Inference in High-Dimensional Linear Regression via Lattice Basis Reduction and Integer Relation Detection*

David Gamarnik† Eren C. Kızıldağ‡ Ilias Zadik§

Abstract

We consider the high-dimensional linear regression problem, where the algorithmic goal is to efficiently infer an unknown feature vector $\beta^* \in \mathbb{R}^p$ from its linear measurements, using a small number n of samples. Unlike most of the literature, we make no sparsity assumption on β^*, but instead adopt a different regularization: In the noiseless setting, we assume β^* consists of entries, which are either rational numbers with a common denominator $Q \in \mathbb{Z}^+$ (referred to as $Q-$rationality); or irrational numbers taking values in a rationally independent set of bounded cardinality, known to learner; collectively called as the mixed-range assumption. Using a novel combination of the Partial Sum of Least Squares (PSLQ) integer relation detection, and the Lenstra-Lenstra-Lovász (LLL) lattice basis reduction algorithms, we propose a polynomial-time algorithm which provably recovers a $\beta^* \in \mathbb{R}^p$ enjoying the mixed-range assumption, from its linear measurements $Y = X\beta^* \in \mathbb{R}^n$ for a large class of distributions for the random entries of X, even with one measurement ($n = 1$). In the noisy setting, we propose a polynomial-time, lattice-based algorithm, which recovers a $\beta^* \in \mathbb{R}^p$ enjoying the $Q-$rationality property, from its noisy measurements $Y = X\beta^* + W \in \mathbb{R}^n$, even from a single sample ($n = 1$). We further establish that for large Q, and normal noise, this algorithm tolerates information-theoretically optimal level of noise. We then apply these ideas to develop a polynomial-time, single-sample algorithm for the phase retrieval problem. Our methods address the single-sample ($n = 1$) regime, where the sparsity-based methods such as the Least Absolute Shrinkage and Selection Operator (LASSO) and the Basis Pursuit are known to fail. Furthermore, our results also reveal algorithmic connections between the high-dimensional linear regression problem, and the integer relation detection, randomized subset-sum, and shortest vector problems.

Keywords: High-dimensional statistical inference, linear regression, phase retrieval, lattices, integer relation detection, subset sum problem, polynomial-time algorithm, single-sample recovery, information-theoretic thresholds, channel capacity.

---

*Parts of this paper have been presented at 2018 Conference on Neural Information Processing Systems (NeurIPS) [1] and 2019 IEEE International Symposium on Information Theory (ISIT) [2].

†MIT; e-mail: gamarnik@mit.edu. Support from ONR Grant N00014-17-1-2790 is gratefully acknowledged.

‡MIT; e-mail: kizildag@mit.edu

§NYU; e-mail: zadik@nyu.edu. Research supported by a CDS Moore-Sloan Postdoctoral Fellowship.

Contents

1 Introduction **3**
- 1.1 Summary of the Results . . . 8
- 1.2 Notation, Definitions and Preliminaries . . . 10
  - 1.2.1 Notation . . . 10
  - 1.2.2 Definitions and Preliminaries . . . 11
  - 1.2.3 The LLL Algorithm . . . 13

2 Main Results **15**
- 2.1 Noisy High-Dimensional Linear Regression with $Q-$rational β^* . . . 15
  - 2.1.1 Extended Lagarias-Odlyzko algorithm . . . 15
  - 2.1.2 Applications to High-Dimensional Linear Regression . . . 17
  - 2.1.3 The Model . . . 17
  - 2.1.4 The Lattice-Based Regression (LBR) Algorithm . . . 17
  - 2.1.5 Recovery Guarantees for the LBR algorithm . . . 18
  - 2.1.6 Noise tolerance of the LBR algorithm . . . 19
  - 2.1.7 Information Theoretic Bounds . . . 20
- 2.2 Noiseless High-Dimensional Linear Regression with Irrational β^* . . . 21
  - 2.2.1 Integer-Valued Measurement Matrix X . . . 21
  - 2.2.2 Continuous-Valued Measurement Matrix X . . . 23
- 2.3 Noiseless High-Dimensional Linear Regression with Mixed β^* . . . 24
  - 2.3.1 Integer-Valued Measurement Matrix X . . . 24
  - 2.3.2 Continuous-Valued Measurement Matrix X . . . 26
- 2.4 Application: Phase Retrieval Problem . . . 28
  - 2.4.1 Discrete-Valued X . . . 28
  - 2.4.2 Continuous-Valued X . . . 30

3 Numerical Experiments **31**

4 Proofs **32**
- 4.1 Proof of Theorem 2.1 . . . 32
- 4.2 Proof of Theorem 2.4.A . . . 40
- 4.3 Proof of Theorem 2.4.B . . . 42
- 4.4 Proof of Theorem 2.8 . . . 43
- 4.5 Proof of Theorem 2.9 . . . 46
- 4.6 Proof of Theorem 2.10 . . . 47
- 4.7 Proof of Theorem 2.11 . . . 48
- 4.8 Proof of Theorem 2.12 . . . 49
- 4.9 Proof of Theorem 2.13 . . . 50
- 4.10 Proof of Proposition 2.14 . . . 51
- 4.11 Proof of Theorem 2.15 . . . 53
- 4.12 Proof of Proposition 2.5 . . . 55
- 4.13 Proof of Proposition 2.6 . . . 56
- 4.14 Proof of Proposition 2.7 . . . 57

1 Introduction

We study the following high-dimensional linear regression model. Consider n linear measurements of a vector $\beta^* \in \mathbb{R}^p$ through the measurement model, $Y = X\beta^*$ for some $X \in \mathbb{R}^{n\times p}$; or its noisy version $Y = X\beta^* + W \in \mathbb{R}^n$, where $W \in \mathbb{R}^n$ is an additive noise. Given the knowledge of Y and X the algorithmic goal is to efficiently infer β^* using a small number n of measurements. Throughout the paper we refer to p as the number of features, to X as the measurement matrix, and to n as the number of measurements, or the sample size for short.

We focus on the high-dimensional case where n may be much smaller than p and p grows to infinity, a setting that has been very popular in the literature during the last years [3, 4, 5, 6, 7]. In this case, and under no additional structural assumptions, the inference task becomes impossible (even in the absence of noise), as the underlying linear system becomes underdetermined. Most papers address this issue by imposing a *sparsity assumption* on β^*, which refers to β^* having only a limited number of non-zero entries compared to its dimension [4, 5, 6], an assumption motivated from an empirical observation that many signals (such as medical images and photos) exhibit sparsity in an appropriate transform domain. During the past decades, the sparsity assumption led to a fascinating line of research in statistics and compressed sensing communities, with a wide-range of applications, from single-pixel camera to medical imaging and radar imaging, which established, among other results, that several polynomial-time algorithms, such as the Basis Pursuit Denoising Scheme and the Least Absolute Shrinkage and Selection Operator (LASSO), can efficiently recover a sparse β^* with number of samples much smaller than the number of features [5, 7, 6]. For example, it is established that if β^* is constrained to have at most $k \leq p$ non-zero entries, X has iid $N(0,1)$ entries, W has iid $N(0,\sigma^2)$ entries and n is of the order $k\log\left(\frac{p}{k}\right)$, then both of the mentioned algorithms can recover β^*, up to the level of the noise. Structural assumptions other than the sparsity have also been considered in the literature. For example, a recent paper [8] makes the assumption that β^* lies near the range of an L-Lipschitz generative model $G : \mathbb{R}^k \to \mathbb{R}^p$ and proposes an algorithm which succeeds with $n = O(k\log L)$ samples. Other assumptions that have also been considered in the literature include tree-sparsity and block-sparsity.

A downside of all of the above results is that they provide no guarantee in the case n is much smaller than $k\log\left(\frac{p}{k}\right)$. Consider for example the case where the components of a sparse β^* are binary-valued, and X, W follow the Gaussian assumptions described above. Supposing that σ is sufficiently small, it follows from a straightforward argument that even when $n = 1$, β^* is recoverable from $Y = \langle X, \beta^*\rangle + W$ by a brute-force method with probability tending to one as p diverges to infinity. On the other hand, for sparse and binary-valued β^*, the Basis Pursuit method in the noiseless case [9] and the LASSO in the noisy case [10, 7] have been proven to fail to recover a binary β^* with $n = o(k\log\left(\frac{p}{k}\right))$ samples. This failure to capture the complexity of the problem accurately enough for small sample sizes also lead to an algorithmic hardness conjecture for the regime $n = o(k\log\left(\frac{p}{k}\right))$ [11], [10]. While this conjecture still stands in the general case, as we show in this paper, in the special case where β^* satisfies the so-called mixed-range assumption (elaborated below), and the magnitude of the noise W is sufficiently small or zero, there is no statistical/computational gap; and β^* can be recovered even when $n = 1$.

In this paper, we make no sparsity assumption on the entries of β^*. Instead, we achieve the regularization based on the assumption that the entries of β^* take values in a set $\mathcal{S}$ consisting of irrational and rational elements that are assumed to satisfy certain structural properties,

which we call the *mixed-range* assumption. Specifically, we assume that the rational entries of β^* have denominator equal to some fixed positive integer $Q \in \mathbb{Z}^+$, something we refer to as the *Q-rationality* assumption. For the irrational entries of β^*, we assume that they take values in a set $\{a_1, \ldots, a_{\mathcal{R}}\}$ known to the learner, and the set enjoys the so-called rational independence property: the only rational combination of a_i, $1 \leq i \leq \mathcal{R}$, adding up to 0 is the trivial one, i.e., if for $q_i \in \mathbb{Q}$, $\sum_{i=1}^{\mathcal{R}} q_i a_i = 0$ holds, then $q_i = 0$, for $1 \leq i \leq \mathcal{R}$. We note that while the linearity holds true between Y and the **unconstrained** β^*, the model that we work with is not quite linear due to the imposed assumption. In the sequel, we skip this issue and will still refer to the model as "linear regression".

The assumption that β^* consists of both mixed real and rational-valued entries is partially motivated by applications in signal processing. The most notable example of this comes from the study of global positioning/global navigation satellite systems (GPS/GNSS). To the best of our knowledge, Teunissen was the first to use mixed linear integer model for application to GPS in a series of papers [12, 13, 14, 15]. In particular, he proposed a mixed real/integer model of form $Y = Ax + Bz + W$, with $A \in \mathbb{R}^{n\times p}$, $B \in \mathbb{R}^{n\times q}$, $x \in \mathbb{R}^p$, $z \in \mathbb{Z}^q$, and $W \in \mathbb{R}^n$ being a noise/error term. Here, the vector $z \in \mathbb{Z}^q$ here corresponds to the integer multiples of a certain wavelength. His approach employs certain ideas from an earlier work by Lenstra [16] (who is also a co-inventor of the seminal LLL algorithm [17] that we employ in the present paper), includes in particular the idea of *basis reduction*, and works well in practice [18].

Recasting the system above, $Y = Ax + Bz + W$, as

$$Y = X\beta^* + W \quad \text{with} \quad X = \begin{bmatrix} A & B \end{bmatrix} \in \mathbb{R}^{n\times(p+q)}, \quad \beta^* = \begin{bmatrix} x^T & z^T \end{bmatrix}^T \in \mathbb{R}^{p+q},$$

we observe that β^* consists of both real and rational entries, with rational entries enjoying $1-$rationality assumption, though admittedly the irrational values in this model do not necessarily enjoy the rational independence assumption. Our investigation of the models admitting the rational independence assumption is primarily of theoretical interest.

Another example of application of the mixed real/rational linear model is in the area of geodesy studies, where the goal is carrier-phase based precise positioning: carrier-phase measurements have an unknown integer part (called ambiguity), that is to be estimated to achieve cm-level accuracy on the position of the receiver (which is a real-valued unknown), see [19]. Admittedly though, the irrational entries of this model also do not necessarily satisfy the rational independence we adopt in this paper.

Having mentioned these applications, in particular those from the GPS/GNSS literature; it is worth noting that the idea of *basis reduction* that we employ herein has, in fact, been also considered previously in those contexts. We now elaborate on this point, beginning with the so-called *integer least squares* problem. Consider the GPS signal model above, where we assume, for simplicity, $Y = Bz + W$ where $z \in \mathbb{Z}^q$ is an unknown vector, $B \in \mathbb{R}^{n\times q}$ is some (known) measurement matrix, and W is a noise vector with i.i.d. standard normal coordinates. The goal is to "learn" z using the observation Y and the design matrix B. Clearly, for $W \sim \mathcal{N}(0, I_n)$, the maximum likelihood (ML) estimate for z is obtained by solving the following *integer least squares* problem:

$$\min_{z\in\mathbb{Z}^q} \|Y - Bz\|_2.$$

Solving this problem is crucial also for the case where the unknown vector consists of mixed integer/rational entries, see [18]. Solving this problem, however, is computationally challenging;

in the sense that it is NP-hard both in the worst-case as well as in average-case [20]. Despite this somewhat pessimistic picture; researchers nevertheless have devised algorithms for solving this problem which appear to work well in practice (as was noted already). These algorithms consist mainly of two steps: reduction and search, see e.g. [21, 22, 23]. The goal of the reduction step is to "reduce" the (columns of the) matrix B in a certain sense (without affecting the lattice structure) which will eventually be more amenable for the search step. One of the best ways to ensure B is *reduced* is to apply the LLL basis reduction algorithm [17], which is also central to the oracles we consider below. Furthermore, albeit somewhat indirectly, the LLL algorithms appears to be useful also for the search step. It is noted in [22, 23] (see also [21]) that a common search strategy is to utilize a technique developed by Fincke and Pohst [24]. An inspection of [24, Step 2 in (2.12)] reveals that their algorithm uses crucially also a *basis reduction* step, and they note that this step can be done successfully using, e.g., the LLL algorithm.

In the aforementioned body on GPS/GNSS literature, another line of research (especially in the field of Geodesy) proposed other mixed integer regression estimators. Among others, two important such estimators are the so-called *integer aperture* [25] and *integer equivariant* [26] estimators. These estimators have an important step called *decorrelation*, which is somewhat analogous to the reduction step considered above. One common approach to the decorrelation step is to utilize a certain algorithm by Teunissen called LAMBDA [15]. It is worth highlighting that while the LAMBDA algorithm does not use the LLL algorithm per se; it is an end product of a series of papers by Teunissen [12, 13, 14, 15], and these papers use certain basis reduction ideas from an earlier work by Lenstra [16]. (Moreover as was already mentioned, the paper [16] by Lenstra is related also to his paper on LLL algorithm.) Strictly speaking, the LAMBDA algorithm was devised through certain techniques independent of the LLL algorithm. Nevertheless, later research actually pointed out similarities between these approaches (which is not surprising, in light of the history on the development of LAMBDA described above). In fact, it appears that the bases generated by these methods are quite close to each other. For a much more elaborate comparison and a rigorous theoretical link between these approaches, see [27] and the references therein.

In particular, the broad idea of *basis reduction* in the context of regression is indeed not new; and has already been explored. Numerical experiments accompanying corresponding papers suggest that these algorithms appear to work well in certain practical settings. However, while these algorithms are well supported by numerical experiments, they unfortunately lack the accompanying rigorous theoretical guarantees. One of the main features of our paper is to provide such rigorous theoretical guarantees.

Some of the models considered in our paper adopt a noise-free assumption ($W = 0$). Such models, though less frequently, also appear in practice; and some specific examples are discussed in the book by Foucart and Rauhut [6]. One such example is the so-called single-pixel camera. In this model a vector β corresponds to the color intensities of an image for different pixels, and the model assumes no noise. However, the corresponding regression matrix has i.i.d. $+1/-1$ Bernoulli entries, as opposed to a continuous distribution we will mostly assume.

Noiseless regression model with integer valued regression coefficients were also important in the theoretical development of compressive sensing methods. Specifically, Donoho [4] and Donoho and Tanner [28, 9, 29] consider a noiseless regression model of the form AB where A is a random (say Gaussian) matrix and B is the unit cube $[0, 1]^p$. One of the goals of these papers was to count the number of extreme points of the projected polytope AB in order to explain

the effectiveness of the linear programming based methods. The extreme points of this polytope can only appear as projections of extreme points of B which are all length$-p$ binary vectors, namely one deals with noiseless regression model with binary coefficients – an important special case of the model we consider in our paper. More recently, the model involving the Q-rationality assumption appeared in [30] in the context of studying the power of convex relaxation type methods. Specifically, the 3-point model in this paper is a special case of Q-rational models with values $0, 0.2, 1$ (Section 1.1 in [30]) and $-1, 0, 1$ (Section 1.2 in [30]).

In the Bayesian setting, where the ground truth β^* is sampled according to a discrete distribution, [31] proposes a low-complexity algorithm which provably recovers β^* with $n = o(p)$ samples. This algorithm uses the technique of approximate message passing (AMP) and is motivated by ideas from statistical physics [32]. Even though the result from [31] applies to the general discrete case for β^*, it requires the matrix X to be spatially coupled, a property that in particular does not hold for X with i.i.d. standard Gaussian entries. Furthermore the required sample size for the algorithm to work is only guaranteed to be sublinear in p, a sample size potentially much bigger than the information-theoretic limit for recovery under sufficiently small or zero noise ($n = 1$), a gap which is filled in this paper. In the present paper, where β^* satisfies the Q-rationality assumption, we propose a polynomial-time algorithm which applies for a large class of continuous distributions for the i.i.d. entries of X, including the normal distribution, and provably works even when $n = 1$.

As noted earlier, our approach focuses both on noiseless and noisy setups, where we adopt the mixed-range assumption in the noiseless case, and $Q-$rationality assumption in the noisy case. In the setting with noise we establish an information-theoretical noise optimality in case of normal noise and large Q. Our techniques, however, do not transfer to $\beta^* \in \mathbb{R}^p$ under our mixed-range assumption with noise, as structures utilized by the building blocks of our algorithms are too fragile under the noisy measurement model. We then apply our ideas to the so-called phase retrieval problem: a noiseless regression type model which has been studied extensively in the literature. Here the coefficients of the regression vector β^* and the entries of the regression matrix X are complex valued, but the observation vector $Y = X\beta^*$ is only observed through absolute values, that is, $Y_i = |\langle X_i, \beta^* \rangle|$, where X_i is the i^{th} row of the measurement matrix X, and $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. This model has many applications, including crystallography, see [33]. The aforementioned paper provides many references to phase retrieval model including the cases when the entries of β^* take values in a finite set. In this paper, we also study the phase retrieval problem, under a random measurement model X, and complex-valued β^*, which we assume to consist of complex numbers taking values in a known set $\mathcal{S}$, where a certain set $\mathcal{S}'$, obtained from $\mathcal{S}$ enjoys the rational independence property. We develop an algorithm which provably recovers $\beta^* \in \mathbb{C}^p$ whp, even with a single observation ($n = 1$), after polynomial in p and $\mathcal{R}$ many operations over real numbers, where $\mathcal{R}$ is the size of the set, $\mathcal{S}$.

The algorithms we propose are inspired by the algorithm introduced in [34] which solves, in polynomial time, a certain version of the so-called randomized Subset-Sum problem, arising in cryptography. To be more specific, consider the following algorithmic problem. Given a set of values $x_i \in \mathbb{Z}^+$, $1 \le i \le p$ and a $y = \sum_{i \in S} x_i \in \mathbb{Z}^+$ for some $S \subset [p]$, the algorithmic goal is to recover S, from x_i, $1 \le i \le p$ and y. Here, one can interpret $\{x_i\}_{i=1}^p$ as a public information and Y to be the ciphertext, where the plaintext $\mathbf{e} = (e_1, \ldots, e_p) \in \{0, 1\}^p$ is encrypted using $Y = \sum_{k=1}^n x_k e_p$. Over 30 years ago, this problem received a lot of attention in the field of cryptography, based on the belief that the problem would be hard to solve in many "real"

instances. This would imply that several already built public key cryptosystems, called knapsack public key cryptosystems, could be considered safe from attacks [35], [36]. This belief though was refuted by several papers in the early 80s, see e.g. [37]. Motivated by this line of research, Lagarias and Odlyzko [34], and a year later Frieze [38] (using a cleaner and shorter argument), proved a surprising fact: if x_i, $1 \leq i \leq p$ follow an i.i.d. uniform distribution on $[2^{cp^2}] \triangleq \{1, 2, 3, \ldots, 2^{cp^2}\}$ for some $c > 1/2$, then there exists a polynomial-in-p time algorithm which solves the subset-sum problem whp as $p \to \infty$. In other words, even though the problem is NP-hard in the worst-case, assuming a quadratic-in-p number of bits for the coordinates of x, the algorithmic complexity of a typical such problem is polynomial in p. This successful efficient algorithm is based on an elegant application of a seminal algorithm from the computational study of lattices called the Lenstra-Lenstra-Lovász (LLL) algorithm, introduced in [17]. This algorithm receives as an input a basis $\{b_1, \ldots, b_m\} \subset \mathbb{Z}^m$ of a full-dimensional lattice $\mathcal{L}$ and returns in time polynomial in m and $\max_{i=1,2,\ldots,m} \log \|b_i\|_\infty$ what is known as a 'reduced basis', from which one can, in particular, obtain a non-zero vector $\hat{z}$ in the lattice, such that $\|\hat{z}\|_2 \leq 2^{\frac{m}{2}} \|z\|_2$, for all $z \in \mathcal{L} \setminus \{0\}$, which therefore solves $2^{m/2}$-approximate shortest vector problem on $\mathcal{L}$, in a time polynomial in m and $\max_i \log \|b_i\|_\infty$. This is elaborated below in Section 1.2.3.

Besides its significance in cryptography, the results of [34] and [38] admit an interesting linear regression interpretation as well. In particular, one can show that under the assumption that x_i, $1 \leq i \leq p$, are i.i.d. (discrete) uniform in $[2^{\frac{1}{2}(1+\epsilon)p^2}]$, there exists exactly one set S^* with $y = \sum_{i \in S^*} x_i$ with high probability, as p tends to infinity, and moreover, this set can be recovered, with high probability, in polynomial time. Therefore if β^* is the indicator vector of this unique set S^*, that is $\beta^*_i = 1$ iff $i \in S$ for $1 \leq i \leq p$, we have that $y = \sum_i x_i \beta^*_i = \langle x, \beta^* \rangle$ where $x := (x_i : 1 \leq i \leq p)$. Furthermore using only the knowledge of y and x as input to the Lagarias-Odlyzko algorithm, we obtain a polynomial in p time algorithm which recovers exactly β^* whp as $p \to \infty$. Written in this form, and given our earlier discussion on high-dimensional linear regression, this statement is equivalent to the statement that the noiseless high-dimensional linear regression problem with binary β^* and X generated with i.i.d. elements from Unif$[2^{\frac{1}{2}(1+\epsilon)p^2}]$ is solvable in polynomial time even from one sample ($n = 1$), w.h.p. as p diverges to infinity. One of the main focuses of this paper is to extend this result to β^* satisfying the Q-rationality assumption, irrationality with rational independence assumption, continuous distributions on the iid entries of X, and non-zero noise levels.

In addition to its connection with the subset-sum problem, our irrationality assumption on β^* also reveals an interesting algorithmic connection between the linear regression problem and the well-known integer relation detection problem. Given a vector, $\mathbf{x} = (x_i : 1 \leq i \leq n) \in \mathbb{R}^n$, an integer relation for $\mathbf{x}$ is a vector $\mathbf{m} = (m_1, m_2, \ldots, m_n) \in \mathbb{Z}^n$ of integers such that $\sum_{k=1}^n m_k x_k = 0$, and $\mathbf{m}$ is not identically 0. Given an input $\mathbf{x}$, the goal of the integer relation detection problem is then to find an integer relation $\mathbf{m}$ for a given input $\mathbf{x} \in \mathbb{R}^n$. Note that, such a relation is not necessarily unique. Finding such a relation has been a long quest in science going back to Euclid. In fact, for the simplest case of finding a relation for a vector with only two entries x_1 and x_2, one may apply Euclidean algorithm to continued fraction expansion of x_1/x_2. The case $n \geq 3$ was attempted by Euler, Perron, Minkowski, Bernstein, Brun, and many others; however none of them was provably working for $n \geq 3$. The first successful algorithm that works for $n \geq 3$ is devised by Ferguson and Forcade [39]. In successive years, Ferguson provided an alternative variant of an algorithm in [39] in [40]. After the introduction of the seminal Lenstra-Lenstra-Lovász (LLL) lattice basis reduction algorithm [17], Hastad et al. [41] devised the HJLS

algorithm, which is based on the LLL lattice basis reduction algorithm [17], and provably recovers an integer relation for a given $\mathbf{x} \in \mathbb{R}^n$ in time polynomial in the dimension n of $\mathbf{x}$, and polynomial in $\log \|\mathbf{m}\|$, where $\mathbf{m} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}$ is a relation for $\mathbf{x}$ with the smallest $\|\mathbf{m}\|$. The HJLS algorithm is the first algorithm proven to have this property. However, it has a drawback that it is numerically unstable; and this issue has further been resolved in PSLQ integer relation algorithm, which was introduced in Ferguson and Bailey [42], whose analysis and further refinement was provided by Ferguson, Bailey and Arno [43].

PSLQ integer relation detection algorithm (where PS stands for partial-sum of squares, and LQ stands for both Least Squares as well as the LQ decomposition) has been included in the list "Top Ten Algorithms of the Century" by the January/February 2000 issue of *Computing in Science and Engineering*, which is jointly published by American Institute of Physics and the IEEE Computer Society [44]. The fundamental property of this algorithm that we will make use of in this paper is that, PSLQ algorithm provably returns a relation for a given $\mathbf{x} \in \mathbb{R}^n$, if it exists, after a number of operations that is polynomial both in the size n of the input vector $\mathbf{x}$ as well as the number of bits required to represent the smallest such relation [43]. Furthermore, it does not suffer from the numerical instability of HJLS algorithm [41], [43].

Since its introduction, PSLQ algorithm has been successfully used in many applications. For instance, it lead to the discovery of new formulas for certain transcendental numbers, such as π (the so-called Bailey-Borwein-Plouffe formula) and $\log(2)$, through which n^{th} hexadecimal or binary digit of the number can be computed in a reasonable time, without computing any of the previous $n-1$ digits [45]. Yet another application of this algorithm is to determine whether a given number α is algebraic of some degree n or less; namely to determine if there exists a polynomial of degree at most n, with integer coefficients, admitting α as one of its roots. Such a polynomial exists iff there exists an integer relation for the vector, $(1, \alpha, \alpha^2, \ldots, \alpha^n)$. This algorithm can be successfully used to recover such a polynomial, if it exists; or to identify a bound, such that, there is no such polynomial with integer coefficients, whose Euclidean norm is less than this bound [46].

The results of the present paper reveal an intriguing algorithmic connection between the high-dimensional regression problem, and the integer-relation detection, subset-sum and consequently, the γ-approximate shortest vector problems.

1.1 Summary of the Results

We now summarize our main contributions.

(a) We first study the high-dimensional linear regression problem, where the algorithmic goal is to efficiently recover an unknown feature vector $\beta^* \in \mathbb{R}^p$, from its noisy linear measurements $Y = X\beta^* + W \in \mathbb{R}^n$, under the assumptions that entries of β^* enjoy the so-called $Q-$rationality assumption, that is, entries of β^* are rational numbers with a common denominator $Q \in \mathbb{Z}^+$; $\|\beta^*\|_\infty \leq \widehat{R}$ for some $\widehat{R} > 0$, X has bounded density and finite expectation, and W is either adversarial with $\|W\|_\infty \leq \sigma$ or i.i.d. mean-zero noise with variance σ^2. Note that after multiplying everywhere by Q, one can assume to deal with a integer-valued regression vector, and in fact, this will precisely be our approach.

We propose an efficient algorithm, which, under some explicit parameter assumptions, provably recovers $\beta^* \in \mathbb{R}^p$ in time polynomial in $n, p, \sigma, \widehat{R}$, and Q. Moreover, our analysis reveals that the single-sample recovery ($n = 1$) is indeed possible, provided that the

parameters satisfy an explicitly stated condition. We complement our analysis with the information-theoretic limits of the problem, and establish for large Q and normal noise, we have near optimal noise tolerance, using results on the capacity of Gaussian channel with power constraint. A crucial step towards our main result is the extension of the Lagarias-Odlyzko algorithm [34], [38] to not necessarily binary, integer vectors $\beta^* \in \mathbb{Z}^p$, for measurement matrix $X \in \mathbb{Z}^{n\times p}$ with iid entries not necessarily from the uniform distribution, and finally, for non-zero noise vector W. As in [34] and [38], the algorithm we construct depends crucially on building an appropriate lattice and applying the LLL algorithm on it. However, unlike these prior papers where the recovered vector is a multiple $\lambda\beta^*$ of a binary vector and that, the corresponding multiplicity can be read off directly; we need an extra additional step: We translate the observations $Y = X\beta^* + Z$ by XZ, then establish, using an analytic number theory argument, that the entries of $\beta^* + Z$ has greatest common divisor equal to one, with high probability. This argument extends an elegant result from probabilistic number theory (see for example, Theorem 332 in [47]) according to which

$$\lim_{m\to\infty} \mathbb{P}_{P,P'\sim \mathrm{Unif}\{1,2,\ldots,m\}, P \perp\!\!\!\perp P'} \left[\gcd\left(P, P'\right) = 1\right] = \frac{6}{\pi^2}$$

where $P \perp\!\!\!\perp P'$ refers to P, P' being independent random variables. A key implication of this result for us is the fact that the limit above is strictly positive.

(b) Our next focus is the noiseless high-dimensional linear regression problem, where the algorithmic goal of the learner is to efficiently recover an unknown feature vector $\beta^* \in \mathbb{R}^p$, this time consisting of both rational as well as irrational entries; from its noiseless linear measurements, $Y = X\beta^* \in \mathbb{R}^n$. We study this problem for a $\beta^* \in \mathbb{R}^p$, which satisfies the so-called mixed-range assumption, an assumption that has been concretely defined as the Assumption 1 in the main body of this paper. We propose an efficient algorithm, which, under some explicitly stated parameter assumptions, provably recovers a $\beta^* \in \mathbb{R}^p$ enjoying the mixed-range assumption from its noiseless linear measurements $Y = X\beta^* \in \mathbb{R}^n$, in time polynomial in the problem parameters. More formally, we show that if the i.i.d. entries of $X \in \mathbb{R}^{n\times p}$ are drawn from a continuous distribution with bounded density and finite expected value (or from a discrete distribution taking values in a large enough set), the entries of β^* take values in a set $\mathcal{S}$, where the irrational elements $\{a_1, \ldots, a_{\mathcal{R}}\}$ of $\mathcal{S}$ are rationally independent and known to the learner; and the rational elements share a common denominator $Q \in \mathbb{Z}^+$; the proposed algorithm recovers a β^* with high probability, after a number of operations, polynomial in the number n of measurements, the dimension p of the feature vector β^*, the number $\mathcal{R}$ of the irrational elements of the set $\mathcal{S}$, and in $\log Q$, where $Q \in \mathbb{Z}^+$ is the common denominator of the rational entries of the set $\mathcal{S}$ that the entries of β^* take values in. The algorithm that we propose is obtained using a novel combination of the PSLQ integer relation detection [43], and LLL lattice basis reduction [17] algorithms. Moreover, analogous to the previous setting, our analysis reveals also that the efficient recovery is indeed possible, even when the learner has access to only one measurement ($n = 1$); provided that the parameters satisfy an explicitly stated condition. In particular, our algorithms for the high-dimensional linear regression problem address the extreme regime when the learner has access to only one measurement ($n = 1$), a regime where the sparsity-based methods, such as LASSO and Basis Pursuit are known

to fail. In particular, in the aforementioned setting, our analysis reveals that there is no *statistical-computational gap*.

(c) We then apply our ideas to the phase retrieval problem, where a complex-valued vector β^* is observed through the model, $Y = |\langle X, \beta^* \rangle|$. We study this problem for both discrete-valued and continuous-valued measurement matrix X with random entries; and complex-valued β^*, whose entries take values in a set known to the learner, with cardinality $\mathcal{R}$. We establish that, under a rational independence assumption on a certain predetermined set, generated from the set that the entries of β^* take values in, one can recover β^* after polynomial in p and $\mathcal{R}$ many arithmetic operations on real numbers, with high probability. Moreover, the proposed method transfers almost immediately, to more generalized observation models, where a complex-valued β^* is observed through $Y = f(|\langle X, \beta^* \rangle|)$, provided that f belongs to a class of functions for which, for any real number r, the roots of the equation, $f(x) = r$ can be computed, in a sense to discussed. An example of such an f is a polynomial with degree at most 4, where there exists formulas for the roots of such polynomials. Towards the goal of devising an efficient algorithm to the phase retrieval problem; we show that the LLL algorithm run on an appropriate lattice yields an efficient algorithm for the random subset-sum problem with dependent inputs, therefore, extending Lagarias-Odlyzko algorithm [34] and generalizing Frieze's result [38] in a different direction. In particular, we show that if $X_1, \ldots, X_p$ are i.i.d. random variables, taking values from a large enough discrete set; and $Y = \sum_{i<j} X_i X_j \xi_{ij}$ with $\xi_{ij} \in \{0, 1\}$, there exists an algorithm, which admits $Y, X_1, \ldots, X_p$ as its inputs and recovers the binary variables ξ_{ij} with high probability as $p \to \infty$.

(d) Our analysis, in addition to addressing high-dimensional linear regression and phase retrieval problems, also reveals an algorithmic connection between these problems, and the well-known, yet not directly related to the inference tasks, integer relation detection, randomized subset-sum, and approximate short vector problems. It is also intriguing that all of these three problems are related to an implicit lattice structure; and admit an efficient, lattice-based algorithm[1].

1.2 Notation, Definitions and Preliminaries

1.2.1 Notation

Let $\mathbb{Z}$ denote the set of integers, and $\mathbb{R}$ denote the set of real numbers. Let $\mathbb{Z}^*$ denote $\mathbb{Z} \setminus \{0\}$. We use $\mathbb{Z}^+$ and $\mathbb{R}^+$ for the set of positive integers, and for the set of positive real numbers, respectively. For $k \in \mathbb{Z}^+$, the set $\{1, 2, \ldots, k\}$ is denoted by $[k]$. For a vector $x \in \mathbb{R}^d$ we define $\mathrm{Diag}_{d\times d}(x) \in \mathbb{R}^{d\times d}$ to be the diagonal matrix with $\mathrm{Diag}_{d\times d}(x)_{ii} = x_i$, for $i \in [d]$. For $1 \leq p < \infty$ by $\mathcal{L}_p$ we refer to the standard p-norm notation for finite dimensional real vectors. Given $x \in \mathbb{R}^d$, $\|x\|$ denotes the Euclidean norm $(\sum_{i=1}^d |x_i|^2)^{1/2}$ of x, and $\|x\|_\infty$ denotes $\max_{1\leq i\leq d} |x_i|$. Given two vectors $x, y \in \mathbb{R}^d$ the Euclidean inner product is $\langle x, y \rangle := \sum_{i=1}^d x_i y_i$. By $\log : \mathbb{R}^+ \to \mathbb{R}$ we refer the logarithm with base 2. The integer lattice generated by a set of linearly independent $b_1, \ldots, b_k \in \mathbb{Z}^k$ is defined as $\{\sum_{i=1}^k z_i b_i | z_1, z_2, \ldots, z_k \in \mathbb{Z}\}$, and will be denoted by Λ, where $\Lambda \subseteq \mathbb{Z}^k$. The collection $b_1, \ldots, b_k$ is called the lattice base. Given any real number r, $\lfloor r \rfloor$

[1]Recall that the earlier HJLS algorithm introduced by Hastad et al. [41] for solving integer relation algorithm is based on LLL lattice basis reduction algorithm.

denotes the largest integer not exceeding r; $\lceil r \rceil$ denotes the smallest integer no less than r; and $\{r\} = r - \lfloor r \rfloor$ denotes the fractional part of r. We denote by x^H the complex conjugate of an $x \in \mathbb{C}$. Throughout the paper we use the standard asymptotic notation, o, O, Θ, Ω for comparing the growth of two real-valued sequences $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$. Finally, we say that a sequence of events $\{A_p\}_{p\in\mathbb{N}}$ holds with high probability (whp) as $p \to \infty$ if $\lim_{p\to\infty} \mathbb{P}(A_p) = 1$.

1.2.2 Definitions and Preliminaries

Definition 1.1. *A set $S = \{a_1, \ldots, a_{\mathcal{R}}\} \subset \mathbb{R}$ is called* **linearly independent over rationals** *or in short* **rationally independent**, *if for every $q_1, \ldots, q_{\mathcal{R}} \in \mathbb{Q}$, $\sum_{k=1}^{\mathcal{R}} q_k a_k = 0$ implies, $q_k = 0$, for every $k \in [\mathcal{R}]$.*

As an example, the set, $S = \{\sqrt{2}, \sqrt{3}\}$ is rationally independent, while the set, $S' = \{\sqrt{2}, \sqrt{3} - \sqrt{2}, \sqrt{3}\}$ is not. Clearly, if S is rationally independent, so does any subset $S' \subset S$; and linear independence over rationals is equivalent to the linear independence over integers, which states that, for $q_1, \ldots, q_{\mathcal{R}} \in \mathbb{Z}$ if $\sum_{k=1}^{\mathcal{R}} q_i a_i = 0$, then $q_i = 0$ for all i.

Definition 1.2. *Let $p, Q \in \mathbb{Z}^+$. We say that a vector $\beta^* \in \mathbb{R}^p$ satisfies the* ***Q-rationality assumption*** *if for all $i \in [p]$, $\beta_i^* = K_i/Q$, for some $K_i \in \mathbb{Z}$. Here, we do not necessarily assume that K_i and Q are coprime.*

Our mixed-range assumption, on the entries of β^* is as follows.

Assumption 1. *Let $Q, \mathcal{R}, \widetilde{R} \in \mathbb{Z}^+$, and let $\mathcal{S}_0 = \{a_1, \ldots, a_{\mathcal{R}}\} \subset \mathbb{R}$ be such that $\mathcal{S}_0 \cup \{1\}$ is rationally independent. A vector $\beta^* \in \mathbb{R}^p$ satisfies the* **mixed-range assumption** *if for all $i \in [p]$,*

(i) Either $\beta_i^* = K_i/Q$ for some $K_i \in \mathbb{Z}$, and $\beta_i^* \in [-\widetilde{R}, \widetilde{R}]$ (namely, β_i^* enjoys the aforementioned Q-rationality assumption).

(ii) Or, $\beta_i^* \in \mathcal{S}_0 = \{a_1, \ldots, a_{\mathcal{R}}\}$.

Note that, under mixed-range assumption, the numerator of the rational entries of β^* are upper bounded by $\widetilde{R}$ in magnitude. The rational independence of $\{a_1, \ldots, a_{\mathcal{R}}, 1\}$ is a slightly stronger assumption than $\{a_1, \ldots, a_{\mathcal{R}}\}$ (to see this, note for instance that $\{2 - \sqrt{2}, \sqrt{2}\}$ is rationally independent, whereas $\{2 - \sqrt{2}, \sqrt{2}, 1\}$ is not rationally independent), and the reason for this will become clear, once we provide the details of the algorithm. Note that, rational independence assumption above implies that a_i's are all irrational.

Remark 1.3. *We note that the high-dimensional case, as noted previously, is in general underdetermined. We now show how the Assumption 1 introduces uniqueness in the solution. Suppose that the entries of β^* takes values in the set $\mathcal{S} \triangleq \{a_i : 1 \le i \le \mathcal{R}\} \cup \{a/Q : -\widetilde{R} \le a \le \widetilde{R}, a \in \mathbb{Z}\}$ (thus $|\mathcal{S}| \le \mathcal{R} + 2\widetilde{R} + 1$), and the measurement matrix $X \in \mathbb{R}^{1\times p}$ has jointly continuous entries. Observe that using a union bound, and the fact that X has jointly continuous entries,*

$$\mathbb{P}(\exists \beta \in \mathcal{S}^p : \beta \neq \beta^*, X\beta = X\beta^*) \le |\mathcal{S}|^p \mathbb{P}(X(\beta - \beta^*) = 0) = 0,$$

since $\beta \neq \beta^*$. Namely, with probability one, the map $\beta \mapsto X\beta$ is injective, declaring uniqueness. Note that this argument still remains valid, even when $X \in \mathbb{R}^{1\times p}$ has i.i.d. discrete entries

taking values in a sufficiently large set (an assumption that we use in our results below for the case X is discrete): in this case a very similar union bound argument yields $\beta \mapsto X\beta$ is with high probability injective. Therefore, single-sample recovery is information-theoretically possible.

Our results will reveal an algorithmic connection between the regression problem, and three well-known problems:

Definition 1.4. *An instance of the* **integer relation detection problem** *is as follows. Given a vector $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$, find an $\mathbf{m} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}$, such that $\langle \mathbf{b}, \mathbf{m} \rangle = \sum_{i=1}^n b_i m_i = 0$. In this case, $\mathbf{m}$ is said to be an* **integer relation** *for the vector, $\mathbf{b}$.*

The problem of finding an integer relation for a given $\mathbf{b} \in \mathbb{R}^n$, as mentioned in the introduction, has been studied since Euclid, and algorithms are available; of which the two most well-known are the celebrated HJLS [41] and PSLQ [43] algorithms. For any given $\mathbf{b} \in \mathbb{R}^n$, the PSLQ algorithm with input $\mathbf{b}$ finds an integer relation $\mathbf{m}' \in \mathbb{Z}^n$ for $\mathbf{b}$ in time polynomial in n and $\log \|\mathbf{m}\|$, where $\mathbf{m}$ is a relation for $\mathbf{b}$ with smallest $\|\mathbf{m}\|$ [43]. More concretely, we have the following theorem:

Theorem 1.5. [43] *Let $\mathbf{m} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}$ be an integer relation for $\mathbf{b} \in \mathbb{R}^n$ with the smallest norm $\|\mathbf{m}\|$. Then, PSLQ algorithm with input $\mathbf{b}$ outputs an integer relation for $\mathbf{b}$ after at most $O(n^3 + n^2 \log \|\mathbf{m}\|)$ arithmetic operations on real numbers.*

While Theorem 1.5 is stated for the case when an integer relation $\mathbf{0} \neq \mathbf{m} \in \mathbb{Z}^n$ exists for the input $\mathbf{b} \in \mathbb{R}^n$, it is also worth noting the following. The iterations of the PSLQ algorithm also produce lower bounds on the norm of **any** integer relation $\mathbf{m} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}$ for a given input $\mathbf{b} \in \mathbb{R}^n$. Thus, the PSLQ algorithm can also be used as a "certificate" to declare no integer relations exists for $\mathbf{b}$ whose norm is smaller than a certain value. See [43], as well as Part 3 of the main theorem in the earlier work [42] for more information on this matter.

Note that, the output of the algorithm is not necessarily $\mathbf{m}$, but can be any other integer relation for $\mathbf{b}$. In the sequel, we assume that this algorithm is at our disposal, and we simply refer to it as the Integer Relation Algorithm (IRA).

The following two algorithmic problems are extensively studied in theoretical computer science and cryptography.

Definition 1.6. *An instance of the* **subset-sum problem** *is as follows. Suppose $Y, X_1, \ldots, X_n$ are integers, such that $Y = \sum_{i \in S^*} X_i$ for some $S^* \subset [n]$. Determine S^* from the knowledge $(Y, X_1, \ldots, X_n)$. Equivalently, given integers $X_1, \ldots, X_n$, and $Y = \sum_{i=1}^n X_i e_i$, where $e \in \{0,1\}^n$ is a hidden vector, find e.*

While this problem is NP-hard in worst case, assuming X consists of i.i.d. samples of an appropriate distribution (*e.g.* uniform over a large enough set of integers) Lagarias and Odlyzko [34] and Frieze [38] developed a polynomial-time algorithm, for recovering the hidden subset S^*. Their algortihm is based on the seminal Lenstra-Lenstra-Lovász (LLL) lattice basis reduction algorithm [17]. We detail this in Section 1.2.3.

Remark 1.7. *As mentioned in the introduction, while the subset-sum problem is NP-hard in the worst case; it becomes solvable in polynomial time when the input X_i are random i.i.d., taking values in a large enough discrete set. Namely, the randomness assumption on X is essential for*

the polynomial running time guarantee. For this same reason, we assume, in the sequel, that the measurement matrix X consists of random entries; and we do not expect the polynomial running time guarantees to hold in the case X is deterministic. It is though worth noting that in the case when X is deterministic, the set $\mathcal{S}$ that the entries of β^* take values in is discrete, and the mapping $\beta \mapsto X\beta$ is injective (ensuring the uniqueness); one can still recover $\beta^* \in \mathcal{S}^p$ by a brute force search in time exponential in the dimension p of β^*.

We now introduce our last definition, which is the computational problem of finding short vectors of an integer lattice.

Definition 1.8. *An instance of* **γ-approximate shortest vector problem** *is as follows. Given a lattice base $b_1, \ldots, b_p \in \mathbb{Z}^p$ for an integer lattice, $\Lambda \subseteq \mathbb{Z}^p$, and a constant $\gamma > 0$; find a vector $\widehat{x} \in \Lambda$, such that*

$$\|\widehat{x}\| \leq \gamma \min_{x \in \Lambda, x \neq 0} \|x\|.$$

LLL lattice basis reduction algorithm solves this problem for $\gamma = 2^{\frac{p-1}{2}}$, in a time polynomial in p and $\log \max_{1 \leq i \leq p} \|b_i\|$. This is elaborated in Section 1.2.3.

Some of our results, where the entries of β^* take values in an irrational-valued set, will operate under the so-called joint continuity assumption.

Definition 1.9. *A random vector $X \in \mathbb{R}^p$ is called* **jointly continuous***, if there exists a measurable function $f : \mathbb{R}^p \to [0, \infty)$, called the joint density of X, such that for every Borel set $\mathcal{B} \subseteq \mathbb{R}^p$,*

$$\mathbb{P}(X \in \mathcal{B}) = \int_{\mathcal{B}} f(x_1, \ldots, x_p)\, d\lambda(x_1, \cdots, x_p).$$

Note that, joint continuity is a more general notion than being iid (and even independent): If $X \in \mathbb{R}^p$ is a random vector with independent coordinates X_i, having density $\varphi_i(\cdot)$, then X is also jointly continuous with $f(x_1, \ldots, x_p) = \prod_{\ell=1}^{p} \varphi_i(x_\ell)$.

Computational Model. The algorithms constructed in this paper will be all polynomial time algorithms, but with varying assumptions regarding the input parameters. For the case of rationally valued regression vector β^* our algorithms will be polynomial-time in bit complexity of entries of β^* (in addition to ambient parameters such as the model dimension) as is the case of the LLL algorithm. For the cases involving irrational values for β^*, however, our algorithms will be polynomial time in the size $\mathcal{R}$ of the set of irrational values, as is dictated by the running complexity of the PSLQ algorithm; as well as the dimension p of the vector β^*.

1.2.3 The LLL Algorithm

The seminal Lenstra-Lenstra-Lovász (LLL) lattice basis reduction algorithm, introduced in [17], is a central component of our algorithms. For this reason, we now provide a description of this algorithm together with its connections to the shortest vector and the subset-sum problems. The material here is mostly verbatim from [17].

Given a positive integer n, a subset L of $\mathbb{R}^n$ is called a *lattice* if there exists a basis b_i, $1 \leq i \leq n$ of $\mathbb{R}^n$ such that L can be expressed in form $L = \left\{\sum_{1 \leq i \leq n} r_i b_i : r_i \in \mathbb{Z}, 1 \leq i \leq n\right\}$. In this case the vectors b_i, $1 \leq i \leq n$ form a basis for L. Next, given any linearly independent $b_i \in \mathbb{R}^n$, $1 \leq i \leq n$,

let $b_i^* \in \mathbb{R}^n$, $1 \le i \le n$ denote the vectors obtained inductively from b_i by applying the Gram-Schmidt process. That is, $b_i^* = b_i - \sum_{1 \le j \le i-1} \mu_{ij} b_j^*$, where $\mu_{ij} = \langle b_i, b_j^* \rangle / \langle b_j^*, b_j^* \rangle$, $1 \le j < i \le n$. A basis $b_i \in \mathbb{R}^n$, $1 \le i \le n$ for a lattice L is called *reduced* if

$$|\mu_{ij}| \le \frac{1}{2}, \quad \text{for } 1 \le j < i \le n \quad \text{and} \quad \left\| b_i^* + \mu_{i,i-1} b_{i-1}^* \right\|^2 \ge \frac{3}{4} \|b_{i-1}^*\|^2, \quad \text{for } 1 \le i \le n.$$

Note that the vectors $b_i^* + \mu_{i,i-1} b_{i-1}^*$ and b_{i-1}^* are respectively the projections of b_i and b_{i-1} onto the orthogonal complement of the set spanned by $(b_j : 1 \le j \le i-2)$. The second property essentially means that b_i^* is not too short, compared to b_{i-1}^*.

Given **any** basis, one can obtain a reduced one. This is precisely the task achieved by the LLL algorithm. We now detail this for the case of a lattice $L \subseteq \mathbb{Z}^n$. The algorithm, as an input, takes $b_i \in \mathbb{Z}^n$, $1 \le i \le n$, a basis of $L \subseteq \mathbb{Z}^n$, and proceeds by applying the Gram-Schmidt process to obtain b_i^*, $1 \le i \le n$. The rest of the procedure consists mainly of two steps.

- The first is the **reduction** step: Fix $2 \le i \le n$, and do the following starting from $i = 2$ and continuing: for $i - 1 \ge j \ge 1$ (starting from $i-1$ and going down to 1), keep replacing b_i with $b_i - \lceil \langle b_i, b_j^* \rangle / \langle b_j^*, b_j^* \rangle \rfloor b_j$, where the operator $\lceil \cdot \rfloor$ rounds to the nearest integer (e.g. $\lceil 3.1 \rfloor = 3$ and $\lceil 6.9 \rfloor = 7$). This rounding ensures $|\mu_{ij}| \le \frac{1}{2}$ in the end.

- The second is the **swap** step: if there is an i such that $\frac{3}{4} \|b_i^*\|^2 \ge \|\mu_{i+1,i} b_i^* + b_{i+1}^*\|^2$, then swap b_i and b_{i+1}; and go to start.

Having repeated these steps, the algorithm provably returns a reduced basis eventually. The **reduction** and **swap** steps take care of the requirements of the *reduced* basis (and the details can be found in the paper by Lenstra, Lenstra, and Lovász [17]). The algorithm with input $b_i \in \mathbb{Z}^n$, $1 \le i \le n$ terminates with a *reduced* basis after $O(n^4 \log B)$ aritmetic operations, where $B \triangleq \max_{1 \le i \le n} \|b_i\|_2^2$, as shown in [17, Proposition 1.26].

As noted in Definition 1.8, the LLL algorithm solves the $\gamma-$approximate shortest problem with $\gamma = 2^{\frac{n-1}{2}}$. This is a consequence of [17, Proposition 1.11] which asserts that if $b_i \in \mathbb{R}^n$, $1 \le i \le n$, is a reduced basis for a lattice $L \subseteq \mathbb{R}^n$, then for every non-zero $x \in L$, $\|b_1\| \le 2^{\frac{n-1}{2}} \|x\|$. Hence the first element b_1 of reduced basis is an "approximate short vector" of L.

We now elaborate on how the LLL algorithm is used for solving the subset-sum problem introduced in Definition 1.6, see [34] and [38] for a more detailed discussion. Given i.i.d. $X_i \in \mathbb{Z}$, $1 \le i \le n$, and $Y = \sum_{1 \le i \le n} X_i e_i$ for some $e = (e_i : 1 \le i \le n) \in \{0,1\}^n$, the goal is to obtain e. The idea is to construct an appropriate lattice one of whose "short" vectors is e. Specifically, we consider the lattice $L \subseteq \mathbb{Z}^{n+1}$ with basis $b_0 = (mY, 0, \ldots, 0)$, $b_1 = (-mX_1, 1, 0, \ldots, 0), \ldots, b_n = (-mX_n, 0, 0, \ldots, 1)$, where m is a (large) number (one can, for instance, take $m = \lceil n 2^{n/2} \rceil$). Observe that $b_0 + \sum_{1 \le i \le n} e_i X_i = (0, e_1, \ldots, e_n)$, which has the norm at most $\sqrt{n}$ (recall that $e_i \in \{0,1\}$). Namely, the vector e (more precisely, the vector obtained by augmenting a zero to e) is a "short" vector of L. In particular, the shortest (non-zero) vector of L has an Euclidean norm of at most $\sqrt{n}$. Following the discussion above, the LLL algorithm produces an output vector (i.e., the first element of the reduced basis) which has the norm at most $M \triangleq 2^{\frac{n}{2}} \sqrt{n}$. One can then establish that the short vector produced by the algorithm, with high probability, is a multiple of e: this is done by showing any $v \in L$ with $\|v\| \le M$ must be an integer multiple of e, provided that the discrete i.i.d. variables X_i take values in a sufficiently large set.

In the remainder of the paper, we assume that we have access to the LLL oracle, where the fact that the algorithm runs in time polynomial in the problem dimension and $\log \max_{1\leq i\leq n} \|b_i\|$ will guarantee that the running time of our algorithms are also polynomial (in respective parameters).

Before we proceed with our main results, it is worth noting the following. While the LLL algorithm provably works in theory, as was established before; one may in practice encounter sensitivity issues. Indeed, recall from the discussion above that the LLL algorithm starts with a Gram-Schmidt orthogonalization. Now, the numbers that we use (for instance, m in the Step 4 of the Algorithm 1 as well as several other relevant parameters in the subsequent algorithms) and the dynamic ranges that we consider (for instance, see Remark 2.3 where the dynamic range is $2^{\Theta(p^2)}$) grow quite large as the dimension of the problem grows, making the input of the LLL algorithm to be potentially ill-conditioned (observe though that the logarithms of these numbers still remain polynomial in the problem parameters, therefore not hurting the polynomial running time guarantees). We present in Section 3 a collection of numerical results though which show that our algorithms do not seem to suffer from the sensitivity issues, and after a reasonable running time recover the signal in large fraction of instances, see in particular Figure 1.

2 Main Results

2.1 Noisy High-Dimensional Linear Regression with $Q-$rational β^*

2.1.1 Extended Lagarias-Odlyzko algorithm

Let $n, p, R \in \mathbb{Z}^+$. Given $X \in \mathbb{Z}^{n\times p}, \beta^* \in (\mathbb{Z} \cap [-R, R])^p$ and $W \in \mathbb{Z}^n$, set $Y = X\beta^* + W$. From the knowledge of Y, X the goal is to infer exactly β^*. For this task we propose the following algorithm which is an extension of the algorithm in [34] and [38]. For realistic purposes the values of $R, \|W\|_\infty$ is not assumed to be known exactly. As a result, the following algorithm, besides Y, X, receives as an input a number $\hat{R} \in \mathbb{Z}^+$ which is an estimated upper bound in absolute value for the entries of β^* and a number $\hat{W} \in \mathbb{Z}^+$ which is an estimated upper bound in absolute value for the entries of W.

We explain here informally the steps of the (ELO) algorithm and briefly sketch the motivation behind each one of them. In the first and second steps the algorithm translates Y by XZ where Z is a random vector with iid elements chosen uniformly from $\{\hat{R}+1, \hat{R}+2, \ldots, 2\hat{R}+\lceil \log p \rceil\}$. In that way β^* is translated implicitly to $\beta = \beta^* + Z$ because $Y_1 = Y + XZ = X(\beta^* + Z) + W$. We will, establish using a number theoretic argument, that $\gcd(\beta) = 1$ whp as $p \to +\infty$ with respect to the randomness of Z, even though this is not necessarily the case for the original β^*. This is an essential requirement for our technique to exactly recover β^* and steps six and seven to be meaningful. In the third step the algorithm gets rid of the significantly small observations. This minor but necessary modification affects the observations in a negligible way.

The fourth and fifth steps of the algorithm provide a basis for a specific lattice in $2n + p$ dimensions. The lattice is built with the knowledge of the input and Y_2, the modified version of Y_1. The algorithm in step five calls the LLL basis reduction algorithm using the columns of A_m as initial basis for the lattice. The fact that Y has been modified to be non-zero on every coordinate is essential here so that A_m is full-rank and the LLL basis reduction algorithm, defined in [17], can be applied. This application of the LLL basis reduction algorithm is similar to the one used in [38] with one important modification. In order to deal here with multiple equations

Algorithm 1 Extended Lagarias-Odlyzko (ELO) Algorithm

Input: $(Y, X, \hat{R}, \hat{W})$, $Y \in \mathbb{Z}^n, X \in \mathbb{Z}^{n\times p}$, $\hat{R}, \hat{W} \in \mathbb{Z}^+$.

Output: $\hat{\beta}^*$ an estimate of β^*

1 Generate a random vector $Z \in \{\hat{R}+1, \hat{R}+2, \ldots, 2\hat{R}+\lceil \log p \rceil\}^p$ with iid entries uniform in $\{\hat{R}+1, \hat{R}+2, \ldots, 2\hat{R}+\lceil \log p \rceil\}$

2 Set $Y_1 = Y + XZ$.

3 For each $i = 1, 2, \ldots, n$, if $|(Y_1)_i| < 3$ set $(Y_2)_i = 3$ and otherwise set $(Y_2)_i = (Y_1)_i$.

4 Set $m = 2^{n+\lceil \frac{p}{2} \rceil + 3} p \left(\hat{R} \lceil \sqrt{p} \rceil + \hat{W} \lceil \sqrt{n} \rceil \right)$.

5 Output $\hat{z} \in \mathbb{R}^{2n+p}$ from running the LLL basis reduction algorithm on the lattice generated by the columns of the following $(2n+p) \times (2n+p)$ integer-valued matrix,

$$A_m := \begin{bmatrix} mX & -m\mathrm{Diag}_{n\times n}(Y_2) & mI_{n\times n} \\ I_{p\times p} & 0_{p\times n} & 0_{p\times n} \\ 0_{n\times p} & 0_{n\times n} & I_{n\times n} \end{bmatrix} \tag{1}$$

6 Compute $g = \gcd(\hat{z}_{n+1}, \hat{z}_{n+2}, \ldots, \hat{z}_{n+p})$, using the Euclid's algorithm.

7 If $g \neq 0$, output $\hat{\beta}^* = \frac{1}{g}(\hat{z}_{n+1}, \hat{z}_{n+2}, \ldots, \hat{z}_{n+p})^t - Z$. Otherwise, output $\hat{\beta}^* = 0_{p\times 1}$.

and non-zero noise, we use $2n+p$ dimensions instead of $1+p$ in [38]. Following though a similar strategy as in [38], it can be established that the $n+1$ to $n+p$ coordinates of the output of the algorithm, $\hat{z} \in \mathbb{Z}^{2n+p}$, correspond to a vector which is a non-zero integer multiple of β, say $\lambda\beta$ for $\lambda \in \mathbb{Z}^*$, w.h.p. as $p \to +\infty$.

The proof of the above result is an important part in the analysis of the algorithm and it is heavily based on the fact that the matrix A_m, which generates the lattice, has its first n rows multiplied by the "large enough" and appropriately chosen integer m which is defined in step four. It can be shown that this property of A_m implies that any vector z in the lattice with "small enough" $\mathcal{L}_2$ norm necessarily satisfies $(z_{n+1}, z_{n+2}, \ldots, z_{n+p}) = \lambda\beta$ for some $\lambda \in \mathbb{Z}^*$ whp as $p \to +\infty$. In particular, using that $\hat{z}$ is guaranteed to satisfy $\|\hat{z}\|_2 \leq 2^{\frac{2n+p}{2}} \|z\|_2$ for all non-zero z in the lattice, it can be derived that $\hat{z}$ has a "small enough" $\mathcal{L}_2$ norm and therefore indeed satisfies the desired property whp as $p \to +\infty$. Assuming now the validity of the $\gcd(\beta) = 1$ property, step six finds in polynomial time this unknown integer λ that corresponds to $\hat{z}$, because $\gcd(\hat{z}_{n+1}, \hat{z}_{n+2}, \ldots, \hat{z}_{n+p}) = \gcd(\lambda\beta) = \lambda$. Finally step seven scales out λ from every coordinate and then subtracts the known random vector Z, to output exactly β^*.

Of course the above is based on an informal reasoning. Formally we establish the following result.

Theorem 2.1. *Suppose*

(1) $X \in \mathbb{Z}^{n\times p}$ *is a matrix with iid entries generated according to a distribution* $\mathcal{D}$ *on* $\mathbb{Z}$ *which for some* $N \in \mathbb{Z}^+$ *and constants* $C, c > 0$, *assigns at most* $\frac{c}{2^N}$ *probability on each element of* $\mathbb{Z}$ *and satisfies* $\mathbb{E}[|V|] \leq C2^N$, *for* $V \overset{d}{=} \mathcal{D}$;

(2) $\beta^* \in (\mathbb{Z} \cap [-R, R])^p$, $W \in \mathbb{Z}^n$;

(3) $Y = X\beta^* + W$.

Suppose furthermore that $\hat{R} \geq R$ *and*

$$N \geq \frac{1}{2n}(2n+p)\left[2n+p+10\log\left(\hat{R}\sqrt{p}+(\|W\|_\infty+1)\sqrt{n}\right)\right]+6\log\left((1+c)\,np\right). \tag{2}$$

For any $\hat{W} \geq \|W\|_\infty$ *the algorithm ELO with input* $(Y, X, \hat{R}, \hat{W})$ *outputs* ***exactly*** β^* *w.p.* $1 - O\left(\frac{1}{np}\right)$ *(whp as* $p \to +\infty$*) and terminates in time at most polynomial in* $n, p, N, \log \hat{R}$ *and* $\log \hat{W}$*.*

The constants C and c are hidden under $O(1/np)$. We defer the proof to Section 4.1.

Remark 2.2. *In the statement of Theorem 2.1 the only parameters that are assumed to grow to infinity are* p *and whichever other parameters among* $n, R, \|W\|_\infty, N$ *are implied to grow to infinity because of (2). Note in particular that* n *can remain bounded, including the case* $n = 1$*, if* N *grows fast enough. Similarly,* $\widehat{R}$*, and* $\|W\|_\infty$ *may remain bounded.*

Remark 2.3. *It can be easily checked that the assumptions of Theorem 2.1 are satisfied for* $n = 1$*,* $N = (1+\epsilon)\frac{p^2}{2}$*,* $R = 1$*,* $\mathcal{D} = \text{Unif}\{1, 2, 3, \ldots, 2^{(1+\epsilon)\frac{p^2}{2}}\}$ *and* $W = 0$*. Under these assumptions, the Theorem's implication is a generalization of the result from [34] and [38] to the case* $\beta^* \in \{-1, 0, 1\}^p$*.*

2.1.2 Applications to High-Dimensional Linear Regression

2.1.3 The Model

The high-dimensional linear regression model we are considering is as follows.

Assumption 2. *Let* $n, p, Q \in \mathbb{Z}^+$ *and* $R, \sigma, c > 0$*. Suppose*

(1) measurement matrix $X \in \mathbb{R}^{n\times p}$ *with iid entries generated according to a continuous distribution* $\mathcal{C}$ *which has density* f *with* $\|f\|_\infty \leq c$ *and satisfies* $\mathbb{E}[|V|] < +\infty$*, where* $V \overset{d}{=} \mathcal{C}$*;*

(2) ground truth vector β^* *satisfies* $\beta^* \in [-R, R]^p$ *and the* Q*-rationality assumption;*

(3) $Y = X\beta^* + W$ *for some noise vector* $W \in \mathbb{R}^n$*. It is assumed that either* $\|W\|_\infty \leq \sigma$ *or* W *has i.i.d. entries with mean zero and variance at most* σ^2*, depending on the context.*

We highlight that the "noise" vector W per Assumption 2 can 1) either be deterministic adversarial vector with $|W_i| \leq \sigma$, $1 \leq i \leq n$; or 2) has centered i.i.d. entries with bounded variance (and is not necessarily normal).

Objective: Based on the knowledge of Y and X the goal is to efficiently recover β^*.

2.1.4 The Lattice-Based Regression (LBR) Algorithm

As mentioned in the Introduction, we propose an algorithm to solve the regression problem, which we call the Lattice-Based Regression (LBR) algorithm. The exact knowledge of $Q, R, \|W\|_\infty$ is not assumed. Instead the algorithm receives as an input, additional to Y and X, $\hat{Q} \in \mathbb{Z}^+$ which is an estimated multiple of Q, $\hat{R} \in \mathbb{Z}^+$ which is an estimated upper bound in absolute value for the entries of β^* and $\hat{W} \in \mathbb{R}^+$ which is an estimated upper bound on $\|W\|_\infty$. Furthermore an integer number $N \in \mathbb{Z}^+$ is given to the algorithm as an input, which, as we will explain,

Algorithm 2 Lattice Based Regression (LBR) Algorithm

Input: $(Y, X, N, \hat{Q}, \hat{R}, \hat{W})$, $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}$ and $N, \hat{Q}, \hat{R}, \hat{W} \in \mathbb{Z}^+$.

Output: $\hat{\beta}^*$ an estimate of β^*

2 Set $Y_N = ((Y_i)_N)_{i\in[n]}$ and $X_N = ((X_{ij})_N)_{i\in[n],j\in[p]}$.

3 Set $(\hat{\beta}_1)^*$ to be the output of the ELO algorithm with input:

$$\left(2^N \hat{Q} Y_N, 2^N X_N, \hat{Q}\hat{R}, 2\hat{Q}\left(2^N \hat{W} + \hat{R}p\right)\right).$$

4 Output $\hat{\beta}^* = \frac{1}{\hat{Q}}(\hat{\beta}_1)^*$.

corresponds to a truncation in the data in the first step of the algorithm. Given $x \in \mathbb{R}$ and $N \in \mathbb{Z}^+$ let $x_N = \text{sign}(x)\frac{\lfloor 2^N|x|\rfloor}{2^N}$, which corresponds to the operation of keeping the first N bits after zero of a real number x.

We now explain informally the steps of the Algorithm 2 (LBR) below. In the first step, the algorithm truncates each entry of Y and X by keeping only its first N bits after zero, for some $N \in \mathbb{Z}^+$. This in particular allows to perform finite-precision operations and to call the ELO algorithm in the next step which is designed for integer input. In the second step, the algorithm naturally scales up the truncated data to integer values, that is it scales Y_N by $2^N\hat{Q}$ and X_N by 2^N. The reason for the additional multiplication of the observation vector Y by $\hat{Q}$ is necessary to make sure the ground truth vector β^* can be treated as integer-valued. To see this notice that $Y = X\beta^* + W$ and Y_N, X_N being "close" to Y, X imply

$$2^N\hat{Q}Y_N = 2^N X_N(\hat{Q}\beta^*) + \text{ "extra noise terms"} + 2^N\hat{Q}W.$$

Therefore, assuming the control of the magnitude of the extra noise terms, by using the Q-rationality assumption and that $\hat{Q}$ is estimated to be a multiple of Q, the new ground truth vector becomes $\hat{Q}\beta^*$ which is integer-valued. The final step of the algorithm consist of rescaling now the output of Step 2, to an output which is estimated to be the original β^*. In the next subsection, we turn this discussion into a provable recovery guarantee.

2.1.5 Recovery Guarantees for the LBR algorithm

We state now our first main result, explicitly stating the assumptions on the parameters, under which the LBR algorithm recovers **exactly** β^* from bounded but **adversarial noise** W.

Theorem 2.4.A. *Under Assumption 2 and assuming $\|W\|_\infty \le \sigma$ for some $\sigma \ge 0$, the following holds. Suppose $\hat{Q}$ is a multiple of Q, $\hat{R} \ge R$ and*

$$N > \frac{1}{2}(2n+p)\left(2n + p + 10\log\hat{Q} + 10\log\left(2^N\sigma + \hat{R}p\right) + 20\log(3(1+c)np)\right). \tag{3}$$

For any $\hat{W} \ge \sigma$, the LBR algorithm with input $(Y, X, N, \hat{Q}, \hat{R}, \hat{W})$ terminates with $\hat{\beta}^* = \beta^*$ w.p. $1 - O\left(\frac{1}{np}\right)$ (whp as $p \to +\infty$) and in time polynomial in $n, p, N, \log\hat{R}, \log\hat{W}$ and $\log\hat{Q}$.

Applying Theorem 2.4.A we establish the following result handling **random noise** W.

Theorem 2.4.B. *Under Assumption 2 and assuming $W \in \mathbb{R}^n$ is a vector with iid entries generating according to an, independent from X, distribution $\mathcal{W}$ on $\mathbb{R}$ with mean zero and variance at most σ^2 for some $\sigma \geq 0$ the following holds. Suppose that $\hat{Q}$ is a multiple of Q, $\hat{R} \geq R$, and*

$$N > \frac{1}{2}(2n+p)\left(2n+p+10\log\hat{Q}+10\log\left(2^N\sqrt{np}\sigma+\hat{R}p\right)+20\log(3(1+c)np)\right). \tag{4}$$

For any $\hat{W} \geq \sqrt{np}\sigma$ the LBR algorithm with input $(Y, X, N, \hat{Q}, \hat{R}, \hat{W})$ terminates with $\hat{\beta}^* = \beta^*$ w.p. $1 - O\left(\frac{1}{np}\right)$ (whp as $p \to +\infty$) and in time polynomial in $n, p, N, \log\hat{R}, \log\hat{W}$ and $\log\hat{Q}$.

Note that in the setting of Theorem 2.4.A, the noise W is only assumed to have $\|W\|_\infty \leq \sigma$, with no randomness imposed. Theorem 2.4.B, with random W, is established, by showing if W is random then with probability at least $1 - O(1/p)$, $\|W\|_\infty \leq \sigma\sqrt{np}$ by a simple Markov inequality; and then by applying Theorem 2.4.A.

The proofs of Theorems 2.4.A and 2.4.B are deferred to Section 4.2, and Section 4.3, respectively.

2.1.6 Noise tolerance of the LBR algorithm

The assumptions (2) and (4) might make it hard to build an intuition for the choice of the truncation level N. For this reason, in this subsection we simplify it and state a Proposition explicitly mentioning the optimal truncation level and hence characterizing the optimal level of noise that the LBR algorithm can tolerate with n samples.

First note that in the statements of Theorem 2.4.A and Theorem 2.4.B the only parameters that are assumed to grow are p and, as an implication N, due to (2) and (4). Therefore, importantly, n does not necessarily grow to infinity. That means that Theorem 2.4.A and Theorem 2.4.B imply non-trivial guarantees for *arbitrary sample size* n. The proposition below shows that if σ is at most exponential in $-(1+\epsilon)\left[\frac{(p+2n)^2}{2n} + (2+\frac{p}{n})\log(RQ)\right]$ for some $\epsilon > 0$, then for appropriately chosen truncation level N the LBR algorithm recovers exactly the vector β^* with n samples. In particular, with one sample ($n = 1$) LBR algorithm tolerates noise level up to exponential in $-(1+\epsilon)[p^2/2 + (2+p)\log(QR)]$ for some $\epsilon > 0$. On the other hand, if $n = \Theta(p)$ and $\log(RQ) = o(p)$, the LBR algorithm tolerates noise level up to exponential in $-O(p)$.

Proposition 2.5. *Under Assumption 2 and assuming $W \in \mathbb{R}^n$ is a vector with iid entries generating according to an, independent from X, distribution $\mathcal{W}$ on $\mathbb{R}$ with mean zero and variance at most σ^2 for some $\sigma \geq 0$, the following holds.*

Suppose for some $\epsilon > 0$, $p \geq \frac{300}{\epsilon}\log\left(\frac{300}{(1+c)\epsilon}\right)$, and $\sigma \leq 2^{-(1+\epsilon)\left[\frac{(p+2n)^2}{2n}+(2+\frac{p}{n})\log(RQ)\right]}$. Then the LBR algorithm with

- *input Y, X, $\hat{Q} = Q$, $\hat{R} = R$ and $\hat{W}_\infty = 1$ and*
- *truncation level N satisfying $\log\left(\frac{1}{\sigma}\right) \geq N \geq (1+\epsilon)\left[\frac{(p+2n)^2}{2n} + (2+\frac{p}{n})\log(RQ)\right]$,*

terminates with $\hat{\beta}^* = \beta^*$ w.p. $1 - O\left(\frac{1}{np}\right)$ (whp as $p \to +\infty$) and in time polynomial in $n, p, N, \log\hat{R}, \log\hat{W}$ and $\log\hat{Q}$.

The assumptions $\hat{Q} = Q$, $\hat{R} = R$, and $\hat{W}_\infty = 1$ are imposed to make the noise level that can be tolerated by n samples more transparent. The proof of Proposition 2.5 is deferred to Section 4.12.

It is worth noticing that in the noisy case ($\sigma > 0$) the above Proposition requires the truncation level N to be upper bounded by $\log(\frac{1}{\sigma})$, which implies the seemingly counter-intuitive conclusion that revealing more bits of the data after some point can "hurt" the performance of the recovery mechanism. Note that this is actually justified because of the presence of adversarial noise of magnitute σ. In particular, handling an arbitrary noise of absolute value at most of the order σ implies that the only bits of each observation that are certainly unaffected by the noise are the first $\log\left(\frac{1}{\sigma}\right)$ bits. Any bit in a later position could have potentially changed because of the noise. This correct middle ground for the truncation level N appears to be necessary also in the analysis of the synthetic experiments with the LBR algorithm; see Section 3 in [1] for corresponding synthetic experiments.

2.1.7 Information Theoretic Bounds

In this subsection, we discuss the maximum noise that can be tolerated information-theoretically in recovering a $\beta^* \in [-R, R]^p$ satisfying the Q-rationality assumption. We establish that under Gaussian white noise, any successful recovery mechanism can tolerate noise level at most exponentially small in $-[p \log(QR)/n]$.

Proposition 2.6. *Suppose that $X \in \mathbb{R}^{n\times p}$ is a vector with iid entries following a continuous distribution $\mathcal{D}$ with $\mathbb{E}[|V|] < +\infty$, where $V \overset{d}{=} \mathcal{D}$, $\beta^* \in [-R, R]^p$ satisfies the Q-rationality assumption, $W \in \mathbb{R}^n$ has iid $N(0, \sigma^2)$ entries and $Y = X\beta^* + W$. Suppose furthermore that $\sigma > R(np)^3 \left(2^{\frac{2p\log(2QR+1)}{n}} - 1\right)^{-\frac{1}{2}}$. Then there is **no** mechanism which, whp as $p \to +\infty$, recovers **exactly** β^* with knowledge of Y, X, Q, R, σ. That is, for any function $\hat{\beta}^* = \hat{\beta}^*(Y, X, Q, R, \sigma)$ we have*

$$\limsup_{p\to+\infty} \mathbb{P}\left(\hat{\beta}^* = \beta^*\right) < 1.$$

The proof of Proposition 2.6 is deferred to Section 4.13.

Sharp Optimality of the LBR Algorithm

Using Propositions 2.5 and 2.6 the following **sharp** result is established.

Proposition 2.7. *Under Assumptions 2 where $W \in \mathbb{R}^n$ is a vector with iid $N(0, \sigma^2)$ entries the following holds. Suppose that $n = o\left(\frac{p}{\log p}\right)$ and $RQ = 2^{\omega(p)}$. Then for $\sigma_0 := 2^{-\frac{p\log(RQ)}{n}}$ and $\epsilon > 0$:*

- *if $\sigma > \sigma_0^{1-\epsilon}$,then the w.h.p. exact recovery of β^* from the knowledge of Y, X, Q, R, σ is impossible.*
- *if $\sigma < \sigma_0^{1+\epsilon}$, then the w.h.p. exact recovery of β^* from the knowledge of Y, X, Q, R, σ is possible by the LBR algorithm.*

The proof of Proposition 2.7 is deferred to Section 4.14. Two remarks are now in order.

First, note that the impossibility results per Propositions 2.6 and 2.7 operate under the Gaussian noise assumption, whereas the noise assumption for the algorithm (see part (3) of Assumption 2) is less restrictive. While we do not pursue in the present paper, it is conceivable that one can still establish information-theoretic bounds for other noise models as well, provided an analogue channel capacity under power constraint result for that noise model exists and is appropriately employed.

Second, it is worth mentioning that the impossibility result per Proposition 2.6 and the sharp optimality per Proposition 2.7 pertain the exact recovery, and show that the **exact** recovery is impossible once the noise (variance) exceeds a certain threshold. They do not, however, rule out the impossibility of the **approximate** recovery. We leave this direction as a very interesting open problem for future work.

2.2 Noiseless High-Dimensional Linear Regression with Irrational β^*

We now consider the noiseless setting, where the learner has access to the noiseless linear measurements $Y = X\beta^* \in \mathbb{R}^n$ of β^*. We establish that, under the mixed-range assumption (Assumption 1) on the entries of β^*, efficient recovery is possible for a large class of distributions, whose iid entries constitute the measurement matrix X. Furthermore, we show that, the efficient recovery is possible, even when the learner has access to only one measurement ($n = 1$), under explicitly stated conditions on the distributions of the entries of X. The algorithms that we propose are obtained by using a novel combination of the LLL lattice basis reduction algorithm [17] employed in previous subsections, together with PSLQ integer relation [43] algorithm.

To demonstrate our techniques, we first start in this section with the case where $\beta^* \in \mathbb{R}^p$ consists only of irrational entries, and the measurement matrix consists of either integer or real-valued random entries. The next section will address the recovery problem in the case when $\beta^* \in \mathbb{R}^p$ enjoys the mixed-range assumption.

2.2.1 Integer-Valued Measurement Matrix X

The setup we consider is as follows. The learner has access to n noiseless linear measurements $Y = X\beta^* \in \mathbb{R}^n$, of an irrational-valued feature vector β^* consisting of entries $\beta_i^* \in \mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\}$. The set $\mathcal{S}$ which is known to the learner consists of rationally independent elements; and $X \in \mathbb{Z}^{n\times p}$ consists of iid entries. We propose the JIRSS algorithm to solve this problem. The informal details of this algorithm are as follows.

Note that, for any fixed i, using $Y_i = \langle X_i, \beta^* \rangle$, where X_i is the i^{th} row of X, one can observe that, Y_i can be written as an integral combination of the elements of $\mathcal{S}$, namely, we will establish that, $Y_i = \sum_{j=1}^{\mathcal{R}} \theta_{ij}^* a_j$, for every $i \in [n]$, where integers θ_{ij}^* are defined in the JIRSS algorithm. The algorithm starts by calling IRA[2] with input $(Y_i, a_1, \ldots, a_{\mathcal{R}})$, recovers a relation, and performs rescaling. Due to rational independence, it is not hard to establish that any integer relation for the vector consisting of Y_i, and the elements of $\mathcal{S}$ is an integer multiple of a fixed vector, and the rescaling, as we will show, takes out this constant and reveals the fixed vector. Due to the structure of the relation, this multiple can be obtained almost immediately, unlike the corresponding result for rational-valued β^*, where Y had to be translated by setting $Y_1 = Y + XZ$ in order to ensure that the greatest common divisor of its entries is 1, see line 2 in Algorithm 1,

[2]IRA stands for the integer relation algorithm, see the comment following Theorem 1.5.

and the associated Theorem 2.1. The remainder of the algorithm relies on an observation that, the task of obtaining the underlying values of β^* from the coefficients of the integer relation can be achieved by solving randomized subset-sum problems, where the subset membership is defined by the corresponding values of the feature vector β_i^*. More concretely, if $\xi^{(k)} \in \{0,1\}^p$ is a binary vector, whose i^{th} entry is 1, if and only if $\beta_i^* = a_k$ for $k = 1, 2, \ldots, \mathcal{R}$; then the coefficient θ_{ik}^*, which is the coefficient in the relation between Y_i and the elements of $\mathcal{S}$ corresponding to a_k, can be represented as $\theta_{ik}^* = \langle X_i, \xi^{(k)} \rangle$. This, indeed, is an instance of the subset-sum problem, with the hidden vector observed through multiple channels. Using a slight modification of the algorithm of Frieze [38], and running LLL on an appropriate lattice whose approximate shortest vectors are integer multiples of $\xi^{(k)}$, we will establish that one can recover $\xi^{(k)}$, and hence, the underlying β^*.

Algorithm 3 Joint Integer Relation and Subset Sum Algorithm (JIRSS)

Input: $(Y, \mathcal{S}, X)$, $Y \in \mathbb{R}^n$, $X \in \mathbb{Z}^{n \times p}$, $\mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\} \subset \mathbb{R}$.

Output: $\widehat{\beta^*}$.

Run IRA with inputs $(Y_i, a_1, \ldots, a_{\mathcal{R}})$, and the output $b^{(i)} = (b_0^{(i)}, b_1^{(i)}, \ldots, b_{\mathcal{R}}^{(i)})$, for $i \in [n]$.

Set $\theta_i^* = (\theta_{i1}^*, \ldots, \theta_{i\mathcal{R}}^*) = (-b_1^{(i)}/b_0^{(i)}, \ldots, -b_{\mathcal{R}}^{(i)}/b_0^{(i)})$.

Set $m = p2^{\lceil \frac{p+n}{2} \rceil}$.

Set $\Theta_j = \begin{bmatrix} \theta_{1j}^* & \theta_{2j}^* & \cdots & \theta_{nj}^* \end{bmatrix}^T$, for each $j \in [\mathcal{R}]$.

For each $j = 1, 2, \ldots, \mathcal{R}$, run LLL lattice basis reduction algorithm on the lattice generated by the columns of the following $(n+p) \times (n+p)$ integer-valued matrix,

$$A_j = \begin{bmatrix} m\mathrm{diag}_{n\times n}(\Theta_j) & -mX_{n\times p} \\ 0_{p\times n} & I_{p\times p}, \end{bmatrix}, \quad j = 1, 2, \ldots, \mathcal{R};$$

with outputs, $\gamma^{(1)}, \ldots, \gamma^{(\mathcal{R})} \in \mathbb{Z}^{p+n}$.

For each $j = 1, 2, \ldots, \mathcal{R}$, compute $g_j = \gcd(\gamma_1^{(j)}, \ldots, \gamma_{n+p}^{(j)})$ using Euclid's algorithm.

Set $e^{(j)} = \frac{1}{g_j}\gamma^{(j)}$ for each $j = 1, 2, \ldots, \mathcal{R}$.

For each $i = 1, 2, \ldots, p$, set $\widehat{\beta_i^*} = a_j$ for the smallest $j \geq 1$ such that $e_i^{(j)} = 1$. Output $\widehat{\beta^*}$.

The analysis we pursued imply also that, the high-dimensional linear regression problem in the aforementioned setup is a simultaneous instance of an integer relation detection, and modified subset-sum problems, both of which admit polynomial-in-p time algorithms. In particular, these algorithms turn out to be the fundamental building blocks of our algorithm for recovering β^* exactly and efficiently.

We establish the following formal performance guarantee for the JIRSS algorithm.

Theorem 2.8. *Let $Y = X\beta^*$, under the following assumptions:*

- *$X \in \mathbb{Z}^{n\times p}$ consisting of iid entries, drawn from a distribution $\mathcal{D}$ on $\mathbb{Z}$, where there exists constants $c, C > 0$ such that, $\mathcal{D}$ assigns at most $c/2^N$ probability to each integer, and $\mathbb{E}[|V|] \leq C2^N$, where $V \overset{d}{=} \mathcal{D}$.*
- *$\beta^* \in \mathbb{R}^{p\times 1}$ such that, $\beta_i^* \in \mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\}$, where $\mathcal{S}$ is rationally independent, and is known to the learner.*

Then, JIRSS algorithm with input (Y, S, X) *terminates with* $\widehat{\beta^*} = \beta^*$ *with high probability, as* $p \to \infty$*, in time at most polynomial in* p *and* $\mathcal{R}$*, provided*

$$\lim_{p\to\infty} \left(n + p + n\log(n^2 p) + \frac{n+p}{2}\log p + \frac{(n+p)^2}{2} - n\log c - nN \right) = -\infty.$$

The proof of Theorem 2.8 is deferred to Section 4.4.

We now make the following remarks. First, one can arrive at a threshold for N, in terms of p and n, which reveals roughly the correct order at which N should be growing, so that the parameter assumption of Theorem 2.8 is satisfied. For instance, if $N \geq \frac{1}{2n}(n+p)(n+p+\epsilon\log p)$, then the limit is always $-\infty$. In particular, when n is fixed, N suffices to be of order at least $p^2/2n$, in order to ensure the required limiting behaviour. Moreover; the limiting condition between the sample size n, discretization N, and the dimension p is independent of the size $\mathcal{R}$ of the irrational-valued set $\mathcal{S}$ that the (irrational) entries of β^* take values in. Second, the run time is polynomial in both p, and $\mathcal{R}$. In particular, one can ensure that the overall process runs in time polynomial in p, if $\mathcal{R}$ is at most polynomial in p, that is, $\mathcal{R} = p^{O(1)}$.

Note that, with an appropriate choice of input parameters, the efficient recovery is possible even when the learner has access to only one measurement ($n = 1$). Quantitatively, any discrete distribution for which, there exists a constant $c > 0$, and a parameter $N \geq (1/2 + \epsilon)p^2$ with $\epsilon > 0$ being bounded away from zero, such that the probability mass of each point is bounded above by $c2^{-N}$, works. As a concrete example, one can consider the uniform distribution on $\{1, 2, \ldots, 2^{cp^2}\}$ where $c = 1/2 + \epsilon > 1/2$, a constant, which is the distribution was studied by Frieze [38].

2.2.2 Continuous-Valued Measurement Matrix X

In this part, we focus our attention on the model, $Y = \langle X, \beta^* \rangle$ with $X \in \mathbb{R}^{1\times p}$ and $\beta^* \in \mathbb{R}^p$, such that, $\beta_i^* \in \mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\}$, a rationally-independent set known to the learner. Note that, in this scenario, observing only measurement ($n = 1$) suffice. The reason for this will become clear soon, once the details of the associated algorithm is presented.

We propose the following IHDR algorithm to address this problem.

Algorithm 4 Irrational High-Dimensional Regression (IHDR)

Input: $(Y, \mathcal{S}, X)$, $Y \in \mathbb{R}$, $X \in \mathbb{R}^{1\times p}$, $\mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\} \subset \mathbb{R}$.

Output: $\widehat{\beta^*}$.

1 Set $\mathcal{L} = \{X_i a_j : i \in [p], j \in [\mathcal{R}]\}$.

2 Run the IRA, with inputs $(Y, \mathcal{L})$, denote the output by $(b_0, b_{ij} : i \in [p], j \in [\mathcal{R}])$.

3 If $b_0 = 0$, set $\widehat{\beta^*} = 0$, and halt.

4 If $b_0 \neq 0$, then set $\mathbf{c} = (c_{ij} : i \in [p], j \in [\mathcal{R}]) = (-b_{ij}/b_0 : i \in [p], j \in [\mathcal{R}])$.

5 For each $i \in [p]$, set $\widehat{\beta_i^*} = a_{j(i)}$, where $j(i) = \min_{1\leq j\leq R} |b_{ij}| > 0$.

6 Output $\widehat{\beta^*}$.

The main idea of the algorithm is as follows. Since $X_i \notin \mathbb{Z}$, there exists no integer relation for the vector consisting of the observation Y, and the elements of $\mathcal{S}$. There is, however, a relation between Y, and the elements of the set $\mathcal{L} = \{X_i a_j : i \in [p], j \in [\mathcal{R}]\}$, which is generated from X

and $\mathcal{S}$, by using $p\mathcal{R}$ (which is polynomial in p and $\mathcal{R}$) arithmetic operations on real numbers. We will establish that provided X is jointly continuous (see Definition 1.9); the set $\mathcal{L}$ is rationally independent with probability one, which implies that, any relation for the vector $(Y, \mathcal{L})$ is a multiple of a fixed vector. Then, running IRA with input $(Y, \mathcal{L})$ will allow us to recover this fixed vector, from which the entries of β^* can be decoded. The algorithm that we propose works for any $X \in \mathbb{R}^p$, as long as X is a jointly continuous random vector.

The following result provides the performance guarantee for the IHDR algorithm.

Theorem 2.9. *Suppose, $Y = X\beta^*$ with,*

- *$X \in \mathbb{R}^{1\times p}$, a jointly continuous random vector.*
- *For every $i \in [p]$, $\beta_i^* \in \mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\}$, where $\mathcal{S}$ is rationally independent set known to the learner.*

Then, the IHDR algorithm with inputs $(Y, \mathcal{S}, X)$ terminates with $\widehat{\beta^*} = \beta^*$ with probability 1, after at most polynomial in p and $\mathcal{R}$ number of aritmetic operations on real numbers.

The proof of Theorem 2.9 is deferred to Section 4.5.

Before we close this section, we make the following remarks. The recovery guarantee holds for any jointly continuous random vector $X \in \mathbb{R}^p$, which is more general than mere iid inputs; and for a single measurement, $n = 1$. Second, under the aforementioned assumptions, the noiseless linear regression problem is simply an instance of the integer relation detection problem.

2.3 Noiseless High-Dimensional Linear Regression with Mixed β^*

2.3.1 Integer-Valued Measurement Matrix X

Let $n, p, \widetilde{R} \in \mathbb{Z}^+$; and $Y = X\beta^* \in \mathbb{R}^n$ be n noiseless linear measurements of a vector $\beta^* \in \mathbb{R}^p$, whose entries satisfy Assumption 1. We assume that only an upper bound $\widehat{R} \geq \widetilde{R}$ and a positive integer $\widehat{Q}$ that is divisible by Q are known to learner. The algorithmic goal is to recover β^* exactly, using the information, Y and X. Under these assumptions, we propose the following algorithm.

Algorithm 5 MIRR (Mixed Irrational-Rational Regression) Algorithm

Input: $(Y, X, \widehat{R}, \widehat{Q}, \mathcal{S})$, $Y \in \mathbb{R}^n$, $X \in \mathbb{Z}^{n\times p}$, $\mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\} \subset \mathbb{R}$, $\widehat{R}, \widehat{Q} \in \mathbb{Z}^+$.
Output: $\widehat{\beta^*}$.

1 Run JIRSS algorithm, with inputs $(\widehat{Q}Y, \widehat{Q}X, \mathcal{S})$. Denote the corresponding output as $\widehat{\beta_1^*}$.
2 For each $i \in [p]$ such that, $(\widehat{\beta_1^*})_i \in \mathcal{S}$, set $(\widehat{\beta^*})_i = (\widehat{\beta_1^*})_i$.
3 Set $\widetilde{Y} = Y - X\widehat{\beta_1^*}$, construct $\widetilde{X}$, by erasing i^{th} column of X, if $(\widehat{\beta_1^*})_i \neq 0$. Let the column $n(i)$ of X correspond to the column i of $\widetilde{X}$.
4 If $\widetilde{Y} \notin \mathbb{Z}^n$, halt, and set $\widehat{\beta^*} = 0$.
5 If not, and $\widetilde{Y} \in \mathbb{Z}^n$, run ELO algorithm, with input $(\widehat{Q}Y, X, \widehat{Q}\widehat{R}, 0)$. Denote the output by $\widehat{\beta_2^*}$.
6 Set $(\widehat{\beta^*})_{n(i)} = \frac{1}{Q}(\widehat{\beta_2^*})_i$. Return $\widehat{\beta^*}$.

The Mixed Irrational-Rational Regression Algorithm (MIRR) builds on JIRSS and ELO algorithms, described earlier. We next briefly and informally sketch the steps of MIRR algorithm. Note that, $Y = X\beta^*$ implies that, $\widehat{Q}Y = \widehat{Q}\beta^*$, which can be written as,

$$\widehat{Q}Y_i = \sum_{j=1}^{p} \widehat{Q}X_{ij}\beta_j^* = \sum_{j=1}^{\mathcal{R}} \theta_{ij}^* a_j + \sum_{j:\beta_j^* \in \mathbb{Q}} X_{ij}(\widehat{Q}\beta_j^*) \quad \forall i \in [n],$$

where $\theta_{ij}^* = Q \sum_{k:\beta_k^*=a_j} X_{ik}$. Observe that, $\sum_{j:\beta_j^* \in \mathbb{Q}} \widehat{Q}X_{ij}\beta_j^* \in \mathbb{Z}$, due to the Q-rationality assumption. In particular, for each i, $\widehat{Q}Y_i$ is an integral combination of the elements of $\mathcal{S}$, and 1. The JIRSS step of the algorithm finds, for each i, an integer relation for the vector $(\widehat{Q}Y_i, a_1, \ldots, a_{\mathcal{R}}, 1)$, and then using this relation, recovers the irrational-valued entries β_i^*, and we complete recovering the irrational entries of β^*.

The second step of the algorithm is based on the following decomposition of β^* into its rational and irrational entries: $\beta^* = \beta_I^* + \beta_R^*$, where, $(\beta_I^*)_i = \beta_i^*$ if $\beta_i^* \notin \mathbb{Q}$, and is 0 otherwise; and $(\beta_R^*)_i = \beta_i^*$ if $\beta_i^* \in \mathbb{Q}$, and is 0, otherwise (namely, β_I^* stands for the irrational part of β^*, whereas β_R^* stands for its rational part). With this, we notice $Y = X\beta_R^* + X\beta_I^*$, and establish that β_I^* coincides with the output $\widehat{\beta_1^*}$ of the JIRSS algorithm, with high probability. Hence, $\widetilde{Y} = Y - X\widehat{\beta_1^*}$ with high probability obeys $\widetilde{Y} = \widetilde{X}\widetilde{\beta}$, where $\widetilde{X}$ is obtained by retaining the columns, corresponding to 0 entries in $\widehat{\beta_1^*}$, and $\widetilde{\beta}$ is simply the vector obtained by erasing the entries that are 0 in β_R^*. From here, the problem of recovery of β_R^* is nothing but a regression problem with integer-valued measurement matrix X, and Q-rational feature vector, $\widetilde{\beta}$. This has been discussed in Section 2.1.1. The details of ELO and JIRSS algorithms can be found respectively in Theorem 2.1 and Theorem 2.8.

Formally, we establish the following recovery guarantee.

Theorem 2.10. *Suppose $Y = X\beta^*$, where*

- *$X \in \mathbb{Z}^{n\times p}$ consisting of iid entries, drawn from a distribution $\mathcal{D}$ on $\mathbb{Z}$, where there exists constants $c, C > 0$ such that, $\mathcal{D}$ assigns at most $c/2^N$ probability to each integer, and $\mathbb{E}[|V|] \le C2^N$, where $V \overset{d}{=} \mathcal{D}$;*
- *$\beta^* \in \mathbb{R}^{p\times 1}$ such that, entries of β^* satisfy the mixed-range assumption (Assumption 1), and for $\beta_i^* \in \mathbb{Q}$, it holds that $|\beta_i^*| \le \widetilde{R}$.*

Suppose, the learner has access to $\widehat{R}$ where $\widehat{R} \ge \widetilde{R}$, and $\widehat{Q}$, a multiple of Q. Then, the MIRR algorithm with inputs $(Y, X, \widehat{R}, \widehat{Q}, \mathcal{S})$ recovers β^* whp, in at most polynomial in $n, p, N, \mathcal{R}, \log \widehat{R}, \log \widehat{Q}$ number of operations, provided N satisfies,

$$N \ge \frac{1}{2n}(2n+p)\left[2n + p + 10\log\left(\hat{Q}\hat{R}\sqrt{p} + \sqrt{n}\right)\right] + 6\log\left((1+c)\,np\right). \tag{5}$$

The proof of Theorem 2.10 is deferred to Section 4.6.

We note that, for Equation 5 above, a lower bound on N can be replaced with,

$$N \ge \frac{1}{2n}(2n+s)\left[2n + s + 10\log\left(\hat{Q}\hat{R}\sqrt{p} + \sqrt{n}\right)\right] + 6\log\left((1+c)\,ns\right),$$

where $s = |\{i \in [p] : \beta^*_i \in \mathbb{Q}\}|$, the number of rational entries in β^*. We assume herein s is not available to the learner; and use instead the (worst-case) bound of Equation 5.

We pause to observe again that the single-sample recovery is indeed possible. Note that, provided N is roughly greater than, $(1/2+\epsilon)p^2$, e.g., when X is drawn from Unif$\{1, 2, \ldots, 2^{(1/2+\epsilon)p^2}\}$, one can indeed, with high probability, recover β^* exactly and efficiently, even with a single observation, $n = 1$.

2.3.2 Continuous-Valued Measurement Matrix X

In this section, we focus on the same noiseless regression problem, $Y = X\beta^* \in \mathbb{R}^n$, this time, under the assumption that the measurement matrix, X, consists of samples of a continuous distribution. As in previous section, we assume that the entries of β^* obey the mixed-range assumption (Assumption 1), and for realistic purposes we assume the values of $\widetilde{R}$ and Q are not known explicitly, but rather, an upper bound $\widehat{R}$ for $\widetilde{R}$ and a positive integer $\widehat{Q}$ divisible by Q are known.

The algorithm that we develop is the MIRR-C algorithm, given below.

Algorithm 6 MIRR-C (Mixed Irrational-Rational Regression, Continuous) Algorithm

Input: $(Y, X, N, \widehat{R}, \widehat{Q}, S)$, $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, $S = \{a_1, \ldots, a_{\mathcal{R}}\} \subset \mathbb{R}$, $\widehat{R}, \widehat{Q} \in \mathbb{Z}^+$.

Output: $\widehat{\beta^*}$.

For each $i \in [d]$, set S_i to be the set, $S_i = \{X_{ij}a_k : j \in [p], k \in [\mathcal{R}]\} \cup \{X_{ij} : j \in [p]\}$.

Run integer relation detection algorithm, with input $(\widehat{Q}Y_i, S_i)$. Denote the corresponding output by $b = (b_0, b_{jk}, b_{ij} : j \in [p], k \in [\mathcal{R}])$, corresponding to S_i.

Set $c_{jk} = -b_{jk}/b_0$, for every $j \in [p], k \in [\mathcal{R}]$.

For every $j \in [p]$, set $\widehat{(\beta^*_1)_j} = a_k$, for the smallest $k \in [\mathcal{R}]$, such that $b_{jk} \neq 0$.

Set $\widetilde{Y} = Y - X\widehat{\beta^*_1}$. If $\widetilde{Y} \notin \mathbb{Q}^n$, halt, and output $\widehat{\beta^*} = 0_{p\times 1}$.

Set $\widetilde{X}$ by retaining column i of X, whenever $\widehat{(\beta^*_1)_i} = 0$. Let $n(i)$ be the column X, corresponding to column i of $\widetilde{X}$.

Run LBR algorithm with inputs $(\widetilde{Y}, \widetilde{X}, N, \widehat{Q}, \widehat{R}, 0)$. Denote its output by $\widehat{\beta^*_2}$.

For each i, set $\beta^*_i = \widehat{(\beta^*_1)_i}$, whenever $\widehat{(\beta^*_1)_i} \neq 0$. Then, set $\beta^*_{n(i)} = \widehat{(\beta^*_2)_i}$ for every i, with $\widehat{(\beta^*_2)_i} \neq 0$.

Return $\widehat{\beta^*}$.

The algorithm is two-fold, and uses a mixture of the real-valued computation model for integer relation detection and recovering irrational entries, and a finite precision arithmetic model for recovering the Q-rational part. The algorithm receives as an input, $Y, X, \mathcal{S}$, as well as the parameters $\widehat{Q}, \widetilde{R}$, and the truncation level, N, where we assume, for $x \in \mathbb{R}$, $x_N = \text{sign}(x)\frac{\lfloor 2^N|x|\rfloor}{2^N}$, namely, x_N is the number obtained by keeping the first N-bits of x after binary point, and discarding the rest.

The rationale behind the algorithm is as follows. There is no integer relation, as is, between Y_i and $a_1, \ldots, a_{\mathcal{R}}$, as employed in Theorem 2.10, since the entries of X are not integer-valued. There is, however, a relation, between Y_i, and a certain set, $\mathcal{S}_i$, defined by

$$\mathcal{S}_i = \{X_{ij}a_k : j \in [p], k \in [\mathcal{R}]\} \cup \{X_{ij} : j \in [p]\}, \tag{6}$$

With this, we note that,

$$\widehat{Q}Y_i = \sum_{j=1}^{p} X_{ij}(Q\beta_j^*) = \sum_{j=1}^{p}\sum_{k=1}^{\mathcal{R}} X_{ij}a_k(Qe_{jk}) + \sum_{j:\beta_j^*\in\mathbb{Q}} X_{ij}(Q\beta_j^*),$$

where $e_{jk} \in \{0,1\}$, and is 1 if and only if $\beta_j^* = a_k$. This decomposition shows indeed that, $\widehat{Q}Y_i$ is an integral combination of the members of the set $\mathcal{S}_i$, defined in (6). We will establish that, with probability 1, $\mathcal{S}_i$ is rationally independent; and consequently, any integer relation for the vector consisting of $\widehat{Q}Y_i$ and the elements of $\mathcal{S}_i$, are integer multiples of a certain fixed vector. These observations will allow the recovery of the irrational entries of β_i^*. This is the vector, $\widehat{\beta_1^*}$, defined in the internal steps of the MIRR-C algorithm.

Given this, we turn our attention to recovering the entries of β^*, that are Q-rational. This is done by first, truncating the continuous-valued input matrix, and then running LBR algorithm, as in Theorem 2.4.A. The following theorem establishes the performance guarantee of the MIRR-C algorithm.

Theorem 2.11. *Suppose, $Y = X\beta^*$ where,*

- *The measurement matrix $X \in \mathbb{R}^{n\times p}$ consists of iid entries, drawn from a continuous distribution $\mathcal{C}$ which has density f with $\|f\|_\infty \leq c$, such that, $\mathbb{E}[|V|] < +\infty$, where $V \stackrel{d}{=} \mathcal{C}$;*
- *The vector $\beta^* \in \mathbb{R}^p$ obeys mixed-range assumption (Assumption 1).*

Suppose that, the learner has access to $\widehat{Q}, \widehat{R}, N \in \mathbb{Z}^+$, such that $\widehat{Q}$ is divisible by Q and $\widehat{R} \geq \widetilde{R}$. Then, MIRR-C algorithm with input $(Y, X, N, \widehat{R}, \widehat{Q}, S)$ terminates with $\widehat{\beta^*} = \beta^*$ with high probability, in time that is at most polynomial in $n, p, N, s, \log\widehat{R}, \log\widehat{Q}, \mathcal{R}$; provided the truncation level N satisfies,

$$N > \frac{1}{2}(2n+p)\left(2n+p+10\log\hat{Q}+10\log\left(\hat{R}p\right)+20\log(3\,(1+c)\,np)\right).$$

The proof of Theorem 2.11 is deferred to Section 4.7.

As in the previous setting, it is possible to replace the p's in the lower bound above with s's, where $s = |\{i \in [p] : \beta_i^* \in \mathbb{Q}\}|$. For realistic purposes, we assume, however, s is not available to the learner, and resort to a bound with p, as stated in the preamble of Theorem 2.11.

A further and detailed look at the noiseless regression problem reveal also that, if $\beta^* \in \mathbb{R}^p$ enjoys the mixed-range assumption (Assumption 1), then the problem of recovering β^* can also be cast directly as an integer relation detection problem, and consequently, integer relation oracle alone can be used for recovery. For completeness, we concretize this observation in the following theorem:

Theorem 2.12. *Suppose $Y = X\beta^* \in \mathbb{R}$, where*

- *The measurement vector $X \in \mathbb{R}^{1\times p}$ is a jointly continuous random vector.*
- *The feature vector $\beta^* \in \mathbb{R}^p$ obeys the mixed-range assumption (Assumption 1).*

Suppose that, the learner has access to $\widehat{Q}, \widehat{R} \in \mathbb{Z}^+$, *such that* $\widehat{Q}$ *is divisible by* Q, *and* $\widehat{R} \geq \widetilde{R}$. *Then, IRA with input* $(\widehat{Q}Y, X_i a_j, X_i : i \in [p], j \in [\mathcal{R}])$ *recovers* β^* *with probability one, in time that is at most polynomial in* $p, \mathcal{R}, \log \widehat{Q}$ *and* $\log \widehat{R}$.

Note that, Theorem 2.12 is a single-sample recovery guarantee as promised, and the statement of the theorem holds as long as the measurement vector $X \in \mathbb{R}^p$ is jointly continuous. The proof of this result is deferred to Section 4.8.

2.4 Application: Phase Retrieval Problem

In this section, we address the so-called phase retrieval problem, using the ideas outlined in earlier sections. The goal of the learner is to recover a vector $\beta^* \in \mathbb{C}^p$ efficiently, using the following magnitude-only observations,

$$Y_i = |\langle X_i, \beta^* \rangle|, \quad i = 1, 2, \ldots, n$$

with a small number n of measurements, where, for each i, X_i consists of random samples of a known distribution $\mathcal{D}$. We address this problem, when the elements of β^* take values in a bounded cardinality set satisfying a certain rational independence assumption.

2.4.1 Discrete-Valued X

We establish the JIRSS-based phase retrieval algorithm (see below), which provably recovers $\beta^* \in \mathbb{C}^p$ in polynomial time. The algorithm is given below, and its performance guarantee is the subject of Theorem 2.13.

Theorem 2.13. *Let* $Y = |\langle X, \beta^* \rangle|$ *with,*

- $X \in \mathbb{Z}_+^{1 \times p}$, *with iid entries drawn from a distribution* $\mathcal{D}$ *on* $\mathbb{Z}_+$, *where there exists constants* $c, C > 0$, *such that* $\mathcal{D}$ *assigns at most* $c/2^N$ *probability to each element of* $\mathbb{Z}_+$, *and* $\mathbb{E}[X_1] \leq C 2^N$.
- $\beta_i^* \in \mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\} \subset \mathbb{C}$, *such that the set,* $\mathcal{S}' = \{|a_d|^2 : d \in [\mathcal{R}]\} \cup \{a_i^H a_j + a_i a_j^H : 1 \leq i < j \leq \mathcal{R}\}$ *is rationally independent.*

Then, the JIRSS-based phase retrieval algorithm with input $(Y, \mathcal{S}, X)$ *terminates with* $\widehat{\beta^*} = \beta^*$ *with high probability after at most polynomial in* p *and* $\mathcal{R}$ *number of arithmetic operations, provided* $N \geq (1/8 + \epsilon)p^4$, *for some* $\epsilon > 0$.

The proof of Theorem 2.13 is deferred to Section 4.9.

The high-level idea is similar to that of Theorem 2.8. We first observe that Y^2 can be expressed via integer combinations of the elements of $\mathcal{S}'$, since

$$Y^2 = \sum_{i=1}^{p} X_i^2 |\beta_i^*|^2 + \sum_{1 \leq i < j \leq p} X_i X_j ((\beta_i^*)^H \beta_j^* + \beta_i^* (\beta_j^*)^H).$$

Using similar ideas, one can prove that, any integer relation for the vector consisting of Y and the elements of $\mathcal{S}'$ is an integer multiple of a fixed vector. However, the corresponding coefficients

Algorithm 7 JIRSS-based Phase Retrieval Algorithm

Input: $(Y, \mathcal{S}, X)$, $Y \in \mathbb{R}$, $X \in \mathbb{Z}_+^{1\times p}$, $\mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\} \subset \mathbb{C}$.

Output: $\widehat{\beta^*}$.

Construct the set, $\mathcal{S}' = \{|a_i|^2 : 1 \le i \le \mathcal{R}\} \cup \{a_\alpha^H a_\beta + a_\alpha a_\beta^H : 1 \le \alpha < \beta \le \mathcal{R}\}$.

Run integer relation detection algorithm with inputs (Y, S'), output $b = (b_0, b_1, \ldots, b_{\mathcal{R}}, b_{\alpha\beta} : 1 \le \alpha < \beta \le \mathcal{R})$.

Set $\theta^* = (\theta_1^*, \ldots, \theta_{\mathcal{R}}^*, \theta_{\alpha\beta}^* : 1 \le \alpha < \beta \le \mathcal{R}) = (-b_1/b_0, \ldots, -b_{\mathcal{R}}/b_0, -b_{\alpha\beta}/b_0 : 1 \le \alpha < \beta \le \mathcal{R})$.

Set $m = p2^{\lceil (c'+1/2)p \rceil}$, and $L = \binom{p}{2}$.

Set $Y = (Y_1, \ldots, Y_L) = (X_i X_j : 1 \le i < j \le p)$, using any ordering, e.g. lexicographical.

For each $(\alpha, \beta) : 1 \le \alpha < \beta \le \mathcal{R}$, run LLL lattice basis reduction algorithm on the lattice generated by the columns of the following $(L+1) \times (L+1)$ integer-valued matrix,

$$A_{\alpha,\beta} = \begin{bmatrix} m\theta_{\alpha\beta}^* & -mY \\ 0_{L\times 1} & I_{L\times L} \end{bmatrix},$$

with outputs, $\gamma^{(\alpha,\beta)} \in \mathbb{Z}^{L+1}$ for $1 \le \alpha < \beta \le \mathcal{R}$.

For each $1 \le \alpha < \beta \le \mathcal{R}$, compute $g_{\alpha\beta} = \gcd(\gamma_0^{(\alpha,\beta)}, \ldots, \gamma_N^{(\alpha,\beta)})$ using Euclid's algorithm.

Set $e^{(\alpha,\beta)} = \frac{1}{g_{\alpha\beta}} \gamma^{(\alpha,\beta)}$ for each $1 \le \alpha < \beta \le \mathcal{R}$.

For each $1 \le \alpha < \beta < \delta \le \mathcal{R}$, add all indices i, j with $1 \le i < j \le p$ to $S_{\alpha,\beta}$ whenever the corresponding $\gamma_k^{(\alpha,\beta)} = 1$. For every $i \in S_{\alpha,\beta} \cap S_{\alpha,\delta}$ set $\widehat{\beta_i^*} = a_\alpha$. Break ties arbitrarily.

are not exactly subset-sums of X_i's, but are more involved. Nevertheless, a variation of the result of Frieze [38] allows us to overcome with this obstacle, which, as a by-product, yields an extension of the Frieze's result on the randomly-generated subset-sum problem to random variables with some dependence. This auxiliary result may be of independent interest, and is isolated in Proposition 2.14.

Proposition 2.14. *Let*

- *$X_1, \ldots, X_p$ be i.i.d. random variables, drawn from a distribution $\mathcal{D}$ taking values in $\mathbb{Z}_+$, where there exists constants $c, C > 0$ such that, for each $k \in \mathbb{Z}^+$, $\mathbb{P}(X_1 = k) \le c2^{-N}$, and $\mathbb{E}[X_1] \le C2^N$.*
- *Let $Y = \sum_{1\le i<j\le p} Y_{ij}\xi_{ij}$ with $\xi_{ij} \in \{0, 1\}$, where $Y_{ij} = X_i X_j$.*

Then, there exists an algorithm, which admits Y and $\{X_i\}_{i=1}^p$ as its inputs, and recovers ξ_{ij}'s with high probability, as $p \to \infty$, in polynomial in p many bit operations, provided that $N \ge (1/8+\epsilon)p^4$, for any $\epsilon > 0$.

The proof of Proposition 2.14 is deferred to Section 4.10.

This algorithm is again based on running LLL algorithm on an appropriate lattice, in a similar way as in [38], such that approximate short vectors of this lattice are integer multiples of the binary vector ξ we wish to recover. The lattice is similar to the ones used in earlier results, and will become apparent from the proof. However, due to the presence of dependence, the details of the proof are more involved.

2.4.2 Continuous-Valued X

We now continue with our study of the phase retrieval problem, this time with continuous-valued measurement matrix X, where the entries of hidden $\beta^* \in \mathbb{C}$ take values in a finite cardinality subset $\mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\} \subset \mathbb{C}$ known to learner, such that the set, $\mathcal{S}' = \{|a_i|^2 : i \in [\mathcal{R}]\} \cup \{a_i^H a_j + a_i a_j^H : 1 \leq i < j \leq \mathcal{R}\}$, which can be obtained from $\mathcal{S}$ after at most $O(\mathcal{R}^2)$ arithmetic operations on real numbers, is rationally independent. Using similar ideas as above, one can observe, as a consequence of $Y = |\langle X, \beta^* \rangle|$, that Y^2 can be represented as a linear combination of the elements of a set $\mathcal{L}$, where $\mathcal{L} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$, with

- $\mathcal{S}_1 = \{X_i^2 |a_k|^2 : i \in [p], k \in [\mathcal{R}]\}$,
- $\mathcal{S}_2 = \{X_i X_j |a_k|^2 : 1 \leq i < j \leq p, k \in [\mathcal{R}]\}$,
- $\mathcal{S}_3 = \{X_i X_j (a_k^H a_\ell + a_k a_\ell^H) : 1 \leq i < j \leq p, 1 \leq k < \ell \leq \mathcal{R}\}$.

Note that, $|\mathcal{S}_1| = p\mathcal{R}$, $|\mathcal{S}_2| = O(p^2\mathcal{R})$ and $|\mathcal{S}_3| = O(p^2\mathcal{R}^2)$, and therefore, $|\mathcal{L}| \leq O(p^2\mathcal{R}^2)$. Thus, $\mathcal{L}$ can be obtained from $\mathcal{S}$ and $\{X_i\}_{i=1}^p$ in at most polynomial in p and R many arithmetic operations on real numbers. The most crucial step is to show, $\mathcal{L}$ is rationally independent, with probability one, and this is achieved by combining several ideas from the proof of Theorems 2.9 and 2.13. Having establish that $\mathcal{L}$ is rationally independent, we then run IRA to recover β^*, similar to earlier IHDR algorithm (Algorithm 4) proposed for recovering an irrational-valued β^* taking values in a known, bounded cardinality, rationally independent set, from its noiseless linear measurements $Y = X\beta^*$.

Theorem 2.15. *Let $Y = |\langle X, \beta^* \rangle|$ with,*

- *The measurement matrix $X \in \mathbb{R}^{1\times p}$ is a jointly continuous random vector.*
- *$\beta_i^* \in \mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\} \in \mathbb{C}$, where $\mathcal{S}$ is known to learner, and $\mathcal{S}' = \{|a_i|^2 : i \in [\mathcal{R}]\} \cup \{a_i^H a_j + a_i a_j^H : 1 \leq i < j \leq \mathcal{R}\}$ is rationally independent.*

Then, there exists an algorithm which admits $(Y, \mathcal{S}, X)$ as its input, and terminates with $\widehat{\beta^*} = \beta^*$ with high probability, after at most polynomial in p and $\mathcal{R}$ many arithmetic operations on real numbers.

Observe that, the assertion of the theorem is valid as long as $X \in \mathbb{R}^p$ is a jointly continuous random vector, which is a milder requirement than $X \in \mathbb{R}^p$ having iid coordinates; and the recovery guarantee is provided for a single measurement.

The proof of Theorem 2.15 is deferred to Section 4.11.

We note that, our approach is not limited to phase retrieval setup; and transfers to more general measurement models. Consider an observation model, $Y = f(|\langle X, \beta^* \rangle|)$, where $f(\cdot)$ is an polynomial with $\deg(f) = d$, and $\beta^* \in \mathbb{C}^p$. Note that, $f(t) = Y$ implies, t is a root of the polynomial, $g(t) = f(t) - Y$, where $\deg(g) = d$. Denote by $H = \{t_1, \ldots, t_d\} \subset \mathbb{C}$ the set of all roots of g. Assume that these roots can be found, e.g., by means of an explicit formula, as in the case for $d \leq 4$. Now, for each i with $t_i \in \mathbb{R}_{\geq 0}$, we solve for $(\beta^{(i)})^* \in \mathbb{C}^p$, by running phase retrieval solver with input $(t_i, \mathcal{S}, X)$, that is, we recover $(\beta^{(i)})^*$, such that, $t_i = |\langle X, (\beta^{(i)})^* \rangle|$. It is clear that, the entire process runs in time that is at most polynomial in p, R, and d.

Yet another class of functions $f(\cdot)$ for which our methods immediately extend, is the class of strictly monotone functions. Provided that $f^{-1}(Y)$ can be computed (e.g., again either exactly or by means of a formula), we can simply run the phase retrieval solver with inputs $(f^{-1}(Y), \mathcal{S}, X)$ to recover β^*. Namely, let $\mathcal{C}$ be a class of functions $f : \mathbb{R} \to \mathbb{R}$, where $f \in \mathcal{C}$, if for every $r \in \mathbb{R}$, the set $\{t : f(t) = r\}$ is finite, and can be computed, in the aforementioned sense. Then, our approach extends immediately, when a feature vector β^* is observed through the mechanism, $Y = f(|\langle X, \beta^* \rangle|)$, provided β^* satisfies a similar rational independence assumption.

3 Numerical Experiments

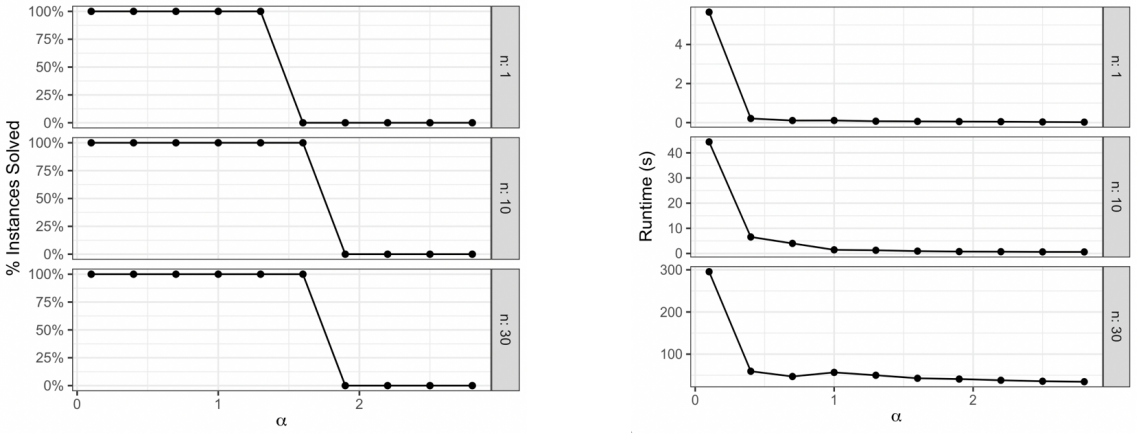

Figure 1: Average performance and the runtime of ELO over 20 instances with $p = 30$ features and $n = 1, 10, 30$ samples.

In this section we present our numerical experiments pertaining the ELO algorithm (Algorithm 1) and the LBR algorithm (Algorithm 2).

The ELO Algorithm

The details of our experiments are as follows. We set $p = 30$ to be the number of features, with sample sizes $n = 1, 10$, and 30; $R = 100$; and zero noise ($W = 0$). Each entry of β^* is uniform in the set $\{1, 2, \ldots, R\}$. For 10 values of $\alpha \in (0, 3)$, specifically for $\alpha \in \{0.25, 0.5, 0.75, 1, 1.3, 1.6, 1.9, 2.25, 2.5, 2.75\}$, we generate the entries of X i.i.d. uniform from the set $\{1, 2, \ldots, 2^N\}$, where $N = \frac{p^2}{2\alpha n}$. For each combination of n and α; we generate 20 independent instances of inputs. We plot in Figure 1 the fraction of instances where the ELO algorithm outputs β^* exactly; and the average termination time of the algorithm.

We observe that for all instances where $\alpha < 1$, the algorithm recovers β^* correctly, with only $p = 30$ features, even though the theoretical guarantees are for large enough p.

Secondly, Theorem 2.1 indicates that for p large, and $N > (2n + p)^2/2n$; the ELO algorithm recovers β^* with high probability. Our experiments indeed confirm this: in that regime, $\alpha = \frac{p^2}{2nN} < 1$. Furthermore, our experiments demonstrate also that the ELO algorithm works for larger values of α.

Finally, the running time of the algorithm was a minute on average; and five minutes in the worst case, thereby granting it reasonable for many application.

The LBR Algorithm

We run our experiments for $p = 30$ features, $n = 10$ samples; $Q = 1$ and $R = 100$. Each entry of β^* is generated i.i.d. where with probability $1/2$ it is equal to zero; and with probability $1/2$; it is drawn uniformly from the set $\{1, 2, \ldots, R\}$. The entries of X are generated i.i.d. uniformly from $(0, 1)$; and the entries of W are generated i.i.d. uniformly from $(-\sigma, \sigma)$ for $\sigma \in \{0, e^{-20}, e^{-12}, e^{-4}\}$.

The fraction of instances where the output of the LBR algorithm coincides with β^* exactly are plotted in Figure 2. Each color represents a different truncation level N. Along each line (with a certain color), the noise power σ is varied.

The main findings are as follows. First, the LBR algorithm works correctly in many cases for $p = 30$, which is a moderate sample size. Secondly, there indeed is an appropriately tuned truncation level N satisfying $(2n + p)^2/2n < N < \log(1/\sigma)$ for which LBR succeeds, which is in exact agreement with Proposition 2.5.

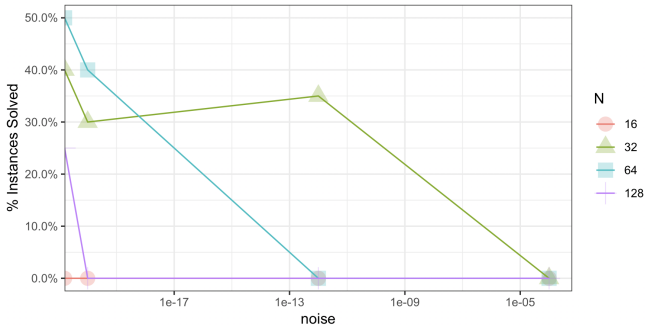

Figure 2: Average performance of LBR algorithm for various noise and truncation levels.

4 Proofs

In what follows below, we often drop the floor and ceiling functions $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ (in particular from the quantities involving logarithms, such as $\log p$) for easiness of notation.

4.1 Proof of Theorem 2.1

Proof. We first observe that directly from (2),

$$\begin{aligned} N &\geq 10 \log\left(\sqrt{p} + \sqrt{n}\left(\|W\|_\infty + 1\right)\right) \\ &\geq 5 \log\left(\sqrt{p}\sqrt{n}\left(\|W\|_\infty + 1\right)\right), \text{ from the elementary } a + b \geq \sqrt{ab} \\ &\geq 2 \log\left(pn\left(\|W\|_\infty + 1\right)\right). \end{aligned}$$

Therefore $2^N \geq \left(pn\left(1 + \|W\|_\infty\right)\right)^2$ which easily implies

$$\frac{\|W\|_\infty}{2^N} \leq \frac{1}{n^2 p^2} = \delta,$$

where we set for convenience $\delta = \delta_p := \frac{1}{n^2 p^2}$.

Lemma 4.1. *For all $i \in [n]$, $|(Y_2)_i| \geq \frac{3}{2}\delta 2^N$, w.p. at least $1 - O\left(\frac{1}{np}\right)$.*

Proof. We divide the proof into two cases based on the magnitude of $\delta 2^N$.

Case 1. Suppose $\delta 2^N < 2$. Then, $|(Y_2)_i| \geq 3 \geq \frac{3}{2}\delta 2^N$ for $1 \leq i \leq n$, due to the second step of the algorithm.

Case 2. Assume now that $\delta 2^N \geq 2$. In this case, observe that

$$Y_1 := Y + XZ = X(\beta^* + Z) + W$$

and therefore from the definition of Y_2, it follows

$$Y_2 = X(\beta^* + Z) + W_1$$

for some $W_1 \in \mathbb{Z}^n$ satisfying

$$\|W_1\|_\infty \leq \|W\|_\infty + 6.$$

Indeed, if i is such that $|(Y_1)_i| \geq 3$, it is the case $(Y_1)_i = (Y_2)_i$; whereas if i is such that $|(Y_1)_i| < 3$, then we set $(Y_2)_i = 3$, hence $|(Y_1)_i - (Y_2)_i| \leq 6$.

Letting $\beta = \beta^* + Z$ we obtain that for all $i \in [n]$, $Y_i = \langle X^{(i)}, \beta \rangle + (W_1)_i$, where $X^{(i)}$ is the i-th row of X. Now, since $(Y_2)_i \geq 0$, it is the case $|(Y_2)_i| = (Y_2)_i$. Consequently, we obtain

$$(Y_2)_i \geq \left|\sum_{j=1}^{p} X_{ij}\beta_j\right| - \|W_1\|_\infty \geq \left|\sum_{j=1}^{p} X_{ij}\beta_j\right| - \|W\|_\infty - 6.$$

Furthermore $\hat{R} \geq R$ implies $\beta \in [1, 3\hat{R} + \log p]^p$.

We now claim that conditional on $\beta \in [1, 3\hat{R} + p]^p$ for $1 \leq i \leq n$,

$$\left|\sum_{j=1}^{p} X_{ij}\beta_j\right| \geq \frac{11}{2}\delta 2^N$$

with probability at least $1 - O\left(\frac{1}{np}\right)$ with respect to the randomness of X. Note that this last inequality alongside with $\|W\|_\infty \leq \delta 2^N$ implies for all i,

$$|(Y_2)_i| \geq \frac{11}{2}\delta 2^N - \delta 2^N - 6 = \frac{9}{2}\delta 2^N - 6 \geq \frac{3}{2}\delta 2^N,$$

since we are inspecting the case $\delta 2^N \geq 2$. Thus, modulo this claim, it follows that with probability at least $1 - O\left(\frac{1}{np}\right)$, $|(Y_2)_i| \geq \frac{3}{2}\delta 2^N$ for every i. Therefore, to conclude the proof of Lemma 4.1, it suffices to prove the claim.

In order to prove the claim, observe that for large enough p,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n}\left\{\left|\sum_{j=1}^{p} X_{ij}\beta_j\right| < \frac{11}{2}\delta 2^N\right\}\right) &\leq \sum_{i=1}^{n}\mathbb{P}\left(\left|\sum_{j=1}^{p} X_{ij}\beta_j\right| < \frac{11}{2}\delta 2^N\right) \\
&= \sum_{i=1}^{n}\sum_{k\in\mathbb{Z}\cap\left[-\frac{11}{2}\delta 2^N, \frac{11}{2}\delta 2^N/2\right]}\mathbb{P}\left(\sum_{j=1}^{p} X_{ij}\beta_j = k\right) \\
&\leq n(11\delta 2^N + 1)\frac{c}{2^N} \\
&\leq n\cdot 12\delta 2^N \cdot \frac{c}{2^N} \\
&= 12\delta nc = O\left(\frac{1}{np}\right).
\end{aligned}$$

We now justify these lines. The first line follows from a union bound over $1 \leq i \leq n$; and the second line follows from a union bound over all integers in $\left[-\frac{11}{2}\delta 2^N, \frac{11}{2}\delta 2^N\right]$, recalling that $\sum_{1\leq j\leq p} X_{ij}\beta_j \in \mathbb{Z}$ for $1 \leq i \leq n$. For the third line, we have used that given $\beta_1 \neq 0$ for $i \in [p]$ and $k \in \mathbb{Z}$ the event $\{\sum_{j=1}^{p} X_{ij}\beta_j = k\}$ implies that the random variable X_{i1} takes a specific value, conditional on the realization of the remaining elements $X_{i2}, \ldots, X_{ip}$ involved in the equations. Therefore by our assumption on the i.i.d. distribution generating the entries of X, each of these events has probability at most $c/2^N$. Note that the choice of β_1, as opposed to choosing some β_i with $i > 1$, was arbitrary in the previous argument. The fourth line uses the assumption $\delta 2^N \geq 2 > 1$, hence $12\delta 2^N > 11\delta 2^N + 1$; and the final line is justified from the choice that $\delta = O(\frac{1}{n^2 p})$ and that $12c$ is a constant. □

Next we use a number-theoretic lemma, which is an extension of a standard result in analytic number theory according to which

$$\lim_{m\to+\infty} \mathbb{P}_{P,Q\sim\mathrm{Unif}\{1,2,\ldots,m\}, P\perp\!\!\!\perp Q}\left[\gcd(P,Q) = 1\right] = \frac{6}{\pi^2},$$

where $P \perp\!\!\!\perp Q$ refers to P, Q being independent random variables. This result is not of clear origin in the literature, but possibly it is attributed to Chebyshev, as mentioned in [48].

Lemma 4.2. *Suppose $q_1, q_2, q \in \mathbb{Z}^+$ with $q \to +\infty$ and $\max\{q_1, q_2\} = o(q^2)$. Then*

$$\left|\{(a,b) \in \mathbb{Z}^2 \cap ([q_1, q_1+q] \times [q_2, q_2+q]) : \gcd(a,b) = 1\}\right| = q^2\left(\frac{6}{\pi^2} + o_q(1)\right).$$

In other words, if we choose independently one uniform integer in $[q_1, q_1+q]$ and another uniform integer in $[q_2, q_2+q]$ the probability that these integers are relatively prime approaches $\frac{6}{\pi^2}$, as $q \to +\infty$.

Proof. We call an integer $n \in \mathbb{Z}^+$ square-free if it is not divisible by the square of a positive integer number other than 1. The **Mobius function** $\mu : \mathbb{Z}^+ \to \{-1, 0, 1\}$ is defined to be

$$\mu(n) = \begin{cases} 1, & \text{n is square-free with an even number of prime factors} \\ -1, & \text{n is square-free with an odd number of prime factors} \\ 0, & \text{otherwise} \end{cases}$$

From now on we ease the notation by always referring for this proof to positive integer variables. A standard property for the Mobius function (see Theorem 263 in [47]) states that for all $n \in \mathbb{Z}^+$,

$$\sum_{1\le d\le n, d|n} \mu(d) = \begin{cases} 1, & n = 1 \\ 0, & \text{otherwise} \end{cases}$$

Therefore using the above identity and switching the order of summation we obtain

$$\begin{aligned} &|(a,b) \in [q_1, q_1+q] \times [q_2, q_2+q], \gcd(a,b) = 1| \\ &= \sum_{(a,b)\in[q_1,q_1+q]\times[q_2,q_2+q]} \left(\sum_{1\le d\le \gcd(a,b), d|\gcd(a,b)} \mu(d) \right) \\ &= \sum_{1\le d\le \max\{q_1,q_2\}+q} \left(\sum_{(a,b)\in[q_1,q_1+q]\times[q_2,q_2+q], d|\gcd(a,b)} \mu(d) \right). \end{aligned}$$

Now introducing the change of variables $a = kd, b = ld$ for some $k, l \in \mathbb{Z}^+$ and observing that the number of integer numbers in an interval of length $x > 0$ are $x + O(1)$, we obtain

$$\begin{aligned} &\sum_{1\le d\le \max\{q_1,q_2\}+q} \left(\sum_{\frac{q_1}{d}\le k\le \frac{q_1+q}{d}, \frac{q_2}{d}\le l\le \frac{q_2+q}{d}} \mu(d) \right) \\ &= \sum_{1\le d\le \max\{q_1,q_2\}+q} \left[\left(\frac{q}{d} + O(1) \right)^2 \mu(d) \right] \\ &= \sum_{1\le d\le \max\{q_1,q_2\}+q} \left[\left(\frac{q}{d} \right)^2 \mu(d) + O\left(\frac{q}{d} \right) \mu(d) + O(1)\mu(d) \right] \end{aligned}$$

Now using $|\mu(d)| \le 1$ for all $d \in \mathbb{Z}^+$, for $n \in \mathbb{Z}^+$,

$$\sum_{d=1}^{n} \frac{1}{d} = O(\log n)$$

and that by Theorem 287 in [47] for $n \in \mathbb{Z}^+$,

$$\sum_{d=1}^{n} \frac{\mu(d)}{d^2} = \frac{1}{\zeta(2)} + o_n(1) = \frac{6}{\pi^2} + o_n(1)$$

we conclude that the last quantity equals

$$q^2 \left(\frac{6}{\pi^2} + \frac{1}{q} O(\log(\max\{q_1, q_2\} + q)) + \frac{\max\{q_1, q_2\} + q}{q^2} + o_q(1) \right).$$

Recalling the assumption $q_1, q_2 = o(q^2)$ the proof is complete. □

Claim 4.3. *The greatest common divisor of the coordinates of $\beta := \beta^* + Z$ equals to 1, w.p. $1 - \exp(-\Theta(p))$ with respect to the randomness of Z.*

Proof. Each coordinate of β is a uniform and independent choice of a positive integer from an interval of length $2\hat{R}+\log p$ with starting point in $[\hat{R}-R+1, \hat{R}+R+1]$, depending on the value of $\beta_i^* \in [-R, R]$. Note though that Lemma 4.2 applies for arbitrary $q_1, q_2 \in [\hat{R}-R+1, \hat{R}+R+1]$ and $q = 2\hat{R}+\log p$ since $q_1, q_2 = o(q^2)$ and $q \to +\infty$. from this we conclude that the probability any two specific coordinates of β have greatest common divisor 1 approaches $\frac{6}{\pi^2}$, as $p \to +\infty$. But the probability the greatest common divisor of all the coordinates is not one implies that the greatest common divisor of the $2i-1$ and $2i$ coordinate is not one, for every $i = 1, 2, \ldots \lfloor \frac{p}{2} \rfloor$. Hence using the independence among the values of the coordinates, we conclude that the greatest common divisor of the coordinates of β is not one with probability at most

$$\left(1 - \frac{6}{\pi^2} + o_p(1)\right)^{\lfloor \frac{p}{2} \rfloor} = \exp\left(-\Theta\left(p\right)\right).$$

□

Given a vector $z \in \mathbb{R}^{2n+p}$, define $z_{n+1:p} := (z_{n+1}, \ldots, z_{n+p})^t$.

Claim 4.4. *The outcome of Step 5 of the algorithm, $\hat{z}$, satisfies*

- $\|\hat{z}\|_2 < m$
- $\hat{z}_{n+1:n+p} = q\beta$, *for some* $q \in \mathbb{Z}^*$, *w.p.* $1 - O\left(\frac{1}{np}\right)$.

Proof. Call $\mathcal{L}_m$the lattice generated by the columns of the $(2n+p) \times (2n+p)$ integer-valued matrix A_m defined in the algorithm; that is $\mathcal{L}_m := \{A_m z | z \in \mathbb{Z}^{2n+p}\}$. Notice that as Y_2 is nonzero at every coordinate, the lattice $\mathcal{L}_m$ is full-dimensional and the columns of A_m define a basis for $\mathcal{L}_m$. Finally, an important vector in $\mathcal{L}_m$ for our proof is $z_0 \in \mathcal{L}_m$ which is defined for $1_n \in \mathbb{Z}^n$ the all-ones vector as

$$z_0 := A_m \begin{bmatrix} \beta \\ 1_n \\ W_1 \end{bmatrix} = \begin{bmatrix} 0_{n\times 1} \\ \beta \\ W_1 \end{bmatrix} \in \mathcal{L}_m. \tag{7}$$

Consider the following optimization problem on $\mathcal{L}_m$, known as the shortest vector problem,

$$\begin{array}{lll} (\mathcal{S}_2) & \min & \|z\|_2 \\ & \text{s.t.} & z \in \mathcal{L}_m, \end{array}$$

If z^* is the optimal solution of $(\mathcal{S}_2)$ we obtain

$$\|z^*\|_2 \leq \|z_0\|_2 = \sqrt{\|\beta\|_2^2 + \|W_1\|_2^2} \leq \|\beta\|_\infty \sqrt{p} + \|W_1\|_\infty \sqrt{n}.$$

and therefore given our assumptions on β, W

$$\|z^*\|_2 \leq \left(3\hat{R} + \log p\right)\sqrt{p} + \left(\|W\|_\infty + 1\right)\sqrt{n}.$$

Using that $\hat{R} \geq 1$ and a crude bound this implies

$$\|z^*\|_2 \leq 4p\left(\hat{R}\sqrt{p} + \left(\|W\|_\infty + 1\right)\sqrt{n}\right).$$

The LLL guarantee and the above observation imply that

$$\|\hat{z}\|_2 \leq 2^{\frac{2n+p}{2}}\|z^*\|_2 \leq 2^{\frac{2n+p}{2}+2}p\left(\hat{R}\sqrt{p} + (\|W\|_\infty + 1)\sqrt{n}\right) := m_0. \tag{8}$$

Now recall that $\hat{W}_\infty \geq \max\{\|W\|_\infty, 1\}$. Since $m \geq 2^{n+\frac{p}{2}+3}p\left(\hat{R}\sqrt{p} + \hat{W}_\infty\sqrt{n}\right)$, we obtain $m > m_0$ and hence $\|\hat{z}\|_2 < m$. This establishes the first part of the Claim.

For the second part, given (8) and that $\hat{z}$ is non-zero it suffices to establish that under the conditions of our Theorem there is no non-zero vector in $\mathcal{L}_m \backslash \{z \in \mathcal{L}_m | z_{n+1:n+p} = q\beta, q \in \mathbb{Z}^*\}$ with L_2 norm less than m_0, w.p. $1 - O\left(\frac{1}{np}\right)$. By construction of the lattice for any $z \in \mathcal{L}_m$ there exists an $x \in \mathbb{Z}^{2n+p}$ such that $z = A_m x$. We decompose $x = (x_1, x_2, x_3)^t$ where $x_1 \in \mathbb{Z}^p, x_2, x_3 \in \mathbb{Z}^n$. It must be true

$$z = \begin{bmatrix} m\left(Xx_1 - \text{Diag}_{n\times n}(Y)x_2 + x_3\right) \\ x_1 \\ x_3 \end{bmatrix}.$$

Note that $x_1 = z_{n+1:n+p}$. We use this decomposition of every $z \in \mathcal{L}_m$ to establish our result.

We first establish that for any lattice vector $z \in \mathcal{L}_m$ the condition $\|z\|_2 \leq m_0$ implies necessarily

$$Xx_1 - \text{Diag}_{n\times n}(Y)x_2 + x_3 = 0. \tag{9}$$

and in particular $z = (0, x_1, x_3)$. If not, as it is an integer-valued vector, $\|Xx_1 - \text{Diag}_{n\times n}(Y)x_2 + x_3||_2 \geq 1$ and therefore

$$m \leq m\|Xx_1 - \text{Diag}_{n\times n}(Y)x_2 + x_3\|_2 \leq \|z\|_2 \leq m_0,$$

a contradiction as $m > m_0$. Hence, necessarily equation (9) and $z = (0, x_1, x_3)$ hold.

Now we claim that it suffices to show that there is no non-zero vector in $\mathcal{L}_m \backslash \{z \in \mathcal{L}_m | z_{n+1:n+p} = q\beta, q \in \mathbb{Z}\}$ with L_2 norm less than m_0, w.p. $1 - O\left(\frac{1}{np}\right)$. Note that in this claim the coefficient q is allowed to take the zero value as well. The reason it suffices to prove this weaker statement is that any non-zero $z \in \mathcal{L}_m$ with $\|z\|_2 \leq m_0$ necessarily satisies that $z_{n+1:n+p} \neq 0$ w.p. $1 - O\left(\frac{1}{np}\right)$ and therefore the case $q = 0$ is not possible w.p. $1 - O\left(\frac{1}{np}\right)$. To see this, we use the decomposition and recall that $x_1 = z_{n+1:n+p}$. Therefore it suffices to establish that there is no triplet $x = (0, x_2, x_3)^t \in \mathbb{Z}^{2n+p}$ with $x_2, x_3 \in \mathbb{Z}^n$ for which the vector $z = A_m x \in \mathcal{L}_m$ is non-zero and $\|z\|_2 \leq m_0$, w.p. $1 - O\left(\frac{1}{np}\right)$. To prove this, we consider such a triplet $x = (0, x_2, x_3)$ and will upper bound the probability of its existence. From equation (9) it necessarily holds $\text{Diag}_{n\times n}(Y)x_2 = x_3$, or equivalently

$$\text{for all } i \in [n], Y_i(x_2)_i = (x_3)_i. \tag{10}$$

From Lemma 4.1 and (10) we obtain that

$$\text{for all } i \in [n], \frac{3}{2}\delta 2^N |(x_2)_i| \leq |(x_3)_i| \tag{11}$$

w.p. $1 - O\left(\frac{1}{np}\right)$. Since z is assumed to be non-zero and $z = A_m x = (0, 0, x_3)$ there exists $i \in [n]$ with $(x_3)_i \neq 0$. Using (10) we obtain $(x_2)_i \neq 0$ as well. Therefore for this value of i it must be

simultaneously true that $|(x_2)_i| \geq 1$ and $|(x_3)_i| \leq m_0$. Plugging these inequalities to (11) for this value of i, we conclude that it necessarily holds that

$$\frac{3}{2}\delta 2^N \leq m_0$$

Using the definition of δ, $\delta = \frac{1}{n^2p^2}$, we conclude that it must hold $\frac{1}{n^2p^2}2^N \leq m_0$, or

$$N \leq 2\log(np) + \log m_0.$$

Plugging in the value of m_0 we conclude that for sufficiently large p,

$$N \leq 2\log(np) + \frac{2n+p}{2} + \log p + \log\left(\hat{R}\sqrt{p} + (\|W\|_\infty + 1)\sqrt{n}\right).$$

This can be checked to contradict directly our hypothesis (2) and the proof of the claim is complete.

Therefore using the decomposition of every $z \in \mathcal{L}_m$, equation (9) and the claim in the last paragraph it suffices to establish that w.p. $1 - O\left(\frac{1}{np}\right)$ there is no triplet (x_1, x_2, x_3) with

(a) $x_1 \in \mathbb{Z}^p, x_2, x_3 \in \mathbb{Z}^n$;

(b) $\|x_1\|_2^2 + \|x_3\|_2^2 \leq m_0$;

(c) $Xx_1 - \mathrm{Diag}_{n\times n}(Y)x_2 - x_3 = 0$;

(d) $\forall q \in \mathbb{Z} : x_1 \neq q\beta$.

We first claim that any such triplet (x_1, x_2, x_3) satifies w.p. $1 - O\left(\frac{1}{np}\right)$

$$\|x_2\|_\infty = O(\frac{m_0 n^2 p^3}{\delta}).$$

To see this let $i = 1, 2, \ldots, n$ and denote by $X^{(i)}$ the i-th row of X. We have because of (c),

$$0 = (Xx_1 - \mathrm{Diag}_{n\times n}(Y)x_2 - x_3)_i = \langle X^{(i)}, x_1\rangle - Y_i(x_2)_i - (x_3)_i,$$

and therefore by triangle inequality

$$|Y_i(x_2)_i| = |\langle X^{(i)}, x_1\rangle - (x_3)_i| \leq |\langle X^{(i)}, x_1\rangle| + |(x_3)_i|. \tag{12}$$

But observe that for all $i \in [n], \|X^{(i)}\|_\infty \leq \|X\|_\infty \leq (np)^2 2^N$ w.p. $1 - O\left(\frac{1}{np}\right)$. Indeed using a union bound, Markov's inequality and our assumption on the distribution $\mathcal{D}$ of the entries of X,

$$\mathbb{P}\left(\|X\|_\infty > (np)^2 2^N\right) \leq np\mathbb{P}\left(|X_{11}| > (np)^2 2^N\right) \leq \frac{1}{2^N np}\mathbb{E}[|X_{11}|] \leq \frac{C}{np} = O\left(\frac{1}{np}\right),$$

which establishes the result. Using this, Lemma 4.1 and (12) we conclude that for all $i \in [n]$ w.p. $1 - O\left(\frac{1}{np}\right)$

$$|(x_2)_i|\frac{3}{2}\delta 2^N \leq (2^N p(np)^2 + 1)m_0$$

which in particular implies

$$|(x_2)_i| \leq O(\frac{m_0 n^2 p^3}{\delta}),$$

w.p. $1 - O\left(\frac{1}{np}\right)$.

Now we claim that for any such triplet (x_1, x_2, x_3) it also holds

$$\mathbb{P}\left(Xx_1 - \mathrm{Diag}_{n\times n}(Y)x_2 - x_3 = 0\right) \leq \frac{c^n}{2^{nN}}. \tag{13}$$

To see this note that for any $i \in [n]$ if $X^{(i)}$ is the i-th row of X because $Y = X\beta + W$ it holds $Y_i = \langle X^{(i)}, \beta\rangle + W_i$. In particular, $Xx_1 - \mathrm{Diag}_{n\times n}(Y)x_2 - x_3 = 0$ implies for all $i \in [n]$,

$$\begin{aligned} &\langle X^{(i)}, x_1\rangle - Y_i(x_2)_i = (x_3)_i \\ \text{or } &\langle X^{(i)}, x_1\rangle - \left(\langle X^{(i)}, \beta\rangle + W_i\right)(x_2)_i = (x_3)_i \\ \text{or } &\langle X^{(i)}, x_1 - (x_2)_i\beta\rangle = (x_3)_i - (x_2)_i W_i \end{aligned}$$

Hence using independence between rows of X,

$$\mathbb{P}\left(Xx_1 - \mathrm{Diag}_{n\times n}(Y)x_2 - x_3 = 0\right) = \prod_{i=1}^{n} \mathbb{P}\left(\langle X^{(i)}, x_1 - (x_2)_i\beta\rangle = (x_3)_i - (x_2)_i W_i\right) \tag{14}$$

But because of (d) for all i, $x_1 - (x_2)_i\beta \neq 0$. In particular, $\langle X^{(i)}, x_1 - (x_2)_i\beta\rangle = (x_3)_i - (x_2)_i W_i$ constraints at least one of the entries of $X^{(i)}$ to get a specific value with respect to the rest of the elements of the row which has probability at most $\frac{c}{2^N}$ by the independence assumption on the entries of X. This observation with (14) implies (13).

Now, we establish that indeed there are no such triplets, w.p. $1-O\left(\frac{1}{np}\right)$. Recall the standard fact that for any $r > 0$ there are at most $O(r^n)$ vectors in $\mathbb{Z}^n$ with L_∞-norm at most r. Using this, (13) and a union bound over all the integer vectors (x_1, x_2, x_3) with $\|x_1\|_2^2 + \|x_3\|_2^2 \leq m_0$, $\|x_2\|_\infty = O(\frac{m_0 n^2 p^3}{\delta})$ we conclude that the probability that there exist a triplet (x_1, x_2, x_3) satisfying (a), (b), (c), (d) is at most of the order

$$(\frac{m_0 n^2 p^3}{\delta})^n m_0^{n+p} \left[\frac{c^n}{2^{nN}}\right].$$

Plugging in the value of m_0 we conclude that the probability is at most of the order

$$\frac{2^{\frac{1}{2}(2n+p)^2 + n\log(cn^2p^3) + n\log(\frac{1}{\delta}) + (2+\log p)(2n+p)} \left[\hat{R}\sqrt{p} + (\|W\|_\infty + 1)\sqrt{n}\right]^{2n+p}}{2^{nN}}.$$

Now recalling that $\delta = \frac{1}{n^2p^2}$ we obtain $\log(\frac{1}{\delta}) = 2\log(np)$ and therefore the last bound becomes at most of the order

$$\frac{2^{\frac{1}{2}(2n+p)^2 + 5n\log(cnp) + (2+\log p)(2n+p)} \left[\hat{R}\sqrt{p} + (\|W\|_\infty + 1)\sqrt{n}\right]^{2n+p}}{2^{nN}}.$$

We claim that the last quantity is $O\left(\frac{1}{np}\right)$ because of our assumption (2). Indeed the logarithm of the above quantity equals

$$\frac{1}{2}(2n+p)\left(2n+p+4+2\log p+2\log\left(\hat{R}\sqrt{p}+(\|W\|_\infty+1)\sqrt{n}\right)\right)+5n\log(cnp)-nN.$$

Using that $\hat{R}\geq 1$ this is upper bounded by

$$\frac{1}{2}(2n+p)\left(2n+p+10\log\left(R\sqrt{p}+(\|W\|_\infty+1)\sqrt{n}\right)\right)+5n\log(cnp)-nN$$

which by our assumption (2) is indeed less than $-n\log(np)<-\log(np)$, implying the desired bound. This completes the proof of claim 4.4. □

Now we prove Theorem 2.1. First with respect to time complexity, it suffices to analyze Step 5 and Step 6. For step 5 we have from [17] that it runs in time polynomial in $n,p,\log\|A_m\|_\infty$ which indeed is polynomial in n,p,N and $\log\hat{R},\log\hat{W}$. For step 6, recall that the Euclid algorithm to compute the greatest common divisor of p numbers with norm bounded by $\|\hat{z}\|_\infty$ takes time which is polynomial in $p,\log\|\hat{z}\|_\infty$. But from Claim 4.4 we have that $\|\hat{z}\|_\infty<m$ and therefore the time complexity is polynomial in $p,\log m$ and therefore again polynomial in n,p,N and $\log\hat{R},\log\hat{W}$.

Finally we prove that the ELO algorithm outputs exactly β^* w.p. $1-O\left(\frac{1}{np}\right)$. We obtain from Claim 4.4 that $\hat{z}_{n+1:n+p}=q\beta$ for $\beta=\beta^*+Z$ and some $q\in\mathbb{Z}^*$ w.p. $1-O\left(\frac{1}{np}\right)$. We claim that the g computed in Step 6 is this non-zero integer q w.h.p. To see it notice that from Claim 4.3 $\gcd(\beta)=1$ w.p. $1-\exp(-\Theta(p))=1-O\left(\frac{1}{np}\right)$ and therefore the g computed in Step 6 satisfies w.p. $1-O\left(\frac{1}{np}\right)$,

$$g=\gcd(\hat{z}_{n+1:n+p})=\gcd(q\beta)=q\gcd(\beta)=q.$$

Hence we obtain w.p. $1-O\left(\frac{1}{np}\right)$.

$$\hat{z}_{n+1:n+p}=g\beta=g\left(\beta^*+Z\right)$$

or w.p. $1-O\left(\frac{1}{np}\right)$

$$\beta^*=\frac{1}{g}\hat{z}_{n+1:n+p}-Z,$$

which implies based on Step 7 and the fact that $g=q\neq 0$ that indeed the output of the algorithm is β^* w.p. $1-O\left(\frac{1}{np}\right)$. The proof of Theorem 2.1 is complete. □

4.2 Proof of Theorem 2.4.A

Proof. We first analyze the algorithm with respect to time complexity. It suffices to analyze step 2 as step 1 runs clearly in polynomial time N,n,p. Step 2 runs the ELO algorithm. From Theorem 2.1 we obtain that the ELO algorithm terminates in polynomial time in $n,p,N,\log\left(\hat{Q}\hat{R}\right),\log\left(2\hat{Q}\left(2^N\hat{W}+\hat{R}p\right)\right)$. As the last quantity is indeed polynomial in the parameters $n,p,N,\log\hat{R},\log\hat{Q},\log\hat{W}$; we are done.

Now we prove that $\hat{\beta^*} = \beta^*$, w.p. $1 - O\left(\frac{1}{np}\right)$. Notice that it suffices to show that the output of Step 3 of the LBR algorithm is exactly $\hat{Q}\beta^*$, as then step 4 gives $\hat{\beta^*} = \frac{Q\beta^*}{Q} = \beta^*$ w.p. $1 - O\left(\frac{1}{np}\right)$.

We first establish that

$$2^N \hat{Q} Y_N = 2^N X_N \hat{Q}\beta^* + W_0 \tag{15}$$

for some $W_0 \in \mathbb{Z}^n$ with $\|W_0\|_\infty + 1 \leq 2\hat{Q}\left(2^N\sigma + Rp\right)$. We have $Y = X\beta^* + W$, with $\|W\|_\infty \leq \sigma$. From the way Y_N is defined, $\|Y - Y_N\|_\infty \leq 2^{-N}$. Hence for $W' = W + Y_N - Y$ which satisfies $\|W'\|_\infty \leq 2^{-N} + \sigma$ we obtain

$$Y_N = X\beta^* + W'.$$

Similarly since $\|X - X_N\|_\infty \leq 2^{-N}$ and $\|\beta^*\|_\infty \leq R$ we obtain $\|\left(X - X_N\right)\beta^*\|_\infty \leq 2^{-N}Rp$, and therefore for $W'' = W' + \left(X - X_N\right)\beta^*$ which satisfies $\|W''\|_\infty \leq 2^{-N} + \sigma + 2^{-N}rp$ we obtain,

$$Y_N = X_N\beta^* + W''$$

or equivalently

$$2^N Y_N = 2^N X_N \beta^* + W''',$$

where $W''' := 2^N W''$ which satisfies $\|W'''\|_\infty \leq 1 + 2^N\sigma + Rp$. Multiplying with $\hat{Q}$ we obtain

$$2^N \hat{Q} Y_N = 2^N X_N \left(\hat{Q}\beta^*\right) + W_0,$$

where $W_0 := \hat{Q}W'''$ which satisfies $\|W_0\|_\infty \leq \hat{Q}\left(1 + 2^N\sigma + Rp\right) \leq 2\hat{Q}\left(2^N\sigma + Rp\right) - 1$. This establishes equation (15).

We now apply Theorem 2.1 for Y our vector $\hat{Q}2^N Y_N$, X our vector $2^N X_N$, β^* our vector $\hat{Q}\beta^*$, W our vector W_0, R our $\hat{Q}R$, $\hat{R}$ our $\hat{Q}\hat{R}$, $\hat{W}$ our quantity $2\hat{Q}\left(2^N\sigma + Rp\right)$ and finally N our truncation level N.

We fist check the assumption (1), (2), (3) of Theorem 2.1. We start with assumption (1). From the definition of X_N we have that $2^N X_N \in \mathbb{Z}^{n\times p}$ and that for all $i \in [n], j \in [p]$,

$$|(2^N X_N)_{ij}| \leq 2^N |X_{ij}|.$$

Therefore for $C = \mathbb{E}[|X_{1,1}|] < \infty$ and arbitrary $i \in [n], j \in [p]$,

$$\mathbb{E}[|(2^N X_N)_{ij}|] \leq 2^N \mathbb{E}[|X_{ij}|] = C2^N,$$

as we wanted. Furthermore, if f is the density function of the distribution $\mathcal{D}$ of the entries of X, recall $\|f\|_\infty \leq c$, by our hypothesis. Now observe for arbitrary $i \in [n], j \in [p]$,

$$\mathbb{P}\left((2^N X_N)_{ij} = k\right) = \mathbb{P}\left(\frac{k}{2^N} \leq X_{ij} \leq \frac{k+1}{2^N}\right) = \int_{\frac{k}{2^N}}^{\frac{k+1}{2^N}} f(u)du \leq \|f\|_\infty \int_{\frac{k}{2^N}}^{\frac{k+1}{2^N}} du \leq \frac{c}{2^N}.$$

This completes the proof that $2^N X_N$ satisfies assumption (1) of Theorem 2.1. For assumption (2), notice that $\hat{Q}\beta^*$ is integer valued, as $\hat{Q}$ is assumed to be a mutliple of Q and β^* satisfies Q-rationality. Furthermore clearly

$$\|\hat{Q}\beta^*\|_\infty \leq \hat{Q}R.$$

For the noise level we have by (15) $W_0 = 2^N \hat{Q} Y_N - 2^N X_N \hat{Q}\beta^*$ and therefore $W_0 \in \mathbb{Z}^n$ as all the quantities $2^N \hat{Q} Y_N$, $2^N X_N$ and $\hat{Q}\beta^*$ are integer-valued. Finally, Assumption (3) follows exactly from equation (15).

Now we check the parameters assumptions of Theorem 2.1. We clearly have

$$\hat{Q}R \leq \hat{Q}\hat{R}$$

and

$$\|W\|_\infty \leq 2\hat{Q}\left(2^N\sigma + Rp\right) = \hat{W}.$$

The last step consists of establishing the relation (2) of Theorem 2.4.A. Plugging in our parameter choice it suffices to prove

$$N > \frac{(2n+p)}{2}\left(2n + p + 10\log\left(\hat{Q}\hat{R}\sqrt{p} + 2\hat{Q}\left(2^N\sigma + Rp\right)\sqrt{n}\right)\right) + 6n\log((1+c)\,np).$$

Using that $\hat{Q}R\sqrt{p} \leq \hat{Q}\left(2^N\sigma + \hat{R}p\right)\sqrt{n}$ and $R \leq \hat{R}$ it suffices to show after elementary algebraic manipulations that

$$N > \frac{(2n+p)}{2}\left(2n + p + 10\log 3 + 10\log\hat{Q} + 10\log\left(2^N\sigma + \hat{R}p\right) + 5\log n\right) + 6n\log((1+c)\,np).$$

Using now that by elementary considerations

$$\frac{(2n+p)}{2}\left(10\log 3 + 5\log n\right) + 4n\log((1+c)\,np) < \frac{(2n+p)}{2}[20\log(3\,(1+c)\,np)] \text{ for all } n \in \mathbb{Z}^+,$$

it suffices to show

$$N > \frac{(2n+p)}{2}\left(2n + p + 10\log\hat{Q} + 10\log\left(2^N\sigma + \hat{R}p\right) + 20\log(3\,(1+c)\,np)\right),$$

which is exactly assumption (3).

Hence, the proof that we can apply Theorem 2.1 is complete. Applying it we conclude that w.p. $1 - O\left(\frac{1}{np}\right)$ the output of LBR algorithm at step 3 is $\hat{Q}\beta^*$, as we wanted. □

4.3 Proof of Theorem 2.4.B

By using a standard union bound and Markov inequality we have

$$\mathbb{P}\left(\|W\|_\infty \leq \sqrt{np}\sigma\right) \geq 1 - \sum_{i=1}^{n}\mathbb{P}\left(|W_i| > \sqrt{np}\sigma\right) \geq 1 - n\frac{\mathbb{E}\left[W_1^2\right]}{np\sigma^2} \geq 1 - \frac{1}{p}.$$

Therefore, conditional on the high probability event $\|W\|_\infty \leq \sqrt{np}\sigma$, we can apply Theorem 2.4.A with $\sqrt{np}\sigma$ instead of σ and conclude the result.

4.4 Proof of Theorem 2.8

Proof. We begin by proving that for any fixed i, all integer relations for the vector $(Y_i, a_1, \ldots, a_{\mathcal{R}})$ are contained in a one-dimensional discrete set. Rational independence results along this spirit are essential to several of our results.

Lemma 4.5. *Let $Y = \langle X, \beta^* \rangle$ where $X \in \mathbb{Z}^p$ and $\beta^* \in \mathbb{R}^p$, such that, $\beta_i \in \mathcal{S} = \{a_1, \ldots, a_{\mathcal{R}}\}$ a rationally independent set. Let $t = (t_0, t_1, \ldots, t_{\mathcal{R}})$ be an integer relation for the vector, $(Y, a_1, \ldots, a_{\mathcal{R}})$. Then, $t \in H$, where*

$$H = \{k(-1, \theta_1^*, \ldots, \theta_{\mathcal{R}}^*) : k \in \mathbb{Z} \setminus \{0\}\},$$

with $\theta_i^* = \sum_{j:\beta_j^* = a_i} X_j$.

Proof. Observe that, Y is an integer combination of $\mathcal{S}$, concretely, $Y = \sum_{i=1}^{\mathcal{R}} \theta_i^* a_i$. Now, let $(t_0, \ldots, t_{\mathcal{R}})$ be an arbitrary relation for the vector $(Y, a_1, \ldots, a_{\mathcal{R}})$, namely, $t_0 Y + \sum_{i=1}^{\mathcal{R}} t_i Y_i = 0$. Observe that, $t_0 \neq 0$, as otherwise we would have obtained that there is a non-trivial integer relation among $a_1, \ldots, a_{\mathcal{R}}$, which is a contradiction. Now, with this, and $Y = \sum_{i=1}^{\mathcal{R}} \theta_i^* a_i$, we have, $\sum_{i=1}^{\mathcal{R}} (t_i + t_0 \theta_i^*) a_i = 0$. Since the set $\mathcal{S}$ is rationally independent, this means the only integer relation among its elements is the trivial one, hence, $t_i + t_0 \theta_i^* = 0$, for every i. This means,

$$t = (t_0, \ldots, t_{\mathcal{R}}) = -t_0(-1, \theta_1^*, \ldots, \theta_{\mathcal{R}}^*) \in H,$$

as claimed. □

We now return to the proof of Theorem 2.8. For each $i \in [n]$, define the coefficients $\theta_{ij}^* = \sum_{k:\beta_k^* = a_j} X_{ik}$, such that:

$$Y_i = \sum_{k=1}^{p} X_{ik} \beta_k^* = \sum_{j=1}^{\mathcal{R}} \theta_{ij}^* a_j.$$

Lemma 4.5 implies that, for each one of the n runs of the integer relation algorithm with inputs $(Y_i, a_1, \ldots, a_{\mathcal{R}})$ with $i \in [n]$, the output of the integer relation detection algorithm is guaranteed to be a multiple of the relation, $(-1, \theta_{i1}^*, \ldots, \theta_{i\mathcal{R}}^*)$, and the corresponding multiplicity can be read off from the first component. Now, recalling Theorem 2.8, we have that the termination time for a single run of the integer relation detection algorithm is $O(\mathcal{R}^3 + \mathcal{R}^2 \log \|\mathbf{m}\|)$, where $\|\mathbf{m}\|$ is the norm of the smallest non-trivial integer relation $\mathbf{m} = (-1, \theta_{i1}^*, \ldots, \theta_{i\mathcal{R}}^*)$. In order to upper bound this quantity, we will first upper bound θ_{ik}^*. Let $X_i \in \mathbb{Z}^{1 \times p}$ be the i^{th} row of X. Notice that, for each $k \in [\mathcal{R}]$, $\theta_{ik}^* = \sum_{j:\beta_j^* = a_k} X_{ij} \implies |\theta_{ik}^*| \leq p \|X\|_\infty$. Now, by a union bound and Markov's inequality,

$$\mathbb{P}(\|X\|_\infty > p^2 n 2^N) \leq pn \mathbb{P}(|X_{11}| > p^2 n 2^N) \leq pn \frac{\mathbb{E}[|X_{11}|]}{p^2 n 2^N} \leq pn \frac{C 2^N}{p^2 n 2^N} = O(\frac{1}{p}).$$

In particular, for all $i \in [n]$ and $j \in [\mathcal{R}]$, $|\theta_{ij}^*| \leq p^3 n 2^N$ with probability at least $1 - O(1/p)$. Therefore, the norm of the smallest relation obeys,

$$\|\mathbf{m}\| \leq \sqrt{1 + \sum_{k=1}^{\mathcal{R}} (\theta_{ik}^*)^2} \leq O(p^3 (\mathcal{R} n)^{1/2} 2^N),$$

hence, with probability at least $1 - O(1/p)$, the vector $\mathbf{m}$ is such that, $\log \|\mathbf{m}\|$ is at most polynomial in $N, p, \mathcal{R}$, and n. Note also that, we make n calls to the integer relation oracle, each taking polynomial in $p, \mathcal{R}, N, n$ many calls. Therefore, we conclude that the overall run time of the integer relation detection step is polynomial in $p, \mathcal{R}, N$ and n, with probability at least $1 - O(1/p)$.

Now, we study the second half of the algorithm, where we make calls to the LLL lattice basis reduction oracle. Fix a $j \in [\mathcal{R}]$. We now show how to recover the entries of β^* which are equal to a_j. Suppose that $e^{(j)}$ is a binary vector with $e^{(j)}_k = 1$ if and only if $\beta^*_k = a_j$, and $e^{(j)}_k = 0$ otherwise, for $1 \leq k \leq p$. Observe that, $\Theta_j = Xe^{(j)}$. Hence, this problem is essentially a subset-sum problem, where we have access to n linear measurements of the hidden binary vector $e^{(j)}$, through the mechanism, $\langle X_i, e^{(j)} \rangle$, where X_i is the i^{th} row of X.

Consider now the $(n+p)$-dimensional integer lattice Λ_j, generated by the columns of the matrix A^j, which we recall

$$A^j = \begin{bmatrix} m\mathrm{Diag}_{n\times n}(\Theta_j) & -mX_{n\times p} \\ 0_{p\times n} & I_{p\times p} \end{bmatrix},$$

with $m = p2^{\lceil \frac{n+p}{2} \rceil}$. Observe that,

$$A_j \begin{bmatrix} 1_{n\times 1} \\ e^{(j)}_{p\times 1} \end{bmatrix} = \begin{bmatrix} 0_{n\times 1} \\ e^{(j)}_{p\times 1} \end{bmatrix} \implies \begin{bmatrix} 0_{n\times 1} \\ e^{(j)}_{p\times 1} \end{bmatrix} \in \Lambda_j.$$

Hence, we have $\min_{z\in\Lambda_m, z\neq 0} \|z\| \leq \sqrt{p}$. Therefore, the LLL lattice basis reduction algorithm, when called on A_m, returns a vector $\widehat{x}$ such that, $\|\widehat{x}\| \leq \sqrt{p}2^{\frac{n+p}{2}} \triangleq m_0$. Note that $m_0 < m$.

We now claim the following. Essentially all 'short' vectors of this lattice satisfy the condition that their $(n+1) : (n+p)$ coordinates are multiples of the hidden vector. This is the subject of the next lemma.

Lemma 4.6. *Define the set $\mathcal{F}^j$ via,*

$$\mathcal{F}^j = \{x \in \Lambda_j : \|x\| \leq m_0, x_{n+1:n+p} \neq ke^{(j)}, \forall k\}.$$

Then, $\mathbb{P}(\mathcal{F} \neq \varnothing) = o(1)$, where the probability is taken with respect to the randomness in X.

Proof. Note that, $x \in \mathcal{F}^j$ implies existence of $x_1 \in \mathbb{Z}^{n\times 1}$ and $x_2 \in \mathbb{Z}^{p\times 1}$ such that,

$$x = A_j \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} m(\mathrm{diag}(\Theta_j)x_1 - Xx_2) \\ x_2 \end{bmatrix}.$$

Since $\|x\| \leq m_0$, we immediately obtain that $|(x_2)_i| \leq m_0$, for each $i \in [p]$.

Next, notice that, for every $i \in [n]$, $\theta^*_{ij}(x_1)_i = \langle X_i, x_2 \rangle$. Indeed, if this is not the case for some $i_0 \in [n]$, then, x_{i_0} is a non-zero integer, divisible by m, and therefore, we have a contradiction through the following chain of inequalities:

$$m_0 \geq \|x\| \geq |x_{i_0}| \geq m > m_0.$$

For any fixed i, and fixed x_1 and x_2, the probability that $\theta^*_{ij}(x_1)_i = \langle X_i, x_2 \rangle$ is equal to, $\mathbb{P}\left(\sum_{j=1}^p X_{ij}e^{(j)}_i(x_1)_i = \sum_{j=1}^p X_{ij}(x_2)_j\right)$, which is upper bounded by $c/2^N$. To see the last deduction, note that since $x \in \mathcal{F}^j$, there is an index $n_0 \in \{n+1, \ldots, n+p\}$ such that, $(x_2)_{n_0-n} \neq$

$e^{(j)}_{n_0}(x_1)_i$ (note that $x_2 \in \mathbb{Z}^p$ with coordinates $1, \ldots, p$), and therefore, the probability above can be expressed as an event, involving X_{n_0}, which by conditioning on the rest is found to be upper bounded by $c/2^N$. Now, using the independence, the probability that $\theta^*_{ij}(x_1)_i = \langle X_i, x_2 \rangle$ for every i is $c^n 2^{-nN}$.

We now show that, with probability at least $1 - O(1/p^2)$, $|\theta^*_{ij}| \geq \frac{2^N}{cnp^2}$, for every $i \in [n]$. Indeed, using a union bound:

$$\mathbb{P}\left(\bigcup_{i=1}^{n}\{|\theta^*_{ij}| < \frac{2^N}{cnp^2}\}\right) \leq n\mathbb{P}(|\theta^*_{ij} < \frac{2^N}{cnp^2}|) = n \sum_{k \in \mathbb{Z} \cap [-\frac{2^N}{cnp^2}, \frac{2^N}{cnp^2}]} \mathbb{P}\left(\theta^*_{ij} = k\right) \leq \frac{2n2^N}{cnp^2}\frac{c}{2^N} = O(\frac{1}{p^2}),$$

since the probability that θ^*_{ij} takes a specific value is at most $c/2^N$. We now claim, $\|X\|_\infty \leq np^2 2^N$ with probability at least $1 - O(1/p)$, where $\|X\|_\infty = \max_{i,j} |X_{ij}|$. To see this, observe that using a union bound and the Markov's inequality, we have:

$$\mathbb{P}(\|X\|_\infty > np^2 2^N) \leq pn\mathbb{P}(|X_{11}| > np^2 2^N) = pn\frac{\mathbb{E}[|X_{11}|]}{np^2 2^N} = O(\frac{1}{p}),$$

since $\mathbb{E}[|X_{11}|] \leq C2^N$ for some constant $C > 0$. Now, recalling the facts $\theta^*_{ij}(x_1)_i = \langle X_i, x_2 \rangle$ with high probability and $|(x_2)_i| \leq m_0$ for each $i \in [p]$; together with the (whp) bounds on $\|X\|_\infty$ and $|\theta^*_{ij}|$, we have the following with high probability:

$$\frac{2^N}{cnp^2}|(x_1)_i| \leq |\theta^*_{ij}(x_1)_i| \leq m_0 p^3 n 2^N \Rightarrow |(x_1)_i| \leq O(m_0 p^5 n^2),$$

for all i. Therefore,

$$\left|\left\{x_1 \in \mathbb{Z}^{n\times 1}, x_2 \in \mathbb{Z}^{p\times 1} : A_j \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathcal{F}\right\}\right| \leq O((2m_0 p^5 n^2 + 1)^n) O((2m_0 + 1)^p).$$

Hence,

$$\mathbb{P}(\mathcal{F} \neq \varnothing) \leq \exp_2\left(n + p + n\log(n^2 p) + \frac{n+p}{2}\log p + \frac{(n+p)^2}{2} - n\log c - nN\right)$$

which is $o(1)$, due to the choice of the parameters, where $\exp_2(\gamma) = 2^\gamma$. □

With the Lemma 4.6 at our disposal, we continue with the proof of Theorem 2.8. Note that due to the lemma, the LLL algorithm, when run on the columns of the matrix A_j, is guaranteed to output a multiple of the vector $e^{(j)}$; thus indicating the indices i such that $\beta^*_i = a_j$. $\mathcal{R}-$executions of this algorithm allow us to recover β^* completely. The runtime of the LLL algorithm is polynomial in n, p and $\log \|A_m\|_\infty$, and therefore, is polynomial in $n, N, p, \mathcal{R}$, keeping in mind that we make $\mathcal{R}$ calls to LLL oracle.

This, together with the previous discussion on the runtime of the integer relation detection step establishes that the overall runtime of the procedure is at most polynomial in $p, n, N, \mathcal{R}$. □

4.5 Proof of Theorem 2.9

Proof. Define the set $\mathcal{L}$ by,

$$\mathcal{L} = \{X_i a_j : 1 \leq i \leq p, 1 \leq j \leq \mathcal{R}\}.$$

Lemma 4.7. *Let X_i's be jointly continuous random variables, drawn from an arbitrary continuous distribution $\mathcal{D}$; and $a_1, \ldots, a_{\mathcal{R}}$ be rationally independent real numbers. Then,*

$$\mathbb{P}(\mathcal{L} \text{ is rationally independent}) = 1.$$

Proof. Note that, the event $\mathcal{L}$ is rationally dependent is precisely equivalent to existence of a collection $\mathbf{v} \in \mathbb{Q}^{p\mathcal{R}} \setminus \{\mathbf{0}\}$, indexed by $v_i^{(j)}$ with $i \in [p], j \in [\mathcal{R}]$, such that,

$$\sum_{j=1}^{\mathcal{R}} a_j \left(\sum_{i=1}^{p} v_i^{(j)} X_i \right) = 0 \iff 0 = \sum_{i=1}^{p} X_i \underbrace{\left(\sum_{j=1}^{\mathcal{R}} a_j v_i^{(j)} \right)}_{\triangleq \gamma_i^{\mathbf{v}}} = \sum_{i=1}^{p} X_i \gamma_i^{\mathbf{v}} = \langle X, \Gamma_{\mathbf{v}} \rangle.$$

where $\Gamma_{\mathbf{v}} = \begin{bmatrix} \gamma_1^{\mathbf{v}} & \gamma_2^{\mathbf{v}} & \cdots & \gamma_p^{\mathbf{v}} \end{bmatrix}$. Recall that $\mathbf{v}$ is not identically zero vector. Thus, there exists a $v_{i_0}^{(j)}$ which is non-zero. In particular, using the fact that $a_1, \ldots, a_{\mathcal{R}}$ are rationally independent, we have $\gamma_{i_0}^{\mathbf{v}} \neq 0$, hence, $\Gamma_{\mathbf{v}} \neq 0$. For each fixed $\mathbf{v} \in \mathbb{Q}^{p\mathcal{R}} \setminus \mathbf{0}$, we have $\mathbb{P}(\langle X, \Gamma_{\mathbf{v}} \rangle = 0) = 0$, as X is a jointly continuous random vector, and $\Gamma_{\mathbf{v}} \neq \mathbf{0}$. Thus, we establish, via countable additivity that

$$\mathbb{P}(\mathcal{L} \text{ is rationally dependent}) \leq \sum_{\mathbf{v} \in \mathbb{Q}^{p\mathcal{R}} \setminus \{\mathbf{0}\}} \mathbb{P}(\langle X, \Gamma_{\mathbf{v}} \rangle = 0) = 0.$$

□

With this, we now continue with the proof of Theorem 2.9. Observe first that, $X_i \beta_i^* = \sum_{k=1}^{\mathcal{R}} \xi_k^{(i)} (X_i a_k)$, where $\xi^{(i)} \in \{0,1\}^{\mathcal{R}}$ is a binary vector, with all but one of its components equal to 0, and the only component that is one is precisely the $k \in [\mathcal{R}]$ such that $\beta_i^* = a_k$. With this, we deduce that, $Y = \sum_{i=1}^{p} \sum_{j=1}^{\mathcal{R}} X_i a_j e_{ij}$, where, $e_{ij} \in \{0,1\}$, and $e_{ij} = 1$ if and only if $\beta_i^* = a_j$. In particular, there exists a non-trivial integer relation between Y and the elements of $\mathcal{L}$. Now, let $(t_0, t_{ij} : i \in [p], j \in [\mathcal{R}])$ be any such relation, with $t_0 Y + \sum_{i=1}^{p} \sum_{j=1}^{\mathcal{R}} X_i a_j t_{ij} = 0$. Clearly, $t_0 \neq 0$ with probability 1; since otherwise we obtain a contradiction with the rational independence of $\mathcal{L}$ established in Lemma 4.7. We have $\sum_{i=1}^{p} \sum_{j=1}^{\mathcal{R}} X_i a_j (t_0 e_{ij} + t_{ij}) = 0$. From here, we proceed exactly in the same way as in the proof of Lemma 4.5, and establish that $t_{ij} = -t_0 e_{ij}$ for every $i \in [p], j \in [\mathcal{R}]$, as a consequence of Lemma 4.7. Namely, all such relations are again contained in a one-dimensional discrete set, spanned by the true relation. Now, letting the error event to be $\mathcal{E} = \{\widehat{\beta^*} \neq \beta^*\}$, and defininining $\mathcal{E}' = \{\mathcal{L} \text{ is rationally independent}\}$, we have, $\mathbb{P}(\mathcal{E}) = \mathbb{P}(\mathcal{E}|\mathcal{E}')\mathbb{P}(\mathcal{E}') + \mathbb{P}(\mathcal{E}|(\mathcal{E}')^c)\mathbb{P}((\mathcal{E}')^c) = 0$, since $\mathbb{P}(\mathcal{E}|\mathcal{E}') = 0$ and $\mathbb{P}((\mathcal{E}')^c) = 0$. Hence, with probability 1, the internal output $\mathbf{c}$ of the IHDR algorithm coincide with the true relation, from which, one can immediately obtain the value of β_i^*.

Note that, the set $\mathcal{L}$ is generated in $O(p\mathcal{R})$ time, which is polynomial in both p and $\mathcal{R}$. Finally, the norm $\|\mathbf{m}\|$ of the smallest relation, $\mathbf{m}$, between Y and the elements of $\mathcal{L}$ is at most $O(\sqrt{p\mathcal{R}})$, hence, the integer relation detection algorithm runs in time $O((p\mathcal{R})^3 + (p\mathcal{R}^2)\log(p\mathcal{R}))$, which is polynomial in p and $\mathcal{R}$. With this, we conclude the proof. □

Remark

A natural question is why this scheme cannot be used to for recovery, in the case of integer-valued X. The reason is as follows. Observe that in this case, $\mathcal{L}$ is no longer rationally independent: for any realization $X_1, \ldots, X_p$ and any $a_1, \ldots, a_{\mathcal{R}}$, one can simply consider $-X_2(X_1 a_1) + X_1(X_2 a_2) + 0 \cdot (\text{the other terms}) = 0$. Hence, $(-X_2, X_1, 0, \ldots, 0)$ is a non-trivial relation, for the set $\mathcal{L}$.

4.6 Proof of Theorem 2.10

We first multiply everywhere by $\widehat{Q}$, and arrive at,

$$\widehat{Q}Y_i = \sum_{j=1}^{p} \widehat{Q}X_{ij}\beta_j^* = \sum_{j=1}^{\mathcal{R}} \theta_{ij}^* a_j + \sum_{j:\beta_j^* \in \mathbb{Q}} \widehat{Q}X_{ij}\beta_j^* \quad \forall i \in [n],$$

with $\theta_{ij}^* = \widehat{Q} \sum_{k:\beta_k^* = a_j} X_{ik}$. Note also that $\sum_{j:\beta_j^* \in \mathbb{Q}} \widehat{Q}X_{ij}\beta_j^* \in \mathbb{Z}$, as Q divides $\widehat{Q}$, and whenever β_j^* is rational, it is also Q-rational.

Next, one can transfer the proof of Lemma 4.5 to obtain that, the only integer relations for the vector, $(\widehat{Q}Y_i, a_1, \ldots, a_{\mathcal{R}}, 1)$ are those, contained in one-dimensional discrete set $\mathcal{H}$, where

$$\mathcal{H} = \{k(-1, \theta_{i1}^*, \ldots, \theta_{i\mathcal{R}}^*, \sum_{j:\beta_j^* \in \mathbb{Q}} \widehat{Q}X_{ij}\beta_j^*) : k \in \mathbb{Z}^*\}.$$

Thus, the integer relations found by the JIRSS step are, indeed, multiples of the true relation, for each i. The fact that integer relation algorithm runs in time polynomial in $n, \mathcal{R}, N, \log \|W\|_\infty$, and $\log \widehat{Q}$, follows immediately, by inspecting and adapting the corresponding lines, in the proof of the Theorem 2.8.

Next, we decompose β^* according to, $\beta^* = \beta_I^* + \beta_R^*$, where,

$$(\beta_I^*)_i = \begin{cases} \beta_i^*, & \text{if } (\beta_I^*)_i \notin \mathbb{Q} \\ 0, & \text{otherwise,} \end{cases}$$

and,

$$(\beta_R^*)_i = \begin{cases} \beta_i^*, & \text{if } (\beta_R^*)_i \in \mathbb{Q} \\ 0, & \text{otherwise.} \end{cases}$$

Note that, $Y = X\beta^* + W = X\beta_I^* + X\beta_R^* + W$. Under the given specifications on its parameters; the output $\widehat{\beta_1^*}$ of the JIRSS step, per Theorem 2.8, satisfies $\widehat{\beta_1^*} = \beta_I^*$ with high probability. In this case, the condition of Theorem 2.8 is satisfied. Hence, $\widetilde{Y} = Y - X\widehat{\beta_1^*}$, satisfies, with high probability, $\widetilde{Y} = X\beta_R^* + W$, under given parameter specifications.

Now, let $\widetilde{s}$ be the number of 0 entries in $\widehat{\beta_1^*}$. Notice that, with high probability, $\widetilde{s} = s$, where $s = |\{i \in [p] : \beta_i^* \in \mathbb{Q}\}|$. We now let $\widetilde{\beta}$ to be the vector, obtained by erasing β_i^*, if $\widehat{\beta_1^*} = 0$. The corresponding matrix, with erased columns, is $\widetilde{X} \in \mathbb{Z}^{n \times \widetilde{s}}$. Now, the problem of recovering the remainder of $\widehat{\beta^*}$ is simply the problem of inferring $\widetilde{\beta}$, from $\widetilde{Y} = \widetilde{X}\widetilde{\beta} + W$, where $\widetilde{\beta} \in \mathbb{Q}^s$, consists of Q-rational numbers. The algorithm 1, per Theorem 2.1, with high probability, recovers $\widetilde{\beta}$, provided that the parameters involved obey the condition (5).

The overall polynomial run-time guarantee follows from the corresponding run time guarantees of the Algorithm 3 per Theorem 2.8 and Algorithm 1 per Theorem 2.1.

4.7 Proof of Theorem 2.11

Fix an $i \in [n]$, and express $\widehat{Q}Y_i = \sum_{j=1}^{p} \widehat{Q}X_{ij}\beta_j^*$ in the following way.

$$\widehat{Q}Y_i = \sum_{j=1}^{p}\sum_{k=1}^{\mathcal{R}} X_{ij}a_k(\widehat{Q}e_{jk}) + \sum_{j:\beta_j^*\in\mathbb{Q}} \widehat{Q}X_{ij}\beta_j^*,$$

where $e_{jk} \in \{0,1\}$ and $e_{jk} = 1$ if and only if $\beta_j^* = a_k$, and is 0, otherwise. This is nothing but a decomposition of $\widehat{Q}Y_i$, as an integral combination of the elements of the set S_i, where

$$S_i = \{X_{ij}a_k : j \in [p], k \in [\mathcal{R}]\} \cup \{X_{ij} : j \in [p]\}.$$

Note that, $|S_i| = p + p\mathcal{R} = O(p\mathcal{R})$, hence it can be generated in polynomial in p and $\mathcal{R}$ time. We now establish that, with probability 1, the elements of S_i are rationally independent. This is proved along the same lines as in Lemma 4.7. In order to see this, define the event, $E_i = \{S_i \text{ is rationally independent}\}$. Note that, E_i^c implies existence of a collection of rationals, not simultaneously zero and indexed for convenience by $\{q_{jk}, r_j \in \mathbb{Q} : j \in [p], k \in [\mathcal{R}]\}$, such that,

$$\sum_{j=1}^{p}\sum_{k=1}^{\mathcal{R}} X_{ij}a_k q_{jk} + \sum_{j=1}^{p} X_{ij}r_j = \sum_{j=1}^{p} X_{ij}\left(\sum_{k=1}^{\mathcal{R}} q_{jk}a_k + r_j\right) = \sum_{j=1}^{p} X_{ij}\gamma_j = 0,$$

where $\gamma_j = \sum_{k=1}^{\mathcal{R}} q_{jk}a_k + r_j$. Now, using the fact that $a_1, \ldots, a_{\mathcal{R}}, 1$ are rationally independent, and q_{jk}, r_k are not all 0, we deduce that, there exists a j_0, such that, $\gamma_{j_0} \neq 0$. Now, using a union bound,

$$\mathbb{P}(E_i^c) \leq \sum_{q_{jk}, r_j \in \mathbb{Q}} \mathbb{P}\left(\sum_{j=1}^{p} X_{ij}\gamma_j = 0\right) = 0,$$

since, for a fixed i_0, the probability of the event, $\left\{X_{i_0} = -\frac{1}{\gamma_{i_0}}\sum_{j\neq i_0} X_{ij}\gamma_j\right\}$ is 0, which is obtained first by conditioning on all random variables except X_{i_0}, and recalling that X_i's follow a continuous distribution.

Using this fact, and proceeding in exact same way, as in the proof of Lemma 4.5, we obtain that integer relations for the vector consisting of $\widehat{Q}Y_i$, and the elements of S_i are those, contained in $\mathcal{H}$, where

$$\mathcal{H} = \{k(-1, e_{jk}, s_j) : k \in \mathbb{Z}^*\},$$

with $e_{jk} = \widehat{Q}e_{jk}$, and $s_j = \widehat{Q}\beta_j^*$, if $\beta_j^* \in \mathbb{Q}$, and is 0, otherwise. In particular, with probability 1, $\widehat{\beta_1^*}$ indeed clashes with the irrational entries of β^*.

In the remaining step, we construct a modified observation model. For doing so, we take away the irrational entries of β^*; and recall a consequence of Theorem 2.4.A (with $W = 0$) that the output of the LBR algorithm coincides with the rational part of β^*, with high probability.

We will now conclude the proof by examining the overall runtime. First, note that, the set S_i can be generated in $O(p\mathcal{R})$ time, which is polynomial in p and $\mathcal{R}$. Next, we know that, the integer relation detection algorithm with k inputs $(x_1, \ldots, x_k)$ recovers a relation in a time $O(k^3 + k^2 \log \|\mathbf{m}\|)$, where $\mathbf{m}$ is a relation for the vector, $(x_1, \ldots, x_k)$ with the smallest $\|\mathbf{m}\|$. In our case, the input set is of cardinality $|S_i| + 1 = O(p\mathcal{R})$, and the smallest relation

is $(-1, e_{jk}, s_j : j \in [p], k \in [\mathcal{R}])$, which satisfies $\|m\| \leq \sqrt{\widehat{Q}^2 p\mathcal{R} + p\widehat{Q}^2\widetilde{R}^2}$, and thus, the overall runtime is $O(p^3\mathcal{R}^3 + p^2\mathcal{R}^2 \log(\widehat{Q} + \log p + \log(\mathcal{R}\widetilde{R})))$, that is at most polynomial in $p, \mathcal{R}$, and $\log \widehat{Q}$. Finally, the LLL algorithm runs in time polynomial in $n, s, N, \log \widehat{R}, \log \widehat{Q}$, and therefore, the overall runtime is indeed polynomial in $n, p, N, s, \log \widehat{R}, \log \widehat{Q}$, and $\mathcal{R}$, the cardinality of the irrational part of the set that the entries of β^* take values in.

4.8 Proof of Theorem 2.12

Proof. Quite analogous to the previous proof, we first note that,

$$\widehat{Q}Y = \sum_{j:\beta_j^* \in \mathbb{Q}} X_j(\widehat{Q}\beta_j^*) + \sum_{i=1}^{p}\sum_{j=1}^{\mathcal{R}} X_i a_j(\widehat{Q}e_{ij}),$$

where $e_{ij} = 1$ if and only if $\beta_i^* = a_j$, and is zero, otherwise. Observe that, for any j with $\beta_j^* \in \mathbb{Q}$, the Q-rationality assumption, together with the fact that $\widehat{Q}$ is divisible by Q yield that $\widehat{Q}\beta_j^* \in \mathbb{Z}$. In particular, there is a an integer relation for the vector,

$$(\widehat{Q}Y, X_i a_j, X_i : i \in [p], j \in [\mathcal{R}]).$$

Namely, $\widehat{Q}Y = \sum_{i=1}^{p} X_i\theta_i^* + \sum_{i=1}^{p}\sum_{j=1}^{\mathcal{R}} X_i a_j(\widehat{Q}e_{ij})$ with $\theta_i^* = 0$ if $\beta_i^* \notin \mathbb{Q}$, and $\theta_i^* = \widehat{Q}\beta_i^*$ if $\beta_i^* \in \mathbb{Q}$. We now claim, similar to the previous results, that the set, $\{X_i a_j : i \in [p], j \in [\mathcal{R}]\} \cup \{X_i : i \in [p]\}$ is rationally independent, with probability one. To see this, let $(r_{ij}, q_i : i \in [p], j \in [\mathcal{R}]) \in \mathbb{Q}^{p\mathcal{R}+p}$ be a $p(\mathcal{R}+1)-$tuple of rationals, not all zero. We first establish,

$$\mathbb{P}\left(\sum_{i=1}^{p}\sum_{j=1}^{\mathcal{R}} X_i a_j r_{ij} + \sum_{i=1}^{p} X_i q_i = 0\right) = 0.$$

Let $\gamma_i = \sum_{j=1}^{\mathcal{R}} r_{ij}a_j + q_i$ for $1 \leq i \leq p$. Note that, since $\{a_1, \ldots, a_{\mathcal{R}}, 1\}$ is a rationally independent set by Assumption 1, we then get $\gamma_i = 0 \Rightarrow r_{ij}, q_i = 0$ for any $j \in [\mathcal{R}]$. In particular, if $\gamma_i = 0$ for every i, we then deduce immediately that $r_{ij}, q_i = 0$ for every $i \in [p]$ and $j \in [\mathcal{R}]$, which contradicts with the choice of this tuple. From here, we then get the vector $\Gamma = (\gamma_1, \ldots, \gamma_p)^T \in \mathbb{R}^p$ is not identically zero. But now, $\mathbb{P}(\langle X, \Gamma\rangle = 0) = 0$ immediately, due to joint continuity. A union bound over all such $p(\mathcal{R}+1)-$tuples of rationals then establish the desired claim.

Now, assume in the remainder we condition on this event. Let $(m_0, m_{ij}, n_i : i \in [p], j \in [\mathcal{R}])$ be a (non-zero) integer relation for the vector $(\widehat{Q}Y, X_i a_j, X_i : i \in [p], j \in [\mathcal{R}])$. We then have,

$$0 = m_0\widehat{Q}Y + \sum_{i=1}^{p}\sum_{j=1}^{\mathcal{R}} m_{ij}X_i a_j + \sum_{i=1}^{p} X_i n_i = \sum_{i=1}^{p} X_i(m_0\theta_i^* + n_i) + \sum_{i=1}^{p}\sum_{j=1}^{\mathcal{R}} X_i a_j(m_0\widehat{Q}e_{ij} + m_{ij}).$$

Since the vector, $(X_i a_j, X_i : i \in [p], j \in [\mathcal{R}])$ is rationally independent by conditioning, we then immediately get, $m_{ij} = -m_0\widehat{Q}e_{ij}$ and $n_i = -m_0\theta_i^*$. Namely, IRA indeed recovers $\beta^* \in \mathbb{R}^p$, since by inspecting the relation coefficients, the recover is immediate.

Finally, we study the runtime of the algorithm. Recall that, the smallest such relation is of form $(-1, \theta_i^*, \widehat{Q}e_{ij} : i \in [p], j \in [\mathcal{R}])$. Each coordinate of this vector is clearly upper bounded by $\widehat{Q}\widehat{R}$. From here, we get that the smallest relation has norm which is at most $O(\widehat{Q}\widehat{R}(p\mathcal{R})^{1/2})$. Finally, using Theorem 1.5, we get that the overall runtime is at most $O(p^3\mathcal{R}^3 + p^2\mathcal{R}^2 \log(\widehat{Q}\widehat{R}(p\mathcal{R})^{1/2}))$, which is poly$(p, \mathcal{R}, \log \widehat{Q}, \log \widehat{Q})$, as claimed. □

4.9 Proof of Theorem 2.13

Proof. We first note that, $Y^2 = \sum_{i=1}^{p} X_i^2|\beta_i^*|^2 + \sum_{1\le i<j\le p} X_iX_j((\beta_i^*)^H\beta_j^* + \beta_i^*(\beta_j^*)^H)$. In particular, it is not hard to see that, $Y^2 = \sum_{d=1}^{\mathcal{R}} \theta_d^*|a_d|^2 + \sum_{1\le i<j\le\mathcal{R}} \theta_{ij}^*(a_i^Ha_j + a_ia_j^H)$, for some integers $\theta_1^*,\ldots,\theta_{\mathcal{R}}^* \in \mathbb{Z}$, and $\theta_{ij}^* \in \mathbb{Z}$ for $1 \le i < j \le \mathcal{R}$. In particular, if $\mathbf{t} = (t_0, t_d, t_{ij} : 1 \le d \le \mathcal{R}, 1 \le i < j \le \mathcal{R})$ is an integer relation for the vector, consisting of Y^2, and the entries of $\mathcal{S}' = \{|a_d|^2 : d \in [\mathcal{R}]\} \cup \{a_i^Ha_j + a_ia_j^H : 1 \le i < j \le \mathcal{R}\}$, then one may proceed, in a similar way, as in proof of Lemma 4.5 and establish that, $\mathbf{t} = -t_0(-1, \theta_d^*, \theta_{ij}^* : 1 \le d \le \mathcal{R}, 1 \le i < j \le \mathcal{R}) = -t_0\mathbf{t}'$, namely, all relations contained in a one-dimensional (discrete) set, spanned by the vector, $\mathbf{t}' = (-1, \theta_d^*, \theta_{ij}^* : 1 \le d \le \mathcal{R}, 1 \le i < j \le \mathcal{R}) \in \mathbb{Z}^{\mathcal{R}(\mathcal{R}+1)/2+1}$. In particular, the integer relation algorithm, with inputs Y^2 and the entries of $\mathcal{S}'$ will terminate with $(-1, \theta_d^*, \theta_{ij}^* : 1 \le d \le \mathcal{R}, 1 \le i < j \le \mathcal{R})$.

We now show, how to decode β^*, using $\mathbf{t}'$, and the solution of a certain subset-sum problem. Observe that, X_iX_j contributes to $\theta_{k\ell}^*$, corresponding to $a_k^Ha_\ell + a_ka_\ell^H$, if and only if, $\{\beta_i^*, \beta_j^*\} = \{a_k, a_\ell\}$. With this, we observe that, for every $1 \le k < \ell \le \mathcal{R}$, and $1 \le i < j \le p$, there exists binary variables $\xi_{ij}^{(k,\ell)} \in \{0, 1\}$, such that, the following holds: $\theta_{k\ell}^* = \sum_{1\le i<j\le p} X_iX_j\xi_{ij}^{(k,\ell)}$. Namely, the coefficients, $\theta_{k\ell}^*$ are subset-sums of $\{X_iX_j : 1 \le i < j \le p\}$. The binary variables, $\xi_{ij}^{(k,\ell)}$ can be recovered in polynomial in p and $\mathcal{R}$ many (bit) operations, using the LLL algorithm, and this is isolated as a separate result, in Proposition 2.14. Taking this to be granted, namely, $\xi_{ij}^{(k,\ell)}$ can be recovered, we now show how to use this information to decode β^*. For convenience of notation, assume $\xi_{ij}^{(k,\ell)} = \xi_{ji}^{(k,\ell)}$.

If $|\{d : \theta_d^* \neq 0\}| = 1$, then $\beta_i^* = a_{d_0}$ for every $i \in [p]$, where d_0 is the unique index with $\theta_{d_0}^* \neq 0$. If, $|\{d : \theta_d^* \neq 0\}| = 2$, then there is a pair k, ℓ of indices, such that, $\beta_i^* \in \{a_k, a_\ell\}$, for all $i \in [p]$. Now, we focus on $\theta_{k\ell}^* = \sum_{i<j} X_iX_j\xi_{ij}^{(k,\ell)}$. Note that, if $\beta_i^* = \beta_j^*$, then $\xi_{ij}^{(k,\ell)} = 0$. Start from an arbitrary index i_0. Set $\beta_{i_0}^* = +$. Now, for any index, $i < i_0$, if $\xi_{i_0i}^{(k,\ell)} = 1$, then set $\beta_i^* = -$. Else, set it to $+$. At the end, it is either the case that, all entries labeled with $+$ are a_k, and those labeled with $-$ are a_ℓ; or vice versa. This can be verified, using the information Y^2, clearly, in polynomial time. Finally, suppose that, $|\{d : \theta_d^* \neq 0\}| \ge 3$. Fix an $i \in [p]$, and suppose the goal is to decode β_i^*. Find, in polynomial time, by brute force search, two pairs $(k, \ell) \neq (k', \ell')$ with $1 \le k < \ell \le \mathcal{R}$, and $1 \le k' < \ell' \le \mathcal{R}$; and indices j and j', such that, $\xi_{ij}^{(k,\ell)} = \xi_{ij'}^{(k,\ell')} = 1$. Note that, since β^* consists at least of three different values from the set $\mathcal{S}$, such pairs and indices indeed exist, and that, $|\{k, \ell\} \cap \{k', \ell'\}| = 1$. Observe now that, $\{\beta_i^*, \beta_j^*\} = \{a_k, a_\ell\}$ and $\{\beta_i^*, \beta_{j'}^*\} = \{a_{k'}, a_{\ell'}\}$. Hence, $\beta_i^* = a_{\{k,\ell\}\cap\{k',\ell'\}}$, and found in polynomial in $p, \mathcal{R}$ many operations. Repeating this p times, we finish decoding β^*.

We now show that the overall run-time of this protocol is polynomial. For this, it suffices to show, both LLL and and integer relation detection steps run in polynomial time. The runtime of LLL step will be established in the proof of Proposition 2.14. Now, we focus on the integer relation detection step. First, note that generating $\mathcal{S}'$ takes at most $O(\mathcal{R}^2)$ arithmetic operations on reals. Now, we focus on bounding the norm of $\mathbf{t}'$, which has the smallest $\|\mathbf{t}'\|$. Observe that,

$$\mathbb{P}(\|X\|_\infty > p^22^N) \le p\mathbb{P}(|X_1| > p^22^N) \le p\frac{\mathbb{E}[|X_1|]}{p^22^N} \le O\left(\frac{1}{p}\right),$$

using a union bound and the Markov inequality. Thus, $\|X\|_\infty \le p^22^N$, with probability at least $1 - O(1/p)$. Next, let $\widetilde{X}$ be a vector of dimension $\binom{p}{2}$, consisting of elements, X_iX_j, for

$1 \leq i < j \leq p$. Now,

$$\mathbb{P}(\|Y\|_\infty > p^3 2^{2N}) \leq p^2 \mathbb{P}(|X_1 X_2| > p^3 2^{2N}) \leq p^2 \frac{\mathbb{E}[|X_1 X_2|]}{p^3 2^{2N}} \leq O\left(\frac{1}{p}\right),$$

using a union bound, the Markov inequality, and independence of X_1 and X_2. Hence, with probability at least $1 - O(1/p)$, $\|Y\|_\infty \leq p^3 2^{2N}$. Recalling that, the IRA terminates with (a multiple of) $\mathbf{t}'$ in time at most $O(T^3 + T^2 \log \|\mathbf{t}'\|)$, which is at most polynomial in $\mathcal{R}, N, p$, where $T = O(\mathcal{R}^2)$ is the size of the input to the problem. Finally, the overall run time is polynomial in p and $\mathcal{R}$, since the integer relation detection step runs in time polynomial in p and $\mathcal{R}$, and we make polynomial-in-$\mathcal{R}$ many calls to the LLL lattice basis reduction oracle, each of which returns an answer in polynomial-in-p time. The proof is completed. □

4.10 Proof of Proposition 2.14

Proof. We will now study the modified subset-sum problem with dependent inputs, where $\theta = \sum_{1 \leq i < j \leq p} X_i X_j \xi_{ij}$ with $\xi_{ij} \in \{0, 1\}$. Let $L = \binom{p}{2}$, and $Y_1, \ldots, Y_L$ be an enumeration of $\{X_i X_j : 1 \leq i < j \leq p\}$. Rewrite the problem as $\theta = \sum_{i=1}^{L} Y_i \xi_i$ with $\xi_i \in \{0, 1\}$, where the algorithmic goal is to recover $\xi \in \{0, 1\}^L$, using θ and $Y_1, \ldots, Y_L$. Set $m = p^2 2^{\lceil p^2/4 \rceil}$. We consider $L+1$ dimensional integer lattice, $\Lambda \subset \mathbb{Z}^{L+1}$, generated by the vectors $b_0, b_1, \ldots, b_L$, defined as, $b_0 = (m\theta, 0_{1\times L})$, and for every $i = 1, 2, \ldots, L$, $b_i = (-mY_i, e_i)$, where e_i is the i^{th} element of the standard basis of L-dimensional Euclidean space, $\mathbb{R}^L$. Namely, this is the lattice generated by the columns of the following matrix:

$$\begin{bmatrix} m\theta & -m\widetilde{Y}_{1\times L} \\ 0_{L\times 1} & I_{L\times L} \end{bmatrix},$$

with $\widetilde{Y} = (Y_1, \ldots, Y_L)$. Observe that, $b_0 + \sum_{i=1}^{L} \xi_i b_i = (0, \xi_1, \ldots, \xi_L) = \Xi \in \Lambda$. In particular, we have the bound $\min_{x \in \Lambda, x \neq 0} \|x\| \leq \|\Xi\| < p$, regarding the norm of the shortest non-zero vector of this lattice. Therefore, running the LLL algorithm on the lattice Λ yields a vector, $\widehat{x}$ such that $\|\widehat{x}\| \leq p 2^{L/2} \triangleq m_0$. We will now show that, the 'short' vectors of this lattice are essentially integer multiples of Ξ, which will therefore establish that Ξ can be obtained from the LLL output, keeping in mind that Ξ is a binary vector.

Define the set, $E \subset \Lambda$ as,

$$E = \{x \in \Lambda : \|x\| \leq m_0, x \neq k\Xi, \forall k \in \mathbb{Z}\}.$$

We claim that, $\mathbb{P}(E \neq \varnothing) = o(1)$, as $p \to \infty$. In particular, all 'short' vectors of this lattice, namely those whose norm is at most m_0, are multiples of Ξ. Since the LLL output is guaranteed to have norm at most m_0, this claim establishes that with high probability, LLL output recovers the desired vector, Ξ.

We now prove the claim. Suppose $E \neq \varnothing$, and let $x = (x_0, \ldots, x_L) \in E$. Then, there exists an integer x_0', such that, $x = x_0' b_0 + \sum_{i=1}^{L} x_i b_i$. Now, observe that, if $\theta x_0' - \sum_{i=1}^{L} Y_i x_i \neq 0$, then $\left|\theta x_0' - \sum_{i=1}^{L} Y_i x_i\right| \geq 1$, since it is an integer. Therefore, $|x_0| \geq m$, hence, $\|x\| \geq |x_0| \geq m > m_0$, a contradiction. Therefore, we deduce $x_0'\theta = \sum_{i=1}^{L} Y_i x_i$. As in the proof of outlined by Frieze, the high level idea of the proof is to use a union bound, over all tuples $(x_0', x_1, \ldots, x_L)$ such that $x_0' b_0 + \sum_{i=1}^{L} x_i b_i \in E$, together with the probability of a certain event, for every fixed tuple. To

that end, we begin by bounding the cardinality of allowed such vectors. Note that, $x \in E$ implies $\|x\| \leq m_0$, and therefore, $|x_i| \leq m_0$ for every $i \in [L]$.

We now turn out attention to bounding $|x_0'|$, which requires a separate analysis. Notice, without loss of generality, we may assume that $\theta \geq \frac{1}{2}\sum_{i=1}^{L} Y_i$. Indeed, if this is not the case, we consider an alternative problem, $\widehat{\theta} = \sum_{i=1}^{L} Y_i - \theta = \sum_{i=1}^{L} Y_i(1-\xi_i)$. Observe that, if $\theta < \frac{1}{2}\sum_{i=1}^{N} Y_i$, we would have, $\widehat{\theta} > \frac{1}{2}\sum_{i=1}^{L} Y_i$, and thus we can equivalently consider the problem of recovering ξ', where $\widehat{\theta} = \sum_{i=1}^{L} Y_i\xi_i'$, with inputs $\widehat{\theta}, Y_1, \ldots, Y_L$, and $\xi_i' = 1 - \xi_i$. Under this assumption, we obtain

$$x_0'\theta = \sum_{i=1}^{L} Y_i x_i \implies |x_0'| \cdot \theta \leq \sum_{i=1}^{L} Y_i|x_i| \leq m_0 \sum_{i=1}^{L} Y_i \implies |x_0'| \leq 2m_0.$$

Thus, $|x_0'| \leq 2m_0$. Hence, the set,

$$\mathcal{S} = \{(x_0', x_1, \ldots, x_L) : x_0' b_0 + \sum_{i=1}^{L} x_i b_i \in E\}$$

has cardinality at most $(4m_0+1)\cdot(2m_0+1)^L \leq O(2^{p^4/8+o(p^4)})$.

Next, for any fixed $(x_0', x_1, \ldots, x_L)$ such that the corresponding vector x is contained in E, we will study the probability of the event, $\left\{x_0'\theta = \sum_{i=1}^{L} Y_i x_i\right\}$ which is the event that $\left\{\sum_{i=1}^{L} Y_i\xi_i x_0' = \sum_{i=1}^{L} Y_i x_i\right\}$. Since $x \in E$, we know that $x \neq k\Xi$ for every $k \in \mathbb{Z}$, therefore, there exists an i_0 such that $\xi_{i_0}x_0' \neq x_{i_0}'$. Note that, the smallest i_0, for which $\xi_{i_0}x_0' \neq x_{i_0}$ is not 0, since $x_0 = 0$, as we have already proven, thanks to $\|x\| \leq m_0$.

Let $Y_{i_0} = X_n X_m$ for some $n < m$. Now, for this fixed n, we introduce the following notation: We say $k \sim n$, if Y_k is of form $X_n X_i$, where $i \in [p]$ is an index. That is,

$$k \sim n \iff k \in \{k_0 : \exists i \in [p], Y_{k_0} = X_n X_i, 1 \leq k_0 \leq L\}.$$

By definition, $i_0 \sim n$. Moreover, say $k \not\sim n$ if $Y_k = X_{i'}X_{i''}$, where $n \notin \{i', i''\}$. Also, denote by $X_{\sim i}$ the collection, $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p)$.

Now, define

$$A = \sum_{1\leq k\leq L: k\sim n} \frac{Y_k}{X_n}(\xi_k x_0' - x_k) \quad \text{and} \quad B = \sum_{1\leq k\leq N: k\not\sim n} Y_k(\xi_k x_0' - x_k).$$

Observe that, the source of randomness in B is $X_{\sim n}$, and thus, X_n and B are independent. With this, we observe that,

$$\left\{\sum_{i=1}^{L} Y_i\xi_i x_0' = \sum_{i=1}^{L} Y_i x_i\right\} \implies \{AX_n + B = 0\}.$$

Now, define the events:

$$\mathcal{E} \triangleq \{A = 0\}, \quad \mathcal{F} \triangleq \{AX_n + B = 0\}, \quad \text{and} \quad A_v = \{X_{\sim n} = v\}$$

Focusing on i_0 with $Y_{i_0} = X_n X_m$, we have:

$$\mathcal{E} = \left\{ X_m(\xi_{i_0} - x_{i_0}) = -\sum_{k\sim n, k\neq i_0} \frac{Y_k}{X_n}(\xi_k x_0' - x_{i_0}) \right\}.$$

Due to the fact that $X_1, \ldots, X_p$ are iid, $\mathbb{P}(\mathcal{E}) = O(2^{-N})$. Using these,

$$\begin{aligned}\mathbb{P}(\mathcal{F}) &= \mathbb{P}(\mathcal{F}\cap\mathcal{E}) + \mathbb{P}(\mathcal{F}\cap\mathcal{E}^c) \\ &\leq \sum_v \underbrace{\mathbb{P}(\mathcal{F}|\mathcal{E}^c\cap A_v)}_{\leq O(2^{-N})} \underbrace{\mathbb{P}(\mathcal{E}^c\cap A_v)}_{\leq \mathbb{P}(A_v)} + \mathbb{P}(\mathcal{E}) \\ &\leq O(2^{-N})\sum_v \mathbb{P}(A_v) + O(2^{-N}) = O(2^{-N}),\end{aligned}$$

using the fact that conditional on $\mathcal{E}^c \cap A_v$, the event $\mathcal{F}$ is the event that X_n takes a unique value, which happens with probability $O(2^{-N})$.

Hence, for any fixed $(x_0', x_1, \ldots, x_L)$, the event $\mathcal{F}$ holds with probability at most $O(2^{-N})$. Now, recalling that, $|\{(x_0', x_1, \ldots, x_L) : x_0' b_0 + \sum_{i=1}^L x_i b_i \in E\}| = O(2^{p^4/8 + o(p^4)})$, a union bound over all admissible $(L+1)-$tuples $(x_0', x_1, \ldots, x_L)$ yields that

$$\mathbb{P}(E \neq \varnothing) \leq O(2^{-(N - p^4/8) + o(p^4)}) = o_p(1),$$

since $N \geq (1/8 + \epsilon)p^4$, by assumption of the theorem. □

4.11 Proof of Theorem 2.15

Proof. Note that, $Y^2 = \sum_{i=1}^p X_i^2 |\beta_i^*|^2 + \sum_{1\leq i<j\leq p} X_i X_j ((\beta_i^*)^H \beta_j^* + \beta_i^* (\beta_j^*)^H)$ yields that, Y^2 is an integer combination of the elements of set $\mathcal{L}$, where $\mathcal{L} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$, such that $\mathcal{S}_1 = \{X_i^2 |a_k|^2 : i \in [p], k \in [\mathcal{R}]\}$, $\mathcal{S}_2 = \{X_i X_j |a_k|^2 : 1 \leq i < j \leq p, k \in [\mathcal{R}]\}$, and $\mathcal{S}_3 = \{X_i X_j (a_k^H a_\ell + a_k a_\ell^H) : 1 \leq i < j \leq p, 1 \leq k < \ell \leq \mathcal{R}\}$; namely, $Y^2 = \sum_{a\in\mathcal{L}} a\theta_a^*$ for some coefficients θ_a^*, where for each $a \in \mathcal{L}$, $\theta_a^* \in \mathbb{Z}$. We now establish that $\mathcal{L}$ is rationally independent.

Lemma 4.8. *Let $\mathcal{L}$ be defined as above. Then, $\mathbb{P}(\mathcal{L}$ is rationally independent$) = 1$, where the probability is taken with respect to joint distribution of $\{X_i\}_{i=1}^p$.*

Proof. (of Lemma 4.8) We begin by the following simple auxiliary result, that will turn out to be useful in the proof:

Theorem 4.9. [49] Let ℓ be an arbitrary positive integer; and $P : \mathbb{R}^\ell \to \mathbb{R}$ be a polynomial. Then, either P is identically 0, or $\{x \in \mathbb{R}^\ell : P(x) = 0\}$ has zero Lebesgue measure, namely, $P(x)$ is non-zero almost everywhere.

As a consequence of Theorem 4.9, observe that if $X = (X_1, \ldots, X_p) \in \mathbb{R}^p$ a jointly continuous random vector with density f; $P : \mathbb{R}^p \to \mathbb{R}$ is a polynomial where there exists $(x_1, \ldots, x_p) \in \mathbb{R}^p$ with $P(x_1, \ldots, x_p) \neq 0$; and $S = \{(x_1, \ldots, x_p) \in \mathbb{R}^p : P(x_1, \ldots, x_p) = 0\}$, then we have:

$$\mathbb{P}(P(X) = 0) = \int_{S\subset\mathbb{R}^p} f(x_1, \ldots, x_p) d\lambda(x_1, \cdots, x_p) = 0,$$

where λ is the (p-dimensional) Lebesgue measure; since the set S has zero Lebesgue measure. In particular, if P is a polynomial that is non-vanishing, then $P(X)$ is non-zero almost everywhere, for any jointly continuous $X \in \mathbb{R}^p$.

Equipped with this, we now return to the proof. For an arbitrary subset $S \subseteq [p]$, denote $X_{\sim S} = \{X_i : i \in [p] \setminus S\}$; and when S is a single element set, $S = \{i\}$, let us use $X_{\sim i}$, in place of $X_{\sim\{i\}}$. Now, note that, the event, $\mathcal{L}$ is not rationally independent implies the existence of a collection $\mathcal{Q}$ of rationals, not all zero:

$$\mathcal{Q} = \{q_{ik} : i \in [p], k \in [\mathcal{R}]\} \cup \{r_{ijk} : 1 \leq i < j \leq p, k \in [\mathcal{R}]\} \cup \{t_{ijk\ell} : 1 \leq i < j \leq p, 1 \leq k < \ell \leq \mathcal{R}\}$$

such that,

$$\sum_{i=1}^{p}\sum_{k=1}^{\mathcal{R}} X_i^2|a_k|^2 q_{ik} + \sum_{1\leq i<j\leq p}\sum_{k=1}^{\mathcal{R}} X_iX_j|a_k|^2 r_{ijk} + \sum_{1\leq i<j\leq p}\sum_{1\leq k<\ell\leq\mathcal{R}} X_iX_j(a_k^H a_\ell + a_k a_\ell^H)t_{ijk\ell} = 0.$$

Our strategy is to show that, for any fixed non-zero collection $\mathcal{Q}$, the probability of the event,

$$E_{\mathcal{Q}} = \left\{\sum_{i=1}^{p}\sum_{k=1}^{\mathcal{R}} X_i^2|a_k|^2 q_{ik} + \sum_{1\leq i<j\leq p}\sum_{k=1}^{\mathcal{R}} X_iX_j|a_k|^2 r_{ijk} + \sum_{1\leq i<j\leq p}\sum_{1\leq k<\ell\leq\mathcal{R}} X_iX_j(a_k^H a_\ell + a_k a_\ell^H)t_{ijk\ell} = 0\right\}$$

is zero. Once this is established, a union bound over all non-zero collection $\mathcal{Q}$ of rationals, of the above form yields,

$$\mathbb{P}(\mathcal{L} \text{ is rationally independent}) \leq \sum_{\mathcal{Q}\in\mathbb{Q}^D\setminus\mathbf{0}} \mathbb{P}(E_{\mathcal{Q}}) = 0,$$

since we take a countable union of measure zero events, where $D = p\mathcal{R} + \binom{p}{2}\mathcal{R} + \binom{p}{2}\binom{\mathcal{R}}{2} = |\mathcal{L}|$. Now, we divide the proof into two sub-cases.

(i) Suppose, there is a pair $(i', k') \in [p] \times [\mathcal{R}]$ with $q_{i'k'} \neq 0$. Note that, rational independence of $\mathcal{S}' = \{|a_k|^2 : k \in [\mathcal{R}]\} \cup \{a_k^H a_\ell + a_k a_\ell^H : 1 \leq k < \ell \leq \mathcal{R}\}$ implies any subset of $\mathcal{S}'$ is also rationally independent, in particular, so does $\{|a_k|^2 : k \in [\mathcal{R}]\}$; and also, $|a_i|^2 > 0$ for every i. Using this, the event, $E_{\mathcal{Q}}$ is equal to, $\{X_{i'}^2\gamma_{i'} + X_{i'}B_{i'} + C_{i'} = 0\}$, where, $\gamma_{i'} = \sum_{k=1}^{\mathcal{R}} |a_j|^2 q_{i'k}$ and both $B_{i'}$ and $C_{i'}$ are functions of $X_{\sim i'}$ only, hence, are independent of $X_{i'}$. Now, observe that $\gamma_{i'} = 0$ if and only if $q_{i'k} = 0$ for every $k \in [\mathcal{R}]$ as $\{|a_k|^2 : k \in [\mathcal{R}]\}$ is rationally independent, but since $q_{i'k'} \neq 0$, it follows that $\gamma_{i'} \neq 0$. In particular, the event of interest is of form $\{P(X) = 0\}$ for some polynomial $P : \mathbb{R}^p \to \mathbb{R}$, where $P \neq 0$. Thus, for this selection of $\mathcal{Q}$, we deduce using Theorem 4.9 that $\mathbb{P}(E_{\mathcal{Q}}) = 0$.

(ii) Now, suppose that, $q_{ik} = 0$, for every $i \in [p]$ and $k \in [\mathcal{R}]$. Now, for every $1 \leq i < j \leq p$, define by ξ_{ij} the number

$$\xi_{ij} = \sum_{k=1}^{\mathcal{R}} |a_k|^2 r_{ijk} + \sum_{1\leq k<\ell\leq\mathcal{R}} (a_k^H a_\ell + a_k a_\ell^H)t_{ijk\ell}.$$

In order not to deal with cases $i < j$ and $i > j$ separately, let us adopt the convention that $\xi_{ji} = \xi_{ij}$ for every $i \neq j$. Note that, since the set, $\mathcal{S}' = \{|a_k|^2 : k \in [\mathcal{R}]\} \cup \{a_k^H a_\ell + a_k a_\ell^H :$

$1 \leq k < \ell \leq \mathcal{R}\}$ is assumed to be rationally independent, $\mathcal{Q}$ is a non-zero collection of rationals, and $q_{ij} = 0$, for every $i \in [p], j \in [\mathcal{R}]$, we deduce that, there exists a pair, (i_0, j_0) such that $\xi_{i_0 j_0} \neq 0$. Keeping this in mind,

$$E_{\mathcal{Q}} = \left\{ \sum_{1 \leq i < j \leq p} X_i X_j \xi_{ij} = 0 \right\} = \{A_{i_0} X_{i_0} + B_{i_0} = 0\} \subset \{A_{i_0} = 0\} \cup \{A_{i_0} X_{i_0} + B_{i_0} = 0, A_{i_0} \neq 0\},$$

where, $A_{i_0} = \sum_{j \neq i_0} X_j \xi_{i_0 j}$ and $B_{i_0} = \sum_{1 \leq i < j \leq p, i, j \neq i_0} X_i X_j \xi_{ij}$. Notice that, both A_{i_0} and B_{i_0} are polynomials in of $X_{\sim i_0}$, and are independent of X_{i_0}. Moreover, due to the fact that $\xi_{i_0, j_0} \neq 0$, we have that A_{i_0} is not vanishing (this can be seen, for instance, by observing that taking $X_j = 0$ for all $j \neq j_0$ and setting $X_{j_0} \neq 0$, the polynomial A_{i_0} evaluates to $\xi_{i_0 j_0} X_{j_0}$, which is clearly non-zero). Thus, using Theorem 4.9, we have $\mathbb{P}(A_{i_0} = 0) = 0$. Next, we claim the probability of the event, $\{A_{i_0} X_{i_0} + B_{i_0} = 0, A_{i_0} \neq 0\}$ is 0. To see this, we proceed by conditioning on $X_{\sim i_0}$ such that $A_{i_0} \neq 0$, and observing that conditional on this, the event, $\{A_{i_0} X_{i_0} + B_{i_0} = 0\}$, is simply the probability that a certain polynomial in X_{i_0} is non-zero, which again by Theorem 4.9 happens with probability 0.

Hence, in both cases, we have $\mathbb{P}(E_{\mathcal{Q}}) = 0$, and therefore, we are done with the proof of the lemma. □

Now, recall $D = |\mathcal{L}|$, and define $\mathbf{m} = (m_0, m_i : 1 \leq i \leq D) \in \mathbb{Z}^{D+1} \setminus \{\mathbf{0}\}$ to be an integer relation, for the vector, $(Y^2, a : a \in \mathcal{L})$. Using the exact same steps, as in proof of Theorem 2.9, we deduce, as a consequence of Lemma 4.8 that, with probability 1, any integer relation for this vector must be a multiple of $(-1, \theta^*_a : a \in \mathcal{L}) \in \mathbb{Z}^{D+1}$ where we have defined θ^*_a earlier as $Y^2 = \sum_{a \in \mathcal{L}} a\theta^*_a$, that is, the expression of Y^2, as an integer combination of the elements of $\mathcal{L}$. Hence, for every i, $\beta^*_i = a_k$, where the coefficient $\theta^*_{X_i^2 |a_k|^2}$ in the integer relation corresponding to the product $X_i^2 |a_k|^2$ is non-zero (with probability 1, there must be a unique $k \in [\mathcal{R}]$ for which $\theta^*_{X_i^2 |a_k|^2} \neq 0$). Finally, since the cardinality of $\mathcal{L}$ is at most $O(p^2 \mathcal{R}^2)$, the overall runtime of the procedure is at most polynomial in p and $\mathcal{R}$ (here, the overall runtime includes generating the sets $\mathcal{S}'$ and $\mathcal{L}$, as well as running integer relation solver on a vector consisting of elements of $\mathcal{L}$ and Y^2). □

4.12 Proof of Proposition 2.5

Proof. If we show that we can apply Theorem 2.4.B, the result follows. Since the model assumptions are identical, we only need to check the parameter assumptions of Theorem 2.4.B. First, note that we assume $\hat{R} = R$, we clearly have for the noise $\sigma \leq W_\infty = 1$ and finally $\hat{Q} = Q$. Now for establishing (4), we first notice that since $N \leq \log\left(\frac{1}{\sigma}\right)$ is equivalent to $2^N \sigma \leq 1$, we obtain $2^N \sigma \sqrt{np} + Rp \leq 2^{\log(np) + \log(Rp)}$. Therefore it suffices to have

$$N > \frac{(2n+p)^2}{2n} + 22 \frac{2n+p}{n} \log(3(1+c)np) + \frac{2n+p}{n} \log(RQ).$$

Now since $p \geq \frac{300}{\epsilon} \log\left(\frac{300}{c\epsilon}\right)$, the following holds for all $n \in \mathbb{Z}^+$:

$$22(2n+p) \log(3(1+c)np) < \frac{\epsilon}{2} \frac{(2n+p)^2}{2}. \tag{16}$$

Indeed, this can be equivalently written as

$$22 < \frac{\epsilon}{4}\frac{2n+p}{\log(3(1+c)np)}.$$

But $\frac{2n+p}{\log(3(1+c)np)}$ increases with respect to $n \in \mathbb{Z}^+$ and therefore it is minimized for $n = 1$. In particular it suffices to have

$$22 < \frac{\epsilon}{4}\frac{2+p}{\log(3(1+c)p)},$$

which can be checked to be true for $p \geq \frac{300}{\epsilon}\log\left(\frac{300}{(1+c)\epsilon}\right)$. Therefore using (16), it suffices to have

$$N > (1+\frac{\epsilon}{2})\frac{(2n+p)^2}{2n} + \frac{2n+p}{n}\log(RQ).$$

But now, we observe the following chain of inequalities/equality:

$$\begin{aligned} N &\geq (1+\epsilon)\left[\frac{p^2}{2n} + 2n + 2p + (2+\frac{p}{n})\log(RQ)\right] \\ &= (1+\epsilon)\left[\frac{(2n+p)^2}{2n} + (\frac{2n+p}{n})\log(RQ)\right] \\ &> (1+\frac{\epsilon}{2})\frac{(2n+p)^2}{2} + (2n+p)\log(RQ). \end{aligned}$$

This concludes the proof of Proposition 2.5. □

4.13 Proof of Proposition 2.6

Proof. We first establish that $\|X\|_\infty \leq (np)^2$ whp as $p \to +\infty$. By a union bound and the Markov inequality, we have

$$\mathbb{P}\left(\max_{i\in[n],j\in[p]} |X_{ij}| > (np)^2\right) \leq np\mathbb{P}\left(|X_{11}| > (np)^2\right) \leq \frac{1}{np}\mathbb{E}[|X_{11}|] = o(1).$$

Therefore with high probability, $\|X\|_\infty \leq (np)^2$. Consider the set $T(R,Q)$ of all the vectors $\beta^* \in [-R,R]^p$ satisfying the Q-rationality assumption. The entries of these vectors are of the form $\frac{a}{Q}$ for some $a \in \mathbb{Z}$ with $|a| \leq RQ$. In particular $|T(R,Q)| = (2QR+1)^p$. Now because the entries of X are continuously distributed, all $X\beta^*$ with $\beta^* \in T(R,Q)$ are distinct with probability 1. Furthermore by the above each one of them has L_2 norm satisfies

$$\|X\beta^*\|_2^2 \leq np^2\|X\|_\infty^2\|\beta^*\|_\infty^2 \leq R^2n^5p^6 < R^2(np)^6,$$

w.h.p. as $p \to +\infty$.

Now we establish the proposition by contradiction. Suppose there exist a recovery mechanism that can recover w.h.p. any such vector β^* after observing $Y = X\beta^* + W \in \mathbb{R}^n$, where W has n iid $N(0,\sigma^2)$ entries. In the language of information theory such a recovery guarantee implies that the Gaussian channel with power constraint $R^2(np)^6$ and noise variance σ^2 needs to have capacity at least

$$\frac{\log|T(R,Q)|}{n} = \frac{p\log(2QR+1)}{n}.$$

On the other hand, the capacity of this Gaussian channel with power $\mathcal{P}$ and noise variance Σ^2 is known to be equal to $\frac{1}{2}\log\left(1+\frac{\mathcal{P}}{\Sigma^2}\right)$ (see for example [50, Theorem 10.1.1]). In particular our Gaussian communication channel has capacity

$$\frac{1}{2}\log\left(1+\frac{R^2(np)^6}{\sigma^2}\right).$$

From this we conclude that

$$\frac{p\log\left(2QR+1\right)}{n} \leq \frac{1}{2}\log\left(1+\frac{R^2(np)^6}{\sigma^2}\right).$$

This implies

$$\sigma^2 \leq R^2(np)^6 \frac{1}{2^{\frac{2p\log(2QR+1)}{n}}-1},$$

or

$$\sigma \leq R(np)^3\left(2^{\frac{2p\log(2QR+1)}{n}}-1\right)^{-\frac{1}{2}},$$

which completes the proof of the Proposition. □

4.14 Proof of Proposition 2.7

Proof. Based on Proposition 2.5 the amount of noise that can be tolerated is

$$2^{-(1+\epsilon)\left[\frac{p^2}{2n}+2n+2p+(2+\frac{p}{n})\log(RQ)\right]},$$

for an arbitrary $\epsilon > 0$. Since $n = o(p)$ and $RQ = 2^{\omega(p)}$ this simplifies asymptotically to $2^{-(1+\epsilon)\left[\frac{p}{n}\log(RQ)\right]}$, for an arbitrary $\epsilon > 0$. Since $\sigma < \sigma_0^{1+\epsilon}$, we conclude that the LBR algorithm works successfully in that regime.

For the first part, it suffices to establish that under our assumptions for p sufficiently large,

$$\sigma_0^{1-\epsilon} > R(np)^3\left(2^{\frac{2p\log(2QR+1)}{n}}-1\right)^{-\frac{1}{2}}.$$

Since $n = o(\frac{p}{\log p})$ implies $n = o(p)$ we obtain that for p sufficiently large,

$$2^{\frac{2p\log(2QR+1)}{n}} - 1 > 2^{2(1-\frac{1}{2}\epsilon)\frac{p\log(2QR+1)}{n}}$$

which equivalently gives

$$\left(2^{\frac{2p\log(2QR+1)}{n}}-1\right)^{-\frac{1}{2}} < 2^{-(1-\frac{1}{2}\epsilon)\frac{p\log(2QR+1)}{n}}$$

or

$$R(np)^3\left(2^{\frac{2p\log(2QR+1)}{n}}-1\right)^{-\frac{1}{2}} < R(np)^3 2^{-(1-\frac{1}{2}\epsilon)\frac{p\log(2QR+1)}{n}}.$$

Therefore it suffices to show

$$R(np)^3 2^{-(1-\frac{1}{2}\epsilon)\frac{p\log(2QR+1)}{n}} \leq \sigma_0^{1-\epsilon} = 2^{-(1-\epsilon)\frac{p\log(QR)}{n}}$$

or equivalently by taking logarithms and performing elementary algebraic manipulations, it suffices to show

$$n \log R + 3n \log(np) \leq \left(1 - \frac{\epsilon}{2}\right) p \log(2 + \frac{1}{RQ}) + \frac{\epsilon}{2} p \log RQ.$$

The condition $n = o(\frac{p}{\log p})$ implies for sufficiently large p, $n \log(np) \leq \frac{\epsilon}{4} p$ and $n \log R \leq \frac{\epsilon}{2} p \log QR$. Using both of these inequalities we conclude that for p sufficiently large,

$$\begin{aligned} n \log R + 3n \log(np) &\leq \frac{\epsilon}{2} p \log QR \\ &\leq \left(1 - \frac{\epsilon}{2}\right) p \log(2 + \frac{1}{RQ}) + \frac{\epsilon}{2} p \log RQ. \end{aligned}$$

This completes the proof.

□

Acknowledgments

The authors would like to gratefully acknowledge the work of Patricio Foncea and Andrew Zheng who conducted the synthetic experiments for the ELO and the LBR algorithms; and would also like to thank the two anonymous reviewers for their careful reviews and remarks that improved the clarity of this paper and bringing the earlier papers by Teunissen to their attention.

References

[1] I. Zadik and D. Gamarnik, "High Dimensional Linear Regression using Lattice Basis Reduction," in *Advances in Neural Information Processing Systems*, 2018, pp. 1842–1852.

[2] D. Gamarnik and E. C. Kızıldağ, "High-Dimensional Linear Regression and Phase Retrieval via PSLQ Integer Relation Algorithm," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1437–1441.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[4] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] E. J. Candes, J. K. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006. [Online]. Available: http://dx.doi.org/10.1002/cpa.20124

[6] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.

[7] M. J. Wainwright, "Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using ℓ_1-Constrained Quadratic Programming (LASSO)," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2183–2202, 2009.

[8] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed Sensing Using Generative Models," in *International Conference on Machine Learning*. PMLR, 2017, pp. 537–546.

[9] D. Donoho and J. Tanner, "Counting Faces of Randomly Projected Polytopes When the Projection Radically Lowers Dimension," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 1–53, 2009.

[10] D. Gamarnik and I. Zadik, "Sparse High-Dimensional Linear Regression. Algorithmic Barriers and a Local Search Algorithm," *arXiv preprint arXiv:1711.04952*, 2017.

[11] ——, "High-Dimensional Linear Regression with Binary Coefficients. Estimating Squared Error and a Phase Transition," in *Conference on Learning Theory (COLT)*, 2017.

[12] P. J. Teunissen, "Least-Squares Estimation of the Integer GPS Ambiguities," in *Invited Lecture, Section IV Theory and Methodology, IAG General Meeting, Beijing, China*, 1993, pp. 1–16.

[13] ——, "A New Method for Fast Carrier Phase Ambiguity Estimation," in *Proceedings of 1994 IEEE Position, Location and Navigation Symposium-PLANS'94*. IEEE, 1994, pp. 562–573.

[14] P. Teunissen, "The Invertible GPS Ambiguity Transformations," *Manuscripta Geodaetica*, vol. 20(6), 1995.

[15] P. J. G. Teunissen, "The Least-Squares Ambiguity Decorrelation Adjustment: A Method for Fast GPS Integer Ambiguity Estimation," *Journal of Geodesy*, vol. 70, no. 1, pp. 65–82, 1995. [Online]. Available: https://doi.org/10.1007/BF00863419

[16] H. W. Lenstra Jr, "Integer Programming with a Fixed Number of Variables," *Mathematics of Operations Research*, vol. 8, no. 4, pp. 538–548, 1983.

[17] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring Polynomials with Rational Coefficients," *Mathematische Annalen*, vol. 261, no. 4, pp. 515–534, 1982.

[18] A. Hassibi and S. Boyd, "Integer Parameter Estimation in Linear Models with Applications to GPS," *IEEE Transactions on Signal Processing*, vol. 46, no. 11, pp. 2938–2952, 1998.

[19] D. Medina, J. Vilà-Valls, E. Chaumette, F. Vincent, and P. Closas, "Cramér-Rao Bound for a Mixture of Real-and Integer-valued Parameter Vectors and its Application to the Linear Regression Model," *Signal Processing*, vol. 179, p. 107792, 2020.

[20] M. Ajtai, "The Shortest Vector Problem in L_2 is NP-hard for Randomized Reductions," in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, 1998, pp. 10–19.

[21] B. Hassibi and H. Vikalo, "On the Expected Complexity of Integer Least-Squares Problems," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2. IEEE, 2002, pp. II–1497.

[22] M. A. Borno, "Reduction in Solving Some Integer Least Squares Problems," *arXiv preprint arXiv:1101.0382*, 2011.

[23] S. Breen, *Integer Least Squares Search and Reduction Strategies*. McGill University (Canada), 2011.

[24] U. Fincke and M. Pohst, "Improved Methods for Calculating Vectors of Short Length in a Lattice, Including a Complexity Analysis," *Mathematics of Computation*, vol. 44, no. 170, pp. 463–471, 1985.

[25] A. Brack and C. Günther, "Generalized Integer Aperture Estimation for Partial GNSS Ambiguity Fixing," *Journal of Geodesy*, vol. 88, no. 5, pp. 479–490, 2014.

[26] P. Teunissen, "Theory of Integer Equivariant Estimation with Application to GNSS," *Journal of Geodesy*, vol. 77, no. 7-8, pp. 402–410, 2003.

[27] A. Lannes, "On the Theoretical Link between LLL-Reduction and LAMBDA-Decorrelation," *Journal of Geodesy*, vol. 87, no. 4, pp. 323–335, 2013.

[28] D. L. Donoho and J. Tanner, "Neighborliness of Randomly Projected Simplices in High Dimensions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9452–9457, 2005.

[29] D. Donoho and J. Tanner, "Observed Universality of Phase Transitions in High-Dimensional Geometry, with Implications for Modern Data Analysis and Signal Processing," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1906, pp. 4273–4293, 2009.

[30] M. Celentano and A. Montanari, "Fundamental Barriers to High-Dimensional Regression with Convex Penalties," *Annals of Statistics*, p. To appear.

[31] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-Theoretically Optimal Compressed Sensing via Spatial Coupling and Approximate Message Passing," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7434–7464, 2013.

[32] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Statistical-Physics-Based Reconstruction in Compressed Sensing," *Physical Review X*, vol. 2, no. 2, p. 021005, 2012.

[33] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase Retrieval via Matrix Completion," *SIAM Review*, vol. 57, no. 2, pp. 225–251, 2015.

[34] J. C. Lagarias and A. M. Odlyzko, "Solving Low-Density Subset Sum Problems," *Journal of the ACM (JACM)*, vol. 32, no. 1, pp. 229–246, 1985.

[35] A. Lempel, "Cryptology in Transition," *ACM Computing Surveys (CSUR)*, vol. 11, no. 4, pp. 285–303, 1979.

[36] R. Merkle and M. Hellman, "Hiding Information and Signatures in Trapdoor Knapsacks," *IEEE Transactions on Information Theory*, vol. 24, no. 5, pp. 525–530, 1978.

[37] A. Shamir, "A Polynomial Time Algorithm for Breaking the Basic Merkle-Hellman Cryptosystem," in *23rd Annual Symposium on Foundations of Computer Science (SFCS 1982)*. IEEE, 1982, pp. 145–152.

[38] A. M. Frieze, "On the Lagarias-Odlyzko Algorithm for the Subset Sum Problem," *SIAM Journal on Computing*, vol. 15, no. 2, pp. 536–539, 1986.

[39] H. R. Ferguson and R. W. Forcade, "Generalization of the Euclidean Algorithm for Real Numbers to all Dimensions Higher than Two," *Bulletin of the American Mathematical Society*, vol. 1, no. 6, pp. 912–914, 1979.

[40] H. R. Ferguson, "A Noninductive $GL(n, Z)$ Algorithm that Constructs Integral Linear Relations for n Z-Linearly Dependent Real Numbers," *Journal of Algorithms*, vol. 8, no. 1, pp. 131–145, 1987.

[41] J. Hastad, B. Just, J. C. Lagarias, and C.-P. Schnorr, "Polynomial Time Algorithms for Finding Integer Relations among Real Numbers," *SIAM Journal on Computing*, vol. 18, no. 5, pp. 859–881, 1989.

[42] H. R. Ferguson and D. H. Bailey, "A Polynomial Time, Numerically Stable Integer Relation Algorithm," NASA Applied Research Branch, NASA Ames Research Center, RNR Technical Report RNR-91-032, March 1992.

[43] H. Ferguson, D. Bailey, and S. Arno, "Analysis of PSLQ, an Integer Relation Finding Algorithm," *Mathematics of Computation*, vol. 68, no. 225, pp. 351–369, 1999.

[44] D. H. Bailey, "Integer Relation Detection," *Computing in Science & Engineering*, vol. 2, no. 1, pp. 24–28, 2000.

[45] D. Bailey, P. Borwein, and S. Plouffe, "On the Rapid Computation of Various Polylogarithmic Constants," *Mathematics of Computation*, vol. 66, no. 218, pp. 903–913, 1997.

[46] D. Bailey and D. Broadhurst, "Parallel Integer Relation Detection: Techniques and Applications," *Mathematics of Computation*, vol. 70, no. 236, pp. 1719–1736, 2001.

[47] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*. Oxford University Press, 1979.

[48] P. Erdös and G. Lorentz, "On the Probability that n and $g(n)$ are Relatively Prime," *Acta Arithmetica*, vol. 5, pp. 35–44, 1959.

[49] R. Caron and T. Traynor, "The Zero Set of a Polynomial," *WSMR Report*, pp. 05–02, 2005.

[50] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.