Introduction

The following is an exposition of the recent papers [IQS17] and [IQS18] which together give an algebraic polynomial time algorithm to compute noncommutative rank of a symbolic matrix over (*sufficiently*) large fields.

The Problem

The problem can be stated in multiple interesting ways and these can be read in [IQS17]. We will just look at one particular statement and work with it. Given a matrix T with each entry a linear polynomial over $\mathbb{F}$ in formal variables $(x_1, \cdots x_m)$ we want to compute its rank over the field of rational functions i.e $\mathbb{F}(x_1, \cdots, x_m)$. If the variables commute this is called the *Edmond*'s problem and the decision version i.e. checking if rank is full or not is the classic SDIT problem which is a subcase of PIT. There is a simple randomized algorithm, which is plug in random values from the field i.e. - $(x_1, \cdots, x_m) \to (\alpha_1, \cdots, \alpha_m)$ and compute rank over $\mathbb{F}$. If the variables don't commute, then formulating the problem becomes trickier. The rank is now not over $\mathbb{F}(x_1, \cdots, x_m)$ but over its non-commutative analog called the *skew field*, denoted as $\mathbb{F}\langle\!\langle x_1, \cdots, x_m \rangle\!\rangle$[1]. However, explicitly constructing this skew field is not easy (see appendix for details) and we will use a definition of non-commutative rank that will not require this.

What is non-commutative rank?

First, we reinterpret the rank of $T = \sum_{i=1}^m x_i B_i, \; B_i \in M(n, \mathbb{F})$. Define $\mathcal{B} = \mathsf{span}_{\mathbb{F}}(B_1, \cdots B_m)$ and $\mathsf{rk}(\mathcal{B})$ as the maximum rank of a matrix in $\mathcal{B}$ i.e $\max\limits_{(a_1,\cdots,a_m) \in \mathbb{F}} \mathsf{rk}(\sum_{i=1}^m a_i B_i)$. Throughout the article, $\mathcal{B}$ will be thus the input as above, unless explicitly mentioned otherwise.

Lemma 1. *If* $\mathbb{F} = \Omega(n)$, $\mathsf{rk}_{\mathbb{F}(X)}(T) = \mathsf{rk}(\mathcal{B})$

Proof. Let $B = \sum_i a_i B_i \in \mathcal{B}$ be a matrix of largest rank. Any $d \times d$ minor, A, of B corresponds to a $d \times d$ minor A' of T such that $\mathsf{det}(A')(a_1, \cdots a_m) = \mathsf{det}(A)$. If $\mathsf{det}(A') = 0$ then any evaluation is 0. Thus, $\mathsf{rk}(T) \geqslant \mathsf{rk}(B)$. Since $\mathbb{F}$ is large enough, if $\mathsf{det}(A') \neq 0$, we can find a point $(a_1, \cdots, a_m)$ on which the determinant doesn't vanish i.e $\mathsf{det}(A) \neq 0$. Thus, $\mathsf{rk}(T) = \mathsf{rk}(\mathcal{B})$. □

This fails for small fields. For example, the polynomial $x(x-1)$ is not identically 0 but over $\mathbb{F}_2$ it vanishes on both points.

This, thus, gives us a definition of rank without using the function field which is what we want. But directly using this will not work for noncommutative rank. Let's look at the following example (from Visu's IAS talk),

$$A = \begin{pmatrix} 0 & 1 & x_1 \\ -1 & 0 & x_2 \\ -x_1 & -x_2 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & x_1x_2 - x_2x_1 \end{pmatrix}$$

$\mathsf{det}(A) = (x_1x_2 - x_2x_1)$ and thus, $\mathsf{rk}_{\mathbb{F}(X)}(A) = 2$ but $\mathsf{rk}_{\mathbb{F}\langle\!\langle X \rangle\!\rangle}(A) = 3$

[1] Typesetting this was the hardest part in writing this! I ended up stealing TEXfrom [GGdOW15]

Shrunk Subspaces

Definition 1. *A subspace* $U \subseteq \mathbb{F}^n$ *is a c-shrunk subspace of* $\mathcal{B}$ *if* $\exists W \subseteq \mathbb{F}^n$ *such that* $\mathsf{dim}(W) \leqslant \mathsf{dim}(U) - c$ *and* $\forall B \in \mathcal{B},\ \ B(U) := \{Bu \mid u \in U\} \subseteq W$

This enables us to redefine noncommutative rank with no reference to the skew field.

Definition 2. $\mathsf{ncrk}(\mathcal{B}) = n - \max\{c \mid \exists$ *c-shrunk subspace of* $\mathcal{B}\}$

Cohn had given a construction of the free field and proved that the $\mathsf{rk}_{\mathbb{F}\langle\!\langle X \rangle\!\rangle}(T) = \min\limits_{s \in \mathbb{Z}^+} s$ such that $T = PQ$ where P, Q are homogeneous linear polynomials of sizes $n \times s$ and $s \times n$, respectively. [FL04] then showed that this is equivalent to the above definition, i.e

Theorem 2 ([FL04] + [Coh95])**.** $\mathsf{rk}_{\mathbb{F}\langle\!\langle X \rangle\!\rangle}(T) = \mathsf{ncrk}(\mathcal{B})$

This also lets us conclude the following lemma

Lemma 3. $\forall \mathcal{B}\ \ \mathsf{rk}(\mathcal{B}) \leqslant \mathsf{ncrk}(\mathcal{B})$

Proof. If we have a c-shrunk subspace U. Then every $B \in \mathcal{B}$ takes $\mathsf{dim}(BU) \leqslant \mathsf{dim}(U) - c$. Then $\mathsf{rk}(B) = \mathsf{dim}(BV) \leqslant \mathsf{dim}(BU) + n - \mathsf{dim}(U) \leqslant n - c$ □

Again for our above example, $B_1 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $B_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$ and $B_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$

Note that for any $0 \neq v \in \mathbb{F}^3$, $\mathsf{rk}(B_1 v, B_2 v, B_3 v) \geqslant 2$. And thus, if U is a shrunk subspace then $U \cong \mathbb{F}^3$ but then the image is all of $\mathbb{F}^3$.

An aside - [FL04] showed that we also have an upper bound $\mathsf{rk}(\mathcal{B}) \leqslant \mathsf{ncrk}(\mathcal{B}) \leqslant 2\mathsf{rk}(\mathcal{B})$ and recently [DM16] proved that this bound is tight by giving an explicit examples where $\mathsf{ncrk}(\mathcal{B})$ is arbitrarily close to $2\mathsf{rk}(\mathcal{B})$.

First steps to an algorithm

A PSPACE algorithm

Given the above definition, the most natural way is to assume say that there is a c shrunk subspace U of dimension d. Assume generic basis of size d. Then we can set up a system of equations which have a solution iff there is a W of dimension $d - c$. We can iterate over each $c \leqslant d \leqslant n$ and thus find the largest c.

A Randomized algorithm

We can try to get an inspiration from the randomized algorithm in the commutative case but that fails as field elements commute but the variables don't. Thus, we need to plug in elements from a noncommutative ring which should also have a multiplication defined with $\mathbb{F}$ (in other words be a vectorspace over $\mathbb{F}$). There's a natural candidate - Matrices over $\mathbb{F}$! But does Schwarz-Zippel type lemma hold? If yes, upto what size matrices do we need to plug in?

Definition 3. *d^{th}- blowup of $\mathcal{B}$ is defined as $\mathcal{B}^{\{d\}} = \mathcal{B} \otimes_{\mathbb{F}} M(d, \mathbb{F}) \subset M(nd, \mathbb{F})$.*

It is a matrix space too and clearly, $\mathsf{rk}(\mathcal{B}^{\{d\}}) \geqslant d \cdot \mathsf{rk}(\mathcal{B})$ (just tensor the max rank matrix in $\mathcal{B}$ with identity). But it can be more. Recall that $T = \sum_i x_i B_i$, if we plug in $d \times d$ matrices for the x_i we are formally tensoring it with elements of $M(d, \mathbb{F})$ and hence the above definition. Now, clearly if for some set of matrices A_i we get a non-zero determinant after plugging in, then the original T had full rank but for what sizes do we need to check to conclude the converse? To reformulate this, let's say we have m, $d \times d$ variable matrices $Y_k = (y^k_{ij})\ \ i, j \in [d]$. We want to check if $D_d = \mathsf{det}(\sum_i Y_i \otimes B_i)$ is identically 0 or not. Let's assume we have a bound N such that if we have $D_d = 0\ \ \forall d \leqslant N$, then we have $\mathsf{ncrk}(\mathcal{B}) < n$. We have thus "reduced" the decision problem of noncommutative rank to N instances of SDIT and therefore have a natural $poly(n, N)$ randomized algorithm. This, makes the result of the paper even more surprising as derandomizing SDIT is considered hard (as it would yield nontrivial circuit lower bounds) but this paper gives a deterministic algorithm for the noncommutative case.(Infact, it computes the rank along with a certificate) But how do we get N? This is where invariant theory kicks in. As we won't be needing this, it is discussed at the end.

A Broad Outline

Now that we have understood the problem, let's look at what the main result is.

Theorem 4. *Let $\mathbb{F}$ be such that $|\mathbb{F}| = n^{\Omega(1)}$. Given $\mathcal{B} = \mathsf{span}_{\mathbb{F}}(B_1, \cdots B_m)$, there exists an algorithm that computes $\mathsf{ncrk}(\mathcal{B})$ in $n^{O(1)}$ arithmetic operations. It also outputs a matrix $A \in \mathcal{B}^{\{d\}},\ \ d \leqslant n+1$ of rank rd (certifying $\mathsf{ncrk}(B) \geqslant r$) and a $(n-r)$-shrunk subspace U (certifying $\mathsf{ncrk}(B) \leqslant r$)*

We'll first see what their algorithm broadly does and then get into the details and explain why and how each step works. The main idea to derandomize the above algorithm by carefully using the structure of the blowups. Instead of going sequentially for each d and we rather start with some rank r matrix, A, (say B_1) in $\mathcal{B}$. With each "blowup" we construct a matrix A' of rank $> rd$ for some d and then round it up to $(r+1)d$. We stop when we can no longer do this and at that point we output a certificate i.e a $n-r$-shrunk subspace which proves that the non-commutative rank is indeed r. Clearly, as rank is at most n we have a polynomially many iterations. There are 3 crucial subroutines in this and we need to ensure that each of these can be done in polynomial time -

1. Wong Sequences - To either output shrunk subspace or create A'

2. Regularity Lemma- To round up the rank

3. Blowup Control - To reduce the blowup back to $\leqslant n+1$

A visual description follows.

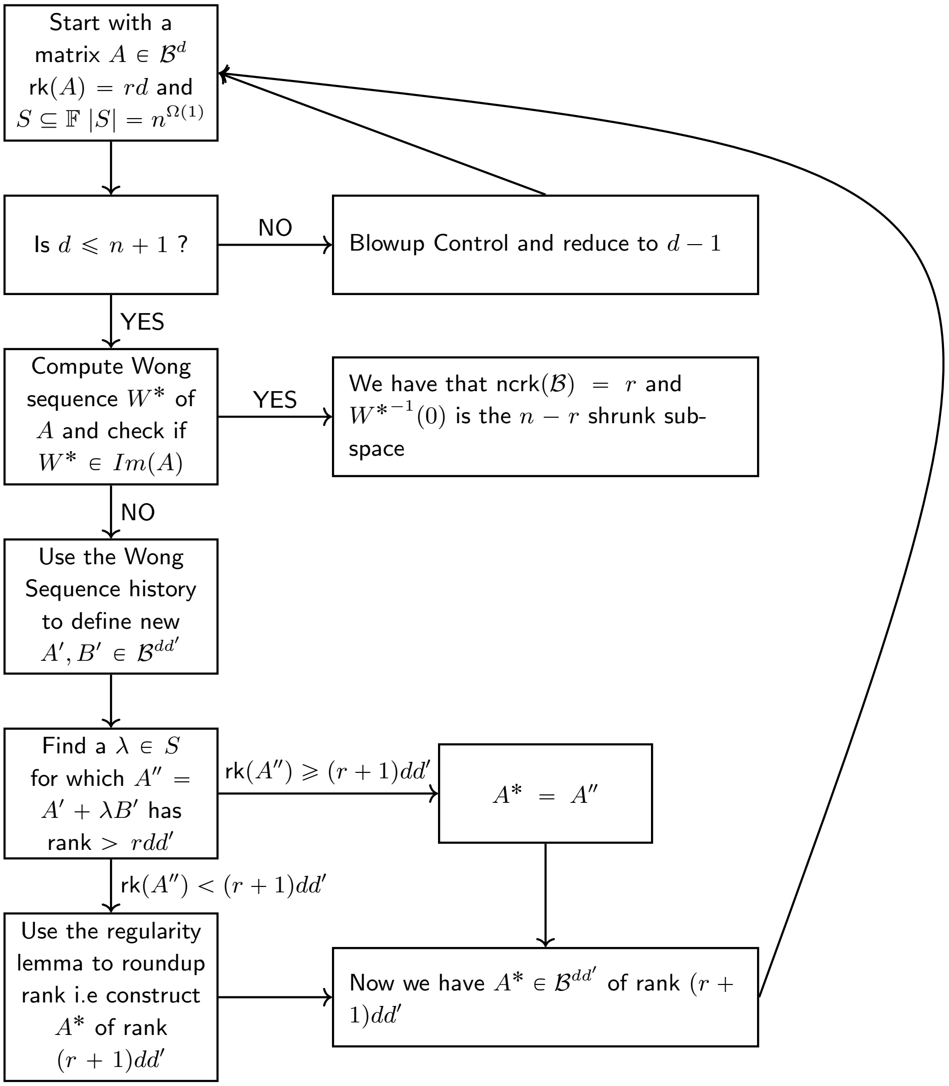
Start with a matrix $A \in \mathcal{B}^d$ rk$(A) = rd$ and $S \subseteq \mathbb{F}$ $|S| = n^{\Omega(1)}$
Is $d \leqslant n + 1$?
NO
Blowup Control and reduce to $d - 1$
YES
Compute Wong sequence W^* of A and check if $W^* \in Im(A)$
YES
We have that ncrk$(\mathcal{B}) = r$ and ${W^*}^{-1}(0)$ is the $n - r$ shrunk subspace
NO
Use the Wong Sequence history to define new $A', B' \in \mathcal{B}^{dd'}$
Find a $\lambda \in S$ for which $A'' = A' + \lambda B'$ has rank $> rdd'$
rk$(A'') \geqslant (r+1)dd'$
$A^* = A''$
rk$(A'') < (r+1)dd'$
Use the regularity lemma to roundup rank i.e construct A^* of rank $(r+1)dd'$
Now we have $A^* \in \mathcal{B}^{dd'}$ of rank $(r+1)dd'$

Figure 1: The overall algorithm

The Wong Sequence

Definition 4. *Given $A \in M(n, \mathbb{F})$, the second Wong sequence of $(A, \mathcal{B})$ is the sequence of subspaces in $\mathbb{F}^n$: $W_0 = (0)$, $W_i = \mathcal{B}(A^{-1}(W_{i-1}))$.*

Clearly if $A \in \mathcal{B}$, $W_i \subset W_{i+1}$. As these are subspaces, the dimension increases by at least 1 (if not 0) and can be at most n. Thus, the sequence has size say ℓ steps where $\ell \leqslant r \leqslant n$ where $r = \mathsf{rk}(A)$. W_ℓ is then called the limit of this sequence, denoted as W^*.

We have the following result from [IKQS15],

Theorem 5. *Let $A \in \mathcal{B}$ of rank r and let W^* be the limit of the second Wong sequence of $(A, \mathcal{B})$. Then, $\mathsf{ncrk}(\mathcal{B}) = r$ if and only if $W^* \subseteq \mathsf{im}(A)$. If this is the case then $A^{-1}(W^*)$ is a $(n-r)$-shrunk subspace of $\mathcal{B}$. In the algebraic RAM model as well as over $\mathbb{Q}$ we can detect whether $W^* \subseteq \mathsf{im}(A)$ and if so compute a shrunk subspace in deterministic polynomial time.*

If we start with a random matrix, with high probability $\mathsf{rk}(A) = crk(\mathcal{B})$. This theorem says that we can compute W^* and thus we have a certificate in case $\mathsf{ncrk}(\mathcal{B}) = crk(\mathcal{B})$. However, if this case doesn't occur what do we do? We might naively try to *augment* i.e find a matrix of higher rank and repeat but alas no such exist! ($r = \mathsf{rk}(A) = crk(\mathcal{B})$). The solution - blowup. As $\mathsf{ncrk}(\mathcal{B}) \geqslant r + 1 \implies \mathsf{ncrk}(\mathcal{B}^{\{d\}}) \geqslant (r + 1)d$. We have $\mathsf{rk}(A \otimes I_d) = rd$ and as there is some slack now we have some hope of constructing a higher rank matrix but it can still happen that $crk(\mathcal{B}^{\{d\}}) = rd$. However, this doesn't occur if $\mathcal{B}$ is 2 dimensional i.e. its generated by 2 matrices. In general, spaces where commutative rank equals noncommutative rank are called *compression spaces.*

Theorem 6. *([AS78]) If $|\mathbb{F}| > n$ and $\mathcal{B} = \mathsf{span}_{\mathbb{F}}(A, B) \subset M(n, \mathbb{F})$. Then $crk(\mathcal{B}) = \mathsf{ncrk}(\mathcal{B})$*

The following is my proof so the reader must be skeptical! A clean, geometric (and a much more general) proof appears in [EH88] but that uses sheaves and vector bundles!!

Proof Sketch. Assume there is matrix $C \in \mathcal{B}$ of highest rank say, k. We can assume it's of the form

$$C = \left[\begin{array}{c|c} C' & 0 \\ \hline 0 & 0 \end{array}\right]$$

where C' is $k \times k$. Complete the basis by picking a D i.e. $\mathcal{B} = \mathsf{span}_{\mathbb{F}}(C, D)$. Take $U = \mathbb{F}^n$ and $W = Im(C)$. If we show that $Im(D) \subset W$, then U is a shrunk subspace and then $\mathsf{ncrk}(\mathcal{B}) = k = crk(\mathcal{B})$. Assume not, and let $v \in \mathbb{F}^n$ such that $Dv \notin W$.

This means that for some l, m l^{th} row of D times v is non-zero for $k < l \leqslant n$ and choose m such that $D(l, m) \neq 0$ Consider $E = (C + \mu D)$. The idea is that if any $k + 1 \times k + 1$ minor in E has a determinant that is not identically 0, we can find a μ that makes it non-zero (as $|\mathbb{F}| > n$) and that would contradict the fact that max rank is k. However, consider the minor, M, formed by first k rows and columns and the l^{th} row and m^{th} column. It is of the form,

$$M = \left[\begin{array}{c|c} C' + \mu D' & \mu \cdot w \\ \hline \mu \cdot u & \mu \cdot x \end{array}\right]$$

But if that has identically 0 determinant so does,

$$M' = \left[\begin{array}{c|c} C' + \mu D' & \mu \cdot w \\ \hline u & x \end{array}\right]$$

Substituting $\mu = 0$ we get $\mathsf{det}(M') = x.\mathsf{det}(C') \neq 0$ as $\mathsf{det}(C') \neq 0$ as C is of rank k and $x = D(l, m) \neq 0$. Therefore, $im(D) \subset W$ and thus we are done. □

Say the length of Wong sequence of $(A, \mathcal{B})$ is $l \leqslant r$ i.e. after ℓ steps, W_l moves out of $im(A)$. We can then find *witness* matrices $C_i \in \mathcal{B}$ such that $C_i(A^{-1}W_{i-1}) \in W_i$, that is to say that $C_l A^{-1} C_{l-1} A^{-1} \cdots C_1 A^{-1}(0) \not\subset im(A)$ we can embed the sequence $C_1 \cdots, C_l \in M(n, \mathbb{F})$ 'suitably' into say $C \in \mathcal{B}^{\{l\}}$. Embed A trivially i.e. $A' = A \otimes I$ and let $\mathcal{B}' = \mathsf{span}_{\mathbb{F}}(A', C) \subset \mathcal{B}^{\{l\}}$. The embedding is created such that the Wong sequence of $(A', \mathcal{B}')$ escapes $Im(A) \iff$ the original one does. Since we have the first one escapes, so does this and therefore $\mathsf{ncrk}(\mathcal{B}') > rl$.

We already know that $\mathsf{ncrk}(\mathcal{B}^{\{l\}}) \geqslant \mathsf{ncrk}(\mathcal{B}) \cdot l \geqslant (r+1)l > rl$. Why did we then create the 2-dimensional $\mathcal{B}'$? Because $\mathcal{B}'$ is 2 dimensional, from Theorem 6, we know that $\mathsf{rk}(\mathcal{B}') = \mathsf{ncrk}(\mathcal{B}') > rd$. Thus, if our set S is big enough i.e $> nl$ we will find a $\lambda \in S$ such that $A' + \lambda C$ has higher rank.

The regularity lemma

An informal statement

Theorem 7. *Given $A \in \mathcal{B}^{\{d\}}$ such that $rd < \mathsf{rk}(A) < (r+1)d$ for some r. There is a poly(n,d) algorithm that outputs $A' \in \mathcal{B}^{\{d\}}$ such that $\mathsf{rk}(A') \geqslant (r+1)d$*

These can be unpacked as 2 claims. One, that there exists always such a matrix which is equivalent to saying that $\mathsf{rk}(\mathcal{B}^{\{d\}})$ is always a multiple of d. This is non-constructive as it is just an existential claim. [DM16] gave another proof of this part of the claim. The second, which is needed for an algorithm, is that we can compute this A' given A. We present the proof below.
Note - There are many technical details in this section and the reader who just wishes to get an overall picture of the proof may read the intuition and skip the details.

Intuition

Another definition of rank is $\mathsf{rk}(A) = \mathsf{dim}(Im(A))$. Say we have a vector space $V = (v_1, \cdots v_{2n}) \cong \mathbb{R}^{2n}$ and $A \in M(2n, \mathbb{R})$. We can view V as a vector space over $\mathbb{C}$ by defining scalar multiplication as follows,

$$i \cdot v_i = v_{i+1}, \quad i \cdot v_{i+1} = -v_i \ \ \forall i\{1, 3, \cdots, 2n-1\}$$

Thus, $V \cong \mathbb{C}^n$. But it might be the case that a $\mathbb{R}^{2n}$ subspace $U \subset V$ may no longer be a subspace over $\mathbb{C}^n$. For example, $U = \mathsf{span}_{\mathbb{R}}(v_1)$ is a $\mathbb{R}^{2n}$ subspace but not a $\mathbb{C}^n$ subspace as $i \cdot v_1 = v_2 \notin U$. Moreover, it's easy to see that if U is a $\mathbb{C}^n$ subspace of dimension k it is a $\mathbb{R}^{2n}$ subspace of dimension $2k$. Thus, a way to prove that $2 | \mathsf{rk}(A)$ is by showing that $\mathsf{im}(A) = \{Av \mid v \in \mathbb{R}^{2n}\}$ is a $\mathbb{C}^n$ subspace. The natural plan now is this, try to embed our ambient vector space $\mathbb{F}^{nd}$ into a suitable $\mathbb{F}'^{nd'}$ such that the image of our matrix space $\mathcal{B}^{\{d\}} \subset M(nd, \mathbb{F})$ is a subspace in $\mathbb{F}'^{nd'}$ and moreover, the *"degree"* of $\mathbb{F}'$ is d over $\mathbb{F}$ just like for $\mathbb{C}$ it was 2.

Making the intuition rigorous

While the idea may sound simple there are many problems we face in making the above work for general $\mathbb{F}$ and d. Let's look at each and resolve them. Each step is progressively more technical so feel free to stop when you are satiated!

Problem 1 - What is the a d-degree analog of $\mathbb{C}$? - Division algebras

Not in general. Say we are working with $\mathbb{R}$, then let's try to see what properties do we need from the degree d extension say D to be able to carry out the above plan. Clearly.

it must be an ring i.e we need addition and multiplication. To make sure we get a vector space over D and not just a module, we need that D has multiplicative inverse. To see this if $d \in D$ doesn't have an inverse, then we run into stuff like, $v \notin \mathsf{span}_D(dv)$. Also, we need to be able to define multiplication between $\mathbb{R}$ and D. Thus, D is an $\mathbb{R}$ algebra. Wait then! You may say that well, isn't D just a degree d field extension of $\mathbb{R}$. For $d = 2$, that is certainly the case but we'll see that doesn't quite work. There are no more *finite* field extensions of $\mathbb{R}$!! Well, we don't really need commutativity here. Sure, the proof becomes harder but it goes through. This lack of constraint does help us as for $d = 4$ we have the famous quaternions which do the job. To summarize, D is a non-commutative field of degree d over $\mathbb{R}$ or in general $\mathbb{F}$. Another name for non-commutative fields is *division algebra.*

Problem 2 - Even division algebras don't exist.
In what might seem like a cruel joke, here's a famous theorem which might appear as a blow to our plans.

Theorem 8 (Frobenius)**.** *If $\mathbb{F}$ is algebraically closed it has no dimension $d > 1$ division algebras. If $\mathbb{F} = \mathbb{R}$ (or real closed), then the only associative division algebras are of dimension* $1, 2, 4$.

So if we can't construct a D over R what do we do. Well, go to a field extension of $\mathbb{R}$ and then construct D. But didn't I say there are no fields over $\mathbb{R}$? I didn't. There are no *finite* extensions, we can always construct $\mathbb{R}(X)$ which is the field of all rational polynomials in 1 variable and then construct D over it. So now, $\mathbb{R} \hookrightarrow \mathbb{R}(X_1, \cdots X_n) \hookrightarrow D$ where the first is not finite but the second is. This is not quite it for a few technical reasons but we will get back to this later.

Problem 3 - Dealing with non-commutativity
We are familiar with vector spaces over fields and the case with division algebras is pretty much the same. However, there are certain places where non-commutativity messes with our intuitions. The main one here is that *scalar* multiplication by D is not a D-linear map. Let's do this slowly. Say, D is a division algebra over $\mathbb{F}$ such that its *index* is d, i.e, $dim_{\mathbb{F}}(D) = d^2$ (This is dimension of D as a vector space over $\mathbb{F}$). It's center $K = \{d \in D \mid d \cdot x = x \cdot d \ \forall x \in D \}$. In the above example with quaternions, $\mathbb{F} = K = \mathbb{R}$. But in general, K would be a field extension over $\mathbb{F}$. D is called *central* if $\mathbb{F} = K$. Define D^{op} to be the opposite division algebra which has the same elements as D but the multiplication is opposite, i.e., $a * b = b \cdot a$ where $*$ is the multiplication in D^{op}. We use the following theorem to make sense of the multiplication in D. See corollary 15.5 in [Lam] for the proof.

Theorem 9. *If D is a central division algebra over $\mathbb{F}$ such that $dim_{\mathbb{F}}(D) = d^2$, then, $D \otimes_{\mathbb{F}} D^{op} \cong M(d^2, \mathbb{F})$.*

Now we can forget about division algebras and view them as matrices.

$$D \hookrightarrow D \otimes_{\mathbb{F}} D^{op} \cong M(d^2, \mathbb{F}) : x \to x \otimes 1 \to M_x$$

Similarly, $y \in D^{op} = 1 \otimes y \to M_y$. The above isomorphism is such that M_x and M_y commute.

The intuitive plan was to take the matrix A and look at its image and compute its dimension. The weird thing due to non-commutativity is that even the image of multiplication by $x \in D$ is not a $D-$subspace. Let x be represented by $M_x \in M(d^2, \mathbb{F})$. $\mathsf{im}(M_x) = \{M_x v \mid v \in \mathbb{F}^{d^2}\}$. But, for an arbitrary $z \in D, z \cdot M_x v = M_z M_x v \notin \mathsf{im}(M_x)$. However, it is a subspace over D^{op} as $\forall\ t \in D^{op},\ \ t \cdot M_x v = M_t M_x v = M_x(M_t v) \in \mathsf{im}(M_x)$. Now we are safe as D^{op} also has $dim_{\mathbb{F}}(D^{op}) = d^2$. (In general $D \ncong D^{op}$ but it holds if index is finite)

Moreover, we can write $\mathbb{F}^{nd^2}$ as n-tuples of $\mathbb{F}^{d^2}$ i.e. $(\mathbb{F}^{d^2})^n$ and for any $B \in M(n, \mathbb{F})$, $x \in D$, $\mathsf{im}(B \otimes_{\mathbb{F}} M_x)$ is a D^{op} subspace. Thus we have,

Lemma 10. *If $A \in M(n, \mathbb{F}) \otimes_{\mathbb{F}} D \hookrightarrow M(nd^2, \mathbb{F})$ then $\mathsf{im}_{\mathbb{F}}(A)$ is a D^{op} subspace and thus, $\mathsf{rk}_{\mathbb{F}}(A)$ is a multiple of d^2*

So, the broad plan is to now

1. Construct a suitable D over our $\mathbb{F}$ of $dim_{\mathbb{F}}(D) = d^2$
2. Convert $A \in \mathcal{B}^{\{d\}}$ to that of the form $B \otimes M_x$ for some $x \in D$
3. Convert $B \otimes_{\mathbb{F}} M_x$ back to an element $A' \in \mathcal{B}^{\{d\}}$
4. Take care that both steps 2 and 3 do not decrease rank.

Step 1 - Constructing the division algebra

There are quite a few details here both mathematical and algorithmic. I'll just show what is being constructed and claim that all this can be carried out efficiently i.e in time $poly(n, d)$. *Add good references and maybe an appendix*

$$\begin{tikzcd}
\mathbb{F} \arrow[r, hook, "\text{Add } \sqrt[d]{1}"] & \mathbb{F}' \arrow[r, hook] & \mathbb{K} := \mathbb{F}'(X) \arrow[r, hook] \arrow[d, hook] & \mathbb{L} := \mathbb{F}'(\sqrt[d]{X}) \arrow[d, hook] \\
 & & \mathbb{K}(Z) \arrow[r, hook] \arrow[d, hook] & \mathbb{L}(Z) \arrow[d, hook] \\
 & & \mathbb{K}(Y) := \mathbb{K}(\sqrt[d]{Z}) \arrow[r, hook] & D := \mathbb{L}(\sqrt[d]{Z})
\end{tikzcd}$$

Thus, D is of degree d^2 over $\mathbb{K}(Z)$ and it is a division algebra follows from Wedderburn's theorem that all finite dimensional associative algebras are $\cong M(n, D)$ for some division algebra D. Thus, D itself has to be of the form $M(d, \mathbb{K}(Z))$ and clearly then, it's dimension is d^2 over $\mathbb{K}(Z)$

The final output is a basis $\Gamma \subset M(d, \mathbb{F}'[X, Y])$ of $M(d, \mathbb{K}(Y))$ such that $\mathsf{span}_{\mathbb{K}(Z)}(\Gamma) = D$. This basis has entries which are polynomials of degree at most d and can be computed in $poly(d)$ time. This is what will allow us to actually use D.

Step 2,3 - Convert to $B \otimes_{\mathbb{F}} D$ and back

Any matrix over $\mathbb{F}$ can be thought of as a matrix over $\mathbb{K}(Y)$. Now, we have a have a basis Γ of $\mathbb{K}(Y)$ and we can thus rewrite the matrix in this basis. As Γ is also a $\mathbb{K}(Z)$ basis for

D we actually get an element of D. But for the form we need, we want the coefficients to come from $\mathbb{F}$ and not $\mathbb{K}(Y)$. This requires a data reduction step which we prove now.

Theorem 11 ([DGIR96])**.** *Let $\mathbb{K}$ be an extension field of $\mathbb{F}$. Say we have a matrix space with basis $B_i \in M(n, \mathbb{F})$. Given $A' = \sum_i a_i B_i \;\; a_i \in \mathbb{K} \; i \in [m]$ of rank r and $S \subset \mathbb{F}$ such that $|S| > r$, we can construct $A = \sum_i c_i B_i \;\; c_i \in S$. This uses poly$(n, r)$ rank computations over matrices $\sum_i d_i B_i \;\; d_i \in S \cup \{a_i\}$*

Proof. Let $A'_j = \sum_{i \neq j} a_i B_i + x_j B_j$. As A' has rank r, there exists a $r \times r$ minor with non-zero determinant. This is now a degree r polynomial in x_j. Thus, it has at most r roots in S. Say $c_j \in S$ is not and substitute $x_j \to c_j$. Now, $\mathsf{rk}(A') = \mathsf{rk}(A'_j)$. Repeat for all j. The algorithm then, is to compute $\mathsf{rk}(A'_j)$ $A'_j = a_i B_i + cB_j$ for all $j \in [m]$ $c \in S$. The number of rank computations needed is thus $m(r + 1)$ □

Note - There are subtleties in performing computations over the extension field $\mathbb{F}'$. In fact, we can't even construct it explicitly. To, get around this we construct a $R = \mathbb{F}'^{\oplus k}$ i.e. R is k copies of $\mathbb{F}'$. Now, as R is no longer a field, computing rank over $\mathbb{F}'$ becomes slighly trickier. See Section 4.4 in [IQS17] for these details.

Let now see the overall picture with yet another diagram (Yes, I love diagrams!).

$$\begin{tikzcd}
\mathcal{B}^{\{d\}} = \mathcal{B} \otimes_{\mathbb{F}} M(d, \mathbb{F}) \arrow[r, hook, "\text{Write in } \Gamma \text{ basis}"] \arrow[dr, phantom, "\circlearrowright"] & \mathcal{B} \otimes_{\mathbb{K}(Y)} D \arrow[d, hook, "\text{Data reduction}"] \\
\mathcal{B} \otimes_{\mathbb{K}(Y)} M(d, \mathbb{F}) \arrow[u, hook, "\text{Data reduction}"] & \mathcal{B} \otimes_{\mathbb{F}} D \arrow[l, "\text{Write in standard basis}"']
\end{tikzcd}$$

Writing all this down more explicitly, we have,

$$\begin{tikzcd}
A = \sum_{i \in [m], j \in [d^2]} \alpha_{ij} B_i \otimes E_j \arrow[r, hook, "\text{Write in } \Gamma \text{ basis}"] & A = \sum_{i \in [m], j \in [d^2]} \lambda_{ij} B_i \otimes C_j, \;\; \lambda_{ij} \in \mathbb{F}(X, Y) \arrow[d, hook, "\text{Data reduction}"] \\
A' = \sum_{i \in [m], j \in [d^2]} \beta_{ij} B_i \otimes E_j, \;\; \beta_{ij} \in \mathbb{F}(X, Y) \arrow[d, hook, "\text{Data reduction}"'] & A' = \sum_{i \in [m], j \in [d^2]} \mu_{ij} B_i \otimes C_j, \;\; \mu_{ij} \in \mathbb{F} \arrow[l, "\text{Write in standard basis}"'] \\
A'' = \sum_{i \in [m], j \in [d^2]} \delta_{ij} B_i \otimes E_j, \;\; \delta_{ij} \in \mathbb{F} &
\end{tikzcd}$$

For all this to work, we have the following chain of dependencies,

Lemma 12. $rd < \mathsf{rk}_{\mathbb{F}}(A) = \mathsf{rk}_{\mathbb{K}(Y)}(A) \leqslant \mathsf{rk}_{\mathbb{K}(Y)}(A') = \frac{\mathsf{rk}_{\mathbb{K}(Z)}(A')}{d} = (r + 1) \cdot d = \mathsf{rk}_{\mathbb{F}}(A'')$

Proof. 1. This is part of the input condition

2. Clearly, $\mathsf{rk}_{\mathbb{F}}(A) \geqslant \mathsf{rk}_{\mathbb{K}(Y)}(A)$ as a dependency $\sum_i a_i C_i \; a_i \in \mathbb{F}$ where $C_i \in \mathbb{F}^{nd}$ are the column vectors, is also a dependency over $\mathbb{K}(Y)$. For the other direction, say we have $\sum_i a_i C_i \; a_i \in \mathbb{K}(Y)$. We can clear denominators by multiplying by their product. Now the $a_i \in \mathbb{F}'[X, Z, \sqrt[d]{Z}]$ are of the form $p_0(x, z) + \cdots p_{d-1}(x, z)y^d$. We can now isolate any non-zero monomial in x, y, z and by requiring it's coefficient to be zero get a $\sum_i b_i C_i \; b_i \in \mathbb{F}$. Note that this works as the elements of $C_i \in \mathbb{F}$

3. This holds due to the the data reduction lemma.

4. As D is a division algebra of dimension d^2 over $\mathbb{K}(Z)$, by lemma 10 we have that $\mathsf{rk}_{\mathbb{K}(Z)}$ is a multiple of d^2 and thus $\mathsf{rk}_{\mathbb{K}(Y)}$ is a multiple of d .

5. Again follows from the data reduction lemma.

□

Blowup Control

This is the part that was introduced in [IQS18] and is a key ingredient that make the algorithm *efficient* i.e. polymonial time. To see this, say, we did not have this step. Assume we started with $A \in \mathcal{B}$ of rank r. Now in each iteration we increase rank by 1 and the blowup by a multiplicative factor of $r + 1$ (or $r + 2$). Thus, if $\mathsf{ncrk}(\mathcal{B}) = n$, then the iteration would have run for $n - r$ steps and the final blowup would be by about $(r + 1) \cdots (n + 1) = \frac{(n+1)!}{r!}$. This is problematic as our subroutines take time $poly(n, d)$ and we need d to be polynomially bounded by n.

A polynomial bound on d was first shown by [DM17] but they gave a nonconstructive bound which was constructivized in [IQS18]. We will not discuss this and the interested reader may consult the original text. The paper though gives another simpler method which we will use.

Lemma 13. *Let $A \in \mathcal{B}^{\{d\}}$ be of rank dn , $d > n + 1$. Then we can compute in deterministic $poly(d, n)$ time $A' \in \mathcal{B}^{\{d-1\}}$ of rank $(d - 1)n$*

Proof. Let A'' be any $(d - 1)n(d - 1)n$ submatrix of A. As there are n rows and columns removed from A. Then, $\mathsf{rk}(A'') \geqslant \mathsf{rk}(A) - 2n = (d - 2)n = (d - 2)(n - 1) + d - 2 > (d - 2)(n - 1) + n - 1 = (d - 2)(n - 1)$ Using regularity lemma we can round up the rank to obtain $A' \in \mathcal{B}^{\{d-1\}}$ of rank $(d - 1)n$ □

At any stage the blowup occurs from d to $dd' \leqslant dn$, Thus, we might need to repeat this step at most n times in each iteration. This is a polynomial increase but it ensures that each step occurs in $poly(n)$ time.

References

[AS78] M. D. Atkinson and N. M. Stephens. Spaces of Matrices of Bounded Rank. *The Quarterly Journal of Mathematics*, 29(2):221–223, 06 1978.

[BD06] Matthias Bürgin and Jan Draisma. The hilbert null-cone on tuples of matrices and bilinear forms. *Mathematische Zeitschrift*, 254(4):785–809, Dec 2006.

[Clns77] P.M. Cohn and London Mathematical Society lecture note series. *Skew Field Constructions*. Lecture note series / London Mathematical Society. Cambridge University Press, 1977.

[Coh95] P. M. Cohn. *Skew Fields: Theory of General Division Rings*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1995.

[DGIR96] Willem De Graaf, Gábor Ivanyos, and Lajos Rónyai. Computing cartan subalgebras of lie algebras. *Applicable Algebra in Engineering, Communication and Computing*, 7(5):339–349, Sep 1996.

[DM16] Harm Derksen and Visu Makam. On non-commutative rank and tensor rank. *CoRR*, abs/1606.06701, 2016.

[DM17] Harm Derksen and Visu Makam. Polynomial degree bounds for matrix semi-invariants. *Advances in Mathematics*, 310:44 – 63, 2017.

[EH88] David Eisenbud and Joe Harris. Vector spaces of matrices of low rank. *Advances in Mathematics*, 70(2):135 – 155, 1988.

[FL04] Marc Fortin and Alain Lascoux. Commutative/noncommutative rank of linear matrices and subspaces of matrices of low rank. 2004.

[GdO18] Ankit Garg and Rafael Mendes de Oliveira. Recent progress on scaling algorithms and applications. *CoRR*, abs/1808.09669, 2018.

[GGdOW15] Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. A deterministic polynomial time algorithm for non-commutative rational identity testing. *CoRR*, abs/1511.03730, 2015.

[IKQS15] Gábor Ivanyos, Marek Karpinski, Youming Qiao, and Miklos Santha. Generalized wong sequences and their applications to edmonds' problems. *J. Comput. Syst. Sci.*, 81(7):1373–1386, 2015.

[IQS17] Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Non-commutative edmonds' problem and matrix semi-invariants. *Computational Complexity*, 26(3):717–763, 2017.

[IQS18] Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Constructive non-commutative rank computation is in deterministic polynomial time. *Computational Complexity*, 27(4):561–593, 2018.

[Lam] Tsit-Yuen Lam. *A First Course in Noncommutative Rings*, volume 131 of *Graduate Texts in Mathematics*. Springer-Verlag New York.

Appendix

Skew Field Shenanigans

This is tricky to construct due to the noncommutativity. Let's review how do we construct it in the commutative case. We start with the ring of polynomials $\mathbb{F}[x_1, \cdots, x_n]$ and then invert all non-zero elements by what is called *localization*. Formally, $\mathbb{F}(x) = \{(p, q) \mid p, q \in \mathbb{F}[x], q \neq 0\}$ with an equivalence relation $(p, q) \sim (f, g) \iff pg = fq$. p/q then denotes the equivalence class of (p, q). Addition is formally defined as $p/q + f/g = (pg + fq)/qg$. and multiplication is $p/q \cdot f/g = (pf)/(qg)$

Now, one can see the problems if variables don't commute. $1/x + 1/y = (x + y)/xy$ but $1/y + 1/x = (x + y)/(yx)$ which are not equal anymore! To remedy this we require the ring R to satisfy the either the left or right *Ore condition.* The (right) Ore condition says that $\forall a, b \in R\ \exists a_b, b_a \in R\backslash\{0\} \quad aa_b = bb_a$. Note that in the commutative world this is true as setting $a_b = b$, $b_a = a$ works. If this holds, we can define, $(a, b) \sim (c, d \iff ab_d = cd_b)$ and $a/b + c/d = (ab_d + cd_b)/bb_d$. Also define, $a/b * c/d = (ab_c)/dc_b$

First let's note that the equivalence makes sense i.e. $(a, b) \sim (ax, bx) \forall\ x \in R$. This is true as $(a, b) \sim (ax, bx) \iff ab_{bx} = (ax)(bx)_b \iff b_{bx} = (x)(bx)_b \iff bb_{bx} = (bx)(bx)_b$ The second implication is true as R is an integral domain and the last holds by definition.

Lemma 14. *Addition is well-defined, commutative, has an inverse and an identity.*

Proof. Firstly, $a/b + 0/1 = (ab_1)/(11_b) = (ab_1)/(bb_1) = a/b$. Now, $(a, b) \sim (p, q) \implies ab_q = pq_b$. $a/b + (-p)/q = (ab_q - pq_b)/bb_q = 0/1$. Commutativity is easy as,
$c/d + a/b = (cd_b + ab_d)/(dd_b) = (ab_d + cd_b)/bb_d$ ($dd_b = bb_d$ by definition) □

Lemma 15. *Multiplication is well-defined, has an inverse and an identity.*

Proof. Firstly, $a/b * 1/1 = (ab_1)/(11_b) = (ab_1)/(bb_1) = a/b$.
Now, $a/b * b/a = (ab_a)/(ab_a) = 1/1$ □

Thus, if R obeys the (right) Ore condition the field of fractions is well-defined. Similar concstruction can be done for the *left* case.

We can start similarly by creating polynomials in non-commuting variables, denoted as $\mathbb{F}\langle x_1, \cdots, x_m\rangle$. The usual way to construct it is by creating a free algebra (words in the variables x_i) i.e we impose that $a \cdot x_i = x_i \cdot a\ \forall a \in \mathbb{F},\ i \in [m]$.

This however, does not satisfy either left or right Ore condition as $\nexists\ a, b$ such that $xa = yb$ Thus, we start not with the free algebra but another ring in which $a \cdot x_i = x_i \cdot a + \overline{a}$. This works and we get a field of fractions. This is not unique as taking the left construction gives us another one. Buut these can be embedded in a universal one which is unique (upto isomorphism) denoted as $\mathbb{F}\langle\!\langle x_1, \cdots, x_m\rangle\!\rangle$. The first chapter in [Clns77] discusses this very well.

Invariant Theory connection

Consider the left-right action of $SL_n(\mathbb{F}) \times SL_n(\mathbb{F})$ on the tuple $(B_1, \cdots B_m)$ i.e (A, C) acts as $(AB_1C^T, \cdots AB_mC^T)$. A polynomial $p \in \mathbb{F}[x_1, \cdots x_{mn^2}]$ is said to be *invariant* if $p(B_1, \cdots B_m) = p(AB_1C^T, \cdots AB_mC^T)\ \forall A, C \in SL_n(\mathbb{F}) \times SL_n(\mathbb{F})$. [2]

Polynomials of the form, $d_M = \mathsf{det}(\sum_i M_i \otimes B_i)$ where $M_i \in M(d, \mathbb{F})$, are invariant as determinant is multiplicative and $\mathsf{det}(A) = \mathsf{det}(C) = 1$. But these are just some of the invariants what about others?

Clearly, the set of all invariant polynomials forms a ring denoted as $R(n, m)$. Hilbert in a landmark result proved that this ring is finitely generated (as an $\mathbb{F}$- algebra), i.e there is a finite set of polynomials $\{p_i \ \ i \in I\}$ such that every invariant polynomial can be written as a polynomial in these i.e. $f = F(p_i)$. Since the set is finite, define $\beta(n, m) = \max\limits_{i \in I}\ deg(p_i)$.

Well, for the above action it turns out that all generators are of this form,

Theorem 16 (First Fundamental Theorem)**.** *The set of generators* $\{p_i \ \ i \in I\} \subset \{d_M | M \in M(n, \mathbb{F})^{\oplus m}\}$

That's great. We have some connection, now that is, the matrices of non-full rank definitely vanish on all the polynomials in $R(n, m)$ and thus we can check only on the finite set of generators. Define $\beta(n, m) = \max\limits_{i \in I}\ deg(p_i)$. This is the bound N that we needed. But the crucial question still remains. Can it be that $p(B_1, \cdots B_m) = 0\ \ \forall i$ but $\mathsf{ncrk}(\mathcal{B}) = n$? No.

Theorem 17. *[BD06] Let* $\mathcal{N}(n, m) = \{(B_1, \cdots B_m) \mid p_i(B_1, \cdots B_m) = 0\ \forall i \in I\}$. *Then,* $(B_1, \cdots B_m) \in \mathcal{N}(n, m) \iff \mathsf{ncrk}(\mathcal{B}) < n$

$\mathcal{N}(n, m)$ is called the *nullcone* with respect to this action and the problem of checking if noncommutative rank is full is equivalent to checking membership in the nullcone. This nullcone membership problem can be studied in more generality for a general group and its actions on any vector space.

Another line of very interesting work originated with [GGdOW15] giving analytic algorithms for these nullcone questions. The algorithm in [GGdOW15] solves this nullcone problem but its analysis uses just the fact that the degree bound is finite and doesn't actually need its value for the proof. Read the beautifully written survey [GdO18] for details on these works.

[2] ***Note*** - As a shorthand, we usually write the polynomial $p(v)$ and not $p(e_1, \cdots, e_n)$. For example, we write $p(B_1, B_2) = det(B_1 + B_2)$ but this is a polynomial in $2n^2$ and not 2 variables.