Dictionary learning from nonlinearly generated data

MacSeNet
Machine Sensing Training Network

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

*Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK

†SIERRA-project team, INRIA, Paris, France

UNIVERSITY OF SURREY

Abstract

Dictionary learning is a popular technique for linear inverse problems such as denoising or inpainting. However in many cases, the measurement process is nonlinear, for example for saturated, quantized or 1-bit data. Here, we propose a dictionary learning algorithm that is able to learn from nonlinearly generated data. We propose a cost function that minimizes the distance to a convex feasibility set, and is thus convex and differentiable with a Lipschitz gradient. We then propose a proximal gradient descent-based sparse coding and dictionary learning algorithm that minimizes the proposed cost. We show how the proposed algorithm can be applied to recover signals from saturated, quantized and binary measurements.

Introduction

A dictionary learning problem with clean or noisy measurements is often formulated as:

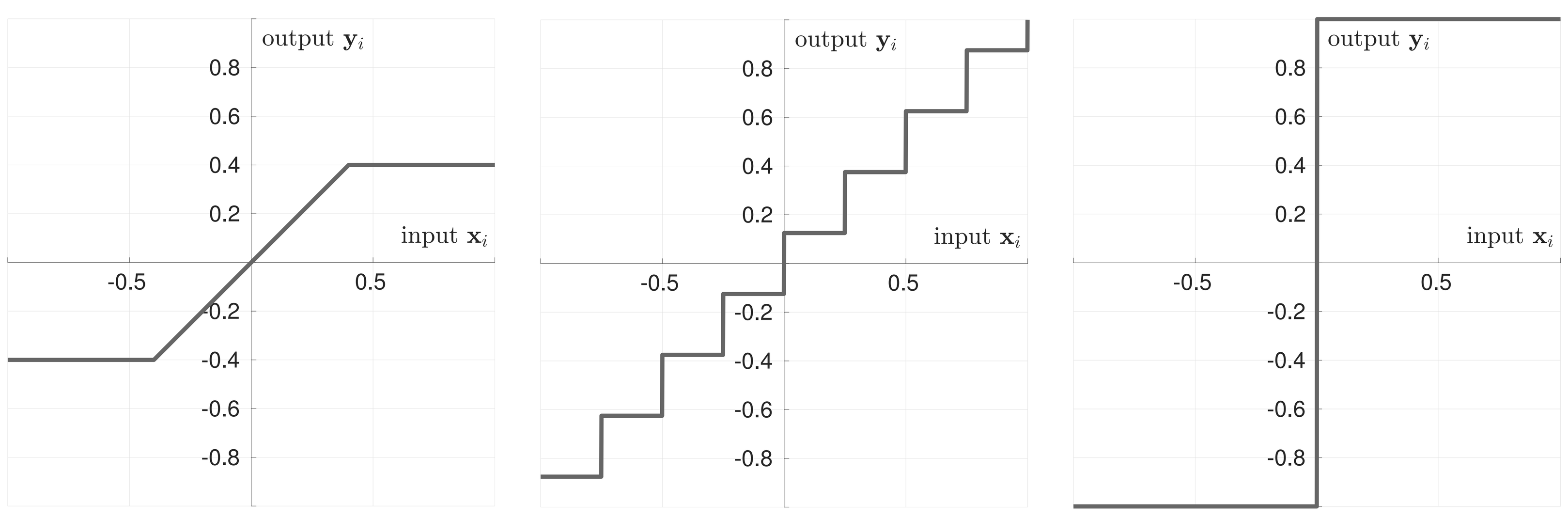
$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_{t=1}^{T} \left[\|\mathbf{x}_t - \mathbf{D}\,\boldsymbol{\alpha}_t\|_2^2 + \lambda\Psi(\boldsymbol{\alpha}_t)\right] \tag{1}$$

where $\{\mathbf{x}_t\}_{1\ldots T}$ is a collection of T signals in $\mathbb{R}^N$, and $\boldsymbol{\alpha}_t$ are the corresponding sparse activation vectors. The dictionary $\mathbf{D}$ is often constrained to be in $\mathcal{D} = \{\mathbf{D}\in\mathbb{R}^{N\times M} | \forall i, \|\mathbf{d}_i\|_2 \leq 1\}$ in order to avoid scaling ambiguity.

However the data is often measured in a nonlinear way:

$$\mathbf{y} = f(\mathbf{x}) \tag{2}$$

where is a known, nonlinear, nonsmooth and noninvertible measurement function. Common examples of nonlinear measurement functions in signal processing are clipping (saturation), quantization and 1-bit compression.



(a) Clipping (b) Quantization (c) 1-bit compression

Figure 1: Visualization of different nonlinear measurement functions f.

Since f is often non-differentiable, solving:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_{t=1}^{T} \left[\|\mathbf{y}_t - f(\mathbf{D}\,\boldsymbol{\alpha}_t)\|_2^2 + \lambda\Psi(\boldsymbol{\alpha}_t)\right] \tag{3}$$

is a challenging problem.

We instead propose to solve:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_{t=1}^{T} \left[\mathrm{d}((\mathbf{D}\,\boldsymbol{\alpha}_t), f^{-1}(\mathbf{y}_t))^2 + \lambda\Psi(\boldsymbol{\alpha}_t)\right], \tag{4}$$

where $f^{-1}(\mathbf{y}_t)$ is the *pre-image* set of $\mathbf{y}_t$ through the measurement f, and $\mathrm{d}(\mathbf{x},\mathcal{C})^2$ is the squared Euclidean distance between $\mathbf{x}$ and the set $\mathcal{C}$, defined as:

$$\mathrm{d}(\mathbf{x},\mathcal{C})^2 = \min_{\mathbf{z}\in\mathcal{C}} \|\mathbf{x}-\mathbf{z}\|_2^2. \tag{5}$$

For a closed convex set $\mathcal{C}$, we recall some properties of $\mathrm{d}(\mathbf{x},\mathcal{C})^2$:

- The data-fidelity term $\mathrm{d}(\mathbf{x},\mathcal{C})$ is convex, as a minimum of a family of convex functions $\|.\|_2$ over a non-empty and convex set.
- $\mathrm{d}(\mathbf{x},\mathcal{C})^2$ is differentiable with gradient:

$$\nabla_{\mathbf{x}}\frac{1}{2}\mathrm{d}(\mathbf{x},\mathcal{C})^2 = \mathbf{x} - \Pi_{\mathcal{C}}(\mathbf{x}), \tag{6}$$

  where $\Pi_{\mathcal{C}}(\mathbf{x})$ is the orthogonal projection of $\mathbf{x}$ onto $\mathcal{C}$.
- The gradient (6) is 1-Lipschitz. This stems from the contraction property of projection operators onto convex sets.

The proposed formulation is thus a problem of minimizing a convex and differentiable data-fidelity term, along with a sparsity-inducing regularizer, which is similar to the classical dictionary learning problem (1). Moreover when the feasibility set is a singleton $f^{-1}(\mathbf{y}) = \{\mathbf{x}_0\}$ (i.e. the signal is unclipped/unquantized) then (4) simplifies to (1).

Algorithm

We propose a gradient-descent based dictionary algorithm to solve (4).

Algorithm 1 Consistent dictionary learning from nonlinear data

Require: $\{\mathbf{y}_t\}_{1\ldots T}$, $\mathbf{D}^0$, n_1, n_2, μ_1, μ_2
  initialize: $\mathbf{D} \leftarrow \mathbf{D}^0$, $\boldsymbol{\alpha}_t \leftarrow \mathbf{0}$
  while stopping criterion not reached **do**
    for $t = 1\ldots T$ **do** ▷ Sparse coding step
      for $i = 1,\ldots,n_1$ **do**

$$\boldsymbol{\alpha}_t \leftarrow \boldsymbol{\alpha}_t + \mu_1\mathbf{D}^T(\Pi_{\mathcal{C}(\mathbf{y}_t)}(\mathbf{D}\,\boldsymbol{\alpha}_t) - \mathbf{D}\,\boldsymbol{\alpha}_t) \tag{7}$$

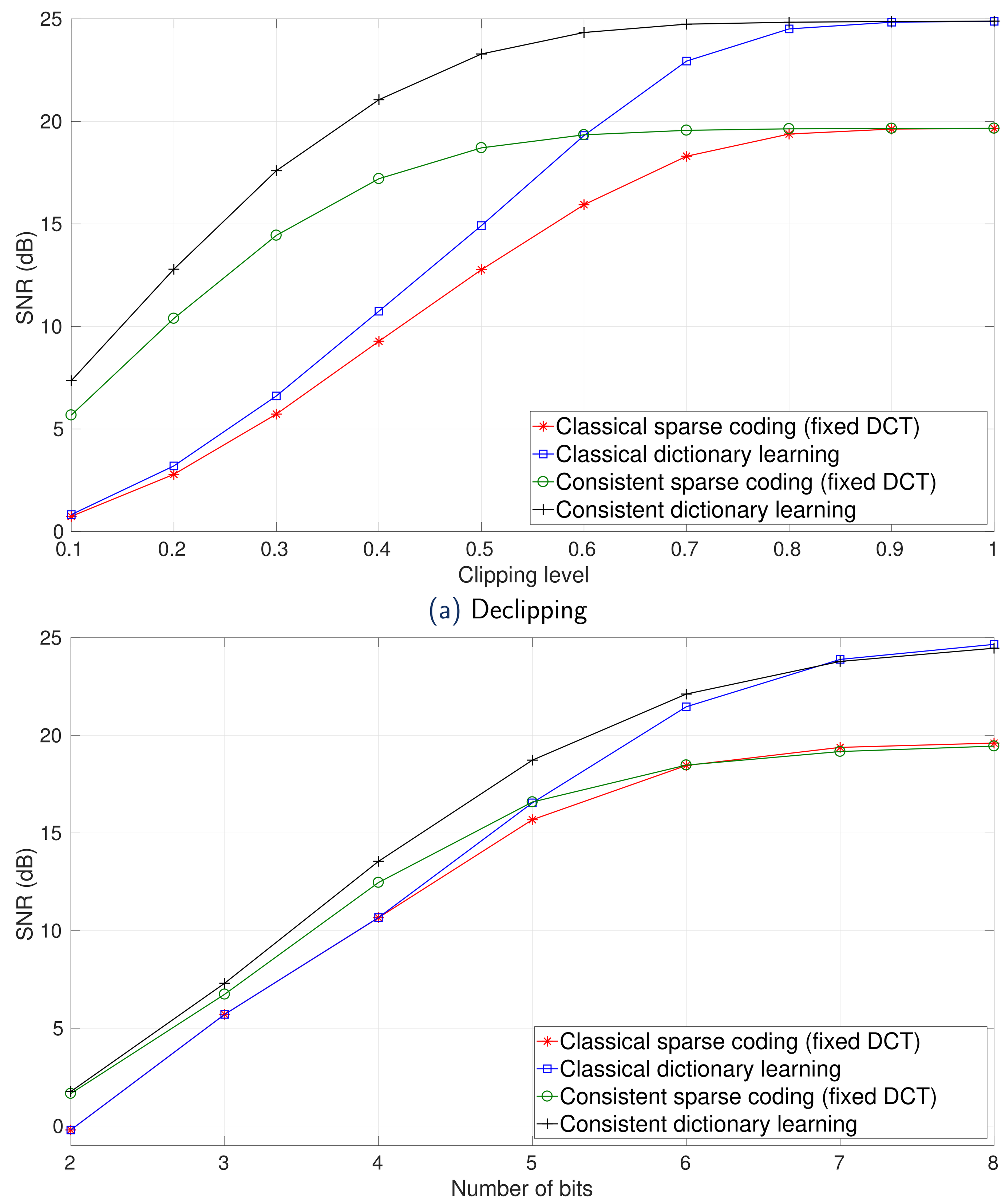
$$\boldsymbol{\alpha}_t \leftarrow \mathcal{H}_K(\boldsymbol{\alpha}_t) \tag{8}$$

    for $j = 1,\ldots,n_2$ **do** ▷ Dictionary update step

$$\mathbf{D} \leftarrow \Pi_{\mathcal{D}}\Big(\mathbf{D} + \mu_2 \sum_t (\Pi_{\mathcal{C}(\mathbf{y}_t)}(\mathbf{D}\,\boldsymbol{\alpha}_t) - \mathbf{D}\,\boldsymbol{\alpha}_t)\,\boldsymbol{\alpha}_t^T\Big) \tag{9}$$

  return $\hat{\mathbf{D}}$, $\{\hat{\boldsymbol{\alpha}}_t\}_{1\ldots T}$

Experiments



(a) Declipping

(b) De-quantization

Figure 2: Comparison of the proposed consistent sparse coding and dictionary learning algorithms, compared to classical sparse coding and dictionary learning.

References

[1] L. Rencker, F. Bach, W. Wang and M.D. Plumbley *Consistent dictionary learning for signal declipping*, Latent Variable Analysis and Source Separation Conference (LVA/ICA), Guildford, UK, 2018. Pre-print available at `http://www.cvssp.org/Personal/LucasRencker/files/consistent_DL_LVA18.pdf`

UNIVERSITY OF SURREY

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