Multi Denoising Approximate Message Passing for Optimal Recovery with Lower Computational Cost

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the 25th European Signal Processing Conference (EUSIPCO)

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Multi Denoising Approximate Message Passing for Optimal Recovery with Lower Computational Cost

Alessandro Perelli*, Mike E. Davies†
Institute for Digital Communications (IDCOM)
The University of Edinburgh (UK)
Email: *a.perelli@ed.ac.uk, †mike.davies@ed.ac.uk

Abstract—An emerging issue in large-scale inverse problems is constituted by the interdependency between computational and recovery performance; in particular in practical application, such as medical imaging, it is crucial to provide high quality estimates given bounds on computational time. While most work in this direction has gone down the lines of improving optimisation schemes, in this paper we are proposing and investigating a different approach based on a multi denoising approximate message passing (MultiD-AMP) framework for Compressive Sensing (CS) image reconstruction which exploits an hierarchy of denoisers by starting with a low fidelity model and then using the estimate as starting point for a higher fidelity models through an iterative reconstruction algorithm. MultiD-AMP achieves lower time complexity and same accuracy compared to using the same most accurate denoiser as in D-AMP. The novelty of our approach is based on exploiting the deterministic state evolution of AMP, which means the predictability of the recovery performances, to design a strategy for selecting the denoiser from a set ordered by both computational complexity and statistical efficiency. We apply the MultiD-AMP framework for image reconstruction given noisy Gaussian random linear measurements. Furthermore we extend and show the applicability of MultiD-AMP for CS to image reconstruction.

I. INTRODUCTION

The enormous growth in data size is becoming an issue, both for high dimensional inverse problem in medical imaging and large dataset in machine learning. Moreover real time applications require a certain level of accuracy to be achieved within a constrained amount of time and it imposes a new perspective in designing reconstruction algorithms compared to classical iterative methods. In machine learning problems where data is generally abundant, some works recently have considered strategies for subsampling or dynamically increase data. In particular in [1] and [2], a dynamically increasing the sample size scheme is proposed for reducing the computational complexity of iterative methods, such as stochastic gradient descent, which generally scales with the size of training samples. For optimization and denoising problems, [3] opens up a new insight to this problem, developing the concept of "algorithmic weakening" and exploiting a hierarchy of convex relaxations, ordered by both computational and statistical efficiency. However, little has been proposed in the context of Compressive Sensing and other imaging problems; some recent works suggest to utilize an hierarchy of models, within proximal gradient methods, for solving composite optimization problems with lower complexity [4]. Another work [5] examined the time-data tradeoff for image interpolation problem, by varying the amount of smoothing applied to the convex optimization problem. However most of these works have addressed the problem of computational and statistical tradeoff in terms of modifying or improving a statistical optimization scheme, while in this work we pursue a different approach based on building a more general family of denoisers. In this paper we are interested in investigating and developing a computationally fast strategy for Denoising Approximate Message Passing (D-AMP) which is a compressed sensing algorithm that exhibits fast convergence and excellent performance through the application of a sequence of sophisticated denoising procedures such as BM3D [6], [7]. D-AMP can be seen as an iterative algorithm that uses sampled data to perform a sequence of updates each of them involves a gradient step, which requires the forward and back projection, and a denoising step. These denoising steps tend to dominate the computational complexity therefore, taking inspiration from [8] we propose to exploit a hierarchy of denoisers to achieve the same overall performance but at reduced computational cost.

A. Main Contribution

D-AMP algorithm [8] has been analysed in terms of inferential accuracy without considering computational complexity. This is an important missing aspect since the denoising is often the computational bottleneck in the D-AMP reconstruction. The approach that is proposed in this paper is different, we aim to derive a mechanism for leveraging a hierarchy of signal approximations and minimize the overall time complexity [3]. The intuition is based on the observation that at earlier iterations, when the estimate $x^k$ is far according to some distance from the true $x$, the algorithm does not benefit significantly from the use of a complicated denoiser, since the attainable structure from the signal is poor. Thus almost the same performance can be achieved through simpler faster denoisers. This leads to the idea of defining a family/hierarchy of denoisers of increased complexity; the main challenge is to define a switching scheme. This is based on an empirical finding that in MultiD-AMP we can predict exactly, in the large system limit, the evolution of the MSE. We can exploit a bound on the state evolution of a training set of images to find a proper switching strategy.
II. Multi Denoising-AMP for CS System Model

In this paper we consider the reconstruction of a structured signal belonging to a certain class \( C \), i.e. \( \mathbf{x}^* \in C \), given the linear measurement

\[
y = \mathbf{A}\mathbf{x}^* + \mathbf{w},
\]

with \( M < N \), where \( y \in \mathbb{R}^M \) is the vector of measurements, \( \mathbf{A} \in \mathbb{R}^{M \times N} \) is a Gaussian random measurement matrix, \( \mathbf{x}^* \in \mathbb{R}^N \) represents the signal of interest and \( \mathbf{w} \in \mathbb{R}^M \) is the noise with \( \mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \). A denoiser is a non-linear mapping

\[
D_{\sigma}(\cdot): \mathbb{R}^N \rightarrow \mathbb{R}^N, \ x \rightarrow D_{\sigma}(x)
\]

that gives an estimate of \( \mathbf{x}^* \) given some noisy measurements \( \mathbf{x} = \mathbf{x}^* + \mathbf{\epsilon}, \mathbf{\epsilon} \sim \mathcal{N}(0, \mathbf{I}) \) and \( \sigma \) is the standard deviation of the noise. The proposed MultiD-AMP algorithm follows the D-AMP iteration:

\[
\begin{align*}
\mathbf{x}^{k+1} &= D_{\sigma_k}^{h_k}(\mathbf{x}^k + \mathbf{A}^T \mathbf{z}^k) \\
\mathbf{z}^k &= y - \mathbf{A}\mathbf{x}^k + \frac{N}{M} < (D_k^k)_{\sigma_k}(\mathbf{x}^k + \mathbf{A}^T \mathbf{z}^k) > \\
\sigma^k &= \frac{||\mathbf{z}^k||^2}{M}
\end{align*}
\]

where \( \mathbf{x}^k \) is the estimate for \( \mathbf{x}^* \), \( \mathbf{z}^k \) is the estimate of the residual in the measurement domain at the \( t \)-th iteration and \( < (D_k^k)_{\sigma_k}(\cdot) > \) denotes the divergence of the denoiser. It is important to stress that the denoising functions \( D_{\sigma_k}^{h_k} \) are dependent on \( t \), i.e. for a sequence of denoisers \( D_{\sigma_k}^{h_k} \), which belong to a set \( \mathcal{D} \), \( D_{\sigma_k}^{h_k} \in \mathcal{D} \), \( \forall k \).

Generally our approach can be formulated in terms of the time-data-risk class

\[
\{\eta(N), M(N), R(N)\}
\]

of estimation problems where we want to estimate a \( N \)-dimensional signal with a risk \( R(N) \), given \( M(N) \) measurements and with computational time \( \eta(N) \).

The main challenge is to develop a switching strategy to achieve the same accuracy as the most accurate denoiser and lower computational time, given the set of denoisers \( \mathcal{D} \) ordered in terms of the risk and time complexity. The key intuition relies on exploiting a method for predicting the performance recovery for each \( t \) for the set of denoisers \( \mathcal{D} \). This can be obtained by the State Evolution (SE) of MultiD-AMP algorithm which represents a tool to predict the expected Mean Square Error (MSE) at each iteration.

While it is possible to use the MSE estimate from the SE within the D-AMP algorithm to select the best performing denoiser at each iteration, this does not account for the computation cost of each denoiser. We therefore exploit the SE given a set of trained images to determine which denoiser to select at each stage.

In the next Section we will detail how to construct an hierarchy of denoisers based on the risk and complexity and we will develop a possible switching rule based on the deterministic SE.

III. Hierarchy of Denoisers

Before introducing the SE we need to define the following Definition 1: The Risk of a denoiser \( D_{\sigma}(\cdot) \) is defined as

\[
R(\mathbf{x}^*, \sigma^2 | D) = \frac{\mathbb{E}[||D_{\sigma}(\mathbf{x}^* + \sigma \mathbf{\epsilon}) - \mathbf{x}^*||^2]}{N}
\]

where the expectation is with respect to \( \mathbf{\epsilon} \sim \mathcal{N}(0, \mathbf{I}) \).

In this work we consider the set \( \mathcal{D} \) which contains \( n \) denoisers with the property to be proper for \( \mathbf{x}^* \in C \), i.e.

\[
\sup_{x^* \in C} R(\mathbf{x}^*, \sigma^2 | D) \leq \kappa \sigma^2
\]

and monotone

\[
\forall \mathbf{x}^*, \forall \sigma^2 \leq \sigma^2, R(\mathbf{x}^*, \sigma^2 | D) \leq R(\mathbf{x}^*, \sigma^2 | D)
\]

For this class of denoisers, we need to define an ordering for the risk and complexity. Each denoiser in the \( n \)-dimensional set \( \mathcal{D} \) is indexed through 2 parameters: its risk \( R \) and the time complexity \( \eta \).

\[
\mathcal{D} = \{ D_1^1(R^1, \eta_1), \ldots, D_n^k(R^k, \eta_k), \ldots, D_n^n(R^n, \eta_n) \}
\]

We propose a nested hierarchy of increasingly tighter approximations ordered over the risk

\[
\sup_{x^* \in C} R(\mathbf{x}^*, \sigma^2 | D^1) < \ldots < \sup_{x^* \in C} R(\mathbf{x}^*, \sigma^2 | D^n)
\]

and we expect that the runtime is inversely proportional to the risk, i.e.

\[
\eta_1 < \ldots < \eta_n
\]

The overall time of the algorithm at iteration \( t \) is

\[
T(k) = \eta_i(k) + T(k - 1)
\]

where \( k \in \mathbb{N}_+ \), \( T \) is a function \( T : \mathbb{N} \rightarrow \mathbb{R} \) and \( \eta_i(k) \) denotes the \( i \)-th denoiser utilized at iteration \( t \).

The hierarchy of denoisers for the class, defined through the relations (9) and (10), is one of the possible choices. For example, if we relax the property of the denoiser to be proper, considering the class of bounded denoisers, i.e. \( ||D_{\sigma}(\mathbf{x}) - \mathbf{x}||^2/N < C \sigma^2 \), \( \forall \mathbf{x} \), or non monotone denoisers, a different ordering in term of risk has to be defined.

Our approach aims to achieve the accuracy given by the denoiser with minimum risk in the set, \( D_{\sigma}^n(R^n, \eta_n) \), with a computational time \( \eta < t \cdot \eta_n \) where \( t \) is the number of iterations. The strategy to determine the \( d \) optimal points

\[
t^* = \{t_1, \ldots, t_d\}
\]

across iterations where to switch denoiser is based on the prediction of the dynamics of MultiD-AMP.
The geometrical interpretation for the switching rule (18) is shown in Fig. 1: for each interval $[T(k-1), T(k)]$, the absolute value of the function (17) represents the slope of the tangent line to the function $\theta$ at a point $T(k_d)$. $T(k-1) < T(k_d) < T(k)$. Therefore the relation (18) indicates that we greedily decide to change denoiser when the rate of decrease of $\theta$ is less than $\frac{1}{4}$ of the previous interval.

![Fig. 1. Geometrical interpretation of the switching rule.](image)
It is worth to highlighting how our strategy yields to exploit the low complexity of DWT at early iterations when the MSE error is generally high and then using the high accurate BM3D denoiser at the cost of higher computational time.

Additionally the proposed MultiD-AMP framework has been applied using as hierarchy of denoisers the Fast Multi-Level Wavelet structure described in [9]. The CS system model is described by the algebraic relation (1) and $x$ is represented through the synthesis wavelet basis $W$, $x = W\xi$. We consider the MultiD-AMP algorithm described by the following update rule

$$x^{k+1} = (W^T_{\sigma^k}W)_{ks}(x^k + A^Tz^k)$$

$$z^k = y - Ax^k + \frac{N}{M} < (W^T_{\sigma^k}W)_{ks}'(x^k + A^Tz^k)$$

$$\sigma^k = \frac{||z^k||_2^2}{M}$$

where $(W^T_{\sigma^k}W)_{ks}$ is dependent on the wavelet sub-bands $s$, which are changing across iterations $k$, by the multilevel V-cycle scheme described in [9] and $T_{\eta^k}$ is the soft thresholding function defined for $x \in C$ as

$$T_{\eta}(x) = \text{sgn}(x) \max(|x| - \eta, 0) \quad (20)$$

The MultiLevel Wavelet AMP has been tested with the same setup used for the previous simulation; Fig. 4 shows the performance of the MSE in actual time for the DWT-AMP and the MultiLevel Wavelet - AMP (both with soft thresholding non linear function). We can notice how exploiting a multigrid procedure in the denoising step, i.e. considering a scheme for selecting sub-bands, yields to a reduction in time complexity especially at earlier iterations; the blue curve correspond to the blue curve in Fig. 3 (DWT) and as expected both the MultiLevel and DWT achieve the same MSE at convergence.
VII. CONCLUSIONS

In this paper we have proposed a new approach in terms of risk-time tradeoffs for reduce the time complexity of the family of D-AMP iterative reconstruction algorithms. In particular a strategy for designing an hierarchy of denoisers in MultiD-AMP has been proposed and successfully tested. It is important to highlight that this framework allows the dynamic switching between denoisers based on the State Evolution which provide a powerful way to predict the performances of MultiD-AMP. Furthermore, we proposed a framework following the idea of [3] for image reconstruction we have build a ordered set of denoisers which constitutes a more general scenario than convex nested set. Finally we have applied the proposed MultiD-AMP concept with a MultiLevel Wavelet denoiser [9] which yields to an improvement in time complexity compared to the full DWT.

ACKNOWLEDGMENT

The authors would like to acknowledge the support from ERC Advanced grant, project 694888, C-SENSE.

REFERENCES