Generalized Approximate Message Passing for Cosparse Analysis Compressive Sensing (GrAMPA)

Mark Borgerding, Philip Schniter, Jeremy Vila: Ohio State   Sundeep Rangan, NYU Poly

The Ohio State University

(Supported by NSF CCF-1218754, CCF-1018368, and 3dB Labs, Inc.)

Problem Statement

Goal: Recover signal $\boldsymbol{x}$, given $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{w}$

- unknown signal $\boldsymbol{x} \in \mathbb{C}^N (or \mathbb{R}^N)$
- known linear measurement operator $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N} (or \mathbb{R}^{M\times N})$
- noise $\boldsymbol{w} \in \mathbb{C}^M (or \mathbb{R}^N)$

Challenge: Underdetermined linear system $M \ll N$

- $\boldsymbol{x}$ cannot be uniquely determined even if $\boldsymbol{w} = \boldsymbol{0}$
- prior information about $\boldsymbol{x}$ can help us navigate the measurement nullspace $\ker(\boldsymbol{\Phi})$

Synthesis CS

- Vector $\boldsymbol{x}$ is assumed to be sparse in an orthonormal dictionary $\boldsymbol{\Psi}$: $\boldsymbol{x} = \boldsymbol{\Psi u}$ for K-sparse $\boldsymbol{u} \in \mathbb{C}^N$.
- The goal is then to find

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_0 \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{\Phi\Psi u}\|_2^2 \le \epsilon, \quad \text{after which} \quad \hat{\boldsymbol{x}} = \boldsymbol{\Psi}\hat{\boldsymbol{u}}. \tag{1}$$

- If the K non-zeros in $\boldsymbol{u}$ can be found and $K \le M$, the system becomes *overdetermined*:

$$\boldsymbol{y} = (\boldsymbol{\Phi\Psi})_\Lambda \boldsymbol{u}_\Lambda + \boldsymbol{w} \text{ , where } (\boldsymbol{\Phi\Psi})_\Lambda \text{ is square or tall}$$

- Solving (1) is NP-hard. Practical approaches include
  - ℓ_1 convex relaxation: LASSO or BPDN
  - Greedy approaches: OMP, CoSaMP, Subspace Pursuit (SP), IHT
  - Bayesian: Sparse Bayesian Learning (SBL), Approximate Message Passing (AMP)
- The ℓ_1 approach works when $\boldsymbol{\Phi\Psi}$ satisfies the Restricted Isometry Property (RIP), which requires $M \gtrsim O(K \log N/K)$ when $\boldsymbol{x}$ is K-sparse.

Analysis CS

- Vector $\boldsymbol{x}$ is assumed to be sparse in an overcomplete dictionary $\boldsymbol{\Psi} \in \mathbb{C}^{N\times D}$ for $D \gg N$.
- Problem: $\boldsymbol{\Phi\Psi}$ does not satisfy RIP $\Rightarrow$ synthesis-CS fails!
- Instead, try "analysis CS" with analysis operator $\boldsymbol{\Omega} = \boldsymbol{\Psi}^\dagger$:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{\Omega x}\|_0 \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2^2 \le \epsilon \tag{2}$$

- If enough zeros in $\boldsymbol{\Omega x}$ can be found, the augmented system becomes *overdetermined*:

$$\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{0}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Phi}\\ \boldsymbol{\Omega}_\Lambda\end{bmatrix}\boldsymbol{x} + \boldsymbol{w} \text{ , where } \begin{bmatrix}\boldsymbol{\Phi}\\ \boldsymbol{\Omega}_\Lambda\end{bmatrix} \text{ is square or tall}$$

- Solving (2) is NP-hard. Practical approaches include
  - ℓ_1 convex relaxation: Generalized LASSO
  - Greedy approaches: Greedy Analysis Pursuit (GAP), Analysis versions of CoSaMP, SP, IHT
  - Bayesian: TV-AMP, SS-AMP
- Some popular choices of $\boldsymbol{\Omega}$ are finite-difference operator and Wavelet transform (concatentations of many).

AMP and GAMP

Approximate Message Passing [Donoho, Maleki, Montanari 10]

- Approximation of loopy belief propagation applied to synthesis CS
- Compute approximate MMSE or MAP estimate of $\boldsymbol{x} \sim \prod_n p_X(x_n)$ from AWGN corrupted $\boldsymbol{z} = \boldsymbol{\Phi x}$
- For i.i.d. sub-Gaussian $\boldsymbol{\Phi\Psi}$, as $M, N \to \infty$ for fixed ratio M/N, state evolution characterizes performance.
- Manifests as an iterative thresholding algorithm with p_X-dependent soft threshold

Generalized AMP [Rangan 11]

- Extension of AMP that handles arbitrary separable likelihood $p(\boldsymbol{y}|\boldsymbol{z}) = \prod_m p_{Y_m|Z_m}(y_m|z_m)$
- Applicable to non-Gaussian/non-additive noise: quantization, phase retrieval, Poisson corruption

Related Publications

1. D.L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction," ITW, 2010.
2. S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inform. Thy.*, (Saint Petersburg, Russia), pp. 2168-2172, Aug. 2011.
3. M. Borgerding and P. Schniter, "Generalized Approximate Message Passing for the Cosparse Analysis Model," *arXiv:1312.3968*,Dec 2013.
4. D. L. Donoho, I. M. Johnstone, and A. Montanari, "Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising," *arXiv:1111.1041*, Nov. 2011.
5. S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "Recovery of cosparse signals with greedy analysis pursuit in the presence of noise," *Proc. IEEE Workshop Comp. Adv. Multi-Sensor Adaptive Process.*, (Puerto Rico), Dec. 2011
6. R. E. Carrillo, J. D. McEwen, D. V. D. Ville, J.-P. Thiran, and Y. Wiaux, "Sparsity averaging for compressive imaging," *IEEE Signal Process. Lett.*, vol. 20, 2013
7. E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. App.*, Dec. 2008
8. J.W. Kang, H. Jung, H.N. Lee, and K. Kim, "One-dimensional piecewise-constant signal recovery via spike-and-slab approximate message-passing," *48th Asilomar Conf*, Nov. 2014

Generalized AMP for Analysis-CS (GrAMPA)

- Formulates analysis-CS via

$$\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{v}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Phi}\\ \boldsymbol{\Omega}\end{bmatrix}\boldsymbol{x} + \begin{bmatrix}\boldsymbol{w}\\ \boldsymbol{0}\end{bmatrix} \tag{3}$$

- Sparse analysis output $\boldsymbol{v}$ modeled via sparsifying prior $p(\boldsymbol{v}) = \prod_i p_V(v_i)$, e.g.,
  - ℓ_1 based prior (Laplacian MAP)
  - Bernoulli-Gaussian prior
  - ℓ_0 mimicking priors (SNIPE...)
- Leverages GAMP's ability to handle non-AWGN likelihood
- Signal $\boldsymbol{x}$ is unconstrained or, depending on application, possibly real or positive

Sparse Non-Informative Parameter Estimator (SNIPE)

- Novel form of soft-thresholding that successfully mimics the desired ℓ_0 minimization
- Consider a Bernoulli-*Something* random variable V, observed in AWGN
  - prior: $p_V(v) = (1-\lambda)\delta(v) + \lambda \frac{p_0(\frac{v}{\sigma})}{\sigma}$
  - AWGN corruption: $q = v + e$ for $e \sim \mathcal{N}(0,\tau)$ in $\mathbb{R}$eal case
- We define SNIPE as the scale-invariant approximation of the MMSE estimator for arbitrarily large σ

$$\hat{v}_{\mathsf{SNIPE}}(q;\tau,\omega) \triangleq \lim_{\sigma\to\infty,\lambda\to 1} \mathrm{E}\left(V|Q=q\right) = \frac{q}{\exp\left(\omega - \frac{q^2}{2\tau}\right)+1} \tag{4}$$

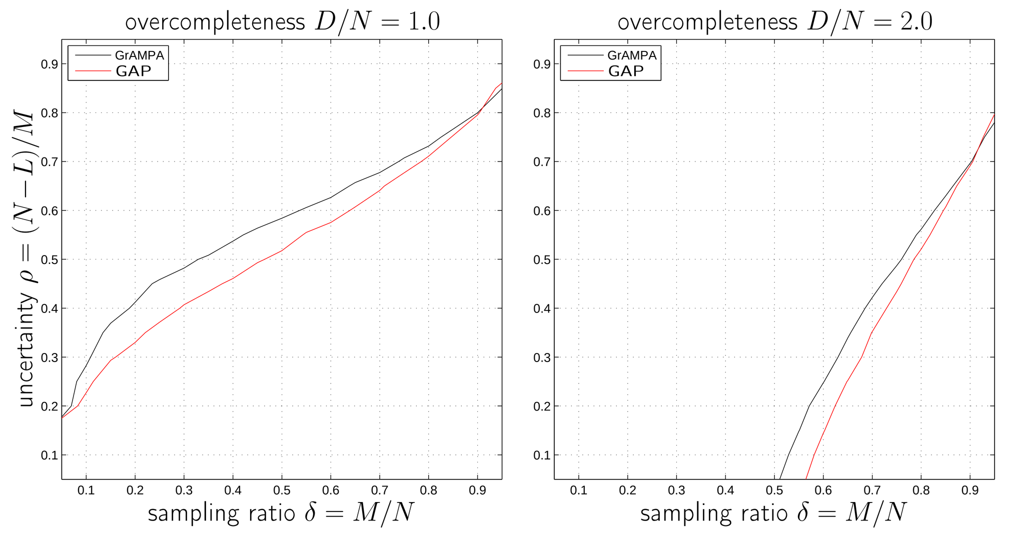
where ω is a tuning parameter (that controls the relative rate at which σ and λ converge).

- Since SNIPE is scale-invariant, ω has a wide "sweet spot".

Phase Transition Curves : Comparison to GAP

- The Phase Transition Curve (PTC) partitions the problem space into solvable vs unsolveable.
- Combinations of (sampling,uncertainty) below the curve succeed with high probabililty, and those above the curve fail.
- The higher the curve, the better the algorithm!

GrAMPA+SNIPE versus GAP under i.i.d Gaussian $\boldsymbol{\Phi}$ and random tight frame $\boldsymbol{\Omega}$.

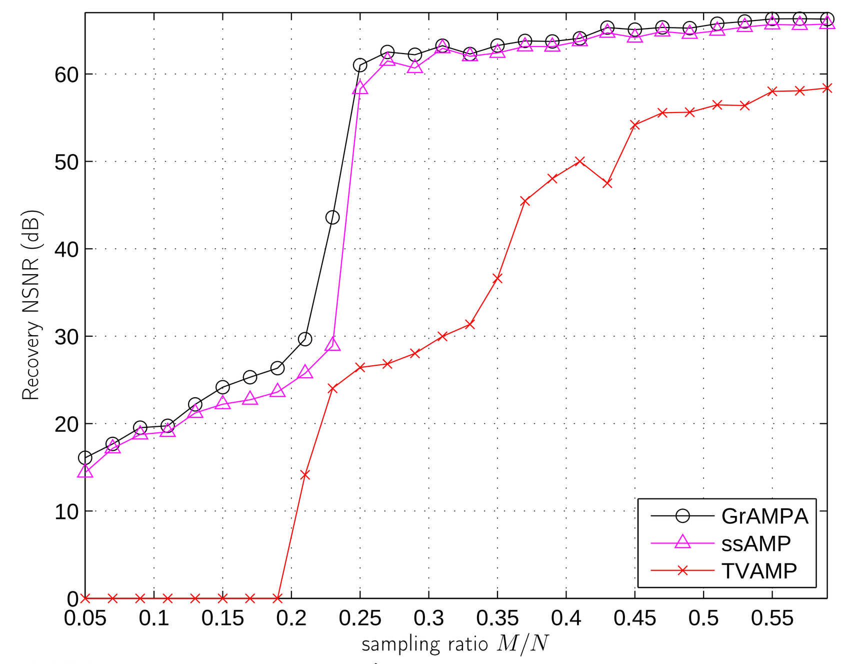


- M: number of measurements
- N: signal dimension
- D: number of atoms(rows) in $\boldsymbol{\Omega}$
- L: number of zeros in $\boldsymbol{\Omega x}$
- $\delta = M/N$: sampling ratio
- $\rho = (N-L)/M$: uncertainty ratio
  $\rho \to 0$ is easy; $\rho > 1$ is impossible

Piecewise Constant Signal Recovery: 1D Difference Dictionary

Others have considered AMP to solve the special case of 1D finite-difference $\boldsymbol{\Omega}$.

- [Donoho, Johnstone, and Montanari 2011] proposed a "TV-AMP" for this application that alternates MAP-AMP with an external TV denoising package like TV-DIP or FLSA
- [Kang, Jung, Lee, and Kim 2014] created a "Spike-and-Slab" (Bernoulli-Gaussian) MMSE-AMP solver.



- i.i.d Gaussian $\boldsymbol{\Phi}$
- 1D finite-difference $\boldsymbol{\Omega}$
- piecewise constant $\boldsymbol{x}$
- noiseless

	GrAMPA	ssAMP	TV-AMP
runtime:	2.1s	**.3s**	.8s

GrAMPA wins in recovery, but not runtime.

Synthetic Image Recovery: 4x Finite-Difference Dictionary

- radially sampled 2D Fourier measurements
- $4\times$ overcomplete $\boldsymbol{\Omega}$: horizontal, vertical, diagonal, antidiagonal differences

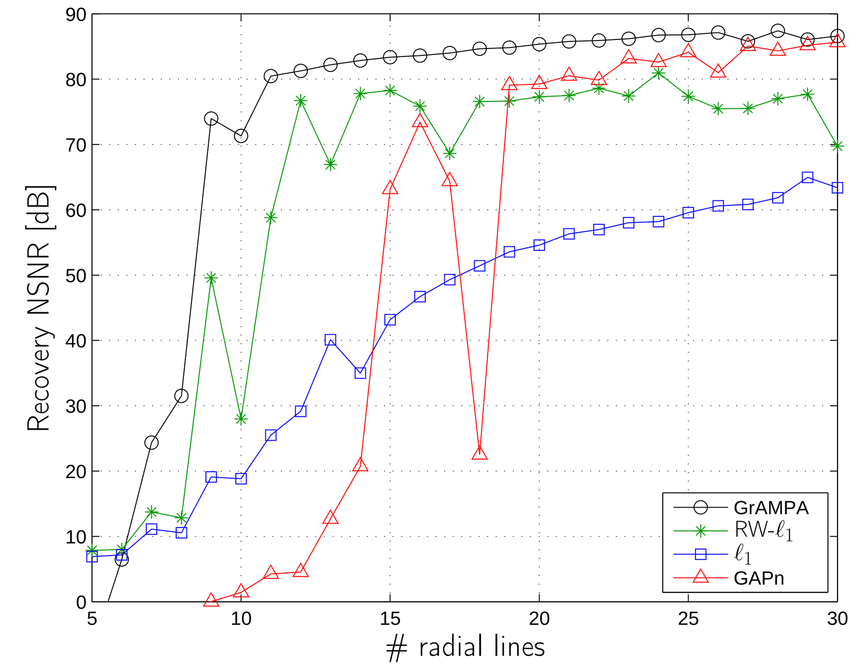


Figure: Shepp-Logan Phantom

	GrAMPA	ℓ_1	RW-ℓ_1	GAP
runtime:	**0.28s**	1.8s	9.7s	30.1s

GrAMPA wins in both runtime and recovery.

Natural Image Recovery: Overcomplete Wavelet Dictionary

- spread-spectrum Fourier measurements
- $8\times$ overcomplete $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_1^T, \ldots, \boldsymbol{\Omega}_8^T]^T$ with $\boldsymbol{\Omega}_i$ = Daubechies-i

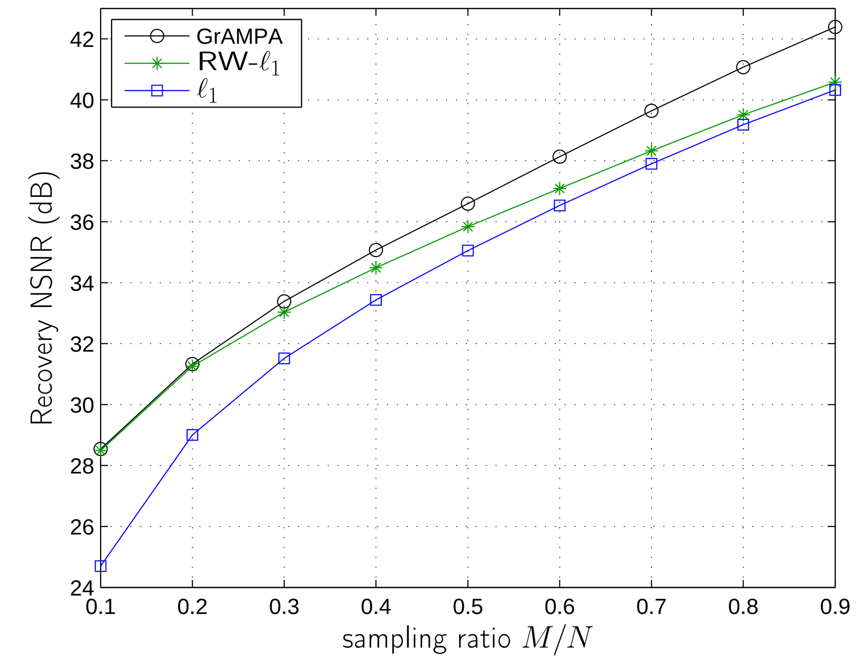


Figure: Lena

	GrAMPA	ℓ_1	RW-ℓ_1
runtime:	**149s**	177s	1994s

GrAMPA wins in both runtime and recovery.

Performance versus Overcompleteness

- sampling ratio fixed at $M/N = 0.5$
- $Q\times$ overcomplete with varying Q using $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_1^T, \ldots, \boldsymbol{\Omega}_Q^T]^T$ with $\boldsymbol{\Omega}_i$ = Daubechies-i

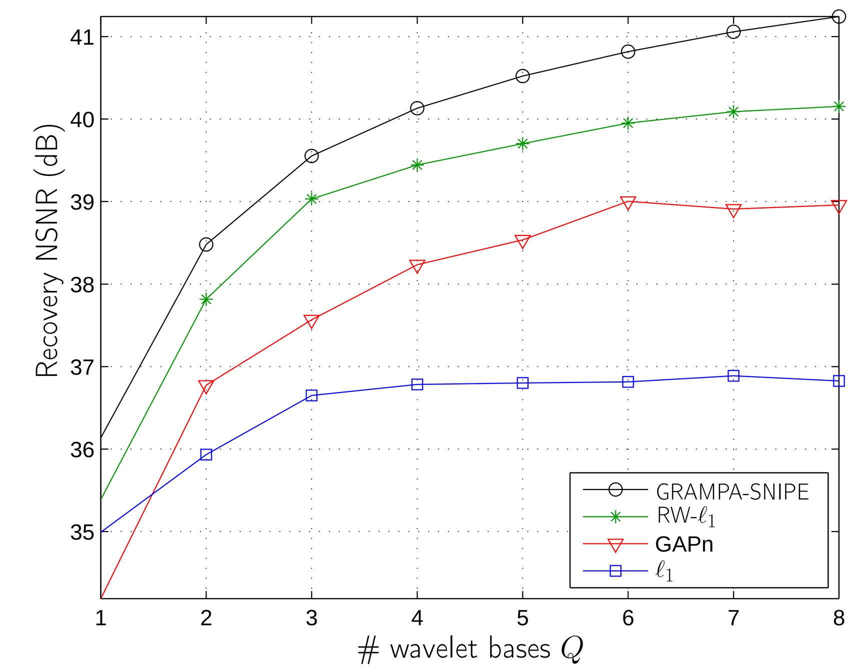


Figure: Knee MRI

Runtime:	Q :	1	2	3	4	5	6	7	8
	GrAMPA(SNIPE)	327s	**358s**	**488s**	**1205s**	**1386s**	1610s	**1902s**	**2366s**
	RW-ℓ_1	3929s	3336s	3199s	3970s	4535s	6855s	9861s	6642s
	GAP	**254s**	730s	829s	1442s	1984s	**1540s**	2959s	2787s