AICR · OPEN PROBLEM

Provable Recovery from Sparse-Plus-Dense Quadratic Systems

The blockwise-random Lin–Matt candidate

Status. Statement: AI-written from Boaz Barak, Samuel B. Hopkins, Aayush Jain, Pravesh Kothari, Amit Sahai, *Sum-of-Squares Meets Program Obfuscation, Revisited*, IACR Cryptology ePrint Archive 2018, not yet checked by a human. Open as a theorem: the paper proves exact polynomial-time recovery when the polynomials are drawn i.i.d. from a *nice* distribution (its Theorem 2), which does not include the blockwise-random distribution below, and reports experiments in which a modified semidefinite program recovers the planted input once $m > 3 \cdot$(number of variables) at $n + n' = 60$ and $n + n' = 120$ variables; no proof for this family is given anywhere in the paper.

Category. *Public-Key Cryptography & Assumptions.*

Conjecture (informal). *The sum-of-squares attack in this paper is proved to work when the published quadratic polynomials are drawn independently from a distribution whose coefficients are normalised and pairwise independent, which covers both random dense and random sparse polynomials. It does not cover one specific published candidate, whose polynomials are a sum of two independent pieces on disjoint variable blocks: a sparse piece supported on a random perfect matching of the first block, and a fully dense random quadratic on the second block. The input is drawn in a way that matches this split, with the first block a half-and-half mixture of a small range and a much larger range. The authors report experiments in which a modified semidefinite program, given one extra constraint that fixes the squared length of the part of the solution belonging to the sparse block, recovers the planted input whenever the number of polynomials exceeds about three times the number of variables. They state that their theoretical analysis does not reach this case but that they believe it can, and conjecture that efficient recovery is provably possible here too. The problem is to turn the experiment into a theorem: give a polynomial-time algorithm, with a proof, that recovers the planted input from this specific structured family at stretch above linear.*

1 The setting

The main theorem of this paper (its Theorem 2) shows that when $q_1, \dots, q_m$ are drawn independently and identically from a *nice* distribution Q — one supported on homogeneous degree-two polynomials whose coefficient vector has ℓ_2 norm within a constant factor of typical, whose coefficient of $x_i x_j$ has variance one, and whose coefficients are pairwise independent — a polynomial-time algorithm recovers a polynomially bounded random input x exactly from $m \geq n(\log n)^{O(1)}$ observations $q_i(x)$. The proof writes $q_i(x) = \langle Q_i, xx^\top \rangle$, so that recovery becomes recovery of a rank-one matrix from $m \ll n^2$ linear measurements, and then verifies the two conditions in Gross's incoherence criterion [Gross] for nuclear-norm minimisation: an approximate-isotropy condition on $\frac{1}{N}\sum_a \mathrm{vec}(Q^{(a)})\mathrm{vec}(Q^{(a)})^\top$, proved by a Bernstein bound, and a small-inner-product condition $|x^\top (Q^{(a)}/n) x| \leq \nu \|xx^\top\|_F / n$, proved by Hanson–Wright. A key step is that in the cryptographic setting

one needs incoherence only for *most* x, not for all x. Niceness covers random dense polynomials and random sparse polynomials, and therefore breaks the candidate generators of Ananth, Jain and Sahai [AJS] and Agrawal [Agr18].

It does not cover the candidate of Lin and Matt [LM], which is recalled in Section 4.2 of the paper and reproduced in Definition 1 and Definition 2 below. There each polynomial is $S_j(x) + MQ_j(x')$: a sparse part supported on a uniformly random perfect matching of the first block of variables, plus an independent fully dense random quadratic on a second, disjoint block. Only one feature takes this candidate outside Definition 1 of the paper, and it concerns the polynomials alone: the coefficient variances are wildly non-uniform across the two blocks, since on the first block only $n/2$ of the $\binom{n+1}{2}$ quadratic monomials are present while on the second every coefficient is, so no single normalisation makes all variances equal to one (compare the paper's footnote 3, PDF p. 5). The input distribution is not an obstruction: the paper notes (PDF p. 6) that the proof of Theorem 2 allows x to be any real random vector with $\mathbb{E}x = 0$ that is $O(\mathbb{E}\|x\|^2/n)$-sub-Gaussian, and the half-and-half mixture on the first block — the paper's $\chi^n_{B_1,B_2}$ — is such a vector. The obstruction the paper names is exactly that each polynomial is the sum of a random dense and a random sparse polynomial: "this combination creates theoretical difficulty in the analysis".

The paper's experimental attack shows what a proof should look for. Because the planted input is polynomially bounded, one can enumerate all candidate values of $\sum_{i\in[n]} x_i^2$ and add, to the trace-norm minimisation semidefinite program, the linear constraint that the trace of the block of the matrix variable corresponding to the sparse variables equals the guessed value. With this one extra constraint, the authors report recovering the planted solution in every experiment once $m > 3(n+n')$, for $n+n' = 60$ and $n+n' = 120$ total variables. So the conjecture is not that recovery is possible in some unspecified way; it is that this concrete semidefinite program, or something like it, admits an analysis in a regime where Gross's criterion and matrix RIP [RFP, Recht] both fail to apply directly.

Settling this would upgrade an experimental break of a published iO assumption to a theorem, and would do so by extending nuclear-norm recovery guarantees to measurement ensembles that are mixtures of very different sparsity levels — a question of interest

to the matrix-recovery literature independently of the cryptographic motivation. It is also the last gap between the paper's theorems and its stated position that it knows of no degree-two candidate that survives its attacks; see also this paper's own Hypothesis 1, "No expanding weak quadratic pseudorandom generators", which conjectures that no expanding weak quadratic generator exists at all.

Progress to date. The paper's Theorem 2 handles the i.i.d.-nice case, including separately the random dense and the random sparse case; what is missing is their sum. Experimentally, adding a single trace constraint on the sparse block (obtained by enumerating the polynomially many possible values of the squared ℓ_2 norm of the sparse part of the planted solution) suffices: the authors report recovery in all their experiments once $m > 3(n + n')$, with correlation to the planted solution already about 0.95 by $m/n \approx 2$ and reaching 1 near $m/n \approx 3$, where the ratio is measured against the total number of variables $n + n'$. They also report that the same works if the sparse part S is replaced by a general random sparse polynomial with $O(n)$ monomials, and that in the three-block variant $S_{j,1}(x_1) + S_{j,2}(x_2) + MQ_j(x_3)$ constraining the sum of the two sparse-block traces recovers the planted solution using about the same number of samples.

2 Notation and parameters

- n is an even positive integer, $[n] := \{1, \ldots, n\}$, and all asymptotics are as $n \to \infty$.
- Variables are split into two blocks of n each, written $x = (x_1, \ldots, x_n)$ and $x' = (x'_1, \ldots, x'_n)$.
- An algorithm $\mathcal{A}$ is *efficient* if it is probabilistic and runs in time polynomial in the bit length of its input; polynomials are given to it by their coefficient vectors.
- A function $P : \mathbb{N} \to \mathbb{N}$ is *polynomially bounded* if $P(n) \leq n^{O(1)}$.

The parameters of the statement below are as follows.

- n: number of variables in each of the two blocks (the paper sets the two block sizes equal); the total input length is $2n$.

- m: number of published polynomials; the conjecture assumes $m \geq n^{1+\varepsilon}$.
- ε: stretch exponent.
- C: bound on the coefficients of the polynomials, which are drawn uniformly from $\{-C, \ldots, C\}$.
- B_1, B_2: the two ranges of the mixture from which each coordinate of the sparse-block input is drawn.
- B': range from which each coordinate of the dense-block input is drawn.
- B: the perturbation magnitude of the intended application; it enters only through the constraint $B_2 \geq \kappa\,(nB^2 + nBB_1)$.
- κ: the constant hidden in the paper's $B_2 = \Omega(nB^2 + nBB_1)$; the conjecture below quantifies over all constants $\kappa > 0$.

3 The conjecture

Definition 1 (blockwise-random polynomial distribution $Q_{n,C}$). Let n be an even positive integer and C a positive integer. A sample from $Q_{n,C}$ is a polynomial in the $2n$ variables (x, x') of the form

$$q(x, x') \;=\; S(x) \;+\; MQ(x'),$$

drawn as follows, with all randomness fresh for each sample.

- *(Dense block.)*

$$MQ(x') \;=\; \sum_{1 \leq i \leq k \leq n} c_{i,k}\, x'_i x'_k \;+\; \sum_{i=1}^{n} c_i\, x'_i,$$

  where every $c_{i,k}$ and every c_i is drawn independently and uniformly from $\{-C, \ldots, C\}$.

- *(Sparse block.)* A permutation σ of $[n]$ is drawn uniformly at

random and

$$S(x) \;=\; \sum_{i=1}^{n/2} \alpha_i \, x_{\sigma(2i)} \, x_{\sigma(2i-1)} + \sum_{i=1}^{n} \beta_i \, x_i \;+\; \gamma,$$

where $\alpha_1,\dots,\alpha_{n/2},\beta_1,\dots,\beta_n,\gamma$ are drawn independently and uniformly from $\{-C,\dots,C\}$.

Thus each sample is quadratic in $2n$ variables, with at most $n/2$ quadratic monomials on the first block, supported on a uniformly random perfect matching, and all $\binom{n+1}{2}$ of them on the second.

Definition 2 (blockwise input distribution $\mathcal{X}_{n,B_1,B_2,B'}$). Let B_1, B_2, B' be positive integers. A sample $(x,x') \in \mathbb{Z}^n \times \mathbb{Z}^n$ is drawn as follows. Each x_i, $i \in [n]$, is drawn independently: with probability $1/2$ uniformly from $\{-B_1,\dots,B_1\}$, and with probability $1/2$ uniformly from $\{-B_2,\dots,B_2\}$. Each x'_i, $i \in [n]$, is drawn independently and uniformly from $\{-B',\dots,B'\}$.

Conjecture 1 (provable recovery for the blockwise-random candidate). *There is an efficient algorithm $\mathcal{A}$ such that the following holds for every constant $\varepsilon > 0$, every constant $\kappa > 0$ and every polynomially bounded $P : \mathbb{N} \to \mathbb{N}$.*
For all sufficiently large even n, every integer $m \geq n^{1+\varepsilon}$, and all positive integers $C, B, B_1, B', B_2 \leq P(n)$ satisfying $B_2 \geq \kappa\,(nB^2 + nBB_1)$: if $q_1,\dots,q_m$ are drawn independently from $\mathcal{Q}_{n,C}$ and (x,x') is drawn from $\mathcal{X}_{n,B_1,B_2,B'}$, independently of the q_j, then with probability $1-o(1)$ over these draws and the coins of $\mathcal{A}$,

$$\mathcal{A}\big(q_1,\dots,q_m,\; q_1(x,x'),\dots,q_m(x,x')\big) \;=\; (x,x').$$

Remark 1 (readings fixed by this write-up). Four choices above are the page author's rather than the paper's. First, Section 4.2

of the paper writes the sparse-part coefficients as $\alpha_i, \beta_i, \gamma$ with no index j, so read literally the same coefficients are shared across all m polynomials and only the matching σ_j varies; Definition 1 takes them to be fresh per polynomial, which is what the paper's own Julia code in Appendix A does (`genmatrixsmq` redraws every coefficient for each of the m samples). A reviewer should confirm this reading against the original Lin–Matt paper, since the shared-coefficient reading gives a genuinely different and much more structured problem. Second, the parameters B_1, B' and C are "set arbitrarily" in the paper, and B — described elsewhere as the magnitude of the allowed perturbations — enters only through $B_2 = \Omega(nB^2 + nBB_1)$, so the constraint effectively just forces B_2 to dominate nB_1; the conjecture quantifies over all settings bounded by a fixed polynomial in n, and over the constant κ hidden in that Ω. The polynomial boundedness is an explicit addition here, though it is forced by the paper's framing, the planted solution having to consist of polynomially bounded integers for the enumeration step to be efficient. Third, the paper says only that "it is possible to recover in this case as well", without naming a success probability or a sample threshold; exact recovery with probability $1 - o(1)$ is stated here by analogy with Theorem 2, which achieves $1 - n^{-\log n}$ from $m \geq n(\log n)^{O(1)}$, and the candidate's own stretch $m \geq n^{1+\varepsilon}$ is kept rather than the theorem's $n(\log n)^{O(1)}$. A reviewer may reasonably prefer the weaker target of distinguishing, or approximate recovery, which is what the experiments actually measure. Fourth, it has not been checked whether the Lin–Matt candidate was subsequently withdrawn or amended, which would not close this question but would change how much anyone cares.

Remark 2 (why the statement is clean). The statement is a single recovery guarantee for one explicitly specified pair of distributions, both of which fit in a few lines and involve nothing but uniform integers, a random permutation, and a two-component mixture. There are no oracles, no idealised models, no security games and no unpublished machinery: the target is an exact-recovery theorem about a semidefinite program, and the paper's own appendix gives the working code. A prover needs no cryptography at all.

4 Bibliography

[Agr18] S. Agrawal. *New methods for indistinguishability obfuscation: bootstrapping and instantiation.* IACR Cryptology ePrint Archive, 2018:633, 2018.

[AJS] P. Ananth, A. Jain, A. Sahai. *Indistinguishability obfuscation without multilinear maps: iO from LWE, bilinear maps, and weak pseudorandomness.* IACR Cryptology ePrint Archive, 2018:615, 2018.

[BBKK] B. Barak, Z. Brakerski, I. Komargodski, P. Kothari. *Limits on low-degree pseudorandom generators (or: sum-of-squares meets program obfuscation).* Electronic Colloquium on Computational Complexity (ECCC), 24:60, 2017.

[Gross] D. Gross. *Recovering low-rank matrices from few coefficients in any basis.* IEEE Transactions on Information Theory, 57(3):1548–1566, 2011.

[LM] H. Lin, C. Matt. *Pseudo flawed-smudging generators and their application to indistinguishability obfuscation.* IACR Cryptology ePrint Archive, 2018:646, 2018.

[Recht] B. Recht. *A simpler approach to matrix completion.* Journal of Machine Learning Research, 12:3413–3430, 2011.

[RFP] B. Recht, M. Fazel, P. A. Parrilo. *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization.* SIAM Review, 52(3):471–501, 2010.