MultiD-AMP: match up Accuracy and Fast Computation by Dynamically Denoising Data

Alessandro Perelli and Mike E. Davies
Institute for Digital Communications (IDCOM), The University of Edinburgh, EH9 3JL, UK

***Abstract*—Denoising AMP (D-AMP) is as an iterative algorithm where at each iteration a non-linear denoising function is applied to the signal estimate. We aim to design a hierarchical denoising model (MultiD-AMP) in order to minimise the complexity given the expected risk. Our intuition is that at earlier iterations we can exploit less complex denoisers since the estimate is far from the true signal. This idea has been tested on i.i.d. random Gaussian measurements with Gaussian noise and the results show the effectiveness of the proposed reconstruction algorithm.**

I. Main Contribution

D-AMP algorithm [1] 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 it 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 [2]. The intuition comes from the observation that at earlier iteration, when the estimate $\mathbf{x}^t$ is far according to some distance to the true $\mathbf{x}^*$, the algorithm does not need a complicated denoiser, since the structure of the signal is poor, but 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 the empirical finding that in MultiD-AMP we can predict exactly, in the large system limit, the evolution of the MSE. We can exploit the State Evolution (SE), evaluated on a set of training images, to find a proper switching strategy.

II. MultiD-AMP

In this paper we consider the reconstruction of a structured signal belonging to a certain class $\mathcal{C}$, i.e. $\mathbf{x}^* \subset \mathcal{C}$, given linear measurement $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{w}$ where $\mathbf{y} \in \mathbb{R}^M$, $\mathbf{x} \in \mathbb{R}^N$, $M < N$ with $\mathbf{w} \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I})$. A denoiser is a non-linear mapping $D_\sigma(\cdot) : \mathbb{R}^N \to \mathbb{R}^N$, $\mathbf{x}^* \to D_\sigma(\mathbf{x})$ that produces an estimate of $\mathbf{x}^*$, given some noisy measurements $\mathbf{x} = \mathbf{x}^* + \sigma\epsilon$ with $\epsilon \sim \mathcal{N}(0, I)$, and σ is the standard deviation of the noise. The proposed MultiD-AMP algorithm follows this iteration:

$$\begin{aligned} \mathbf{x}^{t+1} &= D^t_{\sigma_t}(\mathbf{x}^t + \mathbf{A}^T\mathbf{z}^t), \quad \sigma^t = \frac{||\mathbf{z}^t||_2^2}{M} \\ \mathbf{z}^t &= \mathbf{y} - \mathbf{A}\mathbf{x}^t + \frac{N}{M}\mathbf{z}^{t-1} < (D^t)'_{\sigma_t}(\mathbf{x}^{t-1} + \mathbf{A}^T\mathbf{z}^{t-1}) > \end{aligned} \quad (1)$$

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 complexity, given a set of ξ denoisers, $\mathcal{D}$, ordered in terms of the risk and time complexity.
We need to define a set of denoisers ordered in terms of the risk and time complexity.

Definition 1: The Risk of a denoiser is defined as

$$R(\mathbf{x}^*, \sigma^2|D) = \frac{\mathbb{E}\Big[||D_\sigma(\mathbf{x}^* + \sigma\epsilon) - \mathbf{x}^*||_2^2\Big]}{N} \quad (2)$$

where the expectation is with respect to $\epsilon \sim \mathcal{N}(0, I)$.
We propose a nested hierarchy of increasingly tighter approximations ordered over the risk $\sup_{\mathbf{x}^* \in \mathcal{C}} R(\mathbf{x}^*, \sigma^2|D^\xi) < \ldots < \sup_{\mathbf{x}^* \in \mathcal{C}} R(\mathbf{x}^*, \sigma^2|D^1)$ and we expect that the runtime is inversely proportional to the risk, i.e. $\eta_1 < \ldots < \eta_\xi$. Of course this is one of the possible choices of the set of denoisers. The strategy to determine the optimal points t^* across iterations where to switch denoiser is based on the prediction of the dynamics of MultiD-AMP.

Finding 1: [1] The dynamic of MultiD-AMP is captured by the SE which coincide with the risk of $D^t_{\sigma_t}(\cdot)$

$$\theta^t(\mathbf{x}^*, \delta, \sigma_w^2, D^t_{\sigma^t}) = R(\mathbf{x}^*, (\sigma^t)^2, D^t_{\sigma^t}) = \lim_{n\to\infty} \frac{1}{N}\|\mathbf{x}^t - \mathbf{x}^*\|_2^2$$

where we highlighted the dependency of θ^t on the true signal and the undersampling ratio $\delta = \frac{M}{N}$. Then for finite δ and $N \to \infty$ the SE predicts the MSE of MultiD-AMP
Therefore, the estimation error is connected, in the large system limit, with the risk and the overall complexity is $T(t) = \eta_{i(t)} + T(t-1)$. At this point we are ready to define the switching strategy as a function of the risk $R(\cdot)$ and discrete time $T(t)$, i.e. $g(\theta(\cdot), T(t))$

Given a training set of data, we generate the SE and for each $T(t)$ (time T continuous variable at iteration t discrete variable) we need to evaluate the following function over discrete points

$$g(\theta^{T(t)}(\mathbf{x}^*, D^{T(t)}_{\sigma^{T(t)}}), T(t)) = \frac{\theta^{T(t)}(\cdot) - \theta^{T(t-1)}(\cdot)}{T(t) - T(t-1)} = \theta^{T(t_a)}(\cdot)$$

with $t - 1 < t_a < t$ (Lagrange theorem) as in Fig. (2a) or interpreted as the angular coefficient of the straight line intersecting the 2 points. The chosen greedy criteria is to switch $D_\sigma(\cdot)$ when $|g(\theta^{T(t+1)}(\cdot), T(t+1))| < |g(\theta^{T(t)}(\cdot), T(t))/4|$.

III. Results

MultiD-AMP has been tested on 256×256 Lena image with i.i.d. Gaussian random measurement in Fig. (1), Gaussian noise and 0.2 undersampling ratio. In MultiD-AMP we have used l_1-wavelet and BM3D [3] denoisers. The MSE/time plot is shown in Fig. (2b). In the experiment 4 ”traditional” images, boat, barbara, house, peppers have been used as training images; the switching point t^* is obtained with the greedy procedure on the SE and the mean of t_i^*, selected in the SE of each training image, is the one used in the reconstruction.

IV. Conclusions

A strategy for reducing the computational cost using and hierarchy of denoisers in D-AMP has been proposed and successfully tested.
The main idea is to exploit the SE to find the switching strategy and the sensitivity of this method on the training set will be investigated. We plan to extend MultiD-AMP with multilevel denoising schemes, like with l_1-wavelet [4], and apply to real application such as CT imaging.

Acknowledgment

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

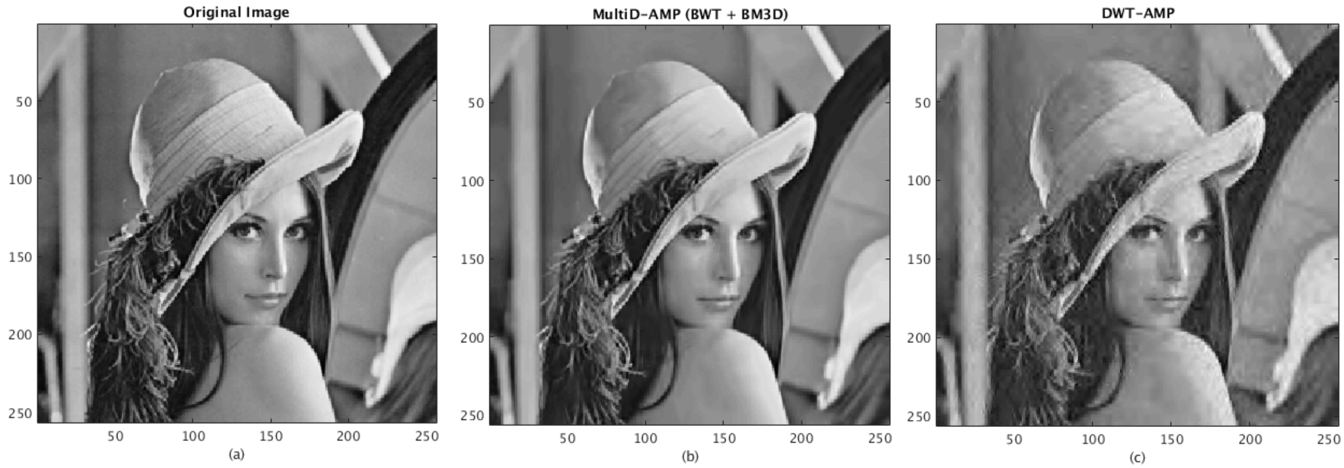
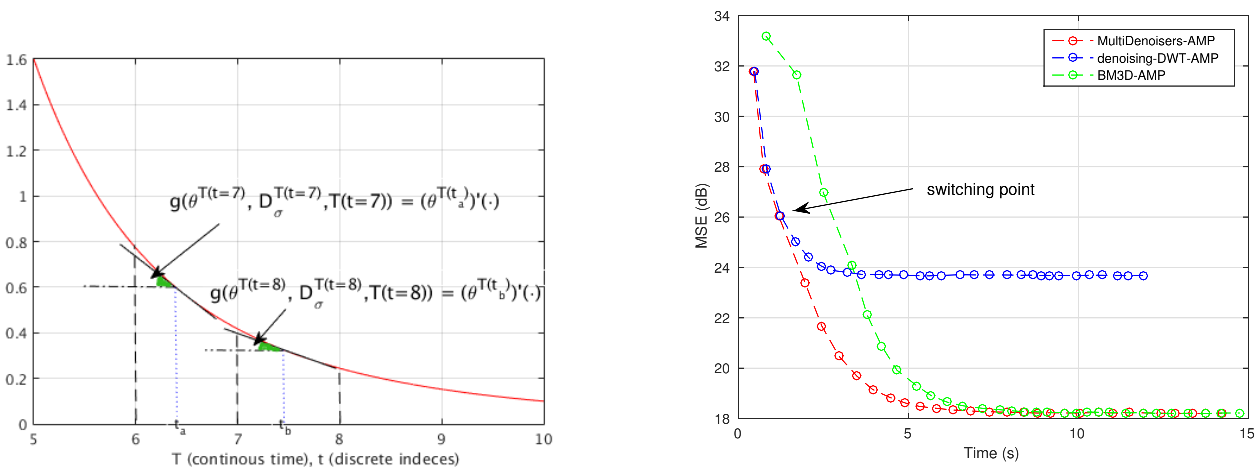


Fig. 1. CS reconstruction problem: (a) Original image, (b) MultiD-AMP, (c) DWT-AMP



(a) (b)

Fig. 2. (a) Geometrical interpretation of the switching rule, (b) Evaluation MSE versus computational time for MultiD-AMP, DWT-AMP, BM3D-AMP.

REFERENCES

[1] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5117–5144, 2016.

[2] Venkat Chandrasekaran and Michael I Jordan, "Computational and statistical tradeoffs via convex relaxation," *Proceedings of the National Academy of Sciences*, vol. 110, no. 13, pp. E1181–E1190, 2013.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[4] Cédric Vonesch and Michael Unser, "A fast multilevel algorithm for wavelet-regularized image restoration," *IEEE Transactions on Image Processing*, vol. 18, no. 3, pp. 509–523, 2009.