Blahut-Arimoto Algorithm and Code Design for Action-Dependent Source Coding Problems

Kasper F. Trillingsgaard*, Osvaldo Simeone†, Petar Popovski* and Torben Larsen*
*Department of Electronic Systems, Aalborg University, Aalborg, Denmark
†CWCSPR, New Jersey Institute of Technology, Newark, NJ, USA

***Abstract*—The source coding problem with action-dependent side information at the decoder has recently been introduced to model data acquisition in resource-constrained systems. In this paper, an efficient Blahut-Arimoto-type algorithm for the numerical computation of the rate-distortion-cost function for this problem is proposed. Moreover, a simplified two-stage code structure based on multiplexing is put forth, whereby the first stage encodes the actions and the second stage is composed of an array of classical Wyner-Ziv codes, one for each action. Leveraging this structure, specific coding/decoding strategies are designed based on LDGM codes and message passing. Through numerical examples, the proposed code design is shown to achieve performance close to the rate-distortion-cost function.**

I. Introduction

The source coding problem in which the decoder can take actions that affect the availability or quality of the side information at the decoder was introduced in [1]. The problem generalizes the well-known Wyner-Ziv set-up and can be used to model data acquisition in resource-constrained systems, such as sensor networks. The rate-distortion-cost function for this problem, which characterizes the achievable tuples of rates, average distortions and average actions costs, was derived in [1]. The analysis in [1] does not provide numerical algorithms for the evaluation of this function and also does not elaborate on practical coding strategies that are able to attain the optimal performance. These two aspects are investigated in this paper. First, an algorithm is proposed that generalizes the classical Blahut-Arimoto (BA) approach [2] to the problem of source coding with action-dependent side information (Sec. III). We recall that the BA algorithm was extended to the Wyner-Ziv problem in [3]. Second, we tackle the aspect of code design by first observing that a layered code structure in which the refinement layer uses multiplexing of separate classical Wyner-Ziv codes, one for each action, is optimal. Based on this structure, LDGM-based codes with message passing encoding are designed and demonstrated, via numerical results, to perform close to the rate-distortion-cost function (Sec. IV and Sec. V). It is noted that, for the Wyner-Ziv problem, code design has been widely studied and efficient codes have been found to include polar codes [4] and compound LDPC/LDGM codes [5].

II. Background

In this section, we recall the definition of source coding problems with action-dependent side information and review the rate-distortion-cost function obtained in [1].

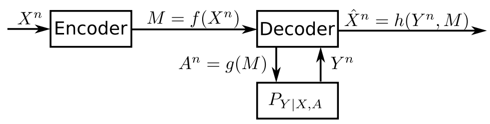


Fig. 1. Source coding with action-dependent side information.

A. System Model

The source coding problem with action-dependent side information introduced in [1] is illustrated in Fig. 1. In this problem, the source $X^n \in \mathcal{X}^n$ is memoryless and each sample is distributed according to the pmf P_X. At the encoder, the encoding function $f : \mathcal{X}^n \to \left[1, \lfloor 2^{nR} \rfloor\right]$ maps the source X^n into a message $M \in \left[1, \lfloor 2^{nR} \rfloor\right]$, where R denotes the rate in bits per sample and $[a, b]$ with $a, b \in \mathbb{Z}$ and $a < b$ denotes the set $\{a, a+1, \dots, b-1, b\}$. At the decoder, an action sequence $A^n \in \mathcal{A}^n$ is chosen according to an action strategy $g : \left[1, \lfloor 2^{nR} \rfloor\right] \to \mathcal{A}^n$, which maps the message M into an action sequence A^n. Based on A^n, the side information $Y^n \in \mathcal{Y}^n$ is conditionally independent and identically distributed (iid) according to the conditional pmf $P_{Y|X,A}$ so that $P_{Y^n|X^n,A^n}(y^n|x^n, a^n) = \prod_{i=1}^n P_{Y|X,A}(y_i|x_i, a_i)$. The decoder makes a reconstruction $\hat{X}^n \in \hat{\mathcal{X}}^n$ of X^n according to the decoding function $h : \left[1, \lfloor 2^{nR} \rfloor\right] \times \mathcal{Y}^n \to \hat{\mathcal{X}}^n$, which maps message M and side information Y^n into the estimate $\hat{X}^n$.

The action cost function $\Delta(a) : \mathcal{A} \to \mathbb{R}_+$ is defined such that $\Delta(a) = 0$ for some $a \in \mathcal{A}$ and $\Delta_{\max} = \max_{a \in \mathcal{A}} \Delta(a) < \infty$, and the distortion function $d(x, \hat{x}) : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}_+$ is defined such that for each $x \in \mathcal{X}$ there is an $\hat{x} \in \hat{\mathcal{X}}$ satisfying $d(x, \hat{x}) = 0$. The rate-distortion-cost tuple (R, D, C) is then said to be achievable if and only if, for all $\varepsilon > 0$, there exist an encoding function f, an action function g and a decoding function h, for all sufficiently large $n \in \mathbb{N}$, satisfying the distortion constraint $\mathbb{E}\left[\sum_{i=1}^n d(X_i, \hat{X}_i)\right] \leq n(D + \varepsilon)$ and the action cost constraint $\mathbb{E}\left[\sum_{i=1}^n \Delta(A_i)\right] \leq n(C + \varepsilon)$. The rate-distortion-cost function, denoted as $R(D, C)$, is defined as the infimum of all rates R such that the tuple (R, D, C) is achievable.

B. Rate-Distortion-Cost Function

The rate-distortion-cost function $R(D, C)$ was derived in [1] and is summarized below.

Lemma 1. *([1, Theorem 1]) The rate-distortion-cost function*

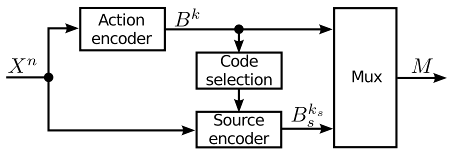


Fig. 2. Optimal encoder for source coding problems with action-dependent side information.

for the source coding problem with action-dependent side information is given as

$$R(D,C) = \min I(X;A) + I(X;U|Y,A), \tag{1}$$

where the joint distribution $P_{X,Y,A,U}$ *is of the form* $P_X(x)P_{U|X}(u|x)\mathbb{1}_{\{\eta(u)=a\}}P_{Y|X,A}(y|x,a)$ *and the minimization is over all pmfs* $P_{U|X}$ *and deterministic functions* $\eta : \mathcal{U} \to \mathcal{A}$ *under which the conditions* $\mathbb{E}[d(X, \hat{X}^{opt}(U,Y))] \le D$ *and* $\mathbb{E}[\Delta(A)] \le C$ *hold. The function* $\hat{X}^{opt} : \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{X}}$ *denotes the best estimate of* X *given* U *and* Y*, i.e.,*

$$\hat{X}^{opt}(u,y) = \arg\min_{\hat{x}\in\hat{\mathcal{X}}} \mathbb{E}[d(X,\hat{x})|U=u, Y=y]. \tag{2}$$

Moreover, the cardinality of the set $\mathcal{U}$ *can be restricted as* $|\mathcal{U}| \le |\mathcal{X}||\mathcal{A}| + 2$.

C. Optimal Coding Strategy

The proof of achievability of the rate-distortion-cost function in [1] shows that an optimal encoder has the structure illustrated in Fig. 2 and consists of the following two steps.

Action Coding: The source sequence X^n is mapped, using the joint typicality criterion, to an action sequence A^n, selected from a codebook $\mathcal{C}_A$ of about $2^{nI(X;A)}$ codewords. The index B^k identifies the selected codeword A^n, and hence consists of k, approximately equal to $nI(X;A)$, bits.

Source Coding: Given the action sequence A^n, a source codebook is chosen out of a set of around $2^{nI(X;A)}$ codebooks, one for each codeword in $\mathcal{C}_A$. The number of codewords in each source codebook is about $2^{nI(X;U|A)}$. Using joint typicality, the source sequence is mapped to a sequence U^n taken from the selected codebook. Each source codebook is divided into around $2^{nI(X;U|A,Y)}$ subcodebooks, or bins, and the index of the bin $B_s^{k_s}$ where the selected codeword U^n lies is produced by the encoder. The index $B_s^{k_s}$ has k_s, approximately equal to $nI(X;U|A,Y)$ bits.

We refer to [1], [6] for further details.

III. Computation of the Rate-Distortion-Cost Function

In this section, we first reformulate the problem in (1) by introducing Shannon strategies. This result is then used to propose a BA-type algorithm for the computation of the rate-distortion-cost function (1).

Algorithm 1 BA-type Algorithm for Computation of the Rate-Distortion-Cost Function

input: Lagrange multipliers $s \le 0$ and $m \le 0$.
output: $R(D_{s,m}, C_{s,m})$ with $C_{s,m}$ and $D_{s,m}$ as in (6)-(7).
initialize: $P_{T|X}$
repeat
  Compute Q_A as P_A.
  Compute $Q_{T,Y}$ as $P_{T,Y}$.
  Minimize $F(P_{T|X}, Q_{T,Y}, Q_A)$ with respect to $P_{T|X}$ using Algorithm 2.
until convergence
$P^*_{T|X} \leftarrow P_{T|X}$

A. Shannon Strategies

We first observe that, from Lemma 1, it is sufficient to restrict the minimization to all joint distributions for which A is a deterministic function $A = \eta(U)$. Moreover, the final estimate of $\hat{X}$ in (2) is a function of both U and Y. Based on these facts, we define a Shannon strategy $T \in \mathcal{T} \subseteq \mathcal{X}^{|\mathcal{Y}|} \times \mathcal{A}$ as a vector of cardinality $|\mathcal{Y}|+1$, in which the first $|\mathcal{Y}|$ elements are indexed by the elements in $\mathcal{Y}$ and $T(y) \in \hat{\mathcal{X}}$ for $y \in \mathcal{Y}$, and the last element is denoted $\mathrm{a}(T) \in \mathcal{A}$. We also define the disjoint sets $\mathcal{T}^a = \{t \in \mathcal{T} : \mathrm{a}(t) = a\}$ for all actions $a \in \mathcal{A}$. The rate-distortion-cost function (1) can be restated in terms of the defined Shannon strategies as formalized in the next proposition.

Proposition 1. *Let* $T \in \mathcal{T} \subseteq \mathcal{X}^{|\mathcal{Y}|} \times \mathcal{A}$ *denote a Shannon strategy vector as defined above. The rate-distortion-cost function in* (1) *can be expressed as*

$$R(D,C) = \min I(X;\mathrm{a}(T)) + I(X;T|Y,\mathrm{a}(T)), \tag{3}$$

where the joint pmf $P_{X,Y,T}$ *is of the form* $P_X(x)P_{T|X}(t|x)P_{Y|A,X}(y|\mathrm{a}(t),x)$*, and the minimization is over all pmfs* $P_{T|X}$ *under the constraints*

$$\mathbb{E}[\Delta(A)] \le C \quad \textit{and} \quad \mathbb{E}[d(X,T(Y))] \le D. \tag{4}$$

Moreover, the cardinality of the alphabet $\mathcal{T}$ *can be restricted as* $|\mathcal{T}| \le |\mathcal{X}||\mathcal{A}| + 2$.

Proof. Given an alphabet $\mathcal{U}$, a pmf $P_{U|X}$ and a function η, the sum of the two mutual informations in (1) can be seen to be equal to the sum of the two mutual informations in (3) and the average distortion and cost in Lemma 1 to be equal to (4), by defining $P_{T|X}$ as follows. For each $u \in \mathcal{U}$, define a strategy t with $P_{T|X}(t|x) = P_{U|X}(u|x)$ such that $\mathrm{a}(t) = \eta(u)$ and $t(y) = \hat{X}^{\mathrm{opt}}(u,y)$ for $y \in \mathcal{Y}$. □

Remark. The characterization in Proposition 1 generalizes the formulation of the Wyner-Ziv rate-distortion function in terms of Shannon strategies given in [3].

B. Computation of the Rate-Distortion-Cost Function

In order to derive a BA-type algorithm to solve the problem in (3), we introduce Lagrange multipliers m for the cost constraint and s for the distortion constraint given in (4). The following proposition provides a parametric characterization of the rate-distortion-cost function in terms of the pair (s, m).

Proposition 2. *Every point (R, D, C) on the rate-distortion-cost function can be obtained for some $s \leq 0$ and $m \leq 0$ as*

$$R(D_{s,m}, C_{s,m}) = sD_{s,m} + mC_{s,m} + \min_{P_{T|X}, Q_{T,Y}, Q_A} F(P_{T|X}, Q_{T,Y}, Q_A), \quad (5)$$

where

$$C_{s,m} = \sum_{t \in \mathcal{T}, x \in \mathcal{X}} P_X(x) P^*_{T|X}(t|x) \Delta(\mathrm{a}(t)), \quad (6)$$

$$D_{s,m} = \sum_{t \in \mathcal{T}, x \in \mathcal{X}, y \in \mathcal{Y}} P_X(x) P^*_{T|X}(t|x) \cdot P_{Y|X,A}(y|x, a) d(t(y), x) \quad (7)$$

and the function $F(P_{T|X}, Q_{T|Y}, Q_A)$ is given by

$$F(P_{T|X}, Q_{T,Y}, Q_A) = D_{KL}(P_{Y,A}||Q_A) - H(Y|X, \mathrm{a}(T)) + \sum_{x \in \mathcal{X}} P_X(x) D_{KL}(P_{Y,T|X}(\cdot, \cdot|x)||Q_{T,Y}) - s\mathbb{E}[d(T(Y), X)] - m\mathbb{E}[\Delta(\mathrm{a}(T))]. \quad (8)$$

The pmf $P^*_{T|X}$ denotes a minimizing pmf in (5) *and $D_{KL}(\cdot||\cdot)$ in* (8) *denotes the Kullback-Leibler divergence [7]. Moreover, the function F is jointly convex in the pmfs $P_{T|X}$, $Q_{T,Y}$ and Q_A.*

Proof. The proof technique is as in [2], and is based on showing that the pmf Q_A minimizing F for fixed $Q_{T,Y}$ and $P_{T|X}$ is P_A and the pmf $Q_{T,Y}$ minimizing F for fixed Q_A and $P_{T|X}$ is $P_{T,Y}$. The convexity of the function $F(\cdot)$ follows from the log-sum inequality [7]. □

Based on Proposition 2, the proposed BA-type algorithm for computation of the rate-distortion-cost function then consists of alternate minimizing (5) with respect to $P_{T|X}$, $Q_{T,Y}$ and Q_A. Due to the convexity of (5), the algorithm is known to converge to the optimal point similar to [3]. The proposed algorithm is summarized in Table Algorithm 1. The step of minimizing $F(P_{T|X}, Q_{T,Y}, Q_A)$ with respect to $P_{T|X}$ is briefly discussed next.

To minimize the function $F(P_{T|X}, Q_{T,Y}, Q_A)$ with respect to $P_{T|X}$ for fixed Q_A and $Q_{T,Y}$, we add a Lagrange multipliers λ_x for each equality constraints $\sum_{t \in \mathcal{T}} P_{T|X}(t|x) = 1$ with $x \in \mathcal{X}$, and resort to the KKT conditions as necessary and sufficient conditions for optimality. We assume $P_X(x) > 0$ without loss of generality, since values of x with $P_X(x) = 0$ can be removed from the alphabet $\mathcal{X}$.

By strong duality, we obtain the following optimization problem

$$\max_{\{\lambda_x\} \in \mathbb{R}^{|\mathcal{X}|}} \min_{P_{T|X}} F(P_{T|X}, Q_A, Q_{T,Y}) + \sum_{x \in \mathcal{X}} \lambda_x \left(\sum_{t \in \mathcal{T}} P_{T|X}(t|x) - 1 \right). \quad (9)$$

Algorithm 2 Algorithm for Minimization of F with respect to $P_{T|X}$

input: $Q_{T,Y}$ and Q_A.
output: $P^*_{T|X}$.
parameters: Subgradient weights $\theta_i = \frac{1}{i}, i \in \mathbb{Z}_+$ and constant $\beta \in (0, 1)$.
initialization: $i = 0$; $\mu_x^{(0)} = 1$ for $x \in \mathcal{X}$; $P^{(0)}_{A|X}(a|x) = \frac{1}{|\mathcal{T}|}$ for $t \in \mathcal{T}, x \in \mathcal{X}$.
repeat
  Perform fixed-point iterations on the system $P_{A|X}(a|x) = g_{a|x}(P_{A|X}, \mu_x)$ for $a \in \mathcal{A}$ and $x \in \mathcal{X}$ with starting point $P^{(i)}_{A|X}$ until convergence to obtain $P^{(i+1)}_{A|X}$.
  Update the subgradients as
  $\mu_x^{(i+1)} = \mu_x^{(i)} + \frac{\theta_i}{P(x)} \left(1 - \sum_{a \in \mathcal{A}} P^{(i+1)}_{A|X}(a|x)\right)$ for $x \in \mathcal{X}$.
  $i \leftarrow i + 1$.
until convergence
Compute $P^*_{T|X}(t|x) = \frac{\alpha_{t,x}}{\alpha_{\mathrm{a}(t),x}} P^{(i)}_{A|X}(\mathrm{a}(t)|x)$.

In the proposed approach, the outer maximization in (9) is then performed using the standard subgradient method, while the inner minimization is instead performed by finding the stationary points of the function. The procedure is summarized in Table Algorithm 2, where we have defined the function $g_{a|x}(P_{A|X}, \mu_x)$ in (10) on next page, and

$$\alpha_{t,x} = Q_A(\mathrm{a}(t)) 2^{m\Delta(\mathrm{a}(t))} \cdot 2^{\sum_{y \in \mathcal{Y}} P_{Y|X,A}(y|x,\mathrm{a}(t))[sd(t(y),x) + \log Q_{T,Y}(t,y)]}, \quad (11)$$

$$\alpha_{a,x} = \sum_{t \in \mathcal{T}^a} \alpha_{t,x}. \quad (12)$$

The procedure is summarized in Table Algorithm 2 and details of the derivation can be found in [6]. Overall, we have the following result.

Proposition 3. *The algorithm in Tables Algorithm 1 and Algorithm 2 converges to the rate-distortion-cost function $R(D_{s,m}, C_{s,m})$ for all $s \leq 0$ and $m \leq 0$.*

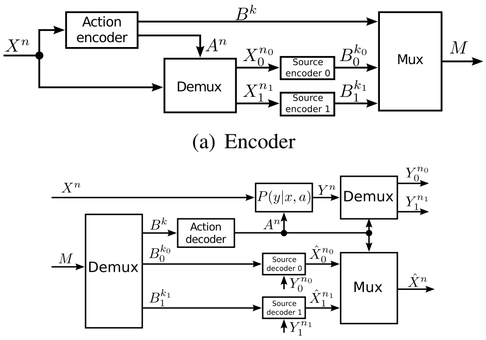
Proof. The proof follows from Proposition 2, from the convergence of the subgradient algorithm [8] and from Banach fixed-point theorem. We refer to [6] for details. □

IV. Code Design

In this section, we aim at designing codes that perform close the rate-distortion-cost function given in Lemma 1 for some fixed pmf $P_{X,Y,A,U}$ (or equivalently in Proposition 1 for some fixed pmf $P_{X,Y,T}$).

A. Achievability via Multiplexing

As explained in Section II-C, the achievability proof in [1] is based on an action codebook $\mathcal{C}_A$ for the action sequences A^n of about $2^{nI(X;A)}$ codewords and $2^{nI(X;A)}$ source codebooks of about $2^{nI(X;U|A)}$ codewords for the sequences U^n, where each source codebook corresponds to an action sequence A^n.



(a) Encoder

(b) Decoder

Fig. 3. Code design for source coding problems with action-dependent side information. The illustration is for $\mathcal{A} = \{0, 1\}$.

Here, we first observe that the code design can be simplified without loss of optimality by using the encoder and decoder structures in Fig. 3. Accordingly, rather than using $2^{nI(X;A)}$ source codebooks, we utilize only $|\mathcal{A}|$ source codebooks $\mathcal{C}_{s,a}$, $a \in \mathcal{A}$. Specifically, the source codebook $\mathcal{C}_{s,a}$ has about $2^{nP_A(a)I(X;U|A=a)}$ codewords, and each codeword in codebook $\mathcal{C}_{s,a}$ has a length of $n_a = \lceil n(P_A(a) + \varepsilon) \rceil$ symbols for some $\varepsilon > 0$.

To elaborate, as seen in Fig. 3(a), after action encoding, which takes place as in [1], the source X^n is demultiplixed into $|\mathcal{A}|$ subsequences, such that the a-th subsequence $X_a^{n_a}$ contains all symbols X_i for which $A_i = a$. Therefore, for sufficiently large n, by the law of large numbers, the number of symbols in $X_a^{n_a}$ is less than n_a with high probability. Appropriate padding is then used to make the length of the sequence exactly n_a symbols. The a-th subsequence $X_a^{n_a}$ is then compressed using the codebook $\mathcal{C}_{s,a}$ with the objective of ensuring that $X_a^{n_a}$ and $U_a^{n_a}$ are jointly typical with respect to the pmf $P_{X,U|A}(\cdot,\cdot|a)$. Binning is performed on each source codebook so that the number of bins is $2^{nP_A(a)I(X;U|Y,A=a)}$. The bin index $B_a^{k_a}$ of $U_a^{n_a}$ is thus of $k_a = \lceil nP_A(a)I(X;U|Y,A=a) \rceil$ bits. Overall, the rate of the message M, consisting of the indices B^k for the action code and $B_{s,a}^{k_a}$ for the source codes with $a \in \mathcal{A}$, is $I(X;A) + \sum_{a\in\mathcal{A}} P_A(a)I(X;U|Y,A=a) = I(X;A) + I(X;U|A,Y)$ as desired.

At the decoder, as seen in Fig. 3(b), the action sequence A^n is reconstructed and is used to measure the side information Y^n. The side information Y^n is demultiplexed into $|\mathcal{A}|$ subsequences, such that the a-th subsequence $Y_a^{n_a}$ contains all symbols Y_i for which $A_i = a$. Each of the subsequences $U_a^{n_a}$ are then reconstructed by using Wyner-Ziv decoding based on the message bits $B_a^{k_a}$ and the side information $Y_a^{n_a}$, and the reconstructed source subsequences $\hat{X}_{a,i}$ are obtained as $\hat{X}_{a,i} = \hat{X}^{\text{opt}}(U_{a,i}, Y_{a,i})$ for $i \in [1, n_a]$, where $\hat{X}_{a,i}$ denotes the i-th symbol of the sequence $X_a^{n_a}$. Finally, the source reconstruction $\hat{X}^n$ is obtained by multiplexing the subsequences $\hat{X}_a^{n_a}$ for $a \in \mathcal{A}$.

B. The Action Code

Based on the encoder structure in Fig. 3(a), we discuss the specific design of the action encoder. The action code $\mathcal{C}_A$ has to ensure that the codewords A^n approximately have the type P_A, and the action encoder must obtain a codeword A^n that is jointly typical with respect to the joint pmf $P_{X,A}$. These conditions are satisfied by optimal source codes [4]. Optimal source codes can be designed using LDGM codes or polar codes as shown in [9] and [4], respectively. Here, we adopt LDGM codes as proposed in [9], [10]. Specifically, in the following, we define an encoder based on message passing. This uses ideas from [9] to handle the general alphabet and pmf P_A, and from [10] to implement message passing and decimation. The key difference with respect to [10] is that there the goal of the encoder is to minimize the Hamming distance, while the aim in this paper is to find an action sequence that is jointly typical with the source.

We use the code described by the factor graph in Fig. 4. The bottom section of the graph is a LDGM code (see, e.g. [9]). The sequence B^k denotes the message bits with $k = \lceil nI(X;A) \rceil$ and $\{g_{\kappa,l} : \kappa \in [1,d], l \in [1,n]\}$ denote the check variables of the LDGM code, where the choice of d is explained later. The objective of the mappings $\psi_l : \{0,1\}^d \times \mathcal{A} \to \{0,1\}$ for $l \in [1,n]$ is to ensure that the types of the codewords, or action variables, are approximately equal to P_A [9]. Specifically, each mapping ψ_l applies to the subset of check variables $\{g_{\kappa,l}\}_{\kappa\in[1,d]}$ and to the symbol a_l and is defined in terms of a mapping $\phi : \{0,1\}^d \to \mathcal{A}$ as $\psi_l(\{g_{\kappa,l}\}_{\kappa\in[1,d]}, a) = \mathbb{1}_{\{\phi(\{g_{\kappa,l}\}_{\kappa\in[1,d]})=a\}}$. Following [9], the value of $d \in \mathbb{Z}_+$ is chosen such that there are integers ν_a for $a \in \mathcal{A}$ satisfying $\sum_{a\in\mathcal{A}} \nu_a = 2^d$ and $P_A(a) \approx \frac{\nu_a}{2^d}$. The mapping ϕ is then arbitrarily chosen such that exactly v_a of the 2^d binary sequences $\{g_{\kappa,l}\}_{\kappa\in[1,d]}$ map to a.

Given the source sequence X^n, the encoder runs the sum-product algorithm with decimation as in [10] in order to obtain the message bits B^k, and hence the action sequence A^n (see [4] for a discussion of the role of decimation in source coding problems).

C. The Source Codes

Based on the proposed encoder structure in Fig. 3(a), the design of each source code $C_{s,a}$ for $a \in \mathcal{A}$ is equivalent to optimal codes for classical Wyner-Ziv problems (see [4], [5]).

$$g_{a|x}(P_{A|X}, \mu_x) = P_{A|X}(a|x)^\beta \left(\frac{2^{\mu_x}\alpha_{a,x}}{\prod_{y\in\mathcal{Y}} \left[\sum_{\tilde{x}\in\mathcal{X}} P_X(\tilde{x}) P_{Y|X,A}(y|\tilde{x},a) P_{A|X}(a|\tilde{x})\right]^{P_{Y|X,A}(y|x,a)}} \right)^{1-\beta} \tag{10}$$

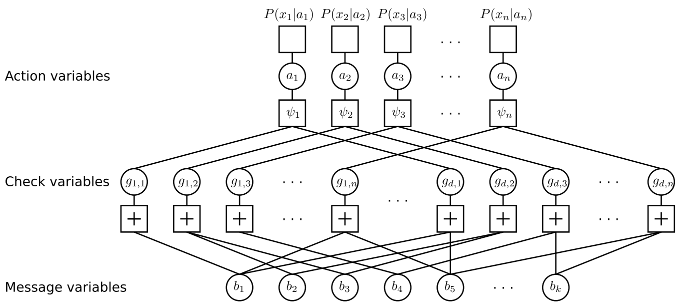


Fig. 4. Factor graph defining the action encoder.

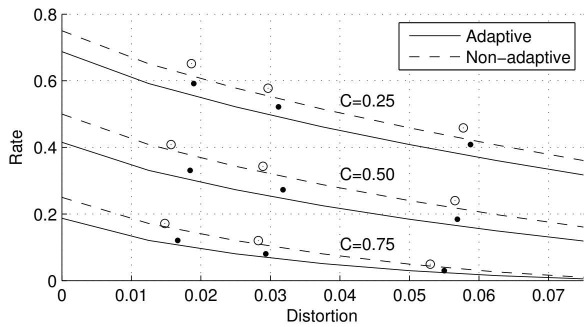


Fig. 5. Rate-distortion-cost function (lines) compared to the performance of the proposed code design (markers) for both the adaptive and non-adaptive action approaches.

V. Numerical Examples

Let $X \in \mathcal{X} = [1, K+1]$ be a random variable with pmf

$$P_X(x) = \begin{cases} \frac{1-q}{K} & \text{if} \quad x \in [1, K] \\ \mathrm{q} & \text{if} \quad x = K+1 \end{cases}, \tag{13}$$

for $q \in [0,1]$. The letters $1, \ldots, K$ denote source outcomes that are relevant for the decoder, and thus should ideally be distinguishable by the latter, while the letter $x = K+1$ represents a source outcome that is irrelevant for the decoder. Examples where this situation arises includes monitoring systems in which the decoder wishes to recover the values of a physical quantity only when above, or below, a certain predetermined threshold. To account for this requirement, the distortion function is given by $d(x, \hat{x}) = \mathbb{1}_{\{x \neq \hat{x} \text{ and } x \in [1,K]\}}$, i.e., the decoder is only penalized if it makes an error when x is a relevant letter.

At each time i, the decoder can choose an action $A_i \in \{0, 1\}$, such that, if $A_i = 0$, the side information is given by $Y_i = \mathrm{e}$, where e denotes an erasure symbol, and if $A_i = 1$, the side information is given by $Y_i = X_i$. The action cost function $\Delta(\cdot)$ is given by $\Delta(a) = \mathbb{1}_{\{a=1\}}$, which implies that the cost constraint with $0 \leq C \leq 1$ enforces that no more than nC samples of the side information X^n can be measured by the receiver. The rate-distortion-cost function for this example can be computed using the proposed BA-type algorithm.

We investigate the problem of code design for the scenario by adopting the approach proposed in Section IV. We consider the case where $q = \frac{1}{2}$, $K = 4$, which yields $d = 2$ for both the action code $\mathcal{C}_A$ and the source code $\mathcal{C}_{s,0}$. We fix a blocklength of $n = 10\,000$, yielding LDGM codes of blocklength, $20\,000$. Each point is averaged over 50 source realizations and LDGM codes. For both codes, we use the sum-product algorithm with decimation in [10]. As in [10], we use damping after 30 iterations and the maximum number of iterations is set to 100. Nodes are decimated if their log-likelihood ratios are larger than 2. Suitable irregular degree distributions optimized for the AWGN channel are obtained from [11]. The results are shown in Fig. 5, where we refer to the optimal approach discussed thus far as "adaptive actions", while labeling as "non-adaptive actions" the simplified class of strategies in which the actions are selected independently of the encoder's message (see [1]). It is seen that the resulting distortions are close the lower bounds for both the adaptive and non-adaptive actions strategies. Moreover, the theoretical gains of the adaptive action strategy versus the non-adaptive one are confirmed by the practical implementation. We refer to [6] for further results.

VI. Conclusion

In this paper, we considered computation of the rate-distortion-code function and code design for source coding problems with action-dependent side information. We formulated the problem using Shannon strategies and proposed a BA-type algorithm that efficiently computes the rate-distortion function. Convergence of this algorithm was proved. Moreover, we proposed a code design based on multiplexing that significantly reduced the complexity compared to previous work. Numerical results demonstrate that the code design achieves performance close to the rate-distortion bound.

References

[1] H. H. Permuter and T. Weissman, "Source coding with a side information "vending machine"," *IEEE Trans. Inform. Theory*, vol. 57, no. 7, pp. 4530–4543, Jul 2011.

[2] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 18, no. 4, pp. 460 – 473, Jul 1972.

[3] F. Dupuis, W. Yu, and F. M. J. Willems, "Blahut-arimoto algorithms for computing channel capacity and rate-distortion with side information," in *Proc. IEEE Symp. Inform. Theory*, Chicago, IL, USA, Jun 2004.

[4] S. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1751 –1768, Apr 2010.

[5] M. Wainwright and E. Martinian, "Low-density graph codes that are optimal for binning and coding with side information," *IEEE Trans. Inform. Theory*, vol. 55, no. 3, pp. 1061 –1079, Mar 2009.

[6] K. F. Trillingsgaard, O. Simeone, P. Popovski, and T. Larsen, "Blahut-arimoto algorithm and code design for action-dependent source coding problems," 2013, available on arXiv.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2006.

[8] B. Polyak, *Introduction to Optimization*. Optimization Software, Inc., 1987.

[9] Z. Sun, M. Shao, J. Chen, K. Wong, and X. Wu, "Achieving the rate-distortion bound with low-density generator matrix codes," *IEEE Trans. Comm.*, vol. 58, no. 6, pp. 1643 –1653, Jun 2010.

[10] T. Filler and J. Fridrich, "Binary quantization using belief propagation with decimation over factor graphs of ldgm codes," in *Proc. Allerton Conf. Comm., Cont. and Comp.*, Monticello, IL, USA, 2007.

[11] A. University of Newcastle. (2012, Dec.) Lopt - online optimisation of ldpc and ra degree distributions. [Online]. Available: http://sonic.newcastle.edu.au/ldpc/lopt