Compressive Phase Retrieval via Generalized Approximate Message Passing

Philip Schniter

THE OHIO STATE UNIVERSITY

Joint work with Sundeep Rangan

Supported in part by NSF grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090.

FFT Workshop – 2/23/2013

Phase Retrieval

- Goal: Recover signal $\boldsymbol{x}_0 \in \mathbb{C}^n$ from m magnitude-only measurements

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}|,$$

where $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ is a known linear transform and $\boldsymbol{w} \in \mathbb{C}^m$ is noise.

- Motivation: In many applications, it feasible to measure the intensity, but not the phase, of the Fourier transform of the signal-of-interest:
  - X-ray crystallography,
  - transmission electron microscopy,
  - coherent diffractive imaging,
  - astronomical imaging, etc.

- Feasibility: To make the solution to $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|$ unique (up to a global phase) w.p.1, $m = 4n - o(n)$ i.i.d Gaussian measurements are necessary [Heinosaari/Mazzarella/Wolf'11] and $m = 4n - 2$ are sufficient [Balan/Casazza/Edidin'06].

Phase Retrieval: Classical Approaches

Most classical approaches are iterative in nature. For example,

- Alternate between...
  - projecting $\boldsymbol{A}\hat{\boldsymbol{x}}$ onto the magnitude constraint $\boldsymbol{y}$, yielding $\hat{\boldsymbol{z}}$,
  - projecting $\boldsymbol{A}^{+}\hat{\boldsymbol{z}}$ onto an apriori known support set, yielding $\hat{\boldsymbol{x}}$.

However, due to the non-convexity of the first projection, it is easy for such algorithms to get trapped in local minima.

Phase Retrieval: Convex Approaches

Recently, some convex relaxations have been proposed.

- Noting that $y_i^2 = |\boldsymbol{a}_i^{\mathsf{H}} \boldsymbol{x}|^2 = \mathrm{tr}(\boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{H}} \boldsymbol{X})$ for $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}$, pose as "$\min_{\boldsymbol{X} \succeq 0} \mathrm{rank}(\boldsymbol{X})$ s.t. $\mathrm{tr}(\boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{H}} \boldsymbol{X}) = y_i^2$ for $i = 1 \ldots m$." (NP hard!)

  Relax to "$\min \mathrm{tr}(\boldsymbol{X})$ s.t. $\mathrm{tr}(\boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{H}} \boldsymbol{X}) = y_i^2$ for $i = 1 \ldots m$," (convex!) known as PhaseLift [Candes/Strohmer/Voroninski'11].
- Another semidefinite program (with similar performance) known as PhaseCut was proposed in [Waldspurger/D'Aspremont/Mallat'12].

It was recently shown [Candes/Li'12] that

- with very high probability, PhaseLift perfectly recovers an arbitrary $\boldsymbol{x}$ from $m \geq c_0 n$ noiseless measurements, where c_0 is a constant,
- and also that PhaseLift can be made robust to noise.

Compressive Phase Retrieval

- Recall that $m \geq 4n - o(n)$ magnitude measurements are needed for $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|$ to have a unique (up to a phase) solution for $\boldsymbol{x} \in \mathbb{C}^n$.
- Sometimes we can only afford $m \ll 4n$ magnitude measurements, in which case the problem becomes one of compressive phase retrieval.
- For successful compressive phase retrieval (CPR), one needs to leverage additional structure in $\boldsymbol{x}$, such as sparsity.

Compressive Phase Retrieval: Prior Work

- Assuming knowledge of $\|\boldsymbol{x}_0\|_1$, [Moravec/Romberg/Baraniuk'07]
  - appended this constraint onto the classical RAAR algorithm, and
  - used RIP-based arguments to establish that $m \gtrsim k^2 \log(n/k^2)$ magnitude measurements suffice for recovery.

  However, the algorithm was prone to local minima and slow convergence. Also, knowledge of $\|\boldsymbol{x}_0\|_1$ is rarely available in practice.

- Taking a convex approach, [Ohlsson/Yang/Dong/Sastry'12] proposed the following generalization of PhaseLift, which they call CPRL:

  $$\min_{\boldsymbol{X}\succeq 0} \operatorname{tr}(\boldsymbol{X}) + \lambda \|\boldsymbol{X}\|_1 + \mu \sum_{i=1}^m \left| \operatorname{tr}(\boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{H}} \boldsymbol{X}) - y_i^2 \right|^2,$$

  and performed both RIP and mutual coherence analyses. Seems promising...

Bring out the GAMP

Zed: Bring out the Gimp.

Maynard: Gimp's sleeping.

Zed: Well, I guess you're gonna have to go wake him up now, won't you?

—Pulp Fiction, 1994.

We propose a new approach to CPR based on generalized approximate message passing (GAMP).

Numerical results show

- excellent phase transitions,
- excellent NMSE & robustness to noise,
- excellent runtime,

enabling, e.g., practical compressive image retrieval.

Preliminary Numerical Results ... as Motivation

For these numerical results we generated random. . .

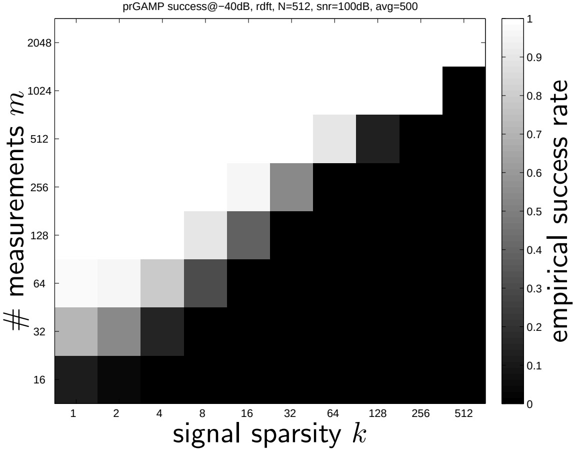
- signals $\boldsymbol{x}_0$ as k-sparse, $n = 512$-length, Bernoulli-circular-Gaussian,
- matrices $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{F}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{m \times n}$ is i.i.d circular Gaussian and $\boldsymbol{F}$ is the $n \times n$ DFT matrix,
- noise $\boldsymbol{w}$ as circular white Gaussian (added prior to taking magnitude),

and we monitored the phase-corrected normalized reconstruction MSE

$$\text{NMSE} \triangleq \min_{\theta} \frac{\|\hat{\boldsymbol{x}} - e^{\mathrm{j}\theta}\boldsymbol{x}_0\|_2^2}{\|\boldsymbol{x}_0\|_2^2}.$$

Phase transition

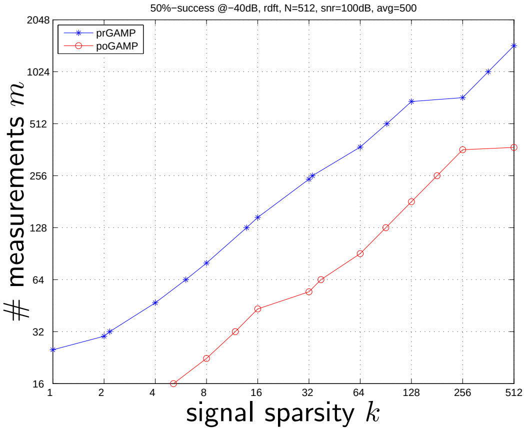
PR-GAMP's empirical success rate, averaged over 500 realizations, was



where success $\triangleq \{\text{NMSE} < 10^{-4}\}$.

Comparison to phase-oracle GAMP

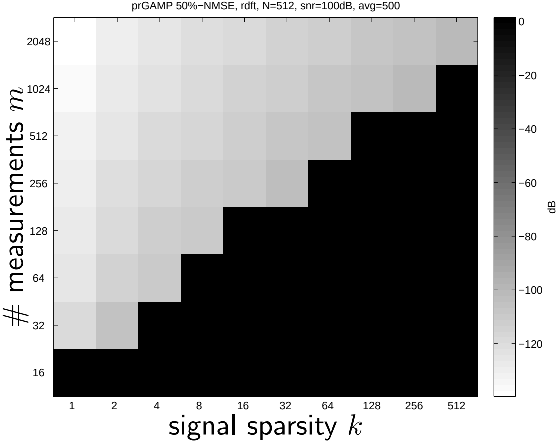
Comparing the 50%-success contours of PR- and phase-oracle GAMP:



we see that PR-GAMP requires about $4\times$ the number of measurements as phase-oracle GAMP. (Very interesting!)

NMSE versus Measurements & Sparsity

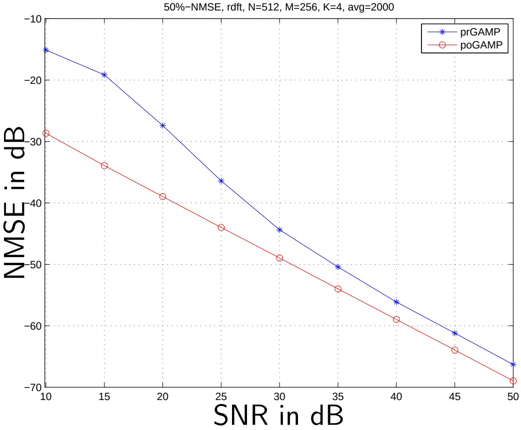
PR-GAMP's median NMSE, measured over the same 500 realizations, was



showing that recovery is very accurate above the phase transition.

Noise Robustness of PR-GAMP

The median NMSE, measured over 2000 realizations:



shows that PR-GAMP loses about 3 dB at medium-to-high SNR.

Comparison to CPRL [Ohlsson/Yang/Dong/Sastry'12]

Empirical success rate (and median runtime) over 100 realizations:

$k = 1$:

	$(m,n) = (20,32)$	$(m,n) = (30,48)$	$(m,n) = (40,64)$
CPRL	0.96 (4.9 sec)	0.97 (51 sec)	0.99 (291 sec)
PR-GAMP	0.83 (0.4 sec)	0.94 (0.3 sec)	0.99 (0.3 sec)

$k = 2$:

	$(m,n) = (20,32)$	$(m,n) = (30,48)$	$(m,n) = (40,64)$
CPRL	0.55 (5.8 sec)	0.55 (58 sec)	0.58 (316 sec)
PR-GAMP	0.72 (0.4 sec)	0.92 (0.3 sec)	1.0 (0.3 sec)

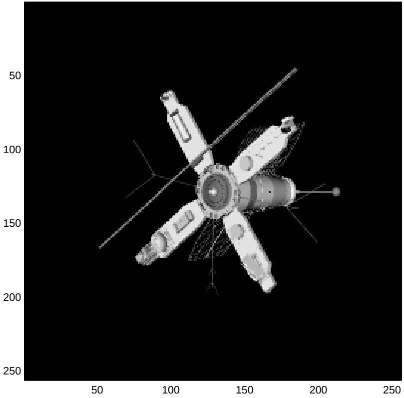
Note:

- CPRL runtime limited us to these toy problems.
- CPRL succeeds when sparsity $k = 1$, but not when $k \geq 2$.
  GAMP instead suffers when problem dimensions are very small.
- CPRL's runtime grows very quickly with problem dimensions!
  GAMP's runtime is invariant to the dimension of these toy problems.

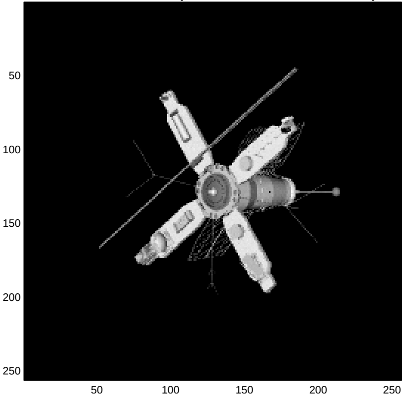
Compressive Image Recovery

65536 image pixels, 32768 measurements, 30dB SNR:

original image



PR-GAMP (-29.7dB NMSE)



PR-GAMP runtime: only 11.1 sec.

Compressive Image Recovery: Details

- Measurements were collected using

$$A = \begin{bmatrix} B_1 & \\ & B_2 \end{bmatrix} \begin{bmatrix} F & \\ & F \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}$$

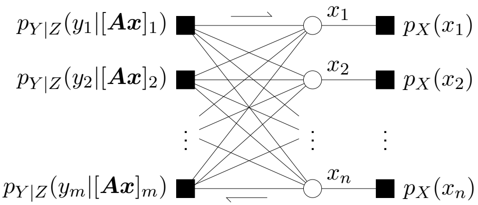
  with banded i.i.d-Gaussian B_i (10 nonzero entries per column), Fourier F, and binary masks M_i.

- Over 100 random measurement & noise realizations, we observed
  - 89% success rate, where "success" meant NMSE<-27 dB, and
  - median runtime of 13.4 sec.

Phase-Retrieval GAMP [Schniter/Rangan'12]

So what's the approach?

1. Formulate as a Bayesian inference problem by assuming
   - $y_i = \big| \underbrace{[\boldsymbol{A}\boldsymbol{x}]_i}_{z_i} + w_i \big| \;\forall i$
   - $w_i \sim \mathcal{CN}(0, \nu^w)$ i.i.d
   - $p(\boldsymbol{x}) = \prod_j p_X(x_j)$ for sparsity promoting p_X

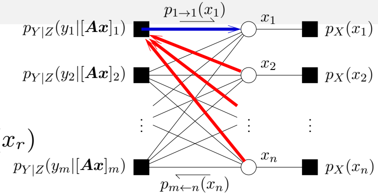


2. Use GAMP, a state-of-the-art loopy belief propagation method, to approximate the marginal posterior pdfs $\{p_{X_j|\boldsymbol{Y}}(\cdot|\boldsymbol{y})\}_{j=1}^n$.

Generalized Approximate Message Passing (GAMP)

- The evolution of GAMP:
  - The original AMP [Donoho/Maleki/Montanari'09] solves the LASSO problem $\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1$ popular in compressive sensing, i.e., MAP estimation of i.i.d Laplacian signal, thru dense $\boldsymbol{A}$, in AWGN.
  - The Bayesian AMP [Donoho/Maleki/Montanari'10] extended the above to a generic i.i.d signal prior and MMSE estimation.
  - The generalized AMP [Rangan'10] extended the above to generic i.i.d likelihoods $p_{Y|Z}(y_i|\boldsymbol{a}_i^{\mathsf{H}}\boldsymbol{x})$, for both MAP and MMSE inference.
- In the end, GAMP produces a sophisticated iterative thresholding alg, whose complexity is dominated by one application of $\boldsymbol{A}$ and $\boldsymbol{A}^{\mathsf{H}}$ per iteration with relatively few iterations (e.g., tens). Very fast!
- Rigorous large-system analyses (under i.i.d sub-Gaussian $\boldsymbol{A}$) have established that GAMP follows a state-evolution trajectory whose fixed-points have nice properties [Rangan'10], [Javanmard/Montanari'12].

GAMP Heuristics (Sum-Product)



1. Message from y_i node to x_j node:

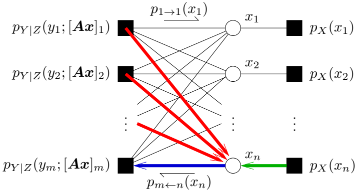
$$p_{i\to j}(x_j) \propto \int_{\{x_r\}_{r\neq j}} p_{Y|Z}\big(y_i; \overbrace{\textstyle\sum_r a_{ir}x_r}^{\approx\,\mathcal{CN}\text{ via CLT}}\big) \prod_{r\neq j} p_{i\leftarrow r}(x_r)$$

$$\approx \int_{z_i} p_{Y|Z}(y_i; z_i)\, \mathcal{CN}\big(z_i; \hat{z}_i(x_j), \nu_i^z(x_j)\big) \approx \mathcal{CN}$$

   To compute $\hat{z}_i(x_j), \nu_i^z(x_j)$, the means and variances of $\{p_{i\leftarrow r}\}_{r\neq j}$ suffice, thus Gaussian message passing!

   Remaining problem: we have $2mn$ messages to compute (too many!).

2. Exploiting similarity among the messages $\{p_{i\leftarrow j}\}_{i=1}^m$, GAMP employs a Taylor-series approximation of their difference, whose error vanishes as $m\to\infty$ for dense $\boldsymbol{A}$ (and similar for $\{p_{i\to j}\}_{j=1}^n$ as $n\to\infty$). Finally, need to compute only $\mathcal{O}(m+n)$ messages!



The GAMP Algorithm

Require: Matrix $\boldsymbol{A}$, sum-prod $\in$ {true,false}, initializations $\hat{\boldsymbol{x}}^0$, $\boldsymbol{\nu}_x^0$

$t = 0, \quad \hat{\boldsymbol{s}}^{-1} = \boldsymbol{0}, \quad \forall ij : S_{ij} = |A_{ij}|^2$

repeat

$\boldsymbol{\nu}_p^t = \boldsymbol{S}\boldsymbol{\nu}_x^t, \quad \hat{\boldsymbol{p}}^t = \boldsymbol{A}\hat{\boldsymbol{x}}^t - \hat{\boldsymbol{s}}^{t-1}.\boldsymbol{\nu}_p^t$ (gradient step)

if sum-prod **then**

$\forall i : \nu_{z_i}^t = \text{var}(Z_i|y_i), \ \hat{z}_i^t = \text{E}(Z_i|y_i)$ for $p_{Z_i|Y_i}(z|y) \propto p_{Y|Z}(y|z)\mathcal{CN}(z; \hat{p}_i^t, \nu_{p_i}^t)$

else

$\forall i : \nu_{z_i}^t = \nu_{p_i}^t \text{prox}'_{-\nu_{p_i}^t \log p_{Y|Z}(y_i,.)}(\hat{p}_i^t) \quad \hat{z}_i^t = \text{prox}_{-\nu_{p_i}^t \log p_{Y|Z}(y_i,.)}(\hat{p}_i^t),$

end if

$\boldsymbol{\nu}_s^t = (1 - \boldsymbol{\nu}_z^t./\boldsymbol{\nu}_p^t)./\boldsymbol{\nu}_p^t, \quad \hat{\boldsymbol{s}}^t = (\hat{\boldsymbol{z}}^t - \hat{\boldsymbol{p}}^t)./\boldsymbol{\nu}_p^t$ (dual update)

$\boldsymbol{\nu}_r^t = 1./(\boldsymbol{S}^T\boldsymbol{\nu}_s^t), \quad \hat{\boldsymbol{r}}^t = \hat{\boldsymbol{x}}^t + \boldsymbol{\nu}_r^t.\boldsymbol{A}^T\hat{\boldsymbol{s}}^t$ (gradient step)

if sum-prod **then**

$\forall j : \nu_{x_j}^t = \text{var}(X_j|\hat{r}_j^t), \ \hat{z}_j^t = \text{E}(X_j|\hat{r}_j^t)$ for $p_{X_j|R_j}(x|r) \propto p_X(x)\mathcal{CN}(x; r, \nu_{r_j}^t)$

else

$\forall j : \nu_{x_j}^{t+1} = \nu_{r_j}^t \text{prox}'_{-\nu_{r_j}^t \log p_X(.)}(\hat{r}_j^t) \quad \hat{x}_j^{t+1} = \text{prox}_{-\nu_{r_j}^t \log p_X(.)}(\hat{r}_j^t),$

end if

$t \leftarrow t+1$

until Terminated

Note connections to Arrow-Hurwicz, primal-dual, ADMM, proximal FB splitting,. . .

GAMP for Phase Retrieval: Likelihood

To apply GAMP to phase retrieval, we need a likelihood function $p_{Y|Z}(\cdot|\cdot)$ relating the noisy magnitude measurements $\{y_i\}_{i=1}^m$ to the corresponding noiseless transform outputs $\{z_i\}_{i=1}^m$ (recalling that $z_i \triangleq [\boldsymbol{Ax}]_i$).

- When Z and W are both circular, one can show that

$$Y = |Z + W| \quad \Leftrightarrow \quad Y = e^{j\Theta}(Z + W)\big|_{\Theta \sim \mathcal{U}[0,2\pi)}$$

  in the sense that both models yield the same $p_{Z|Y}(\cdot|\cdot)$.

- Assuming $W \sim \mathcal{CN}(0, \nu^w)$, we then margin out Θ to obtain

$$p_{Y|Z}(y|z) = \frac{1}{\pi\nu^w} e^{-\frac{(|y|-|z|)^2}{\nu^w}} I_0(\rho) e^{-\rho} \quad \text{for} \quad \rho \triangleq \frac{2|y|\,|z|}{\nu^w},$$

  where $I_0(\cdot)$ is the 0^{th}-order modified Bessel function of the first kind.

Other models are also possible, e.g., $Y = |Z| + W$ or $Y = |Z|^2 + W$.

GAMP for Phase Retrieval: Signal Prior

For *compressive* phase retrieval, we need a structured signal prior $p_{\boldsymbol{X}}(\cdot)$.

- Separable priors constrain $p_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{j=1}^{n} p_X(x_j)$ with, e.g.,
  - sparsity promotion: $p_X(x_j) = \lambda f_X(x_j) + (1-\lambda)\delta(x_j)$
  - real-valuedness: $p_X(x_j)$ supported on $x_j \in \mathbb{R}$
  - non-negativity: $p_X(x_j)$ supported on $x_j \in \mathbb{R}^+ \cup \{0\}$

  and are directly supported by GAMP.

- Non-separable priors model structure across $\{x_j\}$, e.g.,
  - structured sparsity: $\begin{cases} p_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{\boldsymbol{s}\in\{0,1\}^n} p_{\boldsymbol{S}}(\boldsymbol{s}) \prod_{j=1}^{n} p_{X|S}(x_j|s_j) \\ p_{\boldsymbol{S}}(\boldsymbol{s}) = \text{block, Markov field/chain/tree,...} \end{cases}$

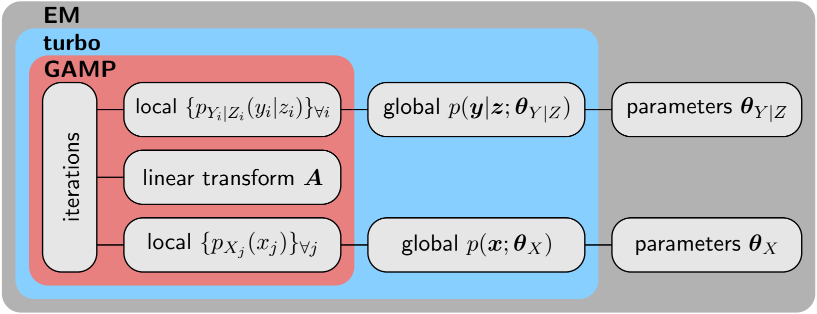
  but are not directly supported by GAMP.

- In any case, we want the assumed $p_{\boldsymbol{X}}(\cdot)$ to match the empirical distribution of the true $\{x_j\}_{j=1}^{n}$, which is apriori unknown.

Making GAMP Practical: EM & turbo Extensions

- The basic GAMP algorithm is limited by two major assumptions:
  1. separable $p(\boldsymbol{y}|\boldsymbol{z}) = \prod_i p_{Y_i|Z_i}(y_i|z_i)$ and $p(\boldsymbol{x}) = \prod_j p_{X_j}(x_j)$
  2. that are well matched to the data.

- The EM-turbo-GAMP framework circumvents these limitations by learning [Vila/Schniter'12] possibly non-separable [Schniter'10] priors:



PR-GAMP: Ongoing Work

PR-GAMP is a work-in-progress. Things we are working on include:

- Derivation of the state evolution.
- Automatic learning of signal prior $p_X(\cdot)$ via the EM-GM approach from [Vila/Schniter'12].
- Exploitation of the hidden-Markov-tree support structure of natural images via the turbo approach from [Som/Schniter'10].
- MAP formulation of PR-GAMP.
- Connections to optimization algorithms.

Conclusions

- (Compressive) phase retrieval is a longstanding problem that is experiencing a rebirth through compressive sensing and convex relaxation.
- We proposed a new approach to CPR based on generalized approximate message passing (GAMP).
- Empirical results show an excellent phase transition ($4\times$meas of phase-oracle), excellent noise robustness (~ 3 dB worse than phase-oracle), and excellent runtime (many orders of magnitude faster than convex relaxation).
- As a practical demonstration, we accurately recovered a 64k-pixel image from 32k noisy measurements in only 11 seconds.

All of these methods are integrated into GAMPmatlab:
`http://sourceforge.net/projects/gampmatlab/`

Thanks!

Bibliography

1. T. Heinosaari, L. Mazzarella, M.M. Wolf, "Quantum tomography under prior information" arXiv:1109.5478, Sep. 2011.
2. R. Balan, P.G. Casazza, D. Edidin, "On signal reconstruction without phase," *ACHA* 2006.
3. E.J. Candès, T. Strohmer, V. Voroninski, "Phase Lift: Exact and stable signal recovery from magnitude measurements via convex programming," arXiv:1109.4499, Sep. 2011.
4. I. Waldspurger, A. D'Aspremont, S. Mallat, "Phase recovery, MaxCut and complex semidefinite programming," arXiv:1206.0102, June 2012.
5. E.J. Candès, X. Li, "Solving quadratic equations via PhaseLift when there are about as many equations as unknowns," arXiv:1208.6247, Aug. 2012.
6. M.L. Moravec, J.K. Romberg, R. Baraniuk, "Compressive phase retrieval,", *Proc. SPIE*, 2007.
7. H. Ohlsson, A.Y. Yang, R. Dong, S.S. Sastry, "Compressive phase retrieval from squared output measurements via semidefinite programming," arXiv:1111.6323, Nov. 2011.
8. P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *Proc. Allerton*, Oct. 2012.
9. D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *PNAS*, Nov. 2009.
10. D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction," *Proc. ITW*, Jan. 2010.
11. S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *Proc. ISIT, Aug. 2011* (see also arXiv:1010.5141)
12. A. Javanmard and A. Montanari, "State Evolution for general approximate message passing algorithms, with applications to spatial coupling," arXiv:1211.5164, Sep. 2012.
13. P. Schniter, "Turbo reconstruction of structured sparse signals," *Proc. CISS*, Mar. 2010.
14. J. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *Proc. CISS*, Mar. 2012 (see also *arXiv:1207.3107*)
15. S. Som and P. Schniter, "Compressive imaging using approximate message passing and a Markov-tree prior," *IEEE Trans. SP*, July 2012.