AMP Methods for Fourier-Structured Operators and Signals

Saurav K. Shastri (OSU), Rizwan Ahmad (OSU), Christopher A. Metzler (UMD), and Philip Schniter (OSU)

THE OHIO STATE UNIVERSITY

UNIVERSITY OF MARYLAND

Supported in part by NSF CCF-1955587 and NIH R01-EB029957

Asilomar Conference on Signals, Systems, and Computers
Oct 30 - Nov 2, 2022

The linear inverse problem

Goal: Recover an unknown signal $\boldsymbol{x}_0 \in \mathbb{C}^N$ from noisy measurements $\boldsymbol{y} \in \mathbb{C}^M$ of the form

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}, \text{ with } \begin{cases} \boldsymbol{A} : \text{linear measurement operator} \\ \boldsymbol{w} : \text{AWGN with precision } \gamma_w \end{cases}$$

Typical methodologies:

- Optimization based algorithms
  - Simple and robust, but not state-of-the-art in accuracy
- Deep networks that recover $\boldsymbol{x}$ from $\boldsymbol{y}$
  - Accurate but may not generalize well to a different $\boldsymbol{A}$
- Plug-and-play algorithms that call deep denoisers
  - Accurate and handles any $\boldsymbol{A}$, but its performance can be improved!

Optimization-based recovery

- The classical approach to recovering $\boldsymbol{x}_0$ is through optimization:

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{ g_1(\boldsymbol{x}) + g_2(\boldsymbol{x}) \right\} \text{ with } \begin{cases} g_1(\boldsymbol{x}): & \text{data fidelity loss} \\ g_2(\boldsymbol{x}): & \text{regularization} \end{cases}$$

- Common choice for data-fidelity term: $g_1(\boldsymbol{x}) = \frac{\gamma_w}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2$

- Common choice for regularization: $g_2(\boldsymbol{x}) = \lambda \|\boldsymbol{\Psi}\boldsymbol{x}\|_1$ with a suitable sparsifying transform $\boldsymbol{\Psi}$ (e.g., wavelet or total-variation) and carefully chosen $\lambda > 0$

Plug-and-play (PnP) image recovery

- A common approach to convex optimization is ADMM: For $k = 1, 2, \ldots$

$$\boldsymbol{x}_k = \arg\min_{\boldsymbol{x}} \left\{ g_1(\boldsymbol{x}) + \frac{\beta}{2} \|\boldsymbol{x} - \boldsymbol{v}_{k-1} + \boldsymbol{u}_{k-1}\|^2 \right\}$$

$$\boldsymbol{v}_k = \arg\min_{\boldsymbol{v}} \left\{ g_2(\boldsymbol{v}) + \frac{\beta}{2} \|\boldsymbol{v} - \boldsymbol{x}_k + \boldsymbol{u}_{k-1}\|^2 \right\} \triangleq \text{prox}_{g_2/\beta}(\boldsymbol{x}_k - \boldsymbol{u}_{k-1})$$

$$\boldsymbol{u}_k = \boldsymbol{u}_{k-1} + \boldsymbol{x}_k - \boldsymbol{v}_k$$

- The prox performs denoising (eg, soft-thresholding when $g_2(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$)

- Bouman et al. proposed PnP[1] ADMM, where the prox is replaced by a sophisticated image denoiser $\boldsymbol{f}(\cdot)$ like BM3D or a deep image denoiser[2]

[1] Venkatakrishnan,Bouman,Wolhberg'13, [2] Meinhardt,Moller,Hazirbas,Cremers'17

Challenges in plug-and-play (PnP) image recovery

- In PnP, the denoiser input-error is difficult to characterize. For example, it is non-white, non-Gaussian, and has iteration-dependent statistics
- PnP algs require careful tuning of parameters (eg, β) and early stopping
- Also, it's unclear how to optimally train the denoiser in PnP
  - Typically the denoiser is trained to remove AWGN
  - Gilton et al. recently proposed[1] to train the denoiser at the PnP equilibrium point, but the result is $\boldsymbol{A}$-dependent and thus may not generalize

[1] Gilton,Ongie,Willet'21

Motivating questions

- Is it possible to design a PnP-style algorithm that presents the denoiser with known error statistics at every iteration?

- Is it possible to construct a deep denoiser that can efficiently leverage those error statistics?

Approximate message passing (AMP) algorithms

- AMP[1] is a family of autotuning PnP algorithms that have remarkable properties for large random $\boldsymbol{A}$:
  - The denoiser input-error is AWGN with predictable variance[2]
  - When used with the MMSE denoiser, AMP algorithms converge[3] to the MMSE estimate of $\boldsymbol{x}_0$ given $\boldsymbol{y}$
- Challenge: In most image recovery problems, $\boldsymbol{A}$ does not satisfy AMP's randomness assumptions!
  - Recent work[4] has studied AMP with nearly deterministic $\boldsymbol{A}$ under i.i.d. $\boldsymbol{x}_0$, but our problems of interest have structured $\boldsymbol{x}_0$

[1] Donoho et al'09, [2] Bayati,Montanari'11, [3] Berthier,et al'19, [4] Dudeja et al'22

AMP for parallel MRI

In this work, we focus on the Fourier-structured matrix and images encountered in parallel magnetic resonance imaging (MRI)

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{M}\boldsymbol{F}\,\mathrm{Diag}(\boldsymbol{s}_1) \\ \vdots \\ \boldsymbol{M}\boldsymbol{F}\,\mathrm{Diag}(\boldsymbol{s}_C) \end{bmatrix} \quad \text{where} \quad \begin{cases} \boldsymbol{M} = \text{sampling mask} \\ \boldsymbol{F} = \text{2D Fourier transform} \\ \boldsymbol{s}_c = \text{ESPIRiT-estimated coil map} \end{cases}$$

For MRI, damped AMP techniques have been proposed:

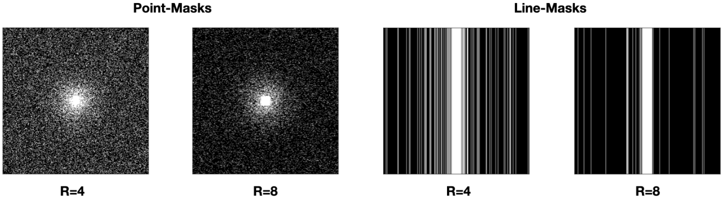
- Denoising AMP (D-AMP)[1]
- Damped denoising vector-AMP (DD-VAMP)[2]

but they are heuristic and don't appear to follow a state evolution

[1]Eksioglu,Tanc'18, [2]Sarkar,Ahmad,Schniter'21

AMP for MRI with 2D point masks

- For MRI with 2D point masks, modified VAMP algs were proposed: VDAMP[1] and P-VDAMP[2]
  - they recover wavelet-domain coefficients, not the image itself
  - they use wavelet thresholding instead of deep denoising
  - they yield AWGN denoiser input error in each subband
- Later the above approaches were extended to deep image denoising by D-VDAMP[3] and PD-VDAMP[4]
- But 2D point masks are impractical and uncommon in 2D MRI

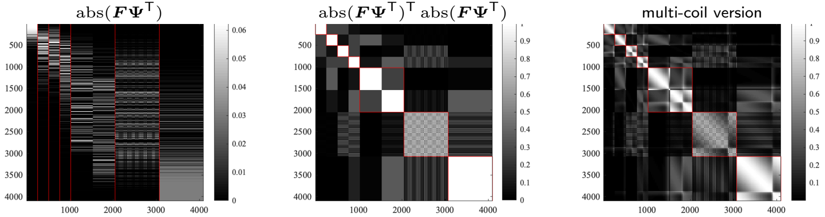


[1]Millard et al'20, [2]Millard et al'22, [3]Metzler,Wetzstein'21, [4]Millard et al'22b

Why recover wavelet coefficients?

- Suppose $\boldsymbol{c}_0 = \boldsymbol{\Psi}\boldsymbol{x}_0$ are coefficients of an orthogonal wavelet transform
- Can rewrite $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}$ as

  $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{c}_0 + \boldsymbol{w}$ with masked Fourier-wavelet matrix $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{\Psi}^{\mathsf{T}}$
- For AMP algorithms, $\boldsymbol{B}$ has desirable behavior:[1]
  - columns of different subbands are relatively decoupled from eachother
  - columns of each subband have a randomizing effect on that subband



[1] Schniter,Rangan,Fletcher'17

Proposed algorithm: Denoising GEC (D-GEC)

We build upon the generalized expectation consistency (GEC) algorithm:[1]

require: $\boldsymbol{f}_1(\cdot)$, $\boldsymbol{f}_2(\cdot)$, and $\mathrm{gdiag}(\cdot)$

initialize: $\boldsymbol{r}_1, \gamma_1$

for $t = 0, 1, 2, \ldots$

$\widehat{\boldsymbol{x}}_1 \leftarrow \boldsymbol{f}_1(\boldsymbol{r}_1, \gamma_1)$ — linear estimation

$\boldsymbol{\eta}_1 \leftarrow \mathrm{Diag}(\mathrm{gdiag}(\nabla \boldsymbol{f}_1(\boldsymbol{r}_1, \boldsymbol{\gamma}_1)))^{-1}\boldsymbol{\gamma}_1$

$\boldsymbol{\gamma}_2 \leftarrow \boldsymbol{\eta}_1 - \boldsymbol{\gamma}_1$

$\boldsymbol{r}_2 \leftarrow \mathrm{Diag}(\boldsymbol{\gamma}_2)^{-1}(\mathrm{Diag}(\boldsymbol{\eta}_1)\widehat{\boldsymbol{x}}_1 - \mathrm{Diag}(\boldsymbol{\gamma}_1)\boldsymbol{r}_1)$ — Onsager

$\widehat{\boldsymbol{x}}_2 \leftarrow \boldsymbol{f}_2(\boldsymbol{r}_2, \boldsymbol{\gamma}_2)$ — denoising

$\boldsymbol{\eta}_2 \leftarrow \mathrm{Diag}(\mathrm{gdiag}(\nabla \boldsymbol{f}_2(\boldsymbol{r}_2, \boldsymbol{\gamma}_2)))^{-1}\boldsymbol{\gamma}_2$

$\boldsymbol{\gamma}_1 \leftarrow \boldsymbol{\eta}_2 - \boldsymbol{\gamma}_2$

$\boldsymbol{r}_1 \leftarrow \mathrm{Diag}(\boldsymbol{\gamma}_1)^{-1}(\mathrm{Diag}(\boldsymbol{\eta}_2)\widehat{\boldsymbol{x}}_2 - \mathrm{Diag}(\boldsymbol{\gamma}_2)\boldsymbol{r}_2)$ — Onsager

[1] Fletcher,Sahraee-Ardakan,Rangan,Schniter'16

Proposed algorithm: Denoising GEC (D-GEC)

- GEC is a version of VAMP[1] that tracks different subsets of coefficients using distinct variances
- Can be interpreted as Peaceman-Rachford ADMM with adaptive vector-valued stepsizes $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$
- The GEC linear estimation stage is preconditioned LS:

$$\boldsymbol{f}_1(\boldsymbol{r},\boldsymbol{\gamma}) = \big(\gamma_w \boldsymbol{B}^{\mathsf{H}}\boldsymbol{B} + \mathrm{Diag}(\boldsymbol{\gamma})\big)^{-1}\big(\gamma_w \boldsymbol{B}^{\mathsf{H}}\boldsymbol{y} + \mathrm{Diag}(\boldsymbol{\gamma})\boldsymbol{r}\big)$$

  which can be implemented using the conjugate gradient method
- For $\boldsymbol{f}_2$, we propose to "plug in" a deep denoiser
- For the MRI application, we will show that D-GEC yields per-subband denoiser input-errors that are AWGN with a predictable variance

[1] Rangan,Fletcher,Schniter'16

D-GEC: Jacobian computation

- $\nabla \boldsymbol{f}_i$ denotes the Jacobian, and $\text{gdiag}(\cdot)$ averages its diagonal across L wavelet subbands using:

$$\text{gdiag}(\boldsymbol{Q}) \triangleq [d_1 \mathbf{1}_{N_1}^{\mathsf{T}}, \ldots, d_L \mathbf{1}_{N_L}^{\mathsf{T}}]^{\mathsf{T}},\ d_\ell = \frac{\text{tr}\{\boldsymbol{Q}_{\ell\ell}\}}{N_\ell},$$

  where N_ℓ is the size of the ℓth subset and $\boldsymbol{Q}_{\ell\ell} \in \mathbb{R}^{N_\ell \times N_\ell}$ is the ℓth diagonal subblock of the matrix input $\boldsymbol{Q}$

- D-GEC approximates the Jacobian using a Monte-Carlo approach[1]
  - For both $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$, we approximate the $\text{tr}\{\boldsymbol{Q}_{\ell\ell}\}$ using

$$\text{tr}\{\boldsymbol{Q}_{\ell\ell}\} \quad \approx \quad \delta^{-1} \boldsymbol{q}_\ell^{\mathsf{H}} \big[\boldsymbol{f}_i(\boldsymbol{r} + \delta \boldsymbol{q}_\ell, \boldsymbol{\gamma}) - \boldsymbol{f}_i(\boldsymbol{r}, \boldsymbol{\gamma}) \big]$$

    where the ℓth coefficient subset in $\boldsymbol{q}_\ell$ is i.i.d. unit-variance Gaussian and the other coefficient subsets are zero

[1] Ramani,Blu,Unser'08

Proposed denoiser: Corr+Corr

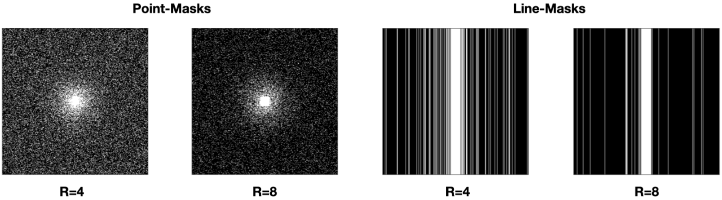
- GEC yields denoiser input-error that is AWGN with known iteration- and subband-dependent precisions γ *in each wavelet subband*
  - In the *pixel* domain, the error is correlated Gaussian with known covariance matrix $\boldsymbol{\Psi}\,\mathrm{Diag}(\boldsymbol{\gamma})^{-1}\boldsymbol{\Psi}^\mathsf{T}$
  - How do we design a deep denoiser to remove this correlated noise?

- We take an arbitrary existing denoiser (e.g., DnCNN) and feed independent realizations of $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}\,\mathrm{Diag}(\boldsymbol{\gamma})^{-1}\boldsymbol{\Psi}^\mathsf{T})$ into extra channels
  - The denoiser learns to extract the error statistics $(\boldsymbol{\Psi}, \boldsymbol{\gamma})$ and use them productively for denoising!
  - In practice, we find that one extra channel suffices

Parallel-MRI experiments

Setup:

- fastMRI[1] brain and knee data
- 8 virtual coils
- acceleration $R = N/M = 4 \ \& \ 8$
- extra AWGN $\boldsymbol{w}$ for noise-robustness study
- variable-density 2D point- and line-masks:



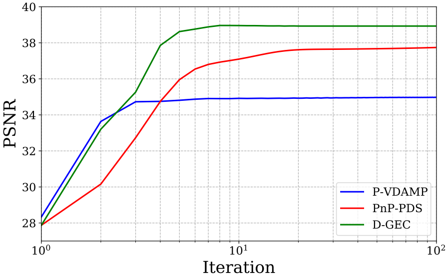
[1] Zbontar et al'18

Average performance results

2D line-mask results averaged over 16 test images:

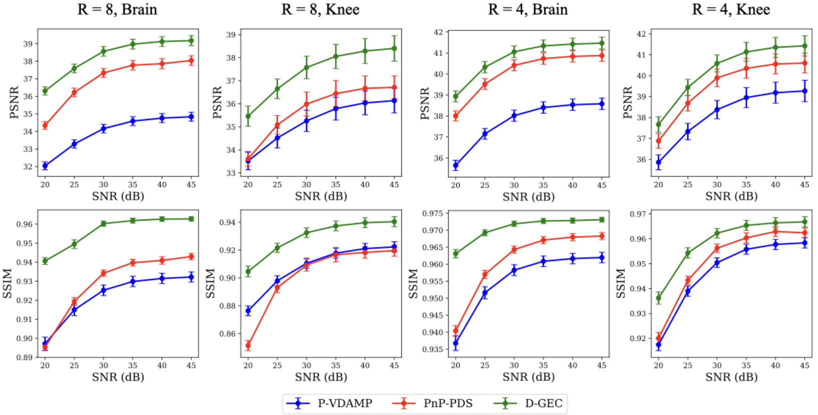
	Knee				Brain			
	$R = 4$		$R = 8$		$R = 4$		$R = 8$	
method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
P-VDAMP	33.84	0.9018	20.34	0.5614	30.30	0.8847	13.51	0.4763
PnP-PDS	36.28	0.9204	32.34	0.8556	38.07	0.9501	28.97	0.8269
D-GEC	**38.82**	**0.9504**	**33.66**	**0.8893**	**39.04**	**0.9631**	**30.61**	**0.9015**

PSNR vs iterations for brain MRI recovery with 2D point-mask at $R = 4$:



Robustness to measurement noise

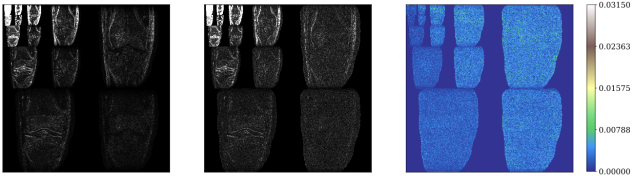
Average PSNR and SSIM versus measurement SNR with 2D point-mask:



Note: PnP-PDS penalty and stopping iteration tuned for every (SNR,R,dataset)

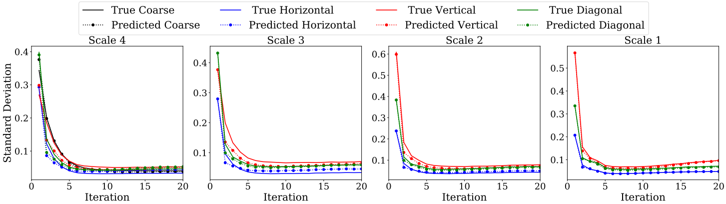
Example D-GEC behavior ($R = 4$, 2D line-mask)

Example wavelet error at iteration 10:



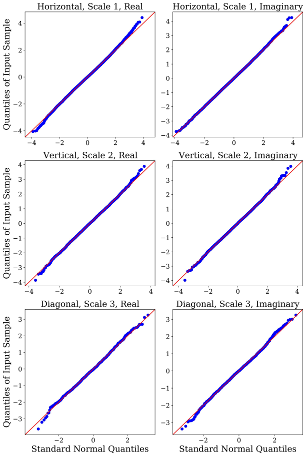
$|\boldsymbol{c}_0|$ $|\boldsymbol{r}_2|$ $|\boldsymbol{c}_0 - \boldsymbol{r}_2|$

Standard deviation of D-GEC denoiser-input error vs iteration:

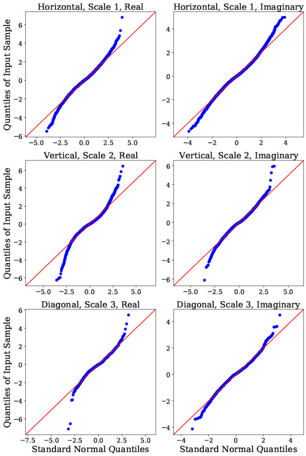


Example behavior of D-GEC vs PnP-PDS

Denoiser input-error QQ plots at iteration 10, demonstrating Gaussianity:



D-GEC



PnP-PDS

Summary

- We proposed a GEC-based PnP algorithm for MRI called D-GEC
- Our algorithm yields denoiser-input error that behaves like AWGN with predictable variance in each wavelet subband
- We proposed a new corr+corr denoiser, which aims to remove the resulting *colored* pixel-domain noise
- Empirical results demonstrate that D-GEC yields significantly better recovery PSNR and SSIM than PnP-PDS and existing AMP-based algorithms on multicoil fastMRI data