IRERE 10.1561/IRERE International Review of Environmental and Resource Economics IRERE 1932-1465 1932-1473 1932-1465 Emerald Publishing Limited 10.1561 10.1561/101.00000046 cat-ECO The Green Paradox and Greenhouse Gas Reducing
Investments Hoel Michael m.o.hoel@econ.uio.no Department of Economics, University of Oslo 25 09 2011 25 09 2011 5 4 353 379 © 2011 M. Hoel 2011 M. Hoel Licensed re-use rights only

If governments cannot commit to future carbon tax rates, investments in greenhouse gas mitigation will be based on uncertain and/or wrong predictions about these tax rates. Predictions about future carbon tax rates are also important for decisions made by owners of nonrenewable carbon resources. The effects of the size of expected future carbon taxes on near-term emissions and investments in substitutes for carbon energy depend significantly on how rapidly extraction costs increase with increasing total extraction. In addition, the time profile of the returns to investments in noncarbon substitutes is important for the effects on emissions and investments.

Climate change carbon tax green paradox commitment exhaustible resources H23 Q30 Q42 Q54 peer-reviewed true academic-content Yes rightslink included

7

This section addresses the rate-distortion (RD) performance of lossy CS from an information-theoretic perspective. A single-sensor setup is assumed due to the analytical tractability. In this setup, a CS based sensor observes a sparse information source indirectly and communicates compressed noisy measurements to a decoder for signal reconstruction with the aim to minimize the MSE distortion. Thus, the obtained theoretical results shed light on, e.g., the compression performance limits of the single-sensor version of the QCS setup in Section 6. The minimum achievable rate for a given distortion fidelity in a QCS setup is represented by the remote rate-distortion function (RDF). We first present an analytically tractable lower bound to the remote RDF by providing support side information to the encoder and decoder. A variant of the Blahut-Arimoto algorithm is devised to numerically approximate the remote RDF. Furthermore, we present several practical symbol-by-symbol QCS algorithms relying on 1) compress-and-estimate, 2) estimate-and-compress, and 3) support-estimation-and-compress strategies. Numerical results illustrate the main RD characteristics of the lossy CS and compare the performance of practical QCS methods against the proposed limits. In particular, an entropy-constrained VQ based estimate-and-compress QCS method is numerically shown to approach the remote RDF.

7.1 Related Works

Since existing symbol-by-symbol quantizer based QCS algorithms are already discussed in Section 6.1, the following sections mainly review the works that focus on information theoretic perspectives of lossy CS.

7.1.1 Lossy CS

In a lossy compression system, the point of interest lies in encoding an in-formation source at the minimum rate so that the source can be reproduced at the destination with a distortion not exceeding a tolerable level. This best achievable compression performance for a given distortion fidelity is given by the rate-distortion function (RDF) of a source [260]. However, the limit is achievable only via excessively complex encoder-decoder pairs with infinitely long block lengths [25, Sect. 3.3] and [134]. Thus, the RD theory is primarily applicable to performance analysis and benchmarking of practical lossy coding methods.

As discussed in Section 6.1, the compression task in a QCS setup falls into remote source coding [85][25, Sect. 3.5, 4.5.4][253, 304, 303, 110]. The compression limit for such a setup is given by the remote RDF. Information-theoretic analysis of QCS – termed lossy CS henceforth – is incomplete, i.e., the remote RDF for lossy CS is unknown. A problem overview along with initial results was first given in [129]. Kipnis et al. [165] analyzed lossy CS under a large system limit (i.e., $M,N \rightarrow \infty$) using the replica method, and derived the minimal achievable per-letter MSE in a general form. Coluccia et al. [69] derived a distortion-rate (DR) lower bound assuming support SI at the decoder, high-rate quantization, a large system regime $N \rightarrow \infty$, and noiseless measurements. As slightly different, yet related works, RD bounds for directly compressing sparse sources were derived in [57, 293, 230], and the compression of (sparse) Bernoulli generalized Gaussian sources via uniform SQ was studied in [113]. Other works on remote compression in (non-CS) setups include [276, 256, 163].

7.1.2 Computation of RDFs

Despite the well-known general definitions, deriving a RDF/remote RDF in closed form is, in general, elusive. To date, remote RDFs have been derived only for few well-behaved source/observation distributions; see, e.g., the works in [85, 304], [25, Sect. 4.5.4] and [256, 276, 163], assuming Gaussian distributions, and [166] considering a remote binary source. When a closed-form solution is unattainable, one recourse is numerical approximation. The bulk of such iterative methods rely on the Blahut-Arimoto (BA) algorithm, which traces back to the pioneering works by Blahut [30] and Arimoto [11]. This algorithm has been adapted to remote sensing scenarios under joint compression and classification in [86], and the chief executive officer problem in [168]. The effect of discretizing a continuous source on the accuracy of the RDF evaluated via the BA algorithm for finite output alphabets was analyzed in [106]; similarly, an approximation of the capacity of a continuous channel was studied in [58]. Other computation methods include a mapping based method akin to deterministic annealing in [249], and Lagrange duality based convex optimization in [65].

7.1.3 Source Coding with Side Information

In some WSN applications, the encoder and/or decoder is intermittently reinforced by various types of prior knowledge, i.e., side information (SI), on the signal of interest. SI on, e.g., sparsity, occupied frequency bands, or magnitude variations of a signal can be obtained via temporal/spatial correlations or inter-sensor collaboration. Since SI reduces the necessary transmission rate, added SI can be used to derive RD lower bounds in closed form. Compression with shared SI at the encoder and decoder follows conditional RD theory introduced by Gray [131, 132]. The case with correlated, but not necessarily identical SI at the encoder and decoder was addressed in, e.g., [73, 62, 140]. Other (non-CS) SI aided compression variants can be found in, e.g., [25, Sect. 6.1], [174, 126, 189], [121, Sect. 11.1] and [124, Sect. 5.6].

7.2 Lossy CS via Remote Source Coding

In this work, the objective is to investigate the RD performance of the model depicted in Fig. 28, where the information source is observed via noisy compressed measurements, encoded with a lossy source code, and communicated to the decoder for signal reconstruction. The transmissions from encoder E to decoder D are assumed to be error-free. The compression task is classified as

Figure 28:

A single-sensor lossy compression of a sparse source from compressive noisy measurements.

remote source coding because the encoder accesses the source only through noisy measurements. The source and the measurement model are defined next, followed by the formal statement of the problem.

7.2.1 Source Model

Let $\{X_n\}_{n=1}^{\infty}$ be a discrete-time memoryless vector source sequence[17] . Each random vector[18] $X_n=[X_{n,1}\cdots X_{n,N}]^T$ is K-sparse, where $K \le N$, i.e., it takes on values in the continuous source alphabet $\mathcal{X}=\{x\in\mathbb{R}^N:\|x\|_0=K\}$. The set χ thus consists of the union of $\binom{N}{K}$ subspaces, i.e., the signal model is nonlinear [101, 81]. It is further assumed that the source sequence is generated from the memoryless sequence of tuples $\{(G_n, B_n)\}_{n=1}^{\infty}$ so that $X_n=G_n \odot B_n$, where $\odot$ denotes the Hadamard product; G_n is a length-N zero mean Gaussian random vector $G_n\sim\mathcal{N}(0, \Sigma_G)$ with a covariance matrix $\Sigma_G\in\mathbb{S}_{++}^N$; B_n is a length-N binary support random vector, independent of G_n, with the discrete alphabet $\mathcal{B}=\{b_1,\dots,b_{|\mathcal{B}|}\}$, where $|\mathcal{B}|=\binom{N}{K}$ is the number of all possible sparsity patterns. Each $b_s=[b_{s,1} \cdots b_{s,N}]^T \in \mathcal{B}$ is unique, contains K ones and N-K zeros, and is associated with the a priori probability $p(b_s) \overset{\Delta}{=} \Pr(B=b_s)$ with $p(b_s) \in [0, 1]$ and $\sum_{s=1}^{|\mathcal{B}|} p(b_s)=1$.

7.2.2 Noisy CS

Let $\Phi \in \mathbb{R}^{M \times N}$ be a fixed and known measurement matrix, $K \leq M \leq N$. The sensor (i.e., the encoder) observes $\{X_n\}_{n=1}^{\infty}$ indirectly [43, 87, 146, 49] as

$$Y_n=\Phi X_n+W_n, \quad n=1,2, \ldots \tag{107}$$

where W_n, n = 1, 2, . . ., are length-M random measurement noise vectors independent of $\{X_n\}_{n=1}^{\infty}$, and each Y_n is a length-M measurement random vector that takes values in the measurement vector space $\mathcal{Y}$. It is assumed that $W_n \sim \mathcal{N}(0, \Sigma_W)$ with a covariance matrix $\Sigma_W \in \mathbb{S}_{++}^M$. No restricting assumptions of Φ are made in the derivations; the impact of Φ on the CS recovery performance is discussed in Section 2.1.

7.2.3 Lossy CS Problem

Let $X^m \overset{\Delta}{=} \{X_n\}_{n=1}^m$ and $x^m \overset{\Delta}{=} \{x_n\}_{n=1}^m$ denote the blocks of m consecutive source random vectors and the corresponding realizations, respectively. Let $\mathcal{X}^m$ denote the m-fold Cartesian product of $\hat{\mathcal{X}}$. Analogous notations are used for the other vectors. Let $\hat{X}$ be the reproduction random vector at the decoder output, taking values in the reproduction alphabet $\hat{\mathcal{X}}$. Finally, define the average per-letter MSE distortion between vectors $x = [x_1 \cdots x_N]^T \in \mathcal{X}$ and $\hat{x}= [\hat{x}_1 \cdots \hat{x}_N]^T \in \hat{\mathcal{X}}$ as

$$d(x,\hat{x}) \triangleq N^{-1}\sum_{k=1}^{N}(x_k-\hat{x}_k)^2 \tag{108}$$

and the average per-letter MSE distortion between blocks $x^m \in \mathcal{X}^m$ and $\hat{x}^m \in \hat{\mathcal{X}}^m$ as

$$d(x^m,\hat{x}^m) \overset{\Delta}{=} (mN)^{-1}\sum_{n=1}^{m}\sum_{k=1}^{N}(x_{n,k}-\hat{x}_{n,k})^2 \tag{109}$$

The lossy source coding system in Fig. 28 operates as follows [134, Sect. 2.1], [74, Sect. 10.2] and [121, Sect. 3.5, 3.6]. The encoder E observes a block of measurements $y^m \in \mathcal{Y}^m$ and compresses it into a message represented by an index $u \in \mathcal{U}$ of rate mNR bits using an encoder mapping

$$g_E^m: \mathcal{Y}^m \rightarrow \mathcal{U} \triangleq \{1,\ldots,2^{mNR}\} \tag{110}$$

where the rate R is defined as the bits/entry of X. The decoder D uses the index to reconstruct an estimate of $X^m \in \mathcal{X}^m$ via a decoder mapping

$$g_D^m: \mathcal{U} \rightarrow \hat{\mathcal{X}}^m \tag{111}$$

A pair (R, D) for distortion $D \geq 0$ is achievable if there exists a sequence of $(2^{mNR},m)$ -RD codes with mappings g_E^m and g_D^m so that $\lim_{m\rightarrow\infty}\mathbb{E}[d(X^m,g_D^m\{g_E^m(Y^m)\})] \leq D$. Let $\mathcal{R}$ be the closure of the set of achievable (R, D) pairs.

Definition 2.

(Lossy CS source coding problem) Amongst all the E-D pairs of mappings (110) and (111), determine the infimum of (achievable) rates R such that X can be reproduced with the average distortion satisfying E $\mathbb{E}[d(Xm,X\hat{}m)]\leq D+\epsilon$ for any positive real number , i.e., define [74, Sect. 10.2]

(112)$RXrem(D)=\inf(R,D)\in RR$

RXrem(D) is called the remote RDF of source X.

It is worth noting that the MSE distortion, in general, leads to non-sparse signal reconstruction, which might be undesirable in certain CS applications.

7.2.4 Remote RDF

The general expression of the remote RDF for a discrete memoryless source with discrete memoryless observations has been derived in [25, Eqs. (3.5.1) –(3.5.5)]. Adapting the result to continuous-valued signals X and Y, RXrem(D) in (112>) can be expressed as

(113a)$RXrem(D) =\min f(x\hat{}|y):\mathbb{E}[d(X,X\hat{})]\leq D1N\ I(Y; X\hat{})$

where the optimization is over the conditional PDF fX^|y , commonly referred to as the test channel, and d(x, X^) is the distortion in (108). The mutual information19 between Y and x̂ is

(113b)$I(Y;X\hat{})=\int y\int x\hat{}f(y)f(x\hat{}|y)\log f(x\hat{}|y)f(x\hat{})dydx\hat{}$

and the average MSE distortion between X and x̂ is

(113c)$\mathbb{E}[d(X,X\hat{})]=(a)\int x\int y\int x\hat{}f(x)f(y|x)f(x\hat{}|y)d(x,x\hat{})dxdydx\hat{}$Figure 29:

A single-sensor lossy CS of a sparse source with support side information

where (a) follows from $f(x\hat{}|y)=f(x\hat{}|y,x)$ because X → Y → X forms a Markov chain. Note that the remote sensing mechanism is captured by the conditional PDF f(y|x) , governed by the measurements in (107).

Due to the time-varying sparsity of $\{Xn\}n=1\infty$ through $\{Bn\}n=1\infty$, the PDFs of X, and consequently, of Y are mixture distributions, which seems to make the optimization over f(x^|y) in (113a) difficult. Hence, we treat the lossy CS problem of Definition 2 with the following two approaches: 1) analytically tractable lower bound to RXrem(D) is derived in Section 7.3, and 2) a method to numerically approximate RXrem(D) is devised in Section 7.4. Note that similar difficulty resides also in the direct compression of X, in which case only RD bounds have been derived [57,293,230].

7.3 Rate-Distortion Lower Bound for Lossy CS

Consider the compression setup of Fig. 29, where, compared to Fig. 28, the encoder E_{si} and decoder D_{si} possess SI on sequence $\{B_n\}_{n=1}^{\infty}$. Such an SI aided setup can be used to derive a lower bound to $R_X^{rem}(D)$ in (113a). Having the support SI at the decoder is often optimistic in practice, but sometimes the encoder may acquire SI on B (i.e., an estimate $\hat{B}$) from the measurements Y at a moderate cost via a sparse signal reconstruction algorithm (see Section 2.2). Nevertheless, the shared support SI allows to derive an analytically tractable lower bound to $R_X^{rem}(D)$ which sheds light on the RD behavior of the original setup in Fig. 28, and establishes a benchmark for practical coding methods.

7.3.1 Lossy CS Problem with Support SI

Owing to the support SI, an informed lossy source code is defined as follows [131, 174, 73] and [126, Sect. 2.3.1]. The encoder E_{si} observes a block of measurements $y^m \in \mathcal{Y}^m$ along with the SI $b^m \in \mathcal{B}^m$ and compresses it to a message index $u \in \mathcal{U}$ using an encoder mapping

$$g_E^{si}: \mathcal{Y}^m \times \mathcal{B}^m \to \mathcal{U}. \tag{114}$$

The decoder D_{si} uses the index and the common SI b^m to reconstruct an estimate of $x^m \in \mathcal{X}^m$ via a decoder mapping

$$g_D^{si}: \mathcal{U} \times \mathcal{B}^m \to \hat{\mathcal{X}}^m \tag{115}$$

A pair (R, D) for distortion D ≥ 0 is achievable if there exists a sequence of informed $(2^{mNR}, m)$ -RD codes with mappings g_E^{si} . and g_D^{si} so that $\lim_{m\to\infty} \mathbb{E}[d(X^m, g_D^{si}(g_E^{si}(Y^m, B^m), B^m))] \le D$. Let $\mathcal{R}_{si}$ be the closure of the set of such achievable (R, D) pairs.

Definition 3.

(Lossy CS source coding problem with support SI) Amongst all the E_{si} - D_{si} pairs of mappings (114) and (115), determine the infimum of (achievable) rates R so that X can be reproduced with the average distortion satisfying $\mathbb{E}[d(X^m, \hat{X}^m)] \le D + \epsilon$ any positive real number e, i.e., define

$$R_{x|B}^{rem}(D) = \inf_{(R,D)\in\mathcal{R}_{si}} R. \tag{116}$$

$R_{x|B}^{rem}(D)$ is called the conditional remote RDF of source X. Clearly, $\mathcal{R} \subseteq \mathcal{R}_{si}$, and $R_{x|B}^{rem}(D)$ establishes a lower bound to the best possible compression performance of the lossy CS as

$$R_x^{rem}(D) \ge R_{x|B}^{rem}(D). \tag{117}$$

Next, we present detailed derivation of $R_{x|B}^{rem}(D)$

7.3.2 Conditional Remote RDF

The conditional RDF for a discrete source along with the respective coding theorems has been derived in [131]. Extending the results to a remote com-pression setup, the conditional remote RDF Rx|Brem(D) can be expressed as

(118a)RX|Brem(D)=min{f(x^|y,bs)}s=1|B|:𝔼[d(X,X^)]≤D1NI(Y;X^|B)

where the optimization is over the |B| different test channels f(X^|y, bs), s = 1, . . . , |B|, the conditional mutual information between Y and X given B is

(118b)I(Y;X^|B)=∑s=1|𝓑|p(bs)I(Y;X^|B=bs)

and the average MSE distortion between X and Xˆ is

(118c)𝔼[d(X,X^)]=∑s=1|𝓑|p(bs)𝔼[d(X,X^)|B=bs]

where, compared to (113c), the expectation is also taken over B. Since B is provided at no cost in the system shown in Fig. 29, RX|Brem(D) determines the complementary information rate that must be conveyed to the decoder D?ᵢ to reconstruct X with fidelity D.

We can observe that (118b) and (118C) decompose with respect to real-izations B = b?, s = 1,..., |B|. The conditional remote RDF RX|Brem(D) in (118a) can thus be expressed as the weighted sum minimization [131, Theorem 5]

(119)RX|Brem(D)=min∑s=1|B|p(bs)Ds=DDs≥0,s=1,...,|B|∑s=1|B|p(bs)RX|Brem(Ds)

with optimization variables D?, s = 1,..., |𝓑|, where RX|bsremDs is the conditional marginal remote RDF of source X for a fixed realization B = b? and distortion D? ≥ 0, given as

(120a)RX|bsrem(Ds)=minf(x^|y,bs):𝔼[d(X,X^)|B=bs]≤Ds1NI(Y;X^|B=bs)

where the mutual information between Y and X, conditioned on B = bE?, is

(120b)I(Y;X^|B=bs)=∫y∫x^f(y|bs)f(x^|y,bs)logf(x^|y,bs)f(x^|bs)dy dx^

and the average MSE distortion between X and x̂, conditioned on B = b?, is

(120c)𝔼[d(X,X^)|B=bs]=(a)∫x∫y∫x^f(x|bs)f(y|x,bs)f(x^|y,bs) d(x,x^)dxdydx^

where (a) follows from f(x̂|y, x, b?) = f(x̂|y, b?) because X → Y → X̂ forms a Markov chain when conditioned on B. Owing to the support SI, all

Figure 30:

A single-sensor lossy CS of a sparse subsource with support side information..

the PDFs above are equivalent to those in (113), except they are conditioned on the realization $B = b_s$.

Based on the above formulations, the characterization of $R^{rem}_{X|B}(D)$ in (119) boils down to deriving each $R^{rem}_{X|b_s}(D), s=1,...,|B|$ in (120a). This is carried out in the next section.

7.3.3 Conditional Marginal Remote RDF

The conditional marginal remote RDF $R^{rem}_{X|b_s}(D), s=1,...,|B|$, in (120a) determines the minimum (achievable) rate R_s so that X can be reproduced with the average distortion satisfying $\mathbb{E}[d(X,\hat{X})|B=b_s]\le D_s$ in the setup depicted in Fig 30, where $\sum_{s=1}^{|\mathcal{B}|}p(b_s)R_s=R$. In order to derive $R^{rem}_{X|b_s}(D_s)$, we introduce the following three definitions.

Definition 4.

(Subsource) Let $\{\bar{X}_{s,n}\}_{n=1}^{\infty}=\{G_n\odot b_s\}_{n=1}^{\infty}$ be the memory less sequence of the sth subsource, consisting of K-sparse source vectors $\{X_n\}_{n=1}^{\infty}$ restricted to a fixed realization $B = b_s$, s = 1,..., $|\mathcal{B}|$. Each subsource $\bar{X}_s$ comprises of two parts: 1) the length-K random vector

$$X_s \triangleq G_{supp(b_s)} \sim N(0,\Sigma_{X_s}) \tag{121}$$

that extracts the entries of $\bar{X}_s$ (i.e., the entries of G) restricted to the support of b_s, where $supp(b_s)\triangleq\{k\in\{1,\dots,N\}|b_{s,k}\neq 0\}$ denotes the sup¬port of vector b_s, $G_{supp(b_s)}$ extracts the entries G_k from G for indices $k\in supp(b_s)$, and the covariance matrix $\Sigma_{X_s}\in S^K_{++}$. extracts the entries $\Sigma_G(k,k')$ from Σ_G for indices $k,k'\in supp(b_s)$; 2) the all-zero vector 0_{N-K} corresponding to the entries of $\bar{X}_s$ for indices $k\in supp(b_s)^c$, where $supp(b_s)^c\triangleq\{1, . . . , N\}\setminus supp(b_s)$ is the complement of $supp(b_s)$.

It is worth noting that the subsources $\bar{X}_s$, are virtual, i.e., not actually present in the system. However, they play an instructive role in the derivations. Due to the decomposability, the subsources can be seen as a composite source[20] [25, Sect. 6.1.1] and [126].

Definition 5.

(Measurements of a subsource) Let $\{Y_{s,n}\}_{n=1}^{\infty}$ be the memoryless sequence of the measurements of form (107) restricted to a fixed realization $B = b_s$, i.e., the measurements of subsource $\bar{X}_s, s=1,...,|B|$, in (121), defined as

$$Y_s\triangleq\Phi\bar{X}_s+W=\Phi_sX_s+W \tag{122}$$

where matrix $\Phi_s \in \mathbb{R}^{M\times K}$ extracts the K columns of Φ with indices $k \in \mathrm{supp}(b_s)$, and thus, $Y_s \sim \mathcal{N}(0, \Sigma_{Y_s})$ with a covariance matrix $\Sigma_{Y_s} = \Phi_s \Sigma_{X_s} \Phi_s^T + \Sigma_W \in \mathbb{S}_{++}^M$.

Definition 6.

(MMSE estimator of a subsource) Let $\bar{Z}_s$ be a length-N random vector representing the MMSE estimator of source X given Y for a fixed realization $B = b_s$, i.e., the MMSE estimator of subsource $\bar{X}_s$ in (121) given Y_s in (122). Each Z_s is given by the conditional expectation as [229, Sect. 8.2]

$$\bar{Z}_s \triangleq \mathbb{E}[X \mid Y, B=b_s], \quad s=1,\ldots,|\mathcal{B}|, \tag{123}$$

which, owing to the sparsity of $\bar{X}_s$ (cf. (121)), splits into two parts: 1) the length-K random vector

$$Z_s \triangleq \mathbb{E}[X_s \mid Y, B=b_s] = \Sigma_{X_sY_s}\Sigma_{Y_s}^{-1}Y_s = F_sY_s \sim \mathcal{N}(0, \Sigma_{Z_s}) \tag{124}$$

that represents the MMSE estimator of X_s given Y and $B = b_s$; 2) 0_{N-K} that corresponds to the MMSE estimator of the zero part of $\bar{X}_s$. For jointly Gaussian random vectors, Z_s is linear [120, Sect. 10.2], where the cross-covariance matrix is $\Sigma_{X_sY_s} = \Sigma_{X_s}\Phi_s^T \in \mathbb{R}^{K\times M}$, the MMSE estimation matrix is $F_s \triangleq \Sigma_{X_sY_s}\Sigma_{Y_s}^{-1} \in \mathbb{R}^{K\times M}$ and $Z_s \sim \mathcal{N}(0, \Sigma_{Z_s})$ with covariance matrix $\Sigma_{Z_s} = F_s\Sigma_{X_sY_s}^T \in \mathbb{S}_+^K$.

$R^{\mathrm{rem}}_{X|b_s}$ in (120a) can be characterized by a two-stage encoding structure, where the encoder first optimally estimates the subsource $\bar{X}_s$ (see (121)) from measurements Y_s (see (122)), and then optimally encodes the constructed estimator $\bar{Z}_s$ in (124). This is elaborated next.

MMSE Distortion Separation

Let $\hat{\bar{X}}_s$ be a length-N random vector representing the reproduction of subsource $\bar{X}_s$ at the decoder output (see Fig. 30). Accordingly, the average conditional MSE distortion $\mathbb{E}[d(\bar{X}_s, \hat{\bar{X}}_s)] \overset{\Delta}{=} \mathbb{E}[d(X, \hat{X}) \mid B=b_s]$ (120c) separates as

$$\mathbb{E}[d(\bar{X}_s,\hat{\bar{X}}_s)] \overset{\Delta}{=} \mathbb{E}[d(X,\hat{X})\mid B=b_s] \; \mathbb{E}[d(\bar{X}_s,\hat{\bar{X}}_s)] = N^{-1}\mathbb{E}[\|\bar{X}_s-\bar{Z}_s+\bar{Z}_s-\hat{\bar{X}}_s\|_2^2] \overset{(a)}{=} N^{-1}\mathbb{E}[\|\bar{X}_s-\bar{Z}_s\|_2^2]+N^{-1}\mathbb{E}[\|\bar{Z}_s-\hat{\bar{X}}_s\|_2^2] \overset{(b)}{=} D_{Z|b_s}+\mathbb{E}[d(\bar{Z}_s,\hat{\bar{X}}_s)] \tag{125}$$

where (a) follows from the MMSE orthogonality principle [229, Sect. 8.2.1] (see Appendix C); (b) follows by denoting the (rate-dependent) average MSE distortion between $\bar{Z}_s$ and $\hat{\bar{X}}_s$ as $\mathbb{E}[d(\bar{Z}_s,\hat{\bar{X}}_s)] = N^{-1}\mathbb{E}[\|\bar{Z}_s-\hat{\bar{X}}_s\|_2^2]$, and defining the (rate-independent) average MMSE estimation error with respect to subsource $\bar{X}_s$ as [120, Sect. 10.2] (see Appendix D)

$$D_{Z|b_s} = N^{-1}\mathrm{Tr}(\Sigma_{X_s} - \Sigma_{Z_s}). \tag{126}$$

Similar separation appears also in, e.g., [85, 253, 304].

Reduced Distortion

Due to the decomposability of (124), the last term in (125) splits as

$$\mathbb{E}[d(\bar{Z}_s,\hat{\bar{X}}_s)]=\mathbb{E}[d(Z_s,\hat{X}_s)]+\mathbb{E}[d(\mathbf{0}_{N-K},\hat{\bar{X}}_{\mathrm{supp}(b_s)^c})] \tag{127}$$

where $\hat{X}_s$ is the length-K reproduction random vector associated with $\bar{X}_s$, and $\hat{\bar{X}}_{\mathrm{supp}(b_s)^c}$ is the reproduction random vector associated with the zero part of $\bar{X}_s$. Since an RDF is a monotonic nonincreasing function of the distortion [25, Sect. 2], it is optimal for $R^{\mathrm{rem}}_{X|b_s}(D_s)$ to set $\hat{\bar{X}}_{\mathrm{supp}(b_s)^c}=\mathbf{0}_{N-K}$, and thus the distortion in (125) reduces to

$$\mathbb{E}[d(\bar{X}_s,\hat{\bar{X}}_s)]=D_{Z|b_s}+\mathbb{E}[d(\bar{Z}_s,\hat{\bar{X}}_s)]. \tag{128}$$

Let $D'_s \geq 0$ be a reduced distortion criterion for the sth subsource as

$$D'_s \triangleq D_s-D_{Z|b_s}\geq 0,\quad s=1,\dots,|\mathcal{B}| \tag{129}$$

where $D_s \geq 0$ is the distortion criterion in (120a), and $D_{Z|b_s}$ is given in (126). Note that according to (128), $\mathbb{E}[d(Z_s,\hat{X}_s)]\leq D'_s$ implies $\mathbb{E}[d(\bar{X}_s,\hat{\bar{X}}_s)]\leq D_s$.

Estimate-and-Compress Separation

Let $R^{\mathrm{dir}}_{Z|b_s}(D'_s)$ denote the direct RDF of the MMSE estimator Z_s defined in (124) for the reduced distortion D'_s in (129), i.e., define

$$R^{\mathrm{dir}}_{Z|b_s}(D'_s)=\min_{f(\hat{z}_s|z_s):\,\mathbb{E}[d(Z_s,\hat{Z}_s)]\leq D'_s}\frac{1}{N}I(Z_s;\hat{Z}_s) \tag{130a}$$

where the minimization is over the test channel $f(\hat{z}_s|z_s)$, $\hat{Z}_s$ is a length-K reproduction random vector for Z_s, the average mutual information between Z_s and $\hat{Z}_s$ is

$$I(Z_s;\hat{Z}_s)=\int_{z_s}\int_{\hat{z}_s}f(z_s)f(\hat{z}_s|z_s)\log\left(\frac{f(\hat{z}_s|z_s)}{f(\hat{z}_s)}\right)dz_s\,d\hat{z}_s \tag{130b}$$

and the average MSE distortion between Z_s and $\hat{Z}_s$ is

$$\mathbb{E}[d(Z_s,\hat{Z}_s)]=\int_{z_s}\int_{\hat{z}_s}f(z_s)f(\hat{z}_s|z_s)d(z_s,\hat{z}_s)\,dz_s\,d\hat{z}_s. \tag{130c}$$

The RDF $R^{\mathrm{dir}}_{Z|b_s}(D'_s)$ can be derived by decorrelating the Gaussian (effective) source Z_s via the Karhunen-Loeve transform, and applying reverse water-filling [74, Sect. 10.3.3]. Accordingly, let $\Sigma_{Z_s}=Q_s\Lambda_sQ_s^T$ be the eigendecomposition, where the diagonal matrix $\Lambda_s\triangleq\mathrm{diag}(\lambda_{s,1},\dots,\lambda_{s,K})$ contains the eigenvalues $\lambda_{s,1}\geq\dots\geq\lambda_{s,K}\geq 0$ of $\Sigma_{Z_s}\in\mathbb{S}^K_+$, and the columns of $Q_s\in\mathbb{R}^{K\times K}$ are the corresponding eigen¬vectors. Consequently, $R^{\mathrm{dir}}_{Z|b_s}(D'_s)$ is given as

$$R^{\mathrm{dir}}_{Z|b_s}(D'_s)=\min_{\substack{\sum_{k=1}^K D'_{s,k}=D'_s\\ D'_{s,k}\geq 0,\ k=1,\dots,K}}\frac{1}{N}\sum_{k=1}^K\max\left\{0,\ \frac{1}{2}\log\frac{\lambda_{s,k}}{D'_{s,k}}\right\} \tag{131}$$

where $D'_{s,k}$, $k = 1,\dots,K$, are the optimization variables.

The following proposition gives an expression for the conditional marginal remote RDF $R^{\mathrm{rem}}_{X|b_s}(D_s)$.

Proposition 6:

The conditional marginal remote RDF of $\bar{X}_s$ in (120a) is given as the (direct) RDF of the MMSE estimator Z in (131), i.e.,

(132) $R^{rem}_{X|b_s}(D_s)=R^{dir}_{Z|b_s}(D'_s)$, $s=1,\ldots,|B|$,

where $D'_s=D_s-D_{Z|b_s}\geq 0$ is the reduced distortion in (129), and $D_{Z|b_s}$ is given in (126).

Proof: The proposition follows from the proofs in [85, 304].

Figure 31.

A forward channel to illustrate the optimal compression structure with respect to the conditional marginal remote RDF $R^{rem}_{X|b_s}(D_s)$ in (120a).

According to Proposition 6, the remote source coding problem of Definition 3 separates into 1) the MMSE estimation of $\bar{X}_s$ given Y, and 2) the derivation of the RDF of the resultant estimator. On this account, the best encoder E_i comprises of the MSE-optimal extraction of the subsources $\bar{X}_s$ from the noisy linear measurements Y in (122), s = 1,..., $|\mathcal{B}|$, followed by the optimal coding of the extracted messages. The estimate-and-compress separation is illustrated in Fig. 31.

Remark 7.1.

The two expressions (130) and (131) for $R^{dir}_{Z|b_s}(D'_s)$ are interrelated by Gaussian forward channels depicted in Fig. 31. Namely, the optimal conditional PDF $f(\hat{z}_s|z_s)$ can be described via $\hat{Z}'_{s,k}\triangleq\theta_{s,k}Z'_{s,k}+V_{s,k}$, $k=1,\ldots,K$ with parameters $\theta_{s,k}=\frac{\lambda_{s,k}-D'_{s,k}}{\lambda_{s,k}}$ and $\sigma^2_{V_{s,k}}=\theta_{s,k}D'_{s,k}$, where $D'_{s,k}$ are optimal variables for (131), $Z'_{s,k}$ is the k-th element of the decorrelated MMSE estimator $Z'_{s,k}=Q_s^TZ_{s,k}$, and $V_{s,k}\sim N(0,\sigma^2_{V_{s,k}})$ is a zero mean Gaussian random variable independent of $Z'_{s,k}$ [25, Theorem 4.3.2]. Thus, the forward channel model provides a practical way to realize the optimal conditional PDF and the respective reproduction random vectors.

Remark 7.2.

If $\mathrm{rank}(\Sigma_{Z_s})<K$, then the covariance matrix Σ_{Z_s} has a nullspace. Consequently, the random vector Z is a degenerate Gaussian vector [33, Lecture 7], and Z' contains a deterministic zero part. This case is inher¬ently handled in (131) by allocating $D'_{s,k}=0$ for $k=\mathrm{rank}(\Sigma_{Z_s})+1,\ldots,K$. The null-space may be caused for example by the rank deficiency of matrix Φ

Remark 7.3.

A proof of the optimality of the two-step coding structure is implicitly present in the seminal work by Dobrushin and Tsybakov [85, Sect. 5] for the case with frequency-weighted MSE distortion where the source and observable processes are jointly Gaussian and stationary. Furthermore, they proved such optimality explicitly for the MSE distortion in the case where observations are noisy versions of the signal (i.e., no dimension reduction) [85, Sect. 7]. Later, Wolf and Ziv [304] addressed a DR framework and proved that separation holds for MSE distortion under more general conditions (i.e., Gaussianity is not required). Consequently, the decomposition principle of Proposition 6 is also valid for non-Gaussian sources/observations; however, finding analytical expressions for $R^{\mathrm{dir}}_{Z|b_s}(D'_s)$ and $D_{Z|b_s}$ may be difficult. Similar separation results appear in, e.g., [303, 256, 163] and [25, Sect. 4.5.4].

Remark 7.4.

$R^{\mathrm{rem}}_{X|b_s}(D_s)$ is an upper bound to the conditional marginal remote RDF of a subsource $\bar{X}_s=\tilde{G}\odot b_s$, where $\tilde{G}$ is a non-Gaussian random vector with covariance matrix $\Sigma_{\tilde{G}} = \Sigma_G$ [25, p. 130].

7.3.4 Characterization of the Conditional Remote RDF

Let $D_{Z|B}\geq 0$ denote the total average MMSE estimation error over all sub-sources X_s, s = 1,..., $|\mathcal{B}|$, with the support SI, i.e.,

$$D_{Z|B}\triangleq\sum_{s=1}^{|\mathcal{B}|}p(b_s)D_{Z|b_s} \overset{(a)}{=} N^{-1}\sum_{s=1}^{|\mathcal{B}|}p(b_s)\mathrm{Tr}(\Sigma_{x_s}-\Sigma_{z_s}) \tag{133}$$

where (a) follows from (126). The conditional remote RDF $R^{\mathrm{rem}}_{X|B}(D)$ is given by the following theorem.

Theorem 1.

For distortion range $D_{Z|B}\leq D\leq \frac{1}{N}\sum_{s=1}^{|\mathcal{B}|}p(b_s)\mathrm{Tr}(\Sigma_{x_s})$, $R^{\mathrm{rem}}_{X|B}(D)$ is positive and can be evaluated via the convex minimization problem as

$$R^{\mathrm{rem}}_{X|b_s}(D_s)=\min_{\substack{\sum_{s=1}^{|\mathcal{B}|}p(b_s)\sum_{K=1}^{K}D'_{s,k}=D-D_{Z|\mathcal{B}}\\ D'_{s,k}\geq 0,\,k=1,\ldots,K}} N^{-1}\sum_{s=1}^{|\mathcal{B}|}p(b_s)\sum_{K=1}^{K}\max\left\{0,\ \frac{1}{2}\log\frac{\lambda_{s,k}}{D'_{s,k}}\right\} \tag{134}$$

where Σ_{x_s} is the covariance matrix of X_s in (121); $\lambda_{s,1},\ldots,\lambda_{s,K}$ are the eigenvalues of the covariance matrix Σ_{x_s} of Z_s in (124); $D'_{s,i}$ are the optimization variables, $k=1,\ldots,K$, $s=1,\ldots,|\mathcal{B}|$. If the distortion values satisfy $D\geq \frac{1}{N}\sum_{s=1}^{|\mathcal{B}|}p(b_s)\mathrm{Tr}(\Sigma_{x_s})$ then $R^{\mathrm{rem}}_{X|B}(D)$ is zero.

Proof: The proof is given in Appendix E.

Remark 7.5.

$R^{\mathrm{rem}}_{X|B}(D)$ is an upper bound to the conditional remote RDF of a source $\bar{X}=\tilde{G}\odot B$, where $\tilde{G}$ is a non-Gaussian random vector with covariance matrix $\Sigma_{\tilde{G}}=\Sigma_G$ [25, p. 130].

In Theorem 1, $R_{X|B}^{rem}(D)$ is determined by a weighted sum of the RDFs of the MMSE estimators Z_s under a reduced distortion criterion, where the weights, i.e., the prior probabilities of the sparsity patterns $p(b_s)$, $s = 1,\ldots, |\mathcal{B}|$, represent the "appearance frequencies" of such estimators. In particular, (134) involves finding the optimal allocation of the distortion components not only for the $|\mathcal{B}|$ different sparsity patterns, but also for the K entries of each decorrelated random vector Z'_s. This type of weighted minimization is discernibly a consequence of the composite source structure.

Furthermore, $R_{X|B}^{rem}(D)$ reflects the remote sensing nature of the lossy CS: regardless of the rate, the lowest achievable distortion is ultimately dictated by $D_{Z|B}$ which is a constant term solely governed by the noisy measurement model in (122). This unavoidable degradation in compression performance, which is caused by the indirect observations of the source, distinguishes the lossy CS from directly compressing X; see, e.g., the works in [57, 293, 230] which derive RD bounds for compressing sparse sources. Note that a constant distortion floor occurs whether or not the support SI is available and only the respective levels for $R_{X|B}^{rem}(D)$ and $R_X^{rem}(D)$ are different. This is demonstrated by the numerical results in Section 7.6.

7.4 Numerical Approximation of the Remote RDF

Since finding an analytical solution for the lossy CS problem of Definition 2 turned out to be elusive, we develop a method based on the Blahut-Arimoto (BA) algorithm [30, 11] to numerically approximate the remote RDF $R_X^{rem}(D)$ in (113a). The standard BA algorithm is designed for direct (i.e., for intact measurements Y = X) source coding with discrete input/output alphabets.

Hence, the algorithm needs to be adapted to handle 1) continuous-valued signals X and Y, and 2) the remote compression setup. The former is accom-plished by a VQ-optimized alphabet discretization method, and the latter by appropriately modifying the distortion measure.

7.4.1 Discretization of Signal Alphabets

As the first step, the measurement vector space $\mathcal{Y}$ (i.e., the encoder input) and the reproduction alphabet $\hat{\mathcal{X}}$ (i.e., the decoder output) are discretized via a VQ. Let $\mathcal{V} \triangleq \{1,\ldots,|\mathcal{V}|\}$ be an index set. The $|\mathcal{V}|$ -level VQ is determined by: 1) the encoder regions $\mathcal{S}_v$, $v \in \mathcal{V}$, which partition the measurement space, i.e., $\mathcal{S}_v \subseteq \mathcal{Y}$, $\mathcal{S}_v \cap \mathcal{S}_{v'} = \emptyset$ for any $v \neq v'$, and $\bigcup_1^{|\mathcal{V}|} \mathcal{S}_v = \mathcal{Y}$. 2) the reconstruction codebook $\mathcal{X}_q \triangleq \{\hat{X}_1,\ldots,\hat{X}_{|\mathcal{V}|}\}$ with codevectors $\hat{X}_v \in \mathbb{R}^N$. The VQ encoder is a mapping $E_{vq}: \mathcal{Y} \to \mathcal{V}$. For an input y in $\mathcal{S}_v$, it produces an index $E_{vq}(y) = v$. The VQ decoder performs an inverse mapping $D_{vq}: \mathcal{V} \to \mathcal{X}_q$ as $D_{vq}(v) = \hat{X}_v \in \hat{\mathcal{X}}_q$. The random variable V represents the VQ output.

Since the next section elucidates the main role of the VQ in RD approximation, the optimization of the VQ is deferred until Section 7.4.3.

7.4.2 Modified Blahut-Arimoto Algorithm for Lossy CS

Consider now a VQ as described above with p(v)≜Pr(V=v)=∫y∈Svf(y)dy v∈V Consequently, index v∈V represents all the measurement vectors that belong to VQ region Sv.Similarly, let Xq be a discrete reproduction random vector at the output of decoder D with alphabet X^q={X^1,…,X^|V|} (i.e., the VQ codebook). Replacing Y with V and X^ with X^q in (113a), RXrem(D) can be approximated as

135a RX,barem(D)=minp(x^j)|v:𝔼[d(X,X^q)]≤D1NI(V;X^q)

where the optimization is over the conditional probabilities p(X^j|v)≜Pr(Xq=X^j|V=v),vj∈V . The mutual information between V and X^q is

(135b)I(V;X^q)=∑v=1|V|∑j=1|V|p(v)p(x^j|v)log(p(x^j|v)p(x^j))

and the average distortion between X and Xq is

(135c)𝔼[d(X,X^q)]=∑v0=1|V|∑j=1|V|p(v)p(x^j|v)d¯(X,x^j|v)

where d¯(X,x^j|v)≥0 is the modified distortion measure, defined as the average per-letter MSE distortion between X and X_j conditioned on V = v, i.e.,

135dd¯(X,x^j|v)≜𝔼[d(X,x^j)|V=v], v,j∈V=1N∫y𝔼[‖X−x^j‖22|V=v,Y=y]f(y|v)dy=1N∫yp(v|y)p(v)𝔼[‖X−x^j‖22|V=v,Y=y]f(y)dy=(a)1Np(v)∫y∈Sv𝔼[‖X−x^j‖22|Y=y]f(y)dy

where (a) follows from the Markov chain X→Y→Z , and from p (v|y) = 1, if y ∈ Sv, v ∈ V and 0 otherwise. Note that pre-calculated |V|2 quantities d¯(X,x^j|v) remain fixed in the BA algorithm. In the context of discrete remote sources, a distortion measure similar to (135d) appears in, e.g., [25, Sect. 3.5] and [303, 86].

Consider a Lagrangian for (135a) as

136L({p(x^j|v)}v,j=1|V|,λ,{vv}v=1|V|)=1N∑v=1|V|∑j=1|V|p(v)p(x^j|v)logp(x^j|v)p(x^j)+λ∑v=1|V|∑j=1|V|p(v)p(x^j|v)d¯(X,x^j|v)+∑v=1|V|vv∑j=1|V|p(x^j|v)

where λ > 0 is the Lagrange multiplier associated with the sum distortion con-straint, and vV, v G V, are the Lagrange multipliers associated with the valid conditional probability constraints ∑j=1|V|p(x^j|v)=1,∀v∈V . Following a standard BA procedure, an (R, D) point of RX1 ba(D) in (135a) is obtained by sequentially updating the conditional probabilities p(x^j|v) and the repro¬duction probabilities p(x^j) for a fixed A at each iteration t = 1, 2,... as [74 , Sect. 10.8]

$$p(\hat{x}_j|v)^{t+1} := \frac{p(\hat{x}_j)^t \exp\left(-\lambda \bar{d}(X,\hat{x}_j|v)\right)}{\sum_{j'=1}^{|\mathcal{V}|} p(\hat{x}_{j'})^t \exp\left(-\lambda \bar{d}(X,\hat{x}_{j'}|v)\right)}, \quad \forall v,j \in \mathcal{V} \tag{137a}$$

$$p(\hat{x}_j)^{t+1} := \sum_{j=1}^{|\mathcal{V}|} p(\hat{x}_j|v)^{t+1} p(v), \quad \forall j \in \mathcal{V}, \tag{137b}$$

until convergence, and by evaluating the rate according to (135b) and the distortion according to (135c). Hence, different values for λ sweep the curve for $R^{\text{rem}}_{X,\text{ba}}(D)$, which approximates the remote RDF $R^{\text{rem}}_X(D)$ in (113a) with an accuracy that increases with the number $|\mathcal{V}|$.

The proposed BA method is summarized in Algorithm 10. The algorithm can be terminated when the quantities do not significantly change (e.g., when $\sum_{j=1}^{|\mathcal{V}|}\left(p(\hat{x}_j)^t - p(\hat{x}_j)^{t-1}\right)^2 < \epsilon_{\text{ba}}$ for a pre-defined positive constant $\epsilon_{\text{ba}} > 0$. The algorithm inputs $p(v)$, $\hat{x}_v$, and $\bar{d}(X,\hat{x}_j|v), \quad v,j \in \mathcal{V}$, are the outcomes of the VQ optimization. The optimization of the VQ is addressed next.

Vector Quantization Optimization

In Algorithm [10], the accuracy of distortion evaluation through (135c) is ultimately limited by the $|\mathcal{V}|^2$ fixed quantities $\bar{d}(X,\hat{x}_j|v), \quad v,j \in \mathcal{V}$, , v, j ∈ V, of (135d). Taking this into account, we optimize the $\mathcal{E}_{\text{vq}}$–$\mathcal{D}_{\text{vq}}$ pair such that it minimizes the average MSE distortion between the source X and its $|\mathcal{V}|$-level reproduction X_q. Thus, the VQ design aims at finding the encoder regions $\mathcal{S}_v$ and codevectors x_v, $v \in \mathcal{V}$ as.

$$\begin{aligned}
\{\mathcal{S}_v^*, \hat{x}_v^*\}_{v\in\mathcal{V}} &:= \arg\min_{\{\mathcal{S}_v, \hat{x}_v\}_{v\in\mathcal{V}}} \frac{1}{N}\mathbb{E}\left[\|X-\hat{X}_q\|_2^2\right] \\
&:= \arg\min_{\{\mathcal{S}_v, \hat{x}_v\}_{v\in\mathcal{V}}} \frac{1}{N}\sum_{v\in\mathcal{V}}\int_y p(v|y)\, \mathbb{E}\left[\|X-\hat{x}_v\|_2^2 \,|\, V=v, Y=y\right] f(y)\,dy \\
&\overset{(a)}{:=} \arg\min_{\{\mathcal{S}_v, \hat{x}_v\}_{v\in\mathcal{V}}} \frac{1}{N}\sum_{v\in\mathcal{V}}\int_{y\in\mathcal{S}_v} \mathbb{E}\left[\|X-\hat{x}_v\|_2^2 \,|\, Y=y\right] f(y)\,dy \\[8pt]
&\overset{(b)}{:=} \arg\min_{\{\mathcal{S}_v, \hat{x}_v\}_{v\in\mathcal{V}}} \sum_{v\in\mathcal{V}} p(v)\, \bar{d}(X,\hat{x}_v|v)
\end{aligned} \tag{138}$$

where (a) follows from the Markov chain $X \to Y \to V$, and $p(v|y) = 1$, if $y \in \mathcal{S}_v$, $v \in \mathcal{V}$, and 0 otherwise; (b) follows from (135d).

Remark 7.6.

Besides $\bar{d}(X,\hat{x}_j|v)$, $v, j \in \mathcal{V}$, the VQ affects the final distortion in (135c) through the conditional probabilities $p(\hat{x}_j|v)$, $v, j \in \mathcal{V}$ the variables to be optimized in iterative steps (137a) and (137b). In addition, the VQ affects the rate approximation in (135b) through the index probabilities $p(v)$, $v \in \mathcal{V}$. Therefore, a better approximation of $R^{\text{rem}}(D)$ is achievable by

Algorithm 10 A modified Blahut–Arimoto algorithm for approximating the remote RDF $R^{\text{rem}}_X(D)$

Inputs: a) Lagrange multipliers $\lambda > 0$; b) codevectors X_v, and index probabilities $p(v)$, $v \in \mathcal{V}$; obtained as described in Section 7.4.3; c) modified distortion measures $\bar{d}(X,\hat{X}_j|v)$, of form (135d).

Initializations: a) Set $t := 1$; b) set $p(\hat{X}_j)^t := 1/|\mathcal{V}|$, $j \in \mathcal{V}$.

For a given λ

repeat

1) Update the conditional probabilities $p(X_j|v)^{t+1}$, $v, j \in V$, according to (137a).

2) Update the reproduction probabilities $p(X_j)^{t+1}$, $j \in V$, according to (137b).

3) Set $t:=t+1$.

until a pre-defined stopping criterion is met.

4) Compute the rate R_λ according to (135b), and the distortion D_λ according to (135c).

end for

Output: $R_{X,\bar{e}m}(D)$ curve determined by the (R_λ, D_λ) pairs.

incorporating the VQ optimization in the iterative loop of Algorithm 10, and, thus, generating a unique VQ for each λ. For example, the mapping approach in [249] adapts the reproduction points within the optimization loop. Nevertheless, the presented non-adaptive discretization yields decent accuracy for all distortion values D, as demonstrated in Section 7.6.

Remark 7.7.

Note that if one sets the conditional probabilities (137a) as $p(\hat{x}_j|v)=1 \text{ for } v=j$, and 0 otherwise, the reproduction probabilities become $p(\hat{x}_j)=p(v) \text{ for } v=j$, and 0 otherwise, in (137b). Thus, the distortion (135c) for $R_{X,\bar{e}m}(D)$ becomes equal to the VQ distortion in (138), and the rate (135b) becomes $R=-\sum_{v=1}^{|V|} p(v)\log(p(v))=H(V)$, i.e., the entropy of quantization index V. In this sense, $R_{X,\bar{e}m}(D)$ can be seen as a "noisy VQ" that randomizes the mapping $V \rightarrow X_q$ via conditional probabilities $p(\hat{x}_j|v)$ that determine a noisy channel between the encoder output and decoder input.

The joint optimization over S_v and $\hat{x}_v$, $v \in V$, in (138) is intractable, and thus, a common alternating minimization is used to derive necessary optimality conditions [211, 107, 192, 188]. Then, the proposed VQ is equivalent to the VQ in [265] designed for noiseless channels. The optimal encoder regions for fixed codevectors satisfy a generalized nearest-neighbor condition.

$$S_v^*=\{y: \|z-\hat{x}_v\|_2^2 \le \|z-\hat{x}_{v'}\|_2^2,\ \forall v' \ne v\}, \quad \forall v \in V, \tag{139}$$

where $z\in\mathbb{R}^N$ is the MMSE estimate of X given Y = y (i.e., Z is the MMSE estimator of X given Y), defined as [99, 235, 281] and [98, Sect. 11.5]

$$(140)\quad z \triangleq \mathbb{E}[X|Y=y]=\sum_{s=1}^{|\mathcal{B}|}p(b_s|y)\mathbb{E}[X|Y=y,B=b_s]=\sum_{s=1}^{|\mathcal{B}|}\frac{p(b_s)f(y|b_s)}{\sum_{s'=1}^{|\mathcal{B}|}p(b_{s'})f(y|b_{s'})}z$$

where the conditional PDF $f(y|b_s)$ is Gaussian as ᵊA;$(0,\Sigma_{Ys})$ (see Definition 5), and $z_s\veeeq\mathbb{E}[X|Y=y,B=b_s]$ is the MMSE estimate of X given $Y=y$ and $B=b_s$, which, according to Definition 6, comprises vectors $z_s=F_{sys}\in\mathbb{R}^K$ and 0_{N-K} . Similarly, the optimal codevectors for fixed encoder regions satisfy a generalized centroid condition.

$$141\quad \hat{x}_v^*=\frac{1}{p(v)}\int_{y\in S_v}\mathbb{E}[X|Y=y]f(y)dy, \quad \forall v\in V.$$

The VQ can be trained offline via the iterative Lloyd algorithm [107, 192, 188] by successively applying the necessary optimality conditions (139) and (141) for training data sets; see Section 6.3.4 for training principles.

7.5 Practical Symbol-by-Symbol QCS Methods

Recall that approaching $R^{rem}_X(D)$ in (113a) requires encoding (large) blocks of vectors, which is infeasible in practice. To this end, this section presents the design of practical QCS methods relying on symbol-by-symbol quantization under three different compression strategies: 1) compress-and-estimate, 2) estimate-and-compress, and 3) support-estimation-and-compress strategies. The schemes are depicted in Fig. 32. The methods use different quantization schemes, which are introduced in Section 7.5.1. The details of each method are presented in Section 7.5.2 – 7.5.4.

Figure 32:

A single-source QCS setups relying on (a) compress-and-estimate, (b) estimateand-compress, and (c) support-estimation-and-compress strategies.

Quantization Types

Let U be a length-L random vector with PDF $f(u)$. U is the input of a quantizer Q, determined by 1) encoder regions $S_i, i\in I=\{1,...,|I|\}$, which partition the input as $\bigcup_{i=1}^{|\mathcal{I}|}S_i=\mathbb{R}^L$, and 2) a reconstruction codebook $C=\{C_1,...,C_{|\mathcal{I}|}\}$ with codevectors $c_i\in\mathbb{R}^L$. For a realization u, quantizer encoder $Q:\mathbb{R}^L\to I$ produces an index i = Q(u) if $u \in S_i$. For a received index i, quantizer decoder Q^{-1} produces an estimate of u as $\hat{}=Q^{-1}(i)=c_i\in C$

Uniform Scalar Quantization

A uniform scalar quantizer (USQ) Q_u of rate R' bits/U consists of fixed-length intervals as presented in Algorithm 11. The online compression phase is presented in Algorithm 12. Saturation effects are omitted herein.

Lloyd-Max Quantization

A fixed-rate Lloyd–Max quantizer [211, 192] Q_{lm} of rate R' bits/U optimizes the $|\mathcal{I}| = 2'$ quantization regions and the codebook to minimize the MSE distortion.

Algorithm 11 Uniform scalar quantization (offline).

Input: a) Quantization rate R'; b) input range parameters $q_{hi} \geq q_{lo}$.

Output: Reconstruction codebook $C_u=\{c_{1,u},\ldots,c_{2^{R'},u}\}$ generated as: $c_{1,u}=q_{lo}; c_{i,u}=c_{i-1,u}+\frac{q_{hi}-q_{lo}}{2^{R'}-1}$, for all $2 \leq i \leq 2^{R'}-1$.

and $c_{2^{R'},u}=q_{hi}$.

Algorithm 12 Fixed-rate quantization of an input realization u (online).

Input: Reconstruction codebook C_{fix} (C_{lm} or C_u) .

Encoding: 1) Find the encoding index $i^*=\arg\min_{i\in\mathcal{I}} \|u-c_i^{fix}\|_2^2$; 2) communicate the R'-bit representation of i^* to the decoder.

Decoding: Given the received index i^* , obtain an estimate of u as $\hat{u}=c_{i^*}^{fix}\in C_{fix}$. $\in$ C_{fix} .

$$D'=\sum_{i=1}^{2^{R'}} p(i)\mathbb{E}[\|U-c_i^{lm}\|_2^2 \mid I=i], \tag{142}$$

where I represents the quantization index, and $p(i) \overset{\Delta}{=} \Pr(I=i)$. Q_{lm} can be trained by the iterative Lloyd and LBG algorithms [211, 192, 188], as presented in Algorithm 13. The online compression phase is presented in Algorithm 12.

Variable-Rate Quantization

A variable-rate quantizer Q_{vr} minimizes a distortion-rate cost function (cf.(94))

$$(1-\mu')D' - \mu'\sum_{i\in\mathcal{I}} p(i)\log(p(i)) \tag{145}$$

where D' is given in (142), and $\mu' \in [0, 1]$ is a weighting parameter. Following entropy-constrained scalar/vector quantization (ECSQ/ECVQ) [306, 66], Q_{vr} can be trained as presented in Algorithm 14 (see also Algorithm 9 in Section 6). The online phase is presented in Algorithm 15. The source codebook $H \overset{\Delta}{=} \{h_1,\ldots h_{|H|}\}$, contains in general variable-length binary codewords and it can be generated by, e.g., the Huffman coding [149].

Algorithm 13 Training of a Lloyd–Max quantizer (offline).

Input: a) Sequence $\{u(1),u(2),...\}$ sampled from $f(u)$; b) quantization rate R' .

Initialization: Reconstruction codebook C_{lm} .

Repeat until convergence

1) For given codebook, find the optimal regions by classifying {u(1),u(2),...} as(143) $S_i^{lm}=\{u:\|u-c_i^{lm}\|_2 \leq \|u-c_{i'}^{lm}\|_2, \forall i' \neq i, i \in I\}$ (143).

2) For given regions, find the optimal codevectors as the conditional expectations (144) $c_i^{lm}=\mathbb{E}[U|I=i]=\frac{1}{p(i)}\int_{S_i^{lm}} u f(u) du$ (144).

Output: Reconstruction codebook $C^{lm}=\{C^{lm},...,C_{2^{R'}}^{lm}\}$

Algorithm 14 Training of an entropy-constrained quantizer (offline).

Input: a) Sequence {u(1),u(2),...} sampled from f(u) ; b) quantization levels $2^{R'}$; c) weight parameter μ′ ∈ [0, 1].

Initialization: i) Reconstruction codebook C^{vr} , and ii) index probabilities $p(i)=1/2^{R'}$, i ∈ I.

Repeat until convergence

1) For given codebook and rate measures, find the optimal regions by classifying {u(1),u(2),...} as (146)$S_i^{vr}=\{u:(1-\mu')\|u-c_i^{vr}\|_2^2-\mu'\log(p(i)) \leq (1-\mu')\|u-c_{i'}^{vr}\|_2^2-\mu'\log(p(i')), \forall i' \neq i\}, \quad i \in \mathcal{I}$..

2) Update the rate measures μ′log(p(i)) , i ∈ I, given the new regions.

3) For given regions, find the optimal codevectors c_i^{vr} , i ∈ I, equivalently as in (144).

Output: a) Reconstruction codebook $C^{vr}==\{c_1^{vr},...,c_{2^{R'}}^{vr}\}$ b) index probabilities p(i), i ∈ I; c) source codebook $H=\{h_1,\ldots h_{|H|}\}$ generated using p(i), i ∈ I.

Algorithm 15 Variable-rate quantization of an input realization u (online).

Input: a) Reconstruction codebook C^{vr} ; b) index probabilities p(i) , i ∈ I; c) weight parameter μ′ ∈ [0, 1]; d) source codebook $\mathcal{H}$.

Encoding: 1) Find the encoding index $i^*=\arg\min_{i \in \mathcal{I}}((1-\mu')\|u-c_i^{vr}\|_2^2-\mu'\log(p(i)))$; 2) communicate the binary codeword h_* to the decoder.

Decoding: Given the received codeword h_* , find the corresponding index i_* and obtain an estimate of u as $\hat{u}=c_i^{vr} \in C^{vr}$.

Compress-And-Estimate QCS Methods

A compress-and-estimate (C&E) scheme depicted in Fig. 32(a) consists of two stages: 1) a compression stage where the encoder quantizes Y under an MSE distortion criterion that depends only on Y (not on X), and 2) an estimation stage where the decoder estimates X from the decoded quantized measurements $\hat{Y}=Q^{-1}(Q(Y))$. Let $\hat{X}_{C\&E}$ denote a length- N reproduction random vector at the decoder output. The end-to-end MSE distortion can be written as

$$D_{C\&E}=\mathbb{E}[\|X-\hat{X}_{C\&E}\|_2^2]=\mathbb{E}[\|X-R_D(Q^{-1}(Q(Y)))\|_2^2] \tag{147}$$

where $R_D:\hat{Y}\rightarrow\hat{X}_{C\&E}$ is a signal reconstruction algorithm at D. A C&E principle underlies many early QCS algorithms in, e.g., [151, 78]. An information-theoretic study of QCS under C&E scheme is addressed in [164]. Four C&E algorithms with 1) USQ, 2) Lloyd–Max SQ (referred to as "SQ" in the QCS method names henceforth), 3) ECSQ, and 4) VQ are considered. The details of each method are summarized in Table 4. The table lists the used offline training algorithms with the key parameters, i.e., the quantizer input vector U of length L , and the total average rate $R_{tot}=NR$ bits/X , where the rate R is defined as the bits/entry of X .

To summarize the main operations of SQ-based methods 1)–3), the decoder receives M indices, forms $\hat{y}=[c^{C\&E}_{1,i^*}\cdots c^{C\&E}_{M,i^*}]^T$, and estimates the source as $\hat{X}=R_D(\hat{Y})$. In particular, the MSE-optimal R_D for C&Evq can be defined as follows. For C&Evq, the MSE-optimal output vectors of R_D that minimize the distortion in (147) are given as

$$\hat{x}(c_i^{C\&Evq})=\mathbb{E}[X|I=i]=\frac{1}{p(i)}\int_{y\in S_i^{C\&Evq}} z f(y)dy, \quad i\in\mathcal{I} \tag{148}$$

where $z\in\mathbb{R}^N$ is the MMSE estimate of X given $Y = y$ (see (140)).

Estimate-and-Compress QCS Methods

An estimate-and-compress (E&C) scheme depicted in Fig. 32(b) consists of two stages: 1) an estimation stage of X from Y at the encoder, and 2) a compression stage to quantize the resulting signal estimate under an MSE distortion criterion. Let $\hat{X}_{E\&C}$ denote a length- N reproduction random vector at the decoder output. The end-to-end MSE distortion reads as

$$D_{E\&C}=\mathbb{E}[\|X-\hat{X}_{E\&C}\|_2^2]=\mathbb{E}[\|X-Q^{-1}(Q(R_E(Y)))\|_2^2] \tag{149}$$

where $R_E:Y\rightarrow\hat{X}(Y)$ is a reconstruction algorithm at E, and $\hat{X}(Y)$ denotes the estimator of X from Y . Recall that an E&C scheme adheres to the optimal encoding structure for remote source coding [304]. E&C based QCS algorithms have been devised in, e.g., [265, 266, 177, 179, 180, 182, 181] and [184, Sect. V]. The related information-theoretic studies include [165, 183, 184].

Five E&C algorithms with 1) USQ, 2) SQ, 3) ECSQ, 4) VQ, and 5) ECVQ are considered with the details listed in Table 4. For SQ based methods 1)–3), the decoder uses the N received indices to estimate the source as $\hat{y}=[c^{C\&E}_{1,i^*}\cdots c^{C\&E}_{M,i^*}]^T$ for VQ based methods 4)–5), the decoder operates as $\hat{X}=c^{E\&C}_{i^*}$.

Support-Estimation-and-Compress QCS Methods

A support-estimation-and-compress (SE&C) scheme depicted in Fig. 32(c) consists of three stages: 1) a support estimation stage where the encoder estimates B from Y , 2) an estimation stage of X given Y and the support estimator at E, and 3) a two-phase compression stage of the resulting source estimate by a) lossless compression of the support, and b) lossy compression of the non-zero part. Let XSE&Ĉ denote a length- N reproduction random vector at the decoder output. The end-to-end MSE distortion can be expressed as

150DSE&C=𝔼[‖22X−X^SE&C‖22]=𝔼[‖22X−Q−1(Q(SE(Y))+RE(Y|SE(Y)))‖22]Table 4 QCS methods relying on different compression strategies

Class	Method	Training/Online	U	L	Rtot=NR
Compress-and-Estimate	C&Eusq	Alg. 11 / 12	Ym	1	MR′
	C&Esq	Alg. 13 / 12	Rλ Ym	1	MR′
	C&Eecsq	Alg. 14 / 15	Ym	1	∑m=1Mr(HmC&E)
	C&Evq	Alg. 14 / 15	Y	1	R'
Estimate-and-Compress	E&Cusq	Alg. 11 / 12	[X~(Y)]n	1	NR'
	E&Csq	Alg. 13 / 12	[X~(Y)]n	1	N R'
	E&Cecsq	Alg. 14 / 15	[X~(Y)]n	1	∑n=1Nr(HnC&E)
	E&Cvq	Alg. 13 / 12	X^(Y)	N	R'

Class	Method	Training/Online	U	L	Rtot=NR
E&Cecvq	Alg. 14 / 15	X^(Y)	N	r(HE&C)	
Support-Estimation-and-Compress	SE&Cusq	Alg. 11 / 12	[Zs(Y,B˜(Y))]n	1	r(HsuppE&C)+KR'
	SE&Csq	Alg. 13 / 12	[Zs(Y,B˜(Y))]n	1	r(HsuppE&C)+KR'
	SE&Cecsq	Alg. 14 / 15	[Zs(Y,B˜(Y))]n	1	r(HsuppE&C)+∑k=1Kr(HkSE&C)
	SE&Cvq	Alg. 13 / 12	Zs(Y,B˜(Y))	K	r(HkSE&C)+R'

where SE : Y→B~(Y) is a support estimation algorithm at E, and B~(Y) denotes the estimator of B from Y. The sparsity K is assumed to be known by SE(.) . QCS schemes akin to SE&C have been considered in, e.g., [129, 265].

The optimal RE(.) in (150) is the one that minimizes the MSE distortion 𝔼[‖Xs−RE(.)‖22] , where Xs is the sth subsource in Definition 4.

It can be shown that RE(Y|SE(Y))=Zs(Y,B˜(Y)) , where Zs(Y,B˜(Y)) is the MMSE estimator of Xs given Y and B̂(Y) so that (cf. 124).

151zs(y,b˜s)Δ¯¯¯ E[Xs|Y=y,B˜(Y)=b˜s]=ΣXsΦsT(ΦsΣXsΦsT+ΣW)−1y.

Four SE&C algorithms with 1) USQ, 2) SQ, 3) ECSQ, and 4) VQ are considered; see the details in Table 4. Quantizers operate equivalently as in the corresponding E&C schemes. The support estimator B~(Y) is communicated losslessly using a binary source codebook HsuppSE&C , which can be generated by, e.g., the Huffman coding.

7.6 Numerical Results

Numerical results are presented to illustrate the RD behavior of lossy CS, assess the tightness of the lower bound, and compare the performance of the QCS methods of Section 7.5 against the derived limits.

Simulation Setup

Consider setups with $\Sigma_G=\sigma_G^2 I_N$ with $\sigma_G^2=1$, and $\Sigma_W=\sigma_W^2 I_N$ with $\sigma_W^2=0.01$. The following curves are evaluated:

$R^{rem}_{X|B}(D)$:the conditional remote RDF of Theorem 1.

$R^{rem}_{X,ba}(D)$: a numerically approximated remote RDF of Algorithm 10.

$R^{dir}_{X|B}(D)$: the conditional direct RDF of X, corresponding to the lossy compression of X with B available as SI at encoder and decoder, derived in Appendix F (see 293, Sect. VII-A). Note that $R^{rem}_{X|B}(D) \leq R^{dir}_{X|B}(D)$.

$R^{dir}_{X,ba}(D)$: a numerically approximated direct RDF of X representing lossy compression without support SI, obtained via Section 7.4.3 and Algorithm 10 using Y = X. Note that $R^{dir}_{X|B}(D) \leq R^{dir}_{X,ba}(D) \leq R^{rem}_{X,ba}(D)$.

5)$D_{Z|B}$:the average MMSE estimation error in (133) (known B).

6) D_Z :numerically evaluated average MMSE estimation error of X given Y; i.e., $D_Z \overset{\Delta}{=} N^{-1}\mathbb{E}[\|X-Z\|_2^2]$, where the estimator $Z \overset{\Delta}{=} \mathbb{E}[X|Y]$ (unknown B) takes values according to (140).

7) Various QCS methods of Section 7.5 which will be specified for each simulation case later.

The measurement matrix Φ is generated by taking the first M rows of an N × N discrete cosine transform matrix, and normalizing the columns as $\| . \|_2=1$. ECSQs/ECVQs are run with μ' = $0.1/\log(|\mathcal{I}|)$. All binary source codebooks H are generated via the Huffman coding. The distortion is measured as $10\log_{10}(\mathbb{E}[d(X,X_{est})]/N^{-1}\mathbb{E}[\|X\|_2^2])$, where X_{est} is the methoddependent decoded estimate of X, and $\mathbb{E}[\|X\|_2^2]=\sum_{s=1}^{|\mathcal{B}|} p(b_s)\mathrm{Tr}(\Sigma_{X_s})$ The rate is measured as R bits/entry of X. The convex minimization problems are solved via CVX [130].

Complexity: As the complexities of Algorithm 10 and the QCS methods in Section 7.5 increase exponentially with the number of quantization levels, and |B| , the experiments are confined to moderate signal dimensions and quantization rates. It is worth remarking that in respect to any discretized BA algorithm, there is similar complexity issue due to a large number of variables, regardless of the quantization method. In fact, due to the VQ advantages [194], the proposed algorithm enjoys a superior trade-off between the approximation accuracy and the complexity compared to SQ. The complexities of the VQ-based QCS methods can be reduced by low-complexity VQ variants such as tree-structured, multi-step, and lattice VQs [133]. Since Y has a Gaussian mixture density, computationally efficient VQs designed in [272] are also potential candidates.

7.6.2 Rate-Distortion Behavior of Lossy CS

Consider a setup with N=7 , M=5 , K=1 , and equal support probabilities p(bs)=1/|B| , ∀s = 1,..., $|\mathcal{B}|$. Fig.33(d) depicts the average distortion versus the average rate for different compression schemes. The quantization algorithms are run with R'{2,…,212} , and the BA Algorithm 10 with log|V|=12 . Only few VQ-based QCS methods of Section 7.5 are reported here; experiments for all practical QCS methods are presented in Section 7.6.5. In order to obtain high performance, the MMSE estimator in (140) – which is known to have exponential complexity [99] – is used as R_E for E&Cvq and E&Cce-vq . Lower-complexity alternatives are investigated in Section 7.6.5.

Let us first investigate the SI-aided lower bounds RX|Bdir(D),RX,badir(D) , and RX|Brem(D) to the remote RDF RXrem(D) in (113a). Owing to the direct observations with support SI, RX|Bdir(D) appears as the line (in log scale) with slope -6N/K = -42 dB/bit [293] and yields the lowest R for all values of D, as expected. The substantially increased rate for RX,badir(D) compared to RX|Bdir(D) is caused to the necessity of conveying the support of X to the decoder. While RX|Bdir(D) and RX|Brem(D) nearly coincide at high distortion, the curves diverge for moderate to low distortion values. The gradually increasing gap between RX|Brem(D) and RX|Bdir(D) for low D is a fundamental consequence of remote sensing. Note that whereas an arbitrarily small distortion is possible at asymptotically high rates for RX|Bdir(D) and RX,badir(D) (i.e., limD→0RX|Bdir(D)=∞ and limD→0RX,badir(D)=∞), the lowest. achievable distortion for RX|Brem(D) is the MMSE estimation error floor DZ|B (i.e., limD→DZ|BRX|Brem(D)=∞).

Figure 33:

Rate-distortion performance of lossy CS schemes with equal support probabilities for N = 7, K = 1, and the number of measurements (a) M = 2, (b) M = 3, (c) M = 4, and (d) M = 5. The colors and markers of the curves in (b), (c), and (d) are equivalent to those in (a).

Next, focus on the approximate remote RDF RX,barem(D) , i.e., the best achievable performance of any QCS method. The gap between RX|Brem(D) and RX,barem(D) represents the compression loss induced by the random measurements taken without knowing the sparse support in a QCS setup [129]. The tightness of the lower bound is heavily influenced by the signal setup. parameters, as will be exemplified in the subsequent experiments. Despite the gap, the lower bound RX|Brem(D) captures the main peculiarities of the lossy CS: the curve has an almost linear distortion region at low rates, the distortion saturates to the MMSE estimation error floor Dz . As a special remark, the slope of RX,barem(D) at low rates is steeper than the conventional -6 dB/bit due to the sparsity. Note that for a small R, the rate is the most limiting factor to achievable distortion, and, thus, RX,badir(D) nearly coincides with RX,barem(D) ; for higher rates, the impact of noisy compressive measurements increases, thereby degrading the performance of $R_{X,barem}(D)$. Regarding the approximation accuracy of $R_{X,badir}(D)$ and $R_{X,barem}(D)$, observe that the highest obtained rate is $R\approx 1.3$ bits, so the “over-sampling ratio” of the VQ discretization is at least $|V|/2NR\approx 7.5$.

Because the encoder of C&Evq is incognizant of CS, its performance is the worst amongst the QCS methods. The advantages of entropy coding are shown by the E&Cecvq curves which, for moderate rates, approach the compression limit $R_{X,barem}(D)$. As a proof of validity, the E&Cvq method eventually saturates to D_z , which is expected to also happen for the other QCS methods at sufficiently high rates.

7.6.3 Effect of the Number of Measurements

For the setup of Section 7.6.2, Figs. 33(a) – 33(d) illustrate the influence of different numbers of measurements $M=\{2,3,4,5\}$ on the compression performance. As M increases, i.e., the signal-to-noise ratio increases, the level of D_z decreases, and the performance of each method that has no support SI moves closer to the lower bound $R_{X|B}^{rem}(D)$ .The largest gain is achieved when M is increased from M=2 to M=3, whereas the difference between M=4 and M=5 is almost negligible.This matches the CS philosophy: increasing M beyond the value that suffices for accurate CS signal recovery does not bring significant gains. In this respect, provided that M is already at this satisfactory level, it pays off to primarily invest in rate R to meet the given distortion fidelity D. Note that the convergence of the curves to their respective distortion floors is rather similar for all M, and that $R_{X,badir}(D),R_{X|B}^{rem}(D)$, and $D_{Z|B}$ remain unaltered.

Figure 34:

Rate-distortion performance of lossy CS schemes for N = 20, M = 8, K = 2, and power law type support probabilities with parameters (a) α_{pl}= 0.98, (b) α_{pl}= 0.90, and (c) α_{pl}= = 0.72. The colors and markers of the curves in (b) and (c) are equivalent to those in (a).

7.6.4 Effect of Support Probabilities

Consider a setup with N=20,M=80,K=2 and unequal support probabilities as $p(b_s)=\alpha_{pl}^s/\sum_{s'=1}^{|\mathcal{B}|}\alpha_{pl}^{s'},s=1,\ldots,|\mathcal{B}|$. , where $0<\alpha_{pl}\le 1$ is a parameter that adjusts the concentration of the probability mass function of B, and $1>p(b_1)\ge\ldots\ge p(b_{|B|})>0$. For small values of α_{pl}, the probability mass concentrates around a fraction of elements in $B=\{b_1,\ldots,b_{|\mathcal{B}|}\}$, and vice versa. $\alpha_{pl}=1$ corresponds to the uniform distribution, whereas $\alpha_{pl}\to 0$ approaches remote compression of only a single sparse vector. The vectors b_s in alphabet B are ordered so that the decimal number of a binary string represented by b_{s+1} is greater than that of b_s, $s=1,...,|\mathcal{B}|-1$ $(|\mathcal{B}|=190)$

Fig. 34 shows the average distortion versus the average rate for αpl={0.98,0.90,0.72} . Decreasing αpl reduces the uncertainty of the signal support, which improves the compression efficiency. This is seen in the increased decay rate of D for non-SI schemes, the shift of RX,barem(D) toward RX|Brem(D) , and RX,badir(D) towards RX|Bdir(D) , and the diminution of the gap between Dz and DZ|B , which is related to the best possible support recovery for a given setup. This exemplifies that, for a sufficiently concentrated probability mass of B, the E&Cecvq method efficiently encodes sparse vectors from noisy compressive measurements. Its performance approaches the best achievable performance of a support-unaware QCS method (i.e., RX,barem(D)). The result illustrates the MSE separation principle (cf. 125): an efficient QCS method implicitly recovers X from Y and encodes the estimates optimally.

7.6.5 Performance of QCS Methods

Let us now focus on the practical QCS methods. Consider a setup with N=30,M={8,12,16,20},K=2, and equal support probabilities. Basis pursuit denoising (BPDN) [61] is used as a moderate complexity reconstruction algorithm for RD in C&E, RE in E&C, and SE in SE&C; SE forms the support estimate from the indices of the K largest magnitudes of the BPDN output. In other words, the BPDN approximates the MMSE estimate (140) for R_E ; another alternative is the randomized OMP [99]. Note that more sophisticated support recovery algorithms SE could be considered as well.
For C&Eusq and SE&Cusq , qlo and qhi are set as the minimum and maximum codepoints of the

Figure 35:

Rate-distortion performance of QCS schemes with equal support probabilities for N = 30, K = 2, and the number of measurements (a) M = 8, (b) M = 12, (c) M = 16, and (d) M = 20.

corresponding Lloyd–Max SQs. The same procedure is first used for E&Cusq , after which each codepoint is added a constant shift so that E&Cusq contains the codepoint of E&Csq that is closest to zero. The quantization algorithms are run with R′={21,…,210} , and the BA Algorithm 10 with log|V|=10 .

Fig. 35 depicts the average distortion versus the average rate for different QCS schemes for M={8,12,16,20} . The performances of the SQ-based methods can be ranked as E&C < C&E < SE&C, and the VQ-based methods as SE&C < C&E < E&C. Next, we elaborate each of the three classes in more detail.

C&E: C&Eusq and C&Esq perform nearly identically. The entropy coding in C&Eecsq can only slightly improve the performance. For large values of

M, C&Evq slightly outperforms E&Cvq even though E&Cvq uses the optimal E&C structure. This is caused by the fact that RE in E&C is approximated as the BPDN instead of the MMSE estimator in (140). Recall that while C&Evq uses a suboptimal C&E strategy, it uses the MSE-optimal outputs in (148) for a given VQ.

E&C: Among all QCS methods, entropy coding has the most significant impact on E&Csq → E&Cecsq , which stems from the fact that E&Csq quantizes the zero inputs inefficiently. The best practical QCS method is the E&Cecvq method which approaches the compression limit RX,barem(D) .

SE&C: It can be seen that the adaptive compression of the support set and estimated magnitudes is an effective strategy as the SE&C methods outperform all SQ-based C&E and E&C methods. Because all SE&C curves nearly coincide, the support recovery performance of SE seems to be the limiting factor in SE&C.

In general, as M increases, the QCS methods perform better and the gap between the analytical MMSE estimation floor with support side information (the lower horizontal line DZ|B and the error floor of the BPDN reconstruction (the upper horizontal line) decreases. Note that only C&E and E&C methods relying plainly on the BPDN outputs saturate to the upper floor; because SE&C removes the noise outside the (estimated) support of each BPDN output, it can surpass the floor, as illustrated by Figs. 35(b)-(d).

7.7 Conclusions

This section addressed lossy compression of single-sensor CS from the remote source coding perspective. By providing support SI to the encoder and decoder, a conditional remote RDF that establishes a compression lower bound for a finite-rate CS setup was derived. The best such encoder separates into an MMSE estimation step and an optimal transmission step. A modified BA algorithm was developed to numerically approximate the remote RDF, and, thus, to assess the best attainable compression performance of any practical QCS method. The main RD characteristics of the lossy CS were demonstrated by comparing the performance of various practical QCS methods against the derived limits.

Simulation results showed that when SQ is used, an adaptive compression of the support set and estimated magnitudes is an effective strategy. When VQ/ECVQ is used, the estimate-and-compress strategy, as supported by the theory, is the best one. Accordingly, the ECVQ-based estimate-and-compress method was numerically shown to approach the remote RDF.

As illustrated by the numerical experiments, the sparsity of a signal is a feature that provides substantial compression gains, and, thus, energy savings in a CS-based acquisition setup. Accordingly, finding an appropriate trade-off point in terms of the compression rate and tolerable distortion in a low-power sensor application with sparse signals is worth striving for.