POLITECNICO DI TORINO
Repository ISTITUZIONALE

Graded quantization for multiple description coding of compressive measurements

(Article begins on next page)

19 April 2024

Graded quantization for multiple description coding of compressive measurements

Diego Valsesia, *Student Member, IEEE,* Giulio Coluccia, *Member, IEEE,* and
Enrico Magli, *Senior Member, IEEE*

Abstract—Compressed sensing (CS) is an emerging paradigm
for acquisition of compressed representations of a sparse signal.
Its low complexity is appealing for resource-constrained scenarios
like sensor networks. However, such scenarios are often coupled
with unreliable communication channels and providing robust
transmission of the acquired data to a receiver is an issue.
Multiple description coding (MDC) effectively combats channel
losses for systems without feedback, thus raising the interest in
developing MDC methods explicitly designed for the CS frame-
work, and exploiting its properties. We propose a method called
Graded Quantization (CS-GQ) that leverages the democratic
property of compressive measurements to effectively implement
MDC, and we provide methods to optimize its performance. A
novel decoding algorithm based on the alternating directions
method of multipliers is derived to reconstruct signals from
a limited number of received descriptions. Simulations are
performed to assess the performance of CS-GQ against other
methods in presence of packet losses. The proposed method is
successful at providing robust coding of CS measurements and
outperforms other schemes for the considered test metrics.

I. INTRODUCTION

Compressed sensing (CS) [1] is a novel theory for sam-
pling and acquisition of sparse and compressible signals. The
traditional paradigm based on sampling a signal according to
the Nyquist/Shannon theorem followed by compression can
be replaced by the acquisition of a small number of linear
measurements, in the form of random projections. This is
very appealing for low-complexity systems, such as low-power
sensor motes, where classic acquisition followed by compres-
sion could be expensive in terms of energy consumption and
computational demands. Such systems typically need to trans-
mit the acquired data to a receiver over unreliable channels,
thus raising the issue of robustness of the transmission against
channel losses and at the same time imposing a constraint
on the complexity of the adopted methods. The framework
of multiple description coding (MDC) allows to increase the
robustness by creating multiple correlated representations of
the data to be transmitted. The quality of the decoded data
will then depend on the number of descriptions available at
the receiver, where *side decoders* can recover a lower quality
version of the data from few received descriptions, whereas
the *central decoder* can achieve the maximum quality when
all the descriptions are received. A few works studied the
problem of MDC in the CS framework. An early work on the

topic [2] borrows from the theory of sparse representations to
derive a two-description coding scheme through a frame syn-
thesis operator. However, the work does not consider random
projections. The work in [3] presents a method to generate
descriptions of an image by sensing each of them separately,
i.e., preprocessing the image to split it into two sub-images
and then sense the wavelet coefficients of each separately. We
will not consider this kind of approach in the paper because
we want to explicitly avoid any preprocessing of the signal
before acquisition. Indeed, compressive measurements could
be directly obtained by specialized hardware (*e.g.*, [4]–[6]),
thus hindering any preprocessing of the data. Deng *et al.*,
[7] argue that the democratic property of random projections
makes compressive sensing image coding robust to channel
losses. They just partition the measurements into packets so
that the quality of the decoded image will depend on the
number of received packets. However, they do not consider
a practical packetization problem, as they employ very small
packets containing few measurements for each transmission.
Such system is actually not sensible due to the high packeti-
zation overhead.

This paper builds on the MDC mechanism called graded
quantization (CS-GQ), originally proposed in [8], exploiting
the democratic property of compressive measurements. The
principle behind CS-GQ is that multiple copies of the measure-
ment vector can be used as descriptions. Each description is
then partitioned into sets of samples, and each set is quantized
with different quality. The principle is similar to [9]–[19], but
has never been applied in conjunction with CS.

In this paper, we give several novel contributions with
respect to [8]. First, we discuss how methods based on
segmentation of measurements such as [7] are special cases
of CS-GQ and, contrary to the present literature on MDC
for CS, we carefully consider packetization issues. The use
of CS for low-complexity and low-energy sensor networks
motivates us to pay particular attention to the use of the
proposed method with common data link layer protocols such
as IEEE 802.15.4. Moreover, we propose techniques to im-
prove the performance of CS-GQ, to optimize its parameters,
and we introduce a novel algorithm based on the alternating
directions method of multipliers (ADMM) [20] to implement
the side decoder. On the theoretical side, perfomance analysis
is conducted by providing reconstruction guarantees for the
side decoder, using both the actual decoder and an ideal
version based on the oracle decoder, which serves as a limit
case for performance assessment. On the experimental side,
simulations are performed for various scenarios. The paper

is organized as follows: Sec. II explains some introductory background concepts, Sec. III introduces CS-GQ and Sec. IV the side decoding algorithm based on ADMM. Sec. V presents theorems guaranteeing stable side decoding and discusses the performance of oracle-based side and central decoders. Sec. VI deals with the optimization of the CS-GQ parameters as function of the channel loss probability and discusses packetization issues. Finally, Sec. VII reports the results of simulations comparing various MDC methods, the performance of CS-GQ in presence of channel losses using two different channel models and shows that the robustness of multiple descriptions can improve scene recognition accuracy in a sensor network application.

II. BACKGROUND

A. Compressed sensing

In the standard CS framework, introduced in [1], [21], a signal $\mathbf{x} \in \mathbb{R}^{n \times 1}$ which has a sparse representation in some basis $\Psi \in \mathbb{R}^{n \times n}$, *i.e.* $\mathbf{x} = \Psi\boldsymbol{\theta}$, $\|\boldsymbol{\theta}\|_0 = k$, $k \ll n$ can be recovered from a smaller vector of noisy linear measurements $\mathbf{y} = \Phi\mathbf{x} + \boldsymbol{e}$, $\mathbf{y} \in \mathbb{R}^{m \times 1}$ and $k < m < n$, where $\Phi \in \mathbb{R}^{m \times n}$ is the *sensing matrix* and $\boldsymbol{e} \in \mathbb{R}^{m \times 1}$ is the vector representing additive noise such that $\|\boldsymbol{e}\|_2 < \varepsilon$, using [22]

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \quad \|\Phi\Psi\boldsymbol{\theta} - \mathbf{y}\|_2 < \varepsilon \qquad (1)$$

and $\hat{\mathbf{x}} = \Psi\hat{\boldsymbol{\theta}}$, known as Basis Pursuit DeNoising, provided that $m = O(k\log(n/k))$ [21] and that each submatrix consisting of k columns of $\Phi\Psi$ is (almost) distance preserving [23, Definition 1.3]. The latter condition is the *Restricted Isometry Property* (RIP). Formally, the matrix $\Phi\Psi$ satisfies the RIP of order k if $\exists \delta_k \in (0, 1]$ such that, for any $\boldsymbol{\theta}$ with $\|\boldsymbol{\theta}\|_0 \leq k$:

$$(1 - \delta_k)\|\boldsymbol{\theta}\|_2^2 \leq \|\Phi\Psi\boldsymbol{\theta}\|_2^2 \leq (1 + \delta_k)\|\boldsymbol{\theta}\|_2^2,$$

where δ_k is the RIP constant of order k. It has been shown in [24] that when Φ is an i.i.d. random matrix drawn from any subgaussian distribution and Ψ is an orthogonal matrix, $\Phi\Psi$ satisfies the RIP with overwhelming probability. Moreover, using a random matrix as sensing matrix, ensures the *democratic property* of compressive measurements $\mathbf{y}$ [25]. Indeed, it can be seen that each measurement has roughly the same importance as the others, in the sense that none of them improves or degrades the quality of the reconstruction significantly more than the others. This property is the key ingredient of our proposed CS-GQ technique, as described in Section III.

B. Multiple description coding

MDC [26] is a way of coding an information source that is resilient to packet losses. The multiple description technique allows to create multiple correlated representations of the original information source, each carrying enough information to decode the source with a certain fidelity. Losing a description will not make the received data unusable since each description can be decoded separately, albeit providing a limited quality. However, the best decoding quality is obtained when all the descriptions are available. In the framework of MDC, one can
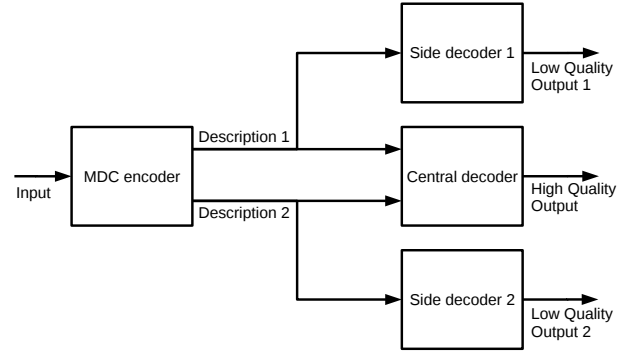

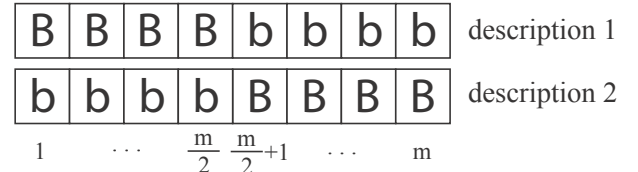Fig. 1: Block diagram of a two-description MDC system.

Fig. 2: CS-GQ

identify the so-called *central* decoder and the *side* decoders. The role of the central decoder is to decode the source when all descriptions have been received, thus achieving the so-called central distortion. The side decoder recovers the source with a lower fidelity (side distortion) since only a subset of the descriptions has been received. This scheme is depicted in Fig. 1 for the simple two-description case. One may seek to create balanced or unbalanced descriptions depending on their individual contribution to the final quality of the recovered source. Several approaches to MDC have been studied in the literature. Among others, it is possible to identify approaches based on transforms, such as the Pairwise Correlating Transform [27], approaches based on channel coding, such as Unequal Error Protection [28] and approaches based on quantization, notably the Multiple Description Scalar Quantizer (MDSQ) [11].

III. GRADED QUANTIZATION FORMULATION

In the simple case of two descriptions, CS-GQ partitions the indices of the measurement vector $\mathbf{y}$ into two sets Ω_1 and Ω_2. Two different quantization step sizes (coarse and fine) are chosen and assigned to each set to generate the first description. The dual assignment is performed for the second description. This is graphically shown in Fig. 2, where we considered $\Omega_1 = \left\{ i \in \left[1, \frac{m}{2}\right] \right\}$ and $\Omega_2 = \left\{ j \in \left[\frac{m}{2} + 1, m\right] \right\}$ but, thanks to the democratic property [25], the same performance is expected for any other definition of the sets with the same cardinality, *e.g.*, the sets of even and odd indices. Before going into further details, we remark that this scheme is amenable to generalizations to an arbitrary number of sets and quantization step sizes and hence of descriptions. In this paper we will not consider this more general case, and we will rather focus on the analysis and usage of the two-descriptions system. The proposed system design may provide either balanced or unbalanced descriptions, depending on how many measurements are
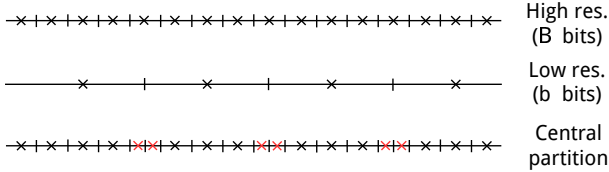
Fig. 3: High and low resolution quantizers are staggered so that the central decoder achieves higher precision.

quantized using the high-rate and the low-rate quantizers. In case of unbalanced descriptions, the cardinality of sets Ω_1 and Ω_2 is not the same. Balanced descriptions offer the same side decoding distortion regardless of which description is received. Otherwise, unbalanced descriptions may have significantly different side distortion depending on the specific description that is received. This is desirable when the descriptions are sent on separate channels having very different packet loss probabilities. In the simplest case, one description is contained in a single packet, *e.g.*, when the number of measurements is small or there is no packet-size limitation. Issues related to packetization strategies and channel losses will be discussed later in Section VI.

From now on, we consider staggered uniform scalar quantizers with 2^B and 2^b levels, and $B \geq b$. The associated quantization step sizes are, respectively, Δ_B and Δ_b, which are linked to the number of levels through the dynamic range r: $\Delta_B = r2^{-B}$ and $\Delta_b = r2^{-b}$. The i-th description, $i = 1, 2$, has then $m_{B,i}$ elements quantized with 2^B levels and $m_{b,i}$ elements quantized with 2^b levels, such that $m_{B,i} + m_{b,i} = m$.

Staggered quantizers are quantizers whose reconstruction levels are shifted with respect to each other. In this work, the staggering involves the low-resolution and the high-resolution quantizers and is motivated by an improvement in the resolution obtained by the central decoder, which receives both the high-resolution and low-resolution versions of each measurement. In fact, when both descriptions are received, each measurement falls inside the intersection of the quantization intervals defined by the two staggered quantizers, which is smaller or equal than Δ_B. Hence, the low-resolution quantizer has its bins shifted by $\frac{\Delta_B}{2}$. The following equations describe how the measurement vector $\mathbf{y}$ is quantized to generate the 2 descriptions:

$$\mathbf{y}^{(B)} = \left\lfloor \frac{\mathbf{y}}{\Delta_B} \right\rfloor \Delta_B + \frac{\Delta_B}{2}$$

$$\mathbf{y}^{(b)} = \left\lfloor \frac{\mathbf{y} - \frac{\Delta_B}{2}}{\Delta_b} \right\rfloor \Delta_b + \frac{\Delta_b}{2} + \frac{\Delta_B}{2} \ .$$

This allows to gain some extra bins of width $\frac{\Delta_B}{2}$ when low-resolution and high-resolution quantizers are combined. This is shown in Fig. 3 for a 4-bit high resolution quantizer and a 2-bit low-resolution quantizer. In general, the high resolution quantizer has 2^B bins, the low resolution quantizer has 2^b bins, while the combined quantizer has $2^B - 2^b + 1$ bins of size Δ_B and $2(2^b - 1)$ bins of size $\frac{\Delta_B}{2}$.

In this work we do not consider more complex quantizers, *e.g.*, the Lloyd-Max method or vector quantization, because

they are regarded as computationally too complex and with little or no performance gain, as shown in [29].

An improvement over classic recovery from noisy measurements (1) can be obtained if we explicitly take into account the two different quantization noise levels. Hence, the side decoder solves the following optimization problem:

$$\hat{\boldsymbol{\theta}}_{\mathsf{S}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \begin{cases} \left\|\mathbf{y}^{(B)} - \Phi^{(B)}\Psi\boldsymbol{\theta}\right\|_2 & \leq \varepsilon_B \\ \left\|\mathbf{y}^{(B)} - \Phi^{(B)}\Psi\boldsymbol{\theta}\right\|_\infty & \leq \frac{\Delta_B}{2} \\ \left\|\mathbf{y}^{(b)} - \Phi^{(b)}\Psi\boldsymbol{\theta}\right\|_2 & \leq \varepsilon_b \\ \left\|\mathbf{y}^{(b)} - \Phi^{(b)}\Psi\boldsymbol{\theta}\right\|_\infty & \leq \frac{\Delta_b}{2} \end{cases} \quad (2)$$

where $\Phi^{(B)}$ and $\Phi^{(b)}$ are the appropriate submatrices of Φ, *i.e.*, Φ restricted to the rows corresponding to the measurements with fine and coarse quantization levels, respectively. The two ℓ_2-norm constraints in (2) take into account the different quantization levels in the two sets, being ε_B and ε_b the expected ℓ_2 norm of quantization noise, which can be estimated as

$$\varepsilon_B = \sqrt{m_{B,i} \frac{\Delta_B^2}{12}} \qquad \text{and} \qquad \varepsilon_b = \sqrt{m_{b,i} \frac{\Delta_b^2}{12}} \ .$$

The ℓ_∞-norm constraints enforce consistency with the quantization intervals - *i.e.*, a reconstruction whose measurements would be quantized to the same intervals as the observed measurements, allowing better dequantization performance [25], [30] -, thus exploiting all the information available regarding the original unknown signal. Note that, while the ℓ_2-norm constraints refer to the *average* norm of the noise, the ℓ_∞-norm constraint is applied individually on every linear measurement. In [8], it is shown that explicitly considering the previously explained structure of the noise in the reconstruction problem provides significant gains in the quality of the reconstruction, with respect to considering an average noise norm.

On the other hand, when both descriptions are received, the central decoder outputs the vector

$$\hat{\boldsymbol{\theta}}_{\mathsf{C}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \ \|\mathbf{y}_{\mathsf{C}} - \Phi\Psi\boldsymbol{\theta}\|_2 \leq \varepsilon_{\mathsf{C}} \ , \qquad (3)$$

where $\mathbf{y}_{\mathsf{C}}$ is obtained in the following manner. For each measurement, the quantization bins of the high- and low-resolution versions are compared to determine the correct bin in the central partition to be used for the dequantization (see Fig. 3). $\varepsilon_{\mathsf{C}}^2$ can be estimated as in (15) to properly take quantizer staggering into account.

We shall also consider a special case of CS-GQ, which we call CS-SPLIT. It is a simple technique that consists in splitting the measurements vector into 2 or more segments, so that instead of transmitting a single packet with all the measurements, two descriptions are created by packetizing half of the measurements in each. Referring to Figure 2, CS-SPLIT can be regarded as a special case of CS-GQ when $b = 0$, providing the best central distortion but worst side distortion as shown in section VI-A.

IV. ADMM-BASED SIDE DECODER

In this section, we propose a method for the resolution of problem (2) based on ADMM. In [8] problem (2) is solved by

CVX [31] [32], a framework to model and solve convex optimization problems, which employs the semidefinite programming (SDP) solver SeDuMi [33]. SeDuMi uses interior point methods to solve general SDP problems, resulting in rather slow performance, especially for large scale problems due to the fact that those are second-order methods[1]. In this paper, we propose a significantly faster reconstruction algorithm based on the Alternating Directions Method of Multipliers (ADMM) tailored to solve problem (2). ADMM [20] is a popular first-order method that allows cheaper iterations, and that combines the robustness of the method of multipliers (a method to solve a constrained optimization problem via unconstrained minimization of an augmented Lagrangian function, including the constraints as penalty terms and an additional ℓ_2 penalty as augmenting term) [35] in terms of convergence, and the decomposition property, similar to dual ascent [36], which also allows for a distributed implementation. It is particularly convenient to solve the side decoding problem (2) because it allows to use alternating updates to the dual variables of an augmented Lagrangian, projecting over the sets defined by the multiple constraints. A proximal gradient method is used to update the primal variable $\boldsymbol{\theta}$ at each iteration. In the following we report the derivation of the algorithm, while the complete procedure is summarized in Algorithm 1.

A. Notation

Let us first introduce some notation that is used during the derivation of the algorithm. First, we define the sets $\mathcal{C}_2$ and $\mathcal{C}_\infty$, which are feasible sets defined by the ℓ_2 and ℓ_∞ constraints of (2), as a function of generic quantities A, $\mathbf{b}$ and c, as

$$\mathcal{C}_2 (A, \mathbf{b}, c) = \{\boldsymbol{\theta} \in \mathbb{R}^n \mid \|A\boldsymbol{\theta} - \mathbf{b}\|_2 \leq c\}$$
$$\mathcal{C}_\infty (A, \mathbf{b}, c) = \{\boldsymbol{\theta} \in \mathbb{R}^n \mid \|A\boldsymbol{\theta} - \mathbf{b}\|_\infty \leq c\} \ ,$$

along with the corresponding indicator functions, namely

$$\mathcal{I}_{\mathcal{C}_2 (A, \mathbf{b}, c)} (\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} \in \mathcal{C}_2 (A, \mathbf{b}, c) \\ \infty & \text{otherwise} \end{cases}$$

$$\mathcal{I}_{\mathcal{C}_\infty (A, \mathbf{b}, c)} (\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} \in \mathcal{C}_\infty (A, \mathbf{b}, c) \\ \infty & \text{otherwise} \end{cases} .$$

The quantities defined above will be used to replace the constrained formulaton of problem (2) with an unconstrained one. Moreover, we define the soft-thresholding operator with threshold λ as

$$\mathcal{S}_\lambda (\mathbf{v}) = \operatorname{sgn}(\mathbf{v}) \odot \max \{\mathbf{0}, |\mathbf{v}| - \lambda\} \ ,$$

where $|\cdot|$ denotes the vector containing the absolute values of $\cdot$ and $\odot$ denotes the elementwise product, and the operators performing projection over ℓ_2 ball

$$\mathcal{P}_{\mathcal{C}_2 (\mathbf{b}, c)} (\mathbf{v}) = \begin{cases} \mathbf{b} + c \frac{\mathbf{v} - \mathbf{b}}{\|\mathbf{v} - \mathbf{b}\|_2} & \text{if } \|\mathbf{v} - \mathbf{b}\|_2 > c \\ \mathbf{v} & \text{otherwise} \end{cases}$$

and ℓ_∞ ball

$$\operatorname{clip}_{[-c, +c]} (\mathbf{v}) = \operatorname{sgn}(\mathbf{v}) \odot \min \{|\mathbf{v}|, c\} \ ,$$

where $\min\{\cdot\}$ and $\max\{\cdot\}$ have to be intended as elementwise operators.

[1]See [34] for details.

B. Unconstrained problem and auxiliary variables

It is possible to use indicator functions to transform (2) into an unconstrained minimization problem. First, we define the following quantities

$$H = \Phi^{(B)} \Psi \quad \text{and} \quad \mathbf{h} = \mathbf{y}^{(B)}$$
$$L = \Phi^{(b)} \Psi \quad \text{and} \quad \mathbf{l} = \mathbf{y}^{(b)} .$$

Then,

$$\begin{aligned} \hat{\boldsymbol{\theta}}_\mathsf{S} = \arg \min_{\boldsymbol{\theta}} \Big\{ & \|\boldsymbol{\theta}\|_1 + \mathcal{I}_{\mathcal{C}_2(\varepsilon_B, H, \mathbf{h})} (\boldsymbol{\theta}) + \mathcal{I}_{\mathcal{C}_2(\varepsilon_b, L, \mathbf{l})} (\boldsymbol{\theta}) \\ & + \mathcal{I}_{\mathcal{C}_\infty(\varepsilon_B, H, \mathbf{h})} (\boldsymbol{\theta}) + \mathcal{I}_{\mathcal{C}_\infty(\varepsilon_b, L, \mathbf{l})} (\boldsymbol{\theta}) \Big\} \\ = \arg \min_{\boldsymbol{\theta}} \Big\{ & \|\boldsymbol{\theta}\|_1 + \mathcal{I}_{\mathcal{C}_2(\varepsilon_B, I, \mathbf{h})} (H\boldsymbol{\theta}) + \mathcal{I}_{\mathcal{C}_2(\varepsilon_b, I, \mathbf{l})} (L\boldsymbol{\theta}) \\ & + \mathcal{I}_{\mathcal{C}_\infty\left(\frac{\Delta_B}{2}, I, \mathbf{h}\right)} (H\boldsymbol{\theta}) + \mathcal{I}_{\mathcal{C}_\infty\left(\frac{\Delta_b}{2}, I, \mathbf{l}\right)} (L\boldsymbol{\theta}) \Big\} \end{aligned}$$

where I is the identity matrix of suitable size. We now introduce auxiliary variables $\mathbf{w}, \mathbf{z}, \mathbf{p}, \mathbf{q}$.

$$\mathbf{w} = H\boldsymbol{\theta}, \quad \mathbf{z} = L\boldsymbol{\theta}, \quad \mathbf{p} = H\boldsymbol{\theta}, \quad \mathbf{q} = L\boldsymbol{\theta}$$

The problem can be now recast as:

$$\begin{aligned} \hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}, \mathbf{w}, \mathbf{z}, \mathbf{p}, \mathbf{q}} \Big\{ & \|\boldsymbol{\theta}\|_1 + \mathcal{I}_{\mathcal{C}_2(\varepsilon_B, I, \mathbf{h})} (\mathbf{w}) + \mathcal{I}_{\mathcal{C}_2(\varepsilon_b, I, \mathbf{l})} (\mathbf{z}) \\ & + \mathcal{I}_{\mathcal{C}_\infty\left(\frac{\Delta_B}{2}, I, \mathbf{h}\right)} (\mathbf{p}) + \mathcal{I}_{\mathcal{C}_\infty\left(\frac{\Delta_b}{2}, I, \mathbf{l}\right)} (\mathbf{q}) \Big\} \\ \text{subject to} \quad & \mathbf{w} = H\boldsymbol{\theta}, \quad \mathbf{z} = L\boldsymbol{\theta}, \quad \mathbf{p} = H\boldsymbol{\theta}, \quad \mathbf{q} = L\boldsymbol{\theta} \end{aligned}$$

We can include these new constraints by building an augmented Lagrangian functional:

$$\begin{aligned} J = & \|\boldsymbol{\theta}\|_1 + \mathcal{I}_{\mathcal{C}_2(\varepsilon_B, I, \mathbf{h})} (\mathbf{w}) + \mathcal{I}_{\mathcal{C}_2(\varepsilon_b, I, \mathbf{l})} (\mathbf{z}) + \mathcal{I}_{\mathcal{C}_\infty\left(\frac{\Delta_B}{2}, I, \mathbf{h}\right)} (\mathbf{p}) \\ & + \mathcal{I}_{\mathcal{C}_\infty\left(\frac{\Delta_b}{2}, I, \mathbf{l}\right)} (\mathbf{q}) + \mathbf{u}^T (\mathbf{w} - H\boldsymbol{\theta}) + \mathbf{v}^T (\mathbf{z} - L\boldsymbol{\theta}) \\ & + \mathbf{s}^T (\mathbf{p} - H\boldsymbol{\theta}) + \mathbf{t}^T (\mathbf{q} - L\boldsymbol{\theta}) + \frac{\rho}{2} \|\mathbf{w} - H\boldsymbol{\theta}\|_2^2 \\ & + \frac{\rho}{2} \|\mathbf{z} - L\boldsymbol{\theta}\|_2^2 + \frac{\rho}{2} \|\mathbf{p} - H\boldsymbol{\theta}\|_2^2 + \frac{\rho}{2} \|\mathbf{q} - L\boldsymbol{\theta}\|_2^2 . \end{aligned}$$

Moreover, we switch to the scaled form by defining the following residuals and scaled dual variables:

$$\begin{array}{llll} \mathbf{r}_u = \mathbf{w} - H\boldsymbol{\theta} & \mathbf{r}_v = \mathbf{z} - L\boldsymbol{\theta} & \mathbf{r}_s = \mathbf{p} - H\boldsymbol{\theta} & \mathbf{r}_t = \mathbf{q} - L\boldsymbol{\theta} \\ \mathbf{u}_\rho = \rho^{-1}\mathbf{u} & \mathbf{v}_\rho = \rho^{-1}\mathbf{v} & \mathbf{s}_\rho = \rho^{-1}\mathbf{s} & \mathbf{t}_\rho = \rho^{-1}\mathbf{t} \end{array}$$

Hence, the augmented Lagrangian can be rewritten as:

$$\begin{aligned} J = & \|\boldsymbol{\theta}\|_1 + \mathcal{I}_{\mathcal{C}_2(\varepsilon_B, I, \mathbf{h})} (\mathbf{w}) + \mathcal{I}_{\mathcal{C}_2(\varepsilon_b, I, \mathbf{l})} (\mathbf{z}) + \mathcal{I}_{\mathcal{C}_\infty\left(\frac{\Delta_B}{2}, I, \mathbf{h}\right)} (\mathbf{p}) \\ & + \mathcal{I}_{\mathcal{C}_\infty\left(\frac{\Delta_b}{2}, I, \mathbf{l}\right)} (\mathbf{q}) + \frac{\rho}{2} \|\mathbf{r}_u + \mathbf{u}_\rho\|_2^2 - \frac{\rho}{2} \|\mathbf{u}_\rho\|_2^2 \\ & + \frac{\rho}{2} \|\mathbf{r}_v + \mathbf{v}_\rho\|_2^2 - \frac{\rho}{2} \|\mathbf{v}_\rho\|_2^2 + \frac{\rho}{2} \|\mathbf{r}_s + \mathbf{s}_\rho\|_2^2 \\ & - \frac{\rho}{2} \|\mathbf{s}_\rho\|_2^2 + \frac{\rho}{2} \|\mathbf{r}_t + \mathbf{t}_\rho\|_2^2 - \frac{\rho}{2} \|\mathbf{t}_\rho\|_2^2 . \end{aligned} \quad (4)$$

Algorithm 1 CS-GQ side decoder.

Require: $\alpha, \lambda, \rho, H, L, \mathbf{h}, \mathbf{l}$

$\boldsymbol{\theta}^{(0,0)} \leftarrow \mathbf{0}, \mathbf{w}^{(0)} = \mathbf{p}^{(0)} \leftarrow \mathbf{0}, \mathbf{z}^{(0)} = \mathbf{q}^{(0)} \leftarrow \mathbf{0}, \mathbf{u}_\rho^{(0)} = \mathbf{s}_\rho^{(0)} \leftarrow \mathbf{0}, \mathbf{v}_\rho^{(0)} = \mathbf{t}_\rho^{(0)} \leftarrow \mathbf{0}$

while Outer stopping criterion is not met **do**

  while Inner stopping criterion is not met **do**

$$\boldsymbol{\theta}^{(i,j+1)} = \mathcal{S}_\lambda \Big(\boldsymbol{\theta}^{(i,j)} - \alpha\rho \Big[H^T \left(H\boldsymbol{\theta}^{(i,j)} - \mathbf{w}^{(\mathbf{i})} - \mathbf{u}_\rho^{(i)} \right) + L^T \left(L\boldsymbol{\theta}^{(i,j)} - \mathbf{z}^{(i)} - \mathbf{v}_\rho^{(i)} \right)$$
$$+ H^T \left(H\boldsymbol{\theta}^{(i,j)} - \mathbf{p}^{(i)} - \mathbf{s}_\rho^{(i)} \right) + L^T \left(L\boldsymbol{\theta}^{(i,j)} - \mathbf{q}^{(i)} - \mathbf{t}_\rho^{(i)} \right) \Big] \Big)$$

    $j \leftarrow j + 1$

  end while

  After stopping: $\boldsymbol{\theta}^{(i+1,0)} = \boldsymbol{\theta}^{(i,j)}$

$$\mathbf{w}^{(i+1)} = \begin{cases} \mathbf{h} + \varepsilon_B \frac{H\boldsymbol{\theta}^{(i+1,0)} - \mathbf{u}_\rho^{(i)} - \mathbf{h}}{\|H\boldsymbol{\theta}^{(i+1,0)} - \mathbf{u}_\rho^{(i)} - \mathbf{h}\|_2} & \text{if } \|H\boldsymbol{\theta}^{(i+1,0)} - \mathbf{u}_\rho^{(i)} - \mathbf{h}\|_2 > \varepsilon_B \\ H\boldsymbol{\theta}^{(i+1,0)} - \mathbf{u}_\rho^{(i)} & \text{otherwise} \end{cases} \quad (5)$$

$$\mathbf{z}^{(i+1)} = \begin{cases} \mathbf{l} + \varepsilon_b \frac{L\boldsymbol{\theta}^{(i+1,0)} - \mathbf{v}_\rho^{(i)} - \mathbf{l}}{\|L\boldsymbol{\theta}^{(i+1,0)} - \mathbf{v}_\rho^{(i)} - \mathbf{l}\|_2} & \text{if } \|L\boldsymbol{\theta}^{(i+1,0)} - \mathbf{v}_\rho^{(i)} - \mathbf{l}\|_2 > \varepsilon_b \\ L\boldsymbol{\theta}^{(i+1,0)} - \mathbf{v}_\rho^{(i)} & \text{otherwise} \end{cases} \quad (6)$$

$$\mathbf{p}^{(i+1)} = \text{clip}_{\left[-\frac{\Delta_B}{2}, \frac{\Delta_B}{2} \right]} \left(H\boldsymbol{\theta}^{(i+1,0)} - \mathbf{s}_\rho^{(i)} - \mathbf{h} \right) + \mathbf{h} \quad (7)$$

$$\mathbf{q}^{(i+1)} = \text{clip}_{\left[-\frac{\Delta_b}{2}, \frac{\Delta_b}{2} \right]} \left(L\boldsymbol{\theta}^{(i+1,0)} - \mathbf{t}_\rho^{(i)} - \mathbf{l} \right) + \mathbf{l} \quad (8)$$

$$\mathbf{u}_\rho^{(i+1)} = \mathbf{u}_\rho^{(i)} + \mathbf{w}^{(i+1)} - H\boldsymbol{\theta}^{(i+1,0)}$$
$$\mathbf{v}_\rho^{(i+1)} = \mathbf{v}_\rho^{(i)} + \mathbf{z}^{(i+1)} - L\boldsymbol{\theta}^{(i+1,0)}$$
$$\mathbf{s}_\rho^{(i+1)} = \mathbf{s}_\rho^{(i)} + \mathbf{p}^{(i+1)} - H\boldsymbol{\theta}^{(i+1,0)}$$
$$\mathbf{t}_\rho^{(i+1)} = \mathbf{t}_\rho^{(i)} + \mathbf{q}^{(i+1)} - L\boldsymbol{\theta}^{(i+1,0)}$$

    $i \leftarrow i + 1$

end while

Ensure: $\hat{\boldsymbol{\theta}}_{\mathsf{S}} \leftarrow \boldsymbol{\theta}^{(i,0)}$

C. Alternating directions minimization

It is now possible to minimize (4) in an iterative fashion using alternating directions, *i.e.*, minimizing over one variable at a time. This constitutes the outer loop of Algorithm 1.

$$\boldsymbol{\theta}^{(i+1)} = \arg\min_{\boldsymbol{\theta}} \Big\{ \|\boldsymbol{\theta}\|_1$$
$$+ \frac{\rho}{2} \|\mathbf{w}^{(i)} - H\boldsymbol{\theta} + \mathbf{u}_\rho^{(i)}\|_2^2 + \frac{\rho}{2} \|\mathbf{z}^{(i)} - L\boldsymbol{\theta} + \mathbf{v}_\rho^{(i)}\|_2^2$$
$$+ \frac{\rho}{2} \|\mathbf{p}^{(i)} - H\boldsymbol{\theta} + \mathbf{s}_\rho^{(i)}\|_2^2 + \frac{\rho}{2} \|\mathbf{q}^{(i)} - L\boldsymbol{\theta} + \mathbf{t}_\rho^{(i)}\|_2^2 \Big\} \quad (9)$$

$$\mathbf{w}^{(i+1)} = \arg\min_{\mathbf{w}} \Big\{ \mathcal{I}_{\mathcal{C}_2(\varepsilon_B, I, \mathbf{h})}(\mathbf{w})$$
$$+ \frac{\rho}{2} \|\mathbf{w} - H\boldsymbol{\theta}^{(i+1)} + \mathbf{u}_\rho^{(i)}\|_2^2 \Big\} \quad (10)$$

$$\mathbf{z}^{(i+1)} = \arg\min_{\mathbf{z}} \Big\{ \mathcal{I}_{\mathcal{C}_2(\varepsilon_b, I, \mathbf{l})}(\mathbf{z})$$
$$+ \frac{\rho}{2} \|\mathbf{z} - L\boldsymbol{\theta}^{(i+1)} + \mathbf{v}_\rho^{(i)}\|_2^2 \Big\} \quad (11)$$

$$\mathbf{p}^{(i+1)} = \arg\min_{\mathbf{p}} \Big\{ \mathcal{I}_{\mathcal{C}_\infty \left(\frac{\Delta_B}{2}, I, \mathbf{h} \right)}(\mathbf{p})$$
$$+ \frac{\rho}{2} \|\mathbf{p} - H\boldsymbol{\theta}^{(i+1)} + \mathbf{s}_\rho^{(i)}\|_2^2 \Big\} \quad (12)$$

$$\mathbf{q}^{(i+1)} = \arg\min_{\mathbf{q}} \Big\{ \mathcal{I}_{\mathcal{C}_\infty \left(\frac{\Delta_b}{2}, I, \mathbf{l} \right)}(\mathbf{q})$$
$$+ \frac{\rho}{2} \|\mathbf{q} - L\boldsymbol{\theta}^{(i+1)} + \mathbf{t}_\rho^{(i)}\|_2^2 \Big\} \quad (13)$$

$$\mathbf{u}_\rho^{(i+1)} = \mathbf{u}_\rho^{(i)} + \mathbf{w}^{(i+1)} - H\boldsymbol{\theta}^{(i+1)}$$
$$\mathbf{v}_\rho^{(i+1)} = \mathbf{v}_\rho^{(i)} + \mathbf{z}^{(i+1)} - L\boldsymbol{\theta}^{(i+1)}$$
$$\mathbf{s}_\rho^{(i+1)} = \mathbf{s}_\rho^{(i)} + \mathbf{p}^{(i+1)} - H\boldsymbol{\theta}^{(i+1)}$$
$$\mathbf{t}_\rho^{(i+1)} = \mathbf{t}_\rho^{(i)} + \mathbf{q}^{(i+1)} - L\boldsymbol{\theta}^{(i+1)}$$

It can be seen that (9) involves minimizing a functional composed of two main parts: a smooth (differentiable) part $f(\boldsymbol{\theta}) = \frac{\rho}{2} \|\mathbf{w}^{(i)} - H\boldsymbol{\theta} + \mathbf{u}_\rho^{(i)}\|_2^2 + \frac{\rho}{2} \|\mathbf{z}^{(i)} - L\boldsymbol{\theta} + \mathbf{v}_\rho^{(i)}\|_2^2 + \frac{\rho}{2} \|\mathbf{p}^{(i)} - H\boldsymbol{\theta} + \mathbf{s}_\rho^{(i)}\|_2^2 + \frac{\rho}{2} \|\mathbf{q}^{(i)} - L\boldsymbol{\theta} + \mathbf{t}_\rho^{(i)}\|_2^2$ whose gradient can be computed analytically and a non-smooth (non-differentiable) part $g(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$. Functionals of this kind can be minimized using the proximal gradient method [37], *i.e.*, an iterative method that takes a step towards the negative gradient of the smooth component and then computes the proximal mapping over the non-smooth part. It is easy to show that the gradient

of the smooth part $f(\boldsymbol{\theta})$ is:

$$
\begin{aligned}
\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \\
= \rho H^T \left(H\boldsymbol{\theta}^{(i)} - \mathbf{w}^{(i)} - \mathbf{u}_\rho^{(i)} \right) + \rho L^T \left(L\boldsymbol{\theta}^{(i)} - \mathbf{z}^{(i)} - \mathbf{v}_\rho^{(i)} \right) \\
+ \rho H^T \left(H\boldsymbol{\theta}^{(i)} - \mathbf{p}^{(i)} - \mathbf{s}_\rho^{(i)} \right) + \rho L^T \left(L\boldsymbol{\theta}^{(i)} - \mathbf{q}^{(i)} - \mathbf{t}_\rho^{(i)} \right) ,
\end{aligned}
$$

while the proximity operator of non-smooth part $g(\boldsymbol{\theta})$ is the soft-thresholding function. Hence, one can iteratively take a step in the gradient direction and apply the soft thresholding function in order to eventually find $\boldsymbol{\theta}^{(i+1)}$, *i.e.*, the minimizer of (9). This constitutes the inner loop of Algorithm 1.

The subsequent minimization problems in (10), (11), (12) and (13) all involve projections over the sets defined using the indicator functions. It can be seen that (10) and (11) correspond to projections over the ℓ_2 balls of radii ε_B and ε_b centered at $\mathbf{h}$ and $\mathbf{l}$ respectively. Similarly, (12) and (13) correspond to projections over the ℓ_∞ balls of diameters Δ_B and Δ_b centered at $\mathbf{h}$ and $\mathbf{l}$ respectively. The results of the projections are (5), (6), (7) and (8). Notice that the projection operations and the update of the dual variables can be performed in parallel.

As stopping criteria of the inner and outer loops, we check the distance between two successive iterations. Hence, two constants $\varepsilon_i^{\mathsf{stop}}$ and $\varepsilon_j^{\mathsf{stop}}$ are properly chosen such that the stopping conditions are

$$
\frac{\|\boldsymbol{\theta}^{(i,j+1)} - \boldsymbol{\theta}^{(i,j)}\|_2}{\|\boldsymbol{\theta}^{(i,j)}\|_2} < \varepsilon_j^{\mathsf{stop}} \qquad \text{(Inner stopping criterion)}
$$

$$
\frac{\|\boldsymbol{\theta}^{(i+1,0)} - \boldsymbol{\theta}^{(i,0)}\|_2}{\|\boldsymbol{\theta}^{(i,0)}\|_2} < \varepsilon_i^{\mathsf{stop}}. \qquad \text{(Outer stopping criterion)}
$$

V. Theoretical Guarantees

In this section we present some theoretical results concerning CS-GQ. First, Theorem 1 provides a guarantee of stable recovery of signals from single descriptions, using the side decoder presented in (2). As it is common in the literature about CS (see for example [23]), this kind of bounds is used as a guarantee that the error does not explode, rather than giving an exact characterization of the error itself.

Theorem 1. (Stable side recovery) *Suppose that $A = \Phi\Psi$ satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2} - 1$ and let $\mathbf{y} = A\boldsymbol{\theta} + \mathbf{e}$ where, without loss of generality, $\mathbf{e} = \begin{bmatrix} \mathbf{e}_B^T & \mathbf{e}_b^T \end{bmatrix}^T$ with $\|\mathbf{e}_B\|_2 \leq \varepsilon_B$, $\|\mathbf{e}_b\|_2 \leq \varepsilon_b$, $\|\mathbf{e}_B\|_\infty \leq \frac{\Delta_B}{2}$, $\|\mathbf{e}_b\|_\infty \leq \frac{\Delta_b}{2}$. Then the solution to (2) obeys*

$$
\begin{aligned}
\|\hat{\boldsymbol{\theta}}_{\mathsf{S}} - \boldsymbol{\theta}\|_2 \leq C_0 \frac{\sigma_k(\boldsymbol{\theta})_1}{\sqrt{k}} \\
+ C_2 \left(\min\left\{ \varepsilon_B, \frac{\Delta_B}{2}\sqrt{m_B} \right\} + \min\left\{ \varepsilon_b, \frac{\Delta_b}{2}\sqrt{m_b} \right\} \right)
\end{aligned}
$$

where $\sigma_k(\boldsymbol{\theta})_1$ is the l_1-norm of the error incurred by approximating $\boldsymbol{\theta}$ with its k largest-magnitude components, m_B and m_b are the number of rows of matrices $\Phi^{(B)}\Psi$ and $\Phi^{(b)}\Psi$ respectively and the constants are

$$
C_0 = 2\frac{1 - \left(1 - \sqrt{2}\delta_{2k}\right)}{1 - \left(1 + \sqrt{2}\delta_{2k}\right)} \qquad C_2 = 4\frac{\sqrt{1 + \delta_{2k}}}{1 - \left(1 + \sqrt{2}\delta_{2k}\right)}.
$$

Proof. We are interested in deriving a bound for $\|\mathbf{d}\|_2 = \|\hat{\boldsymbol{\theta}}_{\mathsf{S}} - \boldsymbol{\theta}\|_2$. We know that $\boldsymbol{\theta} \in \{\mathcal{C}_2(\varepsilon_B, H, \mathbf{h}) \cap \mathcal{C}_2(\varepsilon_b, L, \mathbf{l}) \cap \mathcal{C}_\infty(\frac{\Delta_B}{2}, H, \mathbf{h}) \cap \mathcal{C}_\infty(\frac{\Delta_b}{2}, L, \mathbf{l})\}$ (see Sec. IV for notation). Moreover, the solution to (2) is either $\boldsymbol{\theta}$ or one with lower l_1 norm, so that we can say that $\|\hat{\boldsymbol{\theta}}_{\mathsf{S}}\|_1 \leq \|\boldsymbol{\theta}\|_1$. Using [23, Lemma 1.6] we know that

$$
\|\mathbf{d}\|_2 \leq C_0 \frac{\sigma_k(\boldsymbol{\theta})_1}{\sqrt{k}} + C_1 \frac{|\langle A\mathbf{d}_\Lambda, A\mathbf{d}\rangle|}{\|\mathbf{d}_\Lambda\|_2}
$$

where $C_1 = \frac{2}{1 - (1+\sqrt{2})\delta_{2k}}$, $\Lambda = \Lambda_0 \cup \Lambda_1$, being Λ_0 the set of the k components with largest magnitude of $\boldsymbol{\theta}$ and Λ_1 the set of the k components with largest magnitude of $\mathbf{d}_{\Lambda_0^c}$ (subscript denotes restriction to the components indexed by the subscript set). Notice that

$$
\begin{aligned}
\|A\mathbf{d}\|_2 &= \|A\hat{\boldsymbol{\theta}}_{\mathsf{S}} - A\boldsymbol{\theta}\|_2 = \|A\hat{\boldsymbol{\theta}}_{\mathsf{S}} - \mathbf{y} + \mathbf{y} - A\boldsymbol{\theta}\|_2 \\
&= \left\| \begin{bmatrix} \Phi^{(B)}\Psi\hat{\boldsymbol{\theta}}_{\mathsf{S}} \\ \Phi^{(b)}\Psi\hat{\boldsymbol{\theta}}_{\mathsf{S}} \end{bmatrix} - \begin{bmatrix} \mathbf{y}^{(B)} \\ \mathbf{y}^{(b)} \end{bmatrix} \right. \\
&\quad \left. + \begin{bmatrix} \mathbf{y}^{(B)} \\ \mathbf{y}^{(b)} \end{bmatrix} - \begin{bmatrix} \Phi^{(B)}\Psi\boldsymbol{\theta} \\ \Phi^{(b)}\Psi\boldsymbol{\theta} \end{bmatrix} \right\|_2 \\
&\leq \left\| \Phi^{(B)}\Psi\hat{\boldsymbol{\theta}}_{\mathsf{S}} - \mathbf{y}^{(B)} \right\|_2 + \left\| \Phi^{(b)}\Psi\hat{\boldsymbol{\theta}}_{\mathsf{S}} - \mathbf{y}^{(b)} \right\|_2 \\
&\quad + \left\| \Phi^{(B)}\Psi\boldsymbol{\theta} - \mathbf{y}^{(B)} \right\|_2 + \left\| \Phi^{(b)}\Psi\boldsymbol{\theta} - \mathbf{y}^{(b)} \right\|_2 \\
&\leq 2\min\left\{ \varepsilon_B, \frac{\Delta_B}{2}\sqrt{m_B} \right\} + 2\min\left\{ \varepsilon_b, \frac{\Delta_b}{2}\sqrt{m_b} \right\}
\end{aligned}
$$

Using the Cauchy-Schwarz inequality we write

$$
\begin{aligned}
&|\langle A\mathbf{d}_\Lambda, A\mathbf{d}\rangle| \leq \|A\mathbf{d}_\Lambda\|_2 \|A\mathbf{d}\|_2 \\
&\leq 2\|\mathbf{d}_\Lambda\|_2 \sqrt{1 + \delta_{2k}} \left(\min\left\{ \varepsilon_B, \frac{\Delta_B}{2}\sqrt{m_B} \right\} + \min\left\{ \varepsilon_b, \frac{\Delta_b}{2}\sqrt{m_b} \right\} \right)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\|\mathbf{d}\|_2 \leq C_0 \frac{\sigma_k(\boldsymbol{\theta})_1}{\sqrt{k}} \\
+ C_2 \left(\min\left\{ \varepsilon_B, \frac{\Delta_B}{2}\sqrt{m_B} \right\} + \min\left\{ \varepsilon_b, \frac{\Delta_b}{2}\sqrt{m_b} \right\} \right)
\end{aligned}
$$

$\square$

Next, we characterize the performance of an oracle side decoder for sparse signals so that the result can be used to provide an oracle optimality condition for the optimization of the CS-GQ parameters, as shown in Sec. VI. The oracle decoder is an ideal decoder that knows perfectly the support $\mathcal{S}$ of the sparse signal. For such systems, recovery amounts to computing $\hat{\boldsymbol{\theta}} = A_{\mathcal{S}}^\dagger \mathbf{y}$, where $A_{\mathcal{S}}^\dagger$ is the Moore-Penrose pseudoinverse of matrix A restricted to the columns indexed by $\mathcal{S}$. Often, the oracle receiver is used to evaluate the exact performance of CS reconstruction algorithms, since CS decoding involves a nonlinear reconstruction step, whose distortion performance is hard to characterize exactly.

Theorem 2. (Oracle recovery) *Suppose that $\boldsymbol{\theta}$ is k-sparse and that Φ has i.i.d. Gaussian zero-mean entries having variance $\frac{1}{m}$. Let $\mathbf{y} = \Phi\Psi\boldsymbol{\theta}$ be the vector of measurements to be quantized using staggered low- and high-resolution quantizers.*

Suppose the oracle decoder is used for CS recovery. Then, the expected side distortion in the high-rate regime is:

$$\mathbb{E}\left[\left\|\hat{\boldsymbol{\theta}}_{\mathsf{S}} - \boldsymbol{\theta}\right\|_2^2\right] = \frac{kr^2}{(m-k-1)}\frac{m}{24}\left(2^{-2B} + 2^{-2b}\right)$$

and the expected central distortion is:

$$\mathbb{E}\left[\left\|\hat{\boldsymbol{\theta}}_{\mathsf{C}} - \boldsymbol{\theta}\right\|_2^2\right] = \frac{kr^2}{(m-k-1)}\frac{m}{24}2^{-2B}\frac{2^{B+1} - 2^b + 1}{2^B + 2^b - 1}$$

Proof. According to [38, Theorem 5], the following relation holds:

$$\mathbb{E}\left[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2\right] = \frac{k}{m-k-1}\mathbb{E}\left[\mathbf{e}^T\mathbf{e}\right] \tag{14}$$

where $\mathbf{e} = \mathcal{Q}(\mathbf{y}) - \mathbf{y}$ is the error introduced by quantization $\mathcal{Q}(\cdot)$ In case of side decoding, the receiver has only a single description i with $m_{B,i}$ high resolution measurements and $m_{b,i}$ low resolution measurements. The average error norm is obtained as the mean of the norm of the expected error when description 1 is received ($\mathbf{e}_1$) and when description 2 is received ($\mathbf{e}_2$). Notice that $m_{B,1} = m_{b,2}$ and $m_{b,1} = m_{B,2}$ by design. Hence, in the high-rate regime the following relation holds:

$$\begin{aligned}\mathbb{E}\left[\left\|\hat{\boldsymbol{\theta}}_{\mathsf{S}} - \boldsymbol{\theta}\right\|_2^2\right] &= \frac{k}{m-k-1} \cdot \frac{1}{2}\left(\mathbb{E}\left[\mathbf{e}_1^T\mathbf{e}_1\right] + \mathbb{E}\left[\mathbf{e}_2^T\mathbf{e}_2\right]\right) \\ &= \frac{kr^2}{m-k-1}\frac{1}{24}\Big(m_{B,1}2^{-2B} + m_{b,1}2^{-2b} \\ &\quad + m_{B,2}2^{-2B} + m_{b,2}2^{-2b}\Big) \\ &= \frac{kr^2}{(m-k-1)}\frac{m}{24}\left(2^{-2B} + 2^{-2b}\right)\end{aligned}$$

The central decoder exploits the staggering of the low and high resolution quantizers to obtain a non-uniform central quantizer, having $2^B - 2^b + 1$ bins of size Δ_B and $2(2^b - 1)$ bins of size $\frac{\Delta_B}{2}$ as explained in Sec. III. Thus the expected error norm is:

$$\begin{aligned}\mathbb{E}\left[\mathbf{e}^T\mathbf{e}\right] &= \\ &= \frac{m}{2^B + 2^b - 1}\left[2\left(2^b - 1\right)\frac{\Delta_B^2}{48} + \left(2^B - 2^b + 1\right)\frac{\Delta_B^2}{12}\right] \\ &= \frac{mr^2}{24}2^{-2B}\frac{2^{B+1} - 2^b + 1}{2^B + 2^b - 1}\end{aligned} \tag{15}$$

Substituting it into (14) we obtain the desired expression. $\square$

VI. RATE-DISTORTION OPTIMIZATION AND PACKETIZATION

A. Optimizing CS-GQ

In this section, we focus our analysis on the scenario when one employs identical channel models for both descriptions, *e.g.*, the loss probability p is the same for both descriptions. Under this scenario, balanced descriptions are optimal both instance-wise and on average, *i.e.* the distortion incurred by either side decoder is the same. On the other hand, unbalanced descriptions provide the same average performance, but the distortion incurred by a specific instance depends on which description is received and can be either lower or higher than the distortion in the balanced case. Hereafter, we thus only consider the balanced case. Different values of the parameters

B and b provide different tradeoffs between the distortion at the central decoder and the distortion at the side decoder, for a fixed total rate for both descriptions: $R = B + b$. Moreover, the packet loss rate on the communication links will also affect the expected distortion. On the other hand, the limit case $b = 0$, which we called CS-SPLIT, simply splits the measurements into two sets without inserting any redundancy. This is equivalent to the segmentation performed by any network protocol when the packet size exceeds the maximum size. It is clear that this is the optimal strategy when there are no packet losses, because it is equivalent to generating a single description and no bits are wasted in redundancy. However, if one description fails to reach the receiver, only half of the measurements will be available for decoding, hence significantly degrading the quality or leading to recovery failure when their number is too low. Fig. 4 shows the various operating points enabled by CS-GQ on the side distortion vs. central distortion plane.

It is of interest to study the optimal value of the parameters B and b when the description loss probability is known. In this scenario, one wishes to find the quantization step sizes providing the lowest expected distortion for a given bit budget. Hence we define the average distortion as

$$D = p^2 + 2p\left(1 - p\right) \cdot D_s + (1 - p)^2 \cdot D_c$$

where p is the probability of losing a description, and D_s and D_c are the distortions incurred by the side and central decoder, respectively defined as

$$D_s = \frac{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathsf{S}}\|_2}{\|\boldsymbol{\theta}\|_2} \text{ and } D_c = \frac{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathsf{C}}\|_2}{\|\boldsymbol{\theta}\|_2} .$$

The main problem in the optimization of the parameters is the lack of closed-form expressions for the distortion of the central and side decoders. There are two possible ways of solving this problem. The first way is to resort to an operational curve, in which the operating points on the (D_s, D_c) plane are known in advance (*e.g.*, as a result of experiments).

In this case, for a fixed total rate R, every choice of the low-resolution rate b generates a $(D_s(b), D_c(b))$ point, for a total of $\lfloor \frac{R}{2} \rfloor + 1$ points. The optimal choice of b is therefore:

$$\hat{b} = \underset{b \in \left[0, \frac{R}{2}\right]}{\arg\min}\left[p^2 + 2p \cdot (1 - p)D_s(b) + (1 - p)^2 \cdot D_c(b)\right]$$

(Operational Optimality)

The other possible method is to employ some bounds on the reconstruction performance. We can use the performance of the oracle decoder, as described in Theorem 2, to derive a simple expression that does not require any operational information and can be used to optimize the parameters *a priori*. Hence, the minimization of the average distortion
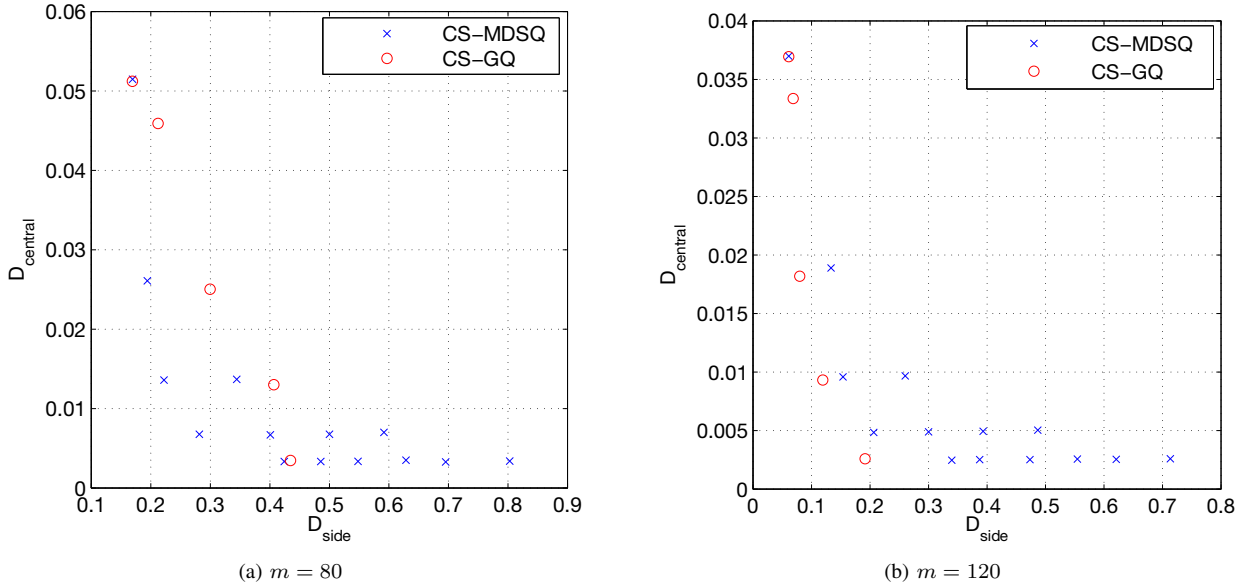
(a) $m = 80$

(b) $m = 120$

Fig. 4: Operating points for CS-GQ (different values of B and b) and for CS-MDSQ (different values of M, refer to [19]). $n = 256$, $k = 10$.

becomes:

$$
\begin{aligned}
\hat{b} &= \operatorname*{arg\,min}_{b \in \left[0, \frac{R}{2}\right]} \left[p^2 + 2p\left(1-p\right) \cdot D_s + \left(1-p\right)^2 \cdot D_c \right] \\
&= \operatorname*{arg\,min}_{b \in \left[0, \frac{R}{2}\right]} \left[2p\left(1-p\right) \frac{kr^2}{(m-k-1)} \frac{m}{24} \left(2^{-2B} + 2^{-2b} \right) \right. \\
&\qquad \left. + \left(1-p\right)^2 \frac{kr^2}{(m-k-1)} \frac{m}{24} 2^{-2B} \frac{2^{B+1} - 2^b + 1}{2^B + 2^b - 1} \right] \\
&= \operatorname*{arg\,min}_{b \in \left[0, \frac{R}{2}\right]} \left[2p\left(2^{-2B} + 2^{-2b}\right) \right. \\
&\qquad \left. + \left(1-p\right)2^{-2B} \frac{2^{B+1} - 2^b + 1}{2^B + 2^b - 1} \right] \\
&= \operatorname*{arg\,min}_{b \in \left[0, \frac{R}{2}\right]} \left[2p\left(2^{-2(R-b)} + 2^{-2b}\right) \right. \\
&\qquad \left. + \left(1-p\right)2^{-2(R-b)} \frac{2^{R-b+1} - 2^b + 1}{2^{R-b} + 2^b - 1} \right]
\end{aligned}
$$

(Oracle Optimality)

Exhaustive search over the feasible integer values of b can be employed to determine the optimal one. We remark that this search just involves evaluating the distortion expression for $\lfloor \frac{R}{2} \rfloor + 1$ values of b (*e.g.*, 5 values when $R = 8$), thus having very low complexity. Unfortunately, the optimality would only hold if we had an oracle decoder, so this choice might be suboptimal. Section VII-A shows a comparison between the average distortion obtained with operational optimization and with oracle optimization in order to show the effectiveness of the latter.
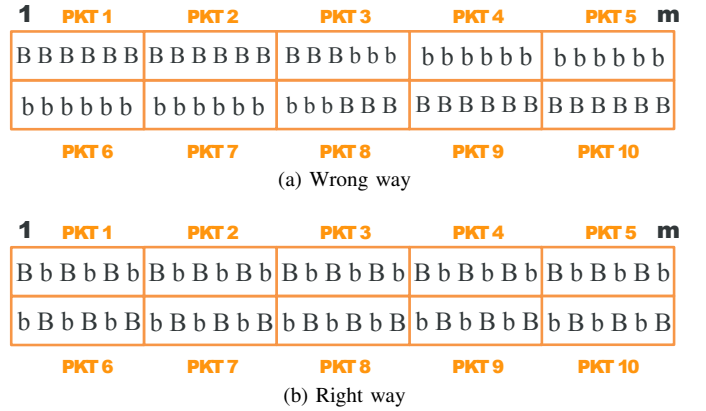

(a) Wrong way

(b) Right way

Fig. 5: Segmenting a long measurements vector.

B. Packetization for high number of measurements

It may happen that the MDC system needs to deal with a large scale problem in which the number of measurements to be acquired is fairly large. In this scenario, the transmitter cannot create packets that are arbitrarly large as their size is capped by the MTU (Maximum Transmission Unit) size of the adopted communication protocol. As an example, the IEEE 802.15.4 protocol, popular in sensor network applications, specifies an MTU equal to 127 bytes with an effective maximum payload size of 104 bytes.

As far as CS-GQ is concerned, a careful segmentation operation must be performed whenever the size of a description exceeds the MTU. The goal is to keep the packets balanced, in order to always get useful information whenever a packet is received. Fig. 5b graphically shows a way to assign quantization step sizes and performing segmentation. Every packet contains an alternating sequence of measurements quantized with B bits and b bits and has a dual packet

with the opposite alternation pattern. Therefore, every received packet is informative, in the sense that it is able to improve the current set of measurements, whether by adding new measurements or by improving their accuracy. To fix ideas, Fig. 5a shows a wrong way to perform segmentation, where the packets containing only low-resolution measurements are not informative if the high resolution dual set is already available. In case of loss of a packet, one could potentially lose more high resolution measurements with respect to the case depicted in Fig. 5b.

VII. NUMERICAL EXPERIMENTS AND APPLICATIONS

Numerical simulations have been performed to evaluate the performance of the proposed technique against other methods to implement MDC for CS and to show the advantages for some practical usage scenarios, inspired by sensor network applications. All the results requiring side decoding have been obtained using the side decoder presented in Sec. IV. We remark that the algorithm solves problem (2) and that the solution returned by the proposed algorithm is the same as the one returned by a solver of convex problems such as CVX. However, the proposed algorithm is specialized for the particular problem considered in this paper and can be significantly faster than CVX. The SeDuMi solver employed by CVX uses second-order methods, such as Newton's method, and the complexity per iteration grows as $\mathcal{O}(n^6)$. ADMM on the other hand is a first order method, thus much less expensive per iteration than a second order method, although in principle having slower convergence. Nevertheless, ADMM enjoys linear convergence rates under mild conditions on the cost function, as shown in [39]. As a practical example, solving a problem with $n = 1024$, $k = 150$ non-zero components, $m = 600$ measurements obtained by a Gaussian sensing matrix, $B = 6$ bits, $b = 4$ bits requires 104 seconds using CVX, while only 9.1 seconds are needed by the proposed algorithm based on ADMM, thus achieving a tenfold speed-up. Moreover, whenever quantization cells are uniform, we add to the reconstruction problem an ℓ_∞ constraint in addition to the ℓ_2 constraint presented in (1). This is to enforce consistent reconstruction as discussed in Sec. III.

A. Effectiveness of optimization via oracle formula

We present an experiment that shows the average distortion obtained via optimization of the parameters with the operational method and with the oracle method discussed in Sec.VI-A. The curves shown in Fig. 6 are obtained for the same system parameters of Fig. 4, *i.e.*, $m = 120$ measurements for a $n = 256$-long, $k = 10$-sparse signal and a memoryless channel with packet loss probability p. It is observed that the optimization via the oracle method provides a good estimate of the parameters, only yielding slightly higher distortion when the packet loss probability is very small. This is due to the ideality of oracle decoding, which estimates lower distortions than actually achieved by practical algorithms such as l_1 minimization.
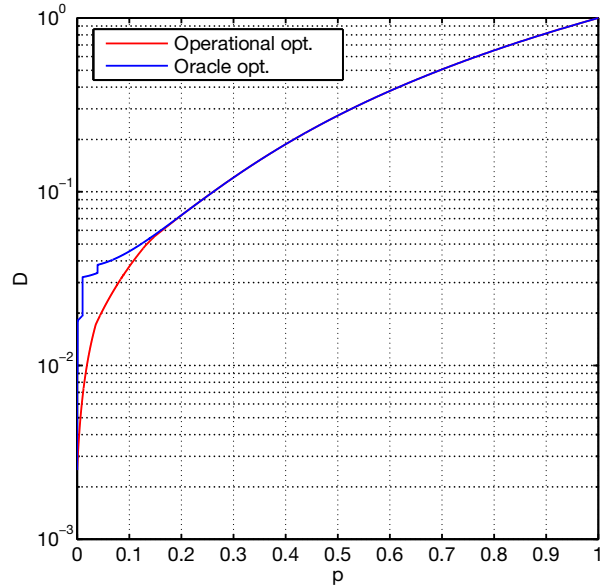

Fig. 6: Average distortion over a memoryless lossy channel with packet loss probability p. Parameters b and B are optimized via oracle/operational formulas. $n = 256$, $m = 120$, $k = 10$.

### B. Comparison with other quantization-based MDC schemes

In this section we compare the performance of CS-GQ against other methods generating multiple descriptions using quantization. Thus, we compare CS-GQ, CS-SPLIT and CS-MDSQ, where CS-MDSQ is an application of the multiple description scalar quantizer (MDSQ) as introduced in [11] to the quantization of compressive measurements. Although MDSQ is a powerful and popular solution for MDC, and provides asymptotic performance close to the rate-distortion bound, its use in CS is not straightforward. In fact, CS reconstruction is a nonlinear process, with no guarantee that lower distortion on the measurements results in better reconstruction quality. Hence, the choice of the number of measurements and quantization step size is a trade-off in order to achieve the best performance after recovery. In our simulations, we use the index assignment technique developed in [19]. As a first comparison we fix the number of measurements and obtain the tradeoff points on the side/central distortion plane for both CS-GQ and CS-MDSQ. In Fig. 4 it is observed that fixing $m = 120$ for a $n = 256$-long, $k = 10$-sparse signal allows CS-GQ to provide better tradeoff points than CS-MDSQ. However, CS-MDSQ outperfroms CS-GQ when a lower number of measurements is chosen. We remark that, theoretically, CS-MDSQ allows 2^{R+1} tradeoff points by modulating the redundancy, while CS-GQ admits $\lfloor \frac{R}{2} \rfloor + 1$ points only. However, observing Fig. 4 we can notice that the redundancy of many tradeoff points of CS-MDSQ is so high that the resulting side distortion is severe (*e.g.*, above 0.4). We remark that the MDSQ can also accomodate the zero-redundancy case when all the entries of the index assignment matrix are used, corresponding to CS-SPLIT. Both zero-redundancy CS-MDSQ and CS-SPLIT obtain the same central distortion. However, side distortion may

be different. CS-SPLIT represents half of the measurements with a fine quantization rate, while CS-MDSQ represents all the measurements with a coarser quantization rate. The effect on CS recovery is that when there are plenty of measurements (depending on n and on the sparsity of the signal), the side distortion after recovery is quantization-limited, so even though CS-MDSQ provides more measurements, their coarser quantization limits the performance. We also notice that the full-redundancy case $(B = b)$ corresponds to a CS-MDSQ with staggered side quantizers and equal size intervals in the central partition.

A further experiment evaluates the dependency of the side and central distortion on the number of measurements. Hence, we fix the number of measurements and an operating point and we analyse how central and side distortion change as a function of the number of measurements. Fig. 7 is obtained with the same signal as in Fig. 4 and choosing $m = 120$, $(B, b) = (6, 2)$ as the operating point of CS-GQ and $M = 2$ as the redundancy of the MDSQ with $R = 8$ total rate (namely, the ratio between the step size of the side quantizer and of the central quantizer is $2M$; refer to [19] for further details). We observe that the distortion of CS-GQ decreases faster than the distortion of CS-MDSQ as the number of measurement increases, while central distortion decreases at the same rate although CS-GQ is marginally better. The reason behind this performance is that CS-MDSQ is more quantization-limited than CS-GQ when the number of measurements is high. Viceversa, for a very low number of measurements, CS-GQ is measurement-limited. The previous experiment considered a rather generous overall budget equal to $Rm = 960$ bits. Indeed, we considered a fixed number of measurements and chose the rate according to the budget. It must be noticed that one could optimize both the value of m and the value of R under the overall budget constraint and this could lead to different choices for CS-GQ and CS-MDSQ. Although we do not report the results for brevity, we observed that if the budget is very large as in the previous case, CS-MDSQ generally provides better tradeoff curves than CS-GQ by using fewer measurements and a finer quantization rate. However, results change under a tight budget constraint. A minimum number of measurements has to be acquired in order to ensure successful reconstruction and due to the tight budget constraint it is not possible to use high quantization rates. Fig.8 compares the tradeoff curves of CS-GQ and CS-MDSQ for the two operating points $m = 50$ and $R = 8$ and for $m = 100$, $R = 4$. Those two choices of m and R are optimal for CS-MDSQ and CS-GQ respectively under a budget of $Rm = 400$ bits. We notice that CS-GQ outperforms CS-MDSQ despite choosing the best combination of m and R. We therefore conclude that CS-GQ can be advantageous with respect to CS-MDSQ when the bit budget is low.

C. Simulations over MTU-limited memoryless and Gilbert channels

In this section we perform some simulations to assess the gain achieved by using CS-GQ when the communication channel is prone to packet loss. We consider two channel models,
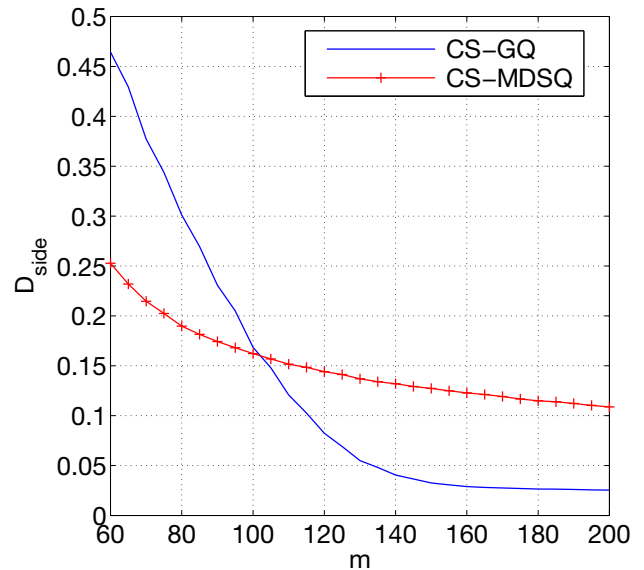
(a) Side Distortion
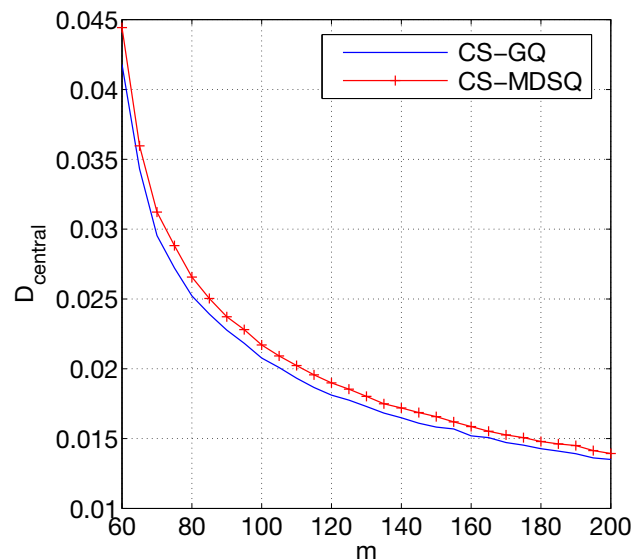
(b) Central Distortion

Fig. 7: Comparison of CS-GQ and CS-MDSQ. $(B, b) = (6, 2)$ for CS-GQ and $M = 2$, $R = 8$ for CS-MDSQ. $n = 256$, $k = 10$.

which are significant for the performance assessment of CS-GQ in a packetization scenario: the memoryless channel, in which the loss trace is a Bernoulli process, and the Gilbert channel [40], [41], where memory is modelled as a two-state Markov chain, as shown in Fig. 9. In this model the communication link can be in either of two states, that we label Good (G) and Bad (B) with a probability p of moving from G to B and probability q of moving from B to G. When in B state the link will drop any packet transmitted. This is a popular model for channels exhibiting burst errors, *e.g.*, due to fading in a wireless system. In both cases we assume that a limit to the maximum packet size is present and it is equal to 100 bytes, as in the MTU of IEEE 802.15.4.
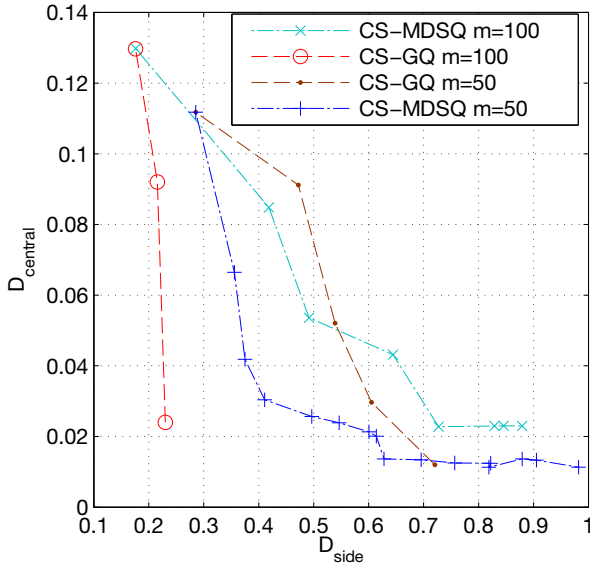
Fig. 8: Comparison under a tight budget $Rm = 400$ bits. Optimal choice for CS-MDSQ is $m = 50, R = 8$. Optimal choice for CS-GQ is $m = 100, R = 4$.
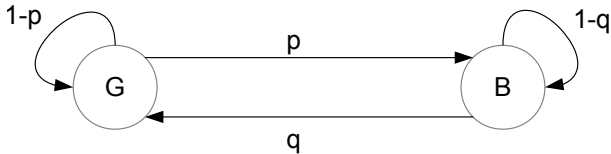

Fig. 9: The Gilbert-Elliot channel model. p is the transition probability from good to bad state.

The simulations over the memoryless channel have been performed with a signal of length $n = 256$, $k = 40$ non-zero components, $m = 160$ measurements and a total rate $R = 10$ bits per measurement over the two descriptions. CS-GQ has been implemented to create two packets of 100 bytes each, in the same fashion of Fig. 5b, and compared against a standard segmentation of the measurements, *i.e.*, splitting half of the measurements into the two packets of 100 bytes. The tests measured the distortion in the reconstructed signal averaged over 10^5 runs. The values of B and b are automatically optimized using the oracle method (see Sec. VI-A). Fig. 10 shows that CS-GQ has superior performance, yielding a lower average distortion than segmentation of the measurements and the gain is more significant when the packet loss probability is high.

The simulations over the Gilbert channel have been performed using a longer signal, in order to correctly test the effect of the memory of the channel. We suppose that a batch of $N_s = 1000$ vectors of measurements have to be transmitted in sequence. The length of the signals is $n = 1000$, $k = 200$ non-zero components, $m = 720$ measurements and a total rate $R = 10$ bits per measurement. Table I reports the average distortion for some values of the transition probabilities p and q. It can be noticed that the proposed scheme allows to achieve lower distortion with respect to the segmentation approach.
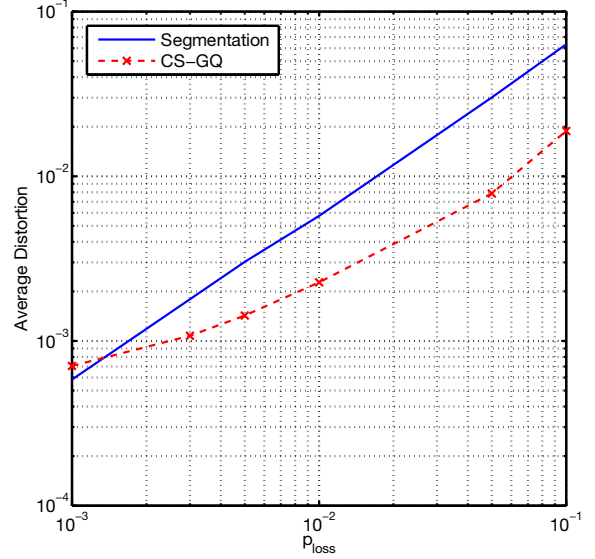

Fig. 10: Simulations over memoryless channel.

TABLE I: Average distortion over the Gilbert-Elliot channel

p	q	Segmentation	CS-GQ
0.05	0.5	0.0257	**0.0192**
0.05	0.3	0.0762	**0.0496**
0.05	0.15	0.1820	**0.1205**
0.01	0.3	0.0160	**0.0102**
0.01	0.15	0.0451	**0.0324**

D. Simulations of object recognition over lossy channel

In this section we consider an application recently proposed in [42] and [43] as a possible scenario in which the techniques proposed in this paper could provide a significant performance improvement. The purpose is to show the effectiveness of CS-GQ in a practical scenario. In this application, a smart camera platform, such as CITRIC [44], computes histograms of image descriptors to be used as features in a scene classification task. Image descriptors are compact representations of an image that provide a certain degree of invariance to transformations such as rotation, scaling, etc., and are widely used in the computer vision field for visual search or scene recognition tasks. The most popular descriptors are the Scale-Invariant Feature Transform (SIFT) [45], which describe each image keypoint, *i.e.*, a signficant point of the image to be included in the descriptor, through a vector of $s = 128$ entries. Each image i has a variable number N_i of keypoints, and thus associated descriptors. A clustering algorithm, such as κ-means, is used to identify κ clusters in $\mathbb{R}^s$ from the N_i original descriptors. A histogram of descriptors is obtained by counting how many keypoint descriptors fall in each cluster. In [42], the authors leverage the sparsity of such histograms to compress them by means of random projections, and transmit them to a remote fusion center. The authors do not consider the problems of quantization of the measurements, nor the possibility of having channel losses. However, we remark that those are key problems for this application. In fact, transmitting the random projections as floating point values requires very large bandwidth, hence quantization could reduce such requirement

TABLE II: Accuracy of scene recognition

p	Segmentation	CS-GQ
0.01	70.63%	**71.25%**
0.05	61.25%	**63.75%**
0.10	54.38%	**61.88%**

and save a sizeable amount of the scarce energy of the mobile platform. Also, channel losses may occur and retransmission of packets may not be feasible due to real-time constraints, and it would indeed require further energy consumption on sensor side. Hence disabling retransmission and providing a scheme that is robust to channel losses could be of interest for such platforms.

In our tests, we build a $10,000$-dimensional dictionary by running hierarchical κ-means, with 10 clusters per level in the hierarchy, over SIFT descriptors extracted from some images in the BMW database [42]. This database comprises photos of the UC Berkeley campus acquired with the CITRIC platform with multiple views of the same scene from different angles. We then use the dictionary of quantized SIFT features to produce histograms of descriptors. Each image is characterized by a histogram of descriptors, which is sparse because the number of descriptors extracted from each image is much less than the number of entries in the dictionary. [42]. A support vector machine is trained from histograms of descriptors of the images in the training set, in order to solve a classification task that amounts to recognizing a scene from a number of possible scenes. In order to simulate a real system, we consider packets with maximum size of 100 bytes (as in the case of the MTU of IEEE 802.15.4), to be transmitted over a memoryless channel. The test metric that we adopt in this case is the accuracy in the scene recognition. Table II shows the accuracy in the scene recognition task as function of the packet loss probability p. We can see that CS-GQ builds robustness into the system and allows to achieve higher recognition rates, when in presence of packet losses.

VIII. CONCLUSIONS

In this paper we considered the problem of robust transmission of CS measurements over unreliable channels. We discussed some strategies based on the Graded Quantization paradigm that enable to increase robustness in a way that directly exploits the democratic property of compressive measurements. Moreover, contrary to many cases in the literature, we considered a realistic packetization scenario and discussed the issues involved in it as well as the performance of the proposed method. Finally, we proposed a novel reconstruction algorithm, based on ADMM, for the specific problem of having measurements obtained from multiple quantizers. Our simulations showed that the proposed technique is a competitive method to implement MDC for CS applications. In particular, we showed that we can improve scene recognition accuracy in an application of a smart camera platform using CS and transmitting over an unreliable channel.

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, april 2006.

[2] T. Petrisor, B. Pesquet-Popescu, and J. Pesquet, "A compressed sensing approach to frame-based multiple description coding," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, vol. 2, 2007, pp. II–709–II–712.

[3] D. Liu, D. Gao, and G. Shi, "Compressive sensing-based multiple description image coding for wireless channel," in *2010 International Conference on Wireless Communications and Signal Processing (WCSP)*, oct. 2010, pp. 1 –5.

[4] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, March 2008.

[5] W. L. Chan, M. L. Moravec, R. G. Baraniuk, and D. M. Mittleman, "Terahertz imaging with compressed sensing and phase retrieval," *Optics letters*, vol. 33, no. 9, pp. 974–976, 2008.

[6] T. Bjorklund and E. Magli, "A parallel compressive imaging architecture for one-shot acquisition," in *Picture Coding Symposium (PCS), 2013*, Dec 2013, pp. 65–68.

[7] C. Deng, W. Lin, B.-S. Lee, and C. Lau, "Robust image coding based upon compressive sensing," *IEEE Transactions on Multimedia*, vol. 14, no. 2, pp. 278–290, 2012.

[8] D. Valsesia, G. Coluccia, and E. Magli, "Graded quantization: Democracy for multiple descriptions in compressed sensing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, May 2013, pp. 5825–5829.

[9] K. P. Subbalakshmi and S. Somasundaram, "Multiple description image coding framework for ebcot," in *International Conference on Image Processing*, vol. 3. IEEE, 2002, pp. 541–544.

[10] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in *Electronic Imaging'99*. International Society for Optics and Photonics, 1998, pp. 998–1008.

[11] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[12] U. Samarawickrama, J. Liang, and C. Tian, "-channel multiple description coding with two-rate coding and staggered quantization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 7, pp. 933–944, 2010.

[13] T. Tillo, E. Baccaglini, and G. Olmo, "Multiple descriptions based on multirate coding for jpeg 2000 and h. 264/avc," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1756–1767, 2010.

[14] S. Dumitrescu, G. Rivers, and S. Shirani, "Unequal erasure protection technique for scalable multistreams," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 422–434, 2010.

[15] E. Akyol, A. M. Tekalp, and M. R. Civanlar, "A flexible multiple description coding framework for adaptive peer-to-peer video streaming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 2, pp. 231–245, 2007.

[16] E. Baccaglini, T. Tillo, and G. Olmo, "A flexible RD-based multiple description scheme for JPEG 2000," *IEEE Signal Processing Letters*, vol. 14, no. 3, pp. 197–200, 2007.

[17] T. Tillo, M. Grangetto, and G. Olmo, "Multiple description image coding based on lagrangian rate allocation," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 673–683, 2007.

[18] I. V. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1211–1225, 2003.

[19] G. Zhang, J. Klejsa, and W. B. Kleijn, "Optimal index assignment for multiple description scalar quantization with translated lattice codebooks," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4444–4451, 2012.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[21] E. Candès and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[22] J. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, March 2006.

[23] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed sensing: theory and applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012, pp. 1–64.

[24] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[25] J. Laska, P. Boufounos, M. Davenport, and R. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 429 – 443, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520311000248

[26] V. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, September 2001.

[27] Y. Wang, M. Orchard, V. Vaishampayan, and A. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Transactions on Image Processing*, vol. 10, no. 3, pp. 351 –366, mar 2001.

[28] A. Mohr, E. Riskin, and R. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 819 –828, june 2000.

[29] W. Dai, H. Pham, and O. Milenkovic, "Information theoretical and algorithmic approaches to quantized compressive sensing," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 1857–1866, July 2011.

[30] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in ir n: analysis, synthesis, and algorithms," *Information Theory, IEEE Transactions on*, vol. 44, no. 1, pp. 16–31, 1998.

[31] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sep. 2012.

[32] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.

[33] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[34] R. D. Monteiro, "First-and second-order methods for semidefinite programming," *Mathematical Programming*, vol. 97, no. 1–2, pp. 209–244, 2003.

[35] M. R. Hestenes, "Multiplier and gradient methods," *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303–320, 1969.

[36] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[37] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, ser. Springer Optimization and Its Applications, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Springer New York, 2011, pp. 185–212.

[38] G. Coluccia, A. Roumy, and E. Magli, "Operational rate-distortion performance of single-source and distributed compressed sensing," *IEEE Transactions on Communications*, vol. 62, no. 6, pp. 2022–2033, June 2014.

[39] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *arXiv preprint arXiv:1208.3922*, 2012.

[40] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, no. 9, pp. 1253–1265, 1960.

[41] E. O. Elliott, "Estimates of Error Rates for Codes on Burst-Noise Channels," *Bell System Technical Journal*, vol. 42, pp. 1977–1997, Sep. 1963.

[42] N. Naikal, A. Yang, and S. Sastry, "Towards an efficient distributed object recognition system in wireless smart camera networks," in *13th Conference on Information Fusion (FUSION)*, 2010, pp. 1–8.

[43] A. Yang, S. Maji, C. Christoudias, T. Darrell, J. Malik, and S. Sastry, "Multiple-view object recognition in band-limited distributed camera networks," in *Third ACM/IEEE International Conference on Distributed Smart Cameras. ICDSC 2009.*, 2009, pp. 1–8.

[44] P. Chen, P. Ahammad, C. Boyer, S.-I. Huang, L. Lin, E. Lobaton, M. Meingast, S. Oh, S. Wang, P. Yan, A. Yang, C. Yeo, L.-C. Chang, J. D. Tygar, and S. Sastry, "CITRIC: A low-bandwidth wireless camera network platform," in *Second ACM/IEEE International Conference on Distributed Smart Cameras. ICDSC 2008.*, 2008, pp. 1–10.

[45] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.