

Polytope of Correct (Linear Programming) Decoding and Low-Weight Pseudo-Codewords

Michael Chertkov

CNLS and Theoretical Division at LANL
& New Mexico Consortium,
Los Alamos, NM 87545, USA
Email: chertkov@lanl.gov

Mikhail Stepanov

Department of Mathematics
University of Arizona
Tucson, AZ 85721, USA
Email: stepanov@math.arizona.edu

Abstract— We analyze Linear Programming (LP) decoding of graphical binary codes operating over soft-output, symmetric and log-concave channels. We show that the error-surface, separating domain of the correct decoding from domain of the erroneous decoding, is a polytope. We formulate the problem of finding the lowest-weight pseudo-codeword as a non-convex optimization (maximization of a convex function) over a polytope, with the cost function defined by the channel and the polytope defined by the structure of the code. This formulation suggests new provably convergent heuristics for finding the lowest weight pseudo-codewords improving in quality upon previously discussed. The algorithm performance is tested on the example of the Tanner [155, 64, 20] code over the Additive White Gaussian Noise (AWGN) channel.

I. INTRODUCTION

Low-Density Parity Check (LDPC) codes are capacity achieving (in the thermodynamic limit) and are easy to decode via message-passing algorithms of the Belief-Propagation (BP) type [1], [2], [3]. However, performance of the efficient decoder on a given finite code is not ideal, resulting in a sizable difference between optimal (Maximum A-Posteriori) and the suboptimal decoders observed in the asymptotics of Bit-Error-Rates (BERs) at the high Signal-to-Noise-Ratios (SNR), in the *error floor* regime [4]. Errors in this extreme regime of the *error floor* are mainly due to special configurations of the channel noise, called instantons [5], correspondent to decoding into pseudo-codewords [6], [7] different from any of the codewords of the code. Analysis of the instantons and pseudo-codewords in the case of LP decoder [8] is of a special interest. LP is a combinatorial (zero-temperature) version of BP, thus admitting convenient description in terms of the pseudo-codeword polytope [8]. The geometric structure associated with the polytope gave rise to new decoding techniques related to graph covers [9], adaptive processing of the polytope constraints [10], and the concept of LP duality [11]. The succinct combinatorial formulation of the coding was also useful in terms of improving LP and thus reducing the gap between the LP and MAP decoders [12], [13], [14], [15], [16].

In [17] we suggested an LP-specific heuristic Pseudo-Codeword Search (PCS) algorithm. The main idea of the algorithm was based on exploring the Wiberg relation, from [6], [7], between pseudo-codeword and an optimal noise

configuration which lies on the median between the pseudo-codeword and zero-codeword. In essence, the algorithm of [17] performs a biased walk over the exterior of the domain of correct LP decoding (surrounding zero codeword) and arrives at the error-surface (boundary of the domain) in a small finite number of steps. The algorithm, tested on some number of codes over the AWGN channel, showed excellent performance. For any noise initiation it always approaches the error-surface monotonically in simulations, even though the monotonicity proof was not provided. Latter the algorithm was generalized to the case of discrete-output channel (specifically Binary Symmetric (BS) channel) in [18], [19], where the monotonicity proof was given. The technique was also extended to discover the most probable configurations of error-vectors in compressed sensing [20].

This paper continues the trend of [17] and analyzes the error-surface and the associated low-weight pseudo-codewords. We study the domain of correct decoding, bounded by the error-surface; formulate the (channel specific) problem of finding the most probable configuration of the noise leading to a failure (and respective pseudo-codeword) as an optimization problem; design an efficient heuristic; and illustrate performance of the algorithm on the exemplary Tanner [155, 64, 20] code [21]. The main statements of the manuscript are:

- The domain of correct decoding is a polytope in the noise space. For a typical code the polytope is likely to be non-tractable, *i.e.*, requiring description exponential in the code size. [Section III.]
- The problem of finding the lowest weight pseudo-codeword of a graphical code over log-concave symmetric (for example AWGN) channel is reduced to maximization of a convex function, associated with the channel, over a polytope, associated with the code and defined as the cross-section of the decoding polytope by a plane. [Section IV.]
- We suggested Majorization Optimization Algorithm (MOA), based on majorization-minimization [22] approximation of the aforementioned optimization formulation. We showed that MOA, as well as previously introduced PCS, are both monotonic in discovering iteratively the low-weight pseudo-codewords (effective

weight decreases with iterations). [Section V.] Performances of MOA and PCS are tested on the Tanner code over AWGN channel in Section VI.

II. PRELIMINARY DISCUSSIONS AND DEFINITIONS

We consider LP decoding [8] of binary LDPC code and discuss the problem of finding the most probable configuration of the noise, so-called instanton, for which the decoding fails [17]. Equivalently stated, this is the problem of finding the lowest weight (closest to the zero codeword) pseudo-codeword of the code.

The technique we discuss here applies to any soft-output, symmetric channels where the transition probability, $\mathcal{P}(\mathbf{x}|\boldsymbol{\sigma})$, from the codeword $\boldsymbol{\sigma}$ to the channel output \mathbf{x} , is a log-convex function of \mathbf{x} , i.e., $-\log(\mathcal{P}(\mathbf{x}|\boldsymbol{\sigma}))$ is a convex function of \mathbf{x} . AWGN channel is our enabling example with

$$\mathcal{P}(\mathbf{x}|\boldsymbol{\sigma}) \propto \exp\left(-2s^2 \sum_{i=1}^N (x_i - \sigma_i)^2\right), \quad (1)$$

where s is the signal-to-noise ratio of the noise, $\boldsymbol{\sigma} = (\sigma_i = 0, 1 | i = 1, \dots, N)$, is the binary N -bits long codeword launched into the channel, and $\mathbf{x} = (x_i \in \mathbf{R} | i = 1, \dots, N)$ is the real valued signal received by the decoder.

Maximum Likelihood decoding can be formulated as an LP optimization over the polytope, \mathcal{P} , spanned by all the codewords of the code C ,

$$\min_{\boldsymbol{\sigma}'} \sum_i (1 - 2x_i) \sigma'_i \Big|_{\boldsymbol{\sigma}' \in \mathcal{P}}. \quad (2)$$

However, the full codeword polytope is exponentially large in the code size and thus it is not tractable. Trading optimality for efficiency the authors of [8] have suggested to relax the full polytope into a tractable one (stated in terms of a polynomial, in the size of the code, number of constraints). The relaxation, coined LP-decoding, is based on decomposition of the code into small individual checks based codes thus assuring (by construction) that the set of original codewords forms a subset of all the corners of the relaxed polytope (so-called set of pseudo-codewords). The LP-decoding can be formulated in multiple ways. Following [17], we choose to start here with the formulation of LP, correspondent to the so-called zero-temperature version of the Bethe Free Energy approach of [23]:

$$LP_p(\mathbf{x}) = \min_b \sum_i (1 - 2x_i) \sum_{\sigma_i=0,1} \sigma_i b_i(\sigma_i) \Big|_{b \in \mathcal{P}_l}, \quad (3)$$

$$\mathcal{P}_l = \left\{ \begin{array}{l} \forall i: \sum_{\sigma_i} b_i(\sigma_i) = 1; \forall \alpha: \sum_{\sigma_\alpha} b_\alpha(\sigma_\alpha) = 1; \\ \forall i, \forall \alpha \sim i: \sum_{\sigma_i} (1 - 2\sigma_i) b_i(\sigma_i) \\ \quad = \sum_{\sigma_\alpha} (1 - 2\sigma_i) b_\alpha(\sigma_\alpha); \\ \forall i: b_i(\sigma_i) \geq 0; \forall \alpha: b_\alpha(\sigma_\alpha) \geq 0 \end{array} \right\},$$

where b are beliefs, i.e., proxies for respective marginal probabilities. \mathcal{P}_l is a polytope, which we call large (LP-decoding) polytope. \mathcal{P}_l only depends on the structure (graph) of the code (and it does not depend on the channel model). There are beliefs of two types associated with two types of nodes

in the parity check graph of the code, \mathcal{G} , bits i and checks α respectively. $\sigma_i = 0, 1$ represent values of the bit i , and the vector $\boldsymbol{\sigma}_\alpha = (\sigma_i | i \sim \alpha; \text{s.t. } \sum_i \sigma_i = 0 \bmod 2)$ stands for one of the allowed local codewords associated with the check α . Of the conditions in the definition of \mathcal{P}_l , the first two equalities are normalizations (for the beliefs/probabilities), the third equality states consistency between beliefs associated with bits and checks. The two last inequalities in \mathcal{P}_l ensure that the beliefs (probabilities) are positive. If the channel noise corrupting the zero codeword is sufficiently weak, i.e., if $|\mathbf{x}| \ll 1$, the LP_p outputs zero, corresponding to successful decoding. However, LP_p confuses another pseudo-codeword (typically non-integer) for the codeword if \mathbf{x} is sufficiently noisy, then giving a strictly negative output, $LP_p < 0$.

Description of the $LP_p(\mathbf{x})$ in Eq. (3) can be restated in terms of a smaller set of beliefs, only bit beliefs $\boldsymbol{\beta} = (\beta_i = b_i(1) | i = 1, \dots, N)$. Then the “small polytope” formulation of Eq. (3) becomes [24], [8]:

$$LP_p(\mathbf{x}) = \min_{\boldsymbol{\beta}} \sum_i (1 - 2x_i) \beta_i \Big|_{\boldsymbol{\beta} \in \mathcal{P}_s}, \quad (4)$$

$$\mathcal{P}_s = \left\{ \begin{array}{l} \forall \alpha \forall I \subseteq I_\alpha, |I| \text{ is odd: } \sum_{i \in I} \beta_i - \sum_{i \in I_\alpha \setminus I} \beta_i \leq |I| - 1 \\ \forall i: 0 \leq \beta_i \leq 1 \end{array} \right\},$$

where I_α is the subset of bit-nodes contributing check α .

The “large polytope” formulation of the LP-decoding (3) can also be restated in terms of its dual (the formulation here is almost identical to DLPD2 of [11])

$$LP_d(\mathbf{x}) = \max_{\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\lambda}} \sum_i \phi_i + \sum_{\alpha} \theta_{\alpha} \Big|_{\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\lambda} \in \mathcal{P}_d}, \quad (5)$$

$$\mathcal{P}_d = \left\{ \begin{array}{l} \forall i, \forall \sigma_i: \sigma_i(1 - 2x_i) - (1 - 2\sigma_i) \sum_{\alpha \sim i} \lambda_{i\alpha} \geq \phi_i \\ \forall \alpha, \forall \sigma_\alpha: \sum_{i \sim \alpha} \lambda_{i\alpha} (1 - 2\sigma_i) \geq \theta_{\alpha} \end{array} \right\},$$

where $\boldsymbol{\phi} = (\phi_i | i = 1, \dots, N)$, $\boldsymbol{\theta} = (\theta_{\alpha} | \alpha = 1, \dots, M)$, $\boldsymbol{\lambda} = (\lambda_{i\alpha} | (i, \alpha) \in \mathcal{G}_1)$ are Lagrangian multipliers (messages) conjugated to the first, second and third conditions in the original LP (3) respectively. According to the main (strong duality) theorem of the convex optimization (see many textbooks, e.g., [25]) the results of the primal problem (3) and the dual problem (5) coincide, $LP_p = LP_d$.

In this manuscript we are mainly concerned with the following practical problem: given a finite code, log-concave channel (for concreteness and without loss of generality we will consider AWGN channel as an example), and the LP-decoding (in its primal or dual versions), to find the most probable configuration (instanton) of the channel noise, \mathbf{x} , imposed on the zero codeword, $\boldsymbol{\sigma}_0 = \mathbf{0}$, which leads to incorrect decoding. Formally, we are solving the following “instanton” problem

$$\min_{\mathbf{x}} \sum_i x_i^2 \Big|_{\mathbf{x} \in \mathcal{D}_{ext}}, \quad (6)$$

where \mathcal{D}_{ext} is defined as an exterior (complement) of the domain, \mathcal{D}_{int} , correspondent to the correct decoding: $LP_p = LP_d = 0$. Thus, $\mathcal{D}_{ext} = \mathbf{R}^N \setminus \mathcal{D}_{int}$.

III. DOMAIN OF CORRECT DECODING IS A POLYTOPE

Let us show that \mathcal{D}_{int} is *actually* a polytope.

Consider the following auxiliary domain of $(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\lambda})$:

$$\mathcal{F}_d = \left\{ \begin{array}{l} \sum_i \phi_i + \sum_{\alpha} \theta_{\alpha} = 0 \\ \forall i, \forall \sigma_i: \sigma_i(1 - 2x_i) - (1 - 2\sigma_i) \sum_{\alpha \sim i} \lambda_{i\alpha} \geq \phi_i \\ \forall \alpha, \forall \sigma_{\alpha}: \sum_{i \sim \alpha} \lambda_{i\alpha}(1 - 2\sigma_i) \geq \theta_{\alpha} \end{array} \right\},$$

constructed from the feasibility region of the dual problem, LP_d , with the zero cost function constraint added. For any $\mathbf{x} \in \mathcal{D}_{int}$ there obviously exists an extended configuration $(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\lambda})$ from \mathcal{F}_d . On the other hand, if $\mathbf{x} \in \mathcal{D}_{ext}$, then $LP_p = LP_d < 0$ (i.e., a pseudo-codeword, different from the zero codeword, is selected by the LP), and since LP_d is defined as a maximum over an extension of \mathcal{F}_d (where the first condition in \mathcal{F}_d is removed) there exists no valid $(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\lambda})$ from \mathcal{F}_d in this case. One concludes that $\mathcal{D}_{int}(\mathbf{x})$ coincides with the projection of \mathcal{F}_d on the \mathbf{x} variable

$$\mathcal{D}_{int} = \text{Proj}(\mathcal{F}_d)_{\mathbf{x}} = \{\exists(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\lambda}) \text{ s.t. } (\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\lambda}) \in \mathcal{F}_d\}. \quad (7)$$

However both \mathcal{F}_d and its projection to \mathbf{x} are polytopes, i.e., \mathcal{D}_{int} is also a convex domain, moreover it is a polytope¹.

Note that the projected polytope is most likely non-tractable, in the sense that the number of constraints required to describe the polytope is expected to be exponential in the dimension of \mathbf{x} (size of the code).

IV. SEARCH FOR LOWEST WEIGHT PSEUDO-CODEWORD AS AN OPTIMIZATION

Noticing, that Eq. (6) is stated in terms of the exterior domain, \mathcal{D}_{ext} , which is a compliment of \mathcal{D}_{int} , one attempts to formulate a closely related problem stated in terms of optimization over a convex sub-domain of \mathcal{D}_{int} :

$$Q(\epsilon) = \min_{\mathbf{x}} LP(\mathbf{x}) \Big|_{\mathbf{x} \in \text{Ball}_{\epsilon}}, \quad (8)$$

where $\text{Ball}_{\epsilon} \equiv \{\boldsymbol{\zeta} \in \mathbf{R}^N: \|\boldsymbol{\zeta}\|_2 \leq \epsilon\}$ is the ball of radius ϵ (which is convex by construction). For sufficiently small ϵ any $LP(\mathbf{x}) = 0$ for any $\mathbf{x} \in \text{Ball}_{\epsilon}$, while a gradual increase in ϵ will eventually lead, at some ϵ_* , to appearance of the closest to the zero codeword (in terms of the l_2 norm of the AWGN channel) noise configuration, \mathbf{x}_{inst} , for which $LP(\mathbf{x}_{inst}) \leq 0$. One concludes that the function of a single parameter, $Q(\epsilon)$, jumps from zero at $\epsilon < \epsilon_*$ to some negative value at $\epsilon = \epsilon_*$. Then, $4\epsilon_*^2$ becomes the effective distance of the code (under the LP-decoding), and the optimal value, \mathbf{x}_* of $Q(\epsilon_*)$, corresponds to the most probable instanton.

Using primal formulation of LP-decoding from Eq. (4) and combining minimization over \mathbf{x} and $\boldsymbol{\beta}$ variables, one reformulates Eq. (8) as the following optimization problem

$$Q(\epsilon) = \min_{\boldsymbol{\beta}, \mathbf{x}} \sum_i (1 - 2x_i) \beta_i \Big|_{\mathbf{x} \in \text{Ball}_{\epsilon}, \boldsymbol{\beta} \in \mathcal{P}_s}. \quad (9)$$

¹We are thankful to P. Vontobel for pointing out, after reading the first version of the manuscript, that the statement above is closely related to these made in [26]. See Fig. 11,12 of [26] as well as preceding and following discussions.

One important advantage of this formulation is in the fact that Eq. (9) is stated as an optimization problem, in contrast with the sequential instanton search optimization of [17], where one optimizes over the noise, then evaluates an internal minimization (the LP decoding itself) for each configuration of the noise. Note that the cost function in Eq. (9) is quadratic and concave.

Eq. (9) can be simplified further. We expect that the extremal value will be achieved (at least for sufficiently large ϵ) “at the surface” of the ball, i.e., at $\sum_i x_i^2 = \epsilon^2$. Replacing $\mathbf{x} \in \text{Ball}_{\epsilon}$ by this equality and performing optimization over \mathbf{x} , we arrive (with the help of the standard Lagrangian multiplier technique, and also assuming that all components of the candidate noise vector are positive) at the following nonlinear optimization problem stated primarily in terms of the beliefs

$$Q(\epsilon) = \min_{\boldsymbol{\beta}} \left(\sum_i \beta_i - 2\epsilon \sqrt{\sum_i \beta_i^2} \right) \Big|_{\boldsymbol{\beta} \in \mathcal{P}_s}. \quad (10)$$

This problem can be solved approximately (but efficiently) via the majorization-minimization iterative method [22], consisting in upper-bounding the cost function by its linearized expression, minimizing the upper-bound, and iterating by shifting the linearization point to the solution received on the previous step. The linearization (for majorization) at each iterative step is justified because of the following obvious inequality

$$\|\boldsymbol{\beta}\|_2 \geq L(\boldsymbol{\beta}; \boldsymbol{\beta}^{(k)}) = (\boldsymbol{\beta} \cdot \boldsymbol{\beta}^{(k)}) / \|\boldsymbol{\beta}^{(k)}\|_2, \quad (11)$$

which holds for any $\boldsymbol{\beta}$. Then the iterative solution of Eq. (10) becomes

$$Q^{(k+1)}(\epsilon) = \min_{\boldsymbol{\beta}} \left(\sum_i \beta_i - 2\epsilon L(\boldsymbol{\beta}; \boldsymbol{\beta}^{(k)}) \right) \Big|_{\boldsymbol{\beta} \in \mathcal{P}_s}, \quad (12)$$

where $k = 0, 1, \dots$ till convergence, $\boldsymbol{\beta}^{(k)}$ is the optimal solution of the optimization found at the k iteration step, and $\boldsymbol{\beta}^{(k+1)}$ becomes the optimal solution at the $(k+1)$ -th iteration. The optimization problem on the rhs of Eq. (12) is an LP, i.e., it can be solved efficiently. We also expect that the iterations over k converge fast. The iterative procedure will depend on the initiation set at $k = 0$, and starting from different initial conditions we sample different local optima.

Note that ϵ_* , defined as the smallest ϵ for which $Q(\epsilon)$ becomes negative, allows useful interpretation in terms of the effective distance of the corresponding pseudo-codeword. Indeed, $4\epsilon_*^2 = w(\boldsymbol{\beta})$, where

$$w(\boldsymbol{\beta}) = \frac{(\sum_i \beta_i)^2}{\sum_i \beta_i^2}. \quad (13)$$

is the weight of the noise (and of the corresponding pseudo-codeword) according to the Wiberg formula from [6], [7], expressing relation between the direction of the optimal noise and the distance along the direction from the zero codeword to the error-surface.

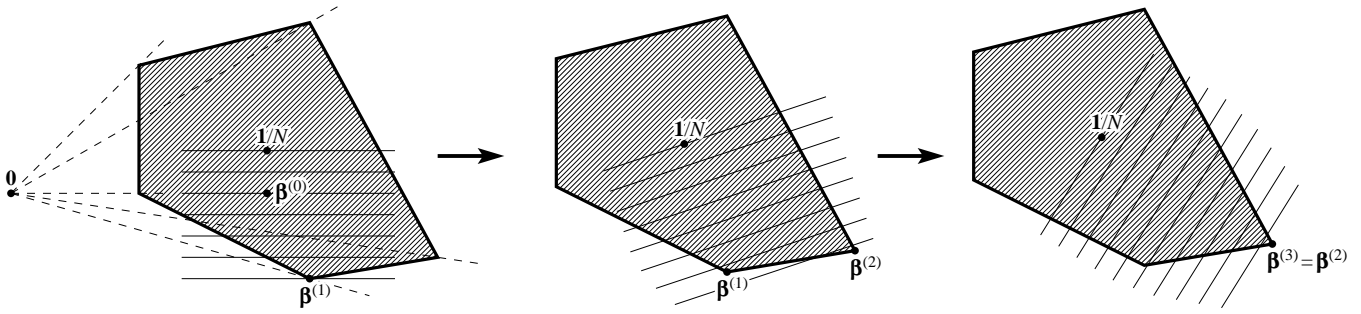


Fig. 1. The Figure illustrates the sequential progress (from left to right) of the majorization-minimization procedure. The shaded area corresponds to $\mathcal{P}_{\text{cone}}$. Dashed lines on the left sub-figure show edges of \mathcal{P}_s containing the origin $\mathbf{0}$. The optimization starts from $\beta^{(0)}$. Thin solid lines are the level curves of the linear function $L(\beta; \beta^{(k)})$, which optimum over β results in $\beta^{(k+1)}$. The procedure continues till convergence, $\beta^{(k+1)} = \beta^{(k)}$ (achieved with $k = 2$ in the illustration).

V. CONE FORMULATION AND MAJORIZATION OPTIMIZATION ALGORITHM

To utilize Eq. (10) for finding the low-weight pseudo-codewords one needs to scan over the values of ϵ , thus making one-parametric optimization (over ϵ) in addition to the (multi-dimensional) optimization contained in Eq. (10). The main result of this Subsection is that this additional degree of freedom in the optimization is unnecessary, thus leading to a simplification of Eq. (10).

Let us first show that: *the vertex of \mathcal{P}_s , correspondent to the pseudo-codeword with the lowest weight, is connected by an edge to the vertex correspondent to the zero-codeword.*

Since the weight-function, $w(\beta)$ from Eq. (13), does not depend on the length of the vector β , one considers β , as the direction in the respective space pointing from the origin, $\mathbf{0} = (0, \dots, 0)$, to a point within the polytope \mathcal{P}_s . It is convenient to parameterize the direction in terms of the projection to the $\sum_i \beta_i = 1$ plane. Pseudo-codewords correspond to special values of the β vector projected to the plane, and to find the pseudo-codeword with the minimum weight we will need to minimize the weight, $w(\beta) = 1/\sum_i \beta_i^2$, over the cross-section of the polytope by the plane (projection). One restates the problem as maximization of $\sum_i \beta_i^2$, which is also equivalent to finding β maximizing the distance to the central point of the plane within the polytope \mathcal{P}_s , $\mathbf{1}/N = (1, \dots, 1)/N$. \mathcal{P}_s is projected through the origin to the plane, thus forming a polytope too (call it cone polytope)

$$\mathcal{P}_{\text{cone}} = \left\{ \begin{array}{l} \forall i: \beta_i \geq 0 \\ \forall \alpha \forall i \sim \alpha: \beta_i \leq \sum_{j \sim \alpha, j \neq i} \beta_j \\ \sum_i \beta_i = 1 \end{array} \right\}. \quad (14)$$

(The projection is understood in the standard projective space sense, with a line connecting a point within the polytope with the point of origin, $(0, \dots, 0)$, projecting to the point where the line crosses the plane.) Note that only faces of \mathcal{P}_s in Eq. (4) with $|I| = 1$ become faces of the cone polytope, $\mathcal{P}_{\text{cone}}$. Further, maximum of $\sum_i \beta_i^2$ is attained at some vertex of the polytope. By construction this vertex corresponds to an edge connecting the point of origin, $(0, \dots, 0)$ with another

vertex of the original polytope \mathcal{P}_s , correspondent to a pseudo-codeword with the lowest weight. All the other vertexes of \mathcal{P}_s , which are not connected to the origin, are projected to interior points of the cone polytope $\mathcal{P}_{\text{cone}}$, thus showing a higher value of the weight.

The choice of the cone cross-section in Eq. (14) is convenient for the purpose of simplifying the optimization problem (10). It guarantees that the first term in the objective of Eq. (10) is constant, and thus the term is inessential for the purpose of optimization. In the result, we arrive at the following reduced version of Eq. (10) (one less degree of freedom and simpler polytope)

$$\tilde{Q} = \max_{\beta} \sqrt{\sum_i \beta_i^2} \Big|_{\beta \in \mathcal{P}_{\text{cone}}}. \quad (15)$$

According to the discussion above, solution of Eq. (15) only describes the optimal direction in the noise space, x , and the respective length is reconstructed from the weight relation $4\epsilon_*^2 = w(\beta)$. Thus our final expression for the optimal noise (instanton), correspondent to the (optimal) solution of Eq. (15) is

$$x = \beta \frac{\sum_i \beta_i}{2 \sum_i \beta_i^2}. \quad (16)$$

The geometrical essence of the cone construction and of the majorization-minimization procedure is illustrated in Fig. 1.

Few remarks are in order. First, note that there is some additional freedom in choosing the objective function in the optimization over β . For example, one can replace, $\sqrt{\sum_i \beta_i^2}$, under the sum in Eq. (15) by $\sum_i (\beta_i - \sum_j \beta_j / N)^2$, and the resulting optimal β stays the same. Second, the majorization-minimization procedure of Eq. (12) for Eq. (10), extends straightforwardly to any appropriate choice of the objective function in the reduced optimization, in particular the choice of Eq. (15), thus resulting in the sequence

$$\beta^{(k+1)} = \arg \max_{\beta} \beta \cdot \beta^{(k)} \Big|_{\beta \in \mathcal{P}_{\text{cone}}}. \quad (17)$$

Third, the sequence (17) is monotonic by construction, *i.e.*, the effective distance can only decrease with the iteration number k , thus proving convergence.

The considerations above suggest the following *Majorization-Optimization Algorithm* (MOA):

- **Start:** Initiate a point $\beta^{(0)}$ inside the cone cross-section $\mathcal{P}_{\text{cone}}$ with a random deviation from the $(1, 1, \dots, 1)/N$. [The sampling step.]
- **Step 1:** Construct a linear function with the gradient vector pointing from $(1, 1, \dots, 1)/N$ to $\beta^{(k)}$, optimize it inside $\mathcal{P}_{\text{cone}}$ according to Eq. (17), and get the new $\beta^{(k+1)}$. [The majorization-minimization step.]
- **Step 2:** If $\beta^{(k+1)} \neq \beta^{(k)}$, then go to **Step 1**.
- **End:** Output the optimal noise configuration according to Eq. (16).

Like PCS of [17], MOA is sensitive to the choice of the initial direction in the β space, and this clarifies importance of repeating sampling step multiple times. Obviously, an individual sampling event outputs only pseudo-codewords sharing an edge in \mathcal{P}_s with the zero-codeword, call them “nearest-neighbors”, thus ignoring other pseudo-codewords, for example these which are “next-nearest-neighbors” to the zero codeword, *i.e.*, ones sharing an edge with a pseudo-codeword which shares an edge with the zero-codeword. Even though the effective distance of these “next-nearest-neighbors” may be smaller than the effective distance of some of the “nearest-neighbors”, MOA guarantees that the exact solution of Eq. (15) can only be a “nearest-neighbor”.

In the remainder of the Section let us briefly compare MOA with PCS. The iterative procedure of PCS is analogous to Eq. (17) and it can be restated as

$$\beta^{(k+1)} = \arg \max_{\beta} \beta \cdot \underbrace{\left(\beta^{(k)} \frac{\sum_i \beta_i^{(k)}}{\sum_i (\beta_i^{(k)})^2} - 1 \right)}_{-h=2x-1} \Big|_{\beta \in \mathcal{P}_s}. \quad (18)$$

Note, that, $\partial w(\beta)/\partial \beta = (2 \sum_i \beta_i)(\sum_i \beta_i^2) \cdot h$, so clearly, PCS aims to approximate $w(\beta)$, linearly inside the polytope, \mathcal{P}_s . The function $w(\beta)$ is a homogeneous function of degree 0. MOA takes advantage of this fact and attempts to minimize $w(\beta)$ in the projective space of β , indexed by the points of $\mathcal{P}_{\text{cone}}$. The value of max in (18) is non-negative, and it is exactly zero at $\beta = \beta^{(k)}$. If $w(\beta) < N$, vector $\partial w(\beta)/\partial \beta$ points away from the central direction $\mathbf{1}$, and thus minimization (18) is not going to increase $w(\beta)$, *i.e.*, under this (weak and easy to realize) condition the PCS is provably monotonic. Also, as $\beta \cdot (\partial w(\beta)/\partial \beta) = 0$, PCS, like MOA, always converges to vertices of \mathcal{P}_s which are the “nearest-neighbors” of the zero-codeword (the cone origin).

Since PCS works with $\partial w(\beta)/\partial \beta$, and not directly with $w(\beta)$ like MOA, it “confuses” $w(\beta)$ for being a homogeneous function of degree 1. Therefore, compared to MOA, PCS has an additional bias away from the cone origin, thus suggesting that its convergence is slower and resulting end-points being further away from the cone origin. This assessment is confirmed in the simulations of the next Section (see, *e.g.*, Fig. 2).

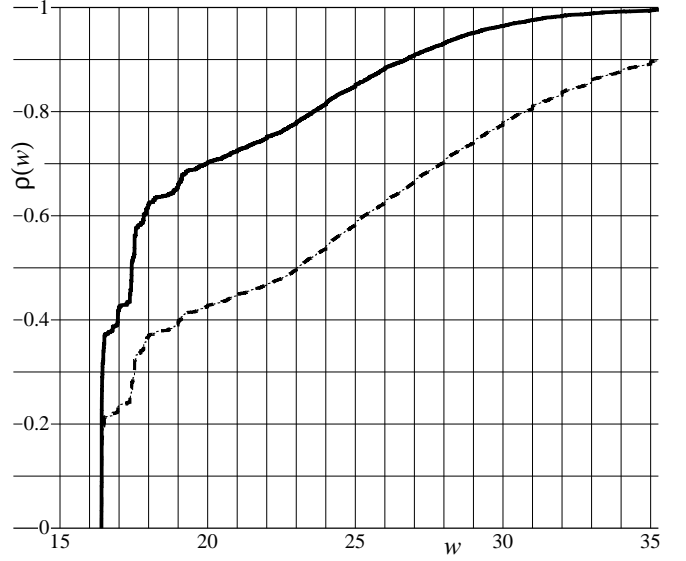


Fig. 2. The probability/frequency of occurrence, $p(w)$, of the pseudo-codewords with effective weight w or smaller for the Tanner [155,64,20] code [21]. Solid and dashed lines represent results of 10^4 trials of MOA and PCS algorithm of [17] respectively.

VI. TANNER CODE TEST

We tested MOA on the popular example of the Tanner [155,64,20] code [21]. The results are shown in Fig. 2. We analyzed effective distance, w , of the pseudo-codewords found in the result of 10^4 trials (different in initial orientation). As in the case of the PCS of [17], the probability (frequency), $p(w)$, of finding pseudo-codeword with effective distance smaller than w , grows monotonically with w . Like PCS, MOA result for the smallest effective distance of the code is, $w_{\min} \approx 16.4037 < 20$, where 20 is the Hamming distance of the code. However, we also observe that MOA is sampling the low-weight “nearest-neighbor” pseudo-codewords more efficiently than PCS, which is seen in a steeper dependence of $p(w)$ as a function of w in Fig. 2. As discussed above, we attribute the better performance of MOA to stronger bias towards the zero codewords convergence, as well as simpler and more homogeneous (in the low weight sector of the pseudo-codewords) initiation procedure.

VII. CONCLUSIONS AND PATH FORWARD

This paper reports new results related to analysis and algorithms discovering the lowest-weight pseudo-codeword(s) of the LP decoding of graphical codes performing over soft-output (log-concave) channels, like the AWGN channel. On the theoretical side, we show here that the set of correct decoding is a polytope in the space of noise. We also formulate the problem of finding the smallest weight noise (instanton) as an optimization problem, Eq. (15), looking for a maximum of a convex function over a convex set (a polytope). The exact solution of the problem is likely non-tractable, and we suggest heuristic iterative algorithmic solution based on the majorization-minimization approach of the optimization theory [22]. We show that convergence of

both MOA and PCS, introduced in [17], is monotonic. We also compare the algorithms in simulations on the standard example of the Tanner [155,64,20] code [21], and observe that MOA is superior in discovering the low-weight part of the pseudo-codeword spectrum.

We plan to extend this research in the future along the following directions:

- Test MOA on other and longer codes.
- Test MOA on other log-concave, but still binary, channels. We also envision extension of the technique to non-binary channels, especially these related to phase modulation in modern fiber optics [27].
- It will be useful to find a version of the majorization-minimization initiation which samples the “nearest-neighbor” pseudo-codewords uniformly, or (preferably) according to a given function of the effective weight.
- The LP-decoding is a close relative of the generally faster but more difficult to analyze iterative BP-decodings. It will be useful to extend the polytope theory and the MOA algorithm discussed in the paper to the case of iterative decodings, for example to the basic min-sum algorithm.
- Our major long-term goal consists in designing better graphical codes. We anticipate that MOA will be instrumental in searching over candidate codes (for example sampled from a properly expurgated ensemble of LDPC codes [3]) for the one showing the lowest error-floor possible.

VIII. ACKNOWLEDGMENTS

The work of MC at LANL was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396. MC acknowledges support of NMC via the NSF collaborative grant CCF-0829945 on “Harnessing Statistical Physics for Computing and Communications.” The work of MS is supported by NSF grant DMS-0807592 “Asymptotic Performance of Error Correcting Codes”.

REFERENCES

- [1] R. G. Gallager, *Low Density Parity Check Codes*. Cambridge, MA: M.I.T. Press, 1963.
- [2] T. J. Richardson and R. Urbanke, “The capacity of low-density parity-check codes under message-passing decoding,” *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
- [3] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008. [Online]. Available: <http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521852296>
- [4] T. J. Richardson, “Error floors of LDPC codes,” in *41st Annual Allerton Conf. on Communications, Control and Computing*, 2003, pp. 1426–1435. [Online]. Available: <http://www.hpl.hp.com/personal/Pascal.Vontobel/pseudocodewords/papers>
- [5] M. Stepanov, V. Chernyak, M. Chertkov, and B. Vasic, “Diagnosis of Weaknesses in Modern Error Correction Codes: A Physics Approach,” *Physical Review Letters*, vol. 95, no. 22, pp. 1–4, Nov. 2005. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevLett.95.228701>
- [6] N. Wiberg, “Codes and decoding on general graphs,” Ph.D., Univ. Linköping, Sweden, Dept. Elec. Eng., 1996.
- [7] G. D. Forney, R. Koetter, F. R. Kschischang, and A. Reznik, “On the effective weights of pseudocodewords for codes defined on graphs with cycles,” in *In Codes, systems and graphical models*. Springer, 2001, pp. 101–112.
- [8] J. Feldman, M. Wainwright, and D. Karger, “Using linear programming to decode binary linear codes,” *Information Theory, IEEE Transactions on*, vol. 51, no. 3, pp. 954 – 972, March 2005.
- [9] R. Koetter and P. O. Vontobel, “Graph covers and iterative decoding of finite-length codes,” in *Proc. of the 3rd Intern. Conf. on Turbo Codes and Related Topics*, Sept. 1-5 2003, pp. 75–82.
- [10] M.-H. Taghavi and P. Siegel, “Adaptive methods for linear programming decoding,” *Information Theory, IEEE Transactions on*, vol. 54, no. 12, pp. 5396 –5410, dec. 2008.
- [11] P. O. Vontobel and R. Koetter, “Towards low-complexity linear-programming decoding,” *CoRR*, vol. abs/cs/0602088, 2006.
- [12] A. Dimakis, A. Gohari, and M. Wainwright, “Guessing facets: Polytope structure and improved lp decoder,” *Information Theory, IEEE Transactions on*, vol. 55, no. 8, pp. 3479 –3487, aug. 2009.
- [13] M. Chertkov and V. Chernyak, “Loop calculus helps to improve belief propagation and linear programming decodings of low-density-parity-check codes,” in *Proc. of the 44th Annual Allerton Conf. on Communications, Control and Computing*, 2006. [Online]. Available: <http://arxiv.org/abs/cs/0609154>
- [14] M. Chertkov, “Reducing the error floor,” *Information Theory Workshop, 2007. ITW '07. IEEE*, pp. 230–235, Sept. 2007.
- [15] K. Yang, X. Wang, and J. Feldman, “Fast ml decoding of spc product code by linear programming decoding,” in *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, nov. 2007, pp. 1577 – 1581.
- [16] J. Yedidia, S. Draper, and Y. Wang, “Multi-stage decoding of ldpc codes,” in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, 28 2009-july 3 2009, pp. 2151 –2155.
- [17] M. Chertkov and M. Stepanov, “An efficient pseudocodeword search algorithm for linear programming decoding of ldpc codes,” *Information Theory, IEEE Transactions on*, vol. 54, no. 4, pp. 1514 –1520, April 2008.
- [18] S. Chilappagari, M. Chertkov, M. Stepanov, and B. Vasic, “Instanton-based techniques for analysis and reduction of error floors of ldpc codes,” *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 6, pp. 855 –865, august 2009.
- [19] S. Chilappagari, B. Vasic, M. Stepanov, and M. Chertkov, “Analysis of error floors of ldpc codes under lp decoding over the bsc,” in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, 28 2009-july 3 2009, pp. 379 –383.
- [20] S. Chilappagari, M. Chertkov, and B. Vasic, “Worst configurations (instantons) for compressed sensing over reals: A channel coding approach,” in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, june 2010, pp. 1978 –1982.
- [21] R. M. Tanner, D. Sridhara, and T. Fuja, “A class of group-structured LDPC codes,” in *ISCTA*, 2001. [Online]. Available: <http://www.soe.ucsc.edu/~tanner/isctaGrpStrLDPC.pdf>
- [22] D. R. Hunter and K. Lange, “A tutorial on mm algorithms,” *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004. [Online]. Available: <http://pubs.amstat.org/doi/abs/10.1198/0003130042836>
- [23] J. Yedidia, W. Freeman, and Y. Weiss, “Constructing free-energy approximations and generalized belief propagation algorithms,” *Information Theory, IEEE Transactions on*, vol. 51, no. 7, pp. 2282 – 2312, July 2005.
- [24] M. Yannakakis, “Expressing combinatorial optimization problems by linear programs,” *Journal of Computer and System Sciences*, vol. 43, no. 3, pp. 441 – 466, 1991. [Online]. Available: <http://www.sciencedirect.com/science/article/B6WJ0-4B4RNRG-GX/2/650991fa0c7a2a>
- [25] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
- [26] P. O. Vontobel and R. Koetter, “Graph-cover decoding and finite-length analysis of message-passing iterative decoding of ldpc codes,” *CoRR*, vol. abs/cs/0512078, 2005.
- [27] M. Nakazawa, K. Kikuchi, and M. Tetsuya, *High Spectral Density Optical Communication Technologies (Optical and Fiber Communications Reports)*. Springer, 2010.