Expectation Consistent Plug-and-Play for MRI

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

The Ohio State University

University of Maryland

Supported by NSF 1955587, NIH 135489, NIH 029957

IEEE International Conference on Acoustics, Speech, and Signal Processing — May 8, 2022

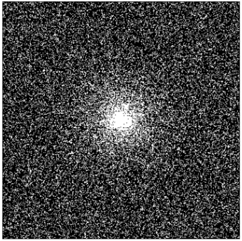
Magnetic Resonance Imaging (MRI)

Challenge:

- An MRI scan can take more than 45 minutes
- To accelerate MRI, it is common to sample far below the Nyquist rate

Measurement model: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}$

- Single coil: $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{F}$
- $\boldsymbol{F} \in \mathbb{C}^{N \times N}$: 2D-DFT matrix
- $\boldsymbol{M} \in \mathbb{R}^{M \times N}$: Sampling mask
- $\boldsymbol{w}$: AWGN with precision γ_w



A variable-density sampling mask $\boldsymbol{M}$ with acceleration $R = \frac{N}{M} = 4$

The Linear Regression Problem

Measurement model: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}$

Goal: Recover the unknown image $\boldsymbol{x}_0 \in \mathbb{C}^N$ from noisy k-space measurements $\boldsymbol{y} \in \mathbb{C}^M$ with $M \ll N$

Typical Methodologies:

- Optimization based algorithms
  - Simple, but poor recovery
- Train a deep network to recover $\boldsymbol{x}$ from $\boldsymbol{y}$
  - Excellent recovery, but may not generalize well to a different $\boldsymbol{A}$
- Hybrid: Plug-and-Play
  - Excellent recovery and handles any $\boldsymbol{A}$, but its performance can be improved!

Optimization-Based Recovery

- A common approach[1] to recovering MRI image $\boldsymbol{x}$ is through optimization:

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

- Typical choice for loss function: $g_1(\boldsymbol{x}) = \frac{\gamma_w}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2$

- Typical 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 $\lambda > 0$

[1] Lustig et al. '08

Optimization-Based Recovery

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

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

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

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

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

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

[2] Venkatakrishnan,Bouman,Wolhberg '13

Plug-and-Play (PnP) Image Recovery

- A more sophisticated deep-net image denoiser can also be used in PnP, which can be trained ...
  - from very few images, using patches
  - independently of $\boldsymbol{A}$, facilitating generalization to any $\boldsymbol{A}$

- Challenge: In PnP, the denoiser input-error statistics are iteration-dependent and difficult to characterize. For example, they are generally non-white and non-Gaussian

- Thus, it's not clear how to train the denoiser for optimal performance in PnP!
  - Typically the denoiser is trained with AWGN
  - Gilton et al. recently proposed[3] to train the denoiser at the PnP equilibrium point, but it's $\boldsymbol{A}$-dependent and thus may not generalize

[3] Gilton, Ongie, Willet '21

Approximate Message Passing (AMP) Algorithms

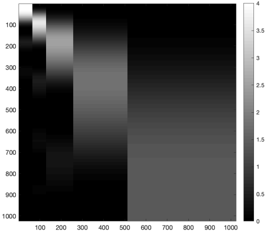
- AMP is a family of PnP algorithms that have remarkable properties for large random $\boldsymbol{A}$:
  - The denoiser input-error is white and Gaussian with predictable variance
  - When used with an MMSE denoiser, AMP algs converge to the MMSE estimate of $\boldsymbol{x}_0$ from $\boldsymbol{y}$

- Challenge: In most image recovery problems, $\boldsymbol{A}$ does not satisfy AMP's randomness assumptions

AMP for Fourier-Structured Matrix

Measurement model: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}$

- Idea: Recover the wavelet coefficients $\boldsymbol{c}_0$, not pixels $\boldsymbol{x}_0$
  - Why? The resulting model becomes $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{c}_0 + \boldsymbol{w}$, where the masked Fourier-wavelet $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{\Psi}^\mathsf{T}$ is approximately block-diagonal with sufficiently randomizing blocks



- With appropriate algorithm design, the denoiser input-error will be white and Gaussian *in each wavelet subband*

- Prior work includes Whitened VAMP [Schniter et al. '17], Variable-Density (VD)-AMP [Millard et al. '20], based on wavelet thresholding, & Denoising-VD-AMP [Metzler et al. '21]

Proposed Algorithm: Denoising GEC (D-GEC)

Our approach builds on the Generalized Expectation Consistent (GEC) algorithm from Fletcher et al. '16:

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

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

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

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

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

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

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

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

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

Proposed Algorithm: Denoising GEC (D-GEC)

- GEC is essentially 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 DNN denoiser

- Note: The algorithm provides well-characterized errors, but a non-standard denoiser is required to exploit them!

Denoising GEC (D-GEC): Jacobian Computation

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

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

  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[4]
  - For both $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$, we approximate the $\mathrm{tr}\{\boldsymbol{Q}_{\ell\ell}\}$ using

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

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

[4] Ramani et al. '08

New Update-Proposed Denoiser: corr+corr

- In the wavelet domain, the denoiser input-error is white and Gaussian in each subband, but with subband-dependent inverse-variances γ that change with the iterations
  - Thus, in the pixel-domain, the error is correlated Gaussian with known covariance matrix $\mathbf{\Psi}\,\mathrm{Diag}(\boldsymbol{\gamma})^{-1}\mathbf{\Psi}^{\mathsf{T}}$
  - How should we inform the denoiser about $(\mathbf{\Psi}, \boldsymbol{\gamma})$?

- We propose to add an extra input channel to an arbitrary denoiser (e.g., DnCNN) and feed it with an independent realization of $\mathcal{N}(\mathbf{0}, \mathbf{\Psi}\,\mathrm{Diag}(\boldsymbol{\gamma})^{-1}\mathbf{\Psi}^{\mathsf{T}})$
  - The denoiser learns to extract the statistics $(\mathbf{\Psi}, \boldsymbol{\gamma})$ from e and use them productively for denoising

New Results: Experimental setup

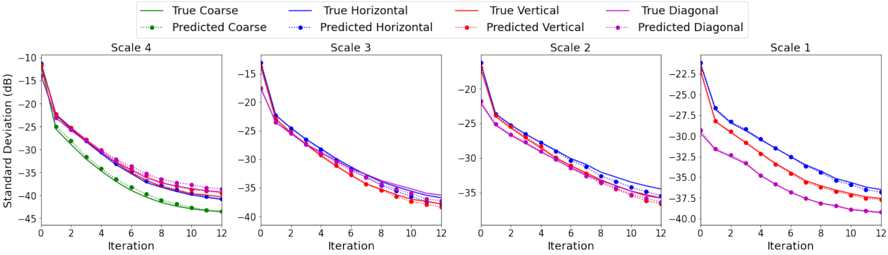
- We consider single coil measurements $\boldsymbol{y} = \boldsymbol{M}\boldsymbol{F}\boldsymbol{x}_0 + \boldsymbol{w}$
- $\boldsymbol{M}$ is a variable density mask
- $\boldsymbol{w}$ is AWGN giving pre-mask SNR $=$ 40 dB
- $\boldsymbol{\Psi}$ is 2D Haar wavelet transform with $D = 4$ levels $\Rightarrow$ 13 subbands
- PnP-PDS uses bias-free white-noise DnCNN and careful tuning
- D-VDAMP uses the modified DnCNN denoiser from [Metzler et al. '21]
- D-GEC uses proposed bias-free corr+corr DnCNN
- training data: 62,000 48x48 patches from 70 training images of the Stanford 2D FSE dataset

New Results: MRI Image Recovery

Avg performance on 10 Stanford 2D FSE 352×352 test images:

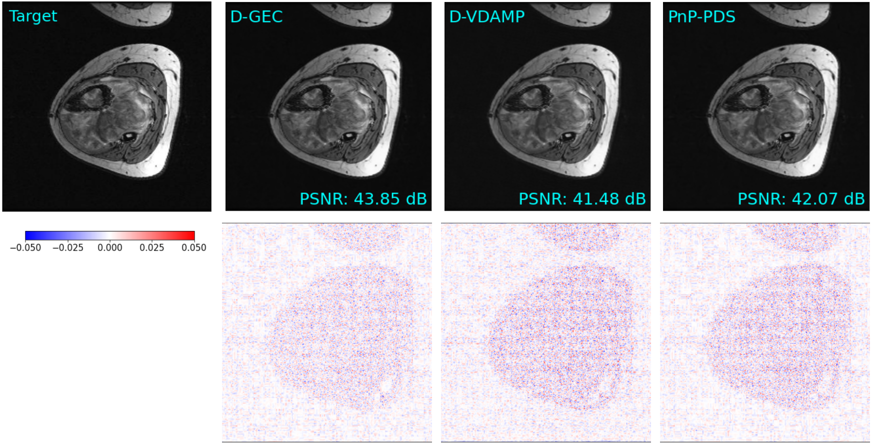
$C = 1$ coil	$M/N = 1/4$		$M/N = 1/8$	
method	PSNR	SSIM	PSNR	SSIM
PnP-PDS	45.97	0.978	41.28	0.957
D-VDAMP	44.61	0.974	38.43	0.901
D-GEC	47.64	0.982	42.42	0.959

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



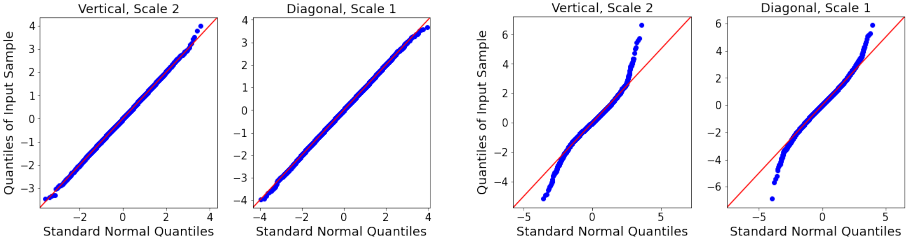
New Results: MRI Image Recovery

Example single-coil recoveries and error maps at $M/N = 1/4$:



New Results: MRI Image Recovery

Example wavelet-error QQ plots at iteration 10:



D-GEC PnP-PDS

Summary

- We designed GEC-based PnP algorithm for MRI called D-GEC
- Our algorithm renders the wavelet sub-band errors white and Gaussian with predictable variance
- We proposed a new Denoiser corr+corr which makes use of the predicted error statistics
- Empirical Results show that D-GEC has better fixed points than PnP-PDS and D-VDAMP