Construction of optimal spectral methods in phase retrieval

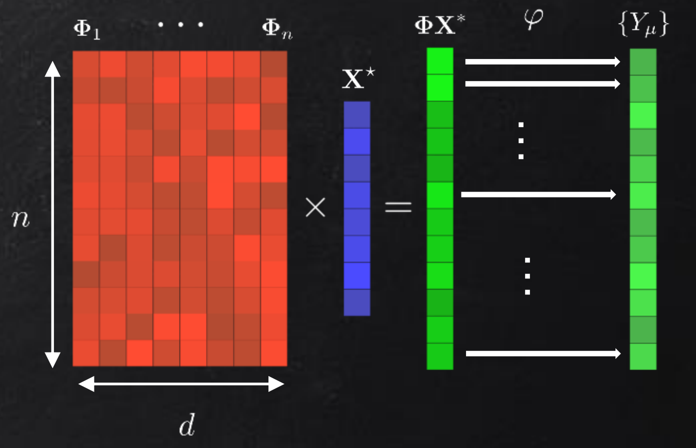
arXiv:2012.04524

Antoine Maillard

Joint work with Florent Krzakala, Yue M. Lu & Lenka Zdeborová

Mathematical and Scientific Machine Learning – August 16th 2021

Phase retrieval



Goal : Recover a d-dimensional signal $\mathbf{X}^\star$ from n data points $\{\mathbf{\Phi}_\mu, Y_\mu\}_{\mu=1}^n$ generated as :

Generalized Linear Model (GLM)

Observations $Y_\mu \in \mathbb{R}$

Real / Complex
$\beta = 1$ $\beta = 2$

$$Y_\mu \sim P_{\rm out}\left(\cdot \middle| \frac{1}{\sqrt{d}} \sum_{i=1}^d \Phi_{\mu i} X_i^\star\right) \quad \mu \in \{1, \cdots, n\}$$

(Probabilistic) channel with possible noise.

Sensing matrix (real/complex)

Signal (real/complex), d-dimensional

In phase retrieval, one measures the modulus $P_{\rm out}(y|z) = P_{\rm out}(y||z|)$, e.g. noiseless $Y_\mu = \frac{1}{d}|(\mathbf{\Phi}\mathbf{X}^\star)_\mu|^2$; Poisson-noise $Y_\mu \sim \mathrm{Pois}(\Lambda|(\mathbf{\Phi}\mathbf{X}^\star)_\mu|^2/d)$.

Arises in signal processing, statistical estimation, optics, X-ray crystallography, astronomy, microscopy...

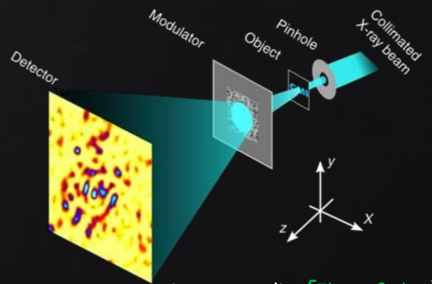


Image credits: [Zhang&al 16]

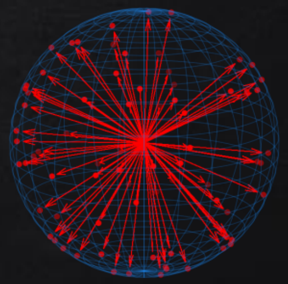
How to solve this problem efficiently in high dimensions ? $n, d \to \infty$

- SDP relaxations [Candès&al '15a&b, Waldspurger&al '15, Goldstein&al '18, ...]
- Non-convex optimization procedures [Netrapalli&al '15, Candès&al '15c, ...]
- Approximate Message-Passing [Barbier&al '19, A.M.&al '20]

Computationally heavy / Need informed initialization → Spectral methods

[Mondelli&al '18, Luo&al '18, Dudeja&al '19,...]

Construction of spectral methods



Our model: The matrix $\mathbf{\Phi}$ is right-orthogonally (unitarily) invariant, i.e. delocalized right-eigenvectors : $\forall \mathbf{U}, \;\; \mathbf{\Phi} \stackrel{d}{=} \mathbf{\Phi U}$

The bulk of eigenvalues of $\mathbf{\Phi}^\dagger \mathbf{\Phi}/d$ converges to a distribution $\nu(x)$, as $n, d \to \infty$ with $n/d \to \alpha > 0$.

Examples: Gaussian matrices, product of Gaussians, random column-orthogonal/unitary, any $\mathbf{\Phi} \equiv \mathbf{USV}^\dagger$ with $S_i^2 \stackrel{\text{i.i.d.}}{\sim} \nu$.

Given a phase retrieval problem, we want an optimal spectral method (among all possible ones) in terms of estimation error:

$$\mathrm{MSE} \equiv \frac{1}{d\rho} \|\mathbf{X}^* - \hat{\mathbf{X}}_{\text{spectral}}\|^2$$

This talk: Three different strategies, related to the statistical physics approach to high-dimensional inference.

- Method I : "Naïve" generalization of what is known for Gaussian matrices.
- Method II : Linearization of message-passing algorithms.
- Method III : Bethe Hessian analysis from the Thouless-Anderson-Palmer [TAP77] free energy.

Method I

[Chen&al '15, Wang&al '16, Zhang&al '17, Mondelli&al '19, Luo&al '19, Dudeja&al '19]

$$y_\mu \sim P_{\text{out}}\Big(\cdot \Big| \frac{1}{\sqrt{d}} \sum_{i=1}^{d} \Phi_{\mu i} X_i^*\Big)$$

Most previous works reduced to methods of the type

$$\mathbf{M}(\mathcal{T}) \equiv \frac{1}{d} \sum_{\mu=1}^{n} \mathcal{T}(y_\mu) \boldsymbol{\Phi}_\mu \boldsymbol{\Phi}_\mu^\dagger$$

For Gaussian matrices $\boldsymbol{\Phi}$ the optimal method in this class is given by

$$\mathcal{T}^\star_{\text{Gaussian}}(y) \equiv \frac{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho)}{1 + \rho \partial_\omega g_{\text{out}}(y_\mu, 0, \rho)}$$

- In noiseless phase retrieval one has $\mathcal{T}^\star_{\text{Gaussian}}(y) = 1 - 1/y$.
- We can naively use it for all matrices: $\mathbf{M}_{\text{naive}} \equiv \mathbf{M}(\mathcal{T}^*_{\text{Gaussian}})$

$$\partial_\omega g_{\text{out}}(y_\mu, 0, \sigma^2) = -\frac{1}{\sigma^2} + \frac{1}{\sigma^4} \frac{\int_{\mathbb{K}} \mathrm{d}x \, e^{-\frac{\beta}{2\sigma^2}|x|^2} \, |x|^2 \, P_{\text{out}}(y_\mu|x)}{\int_{\mathbb{K}} \mathrm{d}x \, e^{-\frac{\beta}{2\sigma^2}|x|^2} \, P_{\text{out}}(y_\mu|x)}$$

$(\mathbb{K} = \mathbb{R}, \mathbb{C})$

Method II

[Schniter&al '16, A.M.&al '20]: For GLMs with rotationally-invariant matrices, the best-known polynomial-time algorithm is *Generalized Vector Approximate Message-Passing* (G-VAMP).

↓ Linearization procedure

Linearized Approximate Message-Passing (LAMP) spectral method.

$$\mathbf{M}_{\text{LAMP}} \equiv \frac{\rho \langle \lambda \rangle_\nu}{\alpha} \Big(\frac{\alpha}{\langle \lambda \rangle_\nu} \frac{\boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger}{d} - \mathrm{I}_n \Big) \mathrm{Diag}(\{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho \langle \lambda \rangle_\nu / \alpha)\}) \longrightarrow \hat{\mathbf{x}} \equiv \frac{\boldsymbol{\Phi}^\dagger \mathrm{Diag}(\{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho \langle \lambda \rangle_\nu / \alpha)\}) \hat{\mathbf{u}}}{\|\boldsymbol{\Phi}^\dagger \mathrm{Diag}(\{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho \langle \lambda \rangle_\nu / \alpha)\}) \hat{\mathbf{u}}\|} \sqrt{d\rho}.$$

$\mathbf{M}_{\text{LAMP}}$ is a $n \times n$ non-Hermitian matrix (complex spectrum).

$\hat{\mathbf{u}}$: top eigenvector of $\mathbf{M}_{\text{LAMP}}$.

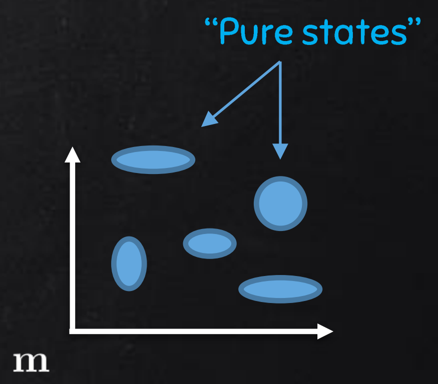
Similar approaches in community detection [Krzakala&al '13], phase retrieval with unitary matrices [Ma&al '21] or spiked matrix estimation [Aubin&al '20].

Method III: TAP landscape and Bethe Hessian

Thouless–Anderson–Palmer approach [TAP77]

- The posterior measure of $\mathbf{x}|\mathbf{Y}$ (the *Gibbs measure*) decomposes along pure states.
- These pure states can be found by "tilting" the measure, imposing $m_i = \langle x_i \rangle$ and $\sigma_i^2 = \mathrm{Var}(x_i)$:

  They are the maxima of the free entropy of this constrained measure, as a function of $(\mathbf{m}, \boldsymbol{\sigma})$.



- TAP free entropy for rotationally-invariant generalized linear models derived in [A.M.&al '19], generalizing [Parisi&Potters '95]:

Involved but explicit!

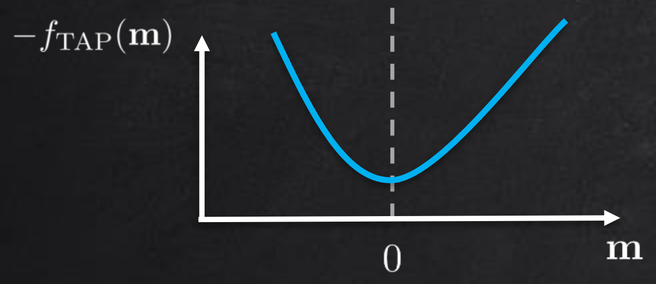
$$f_{\mathrm{TAP}}(\mathbf{m}) = \sup_{\sigma \geq 0} \sup_{\substack{\mathbf{g} \in \mathbb{K}^n \\ r \geq 0}} \mathrm{extr}_{\substack{\boldsymbol{\omega} \in \mathbb{K}^n \\ b \geq 0}} \mathrm{extr}_{\substack{\boldsymbol{\lambda} \in \mathbb{K}^d \\ \gamma \geq 0}} \Big[\frac{\beta}{d} \sum_{i=1}^d \lambda_i \cdot m_i + \frac{\beta\gamma}{2d} \Big(d\sigma^2 + \sum_{i=1}^d |m_i|^2 \Big) - \frac{\beta}{d} \sum_{\mu=1}^n \omega_\mu \cdot g_\mu - \frac{\beta b}{2d} \Big(\sum_{\mu=1}^n |g_\mu|^2 - \alpha d r \Big) + \frac{1}{d} \sum_{i=1}^d \ln \int_{\mathbb{K}} P_0(\mathrm{d}x) e^{-\frac{\beta\gamma}{2}|x|^2 - \beta \lambda_i \cdot x}$$

$$+ \frac{\alpha}{n} \sum_{\mu=1}^n \ln \int_{\mathbb{K}} \frac{\mathrm{d}h}{\left(\frac{2\pi b}{\beta}\right)^{\beta/2}} P_{\mathrm{out}}(y_\mu | h) e^{-\frac{\beta|h - \omega_\mu|^2}{2b}} + \frac{\beta}{d} \sum_{i=1}^d \sum_{\mu=1}^n g_\mu \cdot \Big(\frac{\Phi_{\mu i}}{\sqrt{d}} m_i \Big) + \beta F(\sigma^2, r) \Big].$$

$$F(x, y) \equiv \inf_{\zeta_x, \zeta_y > 0} \Big[\frac{\zeta_x x}{2} + \frac{\alpha \zeta_y y}{2} - \frac{\alpha - 1}{2} \ln \zeta_y - \frac{1}{2} \langle \ln(\zeta_x \zeta_y + \lambda) \rangle_\nu \Big] - \frac{1}{2} \ln x - \frac{\alpha}{2} \ln y - \frac{1 + \alpha}{2}.$$

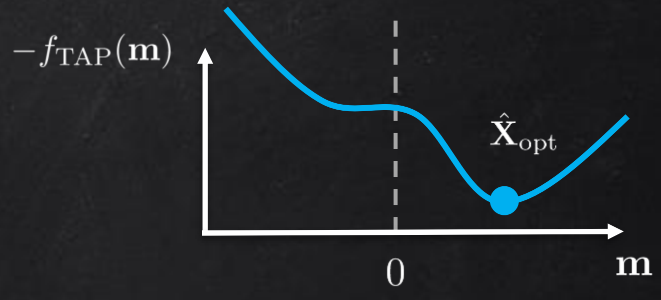
Weak-recovery impossible

Global maximum of f_{TAP} in $\mathbf{m}=0$: the uninformative "paramagnetic" point.



Weak-recovery possible

$\mathbf{m}=0$ is an unstable stationary point of f_{TAP}, which has a global maximum in $\mathbf{m}\neq 0$ (optimal estimator).



A spectral method can only use the physical information available in the uninformative point $\mathbf{m}=0$.

Compute the Hessian of f_{TAP} at the paramagnetic point.

Constructive derivation of a spectral method that is conjectured to be optimal.

TAP – Bethe Hessian spectral method.

$$\mathbf{M}_{\mathrm{TAP}} \equiv -d\nabla^2 f_{\mathrm{TAP}}(\mathbf{m}=0) = -\frac{1}{\rho}\mathrm{I}_d + \frac{1}{d}\sum_{\mu=1}^{n} \frac{\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}{1+\frac{\rho\langle\lambda\rangle_\nu}{\alpha}\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)} \mathbf{\Phi}_\mu \mathbf{\Phi}_\mu^\dagger$$

Similar to previous strategies in community detection. [Saade&al '14]

Optimal spectral method

From the Bethe Hessian analysis

$$\mathrm{M}(\mathcal{T}) \equiv \frac{1}{d}\sum_{\mu=1}^{n} \mathcal{T}(y_\mu)\mathbf{\Phi}_\mu \mathbf{\Phi}_\mu^\dagger$$

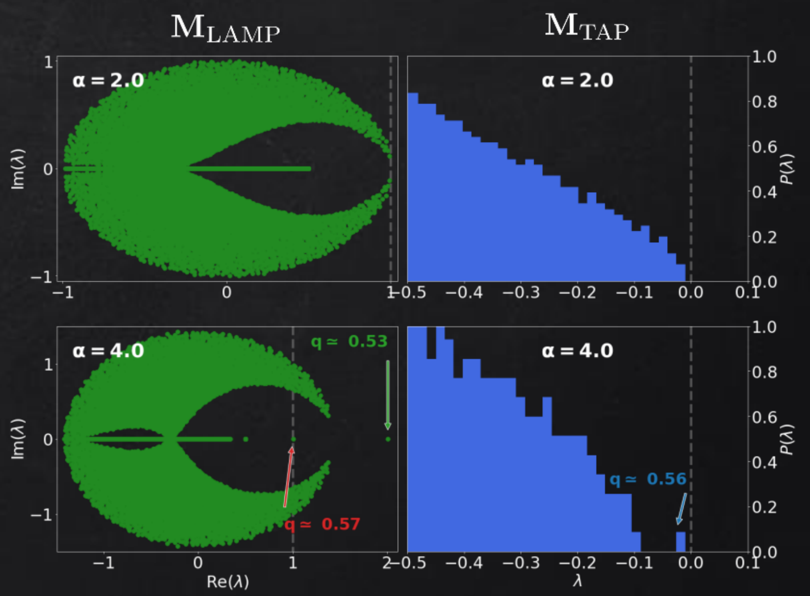
Main conjecture: For any right-orthogonally invariant sensing matrix, the optimal spectral method (in terms of weak-recovery threshold and achieved error) belongs to the class of matrices $\mathrm{M}(\mathcal{T})$ and is attained in:

$$\mathcal{T}^*(y) = \frac{\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}{1 + \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}$$

- The optimal method is the “naïve” generalization of Method I.
- We did not assume anything on the form of the method: we confirm the validity of the restriction of previous works on spectral methods to the class of matrices $\mathrm{M}(\mathcal{T})$!
- The optimal spectral method does not depend on the spectrum of the sensing matrix (apart from a global scaling), nor on the sampling ratio α!

➡ Consequences for practitioners: one only needs to know the observation channel to construct the method!

Complex Gaussian Φ and Poisson noise $P_{\mathrm{out}}(y|z) = e^{-\Lambda|z|^2}\sum_{k=0}^{\infty}\delta(y-k)\frac{\Lambda^k|z|^{2k}}{k!}$

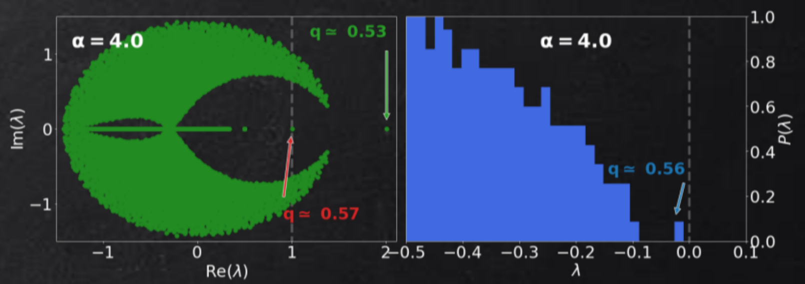


$$\Lambda = 1$$

$$q = \frac{1}{d}\sum_{i=1}^{d} X_i^* \hat{x}_i$$

- The optimal method corresponds to marginal stability in both $\mathrm{M}_{\mathrm{TAP}}$ and $\mathrm{M}_{\mathrm{LAMP}}$.
- Puzzle: the dominant eigenvector of $\mathrm{M}_{\mathrm{LAMP}}$ is a suboptimal estimator!

Spectral methods performance



Noiseless complex phase retrieval $Y_\mu = \frac{1}{d}|\mathbf{\Phi X}^*|^2$

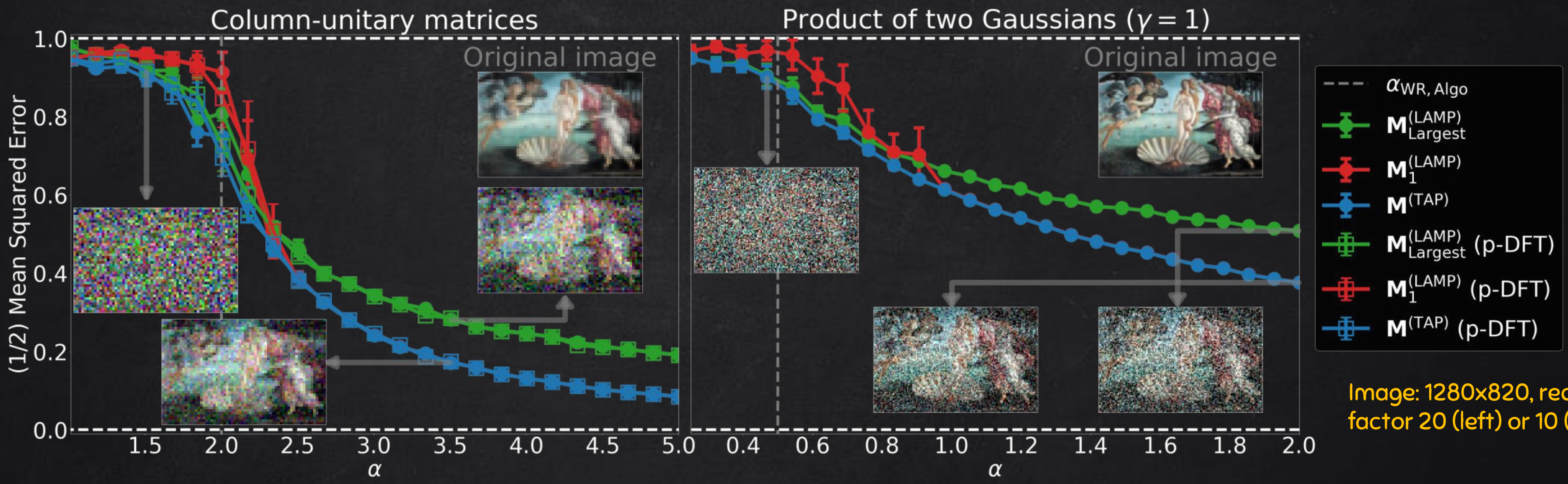


Image: 1280x820, reduced by a factor 20 (left) or 10 (right).

- $\hat{x}_{\mathrm{LAMP}}(\lambda = 1) \iff \hat{x}_{\mathrm{TAP}}$, achieving the best overlap. Otherwise $\hat{x}_{\mathrm{LAMP}}(\lambda_{\max})$ is suboptimal in terms of MSE.
- Our theory stays valid for matrices with controlled structure (partial DFT $\equiv$ randomly subsampled DFT).
- For partial DFT matrices, we use the method as initialization of a gradient-descent procedure: perfect recovery at $\alpha \in (3, 4)$, while the best polynomial-time algorithm achieves $\alpha_{\mathrm{PR}} \simeq 2, 3$ [A.M&al '20]. ➡ Very competitive while computationally cheap!

Conclusion and perspectives

Main contributions

- Constructive derivation of a conjecturally optimal spectral method in generic phase retrieval problems, in a framework that encompasses real/complex variables and a wide variety of sensing matrices.
- Our results apply to randomly subsampled DFT matrices and to real image (i.e. structured signal) recovery.
- We use two fundamentally equivalent approaches – message-passing linearization and Bethe Hessian analysis – that yield the same optimal performance, associated with a marginal stability of the linear dynamics.

Open questions remain, e.g. the "marginality vs instability" puzzle: In $\mathrm{M}_{\mathrm{LAMP}}$ the optimal method is "hidden" inside the bulk and marginally stable, while the dominant eigenvalue is unstable and suboptimal.

THANK YOU !