AICR · OPEN PROBLEM

No Expanding Weak Quadratic Pseudorandom Generators

Arbitrary integer quadratic maps at stretch $n^{1+\varepsilon}$

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. Settled only in the restricted case where $q_1, \ldots, q_m$ are drawn independently and identically from a *nice* distribution — supported on homogeneous (no linear term) degree-two polynomials, with $\|q\|_2^2 \le C \cdot \mathbb{E}\|q\|_2^2$ for a constant C, with $\mathrm{Var}(Q_{i,j}) = 1$ and with the coefficients $Q_{i,j}$ pairwise independent — and where the input is polynomially bounded and sub-Gaussian: there the paper's Theorem 2 gives a polynomial-time algorithm that recovers x exactly from $m \ge n(\log n)^{O(1)}$ samples, which in particular distinguishes. The hypothesis as stated, with arbitrary (possibly correlated) polynomials and an arbitrary bounded input distribution, is open.

Category. *Public-Key Cryptography & Assumptions.*

Conjecture (informal). *Several candidate constructions of indistinguishability obfuscation reduce security to a very weak kind of generator: you publish a list of quadratic polynomials with small integer coefficients, evaluate all of them at one secret input drawn from some distribution, and you only need the resulting list of values to be indistinguishable from that same list with small independent integer noise added to each entry. Nothing needs to look uniformly random, and the secret need not be recoverable even in principle. This paper conjectures that no such object exists once the number of polynomials exceeds the number of variables by any fixed polynomial factor, i.e.* $m \geq n^{1+\varepsilon}$ *for a constant* $\varepsilon > 0$*: whenever the polynomials all have polynomially bounded integer coefficients and the input distribution is also polynomially bounded, there is an efficient distinguisher for that family which achieves a constant advantage, the constant depending only on the stretch and the coefficient bound. The authors prove this in the special case where the polynomials are drawn independently from a distribution on homogeneous degree-two polynomials satisfying mild normalisation and pairwise-independence conditions, using a semidefinite program, and they report that they know of no degree-two candidate their attacks do not break. The general case, where the polynomials may be correlated with each other or chosen adversarially, is left open.*

1 The setting

Candidate constructions of indistinguishability obfuscation (iO) have converged on a common shape: bootstrap iO from a low-degree cryptographic multilinear map, LWE, and a low-degree expanding generator over the integers. Lin and Tessaro [LT] reduced iO to bilinear maps together with a block-wise two-local pseudorandom generator of superlinear stretch. Barak, Brakerski, Komargodski and Kothari [BBKK], and independently Lombardi and Vaikuntanathan [LV], then showed that the Lin–Tessaro candidate, and indeed any generator with their required parameters, can be broken using the degree-two sum-of-squares semidefinite program.

The works of Ananth, Jain and Sahai [AJS], Agrawal [Agr18] and Lin and Matt [LM] proposed a way around that obstacle by weakening the requirement in two directions at once. Structurally,

the outputs need only be *degree-two polynomials* of the input, rather than block-wise local. Security-wise, one asks only that the vector of evaluations $q_1(x), \ldots, q_m(x)$ be indistinguishable from the same vector with small independent integer perturbations added — the ΔRG property of [AJS], and the closely related *pseudo flawed-smudging* property of [LM]. Nothing is required to look uniform, and the input need not be recoverable.

This paper conjectures that the retreat does not help: at stretch $n^{1+\varepsilon}$ no such object exists, for *any* choice of quadratic polynomials with polynomially bounded integer coefficients and *any* polynomially bounded input distribution. The conjecture is stated as a hypothesis about the existence of an efficient distinguisher for each such family, so a proof is an algorithm and a refutation is a construction.

Two obstructions are recorded. First, the obvious route to a distinguisher — recover x from the evaluations — is not available in general: it need not even be information-theoretically possible to recover x, and where it is, doing so is an instance of the NP-hard problem of solving quadratic equations. Second, the paper's own theorem covers only the case where the polynomials are drawn *independently and identically* from a distribution supported on homogeneous degree-two polynomials and satisfying normalisation and pairwise-independence conditions (called *nice*); the conjecture allows arbitrary, possibly heavily correlated, polynomials.

The technical route to the theorem goes through low-rank recovery. Writing a homogeneous quadratic as $q_i(x) = \langle Q_i, xx^\top \rangle$ turns seed recovery into recovering a rank-one matrix from m linear measurements, which nuclear-norm (trace-norm) minimisation provably solves when the measurement matrices satisfy matrix RIP [RFP] or the incoherence condition of Gross [Gross]. Both conditions are randomness conditions on the measurements, and need not hold when the polynomials are adversarially chosen, as the conjecture allows.

Progress to date. Theorem 2 of the paper proves the i.i.d.-nice case: a polynomial-time sum-of-squares/trace-norm algorithm recovers x exactly with probability $1 - n^{-\log n}$ from $m \geq n(\log n)^{O(1)}$ observations, for every nice distribution Q — that is, every distribution supported on homogeneous (no linear term) degree-two polynomials with $\|q\|_2^2 \leq C \cdot \mathbb{E}\|q\|_2^2$ for a constant C, with $\mathrm{Var}(Q_{i,j}) = 1$ and with the coefficients $Q_{i,j}$ pairwise independent — and every

input uniform on $\{-t,\dots,t\}^n$ with $t \le n^{O(1)}$ (and more generally for centred $O(\mathbb{E}\|x\|^2/n)$-sub-Gaussian real inputs, whose coordinates need not be independent). This covers random dense and random sparse polynomials, and hence breaks the Ananth–Jain–Sahai and Agrawal candidates. The paper additionally reports experiments that break every degree-two candidate the authors are aware of, including the Lin–Matt one, and states that it knows of no degree-two construction that escapes a variant of the attack.

2 Notation and parameters

- $n \in \mathbb{N}$ is the number of variables; all asymptotics are as $n \to \infty$, and $[m] := \{1,\dots,m\}$.
- A *quadratic polynomial* is a polynomial $q : \mathbb{R}^n \to \mathbb{R}$ of total degree at most 2. Such a polynomial is presented to an algorithm as its vector of coefficients.
- A function $\Lambda : \mathbb{N} \to \mathbb{N}$ is *polynomially bounded* if $\Lambda(n) \le n^{O(1)}$.
- An algorithm $\mathcal{A}$ is *efficient* if it is probabilistic and runs in time polynomial in the bit length of its input.
- For distributions D_1, D_2 over a common domain, the *advantage* of an algorithm $\mathcal{A}$ is

$$\mathrm{Adv}_{\mathcal{A}}(D_1,D_2) := \Big| \Pr_{z\sim D_1}[\mathcal{A}(z)=1] - \Pr_{z\sim D_2}[\mathcal{A}(z)=1] \Big|.$$

The parameters of the statement below are as follows.

- n: number of variables of each polynomial, i.e. the length of the secret input.
- m: number of polynomials published, i.e. the output length of the generator. No upper bound is imposed; efficiency is measured in the input length, which grows with m.
- ε: stretch exponent; the conjecture assumes $m \ge n^{1+\varepsilon}$.

- $\Lambda(n)$: polynomial bound on the coefficients of the polynomials, on the support of the input distribution, and on the support of the noise.
- $\mathcal{X}$: distribution of the secret input over $\mathbb{Z}^n$.
- Δ_i: distribution of the noise added to the i-th output; required to have no atom of mass 0.9 or more.
- c: the constant distinguishing advantage the conjectured algorithm achieves.

3 The conjecture

Definition 1 (Λ-bounded polynomial). An n-variate polynomial q is Λ-*bounded* if all of q's coefficients are integers in the interval $[-\Lambda, +\Lambda]$.

Definition 2 (Λ-bounded distribution). A distribution $\mathcal{X}$ over $\mathbb{Z}^n$ is Λ-*bounded* if it is supported on $[-\Lambda, +\Lambda]^n$. A distribution Δ over $\mathbb{Z}$ is Λ-*bounded* if it is supported on $[-\Lambda, +\Lambda]$.

Conjecture 1 (no expanding weak quadratic generators). *For every constant $\varepsilon > 0$ and every polynomially bounded $\Lambda : \mathbb{N} \to \mathbb{N}$ there is a constant $c > 0$ such that the following holds for all sufficiently large $n \in \mathbb{N}$.*
Let $m \geq n^{1+\varepsilon}$ be an integer, let $q_1, \ldots, q_m : \mathbb{R}^n \to \mathbb{R}$ be quadratic $\Lambda(n)$-bounded polynomials, let $\mathcal{X}$ be a $\Lambda(n)$-bounded distribution over $\mathbb{Z}^n$, and for each $i \in [m]$ let Δ_i be a $\Lambda(n)$-bounded distribution over $\mathbb{Z}$ such that $\Pr[\Delta_i = z] < 0.9$ for every $z \in \mathbb{Z}$. Define the two distributions

$$\begin{aligned} D_1 &:= \big(q_1, \ldots, q_m,\ q_1(x), \ldots, q_m(x)\big), \\ D_2 &:= \big(q_1, \ldots, q_m, \\ &\qquad q_1(x) + \delta_1, \ldots, q_m(x) + \delta_m\big), \end{aligned}$$

where $x \sim \mathcal{X}$ in both, and where in D_2 the perturbations $\delta_i \sim \Delta_i$ are drawn independently of one another and of x. Then there exists an

efficient algorithm $\mathcal{A}$ with

$$\mathrm{Adv}_{\mathcal{A}}(D_1, D_2) \;\geq\; c.$$

Equivalently: there is no family of quadratic $\Lambda(n)$-bounded polynomials, of stretch $m \geq n^{1+\varepsilon}$, whose evaluations at a $\Lambda(n)$-bounded random input are computationally indistinguishable from those evaluations perturbed by bounded independent noise no one value of which is taken with probability 0.9 or more.

Remark 1 (readings fixed by this write-up). Two choices above are the page author's rather than the paper's. First, the paper writes "then there exists an algorithm $\mathcal{A}$ that can distinguish" without the word *efficient*; polynomial time has been made explicit here, because the very next paragraph of the paper reasons about "an efficient algorithm to recover x", and because the accompanying Theorem 2 supplies a polynomial-time algorithm. Read literally without an efficiency requirement the hypothesis is far weaker, so this is almost certainly the intended reading — but it is a reading, not a transcription. Second, the paper imposes no upper bound on m and none is imposed above; since efficiency is measured in the bit length of the input, which contains all m polynomials, the running time allowed to $\mathcal{A}$ grows with m. Note also that, following the paper's quantifier order, $\mathcal{A}$ is quantified after the instance and so may depend on m, on $q_1, \ldots, q_m$, on $\mathcal{X}$ and on the Δ_i; only the advantage c is uniform over instances.

Remark 2 (scope of the quantification — the page author's commentary, postdating the paper). The following is not drawn from the source and postdates it. The hypothesis fixes the polynomials independently of the input distribution: the same $q_1, \ldots, q_m$ appear in both distributions and are not correlated with x. Later constructions of iO which use degree-two maps evade attacks of this kind by sampling public data jointly with the secret seed, so that the effective polynomials depend on the secret [JLS]; such objects fall outside the quantification above and are not refutations of it. This reading of the later literature has not been verified, and a reviewer should confirm that no refutation has appeared.

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.

[JLS] A. Jain, H. Lin, A. Sahai. *Indistinguishability obfuscation from well-founded assumptions.* STOC, 2021 (recalled from memory; not in this paper's reference list). [UNVERIFIED]

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

[LT] H. Lin, S. Tessaro. *Indistinguishability obfuscation from bilinear maps and block-wise local PRGs.* Cryptology ePrint Archive, Report 2017/250, 2017.

[LV] A. Lombardi, V. Vaikuntanathan. *On the non-existence of blockwise 2-local PRGs with applications to indistinguishability obfuscation.* IACR Cryptology ePrint Archive, 2017:301, 2017.

[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.