Statistical Image Recovery: A Message-Passing Perspective

Philip Schniter*, Sundeep Rangan†, and Alyson Fletcher‡
*Dept. of ECE, The Ohio State University, Columbus, OH 43210, USA.
†Dept. of ECE, Polytechnic Institute of New York University, Brooklyn, NY 11201, USA.
‡Dept. of ECE, The University of California, Santa Cruz, CA 95064, USA.

***Abstract*—We review MAP and MMSE-based approaches to image recovery and their implementation via generalized approximate message-passing (GAMP), highlighting recent results on GAMP convergence for general measurement operators.**

We consider the recovery of image $\boldsymbol{x} \in \mathbb{C}^N$ from noisy outputs $\boldsymbol{y} \in \mathbb{C}^M$ of known linear measurement operator $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N}$. The "statistical" approach to image recovery models the image $\boldsymbol{x}$ as a realization of random $\mathsf{x} \sim p_{\mathsf{x}}$ and the measurements as a realization of random y whose statistics are governed by a likelihood function of the form $p_{\mathsf{y}|\mathsf{z}}(\boldsymbol{y}|\boldsymbol{\Phi}\widehat{\boldsymbol{x}})$. Here, $p_{\mathsf{y}|\mathsf{z}}$ is the pdf of y conditioned on the (hidden) transform outputs $\mathsf{z} = \boldsymbol{\Phi}\mathsf{x}$ and $\widehat{\boldsymbol{x}}$ is a hypothesis of the image. For clarity, we denote random quantities in san-serif font.

In the maximum a posteriori (MAP) approach to statistical image recovery, one computes the most probable estimate of $\boldsymbol{x}$ given $\boldsymbol{y}$, i.e.,

$$\begin{aligned}\widehat{\boldsymbol{x}}_{\mathsf{MAP}} &= \operatorname{argmax}_{\widehat{\boldsymbol{x}}}\, p_{\mathsf{x}|\mathsf{y}}(\widehat{\boldsymbol{x}}|\boldsymbol{y}) = \operatorname{argmax}_{\widehat{\boldsymbol{x}}}\, p_{\mathsf{y}|\mathsf{x}}(\boldsymbol{y}|\widehat{\boldsymbol{x}})p_{\mathsf{x}}(\widehat{\boldsymbol{x}})/p_{\mathsf{y}}(\boldsymbol{y}) \\ &= \operatorname{argmin}_{\widehat{\boldsymbol{x}}}\big\{-\log p_{\mathsf{y}|\mathsf{z}}(\boldsymbol{y}|\boldsymbol{\Phi}\widehat{\boldsymbol{x}}) - \log p_{\mathsf{x}}(\widehat{\boldsymbol{x}})\big\} \qquad (1)\end{aligned}$$

which can be interpreted (from a non-statistical viewpoint) as regularized loss minimization, i.e.,

$$\widehat{\boldsymbol{x}}_{\mathsf{MAP}} = \operatorname{argmin}_{\widehat{\boldsymbol{x}}}\big\{L(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}) + R(\widehat{\boldsymbol{x}})\big\} \qquad (2)$$

using the loss $L(\boldsymbol{z}) \triangleq -\log p_{\mathsf{y}|\mathsf{z}}(\boldsymbol{y}|\boldsymbol{z})$ and regularization $R(\widehat{\boldsymbol{x}}) \triangleq -\log p_{\mathsf{x}}(\widehat{\boldsymbol{x}})$. By choosing $p_{\mathsf{y}|\mathsf{z}}$ and p_{x} so that both $L(\cdot)$ and $R(\cdot)$ are convex, one can readily apply convex optimization algorithms to the image recovery problem. In image recovery, it is popular to use regularizations of the form $R(\widehat{\boldsymbol{x}}) = \|\boldsymbol{\Omega}\widehat{\boldsymbol{x}}\|_1$ for a given matrix $\boldsymbol{\Omega}$.

In the minimum mean-squared error (MMSE) approach to statistical image recovery, the objective is to compute

$$\widehat{\boldsymbol{x}}_{\mathsf{MMSE}} = \mathrm{E}\{\mathsf{x}|\mathsf{y}=\boldsymbol{y}\} = \int_{\mathbb{C}^N} \widehat{\boldsymbol{x}}\, p_{\mathsf{x}|\mathsf{y}}(\widehat{\boldsymbol{x}}|\boldsymbol{y}) d\widehat{\boldsymbol{x}}, \qquad (3)$$

with the hope of mean-square optimal performance. Unfortunately, the high-dimensional integral in (3) is computable in closed-form for only a very narrow class of priors and likelihoods (e.g., Gaussian) and even then may require the inversion of a very large matrix.

For problems with separable loss and regularization, i.e.,

$$L(\widehat{\boldsymbol{z}}) = \textstyle\sum_{i=1}^M L_i(\widehat{z}_i) \text{ and } R(\widehat{\boldsymbol{x}}) = \sum_{j=1}^N R_j(\widehat{x}_j), \qquad (4)$$

a computationally efficient inference methodology that supports either MAP or MMSE recovery was recently proposed under the name of "generalized approximate message passing" (GAMP) [1]. GAMP is an extension of the AMP algorithm [2] from quadratic loss (i.e., $L_i(\widehat{z}_i) = \frac{1}{\nu^w}(y_i - \widehat{z}_i)^2$ for some $\nu^w > 0$) to generic loss $L_i(\cdot)$, as needed for phase retrieval, Poisson noise, or quantized measurements. Interestingly, the behavior of GAMP for large i.i.d $\boldsymbol{\Phi}$ is rigorously characterized by a state evolution whose fixed points, when unique, are MAP or MMSE optimal [3]. Still, important questions remain about the convergence of GAMP for generic $\boldsymbol{\Phi}$, and whether GAMP can be applied to non-separable regularizers like $\|\boldsymbol{\Omega}\widehat{\boldsymbol{x}}\|_1$, which are commonly used in image recovery.

*With support from NSF grants CCF-1018368 and CCF-1218754.

In this talk, we review recent results on the convergence of GAMP for generic $\boldsymbol{\Phi}$. First, for any $\boldsymbol{\Phi}$, we recall that the fixed points of MAP-GAMP are known to coincide with the critical points of the optimization (2) [4]. Meanwhile, the fixed points of MMSE-GAMP are known to coincide with the critical points of the optimization [4]

$$(f_{\mathsf{x}}, f_{\mathsf{z}}) = \operatorname{argmin}_{b_{\mathsf{x}}, b_{\mathsf{z}}} J(b_{\mathsf{x}}, b_{\mathsf{z}}) \text{ s.t. } \mathrm{E}\{\mathsf{z}|b_{\mathsf{z}}\} = \boldsymbol{\Phi}\,\mathrm{E}\{\mathsf{x}|b_{\mathsf{x}}\} \qquad (5)$$

$$J(b_{\mathsf{x}}, b_{\mathsf{z}}) \triangleq D\big(b_{\mathsf{x}}\|p_{\mathsf{x}}\big) + D\big(b_{\mathsf{z}}\|p_{\mathsf{y}|\mathsf{z}}Z^{-1}\big) + H\big(b_{\mathsf{z}}; \boldsymbol{\nu}^p\big), \qquad (6)$$

where J is a high-dimensional approximation of the Bethe free-energy [5]. In (6), $b_{\mathsf{x}}(\boldsymbol{x}) = \prod_j b_{\mathsf{x}_j}(x_j)$ and $b_{\mathsf{z}}(\boldsymbol{z}) = \prod_i b_{\mathsf{z}_i}(z_i)$ are pdfs, $D(\cdot\|\cdot)$ denotes KL divergence, and $H(b_{\mathsf{z}}; \boldsymbol{\nu}^p) \triangleq \sum_{i=1}^M \mathrm{var}\{\mathsf{z}_i|b_{\mathsf{z}_i}\}/\nu_i^p + \ln \pi\nu_i^p$ for $\nu_i^p = \sum_{j=1}^N |\Phi_{ij}|^2 \mathrm{var}\{\mathsf{x}_j|b_{\mathsf{x}_j}\}$. But these fixed points don't tell the whole story, because GAMP may diverge. For quadratic $L_i(\cdot)$ and $R_j(\cdot)$, however, the convergence of GAMP has been fully characterized, and global convergence can be ensured by "damping" [6]. Damping can also be used to ensure local convergence under strictly convex $L_i(\cdot)$ and $R_j(\cdot)$ [6].

We also review recent results on connections between GAMP and convex optimization algorithms. For example, with MAP-GAMP, the variable updates coincide with those of the primal-dual hybrid gradient (PDHG) approach to (2) while the stepsizes are adapted in accordance with the local cost [6]. Meanwhile, with MMSE-GAMP, the mean updates coincide with an application of PDHG to (5) under a local convexification of the augmented Lagrangian, while the variance updates adapt that local convexification. Finally, we describe a recent variant on MMSE-GAMP that guarantees global convergence with generic $\boldsymbol{\Phi}$ for strictly convex F, G with bounded derivatives.

Finally, we describe how GAMP can be configured to use "analysis" non-separable regularizers $R(\widehat{\boldsymbol{x}}) = \sum_{d=1}^D G_d([\boldsymbol{\Omega}\widehat{\boldsymbol{x}}]_d)$ [7].

References

[1] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inform. Thy.*, pp. 2168–2172, Aug. 2011. (full version at *arXiv:1010.5141*).
[2] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, pp. 18914–18919, Nov. 2009.
[3] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, pp. 764–785, Feb. 2011.
[4] S. Rangan, P. Schniter, E. Riegler, A. Fletcher, and V. Cevher, "Fixed points of generalized approximate message passing with arbitrary matrices," in *Proc. IEEE Int. Symp. Inform. Thy.*, pp. 664–668, July 2013. (full version at *arXiv:1301.6295*).
[5] F. Krzakala, A. Manoel, E. W. Tramel, and L. Zdeborová, "Variational free energies for compressed sensing," in *Proc. IEEE Int. Symp. Inform. Thy.*, pp. 1499–1503, July 2014. (see also *arXiv:1402.1384*).
[6] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of generalized approximate message passing with arbitrary matrices," in *Proc. IEEE Int. Symp. Inform. Thy.*, pp. 236–240, July 2014. (full version at *arXiv:1402.3210*).
[7] M. Borgerding, P. Schniter, and S. Rangan, "Generalized approximate message passing for cosparse analysis compressive sensing," *arXiv:1312.3968*, 2013.