THE MATROID OF SUPPORTS OF A LINEAR CODE

Alexander Barg [*]

Abstract

A relation between the Hamming weight enumerator of a linear code and the Tutte polynomial of the corresponding matroid has been known since long ago. It provides a simple proof of the MacWilliams equation (see D. Welsh, *Matroid Theory* (1976)). In this paper we prove analogous results for the support weight distributions of a code.

Keywords: Support weight distributions, Tutte polynomial, Mac-Williams equation.

1 Introduction

Higher weights of linear codes and their distributions have received considerable interest lately, sparked by an application of this concept in cryptography. The state-of-the-art in the study of higher weights has been recently reviewed in [9]. The emphasis in [9] is on the geometric aspect of the problem. Our approach is, rather, more combinatorial.

We refer to [11] for the notions and results from matroid theory used below. Let $0 \le k \le n$ be two integers and $F = \mathbb{F}_q$ a finite field. Let $\mathcal{G}$ be an injective and $\mathcal{H}$ a surjective linear mapping acting as follows:

$$F^k \xrightarrow{\mathcal{G}} F^n \xrightarrow{\mathcal{H}} F^{n-k}$$

such that $\mathcal{H} \circ \mathcal{G} = 0$, and consider two matrices, G and H, where G is the transposed matrix of $\mathcal{G}$ and H the matrix of $\mathcal{H}$. Let S be an n-set identified with the coordinate set of F^n. We shall study the following objects related to this structure:

C $(\widetilde{C})$ a linear space spanned by the rows of G (H);

M $(\widetilde{M})$ a matroid on S represented by the column space of G (H).

Remark 1. $\widetilde{C}$ is the dual code of C and $\widetilde{M}$ the dual matroid of M. Throughout the paper $\widetilde{\ }$ refers to dual objects or their numerical parameters.

Definition 1 *Let $\mathcal{C}$ be a subcode of C. A subset E of S such that*

$$\forall_{e \in S \setminus E} \forall_{\mathbf{c} \in \mathcal{C}} (c_e = 0) \quad \textit{and} \quad \forall_{e \in E} \exists_{\mathbf{c} \in \mathcal{C}} (c_e \neq 0)$$

is called the support *of $\mathcal{C}$, denoted* $\operatorname{supp} \mathcal{C}$.

[*] Faculty of Math. and Computer Science, Eindhoven University of Technology, Postbus 513, Den Dolech 2, 5600 MB, Eindhoven, The Netherlands, and Institute for Information Transmission Problems, Bolshoj Karetnyj 19, Moscow, Russia.

Let

$$A_i^m = \Big|\Big\{ \mathcal{C} \in C \,:\, \dim \mathcal{C} = m,\ |\mathrm{supp}\,\mathcal{C}| = i \Big\}\Big|, \quad 0 \le m \le n.$$

The set $\{A_i^m,\ 0 \le i \le n\}$ is called the *mth support weight distribution* of C. Define the numbers

$$D_i^m = \sum_{u=0}^{m} [m]_u A_i^u, \quad m \ge 0 \tag{1}$$

where

$$[m]_u = \prod_{i=0}^{u-1} (q^m - q^i), \quad [m]_0 := 1,$$

and the polynomials

$$D^m(x) = \sum_{i=0}^{n} D_i^m x^i, \quad m \ge 0.$$

Remark 2. $W(x) := D^1(x) = 1 + \sum_{i=0}^{n} (q-1) A_i^1 x^i$ is simply the Hamming weight enumerator of C. Therefore, $\sum_{i=0}^{n} D_i^1 = q^k$. Generally,

$$\begin{aligned}\sum_{i=0}^{n} D_i^m = \sum_{i=0}^{n}\sum_{u=0}^{m} [m]_u A_i^u = \sum_u [m]_u \sum_i A_i^u = \sum_u [m]_u \begin{bmatrix} k \\ u \end{bmatrix} \\ = q^{mk},\end{aligned} \tag{2}$$

where the last equality follows from the fact that the term $[m]_u \begin{bmatrix} k \\ u \end{bmatrix} = \frac{[m]_u [k]_u}{[u]_u}$ counts the number of $m \times k$ matrices of rank u (see [2]).

Let $E \subseteq S$ and $\mathcal{G}^E = \mathrm{proj}_E\, \mathcal{G}$. Let G^E be the transposed matrix of $\mathcal{G}^E$. Accordingly, let $\mathcal{H}^E = \mathrm{proj}_E\, \mathcal{H}$ and H^E the matrix of $\mathcal{H}^E$. The rank function of $M(\widetilde{M})$ is given by $\rho E = \mathrm{rk}\, G^E$ ($\widetilde{\rho} E = \mathrm{rk}\, H^E$). Note that $\rho S = k$ and $\widetilde{\rho} S = n - k$. The following equation relates the two rank functions: for every $E \subseteq S$,

$$k - \rho(S \setminus E) = |E| - \widetilde{\rho} E. \tag{3}$$

Coding Proof. Suppose E occupies the first part of S and represent G in the form $\left[\begin{array}{c|c} A & 0 \\ \hline B & D \end{array}\right]$, where A has the maximal possible number of rows. Let C^E be the subcode of C equal to $\ker \mathcal{H}^E$, i.e., the subcode generated by A. Clearly,

$$\dim C^E = |E| - \widetilde{\rho} E. \tag{4}$$

Also since A is maximal, $\mathrm{rk}\,(B|D) = \mathrm{rk}\,(D)$, and therefore $\dim C^E = k - \rho(S \setminus E)$. □

Definition 2 *The* Whitney function *of M is defined by*

$$R(M;x,y)=\sum_{E\subset S} x^{\rho S-\rho E}y^{|E|-\rho E} \tag{5}$$

and the Tutte polynomial*, $T(M;x,y)$, by*

$$T(M;x,y)=R(M;x-1,y-1).$$

A relation between the Whitney function of M and that of $\widetilde{M}$ follows immediately from (3) and (5):

$$R(M;x,y)=R(\widetilde{M},y,x). \tag{6}$$

Therefore,

$$T(M;x,y)=T(\widetilde{M},y,x). \tag{7}$$

Collecting monomials in (5), we can write this definition as

$$R(M;x,y)=\sum_u\sum_v R_u^v x^u y^v,$$

where $R_u^v=\Big|\Big\{\ E\subseteq S\ :\ \rho S-\rho E=u, |E|-\rho E=v\ \Big\}\Big|$. Then by (6),

$$R_u^v=\widetilde{R}_v^u,\quad u,v\in\mathbb{Z}. \tag{8}$$

2 MacWilliams Equations for Support Weight Distributions

Proofs of the MacWilliams equations for support weight distributions were given by T. Kløve [5] and J. Simonis [8]. The argument by J. Simonis uses implicitly the matroidal duality (see Sect. 3).

Here we show relations between the support weight enumerators $D^m(x)$ of C and the Tutte polynomial of M, which by (7) enables one to re-establish the MacWilliams equation in a simple form.

Let $M|E$ be a restriction of M on and $M.E$ a contraction of M to a set E (see [11]; in terms of C these operations correspond to puncturing and shortening, respectively). An integer-valued function defined on a class of matroids is called a *Tutte–Grothendieck invariant* if

(i) $\forall_{e\in S} f(M)=f(M|(S\setminus e))+f(M.(S\setminus e))$,

(ii) if M_1 is a connected component of M on $T\subseteq S$, then $f(M)=f(M_1)f(M|(S\setminus T)$.

As in [11, Sect. 15.7], it is not difficult to prove the following.

Lemma 1 *The function*

$$(1-u)^k u^{n-k} D^m(u)$$

is a Tutte–Grothendieck invariant of M.

Then using a powerful theorem of Brylawski (see, e.g., [11]) that any Tutte–Grothendieck invariant of M is an evaluation of its Tutte polynomial, we can prove the sought relations.

Theorem 1

$$D^m(u) = (1-u)^k u^{n-k} T\left(M; \frac{1+(q^m-1)u}{1-u}, \frac{1}{u}\right), \quad 0 \le m \le n. \tag{9}$$

Now the set of MacWilliams equations follows easily.

Theorem 2 (*The MacWilliams equations for support weight distributions*)

$$D^m(x) = q^{-m(n-k)}(1+(q^m-1)x)^n \widetilde{D}^m\Big(\frac{1-x}{1+(q^m-1)x}\Big), \quad m \ge 0. \tag{10}$$

Proof. Starting with $D^m(x)$, apply (9), then (7), and then again (9) in the reverse direction. □

Substituting u/v for x in Eq. (10), we can write it in a more familiar homogeneous form:

$$D^m(u,v) = \frac{1}{|\widetilde{C}|^m}\widetilde{D}^m(u-v, u+(q^m-1)v), \quad m \ge 0.$$

Remark 3. (*The linear programming bound*). Let $\mathbf{D}^m = (D_0^m, D_1^m, \dots, D_n^m)$. Then

$$\mathbf{D}^m = q^{-(n-k)m}\widetilde{\mathbf{D}}^m \mathbf{P}^m, \quad m \ge 0, \tag{11}$$

where $\mathbf{P}^m$ is an $(n+1)\times(n+1)$ Krawtchouk matrix with entries

$$P_{ij}^m = P_j^m(i) := \sum_{\ell=0}^n (-1)^\ell \binom{i}{\ell}\binom{n-i}{j-\ell}(q^m-1)^{j-\ell}. \tag{12}$$

Formally, these Krawtchouk numbers correspond to the alphabet of size q^m, which reflects the connection between higher weights in q-ary codes and the first weight in q^m-ary codes, used for deriving the MacWilliams equation in [5].

Thus, we have the generalized *Delsarte–MacWilliams inequalities* for *linear* codes in the form

$$\sum_{i=0}^n D_i^m P_j^m(i) \ge 0, \quad 0 \le j \le n.$$

Together with (2), this enables one to derive a set of linear programming bounds for linear codes.

Theorem 3 *For any $m = 1, 2, \dots, n$, the size of an $[n,k,d]$ code C can be bounded from above as follows*:

$$|C|^m \le \max\Big\{\sum_{i=0}^n D_i^m \mid D_0^m = 1,\ D_i^m \ge 0,\ D_1^m = 0, \dots, D_{d-1}^m = 0,$$
$$\sum_{i=0}^n D_i^m P_j^m(i) \ge 0 \text{ for } 0 \le j \le n\Big\}.$$

Example 1. Let H be formed by all $n = 2^s - 1$ distinct nonzero binary columns. The row space of H forms the simplex code, and the dual space is the Hamming code. Let $A^r_{s,\ell}$ be the number of $s \times \ell$ matrices of rank r with distinct nonzero columns, when the permutation of columns is not counted. Clearly,

$$R(\widetilde{M}; x, y) = \sum_{u,v} \widetilde{R}^v_u x^u y^v,$$

where $\widetilde{R}^0_s = 1$ and

$$\widetilde{R}^v_u = A^{s-u}_{s,v+s-u}, \quad u \ge 0, v > 0.$$

Numbers $A^r_{s,\ell}$ are calculated in [6], [1] in the form

$$A^r_{s,\ell} = \begin{bmatrix} s \\ r \end{bmatrix} \sum_{j=0}^{r-1} (-1)^j 2^{\frac{1}{2}j(j-1)} \begin{bmatrix} r \\ j \end{bmatrix} \binom{2^{r-j} - 1}{\ell}.$$

This enables one to perform explicit calculations. For instance, let $n = 7, k = 4$. Calculating the numbers $A^r_{3,\ell}$, we obtain

r	ℓ	1	2	3	4	5	6	7
1		7						
2			21	7				
3				28	35	21	7	1

which yields the Whitney function of $\widetilde{M}$ in the form

$$R(\widetilde{M}; x, y) = x^3 + 7x^2 + 21x + 7xy + 35y + 21y^2 + 7y^3 + y^4 + 28.$$

Then

$$T(\widetilde{M}; x, y) = R(\widetilde{M}; x - 1, y - 1) = 4x^2 + x^3 + 7xy + 3y + 6y^2 + 3y^3 + y^4.$$

By (7), (9), this allows us to calculate the matrices $\mathbf{D} = (\mathbf{D}_0, \dots, \mathbf{D}_4)^t$ and $\widetilde{\mathbf{D}} = (\widetilde{\mathbf{D}}_0, \dots, \widetilde{\mathbf{D}}_3)^t$:

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 7 & 7 & 0 & 0 & 1 \\ 1 & 0 & 0 & 21 & 21 & 126 & 42 & 45 \\ 1 & 0 & 0 & 49 & 49 & 882 & 1470 & 1645 \\ 1 & 0 & 0 & 105 & 105 & 4410 & 19110 & 41805 \end{bmatrix} \quad \widetilde{\mathbf{D}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 7 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 21 & 0 & 42 & 0 \\ 1 & 0 & 0 & 0 & 49 & 0 & 294 & 168 \end{bmatrix}$$

From this one recovers, via the triangular system formed by equations (1) for $m \ge 0$, the support weight distributions of C and $\widetilde{C}$:

		i	0	1	2	3	4	5	6	7
		0	1	0	0	0	0	0	0	0
		1		0	0	7	7	0	0	1
A^m_i	m	2			0	0	0	21	7	7
		3				0	0	0	7	8
		4					0	0	0	1

		i	0	1	2	3	4	5	6	7
		0	1	0	0	0	0	0	0	0
$\widetilde{A}^m_i$	m	1		0	0	0	7	0	0	0
		2			0	0	0	0	7	0
		3				0	0	0	0	1

Alternatively, if the matrix $\widetilde{\mathbf{D}}$ is found in a different manner, one can use Krawtchouk matrices to compute $\mathbf{D}$.

Note that though the sequence D_i^m ($\widetilde{D}_i^m$) is by definition semi-infinite, its entries with numbers $m >\ \dim C = 4$ (resp., $m > \dim \widetilde{C} = 3$) do not bear any additional information about the code.

Remark 4. This seems to be one of the few examples when the Whitney function can be found explicitly. Generally, the computation of the Tutte polynomial for vector matroids is known to be difficult [4] and, therefore, does not facilitate the computation of support weight distributions.

3 MacWilliams Identities for Support Weight Distributions

In her original paper [7], F. J. MacWilliams gave two proofs of the main result. The second one uses characters and gives an equation of the form (10), while the first one leads directly to identities between the weights in C and $\widetilde{C}$. J. Simonis [8] mimics this approach in order to relate the support weights in C and $\widetilde{C}$. He starts with the quantities

$$E_i^m = \sum_{v=0}^{n} \binom{n-v}{i} A_v^m, \quad m \ge 0,\ i \ge 0.$$

Then a double counting argument shows that

$$E_i^m = \sum_{\alpha\in\mathbb{Z}} \begin{bmatrix}\alpha\\ m\end{bmatrix} \widetilde{R}^{\alpha}_{n-k-i+\alpha}, \quad m \ge 0,\ i \ge 0. \tag{13}$$

whereupon an application of (8) just in the same manner as in the proof of Theorem 2 gives the MacWilliams-type identities in the form

$$E_{n-i}^m = \sum_{u=0}^{m} q^{u(k-n-i+u-m)} \begin{bmatrix}k-n+i\\ m-u\end{bmatrix} \widetilde{E}_i^u. \tag{14}$$

Unfortunately, the polynomials $\sum_{i=0}^n A_i^m x^i$ for $m \ge 2$ do not satisfy any equation of the form (10).

Taking $m = 1$, we can further simplify (13). Namely, let $W(x) = \sum_{i=0}^n A_i x^{n-i}$ be the *Hamming* weight enumerator of C, i.e., $A_i = A_i^0 + (q-1)A_i^1, 0 \le i \le n$, and denote

$$B_i = \sum_{j=0}^{n} \binom{n-j}{i} A_j, \quad i \in \mathbb{Z}. \tag{15}$$

Then (13) implies

$$B_{n-j} = \sum_{m\in\mathbb{Z}} q^m \widetilde{R}^m_{n-k-j+m}, \quad 0 \le j \le n.$$

Then (14) takes on the following remarkably simple form:

$$B_{n-j} = q^{-(n-k-j)} \widetilde{B}_j, \quad 0 \le j \le n. \tag{16}$$

Note that

$$W(x) = \sum_{i=0}^{n} A_i x^{n-i} = \sum_{i=0}^{n} B_i (x-1)^i,$$

since then (15) follows by putting $x = z + 1$, and therefore, Eq. (16) gives also the MacWilliams equation in a very compact form. Namely, let $w(z) = W(x-1)$, then

$$w(z) = q^k (z/q)^n \widetilde{w}(\frac{q}{z}),$$

or

$$w(qz) = q^k \ \widetilde{w}^*(z),$$

where * means taking the reciprocal.
Example 1 (continued.) For the [7,4,3] code, the polynomial $\widetilde{w}(z)$ has the form

$$\widetilde{w}(z) = 8 + 28z + 42z^2 + 42z^3 + 35z^4 + 21z^5 + 7z^6 + z^7$$

and the reciprocal of $16\widetilde{w}(z)$ is indeed taken by the substitution $z \to \frac{z}{2}$ to the form

$$w(z) = 16 + 56z + 84z^2 + 70z^3 + 42z^4 + 21z^5 + 7z^6 + z^7.$$

Remark 5. Equation (15) implies that

$$B_i = \binom{n}{i}, \quad n - \operatorname{dist} C + 1 \le i \le n.$$

Therefore by (14) also

$$B_i = q^{-(n-k-i)} \binom{n}{i}, \quad 0 \le i \le \operatorname{dist} \widetilde{C} - 1.$$

Observe that introducing the numbers B_i corresponds to writing the polynomial $W(x)$ in a different basis, which makes our treatment of the Hamming weight enumerator close to that in [10, Ch.1].

Eq. (14) gives MacWilliams identities relating the numbers A_i^m and $\widetilde{A}_i^m$. An interesting question is whether it is possible to derive identities for these quantities involving Krawtchouk numbers, as we did in (11) for the numbers D_i^m. The answer is given by the following theorem.

Theorem 4

$$\sum_{i=0}^{n} P_t^m(i) A_i^m = \sum_{u=0}^{m} |C|^u \sum_{v=0}^{n} \binom{n-v}{n-t} \widetilde{A}_v^u \sum_{\ell=0}^{n} (-1)^{t-\ell} q^{(\ell-u)(m-u)} \begin{bmatrix} k-\ell \\ m-u \end{bmatrix} \binom{t-v}{t-\ell}.$$

$$m \ge 0, \ t \ge 0.$$

Proof. We use another expression for the Krawtchouk numbers which is obtained from (12) using the binomial expansion:

$$P_t^m(i)=\sum_{j=0}^{n}(-1)^{t-n+j}q^{m(n-j)}\binom{n-i}{n-j}\binom{j}{n-t}.$$

Multiply both parts of (14) by $(-1)^{t-n+j}\binom{j}{n-t}q^{m(n-j)}$ and sum over all j. Then it turns into

$$\begin{aligned}\sum_{i=0}^{n}P_t^m(i)A_i^m &= \sum_j(-1)^{t-n+j}\binom{j}{n-t}q^{m(n-j)}\sum_{u=0}^{m}q^{u(j-\widetilde{k}+u-m)}\begin{bmatrix}j-\widetilde{k}\\ m-u\end{bmatrix}\sum_{v=0}^{n}\binom{n-v}{j}\widetilde{A}_v^u\\ &=\sum_u\sum_v\widetilde{A}_v^u\sum_j(-1)^{t-n+j}q^{uk+(j+u-n)(u-m)}\begin{bmatrix}j-\widetilde{k}\\ m-u\end{bmatrix}\binom{n-v}{j}\binom{j}{n-t}\\ &=\sum_u|C|^u\sum_v\binom{n-v}{n-t}\widetilde{A}_v^u\sum_j(-1)^{t-n+j}q^{(j+u-n)(u-m)}\begin{bmatrix}j-\widetilde{k}\\ m-u\end{bmatrix}\binom{t-v}{t-n+j}\\ &=\sum_u|C|^u\sum_v\binom{n-v}{n-t}\widetilde{A}_v^u\sum_\ell(-1)^{t-\ell}q^{(\ell-u)(m-u)}\begin{bmatrix}k-\ell\\ m-u\end{bmatrix}\binom{t-v}{t-\ell},\end{aligned}$$

where we have used the identity $\binom{r}{j}\binom{j}{s}=\binom{r}{s}\binom{r-s}{j-s}$. □

Specializing this theorem for $m=1$, we get the usual MacWilliams identities

$$\sum_{i=1}^{n}P_t^1(i)A_i^1=-\binom{n}{t}(q-1)^{t-1}+|C|\widetilde{A}_t^1,\quad 1\le t\le n,$$

where A_i^1 is the number of one-dimensional subcodes of C with support i. The best upper bound for the size of the code, obtained via the linear programming approach [3], rests on the fact that this equation implies the inequality

$$\sum_{i=1}^{n}P_t(i)A_i^1\ge-\binom{n}{t}(q-1)^{t-1}.$$

An attempt to parallelize this technique for higher dimensions leads to rewriting the equation in the theorem, isolating the term with $u=0$:

$$\sum_{i=0}^{n}P_t^m(i)A_i^m=\binom{n}{t}\sum_\ell(-1)^{t-\ell}q^{ml}\begin{bmatrix}k-\ell\\ m\end{bmatrix}\binom{t}{\ell}+\dots. \tag{17}$$

Unfortunately, the discarded terms are not always positive, as shown by the following example.

Example 1 (continued.) The support weight distributions for the $[7,3,4]$ code, calculated above, are $\{A_0^0=1\}$, $\{A_4^1=7\}$, $\{A_6^2=7\}$, $\{A_7^3=1\}$. Taking in (17) $m=2$, $t=4$, the left-hand side equals -315 and the first term on the right-hand side equals 805.

References

[1] L. Carlitz, “Note on a paper by Laksov,” *Math. Scand.*, **19** (1966), 38–40.

[2] S. D. Fisher and M. N. Alexander, “Matrices over a finite field,” *Amer. Math. Monthly*, **79** (1966), 639–641.

[3] P. Delsarte, “Application and generalization of the MacWilliams transform in coding theory,” *Proc.* 15*th Sympos. Inform. Theory in the Benelux* (1994), pp. 9–44.

[4] F. Jaeger, D. Vertigan, and D. J. A. Welsh, “On the computational complexity of the Jones and Tutte polynomials,” *Math. Proc. Cambr. Phil. Soc.*, **108** (1990), 35–53.

[5] T. Kløve, “Support weight distribution of linear codes,” *Discrete Mathematics*, **106/107** (1992), 311–316.

[6] D. Laksov, “Linear recurring sequences over finite fields,” *Math. Scand.*, **16** (1965), 181–196.

[7] F. J. MacWilliams, “A theorem on the distribution of weights in a systematic code,” *Bell Syst. Tech. J.*, **42** (1963), 79–94.

[8] J. Simonis, “The effective length of subcodes,” *Applicable Algebra in Engineering, Communication and Computing*, **5** (1994), 371–377.

[9] M. A. Tsfasman and S. G. Vlăduţ, “Geometric approach to higher weights,” *IEEE Trans. Inform. Theory*, **IT-41** (1995), 1564–1588.

[10] M. A. Tsfasman and S. G. Vlăduţ, *Algebraic Geometric Codes*, Kluwer 1991.

[11] D. J. A. Welsh, *Matroid Theory*, Academic Press, London, 1976.