A constructive proof of Orzech's theorem

Darij Grinberg

version 1.4, 20 November 2016

The purpose of this note is to prove Morris Orzech's theorem on surjective homomorphisms of modules [1, Theorem 1] within constructive mathematics. Our main weapon will be the Cayley-Hamilton theorem.

The LaTeX sourcecode of this note contains additional details of proofs inside "verlong" environments (i. e., between "`\begin{verlong}`" and "`\end{verlong}`"). I doubt they are of any use.

***

Let us begin by stating the theorem:

Theorem 0.1. Let A be a commutative ring with unity. Let M be a finitely generated A-module. Let N be an A-submodule of M, and let $f : N \to M$ be a surjective A-module homomorphism. Then, f is an A-module isomorphism.

Morris Orzech discovered this result [1, Theorem 1] in 1971. It generalizes the following fact, found formerly by Vasconcelos:

Corollary 0.2. Let A be a commutative ring with unity. Let M be a finitely generated A-module. Let $f : M \to M$ be a surjective A-module endomorphism of A. Then, f is an A-module isomorphism.

Corollary 0.2 is a well-known fact (e.g., it appears in [12, Lemma A.3] and in [3]), but most of its proofs in literature do not generalize to Theorem 0.1.

Orzech's original proof of Theorem 0.1 (with the corrections provided in [2], as the original version was shaky) proceeds by reducing the theorem to the case when A is Noetherian, and then using this Noetherianness in an elegant and yet mysterious way. The proof is not constructive and (to my knowledge) cannot easily be made constructive. In this note, I will present a constructive way to prove Theorem 0.1.

Let us first make some preparations. We let $\mathbb{N} = \{0, 1, 2, \ldots\}$. We fix a commutative ring A with unity. For every $n \in \mathbb{N}$, let I_n denote the identity $n \times n$-matrix in $A^{n \times n}$. We reserve a fresh symbol X as an indeterminate for polynomials. We embed A into the polynomial ring $A[X]$ canonically, and we use this to embed the matrix ring $A^{n \times n}$ into $(A[X])^{n \times n}$ canonically for every $n \in \mathbb{N}$. For every $n \in \mathbb{N}$ and any square matrix $M \in A^{n \times n}$, we define the *characteristic polynomial* χ_M of M as the polynomial $\det(X \cdot I_n - M)$. (This is one of the two common ways to define a characteristic polynomial of a matrix M. The other way is to define it as $\det(M - X \cdot I_n)$. These two definitions result in two polynomials which differ only by multiplication by $(-1)^n$.) The famous *Cayley-Hamilton theorem* states the following:

Theorem 0.3. Let $n \in \mathbb{N}$. Let A be a commutative ring with unity. Let $M \in A^{n \times n}$. Then, $\chi_M(M) = 0$. (In words: Substituting the matrix M for X in the characteristic polynomial χ_M of M yields the zero matrix.)

In this exact form, Theorem 0.3 is proven in [8], in [11, Theorem 3.4] and in [5, Theorem 2.5].[1] But there are lots of places where almost complete proofs of Theorem 0.3 can be found. For example, Theorem 0.3 is proven in most standard texts on linear algebra in the case when A is a field. Some of these proofs (e.g., the proof given in [4, Theorem 7.10], or the proof given in [10, Theorem 5.9], or the proofs given in [9], or Straubing's combinatorial proof given in [6][2] and in [7, §3]) can be straightforwardly generalized to the general case. Even if your favorite proof of Theorem 0.3 in the case when A is a field does not generalize to the general case, it is still easy to derive the general case from the case of A being a field (this is what Conrad does in [11, Theorem 3.4]).

We can obtain the following consequence of Theorem 0.3:

Corollary 0.4. Let $n \in \mathbb{N}$. Let A be a commutative ring with unity. Let $M \in A^{n \times n}$. Then, there exists an $(n+1)$-tuple $(c_0, c_1, \ldots, c_n) \in A^{n+1}$ such that $c_0 M^0 + c_1 M^1 + \cdots + c_n M^n = 0$ and $c_n = 1$.

Proof of Corollary 0.4. It is well-known that the characteristic polynomial χ_M of M is a monic polynomial of degree n over A. In other words, there exists an $(n+1)$-tuple $(c_0, c_1, \ldots, c_n) \in A^{n+1}$ such that $\chi_M = c_0 X^0 + c_1 X^1 + \cdots + c_n X^n$ and $c_n = 1$. Consider this $(c_0, c_1, \ldots, c_n)$. Evaluating both sides of the equality $\chi_M = c_0 X^0 + c_1 X^1 + \cdots + c_n X^n$ at $X = M$, we obtain $\chi_M(M) = c_0 M^0 + c_1 M^1 + \cdots + c_n M^n$. Thus, $c_0 M^0 + c_1 M^1 + \cdots + c_n M^n = \chi_M(M) = 0$ (by Theorem 0.3). This proves Corollary 0.4. □

We can now use Corollary 0.4 to prove the following lemma:

Lemma 0.5. Let $n \in \mathbb{N}$. Let $g : A^n \to A^n$ be an A-linear map. Let V be an A-submodule of A^n such that $g^{-1}(V) \subseteq V$. Then, $g(V) \subseteq V$.

Proof of Lemma 0.5. If $n = 0$, then Lemma 0.5 is obviously true (because in this case, $V \subseteq A^n = A^0 = 0$ and thus $V = 0$). Hence, for the rest of this proof,

[1] Of course, the notations in these sources don't exactly match the notations we are using here. For example, the A, the X and the M in our Theorem 0.3 correspond to the $\mathbb{K}$, the t and the A in [5, Theorem 2.5].

[2] We notice that the two displayed equations right before the Lemma in [6, p. 275] should be corrected to

$$p_A^+(A)_{ij} = \sum_{(\sigma,\pi) \in T_{ij}^+} \mu(\sigma)\mu(\pi), \qquad p_A^-(A)_{ij} = \sum_{(\sigma,\pi) \in T_{ij}^-} \mu(\sigma)\mu(\pi).$$

(To be fair, I do not know if they are wrong in the original printed version of [6] or only in Elsevier's dismal scan of the paper.)

we can WLOG assume that $n \geq 1$. Assume this, and notice that this yields $n-1 \in \{0,1,\ldots,n\}$.

Let $(e_1, e_2, \ldots, e_n)$ be the standard basis of the A-module A^n. (Thus, for every $i \in \{1,2,\ldots,n\}$, the vector e_i is the vector in A^n whose i-th coordinate is 1 and whose other coordinates are all 0.) Let $M \in A^{n \times n}$ be the $n \times n$-matrix which represents the A-linear map $g : A^n \to A^n$ with respect to this basis $(e_1, e_2, \ldots, e_n)$ of A^n. Then,

$$Mw = g(w) \qquad \text{for every } w \in A^n. \tag{1}$$

Corollary 0.4 shows that there exists an $(n+1)$-tuple $(c_0, c_1, \ldots, c_n) \in A^{n+1}$ such that $c_0 M^0 + c_1 M^1 + \cdots + c_n M^n = 0$ and $c_n = 1$. Consider this $(c_0, c_1, \ldots, c_n)$. We have $\sum\limits_{k=0}^{n} c_k M^k = c_0 M^0 + c_1 M^1 + \cdots + c_n M^n = 0$.

We shall now show that every $u \in \{0,1,\ldots,n\}$ satisfies

$$\left(\sum_{k=0}^{n-u} c_{u+k} M^k \right)(V) \subseteq V. \tag{2}$$

Proof of (2): We will prove (2) by induction over u:

Induction base: We have

$$\left(\underbrace{\sum_{k=0}^{n-0} \underbrace{c_{0+k}}_{=c_k} M^k}_{=\sum\limits_{k=0}^{n}} \right)(V) = \underbrace{\left(\sum_{k=0}^{n} c_k M^k \right)}_{=0}(V) = 0\,(V) = 0 \subseteq V.$$

In other words, (2) holds for $u = 0$. This completes the induction base.

Induction step: Let $p \in \{0,1,\ldots,n\}$ be such that $p > 0$. Assume that (2) holds for $u = p-1$. We now must show that (2) holds for $u = p$.

We have assumed that (2) holds for $u = p-1$. In other words,

$$\left(\sum_{k=0}^{n-(p-1)} c_{(p-1)+k} M^k \right)(V) \subseteq V. \tag{3}$$

Now,

$$\underbrace{\sum_{k=0}^{n-(p-1)} c_{(p-1)+k}M^k}_{=\sum_{k=0}^{n-p+1}} = \sum_{k=0}^{n-p+1} c_{(p-1)+k}M^k = c_{(p-1)+0}M^0 + \sum_{k=1}^{n-p+1} c_{(p-1)+k}M^k$$

(here, we have split off the addend for $k = 0$ from the sum)

$$= \underbrace{c_{(p-1)+0}}_{=c_{p-1}} \underbrace{M^0}_{=I_n} + \sum_{k=0}^{n-p} \underbrace{c_{(p-1)+(k+1)}}_{=c_{p+k}} \underbrace{M^{k+1}}_{=MM^k}$$

(here, we have substituted $k+1$ for k in the sum)

$$= c_{p-1}I_n + \underbrace{\sum_{k=0}^{n-p} c_{p+k}MM^k}_{=M\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right)} = c_{p-1}I_n + M\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right). \tag{4}$$

Now, let $v \in V$. Then, applying both sides of the equality (4) to v, we obtain

$$\begin{aligned}\left(\sum_{k=0}^{n-(p-1)} c_{(p-1)+k}M^k\right)(v) &= \left(c_{p-1}I_n + M\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right)\right)v \\ &= c_{p-1}\underbrace{I_n v}_{=v} + M\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right)v \\ &= c_{p-1}v + M\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right)v.\end{aligned}$$

Subtracting $c_{p-1}v$ from this equality, we obtain

$$\left(\sum_{k=0}^{n-(p-1)} c_{(p-1)+k}M^k\right)(v) - c_{p-1}v = M\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right)v = g\left(\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right)v\right)$$

(by (1), applied to $w = \left(\sum_{k=0}^{n-p} c_{p+k}M^k\right)v$). Hence,

$$\begin{aligned}g\left(\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right)v\right) &= \left(\sum_{k=0}^{n-(p-1)} c_{(p-1)+k}M^k\right)\left(\underbrace{v}_{\in V}\right) - c_{p-1}\underbrace{v}_{\in V} \\ &\in \underbrace{\left(\sum_{k=0}^{n-(p-1)} c_{(p-1)+k}M^k\right)(V)}_{\substack{\subseteq V \\ \text{(by (3))}}} - c_{p-1}V \subseteq V - c_{p-1}V \subseteq V\end{aligned}$$

(since V is an A-module). Hence, $\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right) v \in g^{-1}(V) \subseteq V$.

Now, let us forget that we fixed v. We thus have shown that $\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right) v \in V$ for every $v \in V$. In other words, $\left(\sum_{k=0}^{n-p} c_{p+k}M^k\right)(V) \subseteq V$. In other words, (2) holds for $u = p$. This completes the induction step. The induction proof of (2) is thus complete.

Now, let us recall that $n-1 \in \{0,1,\ldots,n\}$. Hence, we can apply (2) to $u = n-1$. As a result, we obtain

$$\left(\sum_{k=0}^{n-(n-1)} c_{(n-1)+k}M^k\right)(V) \subseteq V.$$

Since

$$\underbrace{\sum_{k=0}^{n-(n-1)}}_{=\sum_{k=0}^{1}} c_{(n-1)+k}M^k = \sum_{k=0}^{1} c_{(n-1)+k}M^k = \underbrace{c_{(n-1)+0}}_{=c_{n-1}}\underbrace{M^0}_{=I_n} + \underbrace{c_{(n-1)+1}}_{=c_n=1}\underbrace{M^1}_{=M} = c_{n-1}I_n + M,$$

this rewrites as $(c_{n-1}I_n + M)(V) \subseteq V$. Now, let $w \in V$. Then,

$$(c_{n-1}I_n + M)\left(\underbrace{w}_{\in V}\right) \in (c_{n-1}I_n + M)(V) \subseteq V.$$

Since $(c_{n-1}I_n + M)(w) = c_{n-1}\underbrace{I_n w}_{=w} + \underbrace{Mw}_{\substack{=g(w)\\ \text{(by (1))}}} = c_{n-1}w + g(w)$, this rewrites as $c_{n-1}w + g(w) \in V$. Hence,

$$g(w) \in V - c_{n-1}\underbrace{w}_{\in V} \subseteq V - c_{n-1}V \subseteq V \qquad \text{(since } V \text{ is an } A\text{-module)}.$$

Now, let us forget that we fixed w. We thus have shown that $g(w) \in V$ for every $w \in V$. In other words, $g(V) \subseteq V$. This proves Lemma 0.5. □

Our next step is a proof of Theorem 0.3 in the case when N (rather than M) is finitely generated:

Lemma 0.6. Let A be a commutative ring with unity. Let M be an A-module. Let N be an A-submodule of M such that N is finitely generated as an A-module. Let $f : N \to M$ be a surjective A-module homomorphism. Then, f is an A-module isomorphism.

Proof of Lemma 0.6. We know that N is finitely generated. In other words, there exist finitely many elements $a_1, a_2, \ldots, a_n$ of N such that N is generated by $a_1, a_2, \ldots, a_n$ as an A-module. Consider these $a_1, a_2, \ldots, a_n$.

Let $(e_1, e_2, \ldots, e_n)$ be the standard basis of the A-module A^n. (Thus, for every $i \in \{1, 2, \ldots, n\}$, the vector e_i is the vector in A^n whose i-th coordinate is 1 and whose other coordinates are all 0.) Clearly, in order to define an A-linear map from A^n to an A-module, it is enough to specify the images of this map at the basis vectors e_i (and these images can be chosen arbitrarily). Thus, we can define an A-linear map $p : A^n \to N$ by

$$\left(p\left(e_i\right) = a_i \qquad \text{for every } i \in \{1, 2, \ldots, n\}\right).$$

Consider this p.

The generators $a_1, a_2, \ldots, a_n$ of the A-module N are in the image of the map p (since $a_i = p\left(e_i\right)$ for every $i \in \{1, 2, \ldots, n\}$). Thus, the A-linear map $p : A^n \to N$ is surjective. Hence, the map $f \circ p : A^n \to M$ is also surjective (being the composition of the surjective maps f and p). Hence, $M = \left(f \circ p\right)\left(A^n\right)$.

Let us now define n elements $h_1, h_2, \ldots, h_n$ of A^n as follows: For every $i \in \{1, 2, \ldots, n\}$, there exists a vector $h \in A^n$ such that $p\left(e_i\right) = \left(f \circ p\right)(h)$ (since $p\left(e_i\right) \in N \subseteq M = \left(f \circ p\right)\left(A^n\right)$). Pick such an h and denote it by h_i. Thus, for every $i \in \{1, 2, \ldots, n\}$, we have defined a vector $h_i \in A^n$ such that

$$p\left(e_i\right) = \left(f \circ p\right)\left(h_i\right). \tag{5}$$

We have thus constructed n elements $h_1, h_2, \ldots, h_n$ of A^n.

Recall that, in order to define an A-linear map from A^n to an A-module, it is enough to specify the images of this map at the basis vectors e_i (and these images can be chosen arbitrarily). Hence, we can define an A-linear map $g : A^n \to A^n$ by

$$\left(g\left(e_i\right) = h_i \qquad \text{for every } i \in \{1, 2, \ldots, n\}\right).$$

Consider this g. Then, $f \circ p \circ g = p$ [3].

Let V be the A-submodule $\operatorname{Ker}\left(f \circ p\right)$ of A^n. It is straightforward to prove that $g^{-1}\left(V\right) \subseteq V$ [4]. Lemma 0.5 thus shows that $g\left(V\right) \subseteq V$.

[3] *Proof.* Every $i \in \{1, 2, \ldots, n\}$ satisfies

$$\left(f \circ p \circ g\right)\left(e_i\right) = \left(f \circ p\right)\left(\underbrace{g\left(e_i\right)}_{=h_i}\right) = \left(f \circ p\right)\left(h_i\right) = p\left(e_i\right) \qquad \text{(by (5))}.$$

In other words, the A-linear maps $f \circ p \circ g$ and p are equal to each other on each element of the basis $(e_1, e_2, \ldots, e_n)$ of A^n. Consequently, these maps $f \circ p \circ g$ and p must be identical (because if two A-linear maps from some A-module P are equal to each other on each element of a given basis of P, then these two maps must be identical). In other words, $f \circ p \circ g = p$, qed.

[4] *Proof.* Let $w \in g^{-1}\left(V\right)$. Then, $g\left(w\right) \in V = \operatorname{Ker}\left(f \circ p\right)$, so that $\left(f \circ p\right)\left(g\left(w\right)\right) = 0$. Thus,

Let now $w \in \operatorname{Ker} f$ be arbitrary. Then, $w \in N$ satisfies $f(w) = 0$ (since $w \in \operatorname{Ker} f$). But the map p is surjective; thus, $N = p(A^n)$. Hence, $w \in N = p(A^n)$. In other words, there exists some $v \in A^n$ such that $w = p(v)$. Consider this v. We have $(f \circ p)(v) = f\left(\underbrace{p(v)}_{=w}\right) = f(w) = 0$, so that $v \in \operatorname{Ker}(f \circ p) = V$ and thus $g\left(\underbrace{v}_{\in V}\right) \in g(V) \subseteq V = \operatorname{Ker}(f \circ p)$ and thus $(f \circ p)(g(v)) = 0$. Thus, $(f \circ p \circ g)(v) = (f \circ p)(g(v)) = 0$. Since $f \circ p \circ g = p$, this rewrites as $p(v) = 0$. Thus, $w = p(v) = 0$.

Now, let us forget that we fixed w. We thus have proven that $w = 0$ for every $w \in \operatorname{Ker} f$. In other words, $\operatorname{Ker} f = 0$. Hence, the map f is injective. Since f is also surjective, this yields that f is bijective. Thus, f is an A-module isomorphism (since f is an A-module homomorphism). This proves Lemma 0.6. □

Now, we can finally step to the proof of Theorem 0.1:

Proof of Theorem 0.1. We know that M is finitely generated. In other words, there exist finitely many elements $a_1, a_2, \ldots, a_n$ of M such that M is generated by $a_1, a_2, \ldots, a_n$ as an A-module. Consider these $a_1, a_2, \ldots, a_n$.

Notice that $M = f(N)$ (since the map f is surjective).

For every $i \in \{1, 2, \ldots, n\}$, we define an element g_i of N as follows: There exists some $g \in N$ such that $a_i = f(g)$ (since $a_i \in M = f(N)$). Pick such a g and denote it by g_i. Thus, for every $i \in \{1, 2, \ldots, n\}$, we have defined some $g_i \in N$ satisfying

$$a_i = f(g_i). \tag{6}$$

Hence, we have defined n elements $g_1, g_2, \ldots, g_n$ of N.

Let $v \in \operatorname{Ker} f$. We shall prove that $v = 0$.

Let N' be the A-submodule $Av + (Ag_1 + Ag_2 + \cdots + Ag_n)$ of N. Then, the A-module N' is finitely generated (in fact, it is generated by the $n+1$ elements $v, g_1, g_2, \ldots, g_n$) and satisfies $N' \subseteq N \subseteq M$. Also, the A-linear map $f|_{N'}: N' \to M$ is surjective, because its image contains the generators $a_1, a_2, \ldots, a_n$ of M (in fact, for every $i \in \{1, 2, \ldots, n\}$, we have $g_i \in Ag_i \subseteq Av + (Ag_1 + Ag_2 + \cdots + Ag_n) \subseteq$

$0 = (f \circ p)(g(w)) = \underbrace{(f \circ p \circ g)}_{=p}(w) = p(w)$, so that $p(w) = 0$ and thus $(f \circ p)(w) = f\left(\underbrace{p(w)}_{=0}\right) = f(0) = 0$ (since f is A-linear). Consequently, $w \in \operatorname{Ker}(f \circ p) = V$.

Let us now forget that we fixed w. We thus have shown that $w \in V$ for every $w \in g^{-1}(V)$. In other words, $g^{-1}(V) \subseteq V$, qed.

N' and thus $a_i = f\left(\underbrace{g_i}_{\in N'}\right) = (f\mid_{N'})(g_i)$, which shows that the image of $f\mid_{N'}$ contains a_i). Hence, Lemma 0.6 (applied to N' and $f\mid_{N'}$ instead of N and f) yields that $f\mid_{N'}$ is an A-module isomorphism. In particular, $f\mid_{N'}$ is injective. Thus, $\operatorname{Ker}(f\mid_{N'}) = 0$.

But $v \in Av \subseteq Av + (Ag_1 + Ag_2 + \cdots + Ag_n) = N'$ and $(f\mid_{N'})(v) = f(v) = 0$ (since $v \in \operatorname{Ker} f$). Hence, $v \in \operatorname{Ker}(f\mid_{N'}) = 0$. In other words, $v = 0$.

Now, let us forget that we fixed v. We thus have shown that $v = 0$ for every $v \in \operatorname{Ker} f$. In other words, $\operatorname{Ker} f = 0$. Hence, the map f is injective. Since f is also surjective, this yields that f is bijective. Thus, f is an A-module isomorphism (since f is an A-module homomorphism). This proves Theorem 0.1. □

Proof of Corollary 0.2. Corollary 0.2 follows immediately from Theorem 0.1 (applied to $N = M$). □

References

[1] Morris Orzech, *Onto Endomorphisms are Isomorphisms*, Amer. Math. Monthly 78 (1971), pp. 357–362.

[2] *Is Orzech's generalization of the surjective-endomorphism-is-injective theorem correct?*, math.stackexchange question #1065786 and consequent discussion.
http://math.stackexchange.com/questions/1065786

[3] *Surjective endomorphisms of finitely generated modules are isomorphisms*, math.stackexchange question #239364 and consequent discussion.
http://math.stackexchange.com/questions/239364

[4] Joel G. Broida and S. Gill Williamson, *Comprehensive Introduction to Linear Algebra*, Addison-Wesley 1989.
http://cseweb.ucsd.edu/~gill/CILASite/

[5] Darij Grinberg, *The trace Cayley-Hamilton theorem*, 21 November 2016.
http://www.cip.ifi.lmu.de/~grinberg/algebra/trach.pdf

[6] Howard Straubing, *A combinatorial proof of the Cayley-Hamilton theorem*, Discrete Mathematics, Volume 43, Issues 2–3, 1983, pp. 273–279.
http://www.sciencedirect.com/science/article/pii/0012365X83901644

[7] Doron Zeilberger, *A combinatorial approach to matrix algebra*, Discrete Mathematics, Volume 56, Issue 1, September 1985, pp. 61–72.
http://www.sciencedirect.com/science/article/pii/0012365X8590192X

A better scan is available at
http://www.math.rutgers.edu/~zeilberg/mamarimY/DM85.pdf

[8] Chris Bernhardt, *A proof of the Cayley Hamilton theorem*.
http://faculty.fairfield.edu/cbernhardt/cayleyhamilton.pdf

[9] Jerry Shurman, *The Cayley-Hamilton theorem via multilinear algebra*, version 24 May 2015.
http://people.reed.edu/~jerry/332/28ch.pdf

[10] Anthony W. Knapp, *Basic Algebra*, Digital Second Edition, 2016.
http://www.math.stonybrook.edu/~aknapp/download.html

[11] Keith Conrad, *Universal Identities I*, 2013.
http://www.math.uconn.edu/~kconrad/blurbs/

[12] Darij Grinberg, *A note on lifting isomorphisms of modules over PIDs*.
http://www.cip.ifi.lmu.de/~grinberg/algebra/pidisolift.pdf