A First-Order Method for the Multiple-Description ℓ_1-Compression Problem

Tobias Lindstrøm Jensen†, Joachim Dahl‡, Jan Østergaard†, Søren Holdt Jensen†

†Aalborg University
Department of Electronic Systems
Niels Jernesvej 12, 9220 Aalborg,
Denmark
Ph: +45 99408684
Fax: +45 98151583

‡Anybody Technology A/S
Niels Jernesvej 12
9220 Aalborg
Denmark

***Abstract*—In this paper we introduce the multiple-description ℓ_1-compression problem: minimize $\|z_1\|_1 + \lambda\|z_2\|_1$ subject to the three distortion constraints $\|A_1 z_1 - y\|_2 \leq \delta_1$, $\|A_2 z_2 - y\|_2 \leq \delta_2$, and $\|1/2(A_1 z_1 + A_2 z_2) - y\|_2 \leq \gamma$. This problem formulation is interesting in, *e.g.*, ad-hoc networks where packets can be lost. If a description (z_2) is lost in the network and only one description is received (z_1), it is still possible to decode and obtain a signal quality equal or better than described by the parameter δ_1 (and vice versa). If both descriptions are received, the quality is determined by the parameter γ. This problem is difficult to solve using first-order projection methods due to the intersecting second-order cones. However, we show that by recasting the problem into its dual form, one can avoid the difficulties due to conflicting fidelity constraints. We then show that efficient first-order ℓ_1-compression methods are applicable, which makes it possible to solve large scale problems, *e.g.*, multiple-description ℓ_1-compression of video.**

I. INTRODUCTION

There has been great interest in sparse estimation techniques for signal processing based on the convex ℓ_1-approximation of the otherwise intractable cardinality measure, see *e.g.*, [1]. In particular, efficient techniques based on sparse approximations have been successfully applied to such different problems as, *e.g.*, estimation and coding [2], preconditioning [3] and linear prediction [4]. For example, in [5], it is shown how a sparse approximation of an image sequence (video) can be obtained with a higher degree of sparsity than simple thresholding in the transform domain using the discrete cosine transform (DCT).

One way of obtaining a sparse approximation is to solve the so-called ℓ_1-*compression* problem

$$\begin{array}{ll} \text{min.} & \|z\|_1 \\ \text{s.t.} & \|Az - y\|_2 \leq \delta, \end{array} \qquad (1)$$

where $\delta > 0$ is a given reconstruction error, $A \in \mathbb{R}^{M \times N}$ is the 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.

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

Another interesting problem in signal and source coding is the *multiple-description* (MD) problem. The MD problem is about encoding a source into multiple descriptions, which are transmitted over separate channels. The channels may occasionally break down causing 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 extend, the descriptions are also similar to each other. However, if the descriptions are the same, they cannot refine each other.

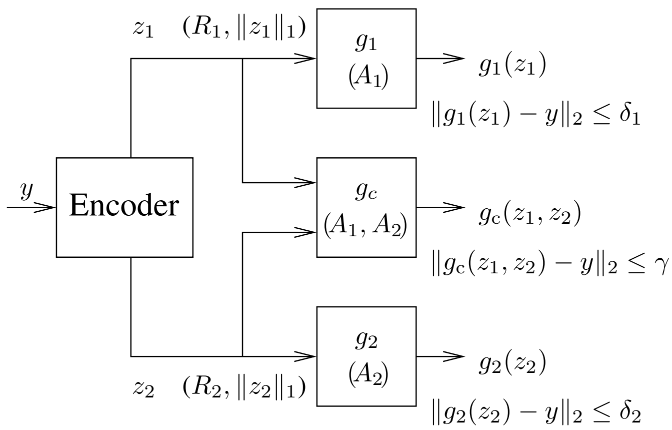

Fig. 1. The MD ℓ_1-compression problem.

The traditional MD coding problem aims at characterizing the set of achievable quintuples $(R_1, R_2, \delta_1, \delta_2, \gamma)$ where R_1 and R_2 denote the minimum coding rates required in order to approximate the source y to within the distortion fidelities δ_1, δ_2, and γ [6]. Specifically, let z_1 and z_2 be the two descriptions representing the source y via their individual reconstructions $g_1(z_1)$ and $g_2(z_2)$ and their joint description $g_c(z_1, z_2)$ satisfying $d(g_1 z_1, y) \leq \delta_1, d(g_2(z_2), y) \leq \delta_2$, and $d(g_c(z_1, z_2), y) \leq \gamma$, where $d(\cdot, \cdot)$ denotes a fidelity measure. The problem is then to construct z_1 and z_2 so that R_1 and R_2 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 a memoryless Gaussian sources [7].

In this paper, we cast the MD problem described above into the framework of ℓ_1-compression (1). Our idea is to form the *multiple-description ℓ_1-compression* problem using linear reconstruction functions, *i.e.*, $g_1(z_1) = A_1 z_1, g_2(z_1) = A_2 z_2$, and $g_c(z_1, z_2) = \frac{1}{2}(A_1 z_1 + A_2 z_2)$.[1] That is

$$\begin{array}{ll} \min. & \|W_1 z_1\|_1 + \lambda \|W_2 z_2\|_1 \\ \text{s.t.} & \|A_1 z_1 - y\|_2 \leq \delta_1 \\ & \|A_2 z_2 - y\|_2 \leq \delta_2 \\ & \|\frac{1}{2}(A_1 z_1 + A_2 z_2) - y\|_2 \leq \gamma \,. \end{array} \tag{2}$$

In this problem, we have introduced $\lambda > 0$ to allow weighting the ℓ_1-norms in order to achieve a desired ratio $\frac{\|z_1\|_1}{\|z_2\|_1}$. Also, even though that the ℓ_1-norm is a well-known approximation of the minimal cardinality solution, it penalizes large coefficients. One method to reduce the penalty is to weight the coefficients with the diagonal matrices $W_1 = \mathbf{diag}(w_1)$, $W = \mathbf{diag}(w_2)$ [11]. The problem is first solved with $W_1 = W_2 = I$. Then the weights w_1, w_2 are selected inversely proportional to the solutions z_1^*, z_2^* and the problem is solved again. These iterations can be executed a number of times.

For simplicity, we will assume that A_1 and A_2 are orthogonal transforms, which means that the central reconstruction represents a two times overcomplete basis, $A_1 z_1 + A_2 z_2 = [A_1\, A_2][z_1;\, z_2]$. The two constraints on the side reconstructions can easily be fulfilled by simply truncating the smallest coefficients to zero in the transform domain separately for each set of coefficients. This will however not guarantee the central reconstruction constraint $\|\frac{1}{2}(A_1 z_1 + A_2 z_2) - y\|_2 \leq \gamma$. So, even by allowing A_1 and A_2 to be orthogonal transforms, the problem is non-trivial. Also, traditionally orthogonal transform such as the DCT and wavelets are used for image and video coding.

It is worth emphasizing that the main difference between the traditional MD problem and the proposed MD ℓ_1-compression problem, is that the former strives at minimizing description rates in a stochastic setting whereas the latter minimizes the ℓ_1-norm in a deterministic setting. Interestingly, it has been shown that for typical transform coding systems, there is a linear relationship between the actual coding bit rate and the sparsity (i.e., the cardinality) of the transformed signal [12]. Moreover, it has been observed that minimizing the ℓ_1-norm usually results in a sparse signal, and there even exist cases where one can mathematically prove that minimizing ℓ_1 is equivalent to minimizing the cardinality [13]. This brings forth the possibility that the proposed MD ℓ_1-compression framework, provides a practical means for solving the otherwise difficult classical information theoretic MD problem.

Related approaches using matching pursuit have been presented, see [14]–[16] and references there in, however, much heuristics are used in order to obtain the two descriptions and to ensure the desired fidelity. For example in [14], they assign the 50 largest coefficients to both descriptions and then alternate between the two descriptions when assigning the remaining smaller coefficients.

The remaining paper is structured as follows: in Sec. II we discuss different approaches to solve the proposed MD problem and in Sec. III we present a numerical method to solve the problem. In Sec. IV we give numerical examples of ℓ_1-compression of a full-size image and an image sequence. Finally, we give discussions in Sec. V.

[1] 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 [8], [9] and [10] for the white and colored cases, respectively.

II. APPROACHES FOR SOLVING THE MD ℓ_1-COMPRESSION PROBLEM

The MD ℓ_1-compression problem can be solved using general-purpose interior point methods. To do so, we need to solve several linear systems of equations of at least size $N \times N$, arising from linearizing first-order optimality conditions. 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 [17]. Another approach is to use first-order projection methods [18]–[20], where the projection is on the feasible set. Such first-order projection methods have shown to be efficient for large scale problems [5], [21]–[24]. But it is difficult to solve the MD ℓ_1-compression problem efficiently because the feasible set is an intersection of the second-order cones (z_1 and z_2 appears in one constraints each and one joint).

In order to illustrate the implications of the overlapping constraints on the feasible set, consider the following simple one-dimensional example. Let $A_1 = A_2 = W_1 = W_2 = 1$ and $\lambda = 1$ so that $A_1 z_1 = z_1$ and $A_2 z_2 = z_2$. 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 $|\frac{1}{2}(z_1 + z_2) - y| \leq \delta$. However, due to the individual constraints on z_1 and z_2, the feasible set is bounded as illustrated in Fig. 2. The problem is to pick a pair (z_1, z_2) from within the shaded region such that the sum $|z_1| + |z_2|$ achieves its minimum. In this particular case, the optimal solution lies on the diagonal line closest to the original as illustrated in Fig. 2.

If we instead consider the dual of the problem (2)

$$\begin{array}{ll} \max\limits_{u_1,\, u_2,\, t} & -\delta_1 \|\frac{1}{2} t + A_1 W_1 u_1\|_2 - \delta_2 \|\frac{1}{2} t + A_2 W_2 u_2\|_2 \\ & \quad -\gamma \|t\|_2 + y^T (A_1 W_1 u_1) + y^T (A_2 W_2 u_2) \\ \text{s.t.} & \|u_1\|_\infty \leq 1 \\ & \|u_2\|_\infty \leq \lambda \,, \end{array} \tag{3}$$

we see that we have simple and non-intersecting constraints. This makes the dual problem interesting for first-order projection methods. The objective in the dual problem (3) is not

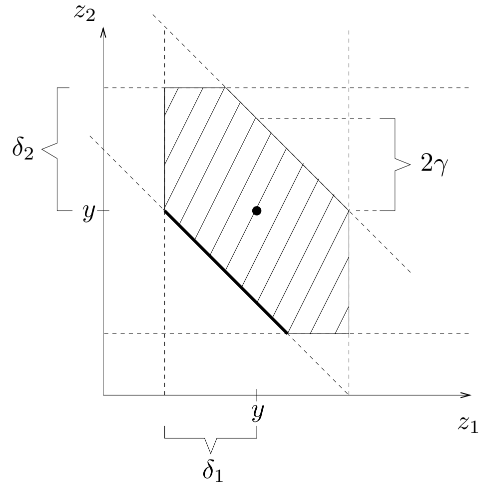

Fig. 2. An example of the feasible set (shaded region) in one dimension. The thick line indicates the optimal solutions for the problem of minimizing $|z_1| + |z_2|$.

smooth[2] because of the $\|\cdot\|_2$-norm. We could then apply an algorithm such as the sub-gradient algorithm with complexity $O(1/\epsilon^2)$, but instead we will try to smooth the problem in order to obtain faster convergence, as suggest in [20]. The primal feasible set has intersecting second-order cone constraints, so we can not efficiently perform smoothing using a prox-function as in [20]. Instead we apply the smoothing directly by modifying the $\|\cdot\|_2$-norm to a function which is very similar.

III. A FIRST-ORDER METHOD

We start by approximating the $\|\cdot\|_2$-norm with $\Psi_\mu(\cdot)$ where

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

and μ is a parameter. For the selection $\mu = 0$ we have $\Psi_0(x) = \|x\|_2$. This approximation is in-fact a result of the prox-function smoothing in [20, ex. 4.2]. The function $\Psi_\mu(x)$ has the (Lipschitz continuous) derivative

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

Fig. 3 shows an example with the $\|\cdot\|_2$-norm and the smooth approximation Ψ_μ in one dimension.

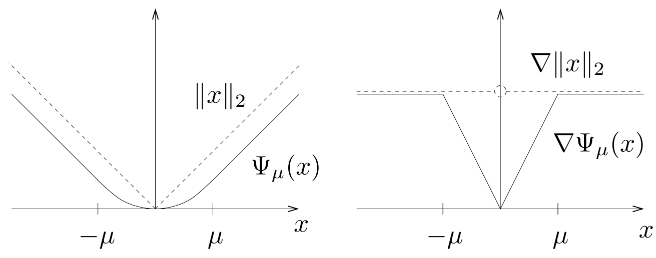

Fig. 3. Example of smoothing in one dimension. In this one-dimensional case, the ℓ_2-norm is equivalent to the ℓ_1-norm.

[2]A smooth function is a function with Lipschitz continuous derivatives [19].

Let $-f$ be the dual objective

$$\begin{aligned} f(u_1, u_2, t) = {} & \delta_1\|\tfrac{1}{2}t + A_1W_1u_1\|_2 + \delta_2\|\tfrac{1}{2}t + A_2W_2u_2\|_2 \\ & + \gamma\|t\|_2 - y^T(A_1W_1u_1) - y^T(A_2W_2u_2) . \end{aligned}$$

and then we have the smooth function f_μ

$$\begin{aligned} f_\mu(u_1, u_2, t) = {} & \delta_1\Psi_\mu(\tfrac{1}{2}t + A_1W_1u_1) + \delta_2\Psi_\mu(\tfrac{1}{2}t + A_2W_2u_2) \\ & + \gamma\Psi_\mu(t) - y^T(A_1W_1u_1) - y^T(A_2W_2u_2) . \end{aligned}$$

The error we make by substituting the 2-norm with the smooth approximation $\Psi_\mu(x)$ is less than

$$\max_{x\in\mathbb{R}^n} \|x\|_2 - \Psi_\mu(x) \leq \frac{\mu}{2} ,$$

and then we have

$$f(u_1, u_2, t) - f_\mu(u_1, u_2, t) \leq \frac{\mu}{2}(\delta_1 + \delta_2 + \gamma) .$$

In order to obtain an ϵ-solution $(u_1^\epsilon, u_2^\epsilon, t^\epsilon)$, *i.e.*,

$$f(u_1^\epsilon, u_2^\epsilon, t^\epsilon) - f(u_1^*, u_2^*, t^*) \leq \epsilon$$

we should not make a larger error on the function f than, say $\epsilon/10$, and we then simply select $\mu = \epsilon/(5(\delta_1+\delta_2+\gamma))$. In this particular case, the gradient with respect to, *e.g.*, u_1 is

$$\nabla_{u_1} f_\mu(u_1, u_2, t) = W_1A_1^T\left(\frac{\delta_1(A_1W_1u_1 + \frac{1}{2}t)}{\max(\|A_1W_1u_1 + \frac{1}{2}t\|_2, \mu)} - y\right) .$$

Without smoothing we would only have $A_1W_1u_1 + \frac{1}{2}t$ in the denominator and the gradient would be undefined at $A_1W_1u_1 + \frac{1}{2}t = 0$.

Now we have a problem on the form

$$\begin{array}{ll} \min & h(x) \\ \text{s.t.} & x \in Q \end{array} \tag{4}$$

where h is a smooth function and Q is a closed convex set. We then apply a gradient projection algorithm with Armijo rule along the projection arc [18] and Barziliai-Borwein strategy for the initial stepsize [25]. This algorithm is known to be practical efficient [21], [22], and outlined in Fig. 4.

We note that projection of u on the set $U = \{u \mid \|u\|_\infty \leq \rho\}$ is given by

$$P_U(u)_i = \begin{cases} \rho & \text{if } u_i \geq \rho \\ -\rho & \text{if } u_i \leq -\rho \\ u_i & \text{else} \end{cases} .$$

A. Obtaining the primal variables

We can now solve the dual problem and obtain an ϵ-solution of the dual variables. We will now show how we can obtain the primal variables, using an approach similar to [26, §5.5.5]. By considering the Lagrangian

$$\begin{aligned} & \mathcal{L}(z_1, z_2, b_1, b_2, b_c, \kappa_1\kappa_2, \kappa_c, t_1, t_2, t_c) = f(z_1, z_2) \\ & +\kappa_1(\|b_1\|_2 - \delta_1) + \kappa_2(\|b_2\|_2 - \delta_2) + \kappa_c(\|b_c\|_2 - \gamma) \\ & +t_1^T(A_1z_1 - y - b_1) + t_2^T(A_2z_2 - y - b_2) \\ & +t_c^T(\tfrac{1}{2}(A_1z_1 + A_2z_2) - y - b_c) \end{aligned}$$

given a feasible $x^{[0]}$

for $k \geq 0$

1. Evaluate $\nabla h(x^{[k]})$
2. Bactracking line search with

   $\beta, \sigma \in (0,1)$, $M \in \mathbb{N}_0$, $h_\mathrm{r} = \max_{i=k-M,\cdots,k} h(x^{[i]})$

   $\alpha = \frac{\|x^{[k]}-x^{[k-1]}\|_2^2}{\langle x^{[k]}-x^{[k-1]}, \nabla h(x^{[k]})-\nabla h(x^{[k-1]})\rangle}$

   $\bar{x} = P_Q(x^{[k]} - \alpha\beta\nabla h(x^{[k]}))$

   while $h(\bar{x}) \geq h_\mathrm{r} - \sigma \nabla h(x^{[k]})^T(x^{[k]} - \bar{x})$

   $\beta := \beta^2$

   $\bar{x} = P_Q(x^{[k]} - \alpha\beta\nabla h(x^{[k]}))$
3. Set $x^{[k+1]} := \bar{x}$

Fig. 4. Outline of a first-order optimization algorithm for problem (4). The algorithm uses gradient projection with Armijo rule along the projection arc [18] and Barziliai-Borwein strategy for the initial stepsize [25].

we get the first-order optimality conditions, where

$$t_1 = -\tfrac{1}{2}t - A_1W_1u_1$$

$$t_2 = -\tfrac{1}{2}t - A_2W_2u_2$$

$$t_\mathrm{c} = t$$

$$\kappa_1 = \|\tfrac{1}{2}t + A_1W_1u_1\|_2$$

$$\kappa_2 = \|\tfrac{1}{2}t + A_2W_2u_2\|_2$$

$$\kappa_\mathrm{c} = \|t\|_2$$

relates the dual variables in the Lagrangian to the dual problem (3). By solving for z_1, z_2 using $\nabla\mathcal{L} = 0$ and the complementary condition for the Lagrange multipliers $\kappa_1, \kappa_2, \kappa_\mathrm{c}$, we obtain different equations for each combination $\kappa_1 = 0$ or $\kappa_1 > 0$, $\kappa_2 = 0$ or $\kappa_2 > 0$, and $\kappa_\mathrm{c} = 0$ or $\kappa_\mathrm{c} > 0$. As an example, we have

$$\kappa_1 > 0, \kappa_2 > 0, \kappa_\mathrm{c} = 0:$$

$$\bar{z}_1 = A_1^T(-\delta_1 \tfrac{A_1W_1u_1}{\|A_1W_1u_1\|_2} + y) \tag{5}$$

$$\bar{z}_2 = A_2^T(-\delta_2 \tfrac{A_2W_2u_2}{\|A_2W_2u_2\|_2} + y)\,,$$

and

$$\kappa_1 > 0, \kappa_2 = 0, \kappa_\mathrm{c} > 0:$$

$$\bar{z}_1 = A_1^T(-\delta_1 \tfrac{+\frac{1}{2}t + A_1W_1u_1}{\|\frac{1}{2}t + A_1W_1u_1\|_2} + y) \tag{6}$$

$$\bar{z}_2 = A_2^T(\gamma \tfrac{t}{\|t\|_2} + 2y - A_1\bar{z}_1)\,.$$

In the algorithm we try out all combinations, and see which are feasible. For all primal feasible variables $(\bar{z}_1, \bar{z}_2)$ we then calculate the objective function $g(\bar{z}_1, \bar{z}_2) = \|\bar{z}_1\|_1 + \lambda\|\bar{z}_2\|_1$, and call the solution with the smallest objective for (z_1, z_2). We terminate the algorithm when we have a duality gap smaller than ϵ, *i.e.*,

$$g(z_1, z_2) + f(u_1, u_2, t) \leq \epsilon \tag{7}$$

and return z_1, z_2 (Note that $-f(u_1, u_2, t)$ is the dual objective).

There is a problem when $\kappa_1 = 0, \kappa_2 = 0, \kappa_\mathrm{c} > 0$, because the linear system of equations obtained from $\nabla\mathcal{L} = 0$ is no longer full rank, and the relation between the dual and primal variables is no longer unique. In this case, the only active constraint is $\|\frac{1}{2}(A_1z_1 + A_2z_2) - y\| \leq \gamma$ and the other two are inactive. To make one of the side constraints active, such that we will have the case (6), we slightly reduce $\mathrm{argmin}(\delta_1, \delta_2)$ and repeatedly run the algorithm again until one of the side constraints is active. This case with $\kappa_1 = 0, \kappa_2 = 0, \kappa_\mathrm{c} > 0$ seems rare, and we can only generate it for small examples. For the simulations in Sec. IV-A and IV-B we do not encounter this case.

IV. SIMULATIONS

We will now show the usage af the algorithm with simulations applied to a still image and an image sequence.

A. A STILL IMAGE EXAMPLE

We will first present an example of using MD ℓ_1-compression on a still image of the *"cameraman"*. We select A_1 as the Symlet16 standard discrete wavelet transform (DWT) with 3 levels and A_2 as the Symlet8 standard DWT with 4 levels. The distortions $\delta_1, \delta_2, \gamma$ are selected such that $PSNR_1 = PSNR_2 = 26\,\mathrm{dB}$, and $PSNR_\mathrm{c} = 38\,\mathrm{dB}$. We select the accuracy as $\epsilon = 2mn\,10^{-3}$ in (7), such that the average accuracy per variable is at least 10^{-3}. Fig 5 illustrates the reconstructed images resulting from the simulations.

Reweighting the ℓ_1-objective has proved a very efficient way to enhance sparsity of the solution [11]. For the example in Fig. 5, the number of non-zero coefficients (NNZC) used in each reweighting iteration are given in Table I. Notice that the sparsity is significantly imporved with only a few iterations of reweighting.

$\times 10^3$	0	1	2	3	4	5
NNZC for z_1	7.59	5.52	5.47	5.28	5.25	5.19
NNZC for z_2	7.15	4.82	4.81	4.71	4.72	4.69

TABLE I
NNZC FOR A DIFFERENT NUMBER OF REWEIGHTING STEPS. THE IMAGE HAS $256^2 = 65.536 \times 10^3$ PIXELS.

We could distribute the coefficients equally among the two channels by thresholding the smallest image coefficients to zero in the transform domains (A_1 and A_2). Let us use the $(5.19 + 4.67) \times 10^3$ largest coefficients, that is the same number of coefficients as with the MD ℓ_1-compression problem example, see Table I. Then we would obtain the distortions $PSNR_1 = 30.6\,\mathrm{dB}$, $PSNR_2 = 30.8\,\mathrm{dB}$ and $PSNR_\mathrm{c} = 32.0\,\mathrm{dB}$. These two descriptions does not refine each other very well

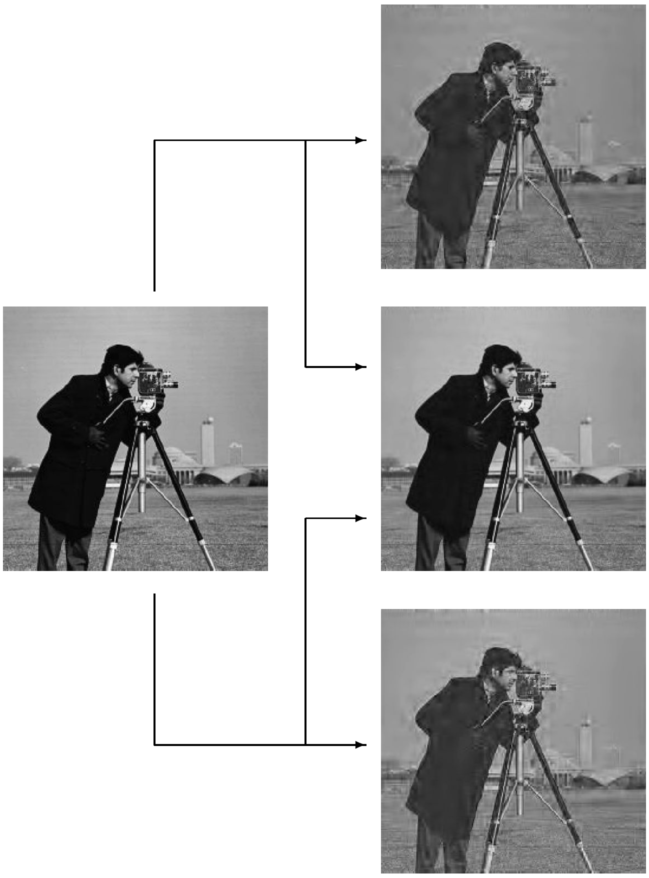

Fig. 5. An example of the MD ℓ_1-compression problem applied to the image *"cameraman"* (256×256). The distortions $\delta_1, \delta_2, \gamma$ are selected such that $PSNR_1 = PSNR_2 = 26\,\text{dB}$, and $PSNR_c = 38\,\text{dB}$.

because the obtained central distortion is not much better than the side distortions. In the MD ℓ_1-compression case the descriptions refine each other well, because the central distortion is much better than the side distortions.

B. AN IMAGE SEQUENCE EXAMPLE

We now apply the MD ℓ_1-compression algorithm on a 32 frame image sequence of *"foreman"* with image dimension 288×352. For A_1 we apply the 3 dimensional DCT and for A_2 the 3 dimensional discrete sine transformation (DST). In this framework we jointly encode multiple frames [5], 8 frames in this example. Hence, the image sequence is divided into blocks of dimension $288 \cdot 352 \cdot 8$, where each block is used as the input y to the MD ℓ_1-compression algorithm. We again use $PSNR_1 = PSNR_2 = 26\,\text{dB}$ and then select 6 different $PSNR_c = \{28, 30, 32, 34, 36, 38\}\,\text{dB}$ and apply 5 reweight iterations. In the introduction we discussed that ℓ_1-norm minimization is used as an approximation of minimizing the NNZC. We are hence interested in the relation between the NNZC and ℓ_1-norms of the solutions. Fig 6 shows the results of this example where we plot $\|z_i\|_1$ vs. NNZC for the two descriptions z_1, z_2.

From Fig. 6 we observe that the lower $\|z_i\|_1$ the lower NNZC and vise versa, an interesting observation when approximating NNZC with the convex ℓ_1-norm. When we increase

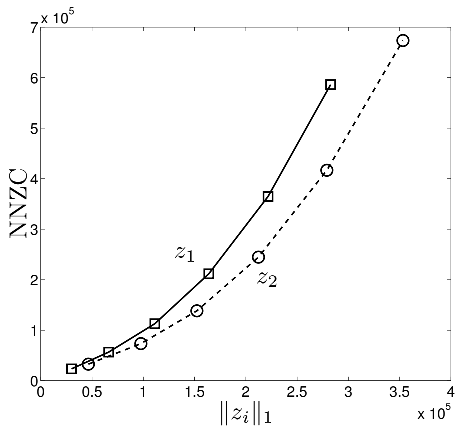

Fig. 6. Plot of NNZC vs. $\|z_i\|_1$ for a 32 frame video of *"foreman"* (288×352). There is in total $27 \cdot 10^5$ pixels in this image sequence. Each data point corresponds to a MD setup with $PSNR_c = \{28, 30, 32, 34, 36, 38\}\,\text{dB}$ and fixed $PSNR_1 = PSNR_2 = 26\,\text{dB}$.

$PSNR_c$ we observe that the NNZC increases faster than $\|z_i\|_1$ because at high PSNR the descriptions z_i contains many small coefficients which will increase the NNZC more than the ℓ_1-norm.

V. DISCUSSIONS

We have presented a problem we call multiple-description (MD) ℓ_1-compression, which combines the ℓ_1-norm approach for finding sparse solutions with constraints found in MD related problems. We demonstrated how to implement an efficient first-order optimization algorithm based on the dual-problem, which makes it possible to solve large scale problems.

An interesting feature with this example of MD is that each vector of coefficients z_1, z_2 are found in a single basis, well known in the image/video compression community. This makes it possible to apply state-of-the-art, of-the-shell coders straightforward on z_1 and z_2.

Generalization of the number of channels and decodings functions is a possible future direction. For example, the side channel decoding function could be with an overcomplete basis, *e.g.*, $g_1(z_1, z_2) = [A_1 A_2][z_1; z_2]$. For more than two channels there will be different joint decoding functions corresponding to each of the possible subsets on the receiver side. For the case with three channels, there are three possible joint decoding functions and one central decoding function.

REFERENCES

[1] R. G. Barniuk, E. Candés, R. Nowak, and M. Vetterli, Eds., *IEEE Sig. Proc. Mag.*, vol. 25, 2008.

[2] S. Mallat, *A Wavelet tour of signal processing*, Academic Press, 1999.

[3] M. Grote and T. Huckle, "Parallel preconditioning with sparse approximate inverses," *SIAM J. Sci. Computing*, vol. 18, no. 3, pp. 838–853, 1997.

[4] D. Giacobello, M. G. Christensen, J. Dahl, S. H. Jensen, and M. Moonen, "Sparse linear predictors for speech processing," in *Proc. 9th INTERSPEECH Conf*, Brisbane, Australia, Sep. 2008.

[5] J. Dahl, J. Østergaard, T. L. Jensen, and S. H. Jensen, "ℓ_1 compression of image sequences using the structural similarity index measure," accepted to the IEEE Data Comp. Conf. (DCC), 2009.

[6] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 851 – 857, Nov. 1982.

[7] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell System Technical Journal*, vol. 59, pp. 1909 – 1921, Dec. 1980.

[8] 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.

[9] J. Østergaard and R. Zamir, "Multiple description coding by dithered delta-sigma quantization," *Submitted to IEEE Trans. Inf. Theory*, 2007.

[10] 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.

[11] E. Candés, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," arXiv:0711.1612v1, 2007.

[12] Z. He 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.

[13] E. J. Candés and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21 – 30, Mar. 2008.

[14] 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 Proc. (ICASSP)*, Hong Kong, Apr. 2003, pp. 425–8.

[15] T. Nguyen and A. Zakhor, "Matching pursuits based multiple description video coding for lossy environments," in *Proc. Int. Conf. of Image Proc. (ICIP)*, Sep. 2003, pp. 57–60.

[16] P. A. Chou, S. Mehrotra, and A. Wang, "Multiple description decoding of overcomplete expansions using projections onto convex sets," in *Proc. IEEE Data Comp. Conf. (DCC)*, Snowbird, Utah, Mar. 1999, pp. 72 – 81.

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

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

[19] Yu. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer Academic Publishers, 2004.

[20] Yu. Nesterov, "Smooth minimization of nonsmooth functions," *Math. Prog. Series A*, vol. 103, pp. 127–152, 2005.

[21] 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.

[22] M. Zhu, S. J. Wright, and T. F. Chan, "Duality-based algorithms for total-variation regularized image restoraration," Computational and Applied Mathematics Report 08-33, UCLA, Oct. 2008.

[23] J. Dahl, P. C. Hansen, S. H. Jensen, and T. L. Jensen, "Algorithms and software for total variation image reconstruction via first-order methods," submitted to Numerical Algorithms, 2008.

[24] J. Dahl, J. Østergaard, T. L. Jensen, and S. H. Jensen, "An efficient first-order method for ℓ_1 compression of images," accepted to the IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP), 2009.

[25] J. Barzilai and J.M. Borwein, "Two point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, pp. 141–148, 1988.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.