Swept Approximate Message Passing for Sparse Estimation

Andre Manoel[IF-USP], Florent Krzakala[LPS-ENS], Eric W. Tramel[LPS-ENS] and Lenka Zdeborová[IPhT CEA]

Compressed Sensing & Inference

Compressed Sensing (CS): Solving for an unknown signal $\mathbf{x}$ of dimensionality N from some set of observations $\mathbf{y}$ of dimensionality $M << N$,

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{w}, \quad w_\mu \sim \mathcal{N}(0, \Delta), \qquad \text{(AWGN)}$$

given that $\mathbf{x}$ is K-sparse, i.e. containing K non-zero entries, with $\Phi_{\mu i} \sim \mathcal{N}(0, \frac{1}{\sqrt{N}})$.

Convex Approach: Approximate an ℓ_0 semi-norm regularization via a convex relaxation to the ℓ_1 norm. Solving

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \Phi\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1, \qquad \text{(LASSO)}$$

recovers $\mathbf{x}$ exactly for sufficient $M > K$, as shown in the early CS literature.

Probabilistic Approach: The posterior probability is given by a prior signal model and a stochastic description of the observation channel. In the case of an AWGN channel,

$$P(\mathbf{x}|\mathbf{y}, \Phi) \propto P_0(\mathbf{x}) P(\mathbf{y}|\mathbf{x}; \Phi),$$
$$\propto \prod_i P_0(x_i) \prod_\mu \frac{1}{\sqrt{2\pi\Delta}} e^{-\frac{(y_\mu - \sum_i \Phi_{\mu i} x_i)^2}{2\Delta}}.$$

Maximum a Posteriori (MAP): maximize over the posterior distribution,

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y}; \Phi), \qquad \text{(MAP)}$$

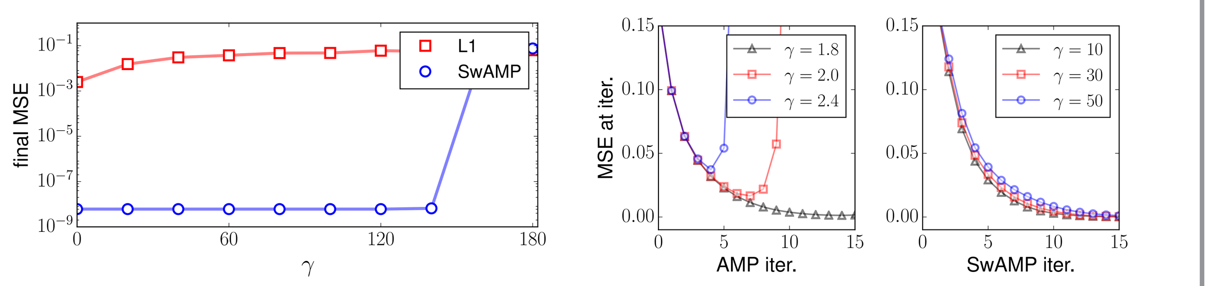
where (LASSO) can be recovered via MAP estimation with $P_0(x_i)$ taken as a Laplace distribution.

Minimum Mean Square Error (MMSE): estimate the average value of $\mathbf{x}$,

$$\hat{\mathbf{x}} = \int \mathrm{d}\mathbf{x}\ \mathbf{x} P(\mathbf{x}|\mathbf{y}; \Phi). \qquad \text{(MMSE)}$$

Exact MMSE estimation is intractable, thus approximate methods are required. In the variational Bayesian approach, one approximates the posterior as fully factorized. A more accurate approximation is obtained via *belief propagation* (BP).

Case I: Non-Zero Mean



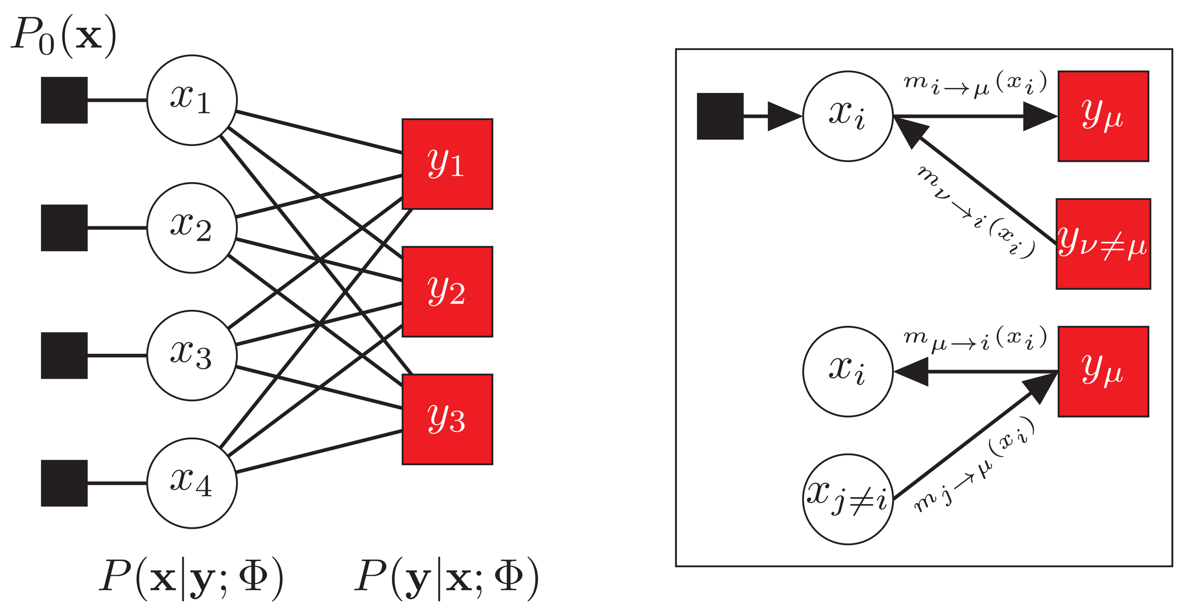
A non-zero constant is added to all projections such that

$$\Phi_{\mu i} \sim \mathcal{N}\left(\frac{\gamma}{N}, \frac{1}{N}\right).$$

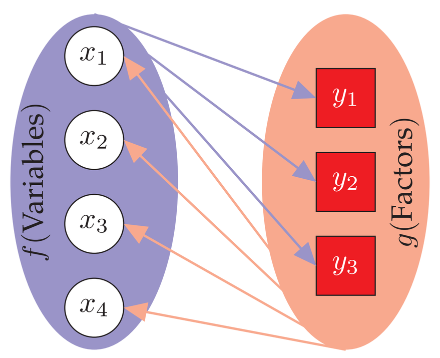
Experiments are conducted for $N = 10^4$, $\rho = 0.2$, $\Delta = 10^{-8}$ and $\alpha = 0.5$. For even small γ, parallel AMP fails to converge for this problem (which is outside the region of successful ℓ_1 recovery.) SwAMP, however, is robust to γ.

Message Passing

Construction: Using a *factor graph* construction, we can write a message passing algorithm via BP to approximate a factorization of the posterior.

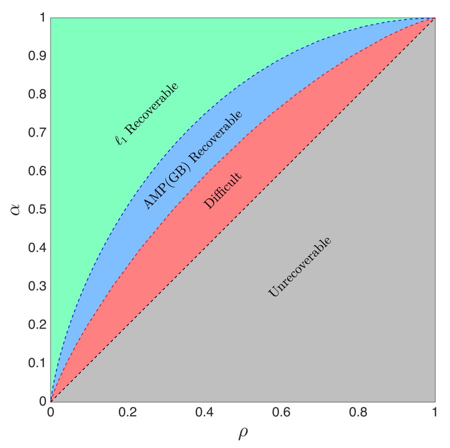


Reducing Complexity: If the entries of Φ are uncorrelated and have magnitudes which scale as $O(1/\sqrt{N})$, i.e. exhibiting *weak interactions*, one can write an *approximate* message passing (AMP) [1] on the variables and factors, $O(M+N)$, rather than performing BP on the *edges* of the factor graph, $O(MN)$ [2].



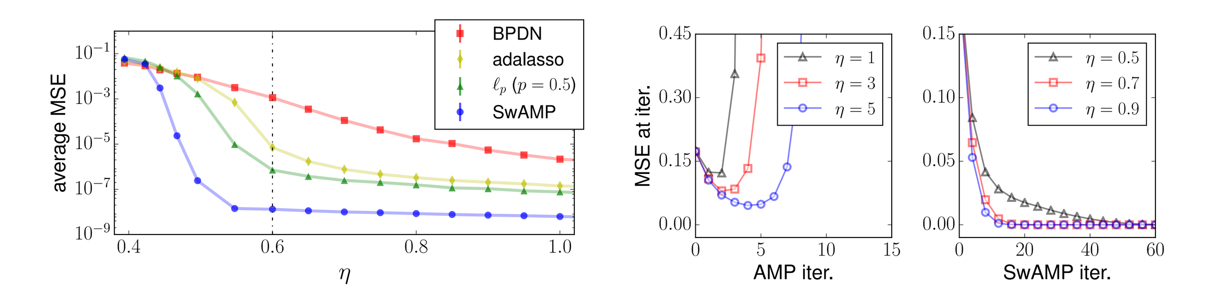
Non-convex Prior: When $\mathbf{x}$ is K-sparse, a non-convex prior can provide superior performance to ℓ_1 regularization. For example, the *Gauss-Bernoulli* (GB) prior,

$$P_0(x_i; \rho, \xi, \sigma^2) = (1-\rho)\delta(x_i) + \rho\mathcal{N}(x_i; \xi, \sigma^2), \qquad \text{(GB)}$$



provides state-of-the-art reconstruction performance for K-sparse signals. We can observe this by looking at the *phase transition* performance, denoting regions of successful and unsuccessful recovery according to the parameters ($\alpha = M/N, \rho = K/N$).

Case II: Correlations

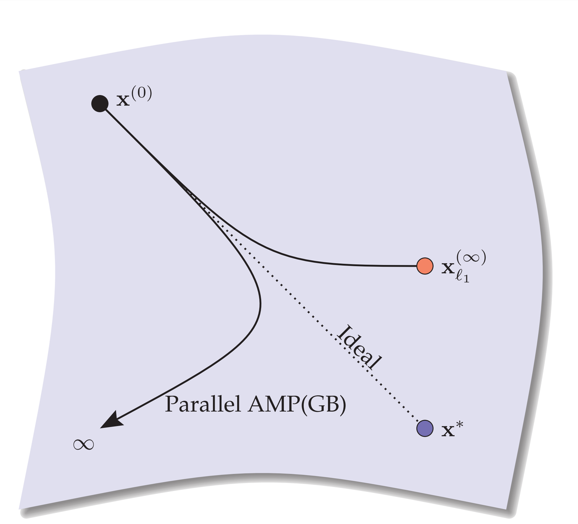


The projection is constructed to be correlated by multiplying two random matrices, $P \in \mathbb{R}^{M \times R}$ and $Q \in \mathbb{R}^{R \times N}$ for $R \triangleq \eta N$,

$$\Phi = \frac{1}{N} PQ \quad P_{\mu k}, Q_{ki} \sim \mathcal{N}(0, 1).$$

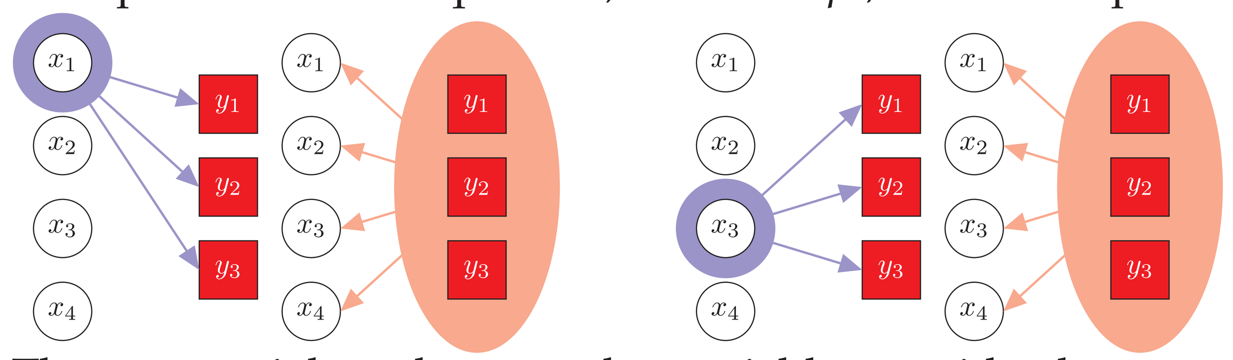
This problem is closely related to sparse linear regression (or feature selection), where Φ represents some observed set of features which may exhibit strong correlations. AMP diverges in such cases, but SwAMP is robust to these correlations. For these experiments $N = 1024$, $\rho = 0.2$, $\alpha = 0.6$, and $\Delta = 10^{-8}$.

Swept AMP (SwAMP)



Breaking AMP: While AMP shows distinct advantages for sparse reconstruction, when Φ does not meet our assumptions the AMP fixed-point iteration can fail to converge, and in fact, can *wildly diverge*. This severely limits its potential application to many practical problems.

Swept Update: Following from the sequential relaxed BP proposed in [3], we attempt the same with AMP. By deriving the proper time indexing, we produce a sequential, or *swept*, AMP update.



The sequential update on the variables avoids the error introduced by the parallel updates when Φ does not match our statistical assumption of wide-scale and weak interactions.

G-SwAMP: Additionally, as in GAMP [4], SwAMP can be written to work on any generalized linear problem of the form

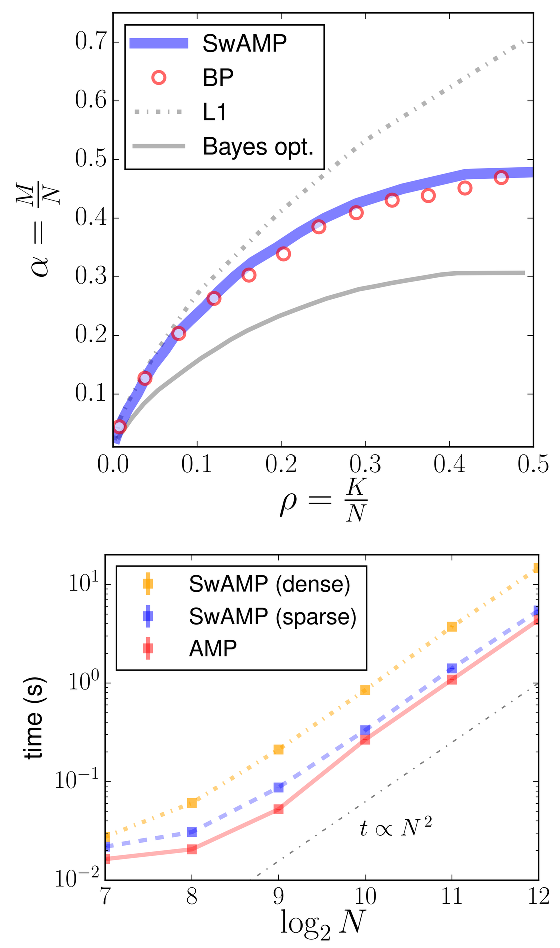
$$\mathbf{y} = f_{\text{out}}(\Phi\mathbf{x}),$$

where $f_{\text{out}}(\cdot)$ is a stochastic model of the output channel. This is accomplished easily by the appropriate change of the updates to g_μ and Σ_i^2.

Algorithm 1 The SwAMP Algorithm

Input: $y, \Phi, \Delta, \theta_{\text{prior}}, t_{\max}, \varepsilon$
$t \leftarrow 0$
Initialize $\{\mathbf{a}^{(0)}, \mathbf{v}^{(0)}\}, \{\omega^{(0;N+1)}, \mathbf{V}^{(0;N+1)}\}$
while $t < t_{\max}$ **and** $\|\mathbf{a}^{(t+1)} - \mathbf{a}^{(t)}\| > \varepsilon$ **do**
  for $\mu = 1$ **to** M **do**
    $g_\mu^{(t)} \leftarrow \frac{y_\mu - \omega_\mu^{(t;N+1)}}{\Delta + V_\mu^{(t;N+1)}}$
    $V_\mu^{(t+1;1)} \leftarrow \sum_i \Phi_{\mu i}^2 v_i^{(t)}$
    $\omega_\mu^{(t+1;1)} \leftarrow \sum_i \Phi_{\mu i} a_i^{(t)} - V_\mu^{(t+1;1)} g_\mu^{(t)}$
  end for
  $\mathbf{S} \leftarrow \text{Permute}([1, 2, \ldots, N])$
  for $k = 1$ **to** N **do**
    $i \leftarrow S_k$
    $\Sigma_i^{2(t+1)} \leftarrow \left[\sum_\mu \frac{\Phi_{\mu i}^2}{\Delta + V_\mu^{(t+1;k)}}\right]^{-1}$
    $R_i^{(t+1)} \leftarrow a_i^{(t)} + \Sigma_i^{2(t+1)} \sum_\mu \Phi_{\mu i} \frac{y_\mu - \omega_\mu^{(t+1;k)}}{\Delta + V_\mu^{(t+1;k)}}$
    $a_i^{(t+1)} \leftarrow f_1(R_i^{(t+1)}, \Sigma_i^{2(t+1)}; \theta_{\text{prior}})$
    $v_i^{(t+1)} \leftarrow f_2(R_i^{(t+1)}, \Sigma_i^{2(t+1)}; \theta_{\text{prior}})$
    for $\mu = 1, m$ **do**
      $V_\mu^{(t+1;k+1)} \leftarrow V_\mu^{(t+1;k)} + \Phi_{\mu i}^2 (v_i^{(t+1)} - v_i^{(t)})$
      $\omega_\mu^{(t+1;k+1)} \leftarrow \omega_\mu^{(t+1;k)} + \Phi_{\mu i} (a_i^{(t+1)} - a_i^{(t)}) - g_\mu^{(t)} (V_\mu^{(t+1;k+1)} - V_\mu^{(t+1;k)})$
    end for
  end for
  $t \leftarrow t + 1$
end while

Case III: Group Testing



Group testing and experiment pooling are another linear problem of interest. Here,

$$\Phi_{\mu i} \in \{0, 1\},$$

with each row summing to a fixed value. In these experiments, the rows sum to 7, making Φ very sparse. AMP cannot solve such problems, while SwAMP matches the performance of BP-based solutions [5]. Additionally, SwAMP can be written efficiently for sparse Φ, approaching the performance of parallel AMP.

Links

Paper — Source Code — SPHINX Team

References

[1] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.

[2] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 8, p. P08009, August 2012.

[3] F. Caltagirone, F. Krzakala, and L. Zdeborová, "On convergence of approximate message passing," in *Information Theory, Proc. IEEE International Symposium on*, Honolulu, HI, June 2014, pp. 1812–1816.

[4] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Information Theory Proceedings, IEEE International Symposium on*, August 2011, pp. 2168–2172.

[5] P. Zhang, F. Krzakala, M. Mézard, and L. Zdeborová, "Non-adaptive pooling strategies for detection of rare faulty items," in *Communications Workshops, Proc. IEEE International Conference on*, Budapest, Hungary, June 2013, pp. 1409–1414.

Funding

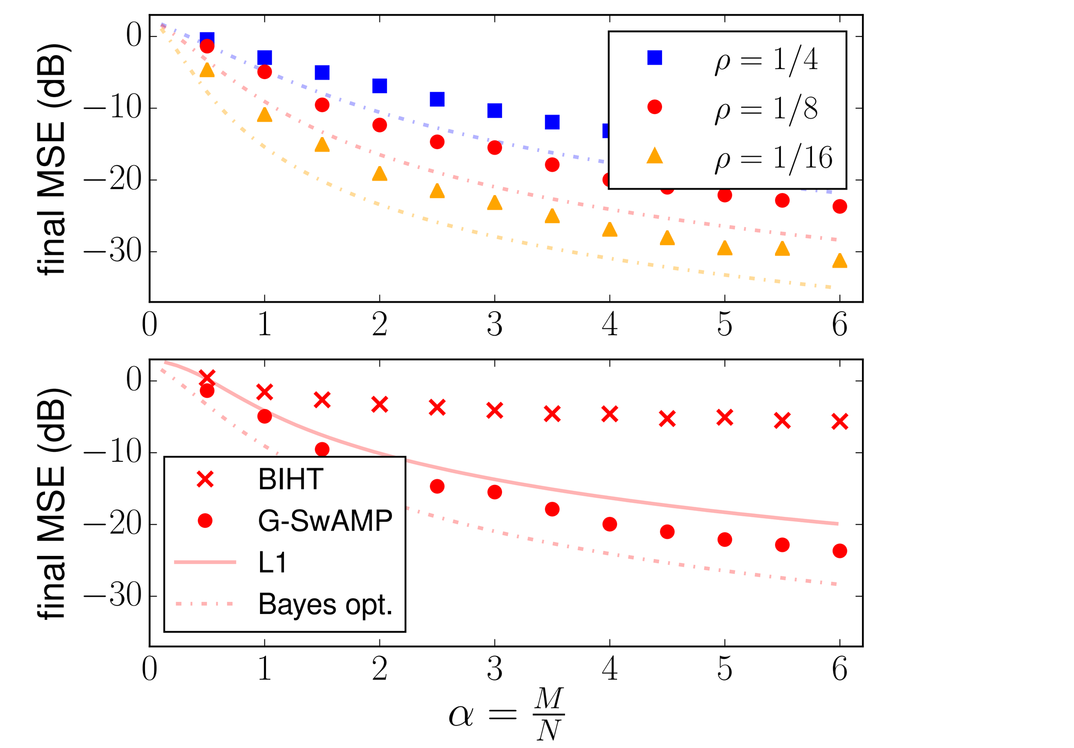
This work has been supported in part by the ERC under the European Union's 7^{th} Framework Programme Grant Agreement 307087-SPARCS, by the Grant DySpaN of "Triangle de la Physique," and by FAPESP under grant 13/01213-8.

Case IV: One-Bit CS

Since SwAMP can be easily adapted to general output channels, we investigate its usefulness for one-bit compressed sensing,

$$\mathbf{y} = \text{sign}(\Phi\mathbf{x}),$$

for Φ with a non-zero mean (as in Case I).



We conduct reconstruction experiments for $N = 512$ averaged over 200 trials. At top, we compare $\gamma = 0$ Bayes-optimal performance (dashed) to the achieved SwAMP performance at $\gamma = 20$. At bottom, we make comparisons against other approaches over α for $\rho = 1/8$. We see that, even on non-AWGN channels, SwAMP provides results superior to convex techniques while remaining robust to the properties of Φ.