



**AALBORG UNIVERSITY**  
DENMARK

**Aalborg Universitet**

## **Multiple-Description I1-Compression**

Jensen, Tobias Lindstrøm; Østergaard, Jan; Dahl, Joachim; Jensen, Søren Holdt

*Published in:*  
I E E E Transactions on Signal Processing

*DOI (link to publication from Publisher):*  
[10.1109/TSP.2011.2145373](https://doi.org/10.1109/TSP.2011.2145373)

*Publication date:*  
2011

*Document Version*  
Early version, also known as pre-print

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Jensen, T. L., Østergaard, J., Dahl, J., & Jensen, S. H. (2011). Multiple-Description I1-Compression. *I E E E Transactions on Signal Processing*, 59(8), 3699 - 3711 . <https://doi.org/10.1109/TSP.2011.2145373>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### **Take down policy**

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# Multiple-Description $l_1$ -Compression

Tobias Lindstrøm Jensen, Jan Østergaard, *Member, IEEE*, Joachim Dahl, Søren Holdt Jensen, *Senior Member, IEEE*

**Abstract**—Multiple descriptions (MDs) is a method to obtain reliable signal transmissions on erasure channels. An MD encoder forms several descriptions of the signal and each description is independently transmitted across an erasure channel. The reconstruction quality then depends on the set of received descriptions. In this paper, we consider the design of redundant descriptions in an MD setup using  $l_1$ -minimization with Euclidean distortion constraints. In this way we are able to obtain sparse descriptions using convex optimization. The proposed method allows for an arbitrary number of descriptions and supports both symmetric and asymmetric distortion design. We show that MDs with partial overlapping information corresponds to enforcing coupled constraints in the proposed convex optimization problem. To handle the coupled constraints, we apply dual decompositions which makes first-order methods applicable and thereby admit solutions for large-scale problems, *e.g.*, coding entire images or image sequences. We show by examples that the proposed framework generates non-trivial sparse descriptions and non-trivial refinements. We finally show that the sparse descriptions can be quantized and encoded using off-the-shell encoders such as the set partitioning in hierarchical trees (SPIHT) encoder, however, the proposed method shows a rate-distortion loss compared to state-of-the-art image MD encoders.

**Index Terms**—Multiple Descriptions, Sparse Decompositions, First-order Methods, Convex Optimization

## I. INTRODUCTION

**A**N important problem in signal processing is the *multiple-description* (MD) problem [1]. The MD problem is on encoding a source into multiple descriptions, which are transmitted over separate channels. The channels may occasionally break down causing description erasures, in which case only a subset of the descriptions are received. Which of the channels that are working at any given time is known by the decoder but not by the encoder. The problem is then to construct a number of descriptions, which individually provide an acceptable quality and furthermore are able to refine each other. It is important to notice the contradicting requirements associated with the MD problem; in order for the descriptions to be individually good, they must all be similar to the source and therefore, to some extent, the descriptions are also similar to each other. However, if the descriptions are the same, they cannot refine each other.

Let  $J$  be the number of channels and let  $\mathcal{J}_J = \{1, \dots, J\}$ . Then  $\mathcal{I}_J = \{\ell \mid \ell \subseteq \mathcal{J}_J, \ell \neq \emptyset\}$  describes the indices of the non-trivial subsets which can be received. Further, let  $z_j$  denote the  $j$ th description and define  $z_\ell = \{z_j \mid j \in \ell\}$ ,  $\forall \ell \in \mathcal{I}_J$ . At the decoder, the descriptions  $z_\ell$ ,  $\ell \in \mathcal{I}_J$ , approximate the source  $y$  via their individual reconstructions  $g_\ell(z_\ell)$  which satisfy the fidelity constraint  $d(g_\ell(z_\ell), y) \leq \delta_\ell$ ,  $\forall \ell \in \mathcal{I}_J$ , with  $d(\cdot, \cdot)$  denoting a distortion measure. An example with  $J = 2$  is illustrated in Fig. 1.

The traditional MD coding problem aims at characterizing the set of achievable tuples  $(R(z_1), R(z_2), \dots, R(z_J), \delta_{\{1\}}, \dots, \delta_{\{1,2,\dots,J\}})$  where  $R(z_j)$  denotes the minimum coding rate for description  $z_j$ ,  $\forall j \in \mathcal{J}_J$ ,

The work is part of the project CSI: Computational Science in Imaging, supported by grant no. 274-07-0065 from the Danish Research Council for Technology and Production Sciences. The work of Jan Østergaard is supported by the Danish Research Council for Technology and Production Sciences, grant no. 274-07-0383.

T. L. Jensen, J. Østergaard and S. H. Jensen are with Aalborg University, Department of Electronic Systems, Aalborg, Denmark. Emails: T. L. Jensen (tlj@es.aau.dk), J. Østergaard (jo@es.aau.dk), and S. H. Jensen (shj@es.aau.dk). J. Dahl is with Mosek Aps, Copenhagen, Denmark. Email: J. Dahl (dahl.joachim@gmail.com).

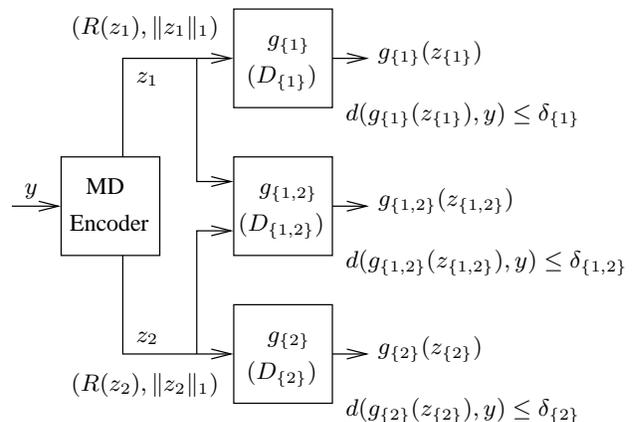


Fig. 1. The MD ( $l_1$ -compression) problem for  $J = 2$ .

required in order to approximate the source  $y$  to within the distortion fidelities  $\delta_\ell$ ,  $\forall \ell \in \mathcal{I}_J$  [1]. The problem is then to construct  $z_\ell$ ,  $\forall \ell \in \mathcal{I}_J$ , so that  $R(z_j)$ ,  $\forall j \in \mathcal{J}_J$ , are minimized and the fidelity constraints are satisfied, *cf.*, Fig. 1. This well-known information theoretic problem remains largely unsolved. In fact, it is only completely solved for the case of two descriptions, with the squared error fidelity criterion and Gaussian sources [2].

Another direction is to form descriptions in a deterministic setting, as opposed to the traditionally MD approach [1]. Algorithms designed for video and image coding may be based on, *e.g.*, Wiener filters with prediction compensation [3], matching pursuit [4], [5] or compressed sensing [6], [7]. Recovery of the latent variables can in compressed sensing be obtained by sparsity driven methods such as  $l_1$ -minimization with known guarantees [8]. There is also results in the case of quantization [9]–[11].

In this paper we propose a convex problem, which can be used to obtain sparse descriptions for MD problems using  $l_1$ -minimization with Euclidean constraints on the distortion of the reconstruction. The proposed MD formulation is flexible in terms of applications (*e.g.*, speech, image and video compression), the number of channels  $J$  as well as supporting both symmetric and asymmetric design. We show how to apply a first-order method to solve the proposed convex optimization problem using dual decomposition and smoothing [12]. Let  $\epsilon$  be the desired accuracy of an approximate solution in function value, in which case the first-order method has iteration complexity  $\mathcal{O}(\frac{1}{\epsilon})$ . The combination of a reasonable iteration complexity and the low complexity of a single iteration in first-order methods makes it possible to apply the proposed MD method to large scale problems such as for entire images or image sequences. The descriptions are for example represented in discrete wavelet dictionaries but arbitrary dictionaries are allowed in the original formulation. For encoding the sparse descriptions it is possible to apply state-of-the-art methods for wavelet encoding, *e.g.*, set partitioning in hierarchical trees (SPIHT) [13]. However, we are not able to obtain state-of-the-art rate-distortion descriptions by the two stage approach of first forming sparse descriptions and then encode.

The organization of the paper is as follows: we will first propose the MD  $l_1$ -compression (MD1C) problem in Sec. II and then analyze

and discuss important properties of the problem in Sec. III. Then, in Sec. IV, we discuss algorithms for solving the proposed convex problem, and present an efficient first-order method. We analyze the sparse descriptions in Sec. V and extend the framework to encoding of MD wavelet coefficient based on well known methods and provide simulations on compression of images and image sequences in Sec. VI.

## II. PROBLEM FORMULATION

An interesting direction of research is in sparse estimation techniques for signal processing based on  $l_1$ -norm heuristics, where, *e.g.*, compressive sampling [8], [14] have gained much attention. The theory is by now well-established and much is known about cases where the  $l_1$ -minimization approach coincides with the solution to the otherwise intractable minimum cardinality solution, see [15] and references therein.

One way of obtaining a sparse approximation  $z$  of the source  $y$  is to solve the so-called  $l_1$ -compression problem

$$\begin{aligned} & \text{minimize} && \|Wz\|_1 \\ & \text{subject to} && \|Dz - y\|_2 \leq \delta, \end{aligned} \quad (1)$$

where  $\delta > 0$  is a given distortion bound,  $D \in \mathbb{R}^{M \times N}$  is an overcomplete dictionary ( $N \geq M$ ),  $z \in \mathbb{R}^N$  is the variable, and  $y \in \mathbb{R}^M$  is the signal we wish to decompose into a sparse representation. In a standard formulation  $W \in \mathbb{R}^{N \times N}$  can be selected as  $W = I$ . To improve the  $l_1$ -minimization approach for minimizing the cardinality it has been proposed to select  $W = \text{diag}(w)$  to reduce the cost of large coefficients [16], see also [17]. To find  $w$ , the unscaled problem (1) is solved first (*i.e.*, with  $w = \mathbb{1}$ ). Then  $w$  is chosen inversely proportional to the solution  $z^*$  of that problem, and (1) is solved again with the new weighting  $w$ . This reweighting scheme can be iterated a number of times.

In this work, we cast the MD problem into the framework of  $l_1$ -compression (1).<sup>1</sup> Let  $z_j \in \mathbb{R}^{\tilde{N}_j \times 1}$ ,  $\forall j \in \mathcal{J}_J$ , be the descriptions of length  $\tilde{N}_j$ . We will define a concatenation operator

$$X = \underset{i \in S}{\mathbf{C}} Y_i = \begin{bmatrix} Y_{S_1} \\ Y_{S_2} \\ \vdots \\ Y_{S_n} \end{bmatrix}$$

where  $Y_i \in \mathbb{R}^{p_i \times q}$ ,  $S = \{S_1, S_2, \dots, S_n\}$  has  $n$  elements and  $X \in \mathbb{R}^{\sum_{i \in S} p_i \times q}$ . Then  $z_\ell = \mathbf{C}_{j \in \ell} z_j \in \mathbb{R}^{\sum_{j \in \ell} \tilde{N}_j \times 1}$ ,  $\forall \ell \in \mathcal{I}_J$ , is the vector concatenation of the descriptions used in the decoding when the subset  $\ell \subseteq \mathcal{J}_J$  is received. For simplicity we will use  $z_j$  with the meaning  $z_{\{j\}}$ ,  $j \in \mathcal{J}_J$ , which also applies to other symbols with subscripted  $\ell$ . The matrix  $D_\ell \in \mathbb{R}^{M \times \sum_{j \in \ell} \tilde{N}_j}$ ,  $\forall \ell \in \mathcal{I}_J$ , is the dictionary associated with the description  $z_\ell$  given as  $D_\ell = \left( \mathbf{C}_{j \in \ell} \bar{D}_{\ell,j}^T \right)^T$  with  $\bar{D}_{\ell,j} \in \mathbb{R}^{M \times \tilde{N}_j}$ . Our idea is to form the *multiple-description  $l_1$ -compression* problem using linear reconstruction functions, *i.e.*,  $g_\ell(z_\ell) = D_\ell z_\ell$  similar to [6] since it preserves convexity [20], and the Euclidean norm as the distortion measure, *i.e.*,  $d(x, y) = \|x - y\|_2$ .<sup>2</sup> Note that its possible to select other distortion measures that is convex, but we choose  $\|\cdot\|_2$  since it relates to the well known peak signal-to-noise-ratio (PSNR). The definition is given below.

<sup>1</sup>This work was presented in part [18], [19].

<sup>2</sup>Interestingly, in the Gaussian case and for the mean squared error fidelity criterion, it has been shown that linear reconstruction functions are sufficient for achieving the MD rate-distortion function, see [21], [22] and [23] for white and colored cases, respectively.

**Definition II.1.** An instance  $\{y, \{\delta_\ell\}_{\ell \in \mathcal{I}_J}, \{D_\ell\}_{\ell \in \mathcal{I}_J}, \{W_j\}_{j \in \mathcal{J}_J}, \{\lambda_j\}_{j \in \mathcal{J}_J}\}$  of the MDIIC problem is defined by

$$\begin{aligned} & \text{minimize} && \sum_{j \in \mathcal{J}_J} \lambda_j \|W_j z_j\|_1 \\ & \text{subject to} && \|D_\ell z_\ell - y\|_2 \leq \delta_\ell, \quad \forall \ell \in \mathcal{I}_J, \end{aligned} \quad (2)$$

for  $\delta_\ell > 0$ ,  $\forall \ell \in \mathcal{I}_J$ ,  $\lambda_j > 0$ ,  $\forall j \in \mathcal{J}_J$  and  $W_j \succ 0$ ,  $\forall j \in \mathcal{J}_J$ . For simplicity we sometime use  $f(z) = \sum_{j \in \mathcal{J}_J} \lambda_j \|W_j z_j\|_1$  for the primal objective and  $Q_P = \{z \mid \|D_\ell z_\ell - y\|_2 \leq \delta_\ell, \forall \ell \in \mathcal{I}_J\}$  for the primal feasible set.

The problem (2) amounts to minimize the number of non-zero coefficients in the descriptions (using convex relaxation) under the constraint that any combination of received descriptions allows a reconstruction error smaller than some quantity. The idea is that the problem (2) can be used to obtain sparse coefficients which obeys certain bounds on the reconstruction error. Since it has been shown that bit rate and sparsity is almost linearly dependent [24], the problem formulation (2) can be used to form descriptions in a MD framework. In Sec. VI we will discuss in detail how to encode the sparse coefficients. Note that since  $|\mathcal{I}_J| = 2^J - 1$ , the number of possible received combinations grows exponential in the number of channels, and thereby the number of constraints in problem (2).

In Definition II.1 we have introduced  $\lambda > 0$  to allow weighting of the  $l_1$ -norms in order to achieve a desired ratio  $\frac{\|W_j z_j\|_1}{\|W_{j'} z_{j'}\|_1}$ ,  $\forall j, j' \in \mathcal{J}_J$ . Note that in the case where we let  $\bar{D}_{j,j}$ ,  $\forall j \in \mathcal{J}_J$ , be orthogonal, we see that the constraints on the side reconstructions can easily be fulfilled by simply truncating the smallest coefficients  $z_j = \bar{D}_{j,j} z_j$ ,  $\forall j \in \mathcal{J}_J$ , to zero separately for the coefficients of each side description. This will, however, not guarantee the joint reconstruction constraint  $\|D_\ell z_\ell - y\|_2 \leq \delta_\ell$ ,  $\forall \ell \in \mathcal{I}_J \setminus \mathcal{J}_J$ . Thus, the problem at hand is non-trivial.

In the following sections we will analyse the MDIIC problem presented in Definition II.1 and give an algorithm to solve large scale instances of this problem.

## III. ANALYSIS OF THE MULTIPLE-DESCRIPTION $l_1$ -COMPRESSION PROBLEM

In this section we will review and discuss some important properties of the proposed MDIIC problem.

**Definition III.1.** (Solvable) The MDIIC problem is solvable if the problem has at least one feasible point.

Remark (Definition III.1) Since the MDIIC problem is always bounded below, this is the same definition as in [20].

**Proposition III.2.** (Solvable conditions) Let  $\bar{D}_{\ell,j} = \rho_{\ell,j} \bar{D}_{j,j}$ ,  $\forall \ell \in \mathcal{I}_J$ ,  $j \in \ell$  with  $\rho_{\ell,j} \in \mathbb{R}$ ,  $\sum_{j \in \ell} \rho_{\ell,j} = 1$ ,  $\forall \ell \in \mathcal{I}_J$ . Furthermore, let  $y \in \text{span}(\bar{D}_{j,j})$ ,  $\forall j \in \mathcal{J}_J$ . Then the MDIIC problem (2) is solvable.

*Proof:* There exists  $z_j$ ,  $\forall j \in \mathcal{J}_J$ , such that  $\bar{D}_{j,j} z_j = y$ ,  $\forall j \in \mathcal{J}_J$ . Then we also have that  $D_\ell z_\ell = \sum_{j \in \ell} \bar{D}_{\ell,j} z_j = \sum_{j \in \ell} \rho_{\ell,j} \bar{D}_{j,j} z_j = y \sum_{j \in \ell} \rho_{\ell,j} = y$ ,  $\forall \ell \in \mathcal{I}_J$ . Hence,  $z = \mathbf{C}_{j \in \mathcal{J}_J} z_j \in Q_P$  is a primal feasible solution and the problem (2) is therefore solvable. ■

One way to obtain the setup used in Proposition III.2 is to use a standard MDIIC setup.

**Definition III.3.** (Standard MD  $l_1$ -compression problem) We denote an MDIIC problem a standard MDIIC problem if

- $\bar{D}_{j,j}$ ,  $\forall j \in \mathcal{J}_J$ , are invertible.
- $\rho_{\ell,j} = \begin{cases} 1 & \text{if } |\ell| = 1 \\ \frac{\sum_{i \in \ell \setminus j} \delta_i^2}{(|\ell| - 1) \sum_{i \in \ell} \delta_i^2}, & \text{otherwise} \end{cases}$ , to weight the contributions

- $\bar{D}_{\ell,j} = \rho_{\ell,j} D_j$ , (combined with above  $\bar{D}_{j,j} = D_j, \forall j \in \mathcal{J}_J$ ).

Remark (Definition III.3) In the general asymmetric case, it's common to weight the reconstruction of the joint reconstructions relative to the distortion of the individual reconstruction [25], [26]. Note that in the symmetric case,  $\delta_\ell = \delta_{\ell'}, \forall \ell, \ell' \in \mathcal{I}_J, |\ell| = |\ell'|$ , we have equal weight  $\rho_{\ell,j} = \rho_{\ell',j}, j, i \in \ell$ .

**Proposition III.4.** (Strong duality) *Strong duality holds for the standard MDIIC problem.*

*Proof:* Since  $\delta_\ell > 0, \forall \ell \in \mathcal{I}_J, \text{span}(D_j) \in y, \forall j \in \mathcal{J}_J$ , and  $\sum_{j \in \ell} \rho_{\ell,j} = 1, \forall \ell \in \mathcal{I}_J$ , there exists a strictly feasible  $\|D_\ell z_\ell - y\|_2 = 0 < \delta_\ell, \forall \ell \in \mathcal{I}_J$ , point  $z$  such that Slater's condition for strong duality holds [20]. ■

In Proposition III.2 we assumed that  $\bar{D}_{\ell,j} = \rho_{\ell,j} D_j, \forall \ell \in \mathcal{I}_J, j \in \ell$ . We will, however, shortly discuss the case where  $\bar{D}_{\ell,j} \neq \rho_{\ell,j} D_j$  for at least one pair  $(\ell, j) \in \mathcal{I}_J \times \mathcal{J}_J$ . The interpretation is that the dictionaries associated with the same description may not be the same in all reconstruction functions. This can be illustrated with an example where we will solve a small problem of size  $M = 2, J = 2$ , with  $D_j, \forall j \in \mathcal{J}_J$ , being the orthogonal discrete cosine transform. We will construct another dictionary in the central reconstruction using a rotation matrix

$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

such that we solve problems on the form

$$\begin{aligned} & \text{minimize} && \|z_1\|_1 + \|z_2\|_1 \\ & \text{subject to} && \|D_1 z_1 - y\|_2 \leq \delta_1 \\ & && \|D_2 z_2 - y\|_2 \leq \delta_2 \\ & && \|\frac{1}{2}(D_1 R_\theta z_1 + D_2 z_2) - y\|_2 \leq \delta_{\{1,2\}} \end{aligned} \quad (3)$$

with solution  $\tilde{z}_\theta^*$ . By considering different  $\theta$ 's we obtain different central decoding functions. In Fig. 2 we show the optimal objective  $f(\tilde{z}_\theta^*)$  from solving problem (3). We investigate  $\theta \in [-\pi/2; \pi/2]$  and only report  $f(\cdot)$  and the cardinality  $\text{card}(\cdot)$  if the problem (3) is solvable. We choose  $\delta_1 = \delta_2 = \{0.2, 0.02\}$  and  $\delta_{\{1,2\}} = 0.01$ . Observe that for both  $\delta_1 = \delta_2 = 0.2$  and  $\delta_1 = \delta_2 = 0.02$ , the objective  $f(\cdot)$  can be reduced if we select  $\theta \neq 0$ , i.e., if the dictionaries associated to the different decoding functions are not equal. For  $\delta_1 = \delta_2 = 0.2$  the cardinality can also be reduced from 3 at  $\theta = 0$  to cardinality 2 at  $\theta \approx -\pi/8$ . If  $\|D_1 z_1 - y\|_2$  is required to be small, we would expect  $|\theta|$  to be small because  $D_1 z_1 \approx y$  and then  $D_1 R_\theta z_1 \approx y$  for  $\theta \approx 0$ . Note that if  $|\theta|$  is too large, then the problem is not solvable.

This example illustrates that it can be useful to have different dictionaries in the decoder associated to the same description. To find such dictionaries a-priori for different applications is signal dependent, and a separate research topic, which will not be treated in this work.

#### IV. SOLVING THE MD $l_1$ -COMPRESSION PROBLEM

The MDIIC problem (2) can be solved using general-purpose primal-dual interior point methods. To do so, we need to solve several linear systems of equations of size  $\mathcal{O}(K) \times \mathcal{O}(K)$ , arising from linearizing first-order optimality conditions, with  $K = M|\mathcal{I}_J| + \sum_{j \in \mathcal{J}_J} N_j$ . This practically limits the size of the problems we can consider to small and medium size, except if the problem has a certain structure that can be used when solving the linear system of equations [27]. Another approach is to use first-order methods [12], [28]–[30]. Such first-order projection methods have shown to be efficient for large scale problems [31]–[34]. However, it is difficult to solve the MDIIC problem efficiently because the feasible set

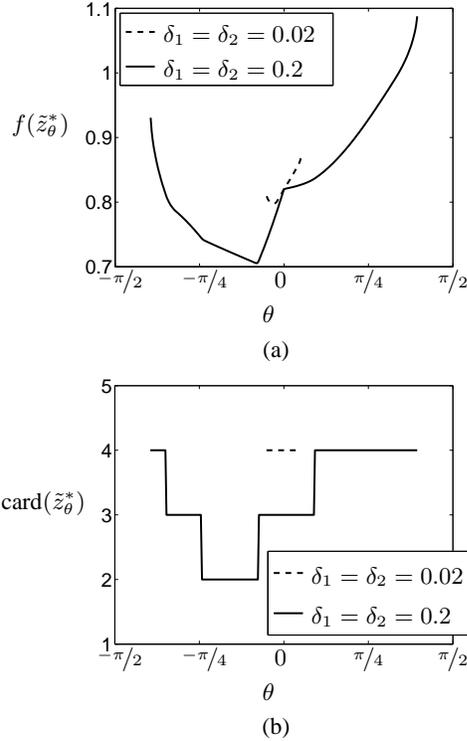


Fig. 2. (a): optimal objective  $f(\tilde{z}_\theta^*)$  and (b): the cardinality  $\text{card}(\tilde{z}_\theta^*)$  from solving the problem (3) with  $M = 2$ . The distortion bounds are  $\delta_1 = \delta_2 = \{0.2, 0.02\}$ ,  $\delta_{\{1,2\}} = 0.01$  and  $D_1 = D_2$ : the discrete cosine transform. We only report  $f(\cdot)$  and  $\text{card}(\cdot)$  if the problem (3) is solvable.

$\|D_\ell z_\ell - y\|_2 \leq \delta_\ell, \forall \ell \in \mathcal{I}_J$ , is an intersection of Euclidean norm balls.

We are interested in solving large-scale instances of problem (2), and in the following subsections IV-A through IV-E, we will present an efficient first-order method to handle problem (2).

#### A. Intersecting Euclidean Norm Balls

In order to illustrate the implications of the overlapping constraints on the feasible set, consider the following simple example. Let  $D_1 = D_2 = W_1 = W_2 = \lambda_1 = \lambda_2 = 1$  so that  $D_1 z_1 = z_1$  and  $D_2 z_2 = z_2$ . From the joint constraint it may be noticed that  $z_1$  and  $z_2$  can be picked arbitrarily large but of different signs and yet satisfy  $|\frac{1}{2}(z_1 + z_2) - y| \leq \delta_{\{1,2\}}$ . However, due to the individual constraints  $|z_1 - y| \leq \delta_1$  and  $|z_2 - y| \leq \delta_2$ , the feasible set is bounded as illustrated in Fig. 3.

#### B. Dual Decomposition

An approach to handle problems with intersecting constraints, sometimes referred to as complicating or coupling constraints, is by dual decomposition [29], [35].

**Proposition IV.1.** (Dual problem) *A dual problem of the standard MDIIC problem can be represented as*

$$\begin{aligned} & \text{maximize} && - \sum_{\ell \in \mathcal{I}_J} \left( \delta_\ell \|t_\ell\|_2 + y^T t_\ell \right) \\ & \text{subject to} && \|u_j\|_\infty \leq \lambda_j, \forall j \in \mathcal{J}_J, \\ & && t_\ell \in \mathbb{R}^{M \times 1}, \forall \ell \in \mathcal{I}_J \setminus \mathcal{J}_J \\ & && t_j = - \left( \bar{D}_{j,j}^{-T} W_j u_j + \sum_{\ell \in c_j(\mathcal{I}_J) \setminus j} \bar{D}_{j,j}^{-T} \bar{D}_{\ell,j}^T t_\ell \right), \forall j \in \mathcal{J}_J, \end{aligned} \quad (4)$$

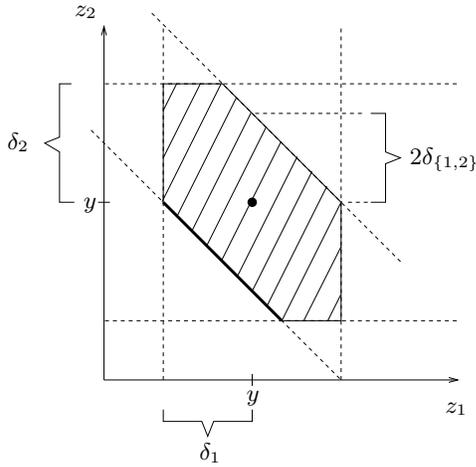


Fig. 3. An example of the feasible set (shaded region) in  $\mathbb{R}^{(1+1)} \times 1$ . The thick line indicates the optimal solutions for the problem of minimizing  $|z_1| + |z_2|$ .

where  $g(t)$  is the dual objective and  $Q_d$  the dual feasible set.

*Proof:* The dual function of (2) is given by

$$\bar{g}(t, \kappa) = \begin{cases} \sum_{j \in \mathcal{J}_J} \bar{g}_j(t) - \left( \sum_{\ell \in \mathcal{I}_J} t_\ell^T y + \delta_\ell \kappa_\ell \right), & \text{if } \|t_\ell\|_2 \leq \kappa_\ell, \forall \ell \in \mathcal{I}_J \\ -\infty, & \text{else} \end{cases}, \quad (5)$$

where

$$t = \{t_\ell\}_{\ell \in \mathcal{I}_J}, \quad \kappa = \{\kappa_\ell\}_{\ell \in \mathcal{I}_J},$$

$$\bar{g}_j(t) = \inf_{z_j} \tilde{g}_j(t) = \inf_{z_j} \left\{ \lambda_j \|W_j z_j\|_1 + \left( \sum_{\ell \in c_j(\mathcal{I}_J)} t_\ell^T \bar{D}_{\ell,j} \right) z_j \right\},$$

and

$$c_j(\mathcal{I}) = \{\ell \mid \ell \in \mathcal{I}, j \in \ell\}.$$

Note that the dual function is now decoupled in the functions  $g_j(t)$  with the implicit constraint  $\|t_\ell\|_2 \leq \kappa_\ell, \forall \ell \in \mathcal{I}_J$ . Furthermore,

$$\bar{g}_j(t) = \begin{cases} 0, & \text{if } \|u_j\|_\infty \leq \lambda_j, u_j = -W^{-1} \sum_{\ell \in c_j(\mathcal{I}_J)} \bar{D}_{\ell,j}^T t_\ell \\ -\infty, & \text{else.} \end{cases} \quad (6)$$

At this point, we change the implicit domain in (5) and (6), to explicit constraints and note that  $\kappa_\ell^* = \|t_\ell^*\|_2, \forall \ell \in \mathcal{I}_J$ , under maximization and thereby obtain the dual problem (4). ■

The equality constraints in (4) are simple because the variables  $t_j, \forall j \in \mathcal{J}_J$ , associated with the side descriptions, only occurs on the left hand side, while the rest of the variables  $t_\ell, \forall \ell \in \mathcal{I}_J \setminus \mathcal{J}_J$ , associated with the joint description, are on the right side. We can then make a variable substitution of  $t_j, \forall j \in \mathcal{J}_J$ , in the objective, but we choose the form (4) for readability.

### C. Smoothing

Since the dual problem has simple and non-intersecting constraints it is possible to efficiently apply first-order projection methods. The objective of the dual problem (4) is differentiable on  $\|t_\ell\|_2 > 0$  and sub-differentiable on  $\|t_\ell\|_2 = 0$ . The objective in the dual problem (4) is hence not smooth.<sup>3</sup> We could then apply an algorithm such as the sub-gradient algorithm with complexity  $\mathcal{O}(1/\epsilon^2)$  or form a

<sup>3</sup>A smooth function is a function with Lipschitz continuous derivatives [30].

smooth approximation and apply an optimal first-order method to the smooth problem and obtain complexity  $\mathcal{O}(\frac{1}{\epsilon})$ , as proposed in [12]. The primal feasible set has intersecting Euclidean norm ball constraints, so we cannot efficiently follow the algorithm [12], since this approach requires projections in both the primal and dual feasible set. We will next show how to adapt the results of [12], in the spirit of [36], using only projection on the dual feasible set and still achieve complexity  $\mathcal{O}(\frac{1}{\epsilon})$ . Consider

$$\|x\|_2 = \max_{\|v\|_2 \leq 1} \{v^T x\} \quad (7)$$

and the approximation

$$\Psi_\mu(x) = \max_{\|v\|_2 \leq 1} \left\{ v^T x - \frac{\mu}{2} \|v\|_2^2 \right\} \quad (8)$$

$$= \begin{cases} \|x\|_2 - \mu/2, & \text{if } \|x\|_2 \geq \mu \\ \frac{1}{2\mu} x^T x, & \text{otherwise} \end{cases}, \quad (9)$$

where  $\Psi_\mu(\cdot)$  is a Huber function with parameter  $\mu \geq 0$ . For  $\mu = 0$  we have  $\Psi_0(x) = \|x\|_2$ . The function  $\Psi_\mu(x)$  has for  $\mu > 0$  the (Lipschitz continuous) derivative

$$\nabla \Psi_\mu(x) = \frac{x}{\max\{\|x\|_2, \mu\}}.$$

The dual objective is

$$g(t) = - \sum_{\ell \in \mathcal{I}_J} \left( \delta_\ell \|t_\ell\|_2 + y^T t_\ell \right)$$

and we can then form a smooth function  $g_\mu$  as

$$g_\mu(t) = - \sum_{\ell \in \mathcal{I}_J} \left( \delta_\ell \Psi_\mu(t_\ell) + y^T t_\ell \right).$$

The Lipschitz constant of the gradient is  $L(\nabla \Psi_\mu(x)) = \frac{1}{\mu}$  and then

$$L_\mu = L(\nabla g_\mu(t)) = \left( \sum_{\ell \in \mathcal{I}_J} \frac{\delta_\ell}{\mu} + 1 \right) = \frac{C}{\mu} + |\mathcal{I}_J|.$$

Also,  $g(t)$  can be bounded as

$$g_\mu(t) \leq g(t) \leq g_\mu(t) + \mu C.$$

Now, fix  $\mu = \frac{\epsilon}{2C}$  and let the  $i$ th iteration  $t^{(i)}$  of an algorithm have the property

$$g_\mu^* - g_\mu(t^{(i)}) \leq \frac{\epsilon}{2}. \quad (10)$$

Then it follows that

$$g^* - g(t^{(i)}) \leq g_\mu^* + \mu C - g_\mu(t^{(i)}) \leq \epsilon. \quad (11)$$

By using an optimal-first order algorithm for  $L$ -smooth problems with complexity  $\mathcal{O}\left(\sqrt{\frac{L}{\epsilon}}\right)$  [30], then  $t^{(i)}$  can be obtained in

$$\begin{aligned} i &= \mathcal{O}\left(\sqrt{\frac{L_\mu}{\epsilon}}\right) = \mathcal{O}\left(\sqrt{\frac{1}{\mu\epsilon} + \frac{1}{\epsilon}}\right) = \mathcal{O}\left(\sqrt{\frac{1}{\epsilon^2} + \frac{1}{\epsilon}}\right) \\ &= \mathcal{O}\left(\sqrt{\frac{1}{\epsilon^2}} + \sqrt{\frac{1}{\epsilon}}\right) = \mathcal{O}\left(\frac{1}{\epsilon}\right) + \mathcal{O}\left(\sqrt{\frac{1}{\epsilon}}\right) = \mathcal{O}\left(\frac{1}{\epsilon}\right) \end{aligned} \quad (12)$$

iterations.<sup>4</sup>

<sup>4</sup>See e.g., [37] for a definition of the big-O notation.

#### D. Recovering primal variables from dual variables

The primal variables can be recovered as a minimizer  $z_j$  of  $\tilde{g}_j(t)$ , see [20, §5.5.5]. But since  $\|\cdot\|_1$  is not strictly convex there will in general be more than one minimizer. We will instead consider a different approach.

The Karush-Kuhn-Tucker (KKT) optimality conditions for the convex problem (2) are

$$\begin{cases} h_2(D_\ell z_\ell^* - y)\|t_\ell^*\|_2 - t_\ell^* \in 0, & \forall \ell \in \mathcal{I}_J \\ \kappa_\ell^* (\|D_\ell z_\ell^* - y\|_2 - \delta_\ell) = 0, & \forall \ell \in \mathcal{I}_J \\ \sum_{\ell \in c_j(\mathcal{I}_J)} D_{\ell,j}^T t_\ell^* + W_j u_j^* = 0, & \forall j \in \mathcal{J} \\ \|D_\ell z_\ell^* - y\|_2 \leq \delta_\ell, & \forall \ell \in \mathcal{I}_J \\ \lambda_j h_1(W_j z_j^*) - u_j^* \in 0, & \forall j \in \mathcal{J} \end{cases} \quad (\|t_\ell^*\|_2 = \kappa_\ell^*) \quad (13)$$

with  $h_a(x) = \partial\|x\|_a$ . We have for  $\delta_\ell > 0, \forall \ell \in \mathcal{I}_J$ ,

$$\begin{cases} h_2(D_\ell z_\ell^* - y)\|t_\ell^*\|_2 - t_\ell^* \in 0 \\ \|t_\ell^*\|_2 (\|D_\ell z_\ell^* - y\|_2 - \delta_\ell) = 0 \end{cases} \Leftrightarrow t_\ell^* = \frac{\|t_\ell^*\|_2}{\delta_\ell} (D_\ell z_\ell^* - y) \text{ if } \|t_\ell^*\|_2 > 0 \quad (14)$$

for all  $\ell \in \mathcal{I}_J$ . The system

$$\begin{cases} h_2(D_\ell z_\ell^* - y)\|t_\ell^*\|_2 - t_\ell^* \in 0, & \forall \ell \in \mathcal{I}_J \\ \kappa_\ell^* (\|D_\ell z_\ell^* - y\|_2 - \delta_\ell) = 0, & \forall \ell \in \mathcal{I}_J \\ \sum_{\ell \in c_j(\mathcal{I}_J)} D_{\ell,j}^T t_\ell^* + W_j u_j^* = 0, & \forall j \in \mathcal{J} \end{cases} \quad (\|t_\ell^*\|_2 = \kappa_\ell^*) \quad (15)$$

is then equivalent to

$$\sum_{\ell \in c_j(\mathcal{I}_J)} \frac{\|t_\ell^*\|_2}{\delta_\ell} \bar{D}_{\ell,j}^T D_\ell z_\ell^* = -W_j u_j^* + \sum_{\ell \in c_j(\mathcal{I}_J)} \frac{\|t_\ell^*\|_2}{\delta_\ell} \bar{D}_{\ell,j}^T y, \quad \forall j \in \mathcal{J}.$$

We can then obtain the equivalent KKT optimality conditions

$$\begin{cases} \sum_{\ell \in c_j(\mathcal{I}_J)} \frac{\|t_\ell^*\|_2}{\delta_\ell} \bar{D}_{\ell,j}^T D_\ell z_\ell^* = \\ -W_j u_j^* + \sum_{\ell \in c_j(\mathcal{I}_J)} \frac{\|t_\ell^*\|_2}{\delta_\ell} \bar{D}_{\ell,j}^T y, & \forall j \in \mathcal{J} \\ \|D_\ell z_\ell^* - y\|_2 \leq \delta_\ell, & \forall \ell \in \mathcal{I}_J \\ \lambda_j h_1(W_j z_j^*) - u_j^* \in 0, & \forall j \in \mathcal{J}. \end{cases} \quad (16)$$

Let  $z^* \in \mathcal{Z}$  be a solution to (16) and let  $\bar{z} \in \bar{\mathcal{Z}}$  be a solution to (16.Δ).

**Proposition IV.2.** (Uniqueness) *If the solution  $\bar{z}$  to the linear system (16.Δ) is unique and there exist a solution  $z^*$  to problem (1), then  $z^* = \bar{z}$ .*

*Proof:* We have from the assumption and the system (16) that  $\emptyset \neq \bar{\mathcal{Z}} \subseteq \mathcal{Z}$ . If  $|\bar{\mathcal{Z}}| = 1$  then  $|\mathcal{Z}| = 1$  such that  $\bar{z} = z^*$ . ■

Proposition IV.2 explains when it is interesting to solve the primal problem by the dual problem and then recover the primal variables by (16.Δ). The reason why we will focus on (16.Δ) is that the remaining equations in the system (16) are sub-differentiable and feasibility equations. These are difficult to handle - especially for large scale problems. On the other hand, the system (16.Δ) is linear in  $z$  and can easily be solved for invertible  $D_j, \forall j \in \mathcal{J}$ .

However, first we will analyze the implication of  $\kappa_\ell^* = 0$ . Let

$$\Omega_J = \{\ell \mid \kappa_\ell^* = \|t_\ell^*\|_2 = 0, \ell \in \mathcal{I}_J\},$$

with  $\Omega_J \subseteq \mathcal{I}_J$ . Then, solving the original primal problem (2) is equivalent to solving [20]

$$\begin{aligned} & \text{minimize} && \sum_{j \in \mathcal{J}} \lambda_j \|W_j z_j\|_1 \\ & \text{subject to} && \|D_\ell z_\ell - y\|_2 \leq \delta_\ell, \quad \forall \ell \in \mathcal{I}_J \setminus \Omega_J, \end{aligned} \quad (17)$$

where we can now remove constraints which are not strongly active. Similarly, if there is an  $i \in \mathcal{J}_J$ , such that

$$c_i(\mathcal{I}_J) \subseteq \Omega_J,$$

then this corresponds to minimization over an unconstrained  $z_i$ . Since by definition  $\lambda_i > 0$  and  $W_i \succ 0$  then  $z_i^* = 0$ . Solving the original primal problem (2) is, hence, equivalent to solving [20]

$$\begin{aligned} & \text{minimize} && \sum_{j \in \mathcal{J}_J \setminus i} \lambda_j \|W_j z_j\|_1 \\ & \text{subject to} && \|D_\ell z_\ell - y\|_2 \leq \delta_\ell, \quad \forall \ell \in \mathcal{I}_J \setminus \Omega_J \\ & && z_i = 0. \end{aligned} \quad (18)$$

**Definition IV.3.** (Trivial instance) *We will call an instance  $\{y, \{\delta_\ell\}_{\ell \in \mathcal{I}_J}, \{D_\ell\}_{\ell \in \mathcal{I}_J}, \{W_j\}_{j \in \mathcal{J}_J}, \{\lambda_j\}_{j \in \mathcal{J}_J}\}$  of the MDIIC problem a trivial instance if it can be reformulated as an MDIIC problem without coupled constraints.*

The reason why we call these trivial instances is because they do not include coupled constraints and therefore do not include the trade-off normally associated with MD problems. All trivial instances can be solved straightforwardly by a first-order primal method because they do not include any coupled constraints, see [38].

**Proposition IV.4.** *We have  $z^* = \bar{z}$  for all non-trivial instances of the standard MDIIC problem with  $J = 2$ .*

*Proof:* Let us represent the system (16.Δ) by  $\tilde{D}_J z^* = \tilde{u}$  and consider the determinant of this system to analyze under which conditions there is a unique solution. By factorizing  $\tilde{D}_J$  and using the multiplicative map of determinants  $\det(AB) = \det(A)\det(B)$ ,  $\det(A^T) = \det(A)$ ,  $\det(\alpha A) = \alpha^h \det(A)$  for  $A \in \mathbb{R}^{h \times h}$  we obtain the determinant

$$\det(\tilde{D}_J) = \left( \prod_{j \in \mathcal{J}_J} \det(D_j)^2 \right) \cdot \det \left( \left[ \sum_{\substack{\ell \in \mathcal{I}_J \\ i,j \in \ell}} \rho_{\ell,j} \rho_{\ell,i} \frac{\kappa_\ell^*}{\delta_\ell} \right]_{i=1, \dots, J; j=1, \dots, J} \right)^M, \quad (19)$$

for the standard MDIIC problem. For our example with  $J = 2$  we have

$$\det(\tilde{D}_2) = \det(D_1)^2 \det(D_2)^2 \left( \frac{\kappa_{\{1\}}^* \kappa_{\{2\}}^* \rho_{\{1\},1}^2 \rho_{\{2\},2}^2}{\delta_{\{1\}} \delta_{\{2\}}} + \frac{\kappa_{\{1\}}^* \kappa_{\{1,2\}}^* \rho_{\{1\},1}^2 \rho_{\{1,2\},2}^2}{\delta_{\{1\}} \delta_{\{1,2\}}} + \frac{\kappa_{\{2\}}^* \kappa_{\{1,2\}}^* \rho_{\{2\},2}^2 \rho_{\{1,2\},1}^2}{\delta_{\{2\}} \delta_{\{1,2\}}} \right)^M.$$

Since  $\rho_{\ell,j} > 0, \forall \ell \in \mathcal{I}_J, j \in \ell, \delta_\ell > 0, \forall \ell \in \mathcal{I}_J$  and  $\det(D_j) \neq 0, \forall j \in \mathcal{J}$ , the condition  $\det(\tilde{D}_2) = 0$  is determined by which  $\ell, \kappa_\ell^* = 0$ . Let

$$O_2 = \{\Omega_2 \mid \det(\tilde{D}_2) = 0\}.$$

Then  $O_2$  is given as equation

$$O_2 = \{ \{ \{1\}, \{2\} \}, \{ \{1\}, \{1,2\} \}, \{ \{2\}, \{1,2\} \}, \{ \{1\}, \{2\}, \{1,2\} \} \}.$$

Let us consider all the cases:

- $\{ \{1\}, \{2\} \}$ . No coupled constraints, which implies a trivial instance.
- $\{ \{1\}, \{1,2\} \}$  or  $\{ \{2\}, \{1,2\} \}$ . Corresponds to  $z_1^* = 0$  or  $z_2^* = 0$ . The primal problem can be solved directly over  $z_2$  or  $z_1$  with no coupled constraints, which implies a trivial instance.

- $\{\{1\}, \{2\}, \{1, 2\}\}$ . Corresponds to an instance with  $z_1^* = 0$  and  $z_2^* = 0$  and no coupled constraints, which implies a trivial instance. ■

Remark (Proposition IV.4): All the cases for  $J = 2$  can be seen as two descriptions transmitted as one description over one channel. It is not easy to analyze  $O_J$  for  $J \geq 3$  in all cases and compare them to the definition of trivial instances. However, we will make the following partial description on the number of active constraints to ensure recovery of optimal primal variables.

**Proposition IV.5.** For a standard MDIIC problem, if

- all side constraints are strongly active,  $\kappa_j^* > 0, \forall j \in \mathcal{J}_J$ , then  $z^* = \bar{z}$
- there are no strongly active constraints  $\kappa_\ell^* = 0, \forall \ell \in \mathcal{I}_J$ , then  $z^* = 0$ .

*Proof:*

- From (14) we have  $t_\ell^* = \frac{\|t_\ell^*\|_2}{\delta_\ell} (D_\ell z_\ell^* - y), \forall \ell \in \mathcal{J}_J$ , which gives a unique solution to  $z^* = \mathbf{C}_{j \in \mathcal{J}_J} D_j^{-1} \left( t_j^* \frac{\delta_j}{\|t_j^*\|_2} + y \right)$ . Since a subsystem ( $\mathcal{J}_J \subseteq \mathcal{I}_J$ ) of (16.Δ) has exactly one point, then  $|\bar{\mathcal{Z}}| \leq 1$ . A standard MDIIC problem is solvable such that  $|\bar{\mathcal{Z}}| \geq |\mathcal{Z}| \geq 1$ . Hence  $|\bar{\mathcal{Z}}| = 1$  and from Proposition IV.2 we then have  $z^* = \bar{z}$ .
- If  $\kappa_\ell^* = \|t_\ell^*\|_2 = 0, \forall \ell \in \mathcal{I}_J$ , then  $g(t^*) = 0$  and  $f(z^*) = 0$  according to strong duality. From the definition,  $W_j \succ 0, \forall j \in \mathcal{J}_J$  and  $\lambda_j > 0, \forall \mathcal{J}_J$ , then  $f(z^*) = 0 \Leftrightarrow z^* = 0$ . ■

Remark (Proposition IV.5): It is always possible to make all the inactive side distortion constraints strongly active by adjusting  $\delta_j, j \in \mathcal{J}_J$ , without significantly changing the original formulation. With this approach we can always recover the primal variables as  $z^* = \bar{z}$ .

### E. Stopping Conditions

Since we implement a primal-dual first-order method and the problem has strong duality, a primal-dual stopping criteria is interesting. From the dual iterates  $(t^{(i)}, u^{(i)})$  we obtain the primal iterate  $z^{(i)}$  as the solution to

$$\sum_{\ell \in c_j(\mathcal{I}_J)} \frac{\|t_\ell^{(i)}\|_2}{\delta_\ell} \bar{D}_{\ell,j}^T D_\ell z_\ell^{(i)} = -W_j u_j^{(i)} + \sum_{\ell \in c_j(\mathcal{I}_J)} \frac{\|t_\ell^{(i)}\|_2}{\delta_\ell} \bar{D}_{\ell,j}^T y,$$

for  $j \in \mathcal{J}_J$ . We then stop the first-order method at iteration  $i$  if

$$f(z^{(i)}) - g(t^{(i)}) \leq \epsilon, \quad z^{(i)} \in Q_p, \quad (t^{(i)}, u^{(i)}) \in Q_d.$$

To ensure scalability in the dimensions of the problem, we select  $\epsilon = JM\epsilon_r$ , where for example we may choose to solve the problem to medium accuracy, e.g.,  $\epsilon_r = 10^{-3}$ .

## V. ANALYZING THE SPARSE DESCRIPTIONS

In this section we will use an image example and analyze the sparse descriptions. In particular, we show that it is possible to obtain a sparse representation which has a lower cardinality for the same PSNR requirement, or better PSNR for same cardinality, using the MDIIC approach compared to the simple approach of thresholding.

For images, let  $y$  be the column major wise stacked version of a two-dimensional image of dimension  $m \times n$ ,  $M = mn$ . The images are normalized such that  $y \in [0; 1]^M$  and the PSNR is

$$\text{PSNR}(\delta) = 10 \log_{10} \left( \frac{1}{\frac{1}{M}\delta^2} \right).$$

We define  $D = \{D_\ell\}_{\ell \in \mathcal{I}_J}$ ,  $\delta = \{\delta_\ell\}_{\ell \in \mathcal{I}_J}$ ,  $\lambda = \{\lambda_j\}_{j \in \mathcal{J}_J}$ . We will denote  $\bar{z} = \Phi_D(y, \delta, \lambda)$  an  $\epsilon$ -suboptimal solution of the problem (2) with  $\bar{z} = \{z_j\}_{j \in \mathcal{J}_J}$  after 4 reweight iterations. Note that in the single channel case  $J = 1$  we will obtain the problem (1). We will also define the function  $\bar{z} = T_{D_j}(y, \delta)$  as the thresholding function of the smallest coefficients of  $D_j^{-1}y$  such that  $\text{PSNR}(\|D_j T_{D_j}(y, \delta) - y\|_2) \approx \delta$ .

We now select the two channel case  $J = 2$ ,  $y$  the Pirate standard image (Grayscale,  $512 \times 512$ ) and as dictionaries  $D_1^{-1}$ : a 2-dimensional orthogonal Symlet16 discrete wavelet transform (DWT) with 7 levels, and  $D_2^{-1}$ : a 2-dimensional biorthogonal CDF 9/7 DWT with 7 levels. The results are reported in Table I, where we for clarity will refer to different approaches using the numbering (1)-(4). First, we obtain  $\bar{z} = \Phi_D(y, \delta, \lambda)$  (1) and then apply thresholding to the same signal such that the side PSNRs are the same (2). Notice that due to the independent thresholding, the refinement  $\ell = \{1, 2\}$  is not much better than the individual descriptions. Considering the same setup, but where we select  $\tilde{\delta}_j$  such that the cardinality of each description is the same  $\text{card}((\Phi_D(y, \delta, \lambda))_j) = \text{card}(T_{D_j}(y, \tilde{\delta}_j))$  (3). In this case, we have a better side PSNR, but the refinement is still poor and the central distortion not as good as in the case of the MDIIC approach. Finally, if we performed thresholding to achieve the same PSNR on the side distortion as the central distortion (4) we see that we need an excessive cardinality. We conclude that the MDIIC framework is able to generate non-trivial descriptions in a MD framework with respect to both the cardinality of the descriptions and the refinement.

With the same setup as used in Table I, we also investigate the first-order iteration complexity for obtaining a solution to the MDIIC problem, including 4 reweight iterations. Each reweight iteration has the worst-case iteration complexity  $\mathcal{O}(\frac{1}{\epsilon})$  given in (12) which results in an overall complexity of  $\mathcal{O}(\frac{1}{\epsilon})$ . The results are shown in Fig. 4. In general, we obtain an empirical complexity slightly (but not significantly) better than the theoretical worst-case iteration complexity  $\mathcal{O}(\frac{1}{\epsilon})$ . For obtaining an  $\epsilon = \epsilon_r JM$ -suboptimal solution with  $\epsilon_r = 10^{-3}$  and 4 reweight iterations requires approximately 700 first-order iterations.

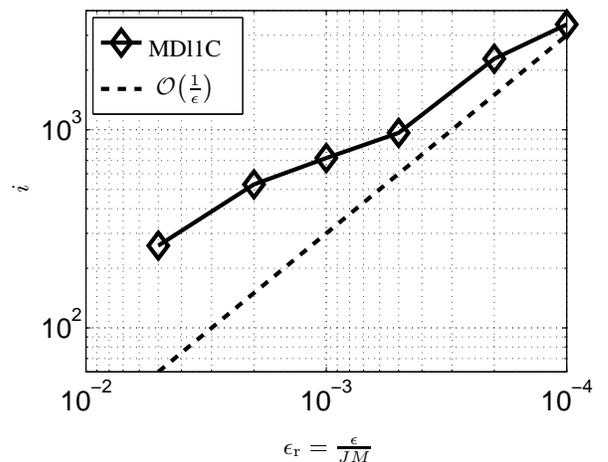


Fig. 4. Number of first-order iterations  $i$  including 4 reweight iterations as a function of of the relative accuracy  $\epsilon_r$ . We also plot the complexity function  $\mathcal{O}(\frac{1}{\epsilon})$  for comparison.

## VI. SIMULATION AND ENCODING OF SPARSE DESCRIPTIONS

In this section we will consider an application of the proposed scheme, where the sparse descriptions are encoded to adapt the sparse

ID	Method	card( $\bar{z}_j$ )/N		PSNR( $\ D_\ell \bar{z}_\ell - y\ _2$ )		
		$j = 1$	$j = 2$	$\ell = \{1\}$	$\ell = \{2\}$	$\ell = \{1, 2\}$
(1)	$\bar{z} = \Phi_D(y, \delta, \lambda)$	0.058	0.059	27.0	27.0	33.1
(2)	$\bar{z}_j = T_{D_j}(y, \delta_j)$	0.030	0.026	27.0	27.0	27.9
(3)	$\bar{z}_j = T_{D_j}(y, \delta_j)$	0.058	0.059	29.3	30.0	30.8
(4)	$\bar{z}_j = T_{D_j}(y, \delta_{\{1,2\}})$	0.128	0.112	33.1	33.1	34.4

TABLE I

COMPARISON BETWEEN MDL1C (1), THRESHOLDING TO SAME SIDE PSNR (2) THRESHOLDING TO SAME CARDINALITY (3) AND THRESHOLDING TO ACHIEVE THE CENTRAL DISTORTION ON EACH SIDE CHANNEL (4).

MD framework to a rate-distortion MD framework.

A state-of-the-art method for encoding images is the SPIHT encoder [13], which efficiently uses the tree structure of the DWT. We then denote the encoding of the coefficients  $\bar{z} = D^{-1}y$  as  $\hat{z} = \Gamma_D(\bar{z}, y, \delta)$ , such that  $\|D\hat{z} - y\|_2 \leq \delta$ . The encoder applies baseline SPIHT encoding. We add half a quantization level to all the significant wavelets transform coefficients when the stopping criteria  $\|D\hat{z} - y\|_2 \leq \delta$  is evaluated, see [39] or [40, Chapt. 6]. The quantized non-zero coefficients and their locations are further entropy coded using the arithmetic coder [41].

In order to illustrate the behaviour of encoding we will also obtain coefficients from solving a series of reweighted  $l_1$ -compression problems

$$\begin{aligned} & \text{minimize} && \|Wz\|_1 \\ & \text{subject to} && \|Dz - y\|_2 \leq \delta. \end{aligned}$$

An  $\epsilon$ -optimal solution from the above problem is denoted  $\bar{z} = \Phi_D(y, \delta, 1)$  (the same notation in the case of  $J = 1$ ).

The encoder can be used in two different ways  $\hat{z} = \Gamma_D(D^{-1}y, y, \delta)$  or  $\hat{z} = \Gamma_D(\Phi_D(y, \delta - \gamma, 1), y, \delta)$ . When encoding the coefficients  $\bar{z} = \Phi_D(y, \delta - \gamma, 1)$ , this corresponds to letting SPIHT encode the (reconstructed) image  $D\bar{z} \approx y$ . Note that we have introduced a modification of the distortion requirement with the parameter  $\gamma$ . This is done because  $\bar{z} = \Phi_D(y, \delta, 1)$  is on the boundary of the ball  $\|D\bar{z} - y\|_2 = \delta$  unless  $\bar{z} = 0$ . However, quantization slightly degrades the reconstruction quality and it is therefore difficult to ensure  $\|D\hat{z} - y\|_2 \leq \delta$  without requiring that the input  $\bar{z}$  to the encoder ensure  $\|D\bar{z} - y\|_2 \leq \delta - \gamma$ ,  $\gamma > 0$ . We use  $\gamma = 0.05\delta$ . It is possible to use any encoder which is based on encoding the coefficients associated to a linear reconstruction function and SPIHT is such an encoder.

In Fig. 5 we illustrate the results from encoding coefficients obtained as  $\hat{z} = \Gamma_D(D^{-1}y, y, \delta)$  or  $\hat{z} = \Gamma_D(\Phi_D(y, \delta - \gamma, 1), y, \delta)$ . As test images we use Lena and Boat (Grayscale,  $512 \times 512$ ) and we select  $D^{-1}$  as a Symlet16 pyramid wavelet transform with 7 levels.

For the simulations we choose different  $\delta$ 's and report the PSNR, the corresponding cardinality and the rate  $R(\hat{z})$  [bits/pixel]. From Fig. 5(a) we can see that for the same rate the reconstruction using the  $l_1$ -minimization approach  $\Phi_D$  shows a smaller PSNR than with the standard approach. This is to be expected, since for an orthogonal transform, SPIHT is designed to minimize the Euclidean distortion  $\|\hat{z} - \bar{z}\|_2$  to the input vector  $\bar{z} = D^{-1}y$  [13], and  $\|\hat{z} - \bar{z}\|_2$  is also our quality criteria. For  $\bar{z} = \Phi_D(y, \delta - \gamma, 1)$ , which implies  $\bar{z} \neq D^{-1}y$  in general, the design argumentation is slightly distorted by the modified input but the quality criteria remains the same. Further, by first forming a sparse coefficient vector using convex relaxation technique and later encode is suboptimal which further add to the loss. We also notice from Fig. 5(b) that the cardinality and bit rate behaves linearly in this setup for the  $l_1$ -minimization approach as also observed in other sparse decompositions, see e.g., [24].

### A. Encoding for Multiple Descriptions

We use the following approach when applying encoding for multiple descriptions, shown in Fig. 6. First the sparse coefficients vectors are formed using the function  $\Phi_D(y, \delta - \gamma, \lambda)$  with  $\gamma = \{\gamma_j\}_{j \in \mathcal{J}_J}$  and  $\gamma_j = 0.05 \min_{i \in \mathcal{J}_J} \delta_i$ ,  $\forall j \in \mathcal{J}_J$ . That is, we aim for at least a 5% better reconstruction in the optimization stage which we use to allow for a loss in the encoding stage. Each description is independently coded using  $\Gamma_{D_j}(D_j^{-1}\bar{z}_j, y, \delta_j - \tilde{\gamma}_j)$  to generate the encoded description vectors  $\hat{z}_j$  with the rate  $R(\hat{z}_j)$ . The parameter  $\tilde{\gamma}_j$  will be discussed shortly. From the received descriptions  $\ell$ , we then apply the decoding function  $g_\ell(\hat{z}_\ell) = D_\ell \hat{z}_\ell$ .

For the encoding function  $\Gamma_{D_j}$  it is easy to ensure  $\|D_j \hat{z}_j - y\|_2 \leq \delta_j$ ,  $\forall j \in \mathcal{J}_J$ , since we can encode each description independently until the side distortion is satisfied as we previously did in the example with  $J = 1$ . It is, however, more complicated to ensure that the coupled constraints are satisfied without requiring an excessive

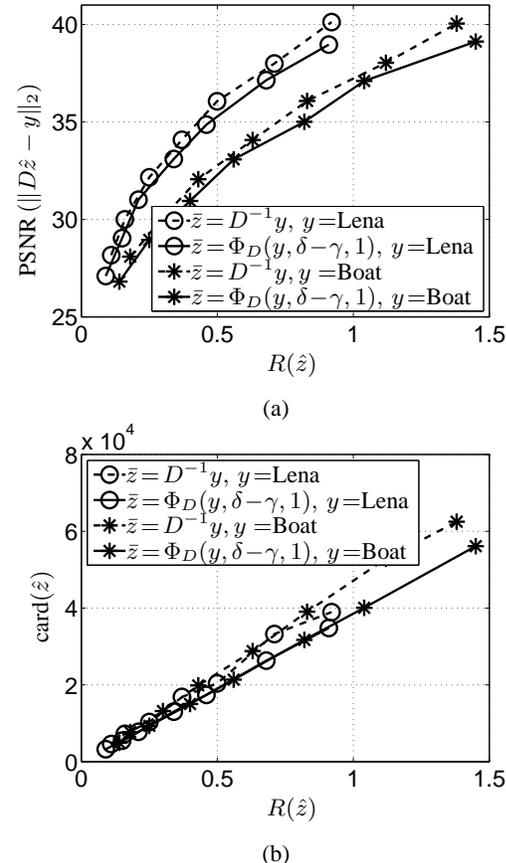


Fig. 5. Encoding of the images Lena and Boat, (Grayscale,  $512 \times 512$ ) using  $\hat{z} = \Gamma_D(\bar{z}, y, \delta)$  with different distortion requirement  $\delta$ . In (a) PSNR versus rate and (b) cardinality versus rate.

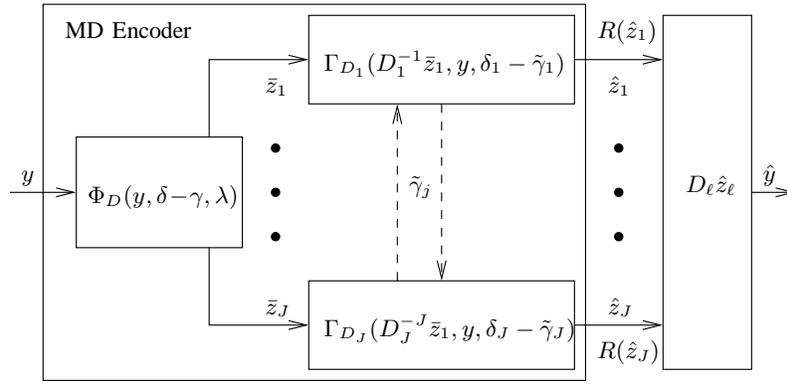


Fig. 6. Encoding of sparse coefficients in a MD setup.

large rate. Setting  $R(\hat{z}_j)$  large, we can always satisfy the distortion requirement, but by locating more appropriate points on the rate-distortion function of  $\hat{z}_j$ , we can achieve a better rate-distortion trade-off. To handle this problem we adjust  $\tilde{\gamma}_j$  by applying the following algorithm:

- Encode the descriptions and independently ensure that  $\|D_j \hat{z}_j - y\|_2 \leq \delta_j, \forall j \in \mathcal{J}_J$ .
- Check all the coupled distortion requirement  $\|D_\ell \hat{z}_\ell - y\|_2 \leq \delta_\ell, \forall \ell \in \mathcal{I}_J/\mathcal{J}_J$ .
  - For the first distortion requirement not fulfilled, check the first-order approximation of the slope  $\text{PSNR}(\|D_j \hat{z}_j - y\|_2)/R(\hat{z}_j)$  for  $j \in \mathcal{I}_J$ . Increase the rate (decrease  $\tilde{\gamma}_j$ ) of the description  $j$  with the largest slope using bisection.
  - If all distortion requirements of the encoded descriptions are satisfied then decrease the rate (increase  $\tilde{\gamma}_j$ ) for the description  $j \in \mathcal{J}_J$  with the smallest slope  $\text{PSNR}(\|D_j \hat{z}_j - y\|_2)/R(\hat{z}_j)$  using bisection.

The purpose of the above algorithm is to find a stable point of the Lagrange rate-distortion function. The process of adjusting the description rate using the description with highest or smallest slope respectively is also applied in [42]. The difference is that the above algorithm target distortion constraint (feasibility  $\hat{z} \in Q_p$ ) while [42] targets a rate constraint.

### B. MD Image Encoding

We perform comparison with state-of-the-art algorithms, specifically MDLT-PC [3], [43] and RD-MDC [42], [44] for the two channel case  $J = 2$  and  $y$  the Pirate image. We adjust the quantization levels such that we achieve a fixed rate and plot the (average) side and central distortion of the schemes under comparison. We also plot the single description performance of  $\Gamma_{D_1}$  and  $\Gamma_{D_2}$  as the distortion obtained at full rate and at half the rate and associate this to the central (horizontal line) and side distortion (vertical line), respectively. This corresponds to the extreme MD setup where there is no requirement for the side or central distortion, in which case a single description encoder is sufficient.

In Fig. 7 (a) and (b) we see that MDLT-PC and RD-MDC are able to obtain better single description PSNR than that of the SPIHT encoder using either  $\Gamma_{D_1}$  or  $\Gamma_{D_2}$ . Further, we observe that MD11C shows larger distortion at same rate, but behaves according to the single description bounds of the SPIHT encoder ( $\Gamma_{D_1}$  and  $\Gamma_{D_2}$ ), however with a gap. This gap is due to the suboptimality of the MD11C approach exemplified in Fig. 5.

We also show an example which demonstrates the flexibility of the proposed method. To this end we select  $J = 3$ , apply non-symmetric

distortion requirement  $\delta_\ell \neq \delta_{\ell'}$  for at least some  $|\ell| = |\ell'|$  with  $\ell, \ell' \in \mathcal{I}_J$ , non-symmetric weights  $\lambda_j \neq \lambda_{j'}$  for at least some  $j, j' \in \mathcal{J}_J$  and both orthogonal and biorthogonal dictionaries.

The results are shown in Fig. 8 where we obtain rates  $R(\hat{z}_1) = 0.32$ ,  $R(\hat{z}_2) = 0.52$  and  $R(\hat{z}_3) = 0.58$  such that  $R(\hat{z}) = 1.42$ . For comparison, if we encode the same image in a single description setup with the coder  $\Gamma_D$  using a biorthogonal Cohen-Daubechies-Feauveau (CDF) 9/7 DWT with 7 levels as dictionary  $D$  we obtain the distortion measure  $\text{PSNR}(\|D\hat{z}' - y\|_2) \approx 27.0$  or  $\text{PSNR}(\|D\hat{z}' - y\|_2) \approx 34.0$  at the rates  $R(\hat{z}') = 0.25$  or  $R(\hat{z}') = 0.81$ , respectively. This example is a large scale problem with  $M \times (|\mathcal{I}_J| + J + J) \approx 3.4 \cdot 10^6$  primal-dual variables. The encoding process required  $v = 13$  iterations and the SPIHT encoder was then applied  $J + (v - 1) = 15$  times since in the first iteration we need to encode all  $J$  descriptions and in the remaining iterations it is only necessary to encode the single description  $j$  for which  $\tilde{\gamma}_j$  was modified in the previous iteration.

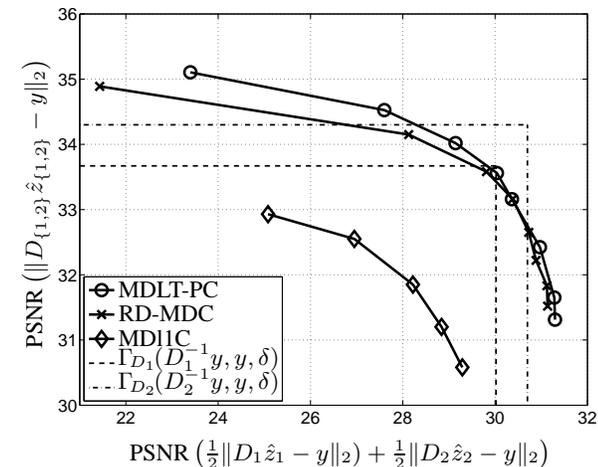
### C. MD Image Sequence Encoding

To show the flexibility of the proposed framework, we give an image sequence example. An image sequence can be seen as a three dimensional signal. If we join  $k$  consequent frames in a single block we obtain a windowed three dimensional signal with dimension  $m \times n \times k$  with  $m \times n$  being the frame dimensions of the video. For image sequences we then form  $y$  as a column major wise stacked version of a each two-dimension frame with  $M = mnk$ . For encoding, we apply 3D SPIHT [45]. To comply the 3D SPIHT framework we also form the dictionaries as three dimensional DWTs. As dictionaries we select  $D_{\{1\}}^{-1}$ : a 3 level orthogonal Haar DWT along the (temporal) dimension associated with  $k$  and a 2-dimensional orthogonal Symlet16 DWT with 5 levels along the (spatial) dimensions associated with  $m, n$ , and  $D_{\{2\}}^{-1}$ : a 3 level orthogonal Haar DWT along the (temporal) dimension associated with  $k$  and a 2-dimensional orthogonal Daub8 DWT with 5 levels along the (spatial) dimensions associated with  $m, n$ .

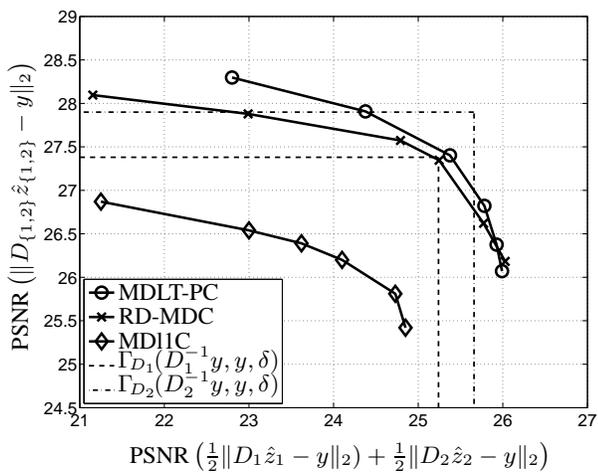
Fig. 9 shows an example for an image sequence where we have defined a frame extraction function  $s(y, \bar{k})$  which takes the  $\bar{k}$ th frame from the image sequence stacked in  $y$ . For this example we obtain the rates  $R(\hat{z}_1) = 0.10$  and  $R(\hat{z}_2) = 0.12$  such that  $R(\hat{z}) = 0.22$ . For comparison, if we had encoded the same image with the coder  $\Gamma_D$  using  $D_{\{2\}} = D$  as dictionary we obtain the distortion measure  $\text{PSNR}(\|D\hat{z}' - y\|_2) \approx 29.3$  or  $\text{PSNR}(\|D\hat{z}' - y\|_2) \approx 34.0$  at the rates  $R(\hat{z}') = 0.05$  or  $R(\hat{z}') = 0.17$ , respectively. This example is a large scale problem with  $M \times (|\mathcal{I}_J| + J + J) \approx 5.7 \cdot 10^6$  primal-dual variables. It is expected that the comparison between the MD11C method and state-of-the-art MD video coder will render similar results as in Fig. 7, that is, overall determined by the single



Fig. 8. Encoding the image Barbara (Grayscale,  $512 \times 512$ ). As dictionaries we use  $D_{\{1\}}^{-1}$ : a 2-dimensional orthogonal Symlet8 DWT with 7 levels,  $D_{\{2\}}^{-1}$ : a 2-dimensional orthogonal Symlet16 DWT with 7 levels, and  $D_{\{3\}}^{-1}$ : a 2-dimensional biorthogonal CDF 9/7 DWT with 7 levels. We have  $\lambda_1 = 1.5$ ,  $\lambda_2 = 1.4$  and  $\lambda_3 = 1.0$ . The distortion requirements and actual distortions are given above the individual images using the notation  $(\text{PSNR}(\|D_\ell \hat{z}_\ell - y\|_2), \text{PSNR}(\delta_\ell))$ . The resulting rates are respectively  $R(\hat{z}_1) = 0.32$ ,  $R(\hat{z}_2) = 0.52$ ,  $R(\hat{z}_3) = 0.58$  such that  $R(\hat{z}) = 1.42$ .



(a)  $R(\hat{z}) = 1$



(b)  $R(\hat{z}) = 0.25$

Fig. 7. Comparison between different MD methods for the image Pirate (Grayscale,  $512 \times 512$ ) at: (a) rate  $R(\hat{z}) = 1$  and (b) rate  $R(\hat{z}) = 0.25$ . The plot shows central distortion versus (average) side distortion. The single description performance using  $\Gamma_{D_1}$  and  $\Gamma_{D_2}$  are also shown as the distortion at full rate and at half the rate which are associated to respectively the central distortion (horizontal lines) and side distortion (vertical lines). As dictionaries we use  $D_{\{1\}}^{-1}$ : a 2-dimensional orthogonal Symlet16 DWT with 7 levels, and  $D_{\{2\}}^{-1}$ : a 2-dimensional biorthogonal CDF 9/7 DWT with 7 levels.

channel encoder and a gap introduced by the suboptimal approach of forming sparse descriptions using convex relaxation. Comparison between 3D SPIHT and standard video encoding schemes are given in, e.g., [45], [46].

## VII. CONCLUSION

We have shown how to use efficient first-order convex optimization techniques in a multiple description framework in order to form sparse descriptions, which satisfies a set of individual and joint distortion constraints. The proposed convex formulation allows for non-symmetric distortions, non-symmetric  $l_1$ -measures, different dictionaries and an arbitrary number of descriptions. We analyzed the sparse descriptions and concluded that the sparse descriptions were non-trivial. When encoding the sparse coefficients and comparing

with state-of-the-art methods it was not possible to achieve the same rate-distortion performance. On the other hand, the proposed method allow for a more flexible formulation and provides an algorithm for applying encoding in sparse signal processing. Efficient encoding of sparse coefficients is generally an open research topic.

## REFERENCES

- [1] A. A. E. Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851 – 857, Nov. 1982.
- [2] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell System Technical Journal*, vol. 59, pp. 1909 – 1921, Dec. 1980.
- [3] U. G. Sun, J. Liang, C. Tian, C. Tu, and T. Tran, "Multiple description coding with prediction compensation," *IEEE Trans. Image Process*, vol. 18, no. 5, pp. 1037–1047, May 2009.
- [4] H. Chan and C. Huang, "Multiple description and matching pursuit coding for video transmission over the internet," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, Hong Kong, Apr. 2003, pp. 425–428.
- [5] T. Nguyen and A. Zakhour, "Matching pursuits based multiple description video coding for lossy environments," in *Proc. Int. Conf. of Image Process. (ICIP)*, Sep. 2003, pp. 57–60.
- [6] T. Petrisor, B. Pesquet-Popescu, and J.-C. Pesquet, "A compressed sensing approach to frame-based multiple description coding," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, Honolulu, Hawaii, Apr. 2007, pp. 709–712.
- [7] Y. Zhang, S. Mei, Q. Chen, and Z. Chen, "A multiple description image/video coding method by compressed sensing theory," in *IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Seattle, Washington, May. 2008, pp. 1830–1833.
- [8] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [9] P. Boufounos and R. Baraniuk, "1-bit compressive sensing," in *Proc. 42nd annual Conf. on Inform. Sci. and Syst. (CISS)*, Princeton, New Jersey, Mar. 2008, pp. 16 – 21.
- [10] J. Z. Sun and V. K. Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Seoul, Korea, Jun.-Jul 2009, pp. 6–10.
- [11] A. Zymnis, S. Boyd, and E. Candès, "Compressed sensing with quantized measurements," *Signal Processing Letters*, vol. 17, no. 2, pp. 149–152, Feb. 2010.
- [12] Y. Nesterov, "Smooth minimization of nonsmooth functions," *Math. Prog. Series A*, vol. 103, pp. 127–152, 2005.
- [13] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
- [14] D. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [15] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21 – 30, Mar. 2008.
- [16] E. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted  $l_1$  minimization," *The Journal of Fourier Analysis and Applications, Special Issue on Sparsity*, vol. 14, no. 5, pp. 877–905, Dec. 2008.
- [17] M. Davies and R. Gribonval, "Restricted isometry constants where  $l_p$  sparse recovery can fail for  $0 < p \leq 1$ ," *IEEE Trans. on Inf. Theory*, vol. 55, no. 5, pp. 2203–2214, May 2009.
- [18] T. L. Jensen, J. Dahl, J. Østergaard, and S. H. Jensen, "A first-order method for the multiple-description  $l_1$ -compression problem," in *Proc. of Signal Processing with Adaptive Sparse Structured Representations (SPARS'09)*, Saint-Malo, France, Apr. 2009, Available at: <http://hal.inria.fr/inria-00369483/PDF/77.pdf>.
- [19] T. L. Jensen, J. Østergaard, J. Dahl, and S. H. Jensen, "Multiple descriptions using sparse decompositions," in *Proc. European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 110–114.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [21] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "Multiple description quantization via Gram-Schmidt orthogonalization," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5197 – 5217, Dec. 2006.

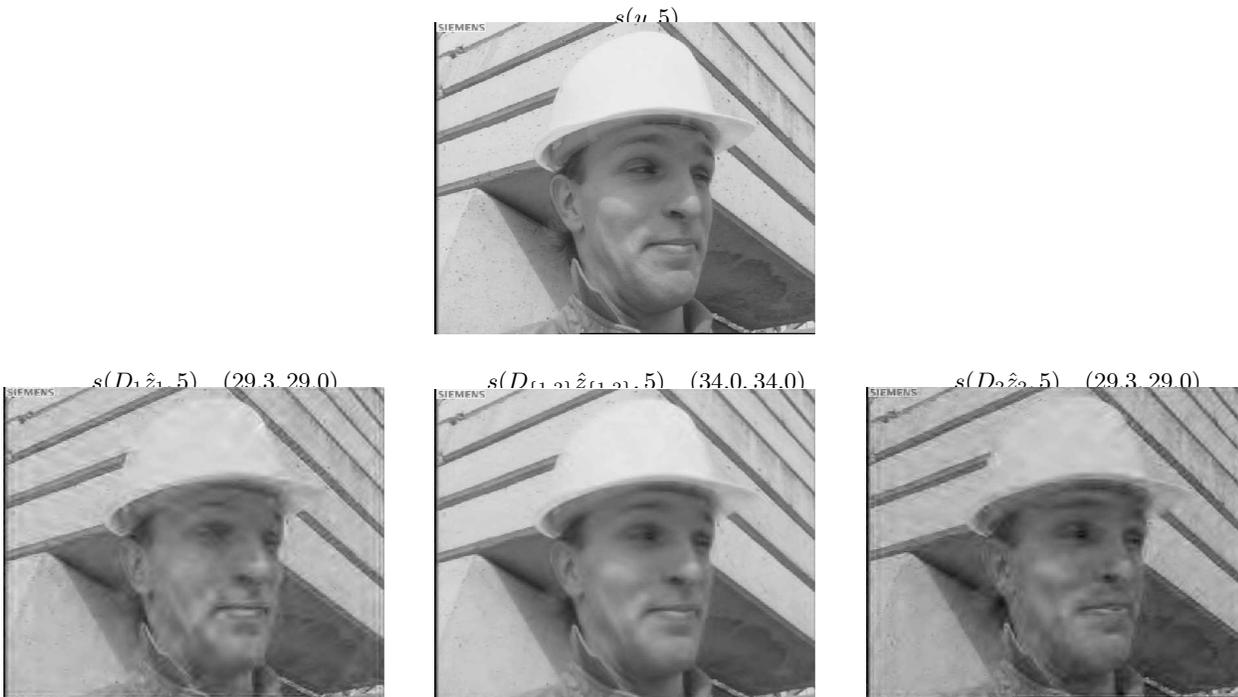


Fig. 9. Encoding the image sequence Foreman (Grayscale,  $288 \times 352$ ). We jointly process  $k = 8$  consecutive frames and let  $\lambda_1 = \lambda_2 = 1.0$ . The dictionaries are  $D_{\{1\}}^{-1}$ : a 3 level orthogonal Haar DWT along the dimension associated with  $k$  and a 2-dimensional orthogonal Symlet16 DWT with 5 levels along the dimensions associated with  $m, n$ , and  $D_{\{2\}}^{-1}$ : a 3 level orthogonal Haar DWT along the dimension associated with  $k$  and a 2-dimensional orthogonal Daub8 DWT with 5 levels along the dimensions associated with  $m, n$ . The distortion requirement and actual distortion are given above the individual images using the notation  $(\text{PSNR}(\|D_\ell \hat{z}_\ell - y\|_2), \text{PSNR}(\delta_\ell))$ . The resulting rates are respectively  $R(\hat{z}_1) = 0.10$  and  $R(\hat{z}_2) = 0.12$  such that  $R(\hat{z}) = 0.22$ .

[22] J. Østergaard and R. Zamir, "Multiple description coding by dithered delta-sigma quantization," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4661–4675, Oct. 2009.

[23] Y. Kochman, J. Østergaard, and R. Zamir, "Noise-shaped predictive coding for multiple descriptions of a colored Gaussian source," in *Proc. IEEE Data Comp. Conf. (DCC)*, Snowbird, Utah, Mar. 2008, pp. 362 – 371.

[24] H. Zhihai and S. K. Mitra, "A linear source model and a unified rate control algorithm for DCT video coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, no. 11, pp. 970–982, Nov. 2002.

[25] S. Diggavi, N. Sloane, and V. Vaishampayan, "Asymmetric multiple description lattice vector quantizers," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 174 – 191, Jan. 2002.

[26] J. Østergaard, R. Heusdens, and J. Jensen, " $n$ -channel asymmetric entropy-constrained multiple-description lattice vector quantization," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6354–6375, Dec. 2010.

[27] S. J. Wright, *Primal-Dual Interior-Point Methods*. SIAM, 1997.

[28] J. Barzilai and J. Borwein, "Two point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, Jan. 1988.

[29] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1995.

[30] Y. Nesterov, *Introductory Lectures on Convex Optimization*. Kluwer Academic Publishers, 2004.

[31] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[32] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 1–39, 2011.

[33] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. on Imaging Sciences*, vol. 2, pp. 183–202, Mar. 2009.

[34] J. Dahl, P. C. Hansen, S. H. Jensen, and T. L. Jensen, "Algorithms and software for total variation image reconstruction via first-order methods," *Numer. Algo.*, vol. 53, no. 1, pp. 67–92, Jan. 2010.

[35] G. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations Research*, vol. 8, no. 1, pp. 101 – 111, Jan.-Feb. 1960.

[36] L. Vandenberghe, "Optimization methods for large-scale systems," Lecture Notes, 2009.

[37] R. Graham, D. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, 1994.

[38] J. Dahl, J. Østergaard, T. L. Jensen, and S. H. Jensen, "An efficient first-order method for  $\ell_1$  compression of images," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 1009–1012.

[39] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 2009.

[40] E. K. Rao and P. Yip, *The Transform and Data Compression Handbook*. CRC Press, 2001.

[41] A. Said, "FastAC: Arithmetic coding implementation," Public available software for arithmetic coding, 2004.

[42] T. Tillo, M. Grangotto, and G. Olmo, "Multiple description image coding based on lagrangian rate allocation," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 673–683, Mar. 2007.

[43] MDLT-PC, Source code, Available at: <http://www.ensc.sfu.ca/~jje1/MDLT-PC.html>.

[44] RD-MDC, Source code, Available at: <http://www.telematica.polito.it/sas-ipl/>.

[45] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proc. IEEE Data Comp. Conf. (DCC)*, Snowbird, Utah, Mar. 1997, pp. 251–260.

[46] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, pp. 1374–1387, Dec. 2000.



**Tobias Lindstrøm Jensen** received his M.Sc. in Electrical Engineering from Aalborg University in 2007 and was awarded the Computational Science in Imaging (CSI) Ph.D. fellowship from the Danish Research Council. Tobias L. Jensen is currently affiliated with the Department of Electronic Systems at Aalborg University. In 2007, he was a research associate at Wipro-NewLogic Technologies in Sophia-Antipolis, France. In 2009 he was a visiting researcher scholar at University of California, Los Angeles (UCLA). His research interests includes

signal processing, convex optimization, estimation and inverse problems.



**Søren Holdt Jensen** received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1988, and the Ph.D. degree in signal processing from the Technical University of Denmark, Lyngby, Denmark, in 1995. Before joining the Department of Electronic Systems of Aalborg University, he was with the Telecommunications Laboratory of Telecom Denmark, Ltd, Copenhagen, Denmark; the Electronics Institute of the Technical University of Denmark; the Scientific Computing Group of Danish Computing Center for Research

and Education, Lyngby; the Electrical Engineering Department of Katholieke Universiteit Leuven, Leuven, Belgium; and the Center for PersonKommunikation (CPK) of Aalborg University. He is Full Professor and is currently heading a research team working in the area of numerical algorithms, optimization and signal processing for speech and audio processing, image and video processing, multimedia technologies, and digital communications. Prof. Jensen was an Associate Editor for the IEEE Transactions on Signal Processing and Elsevier Signal Processing, and is currently Member of the Editorial Board of EURASIP Journal on Advances in Signal Processing. He is a recipient of an European Community Marie Curie Fellowship, former Chairman of the IEEE Denmark Section, and Founder and Chairman of the IEEE Denmark Section's Signal Processing Chapter. Recently, he became member of the Danish Academy of Technical Sciences, and was appointed as member of the Danish Council for Independent Research - Technology and Production Sciences by the Danish Minister for Science, Technology and Innovation.



**Jan Østergaard** (S'98 – M'99) received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1999 and the Ph.D. degree (*cum laude*) in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 2007. From 1999 to 2002, he worked as an R&D engineer at ETI A/S, Aalborg, Denmark, and from 2002 to 2003, he worked as an R&D engineer at ETI Inc., Virginia, United States. Between September 2007 and June 2008, he worked as a post-doctoral researcher in the Centre

for Complex Dynamic Systems and Control, School of Electrical Engineering and Computer Science, The University of Newcastle, NSW, Australia. From June 2008 to March 2011, he worked as a post-doctoral researcher at Aalborg University, Aalborg, Denmark. He has also been a visiting researcher at Tel Aviv University, Tel Aviv, Israel, and at Universidad Técnica Federico Santa María, Valparaíso, Chile. He has received a Danish Independent Research Council's Young Researcher's Award and a fellowship from the Danish Research Council for Technology and Production Sciences. Dr. Østergaard is currently an Associate Professor at Aalborg University, Aalborg, Denmark.



**Joachim Dahl** received the Ph.D. degree in electrical engineering from Aalborg University in 2003. He was a post doctoral researcher at University of Los Angeles, California from 2004 to 2005. He currently works for MOSEK ApS., an optimization software vendor located in Copenhagen, Denmark.