Generalized minimum distance decoding for correcting array errors

Author(s):
Sidorenko, Vladimir R.; Bossert, Martin; Gabidulin, Ernst M.

Publication date:
2010

Permanent link:
https://doi.org/10.3929/ethz-a-006001584

Generalized Minimum Distance Decoding for Correcting Array Errors

Vladimir R. Sidorenko, Martin Bossert
Inst. of Telecommunications and Applied Information Theory
Ulm University, Ulm, Germany,
{vladimir.sidorenko | martin.bossert}@uni-ulm.de

Ernst M. Gabidulin
Moscow Institute (State University) of Physics and Technology
Dolgoprudny, Russia
ernst.gabidulin@gmail.com

***Abstract*— We consider an array error model for data in matrix form, where the corrupted symbols are confined to a number of lines (rows and columns) of the matrix. Codes in array metric (maximum term rank metric) are well suited for error correction in this case. We generalize the array metric to the case when the reliability of every line of the matrix is available. We propose a minimum distance decoder for the generalized metric and estimate the guaranteed error correcting radius for this decoder.**

I. Introduction

Consider transmission of matrices C with elements from the field $\mathbb{F}_q$ over a channel with array (or crisscross) errors. This channel corrupt a number of lines (rows and columns) of the matrix C, i.e., the channel may erase some lines and replace components of some other lines by arbitrary elements of $\mathbb{F}_q$. Array errors can be found in various data storage applications and in OFDM systems. The array metric, which is also known as the maximal-term-rank metric, suits well for the channels with array errors. Array-error-correcting codes, i.e., codes having a distance d in the array metric, were proposed in [1], [2], [3], and in other publications. These codes have algebraic decoders, which are able to correct up to $(d-1)/2$ erroneous lines in the received matrix. More precisely, these decoders correct ε erroneous lines and θ erased lines as soon as

$$\lambda\varepsilon + \theta \le d - 1, \tag{1}$$

where $\lambda = 2$ is the tradeoff rate between errors and erasures for these decoders.

Assume that the decoder has side information about reliabilities of lines in the received matrix. Can we correct more than $(d-1)/\lambda$ erroneous lines in this case?

For the case of correction of independent errors (using codes in Hamming metric) the answer "yes" was done by Forney [4]. He introduced generalized Hamming distance, which is the weighted Hamming distance, where weights are the reliabilities of the received symbols. Forney also suggested a decoding algorithm, which uses an algebraic decoder (with $\lambda = 2$) in multi-trial manner to decode the received vector in generalized metric. Later, Kovalev [5] suggested an adaptive form of the Forney algorithm to decrease twice the number of decoding trials. The Forney-Kovalev decoding algorithm was refined by Weber and Abdel-Ghaffar in [6] and extended for $\lambda \le 2$ in [7].

In this paper we introduce generalized array distance, which is array-distance weighted by reliabilities. We show that this generalized array distance suits well to the channel with array errors and with side reliabilities information. Then we show, that decoding of codes in the new generalized metric can be done by a modification [7] of the Forney-Kovalev algorithm. This allows us to estimate the error correcting radius of the decoding algorithm for all λs.

II. Array-error model and array metric

A. Channel

We consider transmission of $m \times n$ matrix C over $\mathbb{F}_q$. Let us enumerate lines (rows and columns) of C by numbers $1, \ldots, m+n$. The received matrix Y is $Y = C + E$, where error-matrix E is constructed by the channel as follows. The channel selects s different lines of E with probability $P(s)$ and fills these lines randomly by elements of $\mathbb{F}_q$, independently and equiprobable. All the rest components of E are zeros. We assume that $P(s)$ decreases with s.

B. Array metric

The array (or maximal term rank) metric is defined as follows. Assume that all nonzero elements of the matrix $A \in \mathbb{F}_q^{m\times n}$ are contained in t lines with indexes $\{i_1, \ldots, i_t\}$, then we call this set *a covering of* A and denote it by $\mathcal{I}(A) = \{i_1, \ldots, i_t\}$. The *array-weight* (or array-norm) $w^{(\mathrm{a})}(A)$ of a matrix A is defined as follows

$$w^{(\mathrm{a})}(A) \triangleq \min_{\mathcal{I}(A)} |\mathcal{I}(A)|. \tag{2}$$

In other words, the array-weight of a matrix A is the minimum number of lines that contain all nonzero elements of A. The maximum possible array weight of a matrix A is $\min\{m, n\}$.

The *array-distance* $d^{(\mathrm{a})}(A, B)$ between matrices A and B is defined as

$$d^{(\mathrm{a})}(A, B) \triangleq w^{(\mathrm{a})}(A - B). \tag{3}$$

The array-norm (2) satisfies the axioms of a norm, and hence the array distance (3) satisfies the axioms of a distance.

This work of V.R. Sidorenko has been supported by DFG (German Research Council) under grant BO 867/21. V.R. Sidorenko is on leave from IITP, Russian Academy of Sciences, Moscow, Russia.

The work of E.M. Gabidulin has been supported by DFG under grant 436 RUS 113/941/0-1

III. Algebraic codes correcting array-errors

A linear (nm, k, d) code $\mathcal{C}$ of rate $R = \frac{k}{mn}$ is a linear subspace of $\mathbb{F}_q^{m\times n}$ of dimension k, where the array code distance $d^{(\mathrm{a})}(\mathcal{C}) = d$ is the minimum array distance between two different codewords of $\mathcal{C}$.

From now on let us assume without loss of generality that

$$m \geq n. \tag{4}$$

The code-dimension k satisfies the following Singleton-type bound [1]

$$k \leq m(n - d + 1). \tag{5}$$

In [1] the following construction of array-error-correcting codes was proposed. Assume we have an $(n, k, d^{(H)})$ block linear code $\mathcal{C}^{(H)}$ over $\mathbb{F}_q$ with distance $d^{(H)}$ in *Hamming* metric. A code matrix $C = ||c_{i,j}||, i = 1, \ldots, m, j = 1, \ldots, n$ of an array-error-correcting code $\mathcal{C}$ we design as follows. We say that the set $\{c_{(i+j) \bmod m+1, j+1} : \quad j = 0, \ldots, n-1\}$ forms the $(i+1)$st diagonal of the matrix C, $i = 0, \ldots, m-1$. By writing m arbitrary words of the code $\mathcal{C}^{(H)}$ into m diagonals of the matrix C we obtain a codeword of the code $\mathcal{C}$. Notice, that every corrupted line (erased line or line with errors) in C affects at most one symbol of every diagonal of C. As a result we obtain an $(nm, km, d^{(H)})$ code $\mathcal{C}$ with array distance $d^{(H)}$.

Assume we have a decoder of the code $\mathcal{C}^{(H)}$ correcting up to t errors in Hamming metric. Then, by correcting errors in every diagonal of a received matrix Y, we will correct every error matrix E of array-weight up to t. Standard algebraic decoders allow to correct up to $(d^{(H)} - 1)/2$ errors. If the order q of the field is large enough, $q > (n+1)^l$ then we can use l-punctured, $l = 1, 2, ...,$ Reed–Solomon codes [8], which allows to correct up to $\frac{l}{l+1} d^{(H)}$ errors [9]. More precisely the decoder corrects ε errors and θ erasures if (1) holds and fails otherwise. Here, the real number $\lambda = 1 + 1/l$, $1 < \lambda \leq 2$ is the tradeoff rate between errors and erasures for this decoder. This is an example of array-error-correcting (mn, k, d) code $\mathcal{C}$ with array-distance d, which reaches the Singleton-type upper bound (5). If $q > (n+1)^l$ then there is a decoder Φ for this code, which corrects ε errors and θ erasures as soon as (1) is satisfied with $\lambda = 1 + 1/l$.

Another class of array-error-correcting codes is based on codes in rank metric. Rank distance between $m \times n$ matrices A and B is defined as $d^{(\mathrm{r})}(A, B) = \text{rank}\,(A - B)$. Since $\text{rank}\,(A - B) \leq d^{(\mathrm{a})}(A, B)$, every code having distance d in the rank metric has distance at least d in the array metric. There is a class of (mn, k, d) Gabidulin codes [2], [3], which have distance d in rank metric satisfying the Singleton-type upper bound (5) with equality. Hence, every (mn, k, d) Gabidulin code is simultaneously (mn, k, d) code with array-distance d. There are known algebraic decoders of Gabidulin codes, which correct up to $(d-1)/2$ errors in rank metric, and hence in the array metric as well. This is another example of array-error-correcting codes, having the decoder Φ, which corrects ε errors and θ erasures as soon as (1) is satisfied with $\lambda = 2$.

IV. Generalized distance and GMD decoding

A. Generalized weight and distance

Given a vector $h = (h_1, \ldots, h_{n+m})$ of line-reliabilities, where $0 \leq h_i \leq 1$, we define generalized distance as follows. First we define h-weight of a matrix $A \in \mathbb{F}_q^{m\times n}$ as

$$|A|_h = \min_{\mathcal{I}(A)} \sum_{i \in \mathcal{I}(A)} h_i. \tag{6}$$

Theorem 1 *The defined h-weight satisfies the axioms of a seminorm, i.e., for every $A, B \in \mathbb{F}_q^{m\times n}$ holds*

1) $|A|_h \geq 0$,
2) $|A|_h = |-A|_h$,
3) $|A - B|_h \leq |A|_h + |B|_h$.

Proof: The first two properties follow immediately from definition (6). Let us prove the third one. Indeed, $\mathcal{I}(A) \cup \mathcal{I}(B)$ covers $A - B$. Hence

$$|A - B|_h = \min_{\mathcal{I}(A-B)} \sum_{i \in \mathcal{I}(A-B)} h_i \leq \min_{\mathcal{I}(A), \mathcal{I}(B)} \sum_{i \in \mathcal{I}(A) \cup \mathcal{I}(B)} h_i$$

and since $h_i \geq 0$

$$\leq \min_{\mathcal{I}(A)} \sum_{i \in \mathcal{I}(A)} h_i + \min_{\mathcal{I}(B)} \sum_{i \in \mathcal{I}(B)} h_i = |A|_h + |B|_h.$$

■

Notice that the h-weight does not satisfy the axiom of positive definiteness, i.e. the axiom $|A|_h = 0$ iff $A = 0$ does not hold. For example, if $h = 0$ then $|A|_h = 0$ for every matrix A.

Let us modify the h-wight (6) by multiplying it by 2 and adding a fixed (for a fixed h) positive number $m + n - \sum h_i$. The new h-norm remains to be a seminorm. As a result, we obtain the following new definition, which corresponds to a traditional definition of generalized distance.

Definition 1 *For a given vector h of reliabilities and for matrices $A, B \in \mathbb{F}_q^{m\times n}$ a seminorm (or h-norm) $|A|_h$ is defined as follows.*

$$|A|_h = \min_{\mathcal{I}(A)} \left(\sum_{i \in \mathcal{I}(A)} (1 + h_i) + \sum_{i \notin \mathcal{I}(A)} (1 - h_i) \right). \tag{7}$$

A generalized array semidistance (or h-distance) between matrices A and B is defined as

$$d_h(A, B) = |A - B|_h. \tag{8}$$

Notice, for $h = (1, \ldots, 1)$ the h-distance coincides with doubled array distance. For a given h the h-distance $d_h(\mathcal{C})$ of a code $\mathcal{C}$ is defined as the minimum h-distance between two different codewords. For a linear code, h-distance of the code is the minimum h-norm of a nonzero codeword.

Theorem 2 *If array distance of the code $\mathcal{C}$ is $d^{(a)}(\mathcal{C}) = d$ then the minimum h–distance of $\mathcal{C}$ over all h is*

$$\min_h d_h(\mathcal{C}) = d.$$

Proof:

$$\min_h d_h(\mathcal{C}) = \min_h \min_{C:w^{(\mathrm{a})}(C)\geq d} |C|_h = \min_{C:w^{(\mathrm{a})}(C)\geq d} \min_h |C|_h$$
$$= \min_{C:w^{(\mathrm{a})}(C)\geq d} w^{(\mathrm{a})}(C) = d.$$

■

Theorem 2 explains why we modified the definition of the h-norm.

B. Generalized minimum distance decoder

Given a received matrix Y and reliability vector h, the *goal* of the Generalized Minimum Distance (GMD) decoder is to find the list $\mathcal{L}$ of codewords C which are at the minimum h-distance $d_h(Y,C)$ from the received vector Y, i.e., to decode the code $\mathcal{C}$ in the generalized metric.

The guaranteed error correcting radius ρ of a particular GMD decoder is the infimum of real numbers $\tilde{\rho}$, for which there exist two matrices $C \in \mathcal{C}$, $Y \in F_q^{m\times n}$ and a vector $h \in [0,1]^n$, such that $d_h(Y,C) = \tilde{\rho}$, and the GMD decoder fails to decode Y, h, i.e., it outputs a list, which does not contain C. In other words, we guarantee correction of every error of generalized weight less than ρ, where the generalized error-weight is defined to be $|Y-C|_h = d_h(Y,C)$. It follows from Theorem 2 that the error-correcting radius ρ of GMD decoder can not be greater than d.

C. Generalized distance matches the array-error channel

Let $p_i, i = 1,\ldots,m+n$, be the a posteriori probability that the ith line was selected by the channel to be erroneous. Then joint probability that the ith line of length $l_i, l_i \in \{m,n\}$, was selected by the channel and filled by particular l_i symbols from $\mathbb{F}_q$ is $\tilde{p}_i = p_i q^{-l_i}$. Given the received matrix Y and the vector of probabilities $p = (p_1,\ldots,p_{n+m})$, for every codematrix C probability of the error matrix $E = Y - C$ can be estimated neglecting the fact of line-intersection as follows

$$P(E) \approx \sum_{\mathcal{I}(E)} \prod_{i\in\mathcal{I}(E)} \tilde{p}_i \prod_{i\notin\mathcal{I}(E)} (1-p_i)$$
$$\approx \max_{\mathcal{I}(E)} \prod_{i\in\mathcal{I}(E)} \frac{\tilde{p}_i}{1-p_i} \prod_{i=1}^{m+n} (1-p_i). \qquad (9)$$

Denote the second product in (9) by $a(p)$ and

$$h_i = -\ln \frac{\tilde{p}_i}{1-p_i}. \qquad (10)$$

Using definition (6) we obtain

$$P(E) \approx a(p) \max_{\mathcal{I}(E)} \exp\left(-\sum_{i\in\mathcal{I}(E)} h_i\right)$$
$$= a(p) \exp\left(-\min_{\mathcal{I}(E)} \sum_{i\in\mathcal{I}(E)} h_i\right)$$
$$= a(p) \exp\left(-|E|_h\right). \qquad (11)$$

The maximum likelihood decoding rule becomes

$$\arg\max_{C\in\mathcal{C}} P(Y|C) = \arg\max_{C\in\mathcal{C}} P(Y-C) = \arg\min_{C\in\mathcal{C}} |Y-C|_h. \qquad (12)$$

Let us make a realistic assumption that $p_i \leq (1+q^{-l_i})^{-1} \triangleq p_i^{(\max)} \approx 1$. If the assumption is not satisfied then we can replace $p_i > p_i^{(\max)}$ by $p_i^{(\max)}$. Then from (10) it follows that $h_i \geq 0$. Notice that the result of decoding rule (12) will not change if we mutiply every h_i by a positive number. Denote $h_{\max} = \max\{h_i\}$ and divide every h_i by $h_{\max}$ then we have $0 \leq h_i \leq 1$. As a result, up to approximation in (9), maximum likelihood decoder coincides with generalized minimum distance decoder according to definition (6) and hence according to Definition 1 as well, since the result of decoding rule (12) will not change if we replace definition (6) by Definition 1.

V. Forney-Kovalev (FK) decoding

To implement GMD decoding we use the FK algorithm. Given an array-error-and-erasure decoder Φ of the code $\mathcal{C}$, the *FK list decoding* is as follows. For $j = 1,\ldots,s$ we make a trial to decode the received matrix Y in which the τ_j least reliable lines are erased. Performing s decoding trials using decoder Φ we obtain a list $\mathcal{L}$ of codewords. If this list is empty, we declare a decoding failure, otherwise we leave in the list only codewords C having the minimum $d_h(C,Y)$ and output the new list. FK decoders may differ by using different decoders Φ (having different λ) or by different number s of decoding trials or by different selection of the erasure vector $\tau = (\tau_1,\ldots,\tau_s)$. If the erasure vector is fixed we get the Forney algorithm. If the erasure vector is selected adaptive depending on the received vector h of reliabilities, we obtain the Kovalev algorithm, having better performance. Later we consider the adaptive approach only.

Let us estimate the guaranteed error correcting radius ρ of the adaptive FK algorithm. Recall that we consider a FK decoder based on an array-error-correcting algebraic decoder Φ which satisfies (1) with tradeoff rate λ. At the input of the FK decoder we have a received word Y and a vector of reliabilities h. From now on, assume w.l.o.g. that the lines of matrices Y and C are ordered according to their reliabilities as follows

$$0 \leq h_1 \leq h_2 \leq \cdots \leq h_{m+n} \leq 1. \qquad (13)$$

So, we denote by $h = (h_1,...,h_{m+n})$ the vector of *ordered* reliabilities, and by $\mathcal{H}$ the set of all possible real-valued vectors h satisfying (13).

Definition 2 *Given the vector h of reliabilities, by $\delta_\tau(h)$ we denote the minimum h-weight of the error in the channel that causes a failure of the FK decoder with erasing strategy defined by the vector τ. In other words, $\delta_\tau(h)$ is error-correcting radius for fixed h and τ.*

Lemma 3 *Error-correcting radius $\delta_\tau(h)$ is as follows*

$$\delta_\tau(h) = \sum_{j=1}^{m+n}(1-h_j) + 2\sum_{i=1}^{s}\sum_{j=\tau_i+1}^{\tau_i+\varepsilon(\tau_i)-\varepsilon(\tau_{i+1})} h_j, \qquad (14)$$

where we denote the function

$$\varepsilon(\theta) = \left\lfloor \frac{d-\theta-1}{\lambda} \right\rfloor + 1,$$

and τ_{s+1} is formally defined such that $\varepsilon(\tau_{s+1}) = 0$.

Let $\mathcal{T}$ be the set of all integer valued vectors $\tau = (\tau_1, ..., \tau_s)$ such that $0 \le \tau_1 \le \cdots \le \tau_s \le d-1$. To specify a particular FK decoder we are free to select a vector τ. For a given h we will select τ to maximize the error-correcting radius $\delta_\tau(h)$:

$$\tau(h) = \arg\max_{\tau\in\mathcal{T}} \delta_\tau(h). \qquad (15)$$

The algorithm with this $\tau(h)$ we will call *adaptive algorithm* and denote by A. The error correcting radius $\rho_A(\lambda)$ of algorithm A is

$$\rho_A(\lambda) = \inf_{h\in\mathcal{H}} \max_{\tau\in\mathcal{T}} \delta_\tau(h). \qquad (16)$$

To find vector $\tau(h)$ from (15) one should consider $|\mathcal{T}|$ vectors τ, thus the complexity of this step is $\mathcal{O}(d^s)$. Remark, that the decoder should compute $\tau(h)$ for every received h, thus the computation is only feasible for one or two decoding trials, i.e., for $s = 1, 2$. This is a big disadvantage of the adaptive approach using the erasing vector (15).

VI. Decoding algorithm

Fortunately Kovalev suggested a simplification of the adaptive decoding algorithm where vector of erasures $\tau(h)$ should be selected from a set of two vectors only! In [7] this simplified algorithm was extended for all the range of λ and the final decoder is given by Algorithm 1. To compute $\tau(h)$ Algorithm 1 requires $\mathcal{O}(d)$ operations only. Error-correcting radius $\rho_A(\lambda)$ of the initial algorithm A based on $\tau(h)$ given by (15) and radius of the simplified Algorithm 1 coincide!

Theorem 4 ([7]) *The error correcting radius of Algorithm 1 is lower bounded by $\underline{\rho}_A(\lambda)$*

$$\rho_A(\lambda) \ge \underline{\rho}_A(\lambda) = \varepsilon(0) + \varepsilon(\tau_1), \qquad (17)$$

where τ_1 is a solution of recurrent inequalities

$$\tau_i \ge \tau_{i-1} + \varepsilon(\tau_{i-1}) - \varepsilon(\tau_{i+1}), \quad i = 1, \ldots, 2s-1, \qquad (18)$$

with boundary conditions

$$\tau_0 = 0, \quad \tau_{2s} = \lfloor d - 1 + \lambda \rfloor. \qquad (19)$$

The lower bound (17) is nearly tight [7] and can be approximated as follows.

Corollary 5 *For $1 < \lambda < 2$ s-trial decoding radius is*

$$\underline{\rho}_A(\lambda) \approx d\left(1 - \frac{(2-\lambda)(\lambda-1)^{2s}}{\lambda(1-(\lambda-1)^{2s})}\right) \approx d\left(1-(\lambda-1)^{2s}\right). \qquad (20)$$

Algorithm 1: Simplified s-trial adaptive decoding

Precomputations: Solve (18), get vectors $\tau^{(a)} = (\tau_0, \tau_2, ..., \tau_{2(s-1)})$ and $\tau^{(b)} = (\tau_1, \tau_3, ..., \tau_{2s-1})$;
Input: received matrix Y and (ordered) vector h;
Select vector $\tau' = \arg\max_{\tau\in\{\tau_a,\tau_b\}} \delta_\tau(h)$;
for *each j from 1 to s* **do**
  decode Y with erased first τ'_j positions by the decoder Φ of the code $\mathcal{C}$, add obtained codeword (if any) to the list $\mathcal{L}$;
Output:
if *the list $\mathcal{L}$ is empty* **then**
  declare a decoding failure;
else
  leave in $\mathcal{L}$ only codewords nearest to Y in h-metric, output $\mathcal{L}$

To reach $\rho_A(2) = d$ it is sufficient to have $s = \frac{1}{2}\left(\log_{\frac{1}{\lambda-1}} d + 1\right)$ decoding trials.

Corollary 6 *For $\lambda = 2$ s-trial decoding radius is*

$$\rho_A(2) \ge d + 1 - \left\lceil \frac{d+1}{4s} \right\rceil, \qquad (21)$$

which coincides with Kovalev's result. To reach $\rho_A(2) = d$ it is sufficient to have $s = \lceil \frac{d+1}{4} \rceil$ decoding trials.

Notice, to reach $\rho_A(2) = d$, the number s of decoding trials grows linearly with d for the classical case $\lambda = 2$ and only logarithmically for $\lambda < 2$. As a result, for $\lambda < 2$ the error-correcting radius of Algorithm 1 quickly approaches d with increasing number of trials, and 2 or 3 trials are sufficient to reach $\rho_A(2) = d$ in many practical cases.

References

[1] E.M. Gabidulin, B.I. Korjik, "Lattice-error-correcting codes," Izv. Vyssh. Uchebn. Zaved., Radioelektron.,15, no. 4,492-498, 1972.

[2] E.M. Gabidulin, "Theory of codes with maximum rank distance,", Probl. Inform. Transm. 21(1), pp. 3-16, 1985.

[3] R.M. Roth, "Maximum-rank array codes and their application to criss-cross error correction," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 328-336, Mar. 1991.

[4] G. D. Forney Jr., "Generalized minimum distance decoding," *IEEE Trans. Inf. Theory*, vol. 12, pp. 125-131, Apr. 1966.

[5] S. I. Kovalev, "Two classes of minimum generalized distance decoding algorithms," *Probl. Pered. Inform.*, vol. 22, no. 3, pp. 35-42, 1986.

[6] J. H. Weber, K. A. S. Abdel-Ghaffar, "Reduced GMD decoding," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1013-1027, April 2003.

[7] V. Sidorenko, A. Chaaban, Ch. Senger, M. Bossert, On Extended Forney-Kovalev GMD decoding, IEEE International Symposium on Information Theory, June-July, 2009, Seoul, Korea.

[8] V. R. Sidorenko, G. Schmidt, M. Bossert, "Decoding punctured Reed-Solomon codes up to the Singleton bound," in *Proc. of Int. ITG Conference on Source and Channel Coding*, Ulm, January 2008.

[9] G. Schmidt, V. R. Sidorenko, and M. Bossert, "Collaborative decoding of interleaved Reed–Solomon codes and concatenated code designs,"*IEEE Trans. Inf. Theory*, vol. 55, n. 7, pp. 2991-3012, July 2009.