Algorithmic Challenges in High-Dimensional Inference Models. Insights from Statistical Physics

David Gamarnik

MIT
Operations Research Center and IDSS

MADDD Seminar

Joint work with Quan Li (MIT), Ilias Zadik (NYU), Subhabrata Sen (Harvard), Aukosh Jagannath (U of Waterloo)

Computational Challenges in High-Dimensional Infrerence

Computational Challenges in High-Dimensional Infrerence

Modern day inference challenges are characterized by

Computational Challenges in High-Dimensional Infrerence

Modern day inference challenges are characterized by

- Model size (BIG Data)
- Structure of parameters (sparsity, discreteness, low-dimensionality, etc.)

Computational Challenges in High-Dimensional Infrerence

Modern day inference challenges are characterized by

- Model size (BIG Data)
- Structure of parameters (sparsity, discreteness, low-dimensionality, etc.)
- Uncertainty (thus NP-completeness is not useful)

Computational Challenges in High-Dimensional Infrerence

Computational Challenges in High-Dimensional Infrerence

We need a new theory addressing computational hardness and tractability

Computational Challenges in High-Dimensional Infrerence

We need a new theory addressing computational hardness and tractability

- High-dim problems are computationally challenging, giving a rise to the so-called ***Information Theoretic vs Computation gap***.

Computational Challenges in High-Dimensional Infrerence

We need a new theory addressing computational hardness and tractability

- High-dim problems are computationally challenging, giving a rise to the so-called ***Information Theoretic vs Computation gap***.
- What's the reason?

We need a new theory addressing computational hardness and tractability

- High-dim problems are computationally challenging, giving a rise to the so-called ***Information Theoretic vs Computation gap***.
- What's the reason?
- **Statistical Physics** can help: topology of the solution space and phase transition, **Overlap Gap Property (OGP)** phase transition

Example I: Largest Submatrix Problem

Given $m \times n$ matrix J

$$J = \begin{bmatrix} J_{11} & \dots & J_{1n} \\ \vdots & \ddots & \vdots \\ J_{m1} & \dots & J_{mn} \end{bmatrix},$$

Find a $k \times \ell$ submatrix

$$\begin{bmatrix} J_{11} & \dots & J_{1n} \\ \vdots & \begin{bmatrix} * & \dots & * \\ \vdots & J^*_{k,\ell} & \vdots \\ * & \dots & * \end{bmatrix} & \vdots \\ J_{m1} & \dots & J_{mn} \end{bmatrix},$$

with the largest average entry $\text{Ave}(J^*_{n,k}) = \frac{1}{k\ell} \sum J_{i_{\ell_1}, j_{\ell_2}}$

Who cares?..

Who cares?..

- Bi-clustering: applications in **genetics, drug design, energy and social networks**.

Who cares?..

- Bi-clustering: applications in **genetics, drug design, energy and social networks.**
- **Genomics**: J is gene vs expression data.
  $m = 2,500 - 15,000$, $n = 10 - 150$ Madeira & Oliveira survey [2004], Fortunato et al [2010], Shabalin et al [2009].

	A	B	C	D	E	F	G	H
1	Acc ID	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	
2	NM_007818	67540.89	70924.09	80243.76	3501.2	5697.47	2426.72	
3	NM_001105160	811.93	801.36	740.71	128.67	104.42	101.33	
4	NM_028089	190.41	211.06	236.19	9.05	23.33	8.44	
5	NM_016696	66.77	57.56	101.09	750.9	659.84	491.89	
6	NM_013459	3.3	11.29	1.89	735.82	816.46	118.22	
7	NM_007809	45.34	36.12	51.02	245.27	372.13	335.67	
8	NM_009999	103.04	370.21	200.29	17.09	13.33	8.44	
9	NM_133960	7708.78	6976.38	6569.04	1731	1641.81	1853.55	
10	NM_027881	31.32	10.16	24.56	268.39	186.62	135.11	
11	NM_054053	31.32	24.83	19.84	323.68	428.78	116.11	
12	NM_007377	47.81	89.17	70.86	370.93	378.79	279.72	
13	NM_028064	703.95	689.62	662.29	214.11	168.85	144.61	
14	NM_008182	222.56	339.73	226.75	30.16	63.32	26.39	
15	NM_013661	12.36	11.29	8.5	97.51	77.76	71.78	
16	NM_007815	20613.09	25218.13	31540.46	5209.07	7680.3	6312.2	

- ☐ Finding optimal solution takes a lot of time (days)
- ☐ Many, many heuristics used (Div-Conq, Greedy, Clust-Comb, Dist-Ident, Exh-Enum)

Who cares?..

Who cares?..

- **Drug activity.** Chemical compounds vs descriptor, $m \approx 10,000$, $n \approx 30$, Liu & Wang [2003]

Table 2. Observed and calculated log RBA and values of the three selected descriptors, PW2, Mor15m and GAP-10 for the 23 progestins

ID**	name	Ref.	log RBA* Obs.	log RBA Calc.	PW2	Mor15m	GAP-10/ eV
1#	Progesterone	5	1.602	1.194	0.620	0.198	0.094
2	17-Acetoxyprogesterone	6	1.204	1.236	0.620	0.401	0.139
3	17-Hydroxyprogesterone	6	0.079	0.094	0.631	0.535	0.256
4#	21-Hydroxyprogesterone	6	1.049	0.954	0.612	0.424	0.246
5	11β-Hydroxyprogesterone	6	1.158	1.432	0.621	0.252	0.066
6	Methoxyprogesterone acetate	5	2.061	1.418	0.622	0.389	0.095
7	Chloromadinone acetate	7	1.975[a]	2.330	0.622	0.720	0.046
8	Cyproterone acetate	7	1.447[b]	1.228	0.629	0.852	0.187
9	Testosterone	6,8	−0.097	−0.009	0.624	0.227	0.246
10#	5β-Pregnane-3,20-dione	6	0.380	1.092	0.620	−0.046	0.048
11	1,4-Pregnadine-3,20-dione	6	1.318	1.502	0.620	0.341	0.086
12#	4,6-Pregnadine-3,20-dione	6	1.310	1.218	0.620	0.441	0.152
13	Promegestone (R5020)	5	2.000	2.075	0.604	0.673	0.206
14	16α-Ethyl-21-hydroxy-19-nor-4-pregnene-3,20-dione (Organon 2058)	5,9	2.544	2.557	0.597	0.454	0.133
15#	Levonorgestrel	5	2.079	2.474	0.605	0.295	0.046
16	19-Norprogesterone	6,8	1.827	1.794	0.610	0.342	0.118
17	Norethisterone	10	1.866	1.542	0.615	0.267	0.099
18	3-Keto-desogestrel	10	2.827	2.534	0.607	0.523	0.079
19#	Gestodene	10	2.799	2.249	0.605	0.277	0.074
20#	3-Keto-allylestrenol	7,9	1.886[c]	2.188	0.606	0.199	0.055
21	Norethinodrel	7	0.845[d]	1.476	0.615	0.445	0.153
22	19-Nortestosterone	6,8	0.944	0.827	0.613	0.369	0.243
23	Metribolone (R1881)	7	2.146[e]	2.111	0.625	0.945	0.112

*Confidence limits:[a](1.906-2.035);[b](1.322-1.544);[c](1.775 -1.975);[d](0.544 - 1.021);[e](2.021 - 2.276);#The seven test set molecules; **The remaining 16 molecules belong to the training set.

Theory Background

Theory Background

- [Sun, Nobel [2013],[Bhamidi, Dey & Nobel 2013]
  $J = \mathcal{N}(0, I_n)$

Theory Background

- [Sun, Nobel [2013],[Bhamidi, Dey & Nobel 2013]
  $J = \mathcal{N}(0, I_n)$
- Global optimum when $m = n, k = \ell = o(\log n)$

$$\text{Ave}(J^*_{n,k}) \approx 2\sqrt{\frac{\log n}{k}}, \qquad \text{with high probability as } n \to \infty$$

Algorithms

Algorithms

- Several algorithms analyzed in G & Li [2016]

Algorithms

- Several algorithms analyzed in G & Li [2016]
- The best of the three algorithms produces a matrix with average entry

$$\begin{aligned} \text{Ave}(C_{n,k}^{\text{ALG}}) &\approx \frac{4\sqrt{2}}{3}\sqrt{\frac{\log n}{k}} \\ &< 2\sqrt{\frac{\log n}{k}} \\ &\approx \text{OPT}. \end{aligned}$$

Algorithms

- Several algorithms analyzed in G & Li [2016]
- The best of the three algorithms produces a matrix with average entry

$$\begin{aligned}\mathrm{Ave}(C_{n,k}^{\mathrm{ALG}}) &\approx \frac{4\sqrt{2}}{3}\sqrt{\frac{\log n}{k}} \\ &< 2\sqrt{\frac{\log n}{k}} \\ &\approx \mathrm{OPT}.\end{aligned}$$

- as $4\sqrt{2}/3 = 1.885... < 2$.

Algorithms

- Several algorithms analyzed in G & Li [2016]
- The best of the three algorithms produces a matrix with average entry

$$\mathrm{Ave}(C_{n,k}^{\mathrm{ALG}}) \approx \frac{4\sqrt{2}}{3}\sqrt{\frac{\log n}{k}} < 2\sqrt{\frac{\log n}{k}} \approx \mathrm{OPT}.$$

- as $4\sqrt{2}/3 = 1.885... < 2$.
- Nothing better is known.

Algorithms

- Several algorithms analyzed in G & Li [2016]
- The best of the three algorithms produces a matrix with average entry

$$\begin{aligned} \mathrm{Ave}(C_{n,k}^{\mathrm{ALG}}) &\approx \frac{4\sqrt{2}}{3}\sqrt{\frac{\log n}{k}} \\ &< 2\sqrt{\frac{\log n}{k}} \\ &\approx \mathrm{OPT}. \end{aligned}$$

- as $4\sqrt{2}/3 = 1.885... < 2$.
- Nothing better is known.
- Problem has roots in the Largest Clique problem introduced by Karp [1977], still unsolved.

Example II: Spiked Matrix Detection

Example II: Spiked Matrix Detection

- Random matrix with sparse rank-1 signal:

$$J = \lambda\theta^* \cdot (\theta^*)^T + W, \quad \theta^* \in \{0,1\}^n, \; s - \text{sparse}, \quad W \overset{d}{=} \mathcal{N}(0, I_n)$$

Example II: Spiked Matrix Detection

- Random matrix with sparse rank-1 signal:

$$J = \lambda\theta^* \cdot (\theta^*)^T + W, \ \ \theta^* \in \{0,1\}^n, \ s - \text{sparse}, \ \ W \overset{d}{=} \mathcal{N}(0, I_n)$$

$$J = \lambda \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} + \begin{bmatrix} J_{11} & \dots & J_{1n} \\ \vdots & \ddots & \vdots \\ J_{n1} & \dots & J_{nn} \end{bmatrix},$$

Example II: Spiked Matrix Detection

- Random matrix with sparse rank-1 signal:

$$J = \lambda \theta^* \cdot (\theta^*)^T + W, \quad \theta^* \in \{0,1\}^n, \; s-\text{sparse}, \quad W \stackrel{d}{=} \mathcal{N}(0, I_n)$$

$$J = \lambda \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} + \begin{bmatrix} J_{11} & \dots & J_{1n} \\ \vdots & \ddots & \vdots \\ J_{n1} & \dots & J_{nn} \end{bmatrix},$$

- **Problem:** Learn θ^* from J.

Background

Background

- Problem studied widely in recent years:
  Lelarge, Miolane [2009]
  Butucea & Ingster [2013],
  Deshphande & Montanari [2014]
  Lesieur, Krzakala, Zdeborova [2015,2017]
  Dia, Macris, Krzakala, Lesieur, Zdeborova [2016]
  Montanari, Reichman, Zeitouni [2015]
  Jagannath, Lopato, Milane [2018]
  G, Jagannath, Sen [2019]

Background

- Problem studied widely in recent years:
  Lelarge, Miolane [2009]
  Butucea & Ingster [2013],
  Deshphande & Montanari [2014]
  Lesieur, Krzakala, Zdeborova [2015,2017]
  Dia, Macris, Krzakala, Lesieur, Zdeborova [2016]
  Montanari, Reichman, Zeitouni [2015]
  Jagannath, Lopato, Milane [2018]
  G, Jagannath, Sen [2019]
- State of the art understanding

Background

- Problem studied widely in recent years:
  Lelarge, Miolane [2009]
  Butucea & Ingster [2013],
  Deshphande & Montanari [2014]
  Lesieur, Krzakala, Zdeborova [2015,2017]
  Dia, Macris, Krzakala, Lesieur, Zdeborova [2016]
  Montanari, Reichman, Zeitouni [2015]
  Jagannath, Lopato, Milane [2018]
  G, Jagannath, Sen [2019]
- State of the art understanding
  - MLE (computed by **brute force**) recovers θ^* reliably if and only if $\lambda \geq \lambda_{\text{INFO}}$, G, Jagannath, Sen [2019]

Background

- Problem studied widely in recent years:
  Lelarge, Miolane [2009]
  Butucea & Ingster [2013],
  Deshphande & Montanari [2014]
  Lesieur, Krzakala, Zdeborova [2015,2017]
  Dia, Macris, Krzakala, Lesieur, Zdeborova [2016]
  Montanari, Reichman, Zeitouni [2015]
  Jagannath, Lopato, Milane [2018]
  G, Jagannath, Sen [2019]
- State of the art understanding
  - MLE (computed by **brute force**) recovers θ^* reliably if and only if $\lambda \geq \lambda_{\text{INFO}}$, G, Jagannath, Sen [2019]
  - Fast (**poly-time**) algorithm recovers θ^* reliably when $\lambda > \lambda_{\text{COMP}}$ Deshphande & Montanari [2014]

Background

- Problem studied widely in recent years:
  Lelarge, Miolane [2009]
  Butucea & Ingster [2013],
  Deshphande & Montanari [2014]
  Lesieur, Krzakala, Zdeborova [2015,2017]
  Dia, Macris, Krzakala, Lesieur, Zdeborova [2016]
  Montanari, Reichman, Zeitouni [2015]
  Jagannath, Lopato, Milane [2018]
  G, Jagannath, Sen [2019]
- State of the art understanding
  - MLE (computed by **brute force**) recovers θ^* reliably if and only if $\lambda \geq \lambda_{\text{INFO}}$, G, Jagannath, Sen [2019]
  - Fast (**poly-time**) algorithm recovers θ^* reliably when $\lambda > \lambda_{\text{COMP}}$ Deshphande & Montanari [2014]
  - The regime $\lambda_{\text{INFO}} < \lambda < \lambda_{\text{COMP}}$ is not understood. Does a fast algorithm exist? **Info/theory vs Computation** gap

Example III: Sparse High-Dimensional Linear Regression

Example III: Sparse High-Dimensional Linear Regression

- Model $Y = X\beta^* + W$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \vdots \\ \beta_p^* \end{bmatrix} + \begin{bmatrix} W_1 \\ \vdots \\ W_n \end{bmatrix}$$

Example III: Sparse High-Dimensional Linear Regression

- Model $Y = X\beta^* + W$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \vdots \\ \beta_p^* \end{bmatrix} + \begin{bmatrix} W_1 \\ \vdots \\ W_n \end{bmatrix}$$

- **Goal:** Recover β^* from observed X and Y.

Example III: Sparse High-Dimensional Linear Regression

- Model $Y = X\beta^* + W$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \vdots \\ \beta_p^* \end{bmatrix} + \begin{bmatrix} W_1 \\ \vdots \\ W_n \end{bmatrix}$$

- **Goal:** Recover β^* from observed X and Y.
- **Sparsity:** β^* is sparse:

$$\|\beta^*\|_0 \leq k.$$

Background

Background

- MLE: solve (hard) quadratic minimization problem

$$\min_{\beta} \|Y - X\beta\|_2$$

$$\text{Subject to: } \|\beta\|_0 = k.$$

Background

- MLE: solve (hard) quadratic minimization problem

$$\min_{\beta} \|Y - X\beta\|_2$$
$$\text{Subject to: } \|\beta\|_0 = k.$$

- **Questions:**
  (a) Is the optimal solution β_{OPT} a good approximation of the ground truth β^*?
  (b) How to solve this problem fast (poly-time)?

Background

Background

- Spectacular success of convex optimization based techniques, Donoho, Tanner, Wainwright, Hastie, Tibshirani, Candes, Tao:

Background

- Spectacular success of convex optimization based techniques, Donoho, Tanner, Wainwright, Hastie, Tibshirani, Candes, Tao:
- Replace $\|\beta\|_0 = k$ with a convex constraint $\|\beta\|_1 = k$ and solve the convex minimization problem.

Background

- Spectacular success of convex optimization based techniques, Donoho, Tanner, Wainwright, Hastie, Tibshirani, Candes, Tao:
- Replace $\|\beta\|_0 = k$ with a convex constraint $\|\beta\|_1 = k$ and solve the convex minimization problem.
- Suppose entries of X and Y are i.i.d. Gaussian.

Background

- Spectacular success of convex optimization based techniques, Donoho, Tanner, Wainwright, Hastie, Tibshirani, Candes, Tao:
- Replace $\|\beta\|_0 = k$ with a convex constraint $\|\beta\|_1 = k$ and solve the convex minimization problem.
- Suppose entries of X and Y are i.i.d. Gaussian.
- The method recovers the ground truth: $\beta_{\mathrm{OPT}} = \beta^*$.

Background

- Spectacular success of convex optimization based techniques, Donoho, Tanner, Wainwright, Hastie, Tibshirani, Candes, Tao:
- Replace $\|\beta\|_0 = k$ with a convex constraint $\|\beta\|_1 = k$ and solve the convex minimization problem.
- Suppose entries of X and Y are i.i.d. Gaussian.
- The method recovers the ground truth: $\beta_{\text{OPT}} = \beta^*$.
- Works when

$$n > 2k \log p \triangleq n_{\text{Convex}},$$

(assuming $k \ll n \ll p$).

Background

- Spectacular success of convex optimization based techniques, Donoho, Tanner, Wainwright, Hastie, Tibshirani, Candes, Tao:
- Replace $\|\beta\|_0 = k$ with a convex constraint $\|\beta\|_1 = k$ and solve the convex minimization problem.
- Suppose entries of X and Y are i.i.d. Gaussian.
- The method recovers the ground truth: $\beta_{\mathrm{OPT}} = \beta^*$.
- Works when

$$n > 2k \log p \triangleq n_{\mathrm{Convex}},$$

  (assuming $k \ll n \ll p$).
- On the other hand (brute force) MLE solves the problem when

$$n > \frac{2k \log p}{\log\left(1 + \frac{k}{\sigma^2}\right)} = \frac{n_{\mathrm{Convex}}}{\log(1 + k/\sigma^2)} \triangleq n_{\mathrm{INFO}}$$

Challenging Regime

Challenging Regime

- **No algorithms** are known in the regime (except noiseless case)

$$n_{\text{INFO}} \triangleq \frac{n_{\text{Convex}}}{\log\left(1 + \frac{k}{\sigma^2}\right)} < n < n_{\text{Convex}}$$

Challenging Regime

- **No algorithms** are known in the regime (except noiseless case)

$$n_{\mathrm{INFO}} \triangleq \frac{n_{\mathrm{Convex}}}{\log\left(1 + \frac{k}{\sigma^2}\right)} < n < n_{\mathrm{Convex}}$$

- **Info/theory vs Computation** gap.

Who is he culprit? Topological obstruction to algorithms

Who is he culprit? Topological obstruction to algorithms

- Many other examples: Random K-Sat, MaxCut on random graphs, Stochastic Block Model, Planted Clique, Spiked Tensor problem, Sparse Covariance Estimation problem, Locally Computable One Way Functions, etc, etc.

Who is he culprit? Topological obstruction to algorithms

- Many other examples: Random K-Sat, MaxCut on random graphs, Stochastic Block Model, Planted Clique, Spiked Tensor problem, Sparse Covariance Estimation problem, Locally Computable One Way Functions, etc, etc.
- Many of these problems appear to exhibit a complex **solution space topology** in the hard regime and do not exhibit it in the tractable regime

Who is he culprit? Topological obstruction to algorithms

- Many other examples: Random K-Sat, MaxCut on random graphs, Stochastic Block Model, Planted Clique, Spiked Tensor problem, Sparse Covariance Estimation problem, Locally Computable One Way Functions, etc, etc.
- Many of these problems appear to exhibit a complex **solution space topology** in the hard regime and do not exhibit it in the tractable regime
- Topological obstruction come in the form of **Overlap Gap Property (OGP)**

Who is he culprit? Topological obstruction to algorithms

- Many other examples: Random K-Sat, MaxCut on random graphs, Stochastic Block Model, Planted Clique, Spiked Tensor problem, Sparse Covariance Estimation problem, Locally Computable One Way Functions, etc, etc.
- Many of these problems appear to exhibit a complex **solution space topology** in the hard regime and do not exhibit it in the tractable regime
- Topological obstruction come in the form of **Overlap Gap Property (OGP)**
- Discovered in the field of statistical mechanics (spin glass theory)

Overlap Gap Property

Overlap Gap Property

Generic minimization problem with random input $\mathcal{X}$

$$\min_{\theta \in \Theta} \mathcal{L}(\theta, \mathcal{X}).$$

Overlap Gap Property

Generic minimization problem with random input $\mathcal{X}$

$$\min_{\theta \in \Theta} \mathcal{L}(\theta, \mathcal{X}).$$

OGP holds if there exists $R > 0$, such that the set

$$\{\theta : \mathcal{L}(\theta, \mathcal{X}) \leq \min_{\theta \in \Theta} \mathcal{L}(\theta, \mathcal{X}) + R\}$$

is disconnected.

Overlap Gap Property

Generic minimization problem with random input $\mathcal{X}$

$$\min_{\theta \in \Theta} \mathcal{L}(\theta, \mathcal{X}).$$

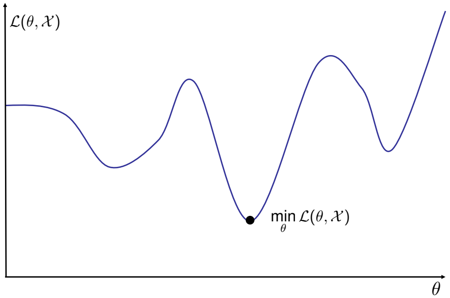
OGP holds if there exists $R > 0$, such that the set

$$\{\theta : \mathcal{L}(\theta, \mathcal{X}) \leq \min_{\theta \in \Theta} \mathcal{L}(\theta, \mathcal{X}) + R\}$$

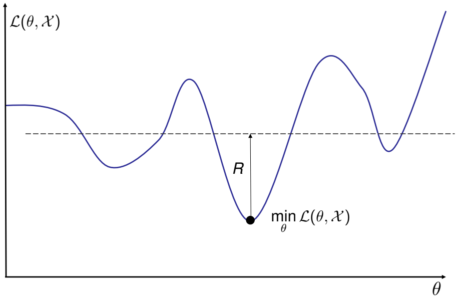
is disconnected.

That is the set of R-optimal solutions is partitioned into separate connected components.

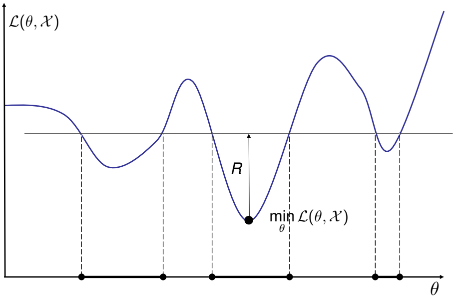
Overlap Gap Property



Overlap Gap Property



Overlap Gap Property



OGP in models with planted signal

OGP in models with planted signal

Ground truth signal θ^* is observed via a noisy observation $\mathcal{X} \stackrel{d}{=} \mathcal{X}|\theta^*$.

OGP in models with planted signal

Ground truth signal θ^* is observed via a noisy observation $\mathcal{X} \stackrel{d}{=} \mathcal{X}|\theta^*$.
Inferring signal θ^* involves solving a minimization problem

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \mathcal{L}(\theta, \theta^*, \mathcal{X}).$$

with the hope that $\hat{\theta}$ is close to θ^*.

OGP in models with planted signal

Ground truth signal θ^* is observed via a noisy observation $\mathcal{X} \stackrel{d}{=} \mathcal{X}|\theta^*$.

Inferring signal θ^* involves solving a minimization problem

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \mathcal{L}(\theta, \theta^*, \mathcal{X}).$$

with the hope that $\hat{\theta}$ is close to θ^*.

When is it "hard" to find signal θ^*?

OGP in models with planted signal

Ground truth signal θ^* is observed via a noisy observation $\mathcal{X} \overset{d}{=} \mathcal{X}|\theta^*$.
Inferring signal θ^* involves solving a minimization problem

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \mathcal{L}(\theta, \theta^*, \mathcal{X}).$$

with the hope that $\hat{\theta}$ is close to θ^*.
When is it "hard" to find signal θ^*?
For every $\zeta \in [0, 1]$ consider

$$\Gamma(\zeta) : \min_{\theta : \|\theta - \theta^*\| = \zeta} \mathcal{L}(\theta, \theta^*, \mathcal{X}).$$

OGP in models with planted signal

Ground truth signal θ^* is observed via a noisy observation $\mathcal{X} \stackrel{d}{=} \mathcal{X}|\theta^*$.
Inferring signal θ^* involves solving a minimization problem

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \mathcal{L}(\theta, \theta^*, \mathcal{X}).$$

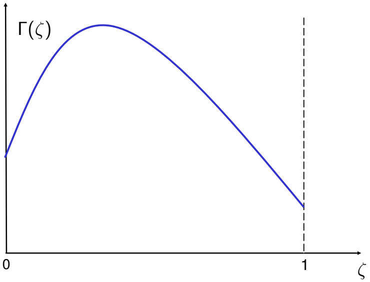
with the hope that $\hat{\theta}$ is close to θ^*.
When is it "hard" to find signal θ^*?
For every $\zeta \in [0, 1]$ consider

$$\Gamma(\zeta) : \min_{\theta : ||\theta - \theta^*|| = \zeta} \mathcal{L}(\theta, \theta^*, \mathcal{X}).$$

Plot Γ.

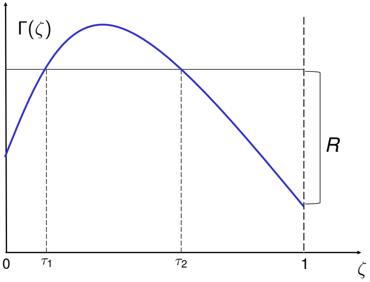
OGP in models with planted signal

Suppose Γ is not monotone.



OGP in models with planted signal

Non-monotonicity leads to OGP: every *R*-optimal solution is either τ_1-close or τ_2-far from the ground signal θ^*.



Overlap Gap Property

Overlap Gap Property

OGP is a provable obstacle to

Overlap Gap Property

OGP is a provable obstacle to

- local algorithms in sparse random graphs G & Sudan [2014,2017], Rahman & Virag [2014], Coja-Oghlan, Haqshenas & Hetterich [2016],

Overlap Gap Property

OGP is a provable obstacle to

- local algorithms in sparse random graphs G & Sudan [2014,2017], Rahman & Virag [2014], Coja-Oghlan, Haqshenas & Hetterich [2016],
- Markov Chain Monte Carlo methods G, Jagannath & Sen [2019], G & Zadik [2019]

Overlap Gap Property

OGP is a provable obstacle to

- local algorithms in sparse random graphs G & Sudan [2014,2017], Rahman & Virag [2014], Coja-Oghlan, Haqshenas & Hetterich [2016],
- Markov Chain Monte Carlo methods G, Jagannath & Sen [2019], G & Zadik [2019]
- Approximate Message Passing type algorithm for finding ground states in p-spin models G & Jagannath [2019]

OGP for the Largest Submatrix Problem

OGP for the Largest Submatrix Problem

Fix $\alpha \in (0, 1)$ and two $k \times k$ submatrices C_1, C_2 with

$$\text{Ave}(C_1) \approx \text{Ave}(C_2) \approx \alpha \text{OPT}.$$

OGP for the Largest Submatrix Problem

Fix $\alpha \in (0,1)$ and two $k \times k$ submatrices C_1, C_2 with

$$\text{Ave}(C_1) \approx \text{Ave}(C_2) \approx \alpha \text{OPT}.$$

Theorem (G & Li [2016])

For each $0 < y_1, y_2 < 1$, the expected number of such pairs C_1, C_2 with $y_1 k$ common rows and $y_2 k$ common columns is

$$\exp\left(f\left(\alpha, y_1, y_2\right) k \log n\right),$$

where

$$f(\alpha, y_1, y_2) = 4 - y_1 - y_2 - \frac{1}{1 + y_1 y_2}\alpha^2.$$

OGP for the Largest Submatrix Problem

Fix $\alpha \in (0, 1)$ and two $k \times k$ submatrices C_1, C_2 with

$$\text{Ave}(C_1) \approx \text{Ave}(C_2) \approx \alpha \text{OPT}.$$

Theorem (G & Li [2016])

For each $0 < y_1, y_2 < 1$, the expected number of such pairs C_1, C_2 with $y_1 k$ common rows and $y_2 k$ common columns is

$$\exp\left(f\left(\alpha, y_1, y_2\right) k \log n\right),$$

where

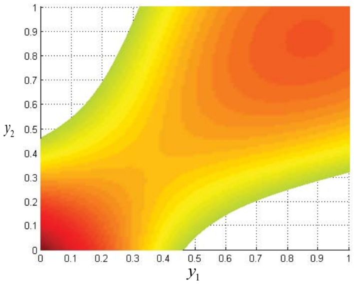
$$f(\alpha, y_1, y_2) = 4 - y_1 - y_2 - \frac{1}{1 + y_1 y_2}\alpha^2.$$

$f(\alpha, y_1, y_2) < 0$ implies no such pairs whp.

No OGP when $0 < \alpha < 0.9625...$

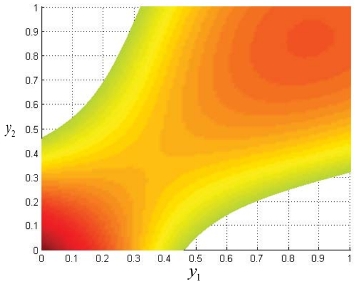
No OGP when $0 < \alpha < 0.9625...$

- Color points correspond to positive value of f.



No OGP when $0 < \alpha < 0.9625...$

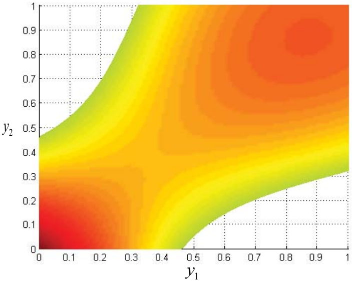
- Color points correspond to positive value of f.



- No OGP when $0 < \alpha < 0.9625...$.

No OGP when $0 < \alpha < 0.9625...$

- Color points correspond to positive value of f.

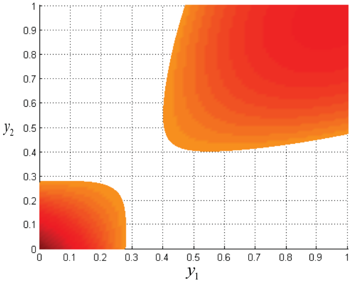


- No OGP when $0 < \alpha < 0.9625...$.
- Includes algorithmically achievable value 0.9425... .

OGP when $0.9625... < \alpha < 1$

OGP when 0.9625... $< \alpha < 1$

The set of overlaps exhibits a gap.



OGP in High-Dimensional Sparse Regression Problem

OGP in High-Dimensional Sparse Regression Problem

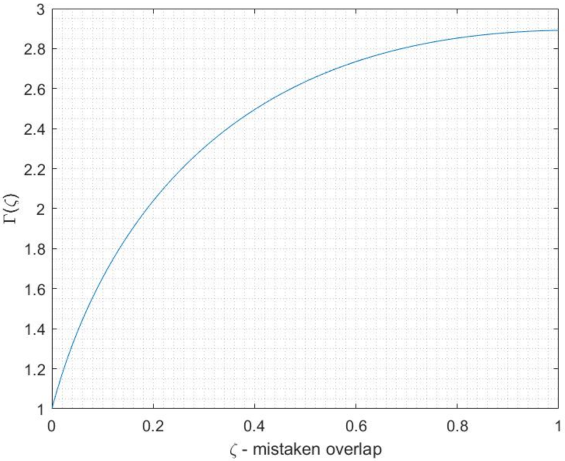
- Consider the optimization problem parametrized by $\zeta \in (0, 1)$

$$\Gamma(\zeta) \triangleq \min_{\beta} \|Y - X\beta\|_2$$

$$\text{Subject to: } \|\beta\|_0 = k, \quad \|\beta^* - \beta\|_0 = 2k\zeta.$$

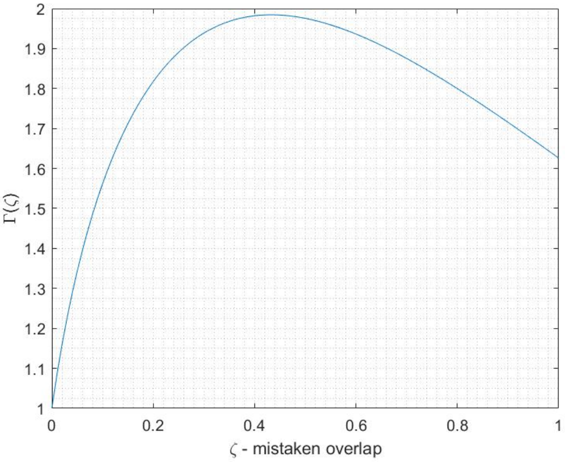
Plot of Γ. Monotonicity

When $n > n_{\text{Convex}}$



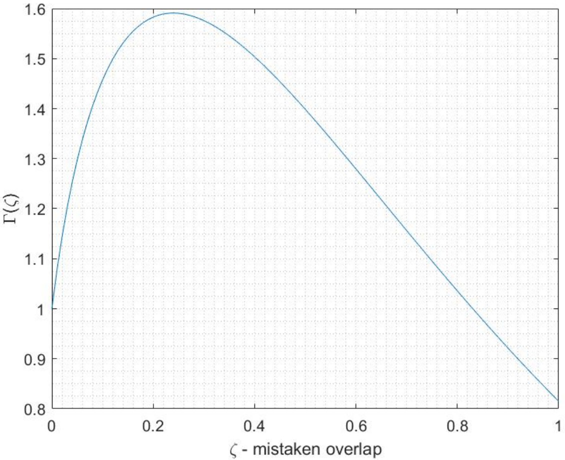
Plot of Γ. Non-monotonicity and OGP

When $n_{\text{INFO}} < n < n_{\text{Convex}}$



Plot of Γ. Non-monotonicity and OGP

When $n < n_{\text{INFO}}$



OGP Theorem

OGP Theorem

Theorem (G & Zadik [2017])

OGP Theorem

Theorem (G & Zadik [2017])

- *The OGP provably takes place when* $n \leq cn_{\mathrm{Convex}}$ *for sufficiently small constant* $c > 0$.

OGP Theorem

Theorem (G & Zadik [2017])

- *The OGP provably takes place when $n \leq cn_{\text{Convex}}$ for sufficiently small constant $c > 0$.*
- *As a consequence, (a variant of) Gradient Descent algorithm fails to find the ground truth regression vector β^*. Conjecturally MCMC fails as well.*

OGP Theorem

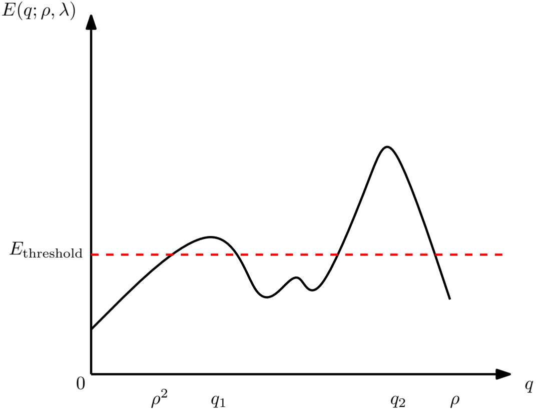
Theorem (G & Zadik [2017])

- *The OGP provably takes place when* $n \leq cn_{\text{Convex}}$ *for sufficiently small constant* $c > 0$.
- *As a consequence, (a variant of) Gradient Descent algorithm fails to find the ground truth regression vector* β^*. *Conjecturally MCMC fails as well.*
- *On the other hand, when* $n \geq cn_{\text{Convex}}$ *and* c *is sufficiently large,*
  - *No local minimum exist except* β^*.
  - *The algorithm based on local improvement finds the ground truth vector fast.*
  - *The model does not exhibit the OGP.*

The OGP in Spiked Matrix Detection

The OGP in Spiked Matrix Detection

For a certain range of values $\lambda_{\text{INFO}} < \lambda < \lambda_{\text{COMP}}$ the model exhibits the OGP



G, Jagannath, Sen [2019]

Conclusions

Conclusions

Lessons learned:

Conclusions

Lessons learned:

- Statistical physics offers a ”new theory” of complexity of inference problems arising in high-dimensional statistical models

Conclusions

Lessons learned:

- Statistical physics offers a ”new theory” of complexity of inference problems arising in high-dimensional statistical models
- Challenge: the link between this new complexity theory and the classical complexity paradigms such as NP-hardness needs to be understood.

Conclusions

Lessons learned:

- Statistical physics offers a ”new theory” of complexity of inference problems arising in high-dimensional statistical models
- Challenge: the link between this new complexity theory and the classical complexity paradigms such as NP-hardness needs to be understood.
- **Conjecture**: optimization problems with random input are hard when they exhibit OGP.

Thank you