Macaulay matrices and Gröbner bases

Ben Clingenpeel

6/15/22

Let $\mathcal{F} = \{f_1, \dots, f_s\}$ be a collection of elements of the polynomial ring $k[x_1, \dots, x_n]$ for a field k, and fix a term order $<$. For a given degree $d \geq 0$, we let $M_{\leq d}$ denote the Macaulay matrix of the system $\mathcal{F}$, as defined in [CG20]. Then if the matrix A is the Macaulay matrix of some collection $\mathcal{B}$ of polynomials, we define $\mathscr{P}(A)$ to be the collection of polynomials corresponding to the nonzero rows of A. If $\{a_{ij}\}$ is some set of entries of the matrix A, we define $\mathscr{P}(\{a_{ij}\})$ to be the set of monomials that index the columns i.

Example 1. Let $f_1 = x$ and $f_2 = x^2 - y$. Then the degree $d = 2$ Macaulay matrix of $\mathcal{F} = \{f_1, f_2\}$ with respect to the degree reverse lexicographic (DRL) order is

$$M_{\leq 2} = \begin{array}{c} \\ f_1 \\ xf_1 \\ yf_1 \\ f_2 \end{array} \begin{array}{c} \begin{array}{cccccc} x^2 & xy & y^2 & x & y & 1 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 & \mathbf{-1} & 0 \end{pmatrix} \end{array}.$$

Then we have that $\mathscr{P}(M_{\leq 2}) = \{f_1, xf_1, yf_1, f_2\} = \{x, x^2, xy, x^2 - y\}$, and we have that $\mathscr{P}(\{a_{41}, a_{12}, a_{23}, a_{14}\}) = \{x, x^2, xy, x\} = \{x^2, xy, x\}$.

Throughout, given a matrix A, we will denote the Reduced Row Echelon Form (RREF) of A by A^r.

Example 2. Using $M_{\leq 2}$ above, we have that

$$M_{\leq 2}^r = \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 \end{pmatrix}.$$

Then $\mathscr{P}(M_{\leq 2}^r) = \{x^2, xy, x, y\}$.

The following Lemma says that the ideal generated by $\mathscr{P}(M_{\leq d})$, the set of polynomials in the Macaulay matrix, is the same as the ideal generated by $\mathscr{P}(M_{\leq d}^r)$, the set of polynomials from the row reduced Macaulay matrix.

Lemma 1. Let M be a Macaulay matrix of some collection of polynomials $\mathcal{F}$ and denote the ideal it generates by I. Then for $d \geq 0$, we have that $(\mathscr{P}(M_{\leq d}^r)) = (\mathscr{P}(M_{\leq d}))$.

Proof. Recall that the elementary row operations performed on a matrix during Gaussian elimination do not change its row space. Since $\text{row}(M_{\leq d}^r) = \text{row}(M_{\leq d})$, any polynomial $f \in \mathscr{P}(M_{\leq d}^r)$ is a k-linear combination of the elements of $\mathscr{P}(M_{\leq d})$, where k is the field over which the polynomial ring is defined. Since ideals in $k[x_1, \ldots, x_n]$ are closed under linear combinations of elements, we have that $\mathscr{P}(M_{\leq d})$ and $\mathscr{P}(M_{\leq d}^r)$ generate the same ideal: any $f \in \mathscr{P}(M_{\leq d}^r)$ can be written as $a_1 p_1 + \cdots + a_m p_m$ for some $a_i \in k$ and $p_i \in \mathscr{P}(M_{\leq d})$, meaning $f \in (\mathscr{P}(M_{\leq d}))$, the ideal generated by $\mathscr{P}(M_{\leq d})$. Since the generators of $(\mathscr{P}(M_{\leq d}^r))$ are all contained in $(\mathscr{P}(M_{\leq d}))$, we have that $(\mathscr{P}(M_{\leq d}^r)) \subset (\mathscr{P}(M_{\leq d}))$, and a similar argument gives the other containment. □

Corollary 1. For an ideal $I = (\mathcal{F})$, there exists a $D \geq 0$ such that for all $d \geq D$, $(\mathscr{P}(M_{\leq d}^r)) = I$.

Proof. Let $\mathcal{F}_{\leq d} = \{f \in \mathcal{F} \mid \deg(f) \leq d\}$ be the set of polynomials in $\mathcal{F}$ with degree less than or equal to d. Then $\mathscr{P}(M_{\leq d})$ is the set of polynomials in the corresponding Macaulay matrix, and so consist of the elements of $\mathcal{F}_{\leq d}$, along with monomial multiples of the elements of $\mathcal{F}_{\leq d}$. Since these multiples are already in the ideal $(\mathcal{F}_{\leq d})$ (because ideals are closed under multiplication by ring elements), we have that $(\mathscr{P}(M_{\leq d})) = (\mathcal{F}_{\leq d})$.

Now choose the degree D to be $D = \max\{\deg(f) \mid f \in \mathcal{F}\}$ so that $\mathcal{F}_{\leq d} = \mathcal{F}$ for all $d \geq D$. Then we apply Lemma 1 to say that if $d \geq D$, then

$$(\mathscr{P}(M_{\leq d}^r)) = (\mathscr{P}(M_{\leq d})) = (\mathcal{F}_{\leq d}) = (\mathcal{F}) = I,$$

as required. □

Definition 1. For an ideal $I = (\mathcal{F})$, we define the **initial ideal of** I to be the ideal $\text{in}(I) = (\{\text{in}(f) \mid f \in I \setminus \{0\}\})$, and we define $\text{in}(\mathcal{F}) = (\{\text{in}(f_i) \mid f_i \in \mathcal{F}\})$ so that $\text{in}(I)$ is the ideal generated by the initial terms of all elements of I and $\text{in}(\mathcal{F})$ is the ideal generated by the initial terms of elements of the generating set $\mathcal{F}$. If $\text{in}(I) = \text{in}(\mathcal{F})$, we say that $\mathcal{F}$ is a **Gröbner basis for** I.

Definition 2. Let $\text{pivots}(A)$ be the set of pivot positions of the matrix A. Then define the monomial ideal $\text{in}(d)$ by $\text{in}(d) = (\mathscr{P}(\text{pivots}(M_{\leq d})))$. That is, $\text{in}(d)$ is the ideal generated by the monomials corresponding to the pivot columns of the Macaulay matrix.

Remark 1. In the definition above, note that the pivot columns of the Macaulay matrix are those columns of $M^r_{\leq d}$ that contain a leading entry 1 of a some row. Since these rows correspond to the polynomials in $\mathscr{P}(M^r_{\leq d})$, the monomials that index these pivot columns therefore correspond exactly to the initial terms of the polynomials in $\mathscr{P}(M^r_{\leq d})$, and so $\text{in}(d) = \text{in}(\mathscr{P}(M^r_{\leq d}))$ as in Definition 1, substituting $\mathscr{P}(M^r_{\leq d})$ for $\mathcal{F}$.

Example 3. Using the DRL ordering, let $\mathcal{F} = \{f_1, f_2\}$ with $f_1 = x^2 - 1$ and $f_2 = xy + x$, as in Example 8 of [CG20]. Then we have that $\text{in}(\mathcal{F}) = (x^2, xy)$, but because

$$y + 1 = -x^2y + y - x^2 + 1 + x^2y + x^2 = -(y+1)f_1 + xf_2 \in I$$

implies that $y \in \text{in}(I) \setminus \text{in}(\mathcal{F})$, we have that $\text{in}(\mathcal{F}) \neq \text{in}(I)$, and therefore $\mathcal{F}$ is not a Gröbner basis for I.

Continuing this example, we can compute that

$$M_{\leq 2} = \begin{array}{c} \\ f_1 \\ f_2 \end{array} \begin{array}{c} \begin{array}{cccccc} x^2 & xy & y^2 & x & y & 1 \end{array} \\ \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & \mathbf{-1} \\ 0 & \mathbf{1} & 0 & \mathbf{1} & 0 & 0 \end{pmatrix} \end{array} = M^r_{\leq 2},$$

which is already reduced. Therefore $\text{pivots}(M_{\leq 2}) = \{a_{11}, a_{22}\}$, and so

$$\text{in}(2) = (\mathscr{P}(\text{pivots}(M_{\leq 2}))) = (\mathscr{P}(\{a_{11}, a_{22}\})) = (x^2, xy),$$

which we see does indeed equal the ideal generated by the initial terms of the polynomials

in $\mathscr{P}(M^r_{\leq 2}) = \{x^2 - 1, xy + x\}$.

By the above, we know then that $\mathscr{P}(M^r_{\leq 2})$ is not a Gröbner basis for I, but if instead we do the same computations with degree $d = 3$, we have the following:

$$M_{\leq 3} = \begin{array}{c} \\ f_1 \\ f_2 \\ xf_1 \\ yf_1 \\ xf_2 \\ yf_2 \end{array} \begin{array}{c} \begin{array}{cccccccccc} x^3 & x^2y & xy^2 & y^3 & x^2 & xy & y^2 & x & y & 1 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & \mathbf{-1} \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{1} & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{-1} & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{-1} & 0 \\ 0 & \mathbf{1} & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 \end{pmatrix} \end{array},$$

and bringing this into RREF, we have

$$M^r_{\leq 3} = \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{-1} & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & \mathbf{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & \mathbf{-1} \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & \mathbf{1} \end{pmatrix},$$

so that $\text{in}(3) = (x^3, x^2y, xy^2, x^2, xy, y) = (x^2, y) = \text{in}(I)$. Therefore $\text{in}(\mathscr{P}(M^r_{\leq 3})) = \text{in}(3) = \text{in}(I)$, and since the maximal degree of the original set of polynomials $\mathcal{F} = \{f_1, f_2\}$ is 2, the proof of Corollary 1 implies that $(\mathscr{P}(M^r_{\leq 3})) = I$. Therefore

$$\mathscr{P}(M^r_{\leq 3}) = \{x^3 - x, x^2y + 1, xy^2 - x, x^2 - 1, xy + x, y + 1\}$$

is a Gröbner basis for I.

In the above example, a Gröbner basis was found by computing $\mathscr{P}(M^r_{\leq d})$ up to degree $d = 3$. In general, ideals may require computing $\mathscr{P}(M^r_{\leq d})$ for degrees much larger than $d = 3$, but we claim that for any ideal $I = (\mathcal{F})$ there is always some choice of degree $d \geq 0$ for which $\mathscr{P}(M^r_{\leq d})$ is a Gröbner basis for I. We will prove this Theorem after the following Lemma, generalizing the observation that in Example 3, in(2) was contained in in(3).

Lemma 2. Given an ideal $I = (\mathcal{F})$ and degrees d_1 and d_2 with $d_1 \leq d_2$, we have that $\text{in}(d_1) \subset \text{in}(d_2)$.

Proof. Since $\text{in}(d_1)$ and $\text{in}(d_2)$ are monomial ideals, it is a Corollary of Lemma 3 of section 2.4 of [CLO13], that $\text{in}(d_1) \subset \text{in}(d_2)$ if and only if all monomials of $\text{in}(d_1)$ lie in $\text{in}(d_2)$. Therefore let $m \in \text{in}(d_1)$ be a monomial. Lemma 2 of the same section in [CLO13] says that $m \in \text{in}(d_1) = (\mathscr{P}(\text{pivots}(M_{\leq d_1})))$, means that there is some generating monomial $m^* \in \mathscr{P}(\text{pivots}(M_{\leq d_1}))$ that divides m. Therefore, if we can show that $m^* \in \text{in}(d_2)$, we will have that $m \in \text{in}(d_2)$. To this end, note that $m^* \in \mathscr{P}(\text{pivots}(M_{\leq d_1}))$ means that the column of $M_{\leq d_1}$ indexed by m^* is a pivot column, so the corresponding column in $M^r_{\leq d_1}$ is all zeros except for a 1 in a row that corresponds to a polynomial p with $\text{in}(p) = m^*$. This row corresponding to p is in the row space $\text{row}(M^r_{\leq d_1}) = \text{row}(M_{\leq d_1})$, and so p can be written as a k-linear combination of the polynomials in $\mathscr{P}(M_{\leq d_1})$. Note that the polynomials involved in this k-linear combination all have degree less than or equal to d_1, which is in turn less than or equal to d_2. Therefore these polynomials also correspond to rows of $M_{\leq d_2}$, and we have that p is a k-linear combination of the elements of $\mathscr{P}(M_{\leq d_2})$.

That p is a k-linear combination of the elements of $\mathscr{P}(M_{\leq d_2})$ means that the row vector r_p corresponding to p with the same length as the length of the rows of $M_{\leq d_2}$ (that is, the row vector representation of p whose entries are indexed by monomials of degree less than or equal to d_2) is in the row space $\text{row}(M_{\leq d_2}) = \text{row}(M^r_{\leq d_2})$, and so this row is a k-linear combination of the nonzero rows of $M^r_{\leq d_2}$, say $r_p = c_1 r_1 + \cdots + c_t r_t$ where $r_1, \ldots, r_t$ are the rows of $M^r_{\leq d_2}$. Recall that $m^* = \text{in}(p)$, and so the entry of r_p indexed by m^* is 1. Because polynomials with initial term less than m^* (with respect to the chosen term ordering) cannot form a k-linear combination with initial term m^*, we see that p must be a k-linear combination of polynomials, some of which must have initial terms at least m^*. Because $M^r_{\leq d_2}$ is in RREF and each nonzero row contains a pivot position, there exists some $1 \leq j < t$ such that the pivots of r_i for $i = 1, \ldots, j$ are all in columns indexed by monomials greater than m^* and the pivots of r_i for $i > j$ are all in columns indexed by monomials no larger than m^*. Then the condition that p is a k-linear combination of polynomials, some of which must have initial terms at least m^* means that r_p is a k-linear combination of the rows r_i, some of which must have pivot positions in columns at least as far to the left as the column corresponding to m^*. However, if $c_i \neq 0$ for some $1 \leq i \leq j$ in the combination $r_p = c_1 r_1 + \cdots + c_t r_t$, then the leading entry of r_p would be the entry c_i in a column corresponding to a monomial greater than m^* since all rows other than r_i contain zeros in this column, and this would mean that the initial term of p would be $c_i m'$ for some $m' > m^*$, a contradiction. Therefore $c_i = 0$ for

all $i = 1, \ldots, j$. But since r_p must be a k-linear combination of rows, some of which must have pivot positions at least as far to the left as the column corresponding to m^*, the only option is that the combination involves some row with a pivot position $a \in \text{pivots}(M_{\leq d_2})$ in the column corresponding to m^*. Therefore this column is a pivot column of $M_{\leq d_2}$, and so $m^* \in \mathscr{P}(\{a\}) \subset \mathscr{P}(\text{pivots}(M_{\leq d_2}))$.

We have shown that if $m \in \text{in}(d_1)$, then there is a monomial in $m^* \in \mathscr{P}(\text{pivots}(M_{\leq d_1}))$ that divides m, and that $m^* \in \mathscr{P}(\text{pivots}(M_{\leq d_2}))$ as well, meaning

$$m \in (m^*) \subset (\mathscr{P}(\text{pivots}(M_{\leq d_2}))) = \text{in}(d_2).$$

Therefore $\text{in}(d_1) \subset \text{in}(d_2)$, as claimed. □

Theorem 1. For an ideal $I = (\mathcal{F})$, there exists a $D \geq 0$ such that $\mathscr{P}(M^r_{\leq D})$ is a Gröbner basis of I.

Proof. By Lemma 2, we have that ascending chain of ideals $\text{in}(0) \subset \text{in}(1) \subset \text{in}(2) \subset \ldots$ in the Noetherian ring $k[x_1, \ldots, x_n]$, and so there exists a $D' \geq 0$ such that for all $d \geq D'$, $\text{in}(d) = \text{in}(D')$. By Corollary 1, there exists some $D^* \geq 0$ such that $(\mathscr{P}(M^r_{\leq d})) = I$ for all $d \geq D^*$. Set $D = \max\{D', D^*\}$. Then for all $d \geq D$, we have that $(\mathscr{P}(M^r_{\leq d})) = I$ and that $\text{in}(d) = \text{in}(D)$. By Remark 1, $\text{in}(D)$ is the set $\text{in}(\mathscr{P}(M^r_{\leq D}))$, and so to show that $\mathscr{P}(M^r_{\leq D})$ is a Gröbner basis for I, it suffices to show that $\text{in}(D) = \text{in}(I)$. Because the ideal $(\mathscr{P}(M^r_{\leq D}))$ is the ideal I, all initial terms of polynomials in $\mathscr{P}(M^r_{\leq D})$ are initial terms of polynomials in I, and so $\text{in}(D) = \text{in}(\mathscr{P}(M^r_{\leq D})) \subset \text{in}(I)$.

To show the other containment, recall from the proof of Lemma 2 that $\text{in}(D)$ and $\text{in}(I)$ being monomial ideals means it is sufficient to show that all monomials $m \in \text{in}(I)$ are in $\text{in}(D)$ in order to show that $\text{in}(I) \subset \text{in}(D)$. Therefore let $m \in \text{in}(I)$ be an arbitrary monomial. Then there exists some $p \in I$ with initial term $\text{in}(p) = m^*$ such that m^* divides m. Then using Lemma 1, $p \in I$ means

$$p \in I = (\mathscr{P}(M^r_{\leq D})) = (\mathscr{P}(M_{\leq D})),$$

so p can be written as $p = \sum_{q \in \mathscr{P}(M_{\leq D})} h_q q$ where $h_q \in k[x_1, \ldots, x_n]$. Then because each h_q is a k-linear combination of monomials, we have that $h_q = \sum_{\alpha \in N_q} a_\alpha x^\alpha$ for some finite index set N_q. Therefore

$$p = \sum_{q \in \mathscr{P}(M_{\leq D})} \sum_{\alpha \in N_q} a_\alpha x^\alpha q.$$

Since $\mathscr{P}(M_{\leq D})$ is finite and each index set N_q is finite, we may set

$$d^* = \max \left\{ \deg(x^\alpha q) \,\middle|\, \alpha \in \bigcup_{q \in \mathscr{P}(M_{\leq D})} N_q,\ q \in \mathscr{P}(M_{\leq D}) \right\} \geq D.$$

Then because each q is a monomial times one of the polynomials $f \in \mathcal{F}$, so is each each $x^\alpha q$, and therefore p as written above is a k-linear combination of polynomials corresponding to rows in the Macaulay matrix $M_{\leq d^*}$. Therefore if r_p is the row vector corresponding to p whose entries are indexed by monomials of degree less than or equal to d^*, we have that r_p is an element of the row space $\text{row}(M_{\leq d^*}) = \text{row}(M^r_{\leq d^*})$. Therefore we have the same situation as in Lemma 2: we have a polynomial p with leading term m^* satisfying $\deg(m^*) \leq d^*$ such that r_p is a k-linear combination of the nonzero rows of $M^r_{\leq d^*}$. As in the proof of Lemma 2, then, we have that the column indexed by m^* is a pivot column of $M_{\leq d^*}$, which means that $m^* \in \mathscr{P}(\text{pivots}(M_{\leq d^*})) = \text{in}(d^*)$. Because $d^* \geq D$, we also have that $m^* \in \text{in}(d^*) = \text{in}(D)$ since the ascending chain $\text{in}(0) \subset \text{in}(1) \subset \text{in}(2) \subset \ldots$ stabilizes at or before D. Because m is divisible by $m^* \in \text{in}(D)$, we have that $m \in \text{in}(D)$, and so $\text{in}(I) \subset \text{in}(D)$. Therefore $\text{in}(D) = \text{in}(I)$. Since $I = (\mathscr{P}(M^r_{\leq D}))$ and $\text{in}(I) = \text{in}(D) = \text{in}(\mathscr{P}(M^r_{\leq D}))$, the set $\mathscr{P}(M^r_{\leq D})$ is a Gröbner basis for I. $\square$

References

[CG20] Alessio Caminata and Elisa Gorla. Solving multivariate polynomial systems and an invariant from commutative algebra. In *International Workshop on the Arithmetic of Finite Fields*, pages 3–36. Springer, 2020.

[CLO13] David Cox, John Little, and Donal OShea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer Science & Business Media, 2013.