MULTIPLE DESCRIPTIONS USING SPARSE DECOMPOSITIONS

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

*Aalborg University
Department of Electronic Systems
Niels Jernesvej 12, 9220 Aalborg
Denmark
{tlj,jo,shj}@.es.aau.dk

**Mosek ApS
Fruebjergvej 3, Box 16
2100 Copenhagen O
Denmark
dahl.joachim@gmail.com

ABSTRACT

In this paper, we consider the design of multiple descriptions (MDs) using sparse decompositions. In a description erasure channel only a subset of the transmitted descriptions is received. The MD problem concerns the design of the descriptions such that they individually approximate the source and furthermore are able to refine each other. In this paper, we form descriptions using convex optimization with l_1-norm minimization and Euclidean distortion constraints on the reconstructions and show that with this method we can obtain non-trivial descriptions. We give an algorithm based on recently developed first-order method to the proposed convex problem such that we can solve large-scale instances for image sequences.

1. INTRODUCTION

Sparse decomposition is an important method in modern signal processing and have been applied to different application such as estimation and coding [1], linear prediction [2] and blind source separation [3]. For estimation and encoding the argument for sparse approaches has been to follow natural statistics, see *e.g.*, [4]. The advent of compressed sensing [5, 6] have further added to the interest in sparse decompositions since the recovery of the latent variables requires a sparse acquisition method.

One method to acquire a sparse decomposition with a dictionary is to solve a convex relaxation of the minimum cardinality problem, that is the l_1-*compression problem*

$$\begin{array}{ll} \text{min.} & \|z\|_1 \\ \text{s.t.} & \|Dz - y\|_2 \le \delta\,, \end{array} \tag{1}$$

where $D \in \mathbb{R}^{M\times N}$ is an overcomplete dictionary, $\delta > 0$ is a selected reconstruction error level (distortion), $(N \ge M)$, $z \in \mathbb{R}^N$ is the latent variable and $y \in \mathbb{R}^M$ is the signal we wish to decompose into a sparse representation. There are several other sparse acquisition methods, including approximations of minimum cardinality and pursuit methods.

In this paper, we apply sparse decomposition to the *multiple-description* (MD) problem [7]. The MD problem is on encoding a source into multiple descriptions and each description is then transmitted over a different channel. Unknown to the encoder, a channel may break which correspond to a description erasure such that only a subset of the transmitted descriptions is received. The problem is to design the descriptions such that the decoded descriptions approximate the source for all possible subsets of descriptions.

The 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.

An important concept for the MD problem is the trade-off associated with the description design; in order for the descriptions to approximate the source, they should be similar to the source, and consequently the descriptions need to be similar to each other. But, if the descriptions are too similar to each other, it is not possible to obtain any refinement when the individual descriptions are combined.

Let J be the number of channels and let $\mathcal{J}_J = \{1, \dots, J\}$. Then $\mathcal{I}_J = \{\ell \,|\, \ell \subseteq \mathcal{J}_J,\ \ell \neq \emptyset\}$ describes the non-trivial subsets of descriptions which can be received. Further, let z_j, $\forall j \in \mathcal{J}_J$, denote the jth description and define $z_\ell = \{z_j \mid j \in \ell\}$, $\forall \ell \in \mathcal{I}_J$. At the decoder, the descriptions z_ℓ, $\ell \in \mathcal{I}_J$, are used to reconstruct an approximation of the source y via the reconstruction functions $g_\ell(z_\ell)$. The approximations satisfy the distortion constraint $d(g_\ell(z_\ell), y) \le \delta_\ell$, $\forall \ell \in \mathcal{I}_J$, with $d(\cdot,\cdot)$ denoting a distortion measure. An example with $J = 2$ is presented in Fig. 1.

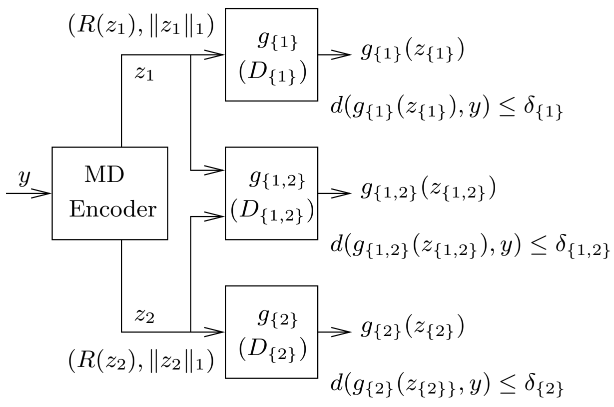

Figure 1: The MD (l_1-compression) problem for $J = 2$.

In a statistical setting, the MD problem is to design the descriptions z_j, $\forall j \in \mathcal{J}_J$, such that the total rate $\sum_{j\in\mathcal{J}_J} R(z_j)$ is minimized and the fidelity constraints are satisfied. This problem is only completely solved with the squared error fidelity criterion, a memoryless Gaussian sources and two descriptions [8]. Another direction is to form descriptions in a deterministic setting. Algorithms specifically designed for video or image coding may be based on, *e.g.*, Wiener filters with prediction compensation [9], matching pursuit [10, 11] or compressed sensing [12, 13].

The remaining part of the paper is organized as follows: in Sec. 2 we present a method to obtain sparse decomposition using convex optimization with constraints on the distortion. Sec. 3 is on a first-order method for solving the proposed convex problem. We provide simulations in Sec. 4 and discussions in Sec. 5.

 ISSN 2076-1465

2. CONVEX RELAXATION

In this work, we cast the MD problem into a similar form as (1).[1] Let $z_j \in \mathbb{R}^{M\times 1}$, $\forall j \in \mathcal{J}_J$, be the descriptions and $z_\ell = \bigcup_{j\in\ell} z_j \in \mathbb{R}^{|\ell|M\times 1}$, $\forall \ell \in \mathcal{I}_J$, be the vector concatenation of the descriptions used in the decoding when the subset $\ell \subseteq \mathcal{J}_J$ is received. We then form the linear reconstruction functions $g_\ell(z_\ell) = D_\ell z_\ell$, $\forall \ell \in \mathcal{I}_J$, see also [12]. The dictionaries are given as $D_\ell = \bigcup_{j\in\ell} \bar{D}_{\ell,j}$ with $D_\ell \in \mathbb{R}^{M\times|\ell|M}$, $\forall \ell \in \mathcal{I}_J$, and $\bar{D}_{\ell,j} = \rho_{\ell,j} D_j$, $\forall \ell \in \mathcal{I}_J, j \in \ell$. We choose:

- the reconstruction weight
$$\rho_{\ell,j} = \begin{cases} 1 & \text{if} \quad |\ell| = 1 \\ \frac{\sum_{i\in\ell\backslash j}\delta_i^2}{(\|\ell\|-1)\sum_{i\in\ell}\delta_i^2}, & \text{otherwise} \end{cases},$$
in order to weight the joint reconstruction relative to the distortion bound on of the individual distortions, see [15],
- D_j, $\forall j \in \mathcal{J}_J$ invertible, the reason for such will become clear in Sec. 3,
- the Euclidean norm as the measure $d(x, y) = \|x - y\|_2$.

With these choices we obtain the *standard multiple-description l_1-compression* (SMDL1C) problem

$$\begin{aligned} \text{min.} \quad & \sum_{j\in\mathcal{J}_J} \lambda_j \|W_j z_j\|_1 \\ \text{s.t.} \quad & \|D_\ell z_\ell - y\|_2 \le \delta_\ell, \qquad \forall \ell \in \mathcal{I}_J, \end{aligned} \tag{2}$$

for $\delta_\ell > 0$, $\forall \ell \in \mathcal{I}_J$, and $\lambda_j > 0, W_j \succ 0$, $\forall j \in \mathcal{J}_J$. The problem (2) is a second-order cone program (SOCP) [16].

For Gaussian sources with the Euclidean fidelity criterion, it has been shown that linear reconstruction functions are sufficient for achieving the MD rate-distortion function, see [17, 18] and [19] for white and colored Gaussian sources, respectively.

In (2) we have introduced $W_j = \mathbf{diag}(w_j)$, $\forall j \in \mathcal{I}_J$, to balance the cost of the coefficients with small and large magnitude [20]. To find w_j, the problem (2) is first solved with $w_j = \mathbb{1}$. Then w_j is chosen approximately inversely proportional to the solution z_j^* of that problem, $w_j(i) \leftarrow 1/(|z_j^*(i)|+\tau)$, for the ith coordinate and with a small $\tau > 0$. The problem (2) is then resolved with the new weighting w_j. This reweighting scheme can be iterated a number of times. The parameter λ_j in (2) allows weighting of the l_1-norms in order to achieve a desired ratio $\frac{\|W_j z_j\|_1}{\|W_{j'} z_{j'}\|_1}$, $\forall j, j' \in \mathcal{J}_J$.

For the SMDL1C problem there is always a solution. Since $D_j z_j = y$ has a solution then $D_\ell z_\ell = \sum_{j\in\ell} \bar{D}_{\ell,j} z_j = \sum_{j\in\ell} \rho_{\ell,j} D_j z_j = y \sum_{j\in\ell} \rho_{\ell,j} = y$, $\forall \ell \in \mathcal{I}_J$. This implies that there exist a strictly feasible point z with $\|D_\ell z_\ell - y\|_2 = 0 < \delta_\ell$, $\forall \ell \in \mathcal{I}_J$, such that Slater's condition for strong duality holds [16].

3. A FIRST-ORDER METHOD

We are interested in solving the SMDL1C problem for image sequences, that is, large instances involving more than 10^6 variables. First-order methods have proved efficient for large scale problems [21–23]. However, such methods requires projection onto the feasible set, which might prove inefficient because the projection on a set of coupled constraints requires yet another iterative method such as alternating projection. Also, if we apply alternating projection then we will only obtain sub-optimal projection which might generate irregularity in the first-order master method.

A problem with coupled constraints is when variable components are coupled in different constraints. To exemplify the coupled constraints, note that in the case where we let $J = 2$, we see that the constraints for $\ell = 1$ or $\ell = 2$ can easily be fulfilled by simply thresholding the smallest coefficients to zero in the transform domain D_ℓ independently. This will, however, not guarantee the joint reconstruction constraint $\|D_{\{1,2\}} z_{\{1,2\}} - y\|_2 \le \delta_{\{1,2\}}$ which then corresponds to the coupling of the variables z_1 and z_2.

3.1 Dual Decomposition

Dual decomposition is a method to decompose coupled constraints if the objective function is decoupled [24,25]. A dual problem of (2) can be represented as

$$\begin{aligned} \text{max.} \quad & -\sum_{\ell\in\mathcal{I}_J} \left(\delta_\ell \|t_\ell\|_2 + y^T t_\ell\right) \\ \text{s.t.} \quad & \|u_j\|_\infty \le \lambda_j, \forall j \in \mathcal{J}_J, \; t_\ell \in \mathbb{R}^{M\times 1}, \forall \ell \in \mathcal{I}_J \backslash \mathcal{J}_J \\ & t_j = -\left(D_j^{-T} W_j u_j + \sum_{\ell\in c_j(\mathcal{I}_J)\backslash j} D_j^{-T} \bar{D}_{\ell,j}^T t_\ell\right), \forall j \in \mathcal{J}_J, \end{aligned} \tag{3}$$

with optimal objective g^* and

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

The equality constraints in (3) are simple because t_ℓ, $\forall \ell \in \mathcal{J}_J$, are isolated on the left hand side, while the remaining variables t_ℓ, $\forall \ell \in \mathcal{I}_\mathcal{J} \backslash \mathcal{J}_\mathcal{J}$, are on the right side. We could then make a variable substitution of t_ℓ, $\forall \ell \in \mathcal{J}_J$, in the objective function. However, we choose the form (3) for clarity. The problem (3) is then decoupled in the constraints but coupled in the objective function which makes the problem (3) appropriate for first-order methods. Note that if the dictionaries D_j, $\forall j \in \mathcal{I}_J$ are not invertible we could not easily make a variable substitution and instead needed to handle the vector equality involving the matrix dictionaries explicitly. Indeed a difficult problem for large scale MD instances.

3.2 Primal recovery

Recovery of optimal primal variables from optimal dual variables can be accomplished if there is a unique solution to the minimization of the Lagrangian, usually in the case of a strictly convex Lagrangian [16, §5.5.5]. Define the primal variables $h_\ell = D_\ell z_\ell - y$, $\forall \ell \in \mathcal{I}_J$, and $x_j = W_j z_j$, $\forall j \in \mathcal{J}_J$, and the Lagrangian at optimal dual variables is then given as

$$\begin{aligned} \mathcal{L}(z, x, h, t^*, u^*, \kappa^*) = & \sum_{j\in\mathcal{J}_J} \lambda_j \|x_j\|_1 + \sum_{\ell\in\mathcal{I}_J} \kappa_\ell^* (\|h_\ell\|_2 - \delta_\ell) \\ & + \sum_{\ell\in\mathcal{I}_J} t_\ell^{*T} (D_\ell z_\ell - y - h_\ell) + \sum_{j\in\mathcal{J}_J} u_j^{*T} (W_j z_j - x_j). \end{aligned}$$

However the Lagrangian associated to the problem is not strictly convex in x due to the $\|\cdot\|_1$-norm. Instead, lets consider the Karush-Kuhn-Tucker (KKT) conditions for the sub-differentiable problem (2) given as

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

with $h_a(x) = \partial\|x\|_a$. We can rewrite the above system using $\delta_\ell > 0$, $\forall \ell \in \mathcal{I}_J$ and obtain the equivalent KKT optimality conditions

$$\begin{cases} \sum_{\ell\in c_j(\mathcal{I}_J)} \frac{\|t_\ell^*\|_2}{\delta_\ell} \bar{D}_{\ell,j}^T D_\ell z_\ell^* = r_j, \quad \forall j \in \mathcal{J}_J \quad (4.\triangle) \\ \|D_\ell z_\ell^* - y\|_2 \le \delta_\ell, \quad \forall \ell \in \mathcal{I}_J \\ \lambda_j h_1(W_j z_j^*) - u_j^* \ni 0, \quad \forall j \in \mathcal{J}_J \end{cases} \tag{4}$$

[1]This work was presented in part for the case of $J = 2$ in [14].

where

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

The equations (4.△) can solved with low complexity for invertible dictionaries. However, the remaining equations are sub-differentiable and feasibility equations and are too difficulty to handle. Especially for large scale problems. Also, for a sub-optimal dual solution it is not possible to find a primal solutions that fulfills (4), because this implies that the dual solution is in fact an optimal dual solution. That is, for a sub-optimal dual solution we can only solve a subset of the KKT system.

Let $z^* \in \mathcal{Z}$ be a solution to (4) and let $\bar{z} \in \bar{\mathcal{Z}}$ be a solution to the square system (4.△). Then the following proposition shows that it is in fact possible to recover optimal primal variables in certain cases.

Proposition 3.1. *(Uniqueness) If the solution $\bar{z}$ to the linear system (4.△) is unique, then $z^* = \bar{z}$ for the SMDL1C problem.*

Proof. Since the SMDL1C problem has a solution and the system (4.△) is a subsystem of (4) then $\emptyset \neq \mathcal{Z} \subseteq \bar{\mathcal{Z}}$. If $|\bar{\mathcal{Z}}| = 1$ then $|\mathcal{Z}| = 1$ such that $\bar{z} = z^*$. □

In the first-order method, from the dual sub-optimal iterates $\left(t^{(i)}, u^{(i)}\right)$, the primal iterate $z^{(i)}$ is obtained as the solution to

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

The algorithm is halted if it is a primal-dual ϵ-solution

$$f(z^{(i)}) - g(t^{(i)}) \leq \epsilon, \quad z^{(i)} \in Q_{\mathrm{p}}, \; \left(t^{(i)}, u^{(i)}\right) \in Q_{\mathrm{d}}\,,$$

where Q_{p} and Q_{d} defines the primal and dual feasible set, respectively. We select $\epsilon = MJ\epsilon_{\mathrm{r}}$ with $\epsilon_{\mathrm{r}} = 10^{-3}$ to scale the accuracy ϵ in the dimensionality of the primal variables.

3.3 Complexity

The objective of the dual problem (3) is differentiable on $\|t_\ell\| > 0$ and sub-differentiable on $\|t_\ell\|_2 = 0$. The objective in the dual problem (3) is hence not smooth. A smooth function is a function with Lipschitz continuous derivatives [26]. We could then apply an algorithm such as the sub-gradient algorithm with complexity $\mathcal{O}(1/\epsilon^2)$ where ϵ is the accuracy in function value. However, it was recently proposed to make a smooth approximation and apply an optimal first-order method to the smooth problem and obtain complexity $\mathcal{O}(\frac{1}{\epsilon})$ [27]. We can not efficiently apply the algorithm in [27], since this requires projections on both the primal and dual feasible set. We will instead show how to adapt the results of [27], similar to [28], using only projection on the dual set and still achieve complexity $\mathcal{O}(\frac{1}{\epsilon})$. Consider

$$\|x\|_2 = \max_{\|v\|_2 \leq 1} \left\{ v^T x \right\}$$

and the approximation

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

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 the smooth function g_μ

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

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

$$L_\mu = L(\nabla g_\mu) = \left(\sum_{\ell \in \mathcal{I}_J} \frac{\delta_\ell}{\mu} + 1 \right) = \frac{C}{\mu} + |\mathcal{I}_J| \,. \tag{5}$$

The smooth function has the approximation

$$g_\mu(t) \leq g(t) \leq g_\mu(t) + \mu C \,. \tag{6}$$

Hence, the parameter μ both controls the level of smoothness (5) and the approximation accuracy (6). Select $\mu = \epsilon/(2C)$ and let the ith iteration $t^{(i)}$ of a first-order method have the property

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

where g_μ^* is the optimal objective for the smooth problem. Then we obtain

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

By using an optimal-first order algorithm for L-smooth problems with complexity $\mathcal{O}\left(\sqrt{\frac{L}{\epsilon}}\right)$ [26], then $t^{(i)}$ can be obtained in i iterations, where

$$i = \mathcal{O}\left(\sqrt{\frac{L_\mu}{\epsilon}}\right) = \mathcal{O}\left(\sqrt{\frac{1}{\epsilon^2} + \frac{1}{\epsilon}}\right) \leq \mathcal{O}\left(\sqrt{\frac{1}{\epsilon^2}} + \sqrt{\frac{1}{\epsilon}}\right) = \mathcal{O}\left(\frac{1}{\epsilon}\right) \,.$$

4. SIMULATIONS

For the simulations we will present an example of obtaining a sparse decomposition in the presented MD framework. As the source we select the grayscale image sequence of "foreman" with height $m = 288$ pixels and width $n = 352$ pixels. We jointly process $k = 8$ consecutive frames [29] and y is formed by stacking each image and scaled such that $y \in [0;1]^M$, $M = mnk$. We select $J = 3$ and as dictionaries

- D_1 : the three dimensional cosine transform,
- D_2 : a two dimensional Symlet16 discrete wavelet transform with 5 levels along the dimensions associated to m,n and a one dimensional Haar discrete wavelet transform with 3 levels along the dimension associated to k,
- D_3 : the three dimensional sine transform.

Let the peak signal-to-noise ratio (PSNR) measure be defined by

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

As distortion constraints we select $\mathrm{PSNR}(\delta_\ell) = 30$, $\forall |\ell| = 1$, $\mathrm{PSNR}(\delta_\ell) = 33$, $\forall |\ell| = 2$ and $\mathrm{PSNR}(\delta_\ell) = 37$, $|\ell| = 3$ with $\ell \in \mathcal{I}_J$. Further we choose equal weights $\lambda_j = 1$, $\forall j \in \mathcal{J}_J$.

If the primal variables where obtained from an algorithm using projection [21] or a method employing a soft-thresholding operator [22], a sub-optimal solution will contain coefficients which are exactly zero. The primal variables are in this approach obtained as the solution to a linear system arising from sub-optimal dual variables and hence there might be many small coefficients which are not exactly zero. To handle this, the distortion requirement are changed by $\bar{\delta}_\ell = \delta_\ell - |\ell|\sigma, \forall \ell \in \mathcal{I}_J$, with $\sigma > 0$ when the SMDL1C problem is solved and the smallest coefficients are afterwards thresholded to zero using the slack introduced by $|\ell|\sigma$ while ensuring the original distortion constraints δ_ℓ. Let $z(r)$ be an ϵ-solution after r reweight iterations of the SMDL1C problem and set $\hat{z} = z(7)$.

4.1 Example

Define a frame extraction function $s(y, i)$ which extract the ith frame from the image sequence stacked in y. In Fig. 2 we show a few examples of the decoded 6th frame for the subset $\ell = \{1\}$, $\ell = \{2, 3\}$ and $\ell = \{1, 2, 3\}$. This example is a large scale problem with $10 \cdot 10^6$ primal-dual variables.

4.2 Reweighting

In Fig. 3 we report the relative cardinality of $z(r)$ as a function of the number of applied reweight iterations r. We observe in Fig. 3 that the cardinality is significantly decreased for $r \in \{0, \cdots, 3\}$, whereupon the decrease is less distinct.

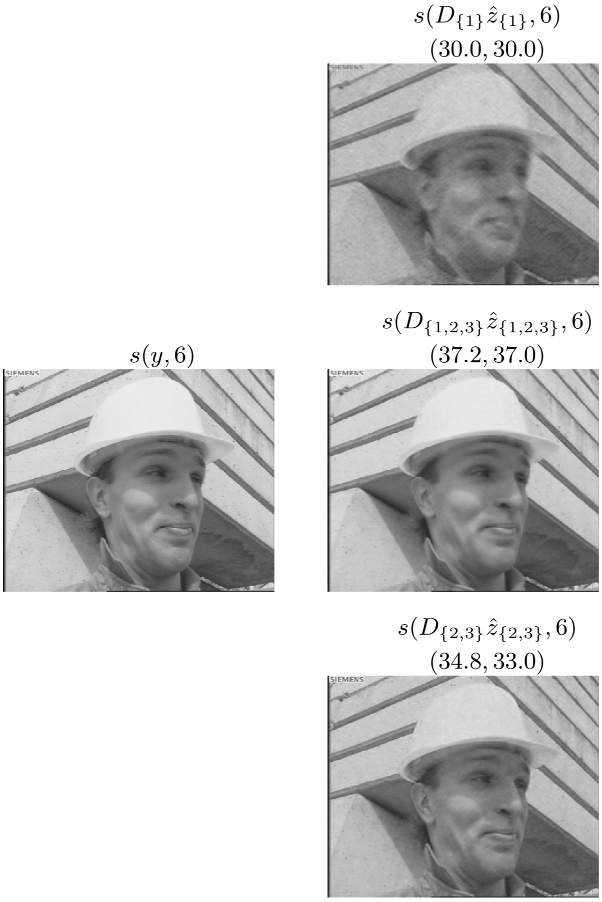

Figure 2: Example using "foreman" (grayscale, 288×352). The images show the 6th frame of the decoded images for $\ell = \{1\}$, $\ell = \{2, 3\}$ and $\ell = \{1, 2, 3\}$. Above the figures are the actual distortion and the distortion bounds reported using the format $(\text{PSNR}(\|D_\ell \hat{z}_\ell - y\|_2), \text{PSNR}(\delta_\ell))$.

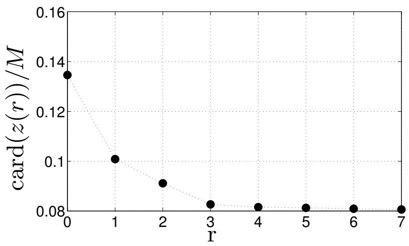

Figure 3: Example of reweighting an image sequence of "foreman" (grayscale, 288×352) by jointly processing $k = 8$ frames.

4.3 Threshold Comparison

For comparison we will obtain sparse decompositions in each basis independently using thresholding. We define the operation $z = T(D, y, \gamma)$ as thresholding the coefficients with the smallest magnitude in the basis D from the source y such that $\text{PSNR}(\|Dz - y\|_2) \approx \gamma$. We report the relative cardinalities card$(z)/M$ and PSNR measures obtained by SMDL1C in Tab. 1 and by independent thresholding for each basis in Tab. 2. When using SMDL1C we observe that from $|\ell| = 1$ to $|\ell| = 2$ the descriptions obtain a refinement in the range $3.2 - 4.8$ dB. For independent thresholding the refinement is smaller, in the range $0.7 - 1.2$ dB. This show that the obtained refinement by the SMDL1C method is non-trivial.

The PSNR measures for thresholding to same cardinality as using SMDL1C are reported in Tab 3. The descriptions are formed independently and refinement is there not guaranteed, which we can observe when the reconstructions at level $|\ell| = 2$ are combined to reconstruction at level $|\ell| = 3$.

card$(z_j)/M$	PSNR$(\|D_\ell z_\ell - y\|_2)$		
$j = 1$	$\ell = \{1\}$	$\ell = \{1, 2\}$	
0.019	30.0	34.4	
$j = 2$	$\ell = \{2\}$	$\ell = \{1, 3\}$	$\ell = \{1, 2, 3\}$
0.025	30.0	33.2	37.2
$j = 3$	$\ell = \{3\}$	$\ell = \{2, 3\}$	
0.037	30.0	34.8	

Table 1: Cardinality and reconstruction PSNR for SMDL1C $(z \leftarrow \hat{z})$, with card$(z)/M = 0.081$.

card$(z_j)/M$	PSNR$(\|D_\ell z_\ell - y\|_2)$		
$j = 1$	$\ell = \{1\}$	$\ell = \{1, 2\}$	
0.012	30.0	30.9	
$j = 2$	$\ell = \{2\}$	$\ell = \{1, 3\}$	$\ell = \{1, 2, 3\}$
0.006	30.0	30.7	31.3
$j = 3$	$\ell = \{3\}$	$\ell = \{2, 3\}$	
0.019	30.0	31.2	

Table 2: Cardinalities and reconstruction PSNRs for thresholding $(z \leftarrow \bigcup_{j \in \mathcal{J}_J} T(D_j, y, 30))$, with card$(z)/M = 0.037$.

card$(z_j)/M$	PSNR$(\|D_\ell z_\ell - y\|_2)$		
$j = 1$	$\ell = \{1\}$	$\ell = \{1, 2\}$	
0.019	31.2	34.8	
$j = 2$	$\ell = \{2\}$	$\ell = \{1, 3\}$	$\ell = \{1, 2, 3\}$
0.025	34.7	32.5	34.5
$j = 3$	$\ell = \{3\}$	$\ell = \{2, 3\}$	
0.037	32.1	35.1	

Table 3: Cardinalities and reconstruction PSNRs for thresholding to same cardinality as using SMDL1C.

The cardinalities for thresholding at PSNR 37.2 dB are given in Tab. 4. By comparing Tab. 1 and 4, we see that the cardinalities of SMDL1C are smaller than that of simple tresholding at 37.2 dB. Also, by comparing Tab. 1 and 2, we see that the cardinalities of SMDL1C are larger than that of simple tresholding at 30.0 dB. These bounds are to be expected for non-trivial descriptions. We also note that if we used the dictionary with the smallest cardinality to achieve the requested PSNR ($j = 2$), it is not possible to duplicate this description at the highest PSNR before the total cardinality exceeds that of $\hat{z}$. This exemplifies that it is not possible to simply transmit the coefficients $T(D_2, y, 37.2)$ over all channels and obtain a comparable cardinality as obtained by the SMDL1C problem.

	$j = 1$	$j = 2$	$j = 3$
$\text{card}(z_j)/M$	0.107	0.048	0.122

Table 4: Cardinalities using thresholding at the highest obtained PSNR by SMDL1C ($z \leftarrow \bigcup_{j \in \mathcal{J}_J} T(D_j, y, 37.2)$).

5. DISCUSSIONS

We presented a multiple description formulation using convex relaxation. In the case of large-scale problems we have proposed a first-order method for the dual problem. The simulations showed that the proposed multiple description formulation renders non-trivial descriptions with respect to both the cardinality and the refinement.

REFERENCES

[1] S. Mallat, *A Wavelet tour of signal processing, Third Edition: The Sparse Way.* Academic Press, 2009.

[2] 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, pp. 1353–1356.

[3] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Comp.*, vol. 13, no. 4, pp. 863–882, 2001.

[4] B. Olshausen and D. Field, "Natural image statistics and efficient coding," *Network: Computation in Neural Systems*, vol. 7, no. 2, pp. 333–339, 1996.

[5] 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, 2006.

[6] D. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[7] 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.

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

[9] U. G. Sun, J. Liang, C. Tian, C. Tu, and T. Tran, "Multiple desciption coding with prediction compensation," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 1037–1047, 2009.

[10] 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.

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

[12] 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.

[13] 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.

[14] T. L. Jensen, J. Dahl, J. Østergaard, and S. H. Jensen, "A first-order method for the multiple-description l1-compression problem," Signal Processing with Adaptive Sparse Structured Representations (SPARS'09), Saint-Malo, France, Apr. 2009.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] E. Candés, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

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

[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," *Numer. Algo.*, vol. 50, no. 1, pp. 67–92, 2010.

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

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

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

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

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

[29] J. Dahl, J. Østergaard, T. L. Jensen, and S. H. Jensen, "ℓ_1 compression of image sequences using the structural similarity index measure," in *Proc. IEEE Data Comp. Conf. (DCC)*, Snowbird, Utah, Mar. 2009, pp. 133 – 142.