possible is equal to* $\det H \cdot \det \begin{pmatrix} M' \\ \hline \varepsilon(M')H^{-1} \end{pmatrix}$*, where* M' *is a submatrix of* $V := UH^{-1}$ *which maximizes* $|\det \begin{pmatrix} M' \\ \hline \varepsilon(M') \end{pmatrix}|$*. Using Hadamard's estimation on* V*, we get an immediate bound. It is the product of the euclidean norm of* $n-1$ *rows of* V *and the norm* $||\varepsilon(M')H^{-1}||_2 \leqslant ||\varepsilon(M')||_2 \cdot ||H^{-1}||_2 = \sqrt{n}||H^{-1}||_2$*.*

4. *Another bound for the quantity of point 3 is given by* $\sqrt{n}||H^{-1}||_2 \cdot \alpha$*, where* α *is the greatest absolute minor of* LL^t *(and* L *is as defined on page 40) that can be obtained with at most* $n-1$ *rows and columns. Take the minimum between the two bounds.*

5. *Denote by* $r'_{k_1}, \ldots, r'_{k_{n-1}}$ *the rows of* V *taken with the same technique as point 1, which are the same used in point 3 to compute Hadamard's bound on* V*. Using those same indices, go back to point 2 with* $\rho_{k_1}, \ldots, \rho_{k_{n-1}}$ *instead of* $\rho_{j_1}, \ldots, \rho_{j_{n-1}}$*, still being rows of* U *(and not of* V*).*

6. *Stop when the bound does not improve after some consecutive steps of this type.*

2.2.3 Comments on the algorithms

To test the algorithm, we can try the code shown in the section "Numerical Examples" of the Appendix.
After running this code several times, it seems that the Hadamard's computation algorithm produces a worse bound overall, but in a very short amount of time. On the other hand, the Local maximums algorithms (Algorithm 6 and Algorithm 7) seem to yield a closer bound in most cases, for example with many sparse matrices, even though they are also more time-consuming (especially Algorithm 7).
While, by construction, Algorithm 7 produces a bound which is never worse (but often

better) than the one from Algorithm 6, it is also true that the former can take more time to run and can lead to memory errors, for big matrices. Also, the improvement of the bound does not seem to be very relevant, when compared to the order of magnitude of the bound itself. This is why Algorithm 6 is probably the best choice, overall. Alternatively, it might be a good idea to try executing Algorithm 7 and, in case it reveals to be too time-consuming, to choose Algorithm 6 instead.
Another option might be adjusting Algorithm 7 by skipping step 4 whenever the matrix L has too many rows.
Another algorithm, which we did not introduce due to its slowness, used Theorem 54 to compute the following Markov basis:

$$\mathcal{B}' := \{v \in \mathrm{Ker}_{\mathbb{Z}}(A) : ||v||_\infty \leqslant n\sqrt{k+1}^{k+1}\beta^{k+1}\},$$

where $\beta = \max_{b\in\mathcal{B}} ||b||_\infty$ and $k := \min(\mathrm{rk}(B), n-1)$. This set corresponds to the set of integer points of a polyhedron (defined by the equations $Av = 0$ and the inequalities $-n\sqrt{k+1}^{k+1}\beta^{k+1} \leqslant v_i \leqslant n\sqrt{k+1}^{k+1}\beta^{k+1}$), which SageMath has a specific implementation for. However, the number of points increases very quickly with n, making this algorithm impractical.
That being said, this method might be a nice starting point to try and define new algorithms that could find a Markov basis, together with the work in Theorem 37 and Section 1.3.

Appendix

Codes

Hadamard's inequality computation (Section 2.2.2)

```
import numpy as np
import heapq


def Indep_rows(U,U_prev,maximum=None):
    #U is the matrix from which we take the rows
    #if it is an iteration, U_prev collects all the rows from the previous blocks
    n=len(U[0])
    l=len(U)
    l_prev=len(U_prev)
    if maximum==None:
        maximum=l

    #count how many vectors from the block U you can add to the ones taken from U_prev
    rk_new=Matrix(QQ,U+U_prev).rank()
    rk_old=Matrix(QQ,U_prev).rank()
    Delta=rk_new-rk_old
    top=min(Delta,maximum)

    #if you can add none, skip to the next block
    if top == 0:
        return []
    #if all the vectors in the block are linearly independent, add as many as you need
    elif Delta == l:
        return [l_prev+i for i in range(top)]
    #otherwise, split into smaller blocks and take from each block as many vectors as you need
    else:
        t=0
        indices=[]
        i=0
        last_row=0
        while t<top:
            indices_new=Indep_rows(U[last_row:min(last_row+top-t,l)],U_prev+U[0:last_row],top-t)
            last_row=min(last_row+top-t,l)
```

```
            i=i+1
            t=t+len(indices_new)
            indices=indices+indices_new
        return indices


def Hadamards_inequality(B):
    n=len(B[0])
    Id=[list(x) for x in np.identity(n)]
    B=[list(x) for x in B if list(x)!=list(np.zeros(n))]

    #check if you need to remove some entries when computing the norms
    because the rows of B are not enough:
    rkB=Matrix(QQ,B).rank()
    if rkB >= n-1:
        eu_norms=[np.linalg.norm(x) for x in B]
    else:
        B_abs=[[abs(y) for y in x] for x in B]
        eu_norms=[np.linalg.norm(heapq.nlargest(rkB+1, x)) for x in B_abs]

    #order the vectors according to the norms you computed, in decreasing
    order:
    order=list(reversed(np.argsort(eu_norms)))
    eu_norms=[eu_norms[i] for i in order]
    B=[B[i] for i in order]

    #run Indep_rows() to undertand which rows have the highest determinant
    indices=Indep_rows(B,[],min(rkB,n-1))

    #compute and return the bound
    product=prod([eu_norms[i] for i in indices])
    Bound=n*np.floor(sqrt(min(n,rkB+1))*product)

    return Bound
```

Listing 2.1: Sage code for Hadamard's inequality comptuation

Local maximums computation (Section 2.2.2)

```
import numpy as np
import heapq



def doubleLocalMax(U,eps,multi):
    n=len(U[0])
    r=len(U)-n

    count_steps=0 #number of useless consecutive steps
    while count_steps < n-1:
        ker=Matrix(QQ, [eps]+[U[multi[i]] for i in range(n-2)]).
    right_kernel().basis()

        #find vector that maximizes the scalar product, and therefore
    determinant
        scal_prod= np.abs(Matrix(QQ,U)*ker[0])
```

```
        ind=np.argmax(scal_prod)

        #if you're substitution did not make a difference, count a useless step
        if scal_prod[ind] == scal_prod[multi[n-2]]:
            count_steps= count_steps+1
        #otherwise, the useless steps go back to zero
        else:
            count_steps=0
        #put the index you just changed in the last position, so it is going to be analyzed again in n-1 steps
        multi = [ind]+multi[0:n-2]

        #do the same but for eps, i.e. find eps(M)
        ker=Matrix(QQ, [U[multi[i]] for i in range(n-1)]).right_kernel().basis()
        eps=[int(x) for x in np.sign(ker[0])]

    return [eps,multi]



def Indep_rows(U,U_prev,maximum=None):
    #U is the matrix from which we take the rows
    #if it is an iteration, U_prev collects all the rows from the previous blocks
    n=len(U[0])
    l=len(U)
    l_prev=len(U_prev)
    if maximum==None:
        maximum=l

    #count how many vectors from the block U you can add to the ones taken from U_prev
    rk_new=Matrix(QQ,U+U_prev).rank()
    rk_old=Matrix(QQ,U_prev).rank()
    Delta=rk_new-rk_old
    top=min(Delta,maximum)

    #if you can add none, skip to the next block
    if top == 0:
        return []
    #if all the vectors in the block are linearly independent, add as many as you need
    elif Delta == l:
        return [l_prev+i for i in range(top)]
    #otherwise, split into smaller blocks and take from each block as many vectors as you need
    else:
        t=0
        indices=[]
        i=0
        last_row=0
        while t<top:
            indices_new=Indep_rows(U[last_row:min(last_row+top-t,l)],U_prev+U[0:last_row],top-t)
            last_row=min(last_row+top-t,l)
```

```
            i=i+1
            t=t+len(indices_new)
            indices=indices+indices_new
        return indices




def Hadamard(U,rkB):
    n=len(U[0])
    l=len(U)
    U_abs=[[abs(y) for y in x] for x in U]

    #check if you need to remove some entries when computing the norms
    because the rows of B are not enough:
    rkB=Matrix(QQ,B).rank()
    if rkB >= n-1:
        eu_norms=[np.linalg.norm(x) for x in U_abs]
    else:
        eu_norms=[np.linalg.norm(heapq.nlargest(rkB+1, x)) for x in U_abs]

    #order the vectors according to the norms you computed, in decreasing
    order:
    order=list(reversed(np.argsort(eu_norms)))
    eu_norms=[eu_norms[i] for i in order]
    U=[U[i] for i in order]

    #run Indep_rows() to undertand which rows have the highest determinant
    indices=Indep_rows(U,[],n-1)

    #compute and return the product
    product=prod([eu_norms[i] for i in indices if eu_norms[i]>1])

    return [product,[order[x] for x in indices]]




def Local_maximums(B,N=None,k=None):
    n=len(B[0])
    r=len(B)
    Id=[list(x) for x in np.identity(n)]
    U=B+Id
    rkB=Matrix(ZZ,B).rank()

    #maximum number of jumps within local maximums
    if N==None:
        N=10
    #maximum number of consecutive useless jumps
    if k==None:
        k=2

    _,init=Hadamard(U,rkB)
    exact=false
    maxDet=np.Infinity
```

```
    eps=[1 for i in range(n)]
    num_bounces=0

    for i in range(N):
        #find a local maximum
        eps,multi=doubleLocalMax(U,eps,init)

        #check "how global" the local maximum is by taking UM^{-1}
        M=Matrix(QQ,[eps]+[U[multi[i]] for i in range(n-1)])
        M_inv=M.inverse()
        V=Matrix(QQ,U)*M_inv

        #check the highest possible determinant from UM^{-1} with Hadamard'
    s inequality
        maximum_ratio,init=Hadamard([list(x) for x in V],rkB)

        #add the norm of the last vector which is eps*M^{-1}
        maximum_ratio=maximum_ratio*sqrt(n)*M_inv.norm()

        #check if you got the biggest determinant, or at least bound it
        if maximum_ratio <= 1:
            exact = true
            maxDet=abs(M.det())
            break
        else:
            newBound=maximum_ratio*abs(M.det())
            #check if the new bound is more effective
            if maxDet <= newBound:
                num_bounces=num_bounces+1
                if num_bounces >= k:
                    break
            else:
                maxDet=newBound
                num_bounces=0


    return n*np.floor(maxDet)
```

Listing 2.2: Sage code for Algorithm 6 (Local maximums computation)

Local maximums computation - v2 (Section 2.2.2)

```
import numpy as np
import heapq
import itertools as it



def doubleLocalMax(U,eps,multi):
    n=len(U[0])
    r=len(U)-n

    count_steps=0 #number of useless consecutive steps
    while count_steps < n-1:
        ker=Matrix(QQ, [eps]+[U[multi[i]] for i in range(n-2)]).
    right_kernel().basis()
```

```
14
15          #find vector that maximizes the scalar product, and therefore
        determinant
16          scal_prod= np.abs(Matrix(QQ,U)*ker[0])
17          ind=np.argmax(scal_prod)
18
19          #if you're substitution did not make a difference, count a useless
        step
20          if scal_prod[ind] == scal_prod[multi[n-2]]:
21              count_steps= count_steps+1
22          #otherwise, the useless steps go back to zero
23          else:
24              count_steps=0
25          #put the index you just changed in the last position, so it is
        going to be analyzed again in n-1 steps
26          multi = [ind]+multi[0:n-2]
27
28          #do the same but for eps, i.e. find eps(M)
29          ker=Matrix(QQ, [U[multi[i]] for i in range(n-1)]).right_kernel().
        basis()
30          eps=[int(x) for x in np.sign(ker[0])]
31
32      return [eps,multi]
33
34
35
36
37  def Indep_rows(U,U_prev,maximum=None):
38      #U is the matrix from which we take the rows
39      #if it is an iteration, U_prev collects all the rows from the previous
        blocks
40      n=len(U[0])
41      l=len(U)
42      l_prev=len(U_prev)
43      if maximum==None:
44          maximum=l
45
46      #count how many vectors from the block U you can add to the ones taken
        from U_prev
47      rk_new=Matrix(QQ,U+U_prev).rank()
48      rk_old=Matrix(QQ,U_prev).rank()
49      Delta=rk_new-rk_old
50      top=min(Delta,maximum)
51
52      #if you can add none, skip to the next block
53      if top == 0:
54          return []
55      #if all the vectors in the block are linearly independent, add as many
        as you need
56      elif Delta == l:
57          return [l_prev+i for i in range(top)]
58      #otherwise, split into smaller blocks and take from each block as many
        vectors as you need
59      else:
60          t=0
61          indices=[]
62          i=0
63          last_row=0
```

```
        while t<top:
            indices_new=Indep_rows(U[last_row:min(last_row+top-t,l)],U_prev
    +U[0:last_row],top-t)
            last_row=min(last_row+top-t,l)
            i=i+1
            t=t+len(indices_new)
            indices=indices+indices_new
        return indices




def Hadamard(U,rkB):
    n=len(U[0])
    l=len(U)
    U_abs=[[abs(y) for y in x] for x in U]

    #check if you need to remove some entries when computing the norms
    because the rows of B are not enough:
    rkB=Matrix(QQ,B).rank()
    if rkB >= n-1:
        eu_norms=[np.linalg.norm(x) for x in U_abs]
    else:
        eu_norms=[np.linalg.norm(heapq.nlargest(rkB+1, x)) for x in U_abs]

    #order the vectors according to the norms you computed, in decreasing
    order:
    order=list(reversed(np.argsort(eu_norms)))
    eu_norms=[eu_norms[i] for i in order]
    U=[U[i] for i in order]

    #run Indep_rows() to undertand which rows have the highest determinant
    indices=Indep_rows(U,[],n-1)

    #compute and return the product
    product=prod([eu_norms[i] for i in indices if eu_norms[i]>1])

    return [product,[order[x] for x in indices]]




def MaxMinor(LLt,maximum):
    n=len(LLt)
    num_rows=min(n,maximum)
    if num_rows==0:
        return 1
    else:
        maxdet=0
        for k in range(num_rows):
            multi=list(it.combinations(range(n), k))
            for i in range(len(multi)):
                Minor=[[LLt[k][j] for j in multi[i]] for k in multi[i]]
                maxdet=max(maxdet,abs(Matrix(RR,Minor).det()))
        return maxdet

```

```
119
120
121
122
123 def Local_maximums_2(B,N=None,k=None):
124     n=len(B[0])
125     r=len(B)
126     Id=[list(x) for x in np.identity(n)]
127     U=B+Id
128     rkB=Matrix(ZZ,B).rank()
129
130     #maximum number of jumps within local maximums
131     if N==None:
132         N=10
133     #maximum number of consecutive useless jumps
134     if k==None:
135         k=2
136
137     _,init=Hadamard(U,rkB)
138     exact=false
139     maxDet=np.Infinity
140     eps=[1 for i in range(n)]
141     num_bounces=0
142
143     for i in range(N):
144         #find a local maximum
145         eps,multi=doubleLocalMax(U,eps,init)
146
147         #check "how global" the local maximum is by taking UM^{-1}
148         M=Matrix(QQ,[eps]+[U[multi[i]] for i in range(n-1)])
149         M_inv=M.inverse()
150         V=Matrix(QQ,U)*M_inv
151
152         #check the highest possible determinant from UM^{-1} with both
    kinds of Hadamard's inequality
153         L=[x for x in V if np.linalg.norm(x)>1]
154         LLt=Matrix(RR,L)*Matrix(RR,L).transpose()
155         LLt=[list(x) for x in LLt]
156         maximum_ratio=sqrt(MaxMinor(LLt,min(len(L),n-1)))
157         maximum_ratio_2,init=Hadamard([list(x) for x in V],rkB)
158         maximum_ratio=min(maximum_ratio,maximum_ratio_2)
159
160         #add the norm of the last vector which is eps*M^{-1}
161         maximum_ratio=maximum_ratio*sqrt(n)*M_inv.norm()
162
163         #check if you got the biggest determinant, or at least bound it
164         if maximum_ratio <= 1:
165             exact = true
166             maxDet=abs(M.det())
167             break
168         else:
169             newBound=maximum_ratio*abs(M.det())
170             #check if the new bound is more effective
171             if maxDet <= newBound:
172                 num_bounces=num_bounces+1
173                 if num_bounces >= k:
174                     break
175             else:
```

```
                maxDet=newBound
                num_bounces=0

    return n*np.floor(maxDet)
```

Listing 2.3: Sage code for Algorithm 7 (Local maximums computation - v2)

Numerical Examples

Using the following code:

```
num=13 #number of examples
S=3 #determine how sparse the matrix has to be
for i in range(num):
    r=np.random.randint(6,17)
    l=np.random.randint(3,r-2)
    A=np.random.randint(0,30,size=(l,r)) #the signs are sill all +
    A=[list(x) for x in A]
    k=0
    while k < r:
        for j in range(l):
            A[j][k]=int(np.floor(np.random.randint(-1,S+1)/S)*A[j][k])
        k=k+1
    B=Matrix(Matrix(A).right_kernel().basis()).transpose()
    B=[list(x) for x in B]

    print('-----------------------------------------------')
    print('-----------------------------------------------')
    print('-----------------------------------------------')
    print('Matrix:')
    print(Matrix(A))
    print('Kernel basis:')
    print(Matrix(B))
    print('Hadamard:')
    start=tm.time()
    print(format(Hadamards_inequality(B),'.1E'))
    end=tm.time()
    print('Requested time:')
    print(end-start)
    print('')
    print('Local maximums:')
    start=tm.time()
    print(format(Local_maximums(B),'.1E'))
    end=tm.time()
    print('Requested time:')
    print(end-start)
    print('')
    print('Local maximums-v2:')
    start=tm.time()
    print(format(Local_maximums_2(B),'.1E'))
    end=tm.time()
    print('Requested time:')
    print(end-start)
```

Listing 2.4: Sage code for generating examples

we can randomly generate some examples to determine how the algorithms perform:

```
-------------------------------------------
-------------------------------------------
-------------------------------------------
Matrix:
[ 19   0   0   0 -25   0 -18   0   7]
[ 25  -2  16   7   5   7   0  -9   2]
[-25   0   0   0  -8 -29  26   0   0]
[-23   0  24   0  -2   0   0   0  20]
Kernel basis:
[    6     0     0     0     0]
[    0     1     0     0     0]
[    1     0     3     0     0]
[    2     2     2     3     0]
[   73   580  1396   580  1740]
[ -106  -862 -2076  -862 -2586]
[  -90  -783 -1886  -783 -2349]
[  -19  -334  -802  -333 -1006]
[   13    58   136    58   174]
Hadamard:
4.1E+14
Requested time:
0.0006425380706787109

Local maximums:
1.1E+07
Requested time:
0.0495297908782959

Local maximums-v2:
1.1E+07
Requested time:
0.05146217346191406
-------------------------------------------
-------------------------------------------
-------------------------------------------
Matrix:
[  0   0  23   0   0 -10   0   0   0   0 -28  14 -26   0 -11   0]
[  0 -13   0   0  29   0 -25 -29   5   0   0 -18   0   0   0   0]
[  0  29   0   0 -21  18 -24   0   0  13   0   4  -4   6   0  23]
[  0   0   0   0   0 -17   0 -20 -14   0   0   0   0   0   0   0]
[  0  15   5   0  -7   0   0  -4   0   0   0  -5  10   0   0   0]
[-19   0   0   0  -7   0   0   0 -22   5   0   0 -22   0   0  19]
[  0   0   0  16   0   0   0   0   0   0   0   0   0  -2 -17   0]
Kernel basis:
[    1     0     0     0     0     0     0     0     0     0]
[    1     2     0     0     0     0     0     0     0     0]
[    0     0     2     0     0     0     0     0     0     0]
[    0     0     0     1     0     0     0     0     0     0]
[    0     0     0     0     1     0     0     0     0     0]
[    2     2     2     0     0     4     0     0     0     0]
[    1     0     1     0     2     0     3     0     0     0]
[ 1035  1245  1000     0   987  1090   735  1260     0     0]
[-1481 -1781 -1431     0 -1410 -1562 -1050 -1800     0     0]
[  969   146  1377  1843   143  1677  1466  1131  2261     0]
[  -33    16   -54   -99     9   -72   -65   -39  -121     0]
```

```
[-2081 -2502 -2010     0 -1983 -2190 -1480 -2530     0]
[ -628  -756  -606     0  -596  -659  -446  -761     0]
[ 9197 12240  8381 -2134  9673  8925  5644 11237 -2618]
[-1082 -1440  -986   252 -1138 -1050  -664 -1322   308]
[-2696 -2976 -2721  -485 -2360 -3013 -2118 -3263  -595]
Hadamard:
6.9E+30
Requested time:
0.0009758472442626953

Local maximums:
1.1E+12
Requested time:
0.20803475379943848

Local maximums-v2:
6.2E+11
Requested time:
0.33669090270996094
-------------------------------------------
-------------------------------------------
-------------------------------------------
Matrix:
[ 10   0   0 -14   1   0  -9 -14   0 -18   0  26]
[  0   0   0   7   0   0   0   0  20 -14   0   0]
[  0   0   0   0   0   0   0   0   0   0   0   0]
[  0   0  -4   0   9   0 -12  29   5   0  23   0]
Kernel basis:
[   1    0    0    0    0    0    0    0    0    0]
[   0    1    0    0    0    0    0    0    0    0]
[   0    0    1    0    0    0    0    0    0    0]
[   0    0    0    2    0    0    0    0    0    0]
[   0    0    0    0    1    0    0    0    0    0]
[   0    0    0    0    0    1    0    0    0    0]
[   0    0    0    0    1    0    2    0    0    0]
[   0    0    0    0    0    0    0    1    0    0]
[ 966    0 1183  161  847    0  336 1687 2093]
[1380    0 1690  231 1210    0  480 2410 2990]
[-210    0 -257  -35 -184    0  -72 -368 -455]
[ 955    0 1170  161  838    0  333 1669 2070]
Hadamard:
2.0E+15
Requested time:
0.0015377998352050781

Local maximums:
8.4E+05
Requested time:
0.16769051551818848

Local maximums-v2:
7.5E+05
Requested time:
0.19174575805664062
-------------------------------------------
-------------------------------------------
-------------------------------------------
Matrix:
```

```
[  0   0  20   0   0  -2   0   0   0   0   0]
[  2   0   0   0   0 -26  21   0   0   0  21]
[-22   0   0 -13   0   0   0   0  -8  25   0]
[-29   0 -16 -11   0   0   0   0 -24   0   0]
[  0   0   0   0   0   0   0   0   0   0   0]
[  0   0   0   0   0   0   4   0 -16   0   6]
[  0   0   0   0   0   0  12   0   0   0   0]
Kernel basis:
[      2       0       0       0       0]
[      0       1       0       0       0]
[   4127       0   11550       0       0]
[ -47814       0 -133800       0       0]
[      0       0       0       1       0]
[  41270       0  115500       0       0]
[      0       0       0       0       0]
[      0       0       0       0       1]
[  19161       0   53625       0       0]
[ -18730       0  -52416       0       0]
[  51096       0  143000       0       0]
Hadamard:
2.4E+11
Requested time:
0.001325845718383789

Local maximums:
8.5E+06
Requested time:
0.0463106632232666

Local maximums-v2:
8.5E+06
Requested time:
0.050093889236450195
-------------------------------------------
-------------------------------------------
-------------------------------------------
Matrix:
[  0   0 -16   0   6   0  19   0 -26   0  -7   0  27]
[ -6  -8  -6  12 -15 -14   0   0   0  24  -7   0   0]
[  0   0   0   0 -22  21   0   0   0  27   0   0  17]
[  0   0   0   0   0  -5   0   0   0   0   0   0   0]
[ -5   0   0   0   0   0   0  -8   0   0   0   0  10]
[  0   0   0   0 -17   0  -8   6   0   0  17   0  -1]
[  0  19   0   0 -12 -23   0   0 -24   0 -28   0   0]
[  0   0   0  28   0 -26   0  10  22   0   0   0 -23]
[  0 -16   0   0   0   0  28 -26   0  29  27   3   0]
Kernel basis:
[        4         0         0         0]
[        0         8         0         0]
[        0         0         1         0]
[  2043697   6672791   3160801   7858077]
[  3991014  13030907   6172544  15345582]
[        0         0         0         0]
[ -7559184 -24681191 -11691112 -29065314]
[  3987785  13020365   6167555  15333180]
[ -1078458  -3521231  -1667953  -4146702]
[  1243274   4059370   1922860   4780428]
[  -786042  -2566471  -1215702  -3022362]
```

```
[100169250 327058651 154922860 385154278]
[  3190230  10416292   4934044  12266544]
Hadamard:
3.7E+24
Requested time:
0.0006687641143798828

Local maximums:
1.5E+11
Requested time:
0.04938173294067383

Local maximums-v2:
1.4E+11
Requested time:
0.05536770820617676
--------------------------------------------
--------------------------------------------
--------------------------------------------
Matrix:
[  0   0 -26  11   0   0  26  -4   7  18   0   0]
[  0   0   0   8   0   0   0 -18   0   0 -24   0]
[ 24  -9   0  27  15   0   0  22   0   0   0 -29]
[-24   0 -16   8   0   9   0  10  10  14   0   0]
[  0 -14   0   0 -13   0 -23   0   0   0   0   0]
Kernel basis:
[     1      0      0      0      0      0      0]
[     0      1      0      0      0      0      0]
[     0      0      1      0      0      0      0]
[     0      0      0      3      0      0      0]
[     0      6      0      0     23      0      0]
[     0      0      0      0      0      2      0]
[     0     -4      0      0    -13      0      0]
[   252   3852   2320   4548   3376   3132   4756]
[  -720 -11104  -6678 -13089  -9774  -9018 -13688]
[   336   5180   3114   6099   4570   4203   6380]
[  -189  -2889  -1740  -3410  -2532  -2349  -3567]
[   192   2925   1760   3453   2573   2376   3608]
Hadamard:
6.4E+22
Requested time:
0.0013802051544189453

Local maximums:
3.4E+08
Requested time:
0.1366126537322998

Local maximums-v2:
2.6E+08
Requested time:
0.15165019035339355
--------------------------------------------
--------------------------------------------
--------------------------------------------
Matrix:
[  0   0   0   0  12   0   0  -7   0   0   0  -1 -11  -9]
[  0   0   0   0   0   0   1   0   0  17  -8   0   0 -26]
```

```
[  0 -17 -29   0   0   0  15   0   0   0   0  -7   2   0]
[  2   0   0   0   0  -8   0 -20 -10   0  -5   0   2 -11]
[ 23  -8   0   0   7  10   0   0   0   0   0   0   6   1]
[  0   0 -27   0 -12 -26   0  11 -17  -2  -5 -24   0 -26]
[  0  20   0   0   0 -18   0   0   0 -27   0   0 -24 -18]
[ 14  29   0  16   0   0   0   0   0   0   0   0   2 -23]
Kernel basis:
[         1          0          0          0          0          0]
[         0          3          0          0          0          0]
[         0          0          1          0          0          0]
[         0          0          0          1          0          0]
[         7          5          0          4         10          0]
[  20542085   21209472   10139220   22982724   11989817   31989510]
[  44216358   45652876   21824424   49469760   25807814   68856636]
[  44647023   46097531   22036991   49951593   26059182   69527295]
[-136182307 -140606650  -67217214 -152362307  -79485684 -212072105]
[  18259642   18852866    9012640   20429092   10657626   28435120]
[  53866182   55616193   26587408   60266090   31440136   83883982]
[  85107132   87872117   42007393   95218818   49674582  132534450]
[ -33747723  -34844135  -16657290  -37757337  -19697568  -52554195]
[  -2934584   -3029921   -1448460   -3283246   -1712832   -4569930]
Hadamard:
1.8E+42
Requested time:
0.0009329319000244141

Local maximums:
3.9E+12
Requested time:
0.11005854606628418

Local maximums-v2:
2.6E+12
Requested time:
0.11481714248657227
-----------------------------------------
-----------------------------------------
-----------------------------------------
Matrix:
[  7   0   0   0   0   0   0   0  17   0   0   0   0]
[  0   0   0   6   0   0   0   0   8   0 -25  21   0]
[  0   0   0   0   0   0  -5   0   0   0  26   0 -15]
[  0   0  -4 -22 -17   0  18   0  -2   0   6   0   0]
[  0   2 -14  28   2  -6   0  12   0   0 -16  25  14]
[  0   0  12   0   0   0   0 -26   0   0 -21   0   0]
[-23  -1   0 -26  -6   0   0   0 -20   0  -7   0   0]
Kernel basis:
[      17         0         0         0         0         0]
[       1         2         0         0         0         0]
[       0         0         1         0         0         0]
[       1         2         0         3         0         0]
[  116282     27746    476770    444522    515970         0]
[ -543394   -129648  -2227230  -2076630  -2410351         0]
[  143059     34137    586503    546840    634725         0]
[   80535     19215    330072    307755    357210         0]
[      -7         0         0         0         0         0]
[       0         0         0         0         0         1]
[  -99710    -23790   -408660   -381030   -442260         0]
```

```
[ -118700   -28322  -486500  -453608  -526500        0]
[ -220517   -52615  -903845  -842732  -978159        0]
Hadamard:
6.8E+31
Requested time:
0.0015988349914550781

Local maximums:
1.7E+10
Requested time:
0.09304547309875488

Local maximums-v2:
1.7E+10
Requested time:
0.10294342041015625
-------------------------------------------
-------------------------------------------
-------------------------------------------
Matrix:
[  0  19   0   0   0   0   0  13  -6   0  27   0   0]
[  0   0   0   0   0   0 -12   0  16  19  -4 -14   0]
[ 12   0   0   0 -27   0   0 -29 -24  -6  24   0 -17]
Kernel basis:
[  1   0   0   0   0   0   0   0   0   0   0]
[  0   1   0   0   0   0   0   0   0   0   0]
[  0   0   1   0   0   0   0   0   0   0   0]
[  0   0   0   1   0   0   0   0   0   0   0]
[  0   0   0   0   1   0   0   0   0   0   0]
[  0   0   0   0   0   1   0   0   0   0   0]
[  0   0   0   0   0   0   1   0   0   0   0]
[  0   2   0   0   0   0   0   3   0   0   0]
[  0   3   0   0   0   0   0   2   9   0   0]
[ 70  82   0   0 140   0 204 186  40 238   0]
[  0  -1   0   0   0   0   0  -1   2   0   0]
[ 95 115   0   0 190   0 276 255  64 323   0]
[-24 -38   0   0 -51   0 -72 -75 -24 -84   0]
Hadamard:
9.3E+10
Requested time:
0.0018160343170166016

Local maximums:
1.3E+06
Requested time:
0.18780875205993652

Local maximums-v2:
1.2E+06
Requested time:
0.2011265754699707
-------------------------------------------
-------------------------------------------
-------------------------------------------
Matrix:
[  0   0   0   0   2  24   0   0]
[ 18   0   0 -17   3   0   0   0]
[-25   0   0   0  16 -23   0   0]
```

```
[  0  10   0   0   0 -20  18   0]
[ -6 -18 -14   0 -17   0   0  21]
Kernel basis:
[ 731    0    0]
[  46   63    0]
[   0    0    3]
[ 954    0    0]
[1020    0    0]
[ -85    0    0]
[-120  -35    0]
[1074   54    2]
Hadamard:
5.7E+06
Requested time:
0.0004973411560058594

Local maximums:
2.4E+06
Requested time:
0.028583049774169922

Local maximums-v2:
9.7E+05
Requested time:
0.0458683967590332
-----------------------------------------
-----------------------------------------
-----------------------------------------
Matrix:
[  0   0 -11   5 -21   0 -29   0   0   0]
[-11   0   0   0  13   0   0  -8 -13   0]
[ 17   0   8  -1 -15   0   0   0  12  14]
[  4   0   0 -28   0   0  16   0   0 -10]
[-21   0   0  18 -15   0   0 -12   0   0]
Kernel basis:
[      1       0       0       0       0]
[      0       1       0       0       0]
[      0       0       1       0       0]
[  59230       0   12612   75324       0]
[-281119       0  -59860 -357508       0]
[      0       0       0       0       1]
[ 213781       0   45521  271872       0]
[ 440242       0   93743  559871       0]
[-552038       0 -117548 -702044       0]
[ 176206       0   37520  224088       0]
Hadamard:
3.3E+18
Requested time:
0.0010342597961425781

Local maximums:
1.5E+07
Requested time:
0.09733843803405762

Local maximums-v2:
1.4E+07
Requested time:
```

```
0.10497784614562988
-------------------------------------------
-------------------------------------------
-------------------------------------------
Matrix:
[  0   0   0   0 -18   0   0   0]
[ 23   0   0   0   4   1   0 -22]
[  0   0   7   0   0   0   0   0]
Kernel basis:
[ 1  0  0  0  0]
[ 0  1  0  0  0]
[ 0  0  0  0  0]
[ 0  0  1  0  0]
[ 0  0  0  0  0]
[21  0  0 22  0]
[ 0  0  0  0  1]
[ 2  0  0  1  0]
Hadamard:
7.6E+02
Requested time:
0.0005559921264648438

Local maximums:
3.4E+02
Requested time:
0.04072856903076172

Local maximums-v2:
3.4E+02
Requested time:
0.04299163818359375
-------------------------------------------
-------------------------------------------
-------------------------------------------
Matrix:
[-13 -11   0 -22   0   0   0 -10]
[  0   0   0   0  11   0   0   0]
[  0   0   0   3   0  13   0   0]
[ 28 -20   6   0  21 -20   6   0]
[ 28 -16   0   0   0   4   0  19]
Kernel basis:
[     1      0      0]
[ 10205  16662      0]
[     0      0      1]
[ -8814 -14391      0]
[     0      0      0]
[  2034   3321      0]
[ 40792  66610     -1]
[  8164  13332      0]
Hadamard:
7.9E+09
Requested time:
0.0008985996246337891

Local maximums:
5.6E+05
Requested time:
0.028952836990356445
```

```
462
463 Local maximums-v2:
464 5.6E+05
465 Requested time:
466 0.0314631462097168
```

Listing 2.5: Examples

Bibliography

[1] B. Sturmfels P. Diaconis. Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, 26:363–397, 1998.

[2] S. Petrovic. What is... a Markov Basis? *Notices of the American Mathematical Society*, 66(7):1088–1092, 2019.

[3] D. Holmes. The norm of the saturation of a binomial ideal, with applications to Markov bases. *Algebraic Statistics*, 11:169–187, 2020.

[4] B. Sturmfels. *Grobner Bases and Convex Polytopes*. University Lecture Series. American Mathematical Society, 1996.

[5] A. Takemura H. Hara, S. Aoki. Running markov chain without markov bases. 08 2011.

[6] J. Vaaler E. Bombieri. On Siegel's lemma. *Inventiones mathematicae*, 73:11–32, 1983.

[7] M. Różański R. Wituła. Hadamard's optimized inequality. *Linear Algebra and its Applications*, 620:109–123, 2021.