

# Sparse Recovery and Dictionary Learning from Nonlinear Compressive Measurements

Lucas Rencker, Francis Bach, Wenwu Wang, Mark D. Plumbley

**Abstract**—Sparse coding and dictionary learning are popular techniques for linear inverse problems such as denoising or inpainting. However in many cases, the measurement process is nonlinear, for example for clipped, quantized or 1-bit measurements. These problems have often been addressed by solving constrained sparse coding problems, which can be difficult to solve, and assuming that the sparsifying dictionary is known and fixed. Here we propose a simple and unified framework to deal with nonlinear measurements. We propose a cost function that minimizes the distance to a convex feasibility set, which models our knowledge about the nonlinear measurement. This provides an unconstrained, convex, and differentiable cost function that is simple to optimize, and generalizes the linear least squares cost commonly used in sparse coding. We then propose proximal based sparse coding and dictionary learning algorithms, that are able to learn directly from nonlinearly corrupted signals. We show how the proposed framework and algorithms can be applied to clipped, quantized and 1-bit data.

**Index Terms**—Sparse coding, dictionary learning, nonlinear measurements, saturation, quantization, 1-bit sensing

## I. INTRODUCTION

Sparse decomposition and dictionary learning are popular techniques for linear inverse problems in signal processing, such as denoising [1], [2], inpainting [3], [4] or super-resolution [5], [6]. Sparse coding aims at finding a sparse set of coefficients  $\alpha \in \mathbb{R}^M$  that accurately represents a signal  $\mathbf{x} \in \mathbb{R}^N$  from a fixed overcomplete dictionary  $\mathbf{D} \in \mathbb{R}^{N \times M}$ , and is often formulated as:

$$\min_{\alpha} \|\mathbf{x} - \mathbf{D}\alpha\|_2^2 + \lambda\Psi(\alpha), \quad (1)$$

where  $\Psi(\cdot)$  is a sparsity inducing regularizer, such as the  $\ell_0$  pseudo-norm or the  $\ell_1$ -norm. Dictionary learning on the other hand, jointly learns the dictionary  $\mathbf{D}$  and sparse coefficients  $\alpha_t$  from a set of training signals  $\{\mathbf{x}_t\}_{t=1,\dots,T}$ :

$$\min_{\mathbf{D}, \alpha_t} \sum_{t=1}^T [\|\mathbf{x}_t - \mathbf{D}\alpha_t\|_2^2 + \lambda\Psi(\alpha_t)]. \quad (2)$$

However, the observed signals are often distorted or measured in a nonlinear way:

$$\mathbf{y} = f(\mathbf{x}), \quad (3)$$

where  $f$  is a nonlinear measurement function, and  $\mathbf{x}$  is the original (unknown) clean signal. Examples of nonlinear

distortions include *clipping* (or saturation) and *quantization*. Clipping is often due to dynamic range limitations in acquisition systems, when a signal reaches a maximum allowed amplitude, and the waveform is truncated above that threshold [7]–[14]. Quantization is a common process in analog-to-digital conversion that maps a signal from a continuous input space to a (finite) discrete space [15]. More recently, 1-bit compression has attracted a lot of interest, as an extreme quantization scheme where samples are coded using only one bit per sample [16], i.e. only measuring the signs of the signal. Clipping and quantization are non-linear, non-smooth, and *compressive* measurements, i.e. the measurement map is non-invertible. For these reasons, the recovery of clipped/quantized signals is a challenging problem.

Recovering a signal from clipped or quantized measurements can be treated as linear inverse problems, by simply ignoring the nonlinearities, i.e. treating clipped samples as missing [8], [17] and quantization error as additive noise [18]. Similarly, 1-bit signals can be tackled by using the sign measurements directly as an input [16], [19]. However using a formulation that is *consistent* with the measurement process, i.e. that takes into account our knowledge about the nonlinear measurement function, has been shown to greatly improve the reconstruction [4], [7]–[14], [16], [19]–[25]. Specially tailored cost functions, constraints, or regularizers have independently been proposed to deal with clipped [4], [7]–[14], quantized [20]–[24] or 1-bit [16], [19], [25] measurements. These formulations often involve solving constrained sparse coding problems, which can be difficult and computationally expensive to solve, since they involve computing expensive non-orthogonal projections at each iteration [13], [24].

Reconstruction methods proposed in the literature assume that the signal is sparse in some orthogonal basis [16], [19]–[23], [23]–[26], or in a fixed dictionary [4], [9], [10], [13]. However it has been shown in a range of applications (when the measurements are linear), that *learning* the dictionary from the observed data greatly improves the reconstruction compared to using a fixed dictionary [1], [2], [27], [28].

## A. Contributions, and main results

Our contributions are as follows:

- We propose a unifying framework for signal recovery from nonlinear measurements such as clipping, quantization and 1-bit measurements, i.e. addressing these three problems in a unified fashion rather than individually. More specifically, we show how these problems can be formulated as minimizing the distance to a convex

L. Rencker, W. Wang and M. D. Plumbley are with the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford, UK.  
F. Bach is with the INRIA-Sierra project-team, Paris, France.

The research leading to these results has received funding from the European Union’s H2020 Framework Programme (H2020-MSCA-ITN-2014) under grant agreement no 642685 MacSeNet.

feasibility set, which models our assumption about the nonlinear measurement process. In particular, the proposed cost generalizes the linear least squares commonly used in sparse coding, as well as several cost functions proposed independently for declipping and 1-bit recovery.

- Using properties of projection operators over convex sets, we show that the proposed cost function is continuous, convex and differentiable with Lipschitz gradient. Our main result uses Danskin’s Min-Max theorem [29], that allows us to derive a closed-form gradient for the proposed cost.
- We propose proximal-based consistent sparse coding, and dictionary learning algorithms, for nonlinear measurements. We show that these algorithms can be applied to clipped, quantized and 1-bit measurements.

### B. Organization of paper

The paper is organised as follows: in Section II, we briefly review sparse recovery and dictionary learning from linear measurements, and some strategies proposed to deal with clipped, quantized and 1-bit measurements. In Section III we propose a unifying cost function for nonlinear measurements, and show some of its properties. In Section IV we propose consistent sparse coding and dictionary learning algorithms using the proposed cost. Applications of the proposed framework, and links to previous work are presented in Section V. The performance of the proposed algorithm is presented in Section VI, before the conclusion is drawn.

### C. Notation

Bold lowercase letters  $\mathbf{x}$  denote vectors and bold uppercase letters  $\mathbf{X}$  denote matrices. The  $i$ -th element of a vector  $\mathbf{x}$  is noted  $x_i$ . The identity matrix is noted  $\mathbf{I}$ . The  $p$ -norm of a vector  $\mathbf{x}$  is  $\|\mathbf{x}\|_p = (\sum x_i^p)^{1/p}$ . The  $\ell_0$  pseudo-norm (i.e. the number of non-zero elements) of  $\mathbf{x}$  is noted  $\|\mathbf{x}\|_0$ . For a matrix  $\mathbf{X}$ ,  $\|\mathbf{X}\|_2$  denotes the matrix 2-norm, i.e. the largest singular value of  $\mathbf{X}$ . We denote  $(\mathbf{x})_+ = \max(\mathbf{0}, \mathbf{x})$  (where  $\max$  is the element-wise maximum), and  $(\mathbf{x})_- = -(-\mathbf{x})_+$ . The floor (i.e. closest lower integer) of a vector is noted  $\lfloor \mathbf{x} \rfloor$  (applied element-wise). The sign (positive or negative) of each element of  $\mathbf{x}$  is noted  $\text{sign}(\mathbf{x})$ . The element-wise multiplication is  $\odot$ . For a set  $\mathcal{C}$ ,  $\text{cl}(\mathcal{C})$  is the closure of  $\mathcal{C}$ , and  $\mathbb{1}_{\mathcal{C}}(\cdot)$  is the indicator function of that set, i.e.  $\mathbb{1}_{\mathcal{C}}(\mathbf{x}) = 0$  when  $\mathbf{x} \in \mathcal{C}$ ,  $+\infty$  otherwise. The notation  $\preceq$  denotes vector-wise inequality.

## II. BACKGROUND

In this paper, we denote observation vectors as  $\mathbf{y} \in \mathbb{R}^L$ , with  $\mathbf{y} = f(\mathbf{x})$ , where  $\mathbf{x} \in \mathbb{R}^N$  is the original un-observed clean signal, and  $f$  is a measurement or distortion function. We further assume that the signal  $\mathbf{x}$  can be decomposed as  $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$  with  $\mathbf{D} \in \mathbb{R}^{N \times M}$  an *overcomplete* dictionary of  $M$  atoms ( $N < M$ ), and  $\boldsymbol{\alpha} \in \mathbb{R}^M$  is a sparse activation vector. In this section we review the different types of linear and nonlinear measurement functions  $f$ , and the associated problem formulations appearing in the literature.

### A. Sparse coding from linear measurements

A widely studied case is when the measurement function is linear, i.e.  $f(\mathbf{x}) = \mathbf{M}\mathbf{x}$  with  $\mathbf{M} \in \mathbb{R}^{L \times N}$ . The corresponding sparse coding problem is often formulated as:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{M}\mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\Psi(\boldsymbol{\alpha}). \quad (4)$$

For example,  $\mathbf{M} = \mathbf{I}$  (the identity matrix) corresponds to clean signals or signals subject to additive Gaussian noise [1], [2]. When  $\mathbf{M}$  is a diagonal binary matrix, (4) corresponds to an inpainting problem [3], [4].

### B. Sparse coding from nonlinear measurements

Sparse coding from linear measurements have been extensively studied in the literature. Signal acquisition systems however often measure signals in a *nonlinear* way, for examples in saturated, quantized or 1-bit measurements. In the following, we review reconstruction strategies proposed in the literature for saturated, quantized and 1-bit measurements.

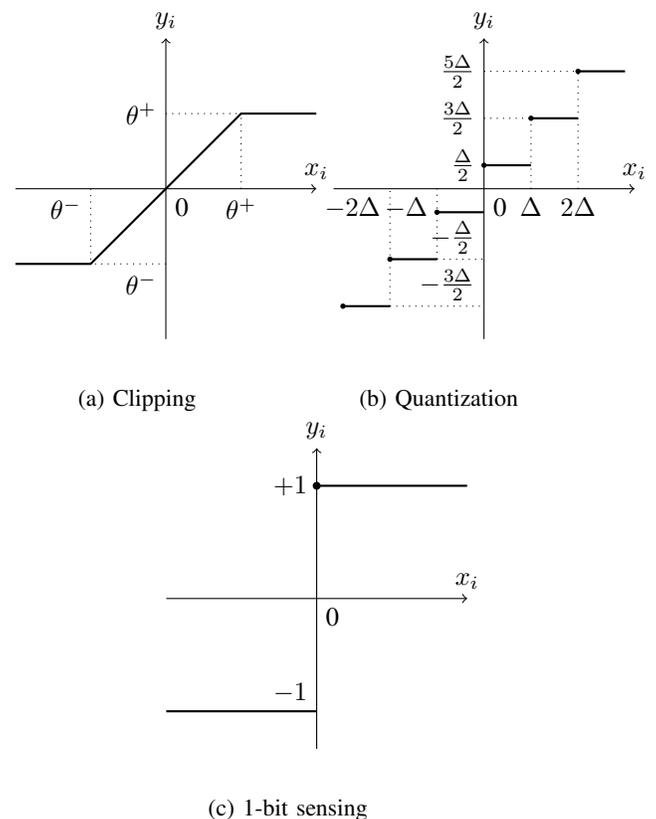


Fig. 1: Visualization of different nonlinear measurement functions  $f$  (output  $y_i = f(x_i)$  versus input  $x_i$ ).

1) *Clipped measurements*: We consider the case of hard clipping, where each sample  $x_i$  is measured as:

$$y_i = f(x_i) = \begin{cases} \theta^+ & \text{if } x_i \geq \theta^+ \\ \theta^- & \text{if } x_i \leq \theta^- \\ x_i & \text{otherwise,} \end{cases} \quad (5)$$

where  $\theta^+ > \theta^-$  are positive and negative clipping thresholds respectively (Figure 1a). This can be written in vector form as:

$$\mathbf{y} = f(\mathbf{x}) = \mathbf{M}^r \mathbf{x} + \theta^+ \mathbf{M}^{c+} \mathbf{1} + \theta^- \mathbf{M}^{c-} \mathbf{1}, \quad (6)$$

where  $\mathbf{1}$  is the all-ones vector in  $\mathbb{R}^N$ , and  $\mathbf{M}^r, \mathbf{M}^{c+}$  and  $\mathbf{M}^{c-}$  are diagonal binary sensing matrices, that define the *reliable* (i.e. unclipped), positive and negative clipped samples respectively and such that  $\mathbf{M}^r + \mathbf{M}^{c+} + \mathbf{M}^{c-} = \mathbf{I}$ . In practice, the clipping thresholds can be estimated from the measurement  $\mathbf{y}$  as (e.g.)  $\theta^+ = \max_i(y_i)$ , and the sensing matrices as  $[\mathbf{M}^{c+}]_{i,i} = 1$  if  $y_i = \theta^+$ , 0 otherwise.

Declicking can be treated as a linear inverse problem by discarding the clipped samples, treating declipping as a linear *inpainting* problem, i.e. solving (4) with  $\mathbf{M} = \mathbf{M}^r$  [4], [8], [17]. However, the reconstruction can be improved by adding extra knowledge about the clipping process. Indeed, we know that the clipped samples should have an amplitude that is greater than the clipping threshold. This extra information can be enforced by solving the following constrained problem [7], [9], [11], [14], [30]:

$$\min_{\alpha} \Psi(\alpha) + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D} \alpha), \quad (7)$$

where:

$$\mathcal{C}(\mathbf{y}) \triangleq \{\mathbf{x} \mid \mathbf{M}^r \mathbf{y} = \mathbf{M}^r \mathbf{x}, \mathbf{M}^{c+} \mathbf{x} \succeq \theta^+ \mathbf{M}^{c+} \mathbf{1}, \mathbf{M}^{c-} \mathbf{x} \preceq \theta^- \mathbf{M}^{c-} \mathbf{1}\} \quad (8)$$

is the *clipping consistency* set. Eqn. (7) is a constrained, non-smooth and possibly non-convex sparse decomposition problem, which can be difficult to solve. An alternating direction method of multipliers (ADMM) [31] based algorithm was proposed in [13] to solve (7). When the dictionary is a tight frame, the algorithm can be computed efficiently [32]. However, for general dictionaries, the algorithm is computationally expensive, since it involves computing non-orthogonal projections at each iteration [13]. A soft consistency metric was used in [8], [10], [12]:

$$\min_{\alpha} \frac{1}{2} [\|\mathbf{M}^r(\mathbf{y} - \mathbf{D} \alpha)\|_2^2 + \|\mathbf{M}^{c+}(\theta^+ \mathbf{1} - \mathbf{D} \alpha)_+\|_2^2 + \|\mathbf{M}^{c-}(\theta^- \mathbf{1} - \mathbf{D} \alpha)_-\|_2^2] + \lambda \Psi(\alpha), \quad (9)$$

The data-fidelity term in (9) is convex and smooth, so methods based on iterative hard thresholding [10], [33] or proximal algorithms [12], [34] can directly be applied.

2) *Quantized measurements*: Quantization maps a continuous input space onto a finite discrete set of codewords  $\mathcal{Y} = \{y_1, \dots, y_p\}$ . A quantization map  $f$  is defined by a set of quantization levels  $\mathcal{R}_q = [l_q, u_q)$  and the relation  $x \in \mathcal{R}_q \Leftrightarrow f(x) = y_q$ , i.e. samples that fall into  $\mathcal{R}_q$  are quantized as  $y_q$ . For example in the case of a uniform mid-riser quantizer,  $\mathcal{R}_q = [\Delta q, \Delta(q+1))$ , and the quantization function can be written as:

$$f(\mathbf{x}) = \Delta \left\lfloor \frac{\mathbf{x}}{\Delta} \right\rfloor + \frac{\Delta}{2}, \quad (10)$$

where  $\Delta > 0$  is the quantization bin width (Figure 1b).

De-quantization can be treated as a simple linear inverse problem by considering quantization error as additive noise, and using a linear sparse model (4) [18]. However it has been

shown that using a more accurate model of the quantization process improves the reconstruction. Bayesian approaches [21],  $\ell_p$ -based data-fidelity terms [22], or specially-tailored cost functions [23] have been proposed in the literature to enforce quantization consistency. Constrained formulations were proposed in [20], [24] in order to enforce consistency:

$$\min_{\alpha} \Psi(\alpha) + \mathbb{1}_{\mathcal{R}}(\mathbf{D} \alpha) \quad (11)$$

where  $\mathcal{R} = \mathcal{R}_{q_1} \times \dots \times \mathcal{R}_{q_N}$  and  $\mathcal{R}_{q_i}$  is the quantization region associated with the  $i$ -th sample  $y_i$ . However similarly to the constrained declipping scenario (7), solving (11) can be computationally expensive.

3) *1-bit measurements*: 1-bit measurement can be seen as an extreme quantization using only one bit per sample, or similarly an extreme saturation where the clipping level tends to zero:

$$f(\mathbf{x}) = \text{sign}(\mathbf{x}). \quad (12)$$

1-bit measurement consistency has also been proposed in [16], [19]:

$$\min_{\alpha} \frac{1}{2} \|\mathbf{y} \odot (\mathbf{D} \alpha)_-\|_2^2 + \lambda \Psi(\alpha) \quad (13)$$

### C. Dictionary learning

Dictionary learning from linear measurements can be formulated as [28]:

$$\min_{\mathbf{D} \in \mathcal{D}, \alpha_t} \sum_{t=1}^T [\|\mathbf{x}_t - \mathbf{D} \alpha_t\|_2^2 + \lambda \Psi(\alpha_t)] \quad (14)$$

where  $\{\mathbf{x}_t\}_{1..T}$  is a collection of  $T$  signals in  $\mathbb{R}^N$ , and  $\alpha_t$  are the corresponding sparse activation vectors. The dictionary  $\mathbf{D}$  is often constrained to be in the convex set  $\mathcal{D} = \{\mathbf{D} \in \mathbb{R}^{N \times M} \mid \forall i, \|\mathbf{d}_i\|_2 \leq 1\}$  in order to avoid scaling ambiguity [28]. Many algorithms have been proposed in the literature to solve (14), such as MOD [35], K-SVD [1] stochastic gradient descent [36], [37], or SimCO [38].

Dictionary learning for 1-bit data have recently been addressed in [39], [40]. To our knowledge, dictionary learning from saturated and quantized measurements, however, has not been addressed in the literature<sup>1</sup>. In the next sections, we propose a unifying and computationally tractable framework for sparse coding and dictionary learning from nonlinear measurements.

## III. A UNIFYING FRAMEWORK FOR NONLINEAR SIGNAL RECONSTRUCTION

Let  $f : \mathcal{X} \mapsto f(\mathcal{X}) = \mathcal{Y}$  be an arbitrary - and possibly nonlinear - measurement function from a clean input space  $\mathcal{X}$  to a measurement space  $\mathcal{Y}$ . For a measured signal  $\mathbf{y} \in \mathcal{Y}$ , we propose a cost function (or data-fidelity term) defined for all  $\mathbf{x} \in \mathcal{X}$  as:

$$\mathcal{L}_f(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}, f^{-1}\{\mathbf{y}\}) \quad (15)$$

where  $f^{-1}\{\mathbf{y}\}$  is the *pre-image* of  $\{\mathbf{y}\}$  under the measurement map  $f$ :

$$f^{-1}\{\mathbf{y}\} \triangleq \{\mathbf{x} \in \mathcal{X} \mid f(\mathbf{x}) = \mathbf{y}\}, \quad (16)$$

<sup>1</sup>We presented some preliminary results on dictionary learning for declipping in [41]

and  $d(\mathbf{x}, \mathcal{C})$  is the distance between  $\mathbf{x}$  and the set  $\mathcal{C}$ , defined for a (pointwise) distance metric  $d(\cdot, \cdot)$  as:

$$d(\mathbf{x}, \mathcal{C}) \triangleq \inf_{\mathbf{z} \in \mathcal{C}} d(\mathbf{x}, \mathbf{z}). \quad (17)$$

The set  $f^{-1}\{\mathbf{y}\}$  can be seen as a *feasibility set*, i.e. the set of all possible input signals  $\mathbf{x} \in \mathcal{X}$  that could have generated  $\mathbf{y}$  when measured through  $f$ . The cost (15) thus measures how “close” a signal  $\mathbf{x}$  is to the feasibility set associated with the measurement  $\mathbf{y}$ . Minimizing (15) thus promotes consistency since it minimizes the distance between an estimate  $\mathbf{x}$  and its feasibility set  $f^{-1}\{\mathbf{y}\}$ . However unlike constrained formulations (7) or (11), here measurement-consistency is enforced in a simple unconstrained way.

### A. Assumptions on $f$ , and choice of distance

Without any assumptions on the feasibility sets  $f^{-1}\{\mathbf{y}\}$  and the metric  $d(\cdot, \cdot)$ ,  $\mathbf{x} \mapsto \mathcal{L}_f(\mathbf{x}, \mathbf{y})$  is in general non-convex and non-smooth, and therefore difficult to optimize. However we show here that under certain conditions, the proposed cost (15) exhibits convenient properties such as convexity, and differentiability with Lipschitz gradient.

The first assumption is that for all  $\mathbf{y} \in \mathcal{Y}$ , the pre-image set  $f^{-1}\{\mathbf{y}\}$  is convex. This assumption is verified by many measurement functions  $f$  found in practice, such as linear measurements, and nonlinear measurements such as clipping, quantization and 1-bit (see Section V). For separable functions  $f(\mathbf{x}) = [f_1(x_1), \dots, f_N(x_N)]$ , a simple sufficient condition such that  $f^{-1}\{\mathbf{y}\}$  is convex for all  $\mathbf{y}$  is that each  $f_i(\cdot)$  is monotonic (but not necessarily strictly monotonic). This condition is however not necessary, and the set of functions with convex pre-images includes a wider range of measurement functions. Convexity of the pre-image sets ensures that the proposed cost (15) is convex and differentiable, as will be shown in this section.

Various metrics can be chosen for the pointwise distance in (17). A popular metric to measure point-wise distances is the squared Euclidean distance  $\|\cdot\|_2^2$ . The squared Euclidean distance, or least squares, is popular in sparse coding and dictionary learning, in part due to its convexity and Lipschitz differentiability. Other pointwise distances are available, such as  $p$ -norms  $\|\cdot\|_p^p$  with  $0 < p \leq 1$ , which might be more appropriate depending on the data at hand. However these come at a cost of non-differentiability and/or non-convexity, which often lead to more difficult optimization problems. In this paper, for simplicity, and in order to favour computationally efficient methods, we focus on the least-squares distance. In fact, we show in this section that the proposed cost (15) benefits from the same properties as – and naturally extends – the linear least-squares cost commonly used in sparse coding.

In the remainder of this paper, we assume  $\mathcal{X} = \mathbb{R}^N$ ,  $d(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$ , and  $f^{-1}\{\mathbf{y}\}$  convex for all  $\mathbf{y} \in \mathcal{Y}$ . Note that for a set  $\mathcal{C}$ ,  $d(\mathbf{x}, \mathcal{C}) = d(\mathbf{x}, \text{cl}(\mathcal{C}))$ , so we can assume without loss of generality that  $f^{-1}\{\mathbf{y}\}$  is closed. Note also that since  $\mathcal{Y} = f(\mathcal{X})$ ,  $f^{-1}\{\mathbf{y}\}$  is non-empty for all  $\mathbf{y} \in \mathcal{Y}$ .

### B. Properties of the proposed cost function

We consider a fixed  $\mathbf{y} \in \mathcal{Y}$ , and review the properties of  $\mathbf{x} \mapsto \mathcal{L}_f(\mathbf{x}, \mathbf{y})$ . Since  $f^{-1}\{\mathbf{y}\}$  is non-empty, closed and convex, the Projection Theorem (Appendix A) ensures existence and uniqueness of a minimizer  $\mathbf{z}^*$  of  $\|\mathbf{x} - \mathbf{z}\|_2^2$  in  $f^{-1}\{\mathbf{y}\}$ . This minimizer is defined as the orthogonal projection  $\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x})$  of  $\mathbf{x}$  on the set  $f^{-1}\{\mathbf{y}\}$ . In particular, the infimum in (17) is attained, and  $\mathcal{L}_f(\mathbf{x}, \mathbf{y})$  can be redefined as:

$$\mathcal{L}_f(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x})\|_2^2. \quad (18)$$

In addition, the continuity property of the projection operator on a convex set ensures that  $\mathcal{L}_f(\cdot, \mathbf{y})$  is a continuous function (as a composition of continuous functions). We now present some properties of  $\mathcal{L}_f(\cdot, \mathbf{y})$ , which make it suitable for a range of optimization algorithms:

**Proposition 1.**  $\mathcal{L}_f(\cdot, \mathbf{y})$  is a convex cost function.

**Proposition 2.**  $\mathcal{L}_f(\cdot, \mathbf{y})$  is differentiable, with gradient:

$$\nabla_{\mathbf{x}} \mathcal{L}_f(\mathbf{x}, \mathbf{y}) = \mathbf{x} - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}). \quad (19)$$

**Proposition 3.** The gradient  $\nabla_{\mathbf{x}} \mathcal{L}_f(\mathbf{x}, \mathbf{y})$  is Lipschitz continuous with constant  $L = 1$ , i.e. for all  $\mathbf{x}_1, \mathbf{x}_2$ :

$$\|\nabla_{\mathbf{x}} \mathcal{L}_f(\mathbf{x}_1, \mathbf{y}) - \nabla_{\mathbf{x}} \mathcal{L}_f(\mathbf{x}_2, \mathbf{y})\|_2^2 \leq \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \quad (20)$$

Proposition 1 is due to the convexity of the set  $f^{-1}\{\mathbf{y}\}$  and of the least-squares cost. Proposition 2 is a direct consequence of Danskin’s Min-Max theorem (Appendix B) and of the uniqueness of the projection operator. Proposition 3 is a consequence of the contraction property of projection onto convex sets. See Appendix C for more detailed proofs.

Continuity, convexity and Lipschitz differentiability makes the proposed cost function suitable for a range of optimization algorithm as will be seen in the next section. Moreover, when  $f$  is the identity map ( $f(\mathbf{x}) = \mathbf{x}$ ), we have  $f^{-1}\{\mathbf{y}\} = \mathbf{y}$  and  $\mathcal{L}_f(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$ . The proposed cost thus generalizes the least squares cost commonly used in sparse coding and dictionary learning. We will show in Section V how the proposed cost also generalizes several cost functions proposed in the literature for inpainting, declipping and 1-bit signals, and how it can be applied to quantized measurements. The proposed cost thus provides a unifying framework to tackle all these problems. In the next section we propose simple proximal-based algorithms for sparse coding and dictionary learning using the proposed cost.

## IV. PROPOSED CONSISTENT SPARSE CODING AND DICTIONARY LEARNING ALGORITHMS

### A. Sparse coding algorithms

For a nonlinear observation  $\mathbf{y}$  and a fixed dictionary  $\mathbf{D}$ , we propose to formulate consistent sparse coding as:

$$\min_{\boldsymbol{\alpha}} \mathcal{L}_f(\mathbf{D}\boldsymbol{\alpha}, \mathbf{y}) + \lambda\Psi(\boldsymbol{\alpha}). \quad (21)$$

Solving (21) is thus a problem of minimizing the sum of a convex, smooth cost function and a non-smooth regularizer. When the regularizer  $\Psi(\cdot)$  is convex, such as the  $\ell_1$ -norm,

this can be classically optimized using *proximal* descent algorithms [34], [42]. The proposed proximal-based sparse coding algorithm is presented in Algorithm 1.

---

**Algorithm 1** Proposed consistent sparse coding algorithm (fixed parameter  $\lambda$ )

---

**Require:**  $f, \mathbf{y}, \mathbf{D}, \alpha^0, \lambda, \mu_1$   
 initialize:  $\alpha \leftarrow \alpha^0$   
**while** stopping criterion not reached **do**  
     **Gradient descent step:**  
      $\alpha \leftarrow \alpha + \mu_1 \mathbf{D}^T (\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{D}\alpha) - \mathbf{D}\alpha)$   
     **Proximal thresholding:**  
      $\alpha \leftarrow \text{prox}_{\lambda\Psi}(\alpha) \triangleq \text{argmin}_{\mathbf{u}} \|\mathbf{u} - \alpha\|_2^2 + \lambda\Psi(\alpha)$   
**return**  $\hat{\alpha}$

---

Note that Algorithm 1 is presented here in its simplest form, however it can easily be accelerated using the same strategy as in [43].

1) *Convergence:* The convergence properties of Algorithm 1 can be summarized as follows:

**Proposition 4.** For a convex penalty  $\Psi(\cdot)$  and a step size  $0 < \mu_1 \leq \frac{1}{\|\mathbf{D}\|_2^2}$  (where  $\|\mathbf{D}\|_2$  is the highest singular value of  $\mathbf{D}$ ), Algorithm 1 converges.

*Proof.* This is a direct consequence of Propositions 1 - 3 and classical results on proximals algorithms with convex and Lipschitz differentiable functions (see, [34], [42] or [43, Theorem 3.1.] for a proof). Here  $\|\mathbf{D}\|_2^2$  is the Lipschitz constant of  $\alpha \mapsto \mathcal{L}_f(\mathbf{D}\alpha, \mathbf{y})$ , as a consequence of Proposition 3.  $\square$

2) *Measurement consistency:* The regularization parameter  $\lambda > 0$  in (21) controls a tradeoff between sparsity and consistency. There is however no analytical formula for the relationship between the parameter  $\lambda$ , and how “consistent” the resulting signal is (i.e., how close it is from its consistency set). In particular the resulting signal is not guaranteed to be *exactly* consistent, i.e. exactly within its consistency set. One way to circumvent this is to solve (21) iteratively for different values of the parameter  $\lambda$ , starting from a large value, and progressively lowering  $\lambda$  until a desired consistency is achieved. This can be done efficiently using a warm-start strategy, i.e. initializing each iteration by the estimate of the previous iteration. Algorithm 2 is a simple modification of Algorithm 1 that implements this strategy ( $\{\lambda^k\}_{k \geq 0}$  is a series of non-increasing and strictly positive values).

---

**Algorithm 2** Proposed consistent sparse coding algorithm (adaptive parameter  $\lambda$ )

---

**Require:**  $f, \mathbf{y}, \mathbf{D}, \alpha^0, \{\lambda^k\}_{k \geq 0}, \mu_1, \epsilon$   
 initialize:  $\alpha \leftarrow \alpha^0, k \leftarrow 0, \lambda \leftarrow \lambda^0$   
**while**  $\mathcal{L}_f(\mathbf{D}\alpha^k, \mathbf{y}) > \epsilon$  **do**  
     **Iterate until convergence:**  
      $\alpha \leftarrow \text{prox}_{\lambda\Psi}(\alpha + \mu_1 \mathbf{D}^T (\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{D}\alpha) - \mathbf{D}\alpha))$   
     **Update  $\lambda$ :**  
      $\lambda \leftarrow \lambda^{k+1}, k \leftarrow k + 1$   
**return**  $\hat{\alpha}$

---

Algorithm 2 is similar to *homotopy* methods proposed for sparse coding [44], [45]. By progressively decreasing  $\lambda$ , the algorithm essentially adds more coefficients to the support set, making the signal less sparse but more consistent. We furthermore have the following:

**Proposition 5.** Let  $F(\lambda, \alpha) \triangleq \mathcal{L}_f(\mathbf{D}\alpha, \mathbf{y}) + \lambda\Psi(\alpha)$ ,  $\lambda^0 > \dots > \lambda^k > \dots > 0$  and  $\alpha^k \triangleq \text{argmin}_{\alpha} F(\lambda^k, \alpha)$ . We have:  
 $\forall k, \Psi(\alpha^{k+1}) \geq \Psi(\alpha^k)$  and  $\mathcal{L}_f(\mathbf{D}\alpha^{k+1}, \mathbf{y}) \leq \mathcal{L}_f(\mathbf{D}\alpha^k, \mathbf{y})$ .

Moreover, if  $\{\lambda_k\} \rightarrow 0$ , then the sequence  $\{F(\lambda^k, \alpha^k)\}_{k \geq 0}$  converges to an optimum of the constrained problem:

$$\min_{\alpha} \Psi(\alpha) \quad \text{s.t.} \quad \mathbf{D}\alpha \in f^{-1}\{\mathbf{y}\}. \quad (22)$$

Proposition 5 and its proof (in Appendix C) are inspired by results on penalty methods for constrained optimization problems [46]. Algorithm 2 can in fact be seen as an unconstrained penalty method to solve the constrained problem (22). Proposition 5 shows that Algorithm 2 leads to asymptotically consistent solutions.

3) *Influence of noise:* In the case when additive noise is present, the signal is measured as:

$$\mathbf{y} = f(\mathbf{D}\alpha^* + \mathbf{n}), \quad (23)$$

i.e.  $\mathbf{D}\alpha^* + \mathbf{n} \in f^{-1}(\mathbf{y})$ , and the original signal  $\mathbf{D}\alpha^*$  is no longer necessarily in the pre-image set of the observation  $\mathbf{y}$ . For this reason, finding a solution that is both sparse and consistent with the measurement  $\mathbf{y}$  might be infeasible. We can however show that if the noise level is small, the cost function does not deviate far from zero, since:

$$\begin{aligned} \mathcal{L}_f(\mathbf{D}\alpha^*, \mathbf{y}) &= \min_{\mathbf{z} \in f^{-1}\{\mathbf{y}\}} \|\mathbf{D}\alpha^* - \mathbf{z}\|_2^2 \\ &\leq \min_{\mathbf{z} \in f^{-1}\{\mathbf{y}\}} \|\mathbf{D}\alpha^* + \mathbf{n} - \mathbf{z}\|_2^2 + \|\mathbf{n}\|_2^2 \\ &= \|\mathbf{n}\|_2^2. \end{aligned} \quad (24)$$

In other words, if the noise level  $\|\mathbf{n}\|_2^2$  is small, then the original signal  $\mathbf{D}\alpha^*$  is *approximately* consistent with the measurement  $\mathbf{y}$ , i.e.  $\mathcal{L}_f(\mathbf{D}\alpha^*, \mathbf{y})$  is small. A good reconstruction strategy is then to solve:

$$\text{argmin}_{\alpha} \Psi(\alpha) \quad \text{s.t.} \quad \mathcal{L}_f(\mathbf{D}\alpha, \mathbf{y}) \leq \|\mathbf{n}\|_2^2, \quad (25)$$

which in its Lagrangian form is equivalent to 21.

## B. Dictionary learning algorithm

For a collection  $\{\mathbf{y}_t\}_{t=1, \dots, T}$  of  $T$  signals measured through the same measurement function  $f$ , consistent dictionary learning can be formulated using the proposed cost as:

$$\min_{\mathbf{D} \in \mathcal{D}, \alpha_t} \sum_{t=1}^T \left( \mathcal{L}_f(\mathbf{D}\alpha_t, \mathbf{y}_t) + \lambda\Psi(\alpha_t) \right) \quad (26)$$

Jointly minimizing  $\mathbf{D}$  and  $\{\alpha_t\}_{t=1, \dots, T}$  in (26) is a non-convex problem. Dictionary learning algorithms typically alternate between a sparse coding, and a dictionary update step [28]. The sparse coding step (with a fixed dictionary) can be solved using the proposed consistent sparse coding algorithm

(Algorithm 1). Once the sparse codes  $\{\alpha_t\}_{1,\dots,T}$  have been updated, the dictionary update step can be formulated as:

$$\min_{\mathbf{D} \in \mathcal{D}} \sum_{t=1}^T \mathcal{L}_f(\mathbf{D} \alpha_t, \mathbf{y}_t), \quad (27)$$

which can be solved using projected gradient descent [46], i.e. alternating between a gradient descent step, and a projection step  $\Pi_{\mathcal{D}}$  which here simply re-normalizes each column  $\mathbf{d}_i$  of  $\mathbf{D}$  as  $\mathbf{d}_i \leftarrow \mathbf{d}_i / \max(\|\mathbf{d}_i\|_2, 1)$ . The proposed dictionary update step is thus similar to classical projected gradient descent approaches already proposed for dictionary learning [36], [47]–[49]. The proposed dictionary learning algorithm is presented in Algorithm 3. The parameter  $\mu_2$  is a step size which can be set as  $\mu_2 = 1/L_2$  where  $L_2 = \|\mathbf{A}\|_2^2$  ( $\mathbf{A} \triangleq [\alpha_1, \dots, \alpha_T]$ ) is the Lipschitz constant of the cost in (27).

---

**Algorithm 3** Proposed consistent dictionary learning algorithm

---

**Require:**  $f, \{\mathbf{y}_t\}_{1\dots T}, \mathbf{D}^0, \{\alpha_t^0\}_{1,\dots,T}, \lambda$   
 initialize:  $\mathbf{D}^{(0)} \leftarrow \mathbf{D}^0, \alpha_t^{(0)} \leftarrow \alpha_t^0, i \leftarrow 0$   
**while** stopping criterion not reached **do**  
      $i \leftarrow i + 1$   
     **Sparse coding step:**  
     **for**  $t = 1\dots T$  **do**  
         Initialize  $\alpha_t \leftarrow \alpha_t^{(i-1)}$ .  
         Update  $\alpha_t^{(i)}$  using Algorithm 1 with  $\mathbf{D} = \mathbf{D}^{(i-1)}$ .  
     **Dictionary update step:**  
     Initialize  $\mathbf{D} \leftarrow \mathbf{D}^{(i-1)}$   
     **while** not converged **do**  
          $\mathbf{D} \leftarrow \mathbf{D} + \mu_2 \sum_t (\Pi_{f^{-1}\{\mathbf{y}_t\}}(\mathbf{D} \alpha_t^{(i)}) - \mathbf{D} \alpha_t^{(i)}) \alpha_t^{(i)T}$   
          $\mathbf{D} \leftarrow \Pi_{\mathcal{D}}(\mathbf{D})$   
      $\mathbf{D}^{(i)} \leftarrow \mathbf{D}$   
**return**  $\hat{\mathbf{D}}, \{\hat{\alpha}_t\}_{1\dots T}$

---

*C. Discussions: extensions to non-convex sets and other distance metrics*

When the pre-image set  $f^{-1}\{\mathbf{y}\}$  is non-convex, the proposed cost is no longer convex. Furthermore, the projection in (44) is no longer necessarily unique, and as a consequence (following Theorem 2), the proposed cost is no longer differentiable. The algorithms proposed in this section could be extended to non-convex sets, however convergence to a global optimum would not be guaranteed, even for  $\ell_1$ -based sparse coding algorithms. For this reason, we focus in this paper on measurement functions with convex pre-image sets (which is the case for many measurement functions).

The proposed optimization problem (21) can be seen as a penalty method to solve the constrained problem (22). Here, the proposed cost used with an  $\ell_2$ -distance in (17) penalizes the samples outside of their consistency set with a quadratic penalty, while the samples inside the set have a cost of zero. However other types of distances have been used for penalty methods. An  $\ell_1$ -norm enforces a softer penalty on larger values, and can be used when large outliers are present. This has been proposed in the context of quantized measurements

in [25] and [50]. However, an  $\ell_1$ -norm leads to a non-differentiable data-fidelity term. Subgradient methods can be derived, however they are known to have a slow convergence rate [51], and require careful tuning of the gradient descent parameter at every iteration. Other penalty functions include the Huber loss, which is differentiable and robust to outliers, but requires tuning of an additional parameter beforehand. Log-barrier functions can also be used to solve constrained problems in an unconstrained way, by forcing the estimates to remain interior to the set and away from the boundary [46]. Log-barriers function however tend to favour solutions that are far away from the boundary, which is not desirable in the context of consistent signal recovery. Finally, other approaches in the literature propose a cost based a maximum-likelihood estimation of the signal, assuming Gaussian additive noise [21], [52], [53]. However, they often involve computing cumulative distribution functions, and therefore lead to more complex formulations than the simple Euclidean distance proposed here. We focus here on the  $\ell_2$ -based data fidelity cost, since it leads to a simpler formulation and algorithms, and naturally extends the linear least-squares used in sparse coding and dictionary learning.

## V. APPLICATIONS AND LINK WITH PREVIOUS WORK

In this section we show how the proposed framework can be applied to linear inverse problems such as denoising and inpainting, and nonlinear inverse problems such as declipping, de-quantization and 1-bit recovery. We give explicit formulations for the proposed cost and show links with costs proposed in the literature.

### A. Linear measurements:

In the linear case  $\mathbf{y} = \mathbf{M}\mathbf{x}$ , the projection operator can be written as:

$$\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}) = \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{z}\|_2^2 \quad \text{s.t.} \quad \mathbf{M}\mathbf{z} = \mathbf{y}, \quad (28)$$

which (when  $\mathbf{M}\mathbf{M}^T$  is invertible) can be computed as:

$$\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}) = \mathbf{x} - \mathbf{M}^T(\mathbf{M}\mathbf{M}^T)^{-1}(\mathbf{M}\mathbf{x} - \mathbf{y}). \quad (29)$$

This shows that our proposed cost can be computed as:

$$\mathcal{L}_f(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \|\mathbf{M}^T(\mathbf{M}\mathbf{M}^T)^{-1}(\mathbf{M}\mathbf{x} - \mathbf{y})\|_2^2, \quad (30)$$

where we recognise the right pseudo-inverse  $\mathbf{M}^T(\mathbf{M}\mathbf{M}^T)^{-1}$  of  $\mathbf{M}$ . The cost (30) can thus be seen as “inverting” the linear measurements and computing the error in the input space. In practice however computing the pseudo-inverse might be expensive or not feasible. In simple cases however, we retrieve classical costs used in the literature. When the measurements are clean or subject to additive Gaussian noise,  $\mathbf{M} = \mathbf{I}$  and we have:

$$\mathcal{L}_f(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2, \quad (31)$$

which is the classical linear least squares commonly used in sparse coding and dictionary learning. When  $\mathbf{M}$  is a diagonal

binary matrix (in the inpainting case), the projection (28) can be computed as:

$$\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}) = \mathbf{y} + (\mathbf{I} - \mathbf{M})\mathbf{x}, \quad (32)$$

and:

$$\mathcal{L}_f(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2, \quad (33)$$

which is the masked least squares commonly used for signal inpainting [4]. This shows in particular that the algorithms proposed in Section IV extend classical algorithms such as ISTA, IHT or gradient-descent based dictionary learning algorithms.

### B. Saturated/clipped measurements

In the case of saturated signals, using the notations of Section II, the feasibility set can be defined in closed form as:

$$f^{-1}\{\mathbf{y}\} = \{\mathbf{x} \mid \mathbf{M}^r \mathbf{y} = \mathbf{M}^r \mathbf{x}, \mathbf{M}^{c+} \mathbf{x} \succeq \mathbf{M}^{c+} \mathbf{y}, \mathbf{M}^{c-} \mathbf{x} \preceq \mathbf{M}^{c-} \mathbf{y}\} \quad (34)$$

which is a convex set. The projection can be computed as:

$$\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{D}\boldsymbol{\alpha}) = \mathbf{M}^r \mathbf{y} + \mathbf{M}^{c+} \max(\mathbf{y}, \mathbf{D}\boldsymbol{\alpha}) + \mathbf{M}^{c-} \min(\mathbf{y}, \mathbf{D}\boldsymbol{\alpha}). \quad (35)$$

This shows that the proposed cost can be written in closed form as:

$$\mathcal{L}_f(\mathbf{D}\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2} [\|\mathbf{M}^r(\mathbf{y} - \mathbf{D}\boldsymbol{\alpha})\|_2^2 + \|\mathbf{M}^{c+}(\mathbf{y} - \mathbf{D}\boldsymbol{\alpha})_+\|_2^2 + \|\mathbf{M}^{c-}(\mathbf{y} - \mathbf{D}\boldsymbol{\alpha})_-\|_2^2]. \quad (36)$$

The proposed cost thus generalizes the soft consistency metric proposed in [8], [10], [12] for declipping. When  $\Psi(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_0$ , Algorithm 1 is thus equivalent to the consistent IHT declipping algorithm proposed in [10]. When  $\Psi(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_1$ , Algorithm 1 is similar to the ISTA-type declipping algorithms proposed in [12].

### C. Quantized measurements

We consider a general quantizer defined by quantization levels  $y_i$  and quantization sets  $f^{-1}\{y_i\} = [l_i, u_i]$  for each sample  $i$ . As commented earlier and discussed in [20], [21], one can assume  $f^{-1}\{y_i\} = [l_i, u_i]$  (the closure of  $[l_i, u_i]$ ) without affecting the cost function. The projection operator for each sample  $x_i$  can be computed as:

$$\Pi_{f^{-1}\{\mathbf{y}\}}(x_i) = \begin{cases} u_i & \text{if } x_i \geq u_i \\ l_i & \text{if } x_i \leq l_i \\ x_i & \text{otherwise.} \end{cases} \quad (37)$$

When concatenating the quantization boundaries  $\mathbf{l} = [l_1, \dots, l_N]$  and  $\mathbf{u} = [u_1, \dots, u_N]$ , the proposed cost can be written as:

$$\mathcal{L}_f(\mathbf{D}\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2} [\|(\mathbf{l} - \mathbf{D}\boldsymbol{\alpha})_-\|_2^2 + \|(\mathbf{u} - \mathbf{D}\boldsymbol{\alpha})_+\|_2^2]. \quad (38)$$

For example in the case of uniform mid-riser quantizer:

$$\mathcal{L}_f(\mathbf{D}\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2} \left[ \left\| \left( \mathbf{y} - \frac{\Delta}{2} - \mathbf{D}\boldsymbol{\alpha} \right)_- \right\|_2^2 + \left\| \left( \mathbf{y} + \frac{\Delta}{2} - \mathbf{D}\boldsymbol{\alpha} \right)_+ \right\|_2^2 \right]. \quad (39)$$

This cost is somewhat similar to the optimization problem in [54] solved at every iteration. However in [54], both the signal and its projection are optimized, solving a quadratic programming problem at every iteration. Here, only the sparse coefficients need to be optimized using simple gradient descent. Interestingly, the authors in [54] also assume convexity of the quantization sets, in order to ensure the solutions to be well-defined and unique. The idea of projecting onto convex sets for image decoding was also used in [55].

### D. 1-bit sensing

In the case of 1-bit measurements, the projection operator can easily be computed for each sample  $x_i$  as:

$$\Pi_{f^{-1}\{\mathbf{y}\}}(x_i) = \begin{cases} x_i & \text{if } \text{sign}(x_i) = y_i \\ 0 & \text{otherwise,} \end{cases} \quad (40)$$

and it can be easily verified that:

$$\mathcal{L}_f(\mathbf{D}\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2} \|(\mathbf{y} \odot (\mathbf{D}\boldsymbol{\alpha}))_-\|_2^2, \quad (41)$$

which shows that the proposed cost is equivalent to the cost (13) proposed for 1-bit signals [16], [19]. We can verify that this indeed corresponds to the clipping consistency cost, since when  $\theta = 0^+$  we have in (36):

$$\begin{aligned} \mathcal{L}_f(\mathbf{D}\boldsymbol{\alpha}, \mathbf{y}) &= \frac{1}{2} [\|\mathbf{M}^{c+}(\mathbf{0}^+ - \mathbf{D}\boldsymbol{\alpha})_+\|_2^2 + \|\mathbf{M}^{c-}(\mathbf{0}^- - \mathbf{D}\boldsymbol{\alpha})_-\|_2^2] \\ &= \frac{1}{2} \|(\text{sign}(\mathbf{y}) \odot (\mathbf{D}\boldsymbol{\alpha}))_-\|_2^2 \end{aligned} \quad (42)$$

Similarly for quantization, taking  $l_i = 0$ ,  $u_i \rightarrow +\infty$  if  $y_i > 0$ ,  $l_i \rightarrow -\infty$ ,  $u_i = 0$  if  $y_i < 0$  in (38), gives the same result.

### E. Summary

The proposed framework unifies cost functions used for denoising, inpainting, declipping and 1-bit recovery. It can also be used for quantization with any quantization map. In particular, performing sparse coding or dictionary learning on clipped, quantized or 1-bit measurements can be done as simply and computationally efficiently as with clean, linear measurement. Experimental results are shown in the next section.

## VI. EVALUATION

We evaluate the performance of the proposed framework on declipping, de-quantization and 1-bit recovery tasks. Each algorithm is evaluated in terms of Signal-to-Noise Ratio (SNR) signal:  $\text{SNR}(\hat{\mathbf{x}}, \mathbf{x}) = 20 \log \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}$  where  $\hat{\mathbf{x}} = \hat{\mathbf{D}}\hat{\boldsymbol{\alpha}}$  is the estimated signal, and  $\mathbf{x}$  is the reference clean signal. However in the case of 1-bit measurements, since the signal can only be recovered up to an amplitude factor, we use the *angular* SNR [23]:

$$\text{SNR}_{\text{angular}}(\hat{\mathbf{x}}, \mathbf{x}) \triangleq 20 \log \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \frac{\|\mathbf{x}\|_2}{\|\hat{\mathbf{x}}\|_2} \hat{\mathbf{x}}\|_2}. \quad (43)$$

### A. Consistent sparse coding

We first evaluate the performance of the proposed consistent sparse coding algorithms. We generate a dictionary  $\mathbf{D} \in \mathbb{R}^{32 \times 64}$ , with i.i.d. normally distributed entries, and unit  $\ell_2$ -norm columns. We then generate  $T = 2000$   $K$ -sparse coefficients  $\alpha_t \in \mathbb{R}^{64}$  with i.i.d normal distribution for the coefficients. We normalize the resulting signals  $\mathbf{x}_t = \mathbf{D} \alpha_t$  to unit  $\ell_\infty$  norm, and artificially clip or quantize the signals as  $\mathbf{y}_t = f(\mathbf{x}_t)$ . We consider clipping with different levels  $\theta$ . For quantization, we consider a uniform mid-rise quantizer that quantizes the input space  $[-1, 1]$  using  $N_b$  bits, i.e. using  $2^{N_b}$  quantization levels of size  $\Delta = 2/2^{N_b}$ .

Figure 2 shows the performance of the proposed algorithms, implemented with  $\Psi(\alpha) = \|\alpha\|_1$ ,  $\lambda = 10^{-2}$  (for Algorithm 1), and the adaptive parameter strategy (Algorithm 2). For the adaptive strategy, we used  $\lambda^0 = \|\nabla \mathcal{L}_f(\mathbf{D}\alpha, \mathbf{y})\|_{\alpha=0} = \|\mathbf{D}^T \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{0})\|_\infty$ , and  $\lambda^{k+1} = \lambda^k/2$ . The algorithm is stopped when a consistency level  $\mathcal{L}_f(\mathbf{D}\alpha, \mathbf{y}) \leq \epsilon$  with  $\epsilon = 10^{-3}$  is reached. We also compare with the ADMM-based algorithm that solves the exact constraint problem (22), as representative of the state-of-the-art. This was recently used for de-quantization in [24], and declipping in [13], although here we use the  $\ell_1$ -norm unlike the  $\ell_0$ -norm used in [13]<sup>2</sup>. The ADMM algorithm is limited to 400 iterations due its computational complexity. Other algorithms are run for a maximum of 400 iterations for a fair comparison. Figure 2 shows that the proposed algorithm with a carefully chosen fixed parameter  $\lambda$  can outperform the exact constrained-based algorithm. This suggests that solutions that are slightly less consistent but sparser, might lead to better reconstruction. However fixing the parameter  $\lambda$  might not be optimal, since in practice the best parameter might depend on the sparsity level, distortion level and/or signal energy. We can see in Figure 2 that using Algorithm 2 and the above described sequence of  $\lambda^k$  performs better than using a fixed parameter  $\lambda$ . In particular the solution reached by the proposed adaptive algorithm performs better than the constrained approach. Another important consideration is the computational time. The ADMM-based algorithm involves computing non-orthogonal projections at each iteration, which have to be computed iteratively, resulting in a high overall computational cost. The proposed algorithms 1 and 2 on the other hand only require gradient computations and element-wise operations, and are thus computationally efficient. The average computational time of each algorithm is shown in Table I.

cpu time (s)	Alg. 2	Alg. 1	ADMM
declipping	<b>1.08</b>	2.75	281.7
dequantization	<b>1.19</b>	2.91	320.3

TABLE I: Average computational time of each algorithm

### B. Consistent dictionary learning

We evaluate the proposed consistent dictionary learning algorithm on real speech signals. The dataset consists of 10

<sup>2</sup>An *analysis* sparsity formulation was also proposed in [13], but we refer here to the synthesis version.

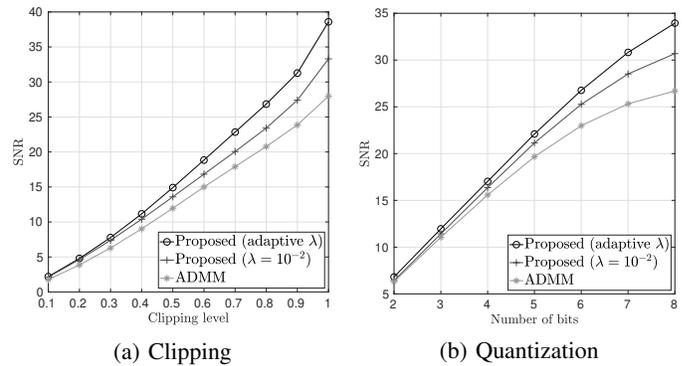


Fig. 2: Comparison of the proposed consistent sparse coding algorithms using the  $\ell_1$ -norm, versus solving the constrained problem with ADMM.

male and female speech signals, taken from the SISEC dataset [56]. Each signal is 10s long, sampled at 16kHz, with 16 bits per sample. Each signal is normalized to unit  $\ell_\infty$  norm, and then processed using overlapping time frames of size  $N = 256$ , with rectangular windows and 75% overlap, for a total of approximately  $T = 2500$  frames per signal. All sparse coding experiments are run using an overcomplete DCT dictionary of size  $M = 512$ . All dictionary learning algorithms are initialized using the same DCT dictionary. To speed up convergence, the sparse coefficients and dictionaries are initialized at every iteration using the estimates from the previous iteration. Similarly, the step size parameters  $\mu_1$  and  $\mu_2$  can be re-estimated at every iteration as  $\mu_1 = 1/\|\mathbf{D}\|_2^2$  and  $\mu_2 = 1/\|\mathbf{A}\|_2^2$  using the current estimates of  $\mathbf{D}$  and  $\mathbf{A}$ .

Figure 3 shows the performance of the proposed framework for consistent sparse coding and dictionary learning, compared to classical linear sparse coding and dictionary learning. In the case of declipping, the classical approach is to discard the clipped samples and treat declipping as an inpainting problem [4]. In the quantization case, the classical approach is to treat the quantized signals as noisy signals with variance  $\frac{\Delta^2}{12}$  [19], which is enforced by stopping the algorithm when an error  $\epsilon = \frac{\Delta^2}{12}$  is reached. For 1-bit signals, we simply use the sign measurements directly as the input [19]. All algorithms are run with an  $\ell_0$ -constraint with fixed  $K = 32$ . In all three cases classical sparse coding is run using IHT [33], and classical dictionary learning alternates between IHT to update the sparse coefficients and gradient descent to update the dictionary. We perform 50 iterations for the sparse coding algorithms, and 50 iterations (with 20 iterations at each inner step) for dictionary learning.

Figure 3a shows the declipping performance, for different clipping levels ranging from  $\theta = 0.1$  (highly clipped) to  $\theta = 1$  (unclipped). Figure 3a demonstrates several things: First, using measurement consistency greatly improves the reconstruction. Consistent sparse coding shows an improvement of up to 8dB compared to classical sparse coding. Consistent dictionary learning shows up to 10dB improvement compared to classical dictionary learning. This improvement is greater when the signals are highly distorted ( $\theta \leq 0.5$ ). As expected, the two frameworks give equivalent results when

TABLE II: Performance of consistent dictionary learning on 1-bit recovery

Angular SNR (dB)	Classical sparse coding (DCT)	Classical dictionary learning	Consistent sparse coding (DCT) (Algorithm 1)	Consistent dictionary learning (Algorithm 3)
Female speech	5.69	5.50	5.91	<b>6.12</b>
Male speech	4.67	4.47	4.50	<b>4.70</b>

$\theta = 1$ , which shows how the proposed consistent framework naturally extends classical sparse coding. Second, we can see that consistent dictionary learning greatly improves the reconstruction performance compared to using a fixed DCT dictionary, since it adapts the dictionary to the signals of interest. This shows that the learned dictionary, although it has been learned using only the low-energy unclipped samples, along with consistency penalties, generalizes well to the unobserved clipped samples. In particular, the proposed consistent dictionary learning algorithm outperforms all the other methods. Finally, it is interesting to point out that when the signals are highly clipped ( $\theta = 0.1$ ), classical dictionary learning does not improve compared to classical sparse coding with DCT. This is probably due to a lack of data to learn from, since most of the data is clipped and discarded. Our consistent dictionary learning algorithm on the other hand, makes use of the clipped data, and is able to learn and improve the performance by 1.7dB.

Figure 3b shows the results for quantization, from highly quantized ( $N_b = 2$  bits) to lightly quantized ( $N_b = 8$ ). Similarly, we can see that using measurement consistency improves the performance when the signals are heavily quantized  $N_b \leq 5$ . As expected, the performance of the consistent framework and of the classical one are comparable when the signals are lightly distorted ( $N_b = 8$ ). Dictionary learning improves the performance (compared to using a fixed dictionary), and the proposed consistent dictionary learning algorithm outperforms the other methods.

The results for 1-bit data are shown in Table II. The results for sparse coding and consistent sparse coding are comparable (note that here for simplicity we don't enforce the signal to be on the unit circle, unlike in [16], [19]). Classical dictionary learning here performs worse than classical sparse coding with a fixed dictionary, presumably because the reconstructed signal overfits the  $\pm 1$  sign measurements. This shows that classical dictionary learning does not perform well with 1-bit data. The proposed consistent dictionary learning however, outperforms consistent sparse coding and classical dictionary learning.

## VII. CONCLUSION

We have presented a unified framework for signal reconstruction from certain types of nonlinear measurements such as clipping, quantization and 1-bit measurements. We proposed a cost function that takes into account prior knowledge about the measurement process, by minimizing the distance to the pre-image of the received signal. When the pre-image is convex, we have shown that the proposed cost is a convex, smooth and Lipschitz differentiable, which makes it ideal for proximal-based algorithms. The proposed cost generalizes the linear least-squares commonly used in sparse coding and dictionary

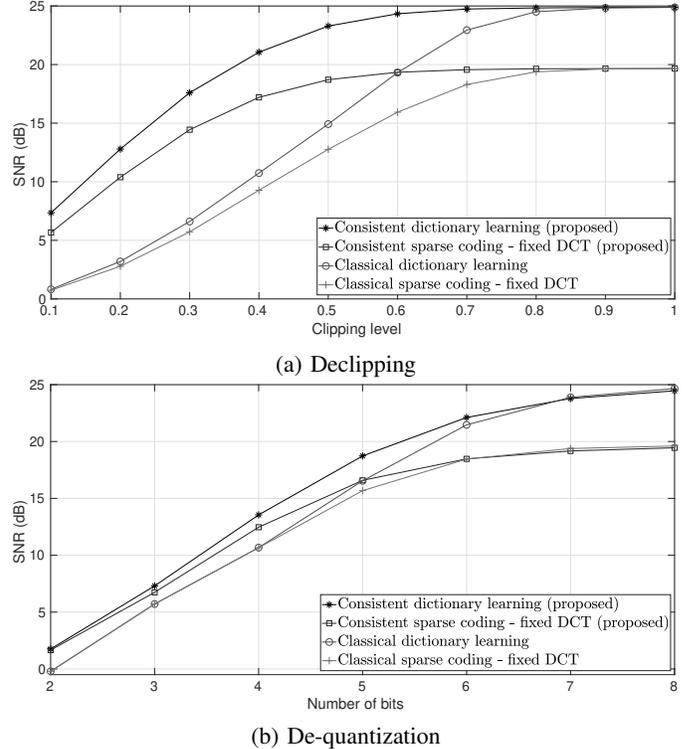


Fig. 3: Comparison of the proposed consistent sparse coding and dictionary learning algorithms, compared to classical sparse coding and dictionary learning.

learning, as well as cost functions proposed for declipping and 1-bit recovery. We proposed proximal based sparse coding and dictionary learning algorithms, which naturally extend classical algorithms and can deal with clipped, quantized and 1-bit measurements.

## APPENDIX A PROJECTION THEOREM

We recall the following theorem [46, Prop. B.11]:

**Theorem 1** (Projection Theorem [46, Prop. B.11]). *Let  $\mathcal{C}$  be a closed convex set in  $\mathbb{R}^N$ . Then, the following hold:*

a) *For every  $\mathbf{x} \in \mathbb{R}^N$ , there exists a unique  $\mathbf{z}^* \in \mathcal{C}$  such that  $\mathbf{z}^*$  minimizes  $\|\mathbf{x} - \mathbf{z}\|_2$  over all  $\mathbf{z} \in \mathcal{C}$ .  $\mathbf{z}^*$  is called the projection of  $\mathbf{x}$  onto  $\mathcal{C}$  and is noted  $\Pi_{\mathcal{C}}(\mathbf{x})$ . In other words:*

$$\Pi_{\mathcal{C}}(\mathbf{x}) \triangleq \underset{\mathbf{z} \in \mathcal{C}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{z}\|_2. \quad (44)$$

b) *For  $\mathbf{x} \in \mathbb{R}^N$ ,  $\mathbf{z}^* = \Pi_{\mathcal{C}}(\mathbf{x})$  if and only if:*

$$(\mathbf{z} - \mathbf{z}^*)^T (\mathbf{x} - \mathbf{z}^*) \leq 0 \quad \forall \mathbf{z} \in \mathcal{C}. \quad (45)$$

c)  *$\mathbf{x} \mapsto \Pi_{\mathcal{C}}(\mathbf{x})$  is continuous and non-expansive, i.e.:*

$$\|\Pi_{\mathcal{C}}(\mathbf{x}_1) - \Pi_{\mathcal{C}}(\mathbf{x}_2)\|_2 \leq \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N. \quad (46)$$

APPENDIX B  
DANSKIN'S MIN-MAX THEOREM

**Theorem 2** (Danskin's Min-Max Theorem [57, Section 4.1], [29]). *Let  $\mathcal{C}$  be a compact<sup>3</sup> set, and  $g(\mathbf{x}) = \min_{\mathbf{z} \in \mathcal{C}} \phi(\mathbf{x}, \mathbf{z})$ . Suppose that for each  $\mathbf{z} \in \mathbb{R}^N$ ,  $\phi(\cdot, \mathbf{z})$  is differentiable with gradient  $\nabla_{\mathbf{x}} \phi(\mathbf{x}, \mathbf{z})$ , and  $\phi(\mathbf{x}, \mathbf{z})$  and  $\nabla_{\mathbf{x}} \phi(\mathbf{x}, \mathbf{z})$  are continuous on  $\mathbb{R}^N \times \mathbb{R}^N$ . Define  $\mathcal{Z}(\mathbf{x}) = \operatorname{argmin}_{\mathbf{z} \in \mathcal{C}} \phi(\mathbf{x}, \mathbf{z})$ . Then  $g$  is directionally differentiable, with derivative in the direction  $\mathbf{h}$ :*

$$\nabla g(\mathbf{x}; \mathbf{h}) = \min_{\mathbf{z} \in \mathcal{Z}(\mathbf{x})} \nabla_{\mathbf{x}} \phi(\mathbf{x}, \mathbf{z})^T \mathbf{h} \quad \forall \mathbf{h}, \quad (47)$$

*In particular, when the minimum is attained at a unique point ( $\mathcal{Z}(\mathbf{x}) = \{\mathbf{z}^*\}$ ),  $g$  is differentiable with gradient:*

$$\nabla g(\mathbf{x}) = \nabla_{\mathbf{x}} \phi(\mathbf{x}, \mathbf{z}^*). \quad (48)$$

In other words, Danskin's Min-Max theorem says that if the minimum over a family of continuous and continuously differentiable functions is attained at a unique point  $\mathbf{z}^*$ , then the gradient of the minimum over this family of functions can be computed by simply evaluating that gradient at the optimum  $\mathbf{z}^*$ .

APPENDIX C  
PROOFS OF PROPOSITION 1-3 AND 5

*Proof of Proposition 1:*  $\mathcal{L}_f(\cdot, \mathbf{y})$  is a minimum of a family of convex functions  $(\mathbf{x}, \mathbf{z}) \mapsto \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2$  over a non-empty convex set, so by [58, Section 3.2.5],  $\mathcal{L}_f(\cdot, \mathbf{y})$  is convex.  $\square$

*Proof of Proposition 2:*  $\mathcal{L}_f(\mathbf{x}, \mathbf{y}) = \min_{\mathbf{z} \in f^{-1}\{\mathbf{y}\}} d(\mathbf{x}, \mathbf{z})$  with  $d(\mathbf{x}, \mathbf{z}) = \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2$ . For all  $\mathbf{z} \in \mathbb{R}^N$ ,  $d(\cdot, \mathbf{z})$  is differentiable with gradient  $\nabla_{\mathbf{x}} d(\mathbf{x}, \mathbf{z}) = \mathbf{x} - \mathbf{z}$ . Furthermore,  $d(\mathbf{x}, \mathbf{z})$  and  $\nabla_{\mathbf{x}} d(\mathbf{x}, \mathbf{z})$  are continuous in  $(\mathbf{x}, \mathbf{z})$ , and  $\mathcal{Z}(\mathbf{x}) = \operatorname{argmin}_{\mathbf{z} \in f^{-1}\{\mathbf{y}\}} d(\mathbf{x}, \mathbf{z})$  is uniquely defined as  $\mathcal{Z}(\mathbf{x}) = \{\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x})\}$  by the Projection Theorem. Using Danskin's Min-Max theorem, we can then conclude that:

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}_f(\mathbf{x}, \mathbf{y}) &= \nabla_{\mathbf{x}} d(\mathbf{x}, \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x})) \\ &= \mathbf{x} - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}). \end{aligned} \quad (49)$$

$\square$

*Proof of Proposition 3:* Let  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$ . By the projection theorem, we have:

$$\begin{aligned} (\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_2) - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_1))^T (\mathbf{x}_1 - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_1)) &\leq 0 \\ (\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_1) - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_2))^T (\mathbf{x}_2 - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_2)) &\leq 0. \end{aligned} \quad (50)$$

Adding and rearranging these two equations gives:

$$\begin{aligned} \|\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_1) - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_2)\|_2^2 \\ \leq (\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_1) - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_2))^T (\mathbf{x}_1 - \mathbf{x}_2). \end{aligned} \quad (51)$$

<sup>3</sup>Note that compactness is only required to ensure existence of a minimum, according to Weierstrass' theorem.

We can then show that:

$$\begin{aligned} &\|\nabla_{\mathbf{x}} \mathcal{L}_f(\mathbf{x}_1, \mathbf{y}) - \nabla_{\mathbf{x}} \mathcal{L}_f(\mathbf{x}_2, \mathbf{y})\|_2^2 \\ &= \|\mathbf{x}_1 - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_1) - (\mathbf{x}_2 - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_2))\|_2^2 \\ &= \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 + \|\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_1) - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_2)\|_2^2 \\ &\quad - 2(\mathbf{x}_1 - \mathbf{x}_2)^T (\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_1) - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_2)) \\ &\leq \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 - \|\Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_1) - \Pi_{f^{-1}\{\mathbf{y}\}}(\mathbf{x}_2)\|_2^2 \\ &\leq \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \end{aligned} \quad (52)$$

where we have used (51) in the third line.  $\square$

*Proof of Proposition 5.* The proof is given assuming exact estimation of the  $\alpha^k$  coefficients, however similar results for approximate estimates are also available [46]. We have by definition of  $\alpha^k$  and  $\alpha^{k+1}$ :

$$\begin{cases} F(\lambda^{k+1}, \alpha^{k+1}) \leq F(\lambda^{k+1}, \alpha^k) \\ F(\lambda^k, \alpha^k) \leq F(\lambda^k, \alpha^{k+1}) \end{cases} \quad (53)$$

Summing and rearranging these two equations gives:

$$(\lambda^{k+1} - \lambda^k)(\Psi(\alpha^{k+1}) - \Psi(\alpha^k)) \leq 0, \quad (54)$$

Since  $\lambda^{k+1} < \lambda^k$ , we have  $\Psi(\alpha^{k+1}) \geq \Psi(\alpha^k)$ , from which follows (by combining with e.g., the first line of (53))  $\mathcal{L}_f(\mathbf{D} \alpha^{k+1}, \mathbf{y}) \leq \mathcal{L}_f(\mathbf{D} \alpha^k, \mathbf{y})$ .

We have furthermore that:

$$\min_{\alpha} F(\lambda^k, \alpha) \leq \min_{\mathbf{D} \alpha \in f^{-1}\{\mathbf{y}\}} F(\lambda^k, \alpha) = \lambda^k \Psi^* \quad (55)$$

Where  $\Psi^*$  is the optimum value of (22). In other words:

$$0 \leq \mathcal{L}_f(\mathbf{D} \alpha^k, \mathbf{y}) + \lambda^k \Psi(\alpha^k) \leq \lambda^k \Psi^* \quad (56)$$

When  $\{\lambda^k\} \rightarrow 0$ , taking the limit in (56) shows that  $\lim_k \mathcal{L}_f(\mathbf{D} \alpha^k, \mathbf{y}) = 0$  and  $\lim_k \Psi(\alpha^k) \leq \Psi^*$ . In particular,  $F(\lambda^k, \alpha^k)$  converges to the optimum value  $\Psi^*$  of the constrained problem (22).  $\square$

REFERENCES

- [1] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [2] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.
- [3] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Computat. Harmon. Anal.*, vol. 19, no. 3, pp. 340–358, 2005.
- [4] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 3, pp. 922–932, 2012.
- [5] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [6] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, 2012.
- [7] H. Mansour, R. Saab, P. Nasiopoulos, and R. Ward, "Color image desaturation using sparse reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 778–781.
- [8] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Appl. Computat. Harmon. Anal.*, vol. 31, no. 3, pp. 429–443, 2011.
- [9] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "A constrained matching pursuit approach to audio de-clipping," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 329–332.

- [10] S. Kitić, L. Jacques, N. Madhu, M. P. Hopwood, A. Spriet, and C. D. Vleeschouwer, "Consistent iterative hard thresholding for signal declipping," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 5939–5943.
- [11] B. Defraene, N. Mansour, S. D. Hertogh, T. van Waterschoot, M. Diehl, and M. Moonen, "Declipping of audio signals using perceptual compressed sensing," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 12, pp. 2627–2637, Dec. 2013.
- [12] K. Siedenburg, M. Kowalski, and M. Dörfler, "Audio declipping with social sparsity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Florence, Italy, May 2014, pp. 1577–1578.
- [13] S. Kitić, N. Bertin, and R. Gribonval, "Sparsity and cosparsity for audio declipping: a flexible non-convex approach," in *Proc. Int. Conf. on Latent Variable Analysis and Signal Separation (LVA/ICA)*, Liberec, Czech Republic, 2015, pp. 243–250.
- [14] S. Foucart and T. Needham, "Sparse recovery from saturated measurements," *Information and Inference: A Journal of the IMA*, vol. 6, no. 2, pp. 196–212, 2016.
- [15] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.
- [16] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *42<sup>nd</sup> Annual Conference on Information Sciences and Systems*, 2008, pp. 16–21.
- [17] J. N. Laska, P. Boufounos, and R. G. Baraniuk, "Finite range scalar quantization for compressive sensing," in *Proc. Sampling Theory and Applications (SampTA)*, 2009.
- [18] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [19] P. T. Boufounos, "Greedy sparse signal reconstruction from sign measurements," in *Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, 2009, pp. 1305–1309.
- [20] W. Dai, H. V. Pham, and O. Milenkovic, "Distortion-rate functions for quantized compressive sensing," in *IEEE Information Theory Workshop on Networking and Information Theory*, 2009, pp. 171–175.
- [21] A. Zymnis, S. Boyd, and E. Candès, "Compressed sensing with quantized measurements," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 149–152, 2010.
- [22] L. Jacques, D. K. Hammond, and J. M. Fadili, "Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 559–571, 2011.
- [23] L. Jacques, K. Degraux, and C. De Vleeschouwer, "Quantized iterative hard thresholding: Bridging 1-bit and high-resolution quantized compressed sensing," in *Proc. 10th Int. Conf. Sampling Theory Appl. (SampTA)*, 2013.
- [24] A. Moshtaghpour, L. Jacques, V. Cambareri, K. Degraux, and C. De Vleeschouwer, "Consistent basis pursuit for signal and matrix estimates in quantized compressed sensing," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 25–29, 2016.
- [25] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2082–2102, 2013.
- [26] P. Boufounos, M. F. Duarte, and R. G. Baraniuk, "Sparse signal reconstruction from noisy compressive measurements using cross validation," in *2007 IEEE/SP 14<sup>th</sup> Workshop on Statistical Signal Processing (SSP 2007)*, 2007, pp. 299–303.
- [27] F. Couznie-Devy, J. Mairal, F. Bach, and J. Ponce, "Dictionary learning for deblurring and digital zoom," *preprint arXiv:1110.0957*, 2011.
- [28] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Found. Trends Comput. Graph. Vis.*, vol. 8, no. 2-3, pp. 85–283, 2014.
- [29] J. M. Danskin, *The theory of max-min and its application to weapons allocation problems*. Ökonometrie und Unternehmensforschung, 1967.
- [30] P. Smaragdis, "Dynamic range extension using interleaved gains," *IEEE transactions on audio, speech, and language processing*, vol. 17, no. 5, pp. 966–973, 2009.
- [31] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [32] P. Zaviscka, P. Rajmic, Z. Prusa, and V. Vesely, "Revisiting synthesis model in sparse audio declipper," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2018, pp. 429–445.
- [33] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Computat. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.
- [34] N. Parikh, S. P. Boyd *et al.*, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [35] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, 1999, pp. 2443–2446.
- [36] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [37] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. Annual Int. Conf. Machine Learning*, 2009, pp. 689–696.
- [38] W. Dai, T. Xu, and W. Wang, "Simultaneous codeword optimization (SimCO) for dictionary update and learning," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6340–6353, 2012.
- [39] H. Zayyani, M. Korki, and F. Marvasti, "Dictionary learning for blind one bit compressed sensing," *IEEE Signal Processing Letters*, vol. 23, no. 2, pp. 187–191, 2016.
- [40] Y. Shen and G. B. Giannakis, "Online dictionary learning from large-scale binary data," in *24<sup>th</sup> European Signal Processing Conference (EUSIPCO)*, 2016, pp. 1808–1812.
- [41] L. Rencker, F. Bach, W. Wang, and M. D. Plumbley, "Consistent dictionary learning for signal declipping," in *Proc. Int. Conf. on Latent Variable Analysis and Signal Separation (LVA/ICA)*, Guildford, UK, 2018, pp. 446–455.
- [42] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point Algorithms for Inverse Problems in Science and Engineering*. Springer, 2011, pp. 185–212.
- [43] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [44] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Homotopy continuation for sparse signal representation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5. IEEE, 2005, pp. v–733.
- [45] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani *et al.*, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [46] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific Belmont, 1999.
- [47] M. Aharon and M. Elad, "Sparse and redundant modeling of image content using an image-signature-dictionary," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 228–247, 2008.
- [48] K. Kavukcuoglu and Y. Lecun, "Fast inference in sparse coding algorithms with applications to object recognition," in *Technical report, Computational and Biological Learning Lab, Courant Institute, NYU*, 2008.
- [49] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, no. Jan, pp. 19–60, 2010.
- [50] D. Valsesia and P. T. Boufounos, "Universal encoding of multispectral images," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 4453–4457.
- [51] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Convex optimization with sparsity-inducing norms," *Optimization for Machine Learning*, vol. 5, 2011.
- [52] S. Bahmani, P. T. Boufounos, and B. Raj, "Robust 1-bit compressive sensing via gradient support pursuit," *arXiv preprint arXiv:1304.6627*, 2013.
- [53] M. A. Davenport, Y. Plan, E. Van Den Berg, and M. Wootters, "1-bit matrix completion," *Information and Inference: A Journal of the IMA*, vol. 3, no. 3, pp. 189–223, 2014.
- [54] W. Dai and O. Milenkovic, "Information theoretical and algorithmic approaches to quantized compressive sensing," *IEEE transactions on communications*, vol. 59, no. 7, pp. 1857–1866, 2011.
- [55] N. T. Thao and M. Vetterli, "Set theoretic compression with an application to image coding," *IEEE transactions on image processing*, vol. 7, no. 7, pp. 1051–1056, 1998.
- [56] A. Liutkus, F.-R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2017, pp. 323–332.
- [57] J. F. Bonnans and A. Shapiro, "Optimization problems with perturbations: A guided tour," *SIAM Review*, vol. 40, no. 2, pp. 228–264, 1998.
- [58] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.



**Lucas Rencker** graduated from Ecole Centrale Marseille, France, in 2015. He is now a PhD candidate at the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, UK, as part of the “Machine Sensing” Innovative Training Network (MacSeNet). His research interests include sparse coding, compressed sensing and dictionary learning, with application to inverse problems in signal processing.



**Francis Bach** is a researcher at INRIA since 2007, in the Computer Science Department at Ecole Normale Supérieure, Paris. He completed his Ph.D. in Computer Science at U.C. Berkeley, and spent two years in the Mathematical Morphology group at Ecole des Mines de Paris. He obtained in 2009 a Starting Grant and in 2016 a Consolidator Grant from the European Research Council, and received the Inria young researcher prize in 2012, the ICML test-of-time award in 2014, as well as the Lagrange prize in continuous optimization in 2018. In 2015,

he was program co-chair of the International Conference in Machine learning (ICML), and general chair in 2018; he is now co-editor-in chief of the Journal of Machine Learning Research.



**Wenwu Wang** (M02SM11) was born in Anhui, China. He received the B.Sc. degree in 1997, the M.E. degree in 2000, and the Ph.D. degree in 2002, all from Harbin Engineering University, China.

He then worked in Kings College London (2002-2003), Cardiff University (2004-2005), Tao Group Ltd. (now Antix Labs Ltd.) (2005-2006), and Creative Labs (2006-2007), before joining University of Surrey, UK, in May 2007, where he is currently a Professor in Signal Processing and Machine Learning, and a Co-Director of the Machine Audition Lab

within the Centre for Vision Speech and Signal Processing. He has been a Guest Professor at Qingdao University of Science and Technology, China, since 2018.

His current research interests include blind signal processing, sparse signal processing, audio-visual signal processing, machine learning and perception, artificial intelligence, machine audition (listening), and statistical anomaly detection. He has (co)-authored over 250 publications in these areas.

He has been a Senior Area Editor (2019-) and an Associate Editor (2014-2018) for IEEE Transactions on Signal Processing. He has been a Publication Co-Chair for ICASSP 2019, Brighton, UK.



**Mark D. Plumbley** (S'88-M'90-SM'12-F'15) received the B.A.(Hons.) degree in electrical sciences and the Ph.D. degree in neural networks from University of Cambridge, Cambridge, U.K., in 1984 and 1991, respectively. Following his PhD, he became a Lecturer at Kings College London, before moving to Queen Mary University of London in 2002. He subsequently became Professor and Director of the Centre for Digital Music, before joining the University of Surrey in 2015 as Professor of Signal Processing. He is known for his work on analysis

and processing of audio and music, using a wide range of signal processing techniques, including matrix factorization, sparse representations, and deep learning. He is a co-editor of the recent book on Computational Analysis of Sound Scenes and Events. He is a Member of the IEEE Signal Processing Society Technical Committee on Signal Processing Theory and Methods, and a Fellow of the IET and IEEE.