



ELSEVIER

Contents lists available at ScienceDirect

Journal of Number Theory

www.elsevier.com/locate/jnt



General Section

Ramanujan graphs and exponential sums over function fields

Naser T. Sardari^a, Masoud Zargar^{b,*}^a Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA^b Fakultät für Mathematik, Universität Regensburg, Universitätsstr. 31, 93040 Regensburg, Germany

ARTICLE INFO

Article history:

Received 23 October 2019

Received in revised form 29 April 2020

Accepted 1 May 2020

Available online 22 June 2020

Communicated by S.J. Miller

Keywords:

Ramanujan graphs

Optimal diameter

Quadratic forms

(Optimal) strong approximation

Function fields

Exponential sums

Stationary phase over function fields

Kloosterman sums

Linnik–Selberg

Morgenstern

ABSTRACT

We prove that $q + 1$ -regular Morgenstern Ramanujan graphs $X^{q,g}$ (depending on $g \in \mathbb{F}_q[t]$) have diameter at most $(\frac{4}{3} + \varepsilon) \log_q |X^{q,g}| + O_\varepsilon(1)$ (at least for odd q and irreducible g) provided that a twisted Linnik–Selberg conjecture over $\mathbb{F}_q(t)$ is true. This would break the 30 year-old upper bound of $2 \log_q |X^{q,g}| + O(1)$, a consequence of a well-known upper bound on the diameter of regular Ramanujan graphs proved by Lubotzky, Phillips, and Sarnak using the Ramanujan bound on Fourier coefficients of modular forms. We also unconditionally construct infinite families of Ramanujan graphs that prove that $\frac{4}{3}$ cannot be improved.

© 2020 Elsevier Inc. All rights reserved.

Contents

1. Introduction	45
2. Lower bound on the diameter	50
3. Recollections on the delta method	52

* Corresponding author.

E-mail addresses: ntalebiz@ias.edu (N.T. Sardari), masoud.zargar@ur.de (M. Zargar).

4. The oscillatory integrals $I_{g,r}(\mathbf{c})$	56
5. The exponential sums $S_{g,r}(\mathbf{c})$	62
6. Strong approximation and Ramanujan graphs	66
Acknowledgments	76
References	77

1. Introduction

We begin by defining Ramanujan graphs. Suppose $k \geq 3$ is a fixed integer, and let G be a k -regular connected graph with adjacency matrix A_G . Since A_G is symmetric, all its eigenvalues are real. Furthermore, it is easy to see that k is the largest eigenvalue of the adjacency matrix A_G . Letting λ_G be the second largest eigenvalue, it is a theorem of Alon–Boppana [Alo86] that $\lambda_G \geq 2\sqrt{k-1} + o(1)$, where $o(1)$ goes to zero as $|G| \rightarrow \infty$. We say that G is a Ramanujan graph if $\lambda_G \leq 2\sqrt{k-1}$. The natural question that arises is if we can construct an infinite sequence of such Ramanujan graphs. Such graphs have been constructed by Lubotzky–Phillips–Sarnak [LPS88], Margulis [Mar88], Morgenstern [Mor94], and others. Though the first two constructions are the same and are $p+1$ -regular, p a prime, those of Morgenstern are $q+1$ -regular, q any prime power. Recently, Marcus–Spielman–Srivastava have proved the existence of d -regular bipartite Ramanujan graphs for arbitrary d [MSS15]; Cohen has shown how to construct such d -regular graphs in polynomial time [Coh16]. The study of such graphs is intimately connected to deep questions in number theory, and is also of interest to computer scientists.

Lubotzky–Phillips–Sarnak [LPS88], and independently Margulis [Mar88], constructed the first examples of Ramanujan graphs; they are Cayley graphs of $\mathrm{PGL}_2(\mathbb{Z}/N\mathbb{Z})$ or $\mathrm{PSL}_2(\mathbb{Z}/N\mathbb{Z})$ with $p+1$ explicit generators, for every prime p and natural number N . We denote them by $X^{p,N}$. The fact that $X^{p,N}$ is a Ramanujan graph follows from the Ramanujan bound on the p -th Fourier coefficients of the weight 2 holomorphic modular forms of level N , hence the naming of *Ramanujan* graphs. This was proved by Lubotzky, Phillips, and Sarnak in [LPS88]. Another result of Lubotzky–Phillips–Sarnak is that the diameter of every k -regular Ramanujan graph G is bounded from above by $2\log_{k-1}|G| + O(1)$. As of today, this is still the best known upper bound on the diameter of a Ramanujan graph. Though it was conjectured by Sarnak that the diameter is bounded from above by $(1+\varepsilon)\log_{k-1}|G|$ as $|G| \rightarrow \infty$ (see [Sar90, Chapter 3]), the first author proved that for some infinite families of LPS Ramanujan graphs $X^{p,N}$ the diameter is bigger than $4/3\log_p|X^{p,N}| + O(1)$ (see [Sar18]).

Let q be a prime power and \mathbb{F}_q be the finite field with q elements. Morgenstern also constructed Ramanujan graphs $X^{q,g}$ by considering a suitable quaternion algebra over $\mathbb{F}_q[t]$, where $g \in \mathbb{F}_q[t]$ and $\gcd(g, t(t-1)) = 1$ [Mor94]. For a discussion of the connection to strong approximation, see the introduction to the authors' paper [SZ20]. The graph-theoretic conjecture with which this paper is concerned is the following.

Conjecture 1.1. *The diameter of $q + 1$ -regular Morgenstern Ramanujan graphs $X^{q,g}$ is bounded from above by*

$$\left(\frac{4}{3} + \varepsilon\right) \log_q |X^{q,g}| + O_\varepsilon(1),$$

at least when q is odd and g is irreducible.

Let us assume from now on that the base field \mathbb{F}_q is of odd characteristic. Consider the following system of equations

$$\begin{cases} F(\mathbf{x}) = f, \\ \mathbf{x} \equiv \boldsymbol{\lambda} \pmod{g}, \end{cases} \quad (1)$$

where F is a quadratic form in 4 variables over $\mathbb{F}_q[t]$, $f, g \in \mathbb{F}_q[t]$, and $\boldsymbol{\lambda} \in \mathbb{F}_q[t]^4$. We say that all the local conditions for the system (1) are satisfied if the system (1) has solutions over $\mathbb{F}_q[t]/\langle gh \rangle$ for any nonzero $h \in \mathbb{F}_q[t]$, in addition to $F(\mathbf{x}) = f$ having a solution over $\mathbb{F}_q((1/t))$. As discussed in the introduction to the authors' paper [SZ20], the above conjecture would follow from the following very general conjecture regarding strong approximation for quadratic forms over $\mathbb{F}_q[t]$ in 4 variables. Throughout this paper, the ideal (a, b) of $\mathbb{F}_q[t]$ generated by polynomials $a, b \in \mathbb{F}_q[t]$ (which is a principal ideal domain) will be identified with its monic generator.

Conjecture 1.2. *Let F be a quadratic form over $\mathbb{F}_q[t]$ in 4 variables and of discriminant $\Delta \neq 0$. Let $f, g \in \mathbb{F}_q[t]$ be nonzero polynomials such that $(f\Delta, g) = 1$, and let $\boldsymbol{\lambda} \in \mathbb{F}_q[t]^4$ be a quadruple of polynomials at least one of whose coordinates is relatively prime to g . Finally, suppose that all local conditions for the system (1) are satisfied. There is a solution $\mathbf{x} \in \mathbb{F}_q[t]^4$ to (1) if $\deg f \geq (4 + \varepsilon) \deg g + O_{\varepsilon, F}(1)$.*

In fact, to obtain Conjecture 1.1, it suffices to prove the above strong approximation result for Morgenstern quadratic forms given by

$$F(x_1, x_2, x_3, x_4) = \eta_1 x_1^2 + \eta_2 x_2^2 + \eta_3 x_3^2 + \eta_4 x_4^2,$$

where $\eta_1 = 1$, $\eta_2 = -\nu$, $\eta_3 = -(t-1)$, $\eta_4 = \nu(t-1)$ and $\nu \in \mathbb{F}_q$ is not a square. In [SZ20], we proved Conjecture 1.2 when the number of variables is $d \geq 5$. Furthermore, we showed that for quadratic forms in $d = 4$ variables, Conjecture 1.2 holds if we strengthen the condition to $\deg f \geq (6 + \varepsilon) \deg g + O_{\varepsilon, F}(1)$ (see [SZ20] for details). However, as we saw in Corollary 1.6 of [SZ20] this implies the weaker upper bound $(2 + \varepsilon) \log_{k-1} |X^{q,g}| + O_\varepsilon(1)$ on the diameter of such graphs. The main purpose of this paper is to show that the $\frac{4}{3}$ -bound is a consequence of a *twisted* Linnik–Selberg conjecture over $\mathbb{F}_q(t)$ that we formulate; the way we show this is that the twisted Linnik–Selberg conjecture below implies Conjecture 1.2 for the Morgenstern quadratic form F .

Working over function fields is not merely a curiosity; as we will see, in the function field case, the oscillatory integrals, once a suitable weight function is chosen, can be computed explicitly for Morgenstern quadratic forms in terms of Kloosterman sums at the infinite place. This uses the function field stationary phase theorem proved by the authors in [SZ20]. It turns out that the exponential sums can also be written in terms of Kloosterman sums at the finite places. Furthermore, these graphs constructed by Morgenstern provide us with $q + 1$ -regular Ramanujan graphs, where q need not be a prime, contrary to the Ramanujan graphs constructed by Lubotzky–Phillips–Sarnak, and independently by Margulis. Most importantly, since the untwisted Linnik–Selberg conjecture is known to be true over function fields [CPS90] (a consequence of the Ramanujan conjecture over function fields proved by Drinfeld), we are hopeful that we will be able to prove at least a variant of the twisted Linnik–Selberg conjecture and at least improve upon the best known upper bound on the diameter of such graphs; see Remark 4. When working over the integers, even if one reduces the optimal strong approximation theorem in four variables to a twisted Linnik–Selberg over the integers, there is little hope in saying much about LPS Ramanujan graphs. Indeed, in the integer case, the Linnik–Selberg conjecture is, as of the writing of this paper, open and is not a consequence of the Ramanujan conjecture. We now devote some time to precisely formulating this conjecture.

In order to formulate the twisted Linnik–Selberg conjecture over function fields, we first define the Kloosterman sums under consideration in this paper. In Subsection 3.2, we define a nontrivial additive character $\psi = \psi_\infty$ on $K_\infty := \mathbb{F}_q((1/t))$ that is trivial on $\mathcal{O} := \mathbb{F}_q[t]$. Given nonzero $r \in \mathbb{F}_q[t]$, we have an additive character

$$\psi_r : \mathbb{F}_q[t]/(r) \rightarrow \mathbb{C}^*$$

given by sending $x \mapsto \psi\left(\frac{x}{r}\right)$. We may extend this to an additive character on the additive structure of the subring \mathcal{O}_r of K with denominator relatively prime to r . The extension we are thinking of here is given by sending $x \in \mathcal{O}_r$ to $\psi\left(\frac{x \bmod r}{r}\right)$.

Definition 1.3. Suppose $r \in \mathbb{F}_q[t]$ is nonzero, and suppose $m, n \in \mathcal{O}_r$. Then we define the Kloosterman sum associated to r, m, n as follows:

$$\text{Kl}_r(m, n) := \sum_{x \in (\mathbb{F}_q[t]/(r))^*} \psi_r(mx + n\bar{x}),$$

where \bar{x} is the multiplicative inverse of x in $\mathbb{F}_q[t]/(r)$. At the infinite place, we have the following definition of the Kloosterman sum. For $\alpha \in K_\infty$, define

$$\text{Kl}_\infty(\psi, \alpha) := \begin{cases} \int_{|x|_\infty = \widehat{l}} \psi\left(\frac{\alpha}{x} + x\right) dx, & \text{if } |\alpha|_\infty = \widehat{l}^2 \text{ for some } l \in \mathbb{Z} \\ 0 & \text{otherwise,} \end{cases}$$

where we are integrating on a subset of K_∞ equipped with the Haar measure normalized such that the unit open ball \mathbb{T} in K_∞ has measure 1, $|\cdot|_\infty$ is the norm on K_∞ induced by the norm $|a/b|_\infty = q^{\deg a - \deg b}$ on K , and $\widehat{l} := q^l$ throughout this paper.

By Weil's estimate on the Kloosterman sums, we have square-root cancellation on the Kloosterman sums. The following is a twisted analogue of the Linnik–Selberg conjecture positing that sums of Kloosterman sums exhibit an additional square-root cancellation.

Conjecture 1.4 (*Twisted Linnik–Selberg conjecture over function fields*). Suppose $g \in \mathbb{F}_q[t]$ is a nonzero polynomial, and let $\delta \in \mathbb{F}_q[t]$ be relatively prime to g . Then for each integer $T \geq 0$, $\alpha \in \mathbb{F}_q[t]$, and nonzero $a, b \in \mathbb{F}_q[t, g^{-1}]$,

$$\left| \sum_{\substack{|r|=\widehat{T} \\ (g,r)=1, \delta|r}} \psi_{g^2}(\alpha r^{-1}) \text{Kl}_r(a, b) \right| \ll_{\varepsilon, \delta} |gab|_\infty^\varepsilon \widehat{T}^{1+\varepsilon}$$

for every $\varepsilon > 0$. Furthermore, for every $\varepsilon > 0$,

$$\left| \sum_{\substack{|r|=\widehat{T} \\ (g,r)=1, \delta|r}} \psi_{g^2}(\alpha r^{-1}) \text{Kl}_r(a, b) \text{Kl}_\infty(\psi, ab/r^2) \right| \ll_{\varepsilon, \delta} |gab|_\infty^\varepsilon \widehat{T}^{1+\varepsilon}.$$

Remark 2. Note that in the above conjecture, it is equivalent to prove the statement by replacing $|r| = \widehat{T}$ with $|r| \leq \widehat{T}$. Indeed, if we have the latter for every T , then we have the former by simply subtracting the terms for $|r| \leq \widehat{T}$ from those with $|r| \leq \frac{\widehat{T}}{q}$. Conversely, if we assume the former for every T , then the latter follows from the triangle inequality and the fact that we are summing $T \ll_\varepsilon \widehat{T}^\varepsilon$ elements.

Remark 3. As previously mentioned, we remark that the untwisted version of the above conjecture is known to be true. For example, see [CPS90] for a proof of the Linnik–Selberg conjecture over function fields. The authors hope to study this twisted Linnik–Selberg conjecture in the future.

We prove the following in Section 6.

Theorem 1.5. Conjecture 1.4 implies Conjecture 1.2 for the Morgenstern quadratic form F at least when g is irreducible. In particular, the twisted Linnik–Selberg conjecture implies that the Morgenstern Ramanujan graphs have diameter at most

$$\left(\frac{4}{3} + \varepsilon\right) \log_q |X^{q,g}| + O_\varepsilon(1)$$

if the Ramanujan graph is constructed (at least) for irreducible g over \mathbb{F}_q of odd characteristic.

We briefly discuss the history behind the proof. Versions of the circle method, as developed over the integers by Heath-Brown [HB96] and later over function fields by Browning and Vishe [BV15], were successfully applied by the first author to prove optimal strong approximation results for quadratic forms in at least five variables over the integers [Sar19] and later joint with the second author over $\mathbb{F}_q[t]$ [SZ20]. These results were achieved with suitable choices of weight functions and delicate divisions of the circles in the two settings. A novelty of the paper [SZ20] was the development of a function field version of the stationary phase theorem that was essential for our calculations there (and here in this paper). That being said, in the case of quadratic forms in four variables (both over the integers and over $\mathbb{F}_q[t]$), which is the case of interest for applications to the covering exponent of S^3 and Ramanujan graphs, the results there are suboptimal. It had been observed in [Sar19, Remark 6.8] that if a certain cancellation in a sum involving exponential sums and oscillatory integrals is true, then optimal strong approximation in the case of four variables would also follow. It was the insight of Browning–Kumarasvamy–Steiner [BKS19] that for the problem regarding optimal covering exponents of S^3 —where the quadratic form is a sum of four squares—the stationary phase theorem can be used to reduce the optimal covering exponent to a twisted Linnik–Selberg conjecture over the integers. We also refer the reader to the work of Steiner [Ste19] on this integral twisted Linnik–Selberg conjecture; though that paper lead to some insights regarding the limitations of the Kuznetsov trace formula for this covering exponent problem, the desired results were not obtained, unfortunately. In the case of Morgenstern Ramanujan graphs, there are two main differences with the case of the covering exponent problem. Firstly, as mentioned above, the stationary phase theorem in the setting of function fields had to be developed by the authors in [SZ20] and applied successfully to the computation of the oscillatory integrals in this paper. Secondly, in addition to the infinite place, the finite places also play a role, complicating the precise computation of the exponential sums under consideration. It is in the Morgenstern case where we are able to relate the exponential sums and oscillatory integrals to Kloosterman sums; for a general quadratic form in four variables, the computations are much more complicated and cannot be written simply in terms of Kloosterman sums. Furthermore, most of the main computations in [SZ20] were done using techniques different from those over the integers; those techniques over function fields are also used in this paper.

Remark 4. Though the twisted Linnik–Selberg Conjecture 1.4 would prove the desired Conjecture 1.2, in Conjecture 1.4 any power $|g|^{\theta+\varepsilon}$ with $0 \leq \theta < 1/2$ would allow us to weaken the $\deg f \geq (6+\varepsilon) \deg g + O_\varepsilon(1)$ condition, as will be clear from the proof of the theorem above in Section 6. In turn, this would allow us to decrease the upper bound given for the Morgenstern Ramanujan graphs.

We also prove the following theorem stating that the coefficient $\frac{4}{3}$ cannot be improved upon; see Section 2 for a proof.

Theorem 1.6. *Suppose that $q \equiv 3 \pmod{4}$. There exist infinitely many $g \in \mathbb{F}_q[t]$ such that the Morgenstern Ramanujan graph $X^{q,g}$ is non-bipartite (or bipartite) and*

$$\text{diam}(X^{q,g}) \geq \frac{4}{3} \log_q |X^{q,g}| + O(1),$$

where $O(1)$ is an absolute constant.

This also gives us a new family of $q+1$ -regular non-bipartite (or bipartite) Ramanujan graphs with large diameter and with $q \equiv 3 \pmod{4}$ any prime power.

2. Lower bound on the diameter

In this section, we prove Theorem 1.6. Our argument is similar to the previous argument of the first author in [Sar18, Theorem 1.2].

Proof. Since $q \equiv 3 \pmod{4}$, -1 is a quadratic non-residue in \mathbb{F}_q , and the Morgenstern quadratic form for $\nu = -1$ is

$$F(x_1, x_2, x_3, x_4) := x_1^2 + x_2^2 - (t-1)(x_3^2 + x_4^2).$$

Let $g(t) \in \mathbb{F}_q[t]$ be any irreducible polynomial relatively prime to $t(t-1)$ such that t is a quadratic non-residue in the finite field $\mathbb{F}_{q^{\deg(g)}} := \mathbb{F}[t]/\langle g \rangle$, and -1 is a quadratic residue in $\mathbb{F}_{q^{\deg(g)}}$, (which means $\deg(g)$ is even). Then it follows from the work of Morgenstern that $X^{q,g}$ is isomorphic to the Cayley graph of $\text{PGL}_2(\mathbb{F}_{q^{\deg(g)}})$ generated by $q+1$ generators. The identification is given by the following map sending the quaternion

$$x_1 + ix_2 + jx_3 + kx_4 \mapsto \begin{bmatrix} x_1 - x_2i & x_3 - x_4i \\ (t-1)(x_3 + x_4i) & x_1 + x_2i \end{bmatrix},$$

where i is a choice of $\sqrt{-1}$ in $\mathbb{F}_{q^{\deg(g)}}$. Moreover the quadratic residue of the determinant gives a bipartite structure on $X^{q,g}$. Let $I := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \in X^{q,g}$ and $W := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \in X^{q,g}$. We show that

$$\text{dist}(I, W) \geq \frac{4}{3} \log_q |X^{q,g}| + O(1).$$

Suppose that there exists a path of minimal length h from I to W . Then, by [Mor94, Theorem 4.5] there exists an integral solution $\mathbf{u} := (u_1, \dots, u_4) \in \mathbb{F}_q[t]^4$ to

$$x_1^2 + x_2^2 - (t-1)(x_3^2 + x_4^2) = t^h,$$

where $g \mid \gcd(u_1, u_3, u_4)$, $\gcd(g, u_2) = 1$ and $t-1 \mid \gcd(u_1-1, u_2)$. This implies that $u_2^2 \equiv t^h \pmod{g^2}$. Here, we are using that g is irreducible and relatively prime to t , and so g cannot divide both $\pm t^h$. Since t is a quadratic non-residue mod g and t^h is a square mod g , h is even. Suppose that $h = 2l$. If $l \geq 2 \deg(g)$ then

$$\text{dist}(I, W) = h \geq 4 \deg(g) = \frac{4}{3} \log_q |X^{q,g}| + O(1).$$

So, we suppose that $l < 2 \deg(g)$. We have $u_2^2 \equiv t^h \equiv t^{2l} \pmod{g^2}$. This implies $u_2 \equiv \pm t^l \pmod{g^2}$. We write $u_2 = ag^2 \pm t^l$ for some $a \in \mathbb{F}_q[t]$. For $a \neq 0$, $\deg(ag^2 \pm t^l) \geq 2 \deg(g)$, and we have

$$2l = \deg(F(\mathbf{u})) \geq 2 \deg(u_2) = 2 \deg(ag^2 \pm t^l) \geq 4 \deg(g) > 2l,$$

because -1 is a non-square residue modulo q . This is a contradiction. So $a = 0$, and this implies $u_1 = 0$, which is a contradiction since $t-1 \mid u_1-1$. This proves Theorem 1.6 in the bipartite cases.

Next, we give an infinite family of non-bipartite Morgenstern Ramanujan graphs with the same lower bound on their diameter. Let $r \in \mathbb{F}_q[t]$ be any irreducible polynomial relatively prime to $t(t-1)$ and such that t and -1 are quadratic residues in the finite field $\mathbb{F}_{q^{\deg(r)}} := \mathbb{F}[t]/\langle r \rangle$. Consider the Morgenstern Ramanujan graph $X^{q, (t^2+1/4)r}$. It follows from the work of Morgenstern that $X^{q, (t^2+1/4)r}$ is isomorphic to the Cayley graph of $\text{PSL}_2(\mathbb{F}[t]/\langle (t^2+1/4)r \rangle)$ generated by $q+1$ generators, and $X^{q, (t^2+1/4)r}$ is a non-bipartite Ramanujan graph. Let I and W be as before and define $I' := \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix}$. Let $\sqrt{-1}$ be a square root of -1 in $\mathbb{F}[t]/\langle (t^2+1/4)r \rangle$. Also let $W' := \sqrt{-1}W \in \text{PSL}_2$. We show that

$$\max(\text{dist}(I, I'), \text{dist}(I, W')) \geq \frac{4}{3} \log_q |X^{q, (t^2+1/4)r}| + O(1).$$

Assume to the contrary that $\max(\text{dist}(I, I'), \text{dist}(I, W')) < 4 \deg(r) = \frac{4}{3} \log_q |X^{q, (t^2+1/4)r}| + O(1)$. Since $\text{dist}(I, I') < 4 \deg(r)$, by [Mor94, Theorem 4.5] it follows that there exists an integral solution

$$a_1^2 + a_2^2 - (t-1)(a_3^2 + a_4^2) = t^{h_1}$$

for some $h_1 < 4 \deg(r)$, where $r \mid \gcd(a_2, a_3, a_4)$, $(t-1) \mid \gcd(a_1-1, a_2)$, and at least one of a_3 or a_4 is non-zero. This implies $a_1^2 \equiv t^{h_1} \pmod{r^2}$. We consider two cases: h_1 is even or h_1 is odd. First, suppose that $h_1 = 2l_1$. Then $a_1 \equiv \pm t^{l_1} \pmod{r^2}$. This implies $a_1 = c_1 g^2 \pm t^{l_1}$ for some $c_1 \in \mathbb{F}_q[t]$. Suppose that $c_1 \neq 0$. Then $\deg(a_1) \geq 2 \deg(r)$, and we have

$$h_1 = \deg(a_1^2 + a_2^2 - (t-1)(a_3^2 + a_4^2)) \geq 2 \deg(a_1) \geq 4 \deg(r).$$

This is a contradiction. Hence, $c_1 = 0$. This implies $a_3 = a_4 = 0$, which is also a contradiction. Therefore, $h_1 = 2l_1 + 1$ is odd.

Similarly, since $\text{dist}(I, W') < 4\deg(r)$ by assumption, it follows that there exists an integral solution

$$b_1^2 + b_2^2 - (t-1)(b_3^2 + b_4^2) = t^{h_2}$$

for some $h_2 < 4\deg(r)$, where $r \mid \gcd(b_1, b_3, b_4)$, $(t-1) \mid \gcd(b_1 - 1, b_2)$. By a similar argument it follows that $h_2 = 2l_2 + 1$ is odd and we have $b_2^2 \equiv t^{h_2} \pmod{r^2}$. Therefore, we have

$$\begin{cases} a_1 \equiv 1 \pmod{t-1}, \text{ and } a_1^2 \equiv t^{2l_1+1} \pmod{r^2}, \text{ and } l_1 < 2\deg(r), \\ b_2 \equiv 0 \pmod{t-1}, \text{ and } b_2^2 \equiv t^{2l_2+1} \pmod{r^2}, \text{ and } l_2 < 2\deg(r). \end{cases}$$

Without loss of generality, suppose that $l_1 \geq l_2$. Then, we have

$$a_1 \equiv \pm t^{l_1-l_2} b_2 \pmod{r^2}.$$

Note that $\deg(a_1) < l_1 + 1/2 < 2\deg(r)$ and $\deg(t^{l_1-l_2} b_2) < l_1 + 1/2 < 2\deg(r)$. Hence,

$$a_1 = \pm t^{l_1-l_2} b_2.$$

This contradicts with $\begin{cases} a_1 \equiv 1 \pmod{t-1}, \\ b_2 \equiv 0 \pmod{t-1}. \end{cases}$ This completes the proof of our theorem. \square

3. Recollections on the delta method

The primary purpose of this section is to collect some of the facts related to the delta method over $\mathbb{F}_q[t]$. This section also serves the purpose of setting the notation for the rest of the paper. For details, the reader may consult Section 2 of the authors' paper [SZ20].

Roughly, the delta method is a procedure by which one rewrites the delta function over integral points inside a region as a weighted sum of characters. In this section, we define a weighted sum $N(w, \boldsymbol{\lambda})$ counting the number of integral solutions to the system (1).

3.1. Notation

As in the authors' paper [SZ20], we let $K = \mathbb{F}_q(t)$ and let $\mathcal{O} = \mathbb{F}_q[t]$ be its ring of integers. We denote the prime at infinity t^{-1} by ∞ . We may equip K with the norm at infinity given by

$$|a/b|_\infty := q^{\deg a - \deg b}.$$

Completing K with respect to this norm gives K_∞ . Henceforth, we drop the subscript ∞ from $|\cdot|_\infty$ and write $|\cdot|$ for simplicity. We can also extend the norm to higher dimensions: for every d , the natural norm on K_∞^d is given by $|\mathbf{a}| := \max_i |a_i|$. These norms equip K_∞^d with the metric topology.

Note that we may identify K_∞ with the field

$$\mathbb{F}_q((1/t)) = \left\{ \sum_{i \leq N} a_i t^i : \text{for } a_i \in \mathbb{F}_q \text{ and some } N \in \mathbb{Z} \right\}.$$

The (open) unit ball in this topological space is

$$\mathbb{T} = \{\alpha \in K_\infty : |\alpha| < 1\} = \left\{ \sum_{i \leq -1} a_i t^i : \text{for } a_i \in \mathbb{F}_q \right\}.$$

3.2. Characters

Let $e_q : \mathbb{F}_q \rightarrow \mathbb{C}^*$ be the nontrivial additive character given by sending $a \in \mathbb{F}_q$ to $e_q(a) = \exp(2\pi i \operatorname{tr}(a)/p)$, where $p := \operatorname{char} \mathbb{F}_q$ and $\operatorname{tr} : \mathbb{F}_q \rightarrow \mathbb{F}_p$ is the trace map. From this, we obtain the non-trivial additive character $\psi : K_\infty \rightarrow \mathbb{C}^*$ given by $\psi(\alpha) = e_q(a_{-1})$ for any $\alpha = \sum_{i \leq N} a_i t^i$ in K_∞ . By construction, $\psi|_{\mathcal{O}}$ is trivial. Furthermore, for any $\gamma \in K_\infty$, the map $\alpha \mapsto \psi(\alpha\gamma)$ is also an additive character on K_∞ . A basic lemma that will be useful in our computations is the following.

Lemma 3.1 (Kubota, Lemma 7 of [Kub74]).

$$\sum_{\substack{b \in \mathcal{O} \\ |b| < \widehat{N}}} \psi(\gamma b) = \begin{cases} \widehat{N}, & \text{if } |(\gamma)| < \widehat{N}^{-1}, \\ 0, & \text{otherwise,} \end{cases}$$

for any $\gamma \in K_\infty$ and any integer $N \geq 0$, where (γ) is the part of γ with all degrees negative.

We also have the following

Lemma 3.2 (Kubota, Lemma 1(f) of [Kub74]). Let $Y \in \mathbb{Z}$ and $\gamma \in K_\infty$. Then

$$\int_{|\alpha| < \widehat{Y}} \psi(\alpha\gamma) d\alpha = \begin{cases} \widehat{Y}, & \text{if } |\gamma| < \widehat{Y}^{-1}, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, if we set $Y = 0$, then we obtain the following expression for the delta function on \mathcal{O} :

$$\delta(x) = \int_{\mathbb{T}} \psi(\alpha x) d\alpha,$$

$$\text{where } \delta(x) = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{otherwise.} \end{cases}$$

3.3. The delta function

The idea now is to decompose \mathbb{T} into a disjoint union of balls (with no minor arcs) which is the analogue of Kloosterman's version of the circle method in this function field setting. This is done via the following lemma of Browning and Vishe [BV15, Lemma 4.2].

Lemma 3.3. *For any $Q > 1$ we have a disjoint union*

$$\mathbb{T} = \bigsqcup_{\substack{r \in \mathcal{O} \\ |r| \leq \widehat{Q} \\ r \text{ monic}}} \bigsqcup_{\substack{a \in \mathcal{O} \\ |a| < |r| \\ (a,r)=1}} \left\{ \alpha \in \mathbb{T} : |r\alpha - a| < \widehat{Q}^{-1} \right\}.$$

The following follows from Lemma 3.3; see [SZ20, Lemma 2.4].

Lemma 3.4. *Let $Q \geq 1$ and $n \in \mathcal{O}$. We have*

$$\delta(n) = \frac{1}{\widehat{Q}^2} \sum_{\substack{r \in \mathcal{O} \\ |r| \leq \widehat{Q} \\ r \text{ monic}}} \sum_{|a| < |r|}^* \psi\left(\frac{an}{r}\right) h\left(\frac{r}{t\widehat{Q}}, \frac{n}{t^2\widehat{Q}}\right), \quad (5)$$

where we henceforth put $\sum_{|a| < |r|}^* := \sum_{\substack{a \in \mathcal{O} \\ |a| < |r| \\ (a,r)=1}}^*$, and h is only defined for $x \neq 0$ as:

$$h(x, y) = \begin{cases} |x|^{-1} & \text{if } |y| < |x| \\ 0 & \text{otherwise.} \end{cases}$$

Moreover,

$$\frac{1}{\widehat{Q}^2} h\left(\frac{r}{t\widehat{Q}}, \frac{n}{t^2\widehat{Q}}\right) = \int_{|\alpha| < |r|^{-1}\widehat{Q}^{-1}} \psi(\alpha n) d\alpha.$$

3.4. Smooth sum $N(w, \lambda)$

Let

$$w(\mathbf{x}) = \begin{cases} 1 & \text{if } |\mathbf{x}| \leq |f|^{1/2}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that

$$w(g\mathbf{t} + \boldsymbol{\lambda}) = \begin{cases} 1 & \text{if } |\mathbf{t}| < \widehat{R}, \\ 0 & \text{otherwise,} \end{cases}$$

where $R := \lfloor \deg(f)/2 - \deg(g) + 1 \rfloor$. Assume that $\mathbf{x} \in \mathcal{O}^d$ satisfies the conditions $F(\mathbf{x}) = f$ and $\mathbf{x} \equiv \boldsymbol{\lambda} \pmod{g}$. We uniquely write $\mathbf{x} = g\mathbf{t} + \boldsymbol{\lambda}$, where $\mathbf{t} \in \mathcal{O}^d$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$ for λ_i of degree strictly less than that of g . Define

$$k := \frac{f - F(\boldsymbol{\lambda})}{g}. \quad (6)$$

If $F(\mathbf{x}) = f$, then $g^2 F(\mathbf{t}) + 2g\boldsymbol{\lambda}^T A\mathbf{t} = f - F(\boldsymbol{\lambda})$ which implies that $g|2\boldsymbol{\lambda}^T A\mathbf{t} - k$. Then, $F(\mathbf{t}) + \frac{1}{g}(2\boldsymbol{\lambda}^T A\mathbf{t} - k) = 0$. We also define

$$G(\mathbf{t}) := \frac{F(g\mathbf{t} + \boldsymbol{\lambda}) - f}{g^2} = F(\mathbf{t}) + \frac{1}{g}(2\boldsymbol{\lambda}^T A\mathbf{t} - k).$$

Finally, we define

$$N(w, \boldsymbol{\lambda}) := \sum_{\mathbf{t}} w(g\mathbf{t} + \boldsymbol{\lambda}) \delta(G(\mathbf{t})),$$

where $\mathbf{t} \in \mathcal{O}^d$. Note that $N(w, \boldsymbol{\lambda})$ is the weighted number of $\mathbf{x} \in \mathcal{O}^d$ satisfying the conditions $F(\mathbf{x}) = f$ and $\mathbf{x} \equiv \boldsymbol{\lambda} \pmod{g}$. We apply the delta expansion in (5) to $\delta(G(\mathbf{t}))$ and follow the computations in [SZ20, Section 2.4], and obtain

$$N(w, \boldsymbol{\lambda}) = \frac{1}{|g|\widehat{Q}^2} \sum_{\substack{r \in \mathcal{O} \\ |r| \leq \widehat{Q} \\ r \text{ monic}}} \sum_{\mathbf{c} \in \mathcal{O}^d} |gr|^{-d} S_{g,r}(\mathbf{c}) I_{g,r}(\mathbf{c}), \quad (7)$$

where $I_{g,r}(\mathbf{c})$ and $S_{g,r}(\mathbf{c})$ are defined by

$$I_{g,r}(\mathbf{c}) := \int_{K_\infty^d} h\left(\frac{r}{tQ}, \frac{G(\mathbf{t})}{t^2Q}\right) w(g\mathbf{t} + \boldsymbol{\lambda}) \psi\left(\frac{\langle \mathbf{c}, \mathbf{t} \rangle}{gr}\right) dt, \quad (8)$$

and

$$S_{g,r}(\mathbf{c}) := \sum_{\substack{\ell \in \mathcal{O} \\ |\ell| < |g|}} \sum_{|a| < |r|}^* S_{g,r}(a, \ell, \mathbf{c}) \quad (9)$$

with

$$S_{g,r}(a, \ell, \mathbf{c}) := \sum_{\mathbf{b} \in (\mathcal{O}/(gr))^d} \psi\left(\frac{(a + r\ell)(2\boldsymbol{\lambda}^T A\mathbf{b} - k) + agF(\mathbf{b}) - \langle \mathbf{c}, \mathbf{b} \rangle}{gr}\right). \quad (10)$$

We henceforth assume that r is always monic without saying so. In the next two sections, we give explicit formulas for $S_{g,r}$ and $I_{g,r}$ when our quadratic form is the Morgenstern quadratic form.

4. The oscillatory integrals $I_{g,r}(\mathbf{c})$

In this section, we give explicit formulas for the oscillatory integrals $I_{g,r}(\mathbf{c})$ in terms of the Kloosterman sums at infinity. Suppose that $F(\mathbf{t}) := \sum_i \eta_i t_i^2$ is the Morgenstern quadratic form and $F^*(\mathbf{c}) := \sum_i c_i^2 / \eta_i$ is its dual. Let $\kappa := \max_i |\frac{c_i}{g}|$. In this section, we assume that $Q = R + 1$.

Proposition 4.1. *Suppose that $|r| \leq \widehat{Q}$. For the Morgenstern quadratic form, we have*

$$I_{g,r}(\mathbf{c}) = \begin{cases} 0, & \text{if } \kappa \geq \widehat{Q}/\widehat{R}, \\ 0, & \kappa = \frac{|r|}{\widehat{R}}, \deg(f) \text{ is even,} \\ & \text{and } \widehat{Q}q^{-3} < |r| \leq \widehat{Q}, \\ I_{g,r}(\mathbf{0}), & \text{if } \kappa < \frac{|r|}{\widehat{R}}, \\ I_{g,r}(\mathbf{0}), & \text{if } \kappa = \frac{|r|}{\widehat{R}}, \\ & \max(|c_3|, |c_4|) > \max(|c_1|, |c_2|), \\ & \text{and } \deg(f) \text{ is even, } |r| \leq \widehat{Q}q^{-3}, \\ -\widehat{Q}^2 |g|^2 |r|^2 |F^*(\mathbf{c})|^{-1} \text{Kl}_\infty \left(\psi, \frac{kF^*(\mathbf{c})}{4r^2 g^3} \right), & \text{otherwise.} \end{cases}$$

We give the proof of Proposition 4.1 at the end of this section. We proceed by citing some general results from [SZ20] that are not specific to the Morgenstern quadratic form. In particular, we do not restrict the number of variables to $d = 4$ for the moment. Recall that

$$G(\mathbf{t}) := \frac{F(g\mathbf{t} + \boldsymbol{\lambda}) - f}{g^2} = F(\mathbf{t}) + \frac{1}{g}(2\boldsymbol{\lambda}^T A\mathbf{t} - k),$$

where $k = \frac{f - F(\boldsymbol{\lambda})}{g}$. We have

$$I_{g,r}(\mathbf{c}) = \int_{K_\infty^d} h\left(\frac{r}{t^Q}, \frac{G(\mathbf{t})}{t^{2Q}}\right) w(g\mathbf{t} + \boldsymbol{\lambda}) \psi\left(\frac{\langle \mathbf{c}, \mathbf{t} \rangle}{gr}\right) d\mathbf{t} = \int_{\substack{|\mathbf{t}| < \widehat{R} \\ |G(\mathbf{t})| < \widehat{Q}|r|}} \frac{\widehat{Q}}{|r|} \psi\left(\frac{\langle \mathbf{c}, \mathbf{t} \rangle}{gr}\right) d\mathbf{t}. \quad (11)$$

We cite [SZ20, Lemma 5.5].

Lemma 4.2. *Let Q and R be as above, and suppose that $|\mathbf{t}| < \widehat{R}$. Then $|G(\mathbf{t})| < \widehat{Q}|r|$ is equivalent to $|F(\mathbf{t}) - k/g| < \widehat{Q}|r|$. Moreover, if $|G(\mathbf{t})| < \widehat{Q}|r|$, then $|G(\mathbf{t} + \boldsymbol{\zeta})| < \widehat{Q}|r|$ for every $\boldsymbol{\zeta} \in K_\infty^d$, where $|\boldsymbol{\zeta}| \leq \min(|r|, \widehat{R})$.*

We cite [SZ20, Lemma 5.6].

Lemma 4.3. Suppose that $\kappa \geq \widehat{Q}/\widehat{R}$ and $|r| \leq \widehat{Q}$. Then, $I_{g,r}(\mathbf{c}) = 0$.

Lemma 4.4. Suppose that $\kappa = \frac{|r|}{\widehat{R}}$, $\deg(f)$ is even, and $\widehat{Q}q^{-3} < |r| \leq \widehat{Q}$. Then, $I_{g,r}(\mathbf{c}) = 0$.

Proof. Since $\deg(f)$ is even, $R = \frac{\deg(f)}{2} - \deg(g) + 1$ and $Q = \frac{\deg(f)}{2} - \deg(g) + 2$. Hence, $\widehat{Q}|r| > |k/g|$, and by Lemma 4.3, $|G(\mathbf{t})| < \widehat{Q}|r|$ is equivalent to

$$|F(\mathbf{t})| < \widehat{Q}|r|.$$

Since $\widehat{Q}q^{-3} < |r| \leq \widehat{Q}$ and $|\mathbf{t}| < \widehat{R}$,

$$|F(\mathbf{t})| \leq \widehat{2R-1} < \widehat{Q}|r|.$$

So, the inequality $|G(\mathbf{t})| < \widehat{Q}|r|$ is satisfied automatically, and we have

$$I_{g,r}(\mathbf{c}) = \frac{\widehat{Q}}{|r|} \int_{|\mathbf{t}| < \widehat{R}} \psi\left(\frac{\langle \mathbf{c}, \mathbf{t} \rangle}{gr}\right) d\mathbf{t} = 0,$$

where we used $\kappa = \frac{|r|}{\widehat{R}}$. This completes the proof of our lemma. \square

Lemma 4.5. Suppose that either

- (1) $\kappa < \frac{|r|}{\widehat{R}}$,
- (2) $\kappa = \frac{|r|}{\widehat{R}}$, $\max(|c_3|, |c_4|) > \max(|c_1|, |c_2|)$, $\deg(f)$ is even, and $|r| \leq \widehat{Q}q^{-3}$,

then $I_{g,r}(\mathbf{c}) = I_{g,r}(\mathbf{0})$.

Proof. Suppose (1). Since $\max_i(|c_i|) < \frac{|gr|}{\widehat{R}}$ and $|\mathbf{t}| < \widehat{R}$, $\psi\left(\frac{\langle \mathbf{c}, \mathbf{t} \rangle}{gr}\right) = 1$. Hence, we have

$$I_{g,r}(\mathbf{c}) = \int_{\substack{|\mathbf{t}| < \widehat{R} \\ |G(\mathbf{t})| < \widehat{Q}|r|}} \frac{\widehat{Q}}{|r|} d\mathbf{t} = I_{g,r}(\mathbf{0}).$$

Suppose (2) and that $\widehat{Q}|r| \leq |k/g|$. By Lemma 4.2,

$$|F(\mathbf{t}) - k/g| < \widehat{Q}|r| \leq |k/g|.$$

Hence, the top degree of $F(\mathbf{t})$ and k/g are the same. Since $\deg(f)$ is even, $\deg(k/g) = \deg(f/g^2)$ is even. The top degree of $F(\mathbf{t})$ is even as well. Hence, $\max(|t_3|, |t_4|) < \max(|t_1|, |t_2|)$. Hence, $|\langle \mathbf{c}, \mathbf{t} \rangle| < |c||\mathbf{t}|$, and

$$\psi\left(\frac{\langle \mathbf{c}, \mathbf{t} \rangle}{gr}\right) = 1,$$

which implies $I_{g,r}(\mathbf{c}) = I_{g,r}(\mathbf{0})$. Finally, suppose (2) and that $\widehat{Q}|r| > |k/g|$. This implies that $|r| = \widehat{Q}q^{-3}$. By Lemma 4.3, $|G(\mathbf{t})| < \widehat{Q}|r|$ is equivalent to $|F(\mathbf{t})| < \widehat{Q}|r| = \widehat{R}^2q^{-1}$. We have

$$2 \deg(\max(|t_3|, |t_4|)) + 1 \leq \deg(F(\mathbf{t})).$$

Hence,

$$\deg(\max(|t_3|, |t_4|)) \leq \frac{\deg(F(\mathbf{t})) - 1}{2} \leq R - 2.$$

Therefore, $\psi\left(\frac{\langle \mathbf{c}, \mathbf{t} \rangle}{gr}\right) = 1$. This implies $I_{g,r}(\mathbf{c}) = I_{g,r}(\mathbf{0})$. \square

4.1. Stationary phase theorem over function fields

In [SZ20, Proposition 4.5], we proved a version of the stationary phase theorem in the function fields setting. We proceed by defining some new notations and cite a special case of [SZ20, Proposition 4.5]. Let $h \in K_\infty$ and define

$$\mathcal{G}(h) := \begin{cases} \min(|h|_\infty^{-1/2}, 1) & \text{if } \text{ord}(h) \text{ is even,} \\ |h|_\infty^{-1/2} \varepsilon_h & \text{if } \text{ord}(h) \geq 1 \text{ and is odd,} \\ 1 & \text{otherwise,} \end{cases} \quad (12)$$

where $\varepsilon_h := \frac{G(h)}{|G(h)|}$ and $G(h) := \sum_{x \in \mathbb{F}_q} e_q(a_h x^2)$ is the gauss sum associated to a_h the top degree coefficient of h . We cite [SZ20, Lemma 4.6].

Lemma 4.6. *For every $f \in K_\infty$, we have*

$$\int_{\mathbb{T}} \psi(fu^2) du = \mathcal{G}(f).$$

For $\alpha \in K$ and $a \in \mathbb{Z}$, define

$$B_\infty(\psi, a, \alpha) := \int_{|x|_\infty = \widehat{a}} \psi\left(\frac{\alpha}{x} + x\right) dx.$$

We write $\alpha = t^{2a+b} \alpha' (1 + \tilde{\alpha})$ and $x = t^a x' (1 + \tilde{x})$ for unique $\tilde{\alpha}, \tilde{x} \in \mathbb{T}$ and $\alpha', x' \in \mathbb{F}_q$. Note that for $b = 0$, we have $B_\infty(\psi, a, \alpha) = \text{Kl}_\infty(\psi, \alpha)$. Let $\text{Kl}(\alpha, \mathbb{F}_q) := \sum_{x \in \mathbb{F}_q^*} e_q\left(\frac{\alpha}{x} + x\right)$. We cite [SZ20, Lemma 5.8].

Lemma 4.7. *We have*

$$B_{\infty}(\psi, a, \alpha) = \begin{cases} (q-1)\hat{a} & \text{if } \max(a+b, a) < -1, \text{ and } b \neq 0, \\ -\hat{a} & \text{if } \max(a+b, a) = -1, \text{ and } b \neq 0, \\ 0 & \text{if } \max(a+b, a) > -1, \text{ and } b \neq 0. \end{cases}$$

$$\text{Kl}_{\infty}(\psi, \alpha) = \begin{cases} (q-1)\hat{a} & \text{if } a < -1, \\ \hat{a}\text{Kl}(\alpha', \mathbb{F}_q) & \text{if } a = -1, \\ \hat{a} \sum_{x'^2=\alpha'} \psi\left(2t^a x'(1+\tilde{\alpha})^{1/2}\right) \mathcal{G}(2x't^a) & \text{if } \alpha' \text{ is a quadratic residue,} \\ 0 & \text{if } \alpha' \text{ is not a quadratic residue.} \end{cases}$$

Proof of Proposition 4.1. In this proof, we assume that we are working with the Morgenstern quadratic form. By Lemma 4.3, we have $I_{g,r}(\mathbf{c}) = 0$ for $\kappa \geq \widehat{Q}/\widehat{R}$ and $|r| \leq \widehat{Q}$. By assuming the conditions of Lemma 4.5, it follows that $I_{g,r}(\mathbf{c}) = I_{g,r}(\mathbf{0})$. Hence, the remaining cases correspond to

- (1) $\frac{|r|}{\widehat{R}} < \kappa$, or
- (2) $\kappa = \frac{|r|}{\widehat{R}}$, and $\max(|c_3|, |c_4|) \leq \max(|c_1|, |c_2|)$ or $\deg(f)$ is odd,

and we proceed to conclude the proposition in these two cases.

By Lemma 4.2, $|G(\mathbf{t})| < \widehat{Q}|r|$ is equivalent to $|F(\mathbf{t}) - k/g| < \widehat{Q}|r|$ for $|\mathbf{t}| < \widehat{R}$. By Lemma 3.2, we have

$$\int_{\mathbb{T}} \psi\left(\frac{\alpha}{rt^Q}(F(\mathbf{t}) - k/g)\right) d\alpha = \begin{cases} 1, & \text{if } |F(\mathbf{t}) - k/g| < \widehat{Q}|r|, \\ 0, & \text{otherwise.} \end{cases}$$

We replace the above integral for detecting $|F(\mathbf{t}) - k/g| < \widehat{Q}|r|$. Hence, by (11)

$$I_{g,r}(\mathbf{c}) = \frac{\widehat{Q}}{|r|} \int_{\mathbb{T}} \int_{|\mathbf{t}| < \widehat{R}} \psi\left(\frac{\langle \mathbf{c}, \mathbf{t} \rangle}{gr} + \frac{\alpha}{rt^Q}(F(\mathbf{t}) - k/g)\right) dt d\alpha.$$

Recall that $F(\mathbf{t}) = \sum_{\eta_i} \eta_i t_i^2$. We have

$$\frac{\langle \mathbf{c}, \mathbf{t} \rangle}{gr} + \frac{\alpha}{rt^Q}(F(\mathbf{t}) - k/g) = \frac{-\alpha k}{rgt^Q} + \frac{1}{r} \left(\sum_i \frac{c_i t_i}{g} + \frac{\alpha \eta_i t_i^2}{t^Q} \right).$$

Hence, $I_{g,r}(\mathbf{c}) = \frac{\widehat{Q}}{|r|} \int_{\mathbb{T}} \psi\left(\frac{-\alpha k}{rgt^Q}\right) I_{g,r}(\alpha, \mathbf{c}) d\alpha$, where

$$I_{g,r}(\alpha, \mathbf{c}) := \prod_{i=1}^4 \int_{|t_i| < \widehat{R}} \psi\left(\frac{1}{r} \left(\sum_i \frac{c_i t_i}{g} + \frac{\alpha \eta_i t_i^2}{t^Q} \right)\right) dt_i.$$

The phase function has a critical point at $t_i = \frac{-c_i t^Q}{2g\eta_i \alpha}$. This critical point is inside the domain $|t_i| < \widehat{R}$, if $|\alpha| > \kappa_i$, where $\kappa_i := \frac{|c_i| \widehat{Q}}{|g| |\eta_i| \widehat{R}}$. Given $\alpha \in \mathbb{T}$, we partition the indices into:

$$CR := \{1 \leq i \leq 4 : |\alpha| > \kappa_i\},$$

$$NCR := \{1 \leq i \leq 4 : |\alpha| \leq \kappa_i\}.$$

For $i \in NCR$, we change the variables to $v_i = t_i + \frac{\alpha \eta_i g}{c_i t^Q} t_i^2$. Note that v_i is an analytic map in terms of t_i and $\frac{\partial v_i}{\partial t_i}(0) = 1$. Hence, by [SZ20, Proposition 4.2], v_i is a bijection from $|t_i| < \widehat{R}$ to $|v_i| < \widehat{R}$. For $i \in CR$, we change the variables to $w_i = t_i + \frac{c_i t^Q}{2g\eta_i \alpha}$. By [SZ20, Section 4], we have

$$I_{g,r}(\alpha, \mathbf{c}) = \prod_{i \in NCR} \int_{|v_i| < \widehat{R}} \psi\left(\frac{c_i v_i}{gr}\right) dv_i \times \prod_{i \in CR} \psi\left(-\frac{t^Q c_i^2}{4rg^2 \eta_i \alpha}\right) \int_{|w_i| < \widehat{R}} \psi\left(\frac{\alpha \eta_i}{rt^Q} w_i^2\right) dw_i.$$

By Lemma 3.2, Lemma 4.6, we have

$$\int_{|v_i| < \widehat{R}} \psi\left(\frac{c_i v_i}{gr}\right) dv_i = \begin{cases} \widehat{R}, & \text{if } \frac{|c_i|}{|g|} < \frac{|r|}{\widehat{R}}, \\ 0, & \text{otherwise,} \end{cases}$$

$$\int_{|w_i| < \widehat{R}} \psi\left(\frac{\alpha \eta_i}{rt^Q} w_i^2\right) dw_i = \widehat{R} \mathcal{G}\left(\frac{\alpha \eta_i t^{2R}}{rt^Q}\right).$$

Therefore,

$$I_{g,r}(\alpha, \mathbf{c}) = \widehat{R}^4 \prod_{i=1}^4 \left(\delta_{|\alpha| \leq \kappa_i} \delta_{\frac{|c_i|}{|g|} < \frac{|r|}{\widehat{R}}} + \delta_{\kappa_i < |\alpha| < 1} \psi\left(-\frac{t^Q c_i^2}{4rg^2 \eta_i \alpha}\right) \mathcal{G}\left(\frac{\alpha \eta_i t^{2R}}{rt^Q}\right) \right). \quad (13)$$

By our assumption, we have $\frac{|r|}{\widehat{R}} \leq \kappa < \widehat{Q}/\widehat{R}$. Note that $Q = R + 1$ and $q\kappa \geq \max \kappa_i \geq \kappa$. This implies that $I_{g,r}(\alpha, \mathbf{c}) = 0$ for every $|\alpha| \leq \kappa$. For $1 > |\alpha| > \kappa \geq \frac{|r|}{\widehat{R}}$, we have

$$I_{g,r}(\alpha, \mathbf{c}) = \widehat{R}^4 \prod_{i=1}^4 \delta_{\kappa_i < |\alpha| < 1} \psi\left(-\frac{t^Q c_i^2}{4rg^2 \eta_i \alpha}\right) \mathcal{G}\left(\frac{\alpha \eta_i t^{2R}}{rt^Q}\right).$$

Hence,

$$I_{g,r}(\mathbf{c}) = \frac{\widehat{Q} \widehat{R}^4}{|r|} \sum_{\substack{\kappa \leq \widehat{l} < 1 \\ |\alpha| = \widehat{l}}} \int \psi\left(\frac{-\alpha k}{rgt^Q}\right) \prod_{i=1}^4 \delta_{\kappa_i < |\alpha| < 1} \psi\left(-\frac{t^Q c_i^2}{4rg^2 \eta_i \alpha}\right) \mathcal{G}\left(\frac{\alpha \eta_i t^{2R}}{rt^Q}\right) d\alpha.$$

By (12) and considering the sign of the Gauss sums, we have for the Morgenstern quadratic form above the equality

$$\prod_i \mathcal{G} \left(\frac{\alpha \eta_i t^{2R}}{rt^Q} \right) = - \prod_i \min \left(1, \left(\frac{\widehat{l} \widehat{R}^2 |\eta_i|}{|r| \widehat{Q}} \right)^{-1/2} \right),$$

a quantity dependent not on α itself, but on the norm of α which is \widehat{l} . Hence,

$$I_{g,r}(\mathbf{c}) = - \frac{\widehat{Q} \widehat{R}^4}{|r|} \sum_{\kappa \leq \widehat{l} < 1} \prod_i \delta_{\kappa_i < \widehat{l}} \min \left(1, \left(\frac{\widehat{l} \widehat{R}^2 |\eta_i|}{|r| \widehat{Q}} \right)^{-1/2} \right) \int_{|\alpha|=\widehat{l}} \psi \left(\frac{-\alpha k}{rgt^Q} \right) \psi \left(-\frac{t^Q F^*(\mathbf{c})}{4rg^2 \alpha} \right) d\alpha,$$

where $F^*(\mathbf{c}) = \sum_i \frac{c_i^2}{\eta_i}$. Since we assume either

- (1) $\frac{|r|}{R} < \kappa$,
- (2) $\kappa = \frac{|r|}{R}$, and $\max(|c_3|, |c_4|) \leq \max(|c_1|, |c_2|)$ or $\deg(f)$ is odd,

we have

$$\deg \left(\frac{k F^*(\mathbf{c})}{4r^2 g^3} \right) \geq -2.$$

Let $a, b \in \mathbb{Z}$ where

$$a = l + \deg \left(\frac{k}{rgt^Q} \right), \text{ and } 2a + b = \deg \left(\frac{k F^*(\mathbf{c})}{4r^2 g^3} \right).$$

Since $2a + b \geq -2$, either $b = 0$ or $\max(a, a + b) > -1$. By Lemma 4.7, we have

$$\begin{aligned} \int_{|\alpha|=\widehat{l}} \psi \left(\frac{-\alpha k}{rgt^Q} \right) \psi \left(-\frac{t^Q F^*(\mathbf{c})}{4rg^2 \alpha} \right) d\alpha &= \left| \frac{rgt^Q}{k} \right| B_\infty \left(\psi, l + \deg \left(\frac{k}{rgt^Q} \right), \frac{k F^*(\mathbf{c})}{4r^2 g^3} \right) \\ &= \begin{cases} \left| \frac{rgt^Q}{k} \right| \text{Kl}_\infty \left(\psi, \frac{k F^*(\mathbf{c})}{4r^2 g^3} \right) & \text{if } 2l = \deg \left(\frac{t^{2Q} F^*(\mathbf{c})}{kg} \right), \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Note that if $2l = \deg \left(\frac{t^{2Q} F^*(\mathbf{c})}{kg} \right)$, then $\widehat{l} > \kappa_i$, and we have

$$I_{g,r}(\mathbf{c}) = -\widehat{Q}^2 |g|^2 |r|^2 |F^*(\mathbf{c})|^{-1} \text{Kl}_\infty \left(\psi, \frac{k F^*(\mathbf{c})}{4r^2 g^3} \right). \quad \square$$

5. The exponential sums $S_{g,r}(\mathbf{c})$

In this section, we explicitly compute our exponential sums. Though our computations can be generalized to all non-degenerate quadratic forms, we focus here on the special case of the Morgenstern quadratic form

$$F(x_1, x_2, x_3, x_4) := \eta_1 x_1^2 + \eta_2 x_2^2 + \eta_3 x_3^2 + \eta_4 x_4^2$$

over $\mathbb{F}_q[t]$, where $\eta_1 = 1$, $\eta_2 = -\nu$, $\eta_3 = -(t-1)$, $\eta_4 = \nu(t-1)$ and $\nu \in \mathbb{F}_q$ is not a square. Its dual quadratic form F^* is obtained by inverting the coefficients η_j . Focusing on the Morgenstern quadratic form is no restriction in our case since we are primarily interested in proving upper bounds for the diameters of Morgenstern Ramanujan graphs. Throughout this section, we let $A = \text{diag}(\eta_1, \dots, \eta_4)$ be the diagonal matrix associated to this quadratic form. Also, by Lemma 3.2 of [SZ20], $S_{g,r}(\mathbf{c}) = 0$ except possibly when $\mathbf{c} \equiv 2\beta(\mathbf{c})A\boldsymbol{\lambda} \pmod{g}$ for some $\beta(\mathbf{c}) \in \mathcal{O}$.

Proposition 5.1. *For the Morgenstern quadratic form above and $g \in \mathbb{F}_q[t]$ not divisible by $t-1$, we have that when $\gcd(r, t-1) \nmid c_3, c_4$, then $S_{g,r}(\mathbf{c})$ is equal to*

$$\begin{aligned} & \frac{|g|^4}{|m|^2} (|\gcd(r, t-1)| \tau_r \tau_{r/(r, t-1)})^2 \psi \left(\frac{-\overline{mr} \beta(\mathbf{c}) \left(\frac{f-F(\boldsymbol{\lambda})}{m} \right) - \overline{m^2 r} \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{(g/m)^2} \right) \psi \left(\frac{\langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2 r} \right) \\ & \cdot \sum_{s \in \mathcal{O}/(m)} \psi \left(\frac{-s \overline{g/m} \beta(\mathbf{c})}{m} \right) \text{Kl}_{m^2 r} \left(\overline{g/m} f - mrs, \frac{1}{4} \overline{g/m}^3 F^*(\mathbf{c}) \right), \end{aligned}$$

where $m := (g, r^\infty)$. If $\gcd(r, t-1) \nmid c_3, c_4$, then $S_{g,r}(\mathbf{c}) = 0$. Note that when \mathbf{c} is such that $|\mathbf{c}| < |gr|$ (when $r \neq 1$ and $\kappa < \widehat{Q}/\widehat{R} = q$, for example), then $\psi \left(\frac{\langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2 r} \right) = 1$.

Proof. Recall that, the exponential sums for quadratic forms in $d = 4$ variables are

$$\begin{aligned} S_{g,r}(\mathbf{c}) &:= \sum_{\substack{\ell \in \mathcal{O} \\ |\ell| < |g|}} \sum_{|a| < |r|}^* \sum_{\mathbf{b} \in (\mathcal{O}/(gr))^4} \psi \left(\frac{(a + r\ell)(2\boldsymbol{\lambda}^T A\mathbf{b} - k) + agF(\mathbf{b}) - \langle \mathbf{c}, \mathbf{b} \rangle}{gr} \right) \\ &= \sum_{\substack{\ell \in \mathcal{O} \\ |\ell| < |g|}} \sum_{|a| < |r|}^* \sum_{\mathbf{b} \in (\mathcal{O}/(gr))^4} \psi \left(\frac{(a + r\ell)(2\boldsymbol{\lambda}^T A\mathbf{b} - k + gF(\mathbf{b})) - \langle \mathbf{c}, \mathbf{b} \rangle}{gr} \right). \end{aligned}$$

Summation over ℓ is zero except possibly when $g|2\boldsymbol{\lambda}^T A\mathbf{b} - k$, and so we may rewrite the latter as

$$S_{g,r}(\mathbf{c}) = \sum_{\substack{|a| < |gr| \\ (a,r)=1}} \sum_{\substack{\mathbf{b} \in (\mathcal{O}/(gr))^4 \\ g|2\boldsymbol{\lambda}^T A\mathbf{b} - k}} \psi \left(\frac{a(2\boldsymbol{\lambda}^T A\mathbf{b} - k + gF(\mathbf{b})) - \langle \mathbf{c}, \mathbf{b} \rangle}{gr} \right).$$

Let $m := (g, r^\infty)$. Using this expression for $S_{g,r}(\mathbf{c})$, we may rewrite the exponential sum in terms of m as follows. First note that $gr = \frac{g}{m} \cdot (mr)$. Additionally, $(a, r) = 1$ is equivalent to $(a, mr) = 1$. Also, since $(m, \frac{g}{m}) = 1$, the condition $g|2\lambda^T A\mathbf{b} - k$ is equivalent to the pair of conditions $\frac{g}{m}|2\lambda^T A\mathbf{b} - k$ and $m|2\lambda^T A\mathbf{b} - k$. Therefore, we have

$$\begin{aligned} S_{g,r}(\mathbf{c}) &= \sum_{\substack{|a| < |gr| \\ (a,r)=1}} \sum_{\substack{\mathbf{b} \in (\mathcal{O}/(gr))^4 \\ g|2\lambda^T A\mathbf{b} - k}} \psi \left(\frac{a(2\lambda^T A\mathbf{b} - k + gF(\mathbf{b})) - \langle \mathbf{c}, \mathbf{b} \rangle}{gr} \right) \\ &= \frac{1}{|m|} \sum_{\substack{|a| < |\frac{g}{m} \cdot (mr)| \\ (a,mr)=1}} \sum_{\substack{\mathbf{b} \in (\mathcal{O}/(\frac{g}{m} \cdot (mr)))^4 \\ \frac{g}{m}|2\lambda^T A\mathbf{b} - k}} \\ &\quad \cdot \sum_{|s| < |m|} \psi \left(\frac{a(2\lambda^T A\mathbf{b} - k + gF(\mathbf{b})) - \langle \mathbf{c}, \mathbf{b} \rangle}{gr} \right) \psi \left(\frac{s(2\lambda^T A\mathbf{b} - k)}{m} \right) \\ &= \frac{1}{|m|} \sum_{|s| < |m|} \sum_{\substack{|a| < |\frac{g}{m} \cdot (mr)| \\ (a,mr)=1}} \\ &\quad \cdot \sum_{\substack{\mathbf{b} \in (\mathcal{O}/(\frac{g}{m} \cdot (mr)))^4 \\ \frac{g}{m}|2\lambda^T A\mathbf{b} - k}} \psi \left(\frac{(a + \frac{grs}{m})(2\lambda^T A\mathbf{b} - k + gF(\mathbf{b})) - \langle \mathbf{c}, \mathbf{b} \rangle}{gr} \right) \\ &= \sum_{\substack{|a| < |\frac{g}{m} \cdot (mr)| \\ (a,mr)=1}} \sum_{\substack{\mathbf{b} \in (\mathcal{O}/(\frac{g}{m} \cdot (mr)))^4 \\ \frac{g}{m}|2\lambda^T A\mathbf{b} - k}} \psi \left(\frac{a(2\lambda^T A\mathbf{b} - k + gF(\mathbf{b})) - \langle \mathbf{c}, \mathbf{b} \rangle}{\frac{g}{m} \cdot (mr)} \right). \end{aligned}$$

Since $\frac{g}{m}$ and mr are coprime, we may write

$$a = (mr)a_1 + \frac{g}{m}a_2$$

and

$$\mathbf{b} = (mr)\mathbf{b}_1 + \frac{g}{m}\mathbf{b}_2,$$

where a_1 ranges modulo $\frac{g}{m}$, a_2 modulo mr coprime to mr , \mathbf{b}_1 modulo $\frac{g}{m}$, \mathbf{b}_2 modulo mr . Furthermore, we also have the condition that $\frac{g}{m}|2\lambda^T A\mathbf{b} - k$ which is equivalent to the condition that $\frac{g}{m}|2mr\lambda^T A\mathbf{b}_1 - k$. Then

$$\begin{aligned} &\psi \left(\frac{a(2\lambda^T A\mathbf{b} - k + gF(\mathbf{b})) - \langle \mathbf{c}, \mathbf{b} \rangle}{\frac{g}{m} \cdot (mr)} \right) \\ &= \psi \left(\frac{a_1(2\lambda^T A(mr)\mathbf{b}_1) - k - \langle \overline{mr}\mathbf{c}, mr\mathbf{b}_1 \rangle}{\frac{g}{m}} \right) \end{aligned}$$

$$\cdot \psi \left(\frac{a_2(2\lambda^T A(\frac{g}{m}\mathbf{b}_2) - k + gF(\frac{g}{m}\mathbf{b}_2)) - \langle \overline{g/m}\mathbf{c}, (g/m)\mathbf{b}_2 \rangle}{mr} \right).$$

Since $\gcd(mr, \frac{g}{m}) = 1$, when summing over \mathbf{b}_1 and \mathbf{b}_2 we may replace $mr\mathbf{b}_1$ and $\frac{g}{m}\mathbf{b}_2$ with \mathbf{b}_1 and \mathbf{b}_2 , respectively. Consequently, we have

$$\begin{aligned} S_{g,r}(\mathbf{c}) &= \sum_{|a_1| < |\frac{g}{m}|} \sum_{\substack{\mathbf{b}_1 \in (\mathcal{O}/(g/m))^4 \\ \frac{g}{m}|2\lambda^T A\mathbf{b}_1 - k}} \psi \left(\frac{a_1(2\lambda^T A(\mathbf{b}_1) - k) - \langle \overline{mr}\mathbf{c}, \mathbf{b}_1 \rangle}{\frac{g}{m}} \right) \\ &\cdot \sum_{|a_2| < |mr|}^* \sum_{\mathbf{b}_2 \in (\mathcal{O}/(mr))^4} \psi \left(\frac{a_2(2\lambda^T A(\mathbf{b}_2) - k + gF(\mathbf{b}_2)) - \langle \overline{g/m}\mathbf{c}, \mathbf{b}_2 \rangle}{mr} \right). \end{aligned}$$

It is easy to see that the first summation is equal to

$$\left| \frac{g}{m} \right|^4 \psi \left(\frac{-\overline{mr}\beta(\mathbf{c})k}{g/m} \right).$$

As a result,

$$\begin{aligned} S_{g,r}(\mathbf{c}) &= \left| \frac{g}{m} \right|^4 \psi \left(\frac{-\overline{mr}\beta(\mathbf{c})k}{g/m} \right) \\ &\cdot \sum_{|a_2| < |mr|}^* \sum_{\mathbf{b}_2 \in (\mathcal{O}/(mr))^4} \psi \left(\frac{a_2(2\lambda^T A(\mathbf{b}_2) - k + gF(\mathbf{b}_2)) - \langle \overline{g/m}\mathbf{c}, \mathbf{b}_2 \rangle}{mr} \right) \\ &= \left| \frac{g}{m} \right|^4 \psi \left(\frac{-\overline{mr}\beta(\mathbf{c})k}{g/m} \right) \sum_{|a| < |mr|}^* \psi \left(\frac{-ak}{mr} \right) \\ &\cdot \prod_{j=1}^4 \sum_{b \in \mathcal{O}/(mr)} \psi \left(\frac{ga\eta_j b^2 + (2a\eta_j \lambda_j - \overline{g/m}c_j)b}{mr} \right). \end{aligned}$$

Let us denote the double summation in this expression by S (temporarily neglect the leading factor). In order to complete our computation of a closed form for S , we use the following easy lemma whose proof we leave to the reader.

Lemma 5.2. Suppose $a, b, c \in \mathbb{F}_q[t]$. If $\gcd(a, c) \nmid b$, then

$$\sum_{x \in \mathcal{O}/(c)} \psi \left(\frac{ax^2 + bx}{c} \right) = 0.$$

Note that $\gcd(mr, ga\eta_j) = m\gcd(r, \eta_j)$. Using Lemma 5.2, we should have $m\gcd(r, \eta_j) | 2a\eta_j \lambda_j - g/mc_j$ for every j in order to have a nonzero value. Since Δ

and g are relatively prime, m and $\gcd(r, \eta_j)$ are also relatively prime. Consequently, the condition that $m \gcd(r, \eta_j) | 2a\eta_j \lambda_j - \overline{g/m} c_j$ for each j can be rewritten as the pair of conditions $m | 2aA\lambda - \overline{g/m} \mathbf{c}$ and $\gcd(r, \boldsymbol{\eta}) | \mathbf{c}$ (that is, for every j , $\gcd(r, \eta_j) | c_j$); otherwise, we have a zero value for $S_{g,r}(\mathbf{c})$. Let us assume for the rest of this section that $\gcd(r, \boldsymbol{\eta}) | \mathbf{c}$. In this case,

$$S = |m|^4 \sum_{\substack{|a| < |mr| \\ m | 2aA\lambda - \overline{g/m} \mathbf{c}}}^* \psi\left(\frac{-ak}{mr}\right) \prod_{j=1}^4 \sum_{b \in \mathcal{O}/(r)} \psi\left(\frac{\frac{g}{m} a \frac{\eta_j}{(r, \eta_j)} b^2 + \frac{(2a \frac{\eta_j}{(r, \eta_j)} \lambda_j - \overline{g/m} \frac{c_j}{(r, \eta_j)})}{m} b}{r/(r, \eta_j)}\right).$$

Completing the square and using the computation of Gauss sums, we may rewrite

$$\begin{aligned} & \sum_{b \in \mathcal{O}/(r)} \psi\left(\frac{\frac{g}{m} a \frac{\eta_j}{(r, \eta_j)} b^2 + \frac{(2a \frac{\eta_j}{(r, \eta_j)} \lambda_j - \overline{g/m} \frac{c_j}{(r, \eta_j)})}{m} b}{r/(r, \eta_j)}\right) \\ &= |\gcd(r, \eta_j)| \left(\frac{(g/m)a(\eta_j/(r, \eta_j))}{r/(r, \eta_j)}\right) \tau_{r/(r, \eta_j)} \\ & \quad \cdot \psi\left(\frac{-\overline{(g/m)a(\eta_j/(r, \eta_j))} \left(\frac{2a \frac{\eta_j}{(r, \eta_j)} \lambda_j - \overline{g/m} \frac{c_j}{(r, \eta_j)}}{2m}\right)^2}{r/(r, \eta_j)}\right), \end{aligned}$$

where $\left(\frac{(g/m)a(\eta_j/(r, \eta_j))}{r/(r, \eta_j)}\right)$ is the Jacobi symbol. Consequently, S is equal to

$$\begin{aligned} & \left(\prod_{j=1}^4 |\gcd(r, \eta_j)| \tau_{r/(r, \eta_j)}\right) |m|^4 \\ & \cdot \sum_{\substack{|a| < |mr| \\ m | 2aA\lambda - \overline{g/m} \mathbf{c}}}^* \left(\prod_{j=1}^4 \left(\frac{(g/m)a(\eta_j/(r, \eta_j))}{r/(r, \eta_j)}\right)\right) \psi\left(\frac{-ak}{mr}\right) \\ & \cdot \psi\left(\frac{-\sum_j \overline{(g/m)a(\eta_j/(r, \eta_j))} \left(\frac{2a \frac{\eta_j}{(r, \eta_j)} \lambda_j - \overline{g/m} \frac{c_j}{(r, \eta_j)}}{2m}\right)^2}{r/(r, \eta_j)}\right). \end{aligned}$$

So far, all our computations were valid for general quadratic forms. In the rest of this proof, we restrict to the Morgenstern quadratic form. In this case, we can write S more explicitly:

$$|m|^4 \left(|\gcd(r, t-1)| \tau_r \tau_{r/(r, t-1)} \right)^2 \psi \left(\frac{\overline{g/m}^2 \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{m^2 r} \right) \\ \cdot \sum_{\substack{|a| < |mr| \\ m|2aA\boldsymbol{\lambda} - \overline{g/m}\mathbf{c}}}^* \psi \left(\frac{-\overline{g/m}fa - \frac{1}{4}\overline{g/m}^3 F^*(\mathbf{c})\bar{a}}{m^2 r} \right),$$

where the last equality follows from $km + \overline{g/m}F(\boldsymbol{\lambda}) \equiv \overline{g/m}f \pmod{m^2 r}$. Here, F^* is the dual of F as before in the computation of the oscillatory integrals. Furthermore, the condition $m|2aA\boldsymbol{\lambda} - \overline{g/m}\mathbf{c}$ is equivalent to $a \equiv \overline{g/m}\beta(\mathbf{c}) \pmod{m}$. We deduce that

$$S_{g,r}(\mathbf{c}) = \frac{|g|^4}{|m|} \left(|\gcd(r, t-1)| \tau_r \tau_{r/(r, t-1)} \right)^2 \psi \left(\frac{-\overline{mr}\beta(\mathbf{c})k}{g/m} \right) \psi \left(\frac{\overline{g/m}^2 \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{m^2 r} \right) \\ \cdot \sum_{\substack{|a| < |m^2 r| \\ a \equiv \overline{g/m}\beta(\mathbf{c}) \pmod{m}}}^* \psi \left(\frac{-\overline{g/m}fa - \frac{1}{4}\overline{g/m}^3 F^*(\mathbf{c})\bar{a}}{m^2 r} \right),$$

where we have changed summation over a modulo mr to modulo $m^2 r$ at the cost of introducing a factor of $\frac{1}{|m|}$. We may replace the congruence condition by a summation modulo m and rewrite $S_{g,r}(\mathbf{c})$ in terms of Kloosterman sums:

$$S_{g,r}(\mathbf{c}) = \frac{|g|^4}{|m|^2} \left(|\gcd(r, t-1)| \tau_r \tau_{r/(r, t-1)} \right)^2 \psi \left(\frac{-\overline{mr}\beta(\mathbf{c})k}{g/m} \right) \psi \left(\frac{\overline{g/m}^2 \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{m^2 r} \right) \\ \cdot \sum_{s \in \mathcal{O}/(m)} \psi \left(\frac{-s\overline{g/m}\beta(\mathbf{c})}{m} \right) \text{Kl}_{m^2 r} \left(\overline{g/m}f - mrs, \frac{1}{4}\overline{g/m}^3 F^*(\mathbf{c}) \right).$$

By the Chinese Remainder Theorem,

$$\psi \left(\frac{-\overline{mr}\beta(\mathbf{c})k}{g/m} \right) \psi \left(\frac{\overline{g/m}^2 \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{m^2 r} \right) = \psi \left(\frac{-\overline{mr}\beta(\mathbf{c}) \left(\frac{f-F(\boldsymbol{\lambda})}{m} - \overline{m^2 r} \langle \boldsymbol{\lambda}, \mathbf{c} \rangle \right)}{(g/m)^2} \right) \psi \left(\frac{\langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2 r} \right),$$

from which the conclusion follows. \square

6. Strong approximation and Ramanujan graphs

In this section, we begin by showing how a certain square-root cancellation in an exponential sum gives us strong approximation for non-degenerate quadratic forms in four variables over $\mathbb{F}_q[t]$. We then proceed to show that assuming the twisted Linnik–Selberg Conjecture 1.4, we do have the desired square-root cancellation for Morgenstern’s quadratic forms used in the construction of Ramanujan graphs with even degree (odd q).

First, let us proceed to estimate the main term contributing to the smooth $N(w, \lambda)$. The following lemmas are true for quadratic forms in more variables, but we restrict here to $d = 4$.

Lemma 6.1. *Suppose $\varepsilon > 0$. With the notation as before with F a non-degenerate quadratic form in 4 variables over $\mathbb{F}_q[t]$, and for $1 \leq |r| \leq \widehat{Q}^{1-\varepsilon}$, we have*

$$I_{g,r}(\mathbf{0}) = C_F \widehat{Q}^4$$

for some non-negative constant C_F and for sufficiently large (depending only on ε and F) \widehat{Q} . $C_F > 0$ if the system under consideration is solvable over K_∞ .

Proof. It follows from equation (11) that

$$I_{g,r}(\mathbf{0}) = \frac{\widehat{Q}}{|r|} \int_{\substack{|\mathbf{t}| < \widehat{R} \\ |G(\mathbf{t})| < \widehat{Q}|r|}} dt = \frac{\widehat{Q}}{|r|} \int_{\substack{|g\mathbf{t} + \lambda| \leq |f|^{1/2} \\ |F(g\mathbf{t} + \lambda) - f| < \widehat{Q}|r||g|^2}} dt.$$

Making the substitution $\mathbf{x} = g\mathbf{t} + \lambda$ gives us the equality

$$I_{g,r}(\mathbf{0}) = \frac{\widehat{Q}}{|r||g|^4} \int_{\substack{|\mathbf{x}| \leq |f|^{1/2} \\ |F(\mathbf{x}) - f| < \widehat{Q}|r||g|^2}} d\mathbf{x}.$$

Write $f = \alpha_f u^2$, where $\alpha_f \in \{1, \nu, t, \nu t\}$ is the quadratic residue of f . By Lemma 3.2 and Fubini, we may rewrite this as

$$\begin{aligned} I_{g,r}(\mathbf{0}) &= \frac{\widehat{Q}}{|r||g|^4} \int_{|\mathbf{x}| \leq |f|^{1/2}} \int_{\mathbb{T}} \psi \left(\frac{(F(\mathbf{x}) - f)}{rg^2 t^Q} \alpha \right) d\alpha d\mathbf{x} \\ &= \frac{\widehat{Q}}{|r||g|^4} \int_{\mathbb{T}} \int_{|\mathbf{x}| < \widehat{D}} \psi \left(\frac{(F(\mathbf{x}) - f)}{rg^2 t^Q} \alpha \right) d\mathbf{x} d\alpha \\ &= \frac{\widehat{Q} \widehat{D}^4}{|r||g|^4} \int_{\mathbb{T}} \int_{\mathbb{T}^4} \psi \left(\frac{(F(\mathbf{x}) - f/(t^2 u^2))}{rg^2 t^Q / (t^2 u^2)} \alpha \right) d\mathbf{x} d\alpha \\ &= \frac{\widehat{Q} \widehat{D}^4}{|r||g|^4} \int_{\mathbb{T}} \int_{\mathbb{T}^4} \psi \left(\frac{(F(\mathbf{x}) - \alpha_f / (t^2))}{rg^2 t^Q / (t^2 u^2)} \alpha \right) d\mathbf{x} d\alpha \end{aligned}$$

where $D := 1 + \deg u$, and the last equality follows from scaling the \mathbf{x} coordinate by a factor of \widehat{D} . Making the substitution $\beta = \frac{\alpha}{rg^2 t^Q / (t^2 u^2)}$, we obtain the equality

$$I_{g,r}(\mathbf{0}) = \frac{\widehat{Q}^2 \widehat{D}^4}{|g|^{2\widehat{D}}} \int_{|\beta| < \frac{\widehat{D}}{\widehat{Q}|r||g|^2}} \int_{\mathbb{T}^4} \psi((F(\mathbf{x}) - \alpha_f/t^2)\beta) d\mathbf{x} d\beta.$$

Note that the integral is equal to

$$\frac{\widehat{D}}{\widehat{Q}|r||g|^2} \text{vol} \left(\left\{ \mathbf{x} \in \mathbb{T}^4 : |F(\mathbf{x}) - \alpha_f/t^2| \leq \frac{\widehat{Q}|r||g|^2}{\widehat{D}} \right\} \right) \geq 0.$$

Note that $\frac{\widehat{D}}{\widehat{Q}|r||g|^2} \gg \widehat{Q}^\varepsilon$. Applying Lemma 6.2 of [SZ20], we can choose \widehat{Q} large enough (depending on ε and the F) so that the integral over $|\beta|$ is constant over balls of radii at least \widehat{Q}^ε . The conclusion follows. \square

As in Lemma 6.3 of [SZ20], we can show that for $\widehat{Q}^{1-\varepsilon} \leq |r| \leq \widehat{Q}$, the contribution of the terms in $N(w, \lambda)$ when $\mathbf{c} = \mathbf{0}$ and such r is small.

Lemma 6.2.

$$\sum_{\widehat{Q}^{1-\varepsilon} \leq |r| \leq \widehat{Q}} |gr|^{-4} |S_{g,r}(\mathbf{0})| |I_{g,r}(\mathbf{0})| \ll_{\varepsilon, F} |g|^\varepsilon \widehat{Q}^{\frac{7}{2}+\varepsilon}$$

Proof. The only difference in the proof of this lemma and that of Lemma 6.3 of [SZ20] is that the definitions of the oscillatory integrals are different. However, we only need the same bound

$$|I_{g,r}(\mathbf{0})| \ll_{\varepsilon, F} \widehat{Q}^{4+\varepsilon}$$

for such r , which is trivial. The rest of the proof is as before; we also need to use Proposition 3.1 of [SZ20]. \square

We now proceed to show that

$$\sum_{r: 1 \leq |r| \leq \widehat{T}} |gr|^{-4} S_{g,r}(\mathbf{0})$$

can be written in terms of local densities. Indeed, by Lemma 6.5 of [SZ20], we have the estimate

$$\sum_{r: 1 \leq |r| \leq \widehat{T}} |gr|^{-4} S_{g,r}(\mathbf{0}) = \sum_r |gr|^{-4} S_{g,r}(\mathbf{0}) + O_{\varepsilon, \Delta}(\widehat{T}^{-1/2+\varepsilon})$$

for every $\varepsilon > 0$. On the other hand, Lemma 6.6 of [SZ20] shows that

$$\sum_r |gr|^{-4} S_{g,r}(\mathbf{0}) = \prod_{\varpi} \sigma_{\varpi} \gg |f|^{-\varepsilon},$$

where ϖ ranges over the monic irreducible polynomials in $\mathbb{F}_q[t]$, and

$$\sigma_{\varpi} := \lim_{k \rightarrow \infty} \frac{|\{\mathbf{x} \bmod \varpi^{k+\nu_{\varpi}(g)} : F(\mathbf{x}) \equiv f \bmod \varpi^{k+\nu_{\varpi}(g)}, \mathbf{x} \equiv \boldsymbol{\lambda} \bmod \varpi^{\nu_{\varpi}(g)}\}|}{|\varpi|^{3k}}.$$

Using the above, estimates, let us take the simple step of showing that conditional on a square-root cancellation we have optimal strong approximation for any non-degenerate quadratic form in four variables over $\mathbb{F}_q[t]$. First, recall from Section 3 that the smooth sum for quadratic forms of four variables is

$$N(w, \boldsymbol{\lambda}) = \frac{1}{|g|\widehat{Q}^2} \sum_{\substack{r \in \mathcal{O} \\ |r| \leq \widehat{Q}}} \sum_{\mathbf{c} \in \mathcal{O}^4} |gr|^{-4} S_{g,r}(\mathbf{c}) I_{g,r}(\mathbf{c}).$$

Lemma 6.3. *Suppose we have a non-degenerate quadratic form F over $\mathbb{F}_q[t]$ in $d = 4$ variables. Additionally, assume that*

$$\left| \sum_{1 \leq |r| \leq \widehat{Q}} \sum_{\mathbf{c} \neq 0} |gr|^{-4} S_{g,r}(\mathbf{c}) I_{g,r}(\mathbf{c}) \right| \ll_{\varepsilon, F} \widehat{Q}^{\frac{7}{2} + \varepsilon} |g|^{\frac{1}{2} + \varepsilon}.$$

Then Conjecture 1.2 is true for the given quadratic form F .

Proof. Using the assumption, Lemma 4.3, and Lemma 6.2, we have

$$N(w, \boldsymbol{\lambda}) = \frac{1}{|g|\widehat{Q}^2} \sum_{\substack{r \in \mathcal{O} \\ |r| \leq \widehat{Q}^{1-\varepsilon}}} |gr|^{-4} S_{g,r}(\mathbf{0}) I_{g,r}(\mathbf{0}) + O_{\varepsilon, F} \left(\widehat{Q}^{\frac{3}{2} + \varepsilon} |g|^{-\frac{1}{2} + \varepsilon} \right).$$

From Lemma 6.1 above, $I_{g,r}(\mathbf{0}) = C_F \widehat{Q}^4$ for some constant $C_F > 0$ and sufficiently large (depending only on ε and F) \widehat{Q} . Hence for such \widehat{Q} ,

$$\frac{1}{|g|\widehat{Q}^2} \sum_{\substack{r \in \mathcal{O} \\ |r| \leq \widehat{Q}^{1-\varepsilon}}} |gr|^{-4} S_{g,r}(\mathbf{0}) I_{g,r}(\mathbf{0}) = \frac{C_F \widehat{Q}^2}{|g|} \sum_{\substack{r \in \mathcal{O} \\ |r| \leq \widehat{Q}^{1-\varepsilon}}} |gr|^{-4} S_{g,r}(\mathbf{0}).$$

From the discussion prior to this lemma, we know that we can express the first sum in terms of local densities:

$$\sum_{\substack{r \in \mathcal{O} \\ |r| \leq \widehat{Q}^{1-\varepsilon}}} |gr|^{-4} S_{g,r}(\mathbf{0}) = \prod_{\varpi} \sigma_{\varpi} + O \left(\widehat{Q}^{-\frac{1}{2} + \varepsilon} \right).$$

Consequently, the smooth sum is equal to

$$\begin{aligned}
N(w, \lambda) &= \frac{C_F \widehat{Q}^2}{|g|} \left(\prod_{\varpi} \sigma_{\varpi} + O\left(\widehat{Q}^{-\frac{1}{2}+\varepsilon}\right) \right) + O_{\varepsilon, F}\left(\widehat{Q}^{\frac{3}{2}+\varepsilon} |g|^{-\frac{1}{2}+\varepsilon}\right) \\
&= \frac{C_F \widehat{Q}^2}{|g|} \prod_{\varpi} \sigma_{\varpi} + O_{\varepsilon, F}\left(\widehat{Q}^{\frac{3}{2}+\varepsilon} |g|^{-\frac{1}{2}+\varepsilon}\right) \\
&= \frac{C_F \widehat{Q}^2}{|g|} \prod_{\varpi} \sigma_{\varpi} \left(1 + O_{\varepsilon, F}\left(\frac{|f|^{\varepsilon} \widehat{Q}^{\frac{3}{2}+\varepsilon} |g|^{\frac{1}{2}+\varepsilon}}{\widehat{Q}^2}\right) \right) \\
&= \frac{C_F \widehat{Q}^2}{|g|} \prod_{\varpi} \sigma_{\varpi} \left(1 + O_{\varepsilon, F}\left(\frac{|g|^{1+\varepsilon}}{|f|^{\frac{1}{4}-\varepsilon}}\right) \right) \\
&= \frac{C_F \widehat{Q}^2}{|g|} \prod_{\varpi} \sigma_{\varpi} \left(1 + O_{\varepsilon, F}\left(\left(\frac{|g|^{4+\varepsilon}}{|f|}\right)^{\frac{1}{4}-\varepsilon}\right) \right).
\end{aligned}$$

Hence, if $|f| \gg |g|^{4+\varepsilon}$, we have strong approximation. In the third equality, we have used that the product of the local densities is $\gg |f|^{-\varepsilon}$. \square

Remark 14. In the proof of the main theorem of [SZ20], the only reason we had to have $|f| \gg |g|^{6+\varepsilon}$ in the case of four variables was that we used the weaker statement

$$\begin{aligned}
\left| \sum_{1 \leq |r| \leq \widehat{Q}} \sum_{\mathbf{c} \neq 0} |gr|^{-d} S_{g,r}(\mathbf{c}) I_{g,r}(\mathbf{c}) \right| &\leq \sum_{1 \leq |r| \leq \widehat{Q}} \sum_{\mathbf{c} \neq 0} |gr|^{-d} |S_{g,r}(\mathbf{c})| |I_{g,r}(\mathbf{c})| \\
&\ll_{\varepsilon, F} \widehat{Q}^{\frac{d+3}{2}+\varepsilon} |g|^{\frac{d-3}{2}+\varepsilon} (1 + |g|^{-\frac{d-5}{2}+\varepsilon})
\end{aligned}$$

proved in Proposition 7.1 of [SZ20]. Indeed, for $d = 4$, this is weaker than what we ask above because then $1 + |g|^{-\frac{d-5}{2}+\varepsilon} \sim |g|^{1/2+\varepsilon}$ and so is not $O(1)$ (in contrast to when $d \geq 5$). The bound we assume in the statement of this lemma is precisely that if we do not take absolute values we get an extra power saving of $|g|^{1/2}$ when in the case of four variables.

Remark 15. In light of the previous remark and the proof of the above lemma, any improvement to the factor $|g|^{1/2}$ in the previous remark would allow us to weaken the condition $\deg f \geq (6 + \varepsilon) \deg g + O_{\varepsilon}(1)$ that was required for the main theorem of [SZ20] in the case of non-degenerate quadratic forms in $d = 4$ variables.

In the rest of this section, we show how the twisted Linnik–Selberg Conjecture 1.4 implies that the above square-root cancellation is true for Morgenstern’s quadratic forms. This in turn implies Conjecture 1.1 giving the upper bound $(\frac{4}{3} + \varepsilon) \log_q |G| + O_{\varepsilon}(1)$ for the diameter of $q + 1$ -regular Morgenstern Ramanujan graphs with q odd.

Recall the notations in Section 4. In order to understand the error in the smooth sum $N(w, \lambda)$ for the Morgenstern quadratic form, we use the explicit formulas for the oscillatory integrals $I_{g,r}(\mathbf{c})$ and exponential sums derived in the last two sections. Note

that by Lemma 4.3, $I_{g,r}(\mathbf{c}) = 0$ when $|r| \leq \widehat{Q}$ and $|\mathbf{c}| \geq \widehat{Q}|g|/\widehat{R}$. Therefore, it suffices to study

$$\sum_{1 \leq |r| \leq \widehat{Q}} \sum_{0 < |\mathbf{c}| < \frac{\widehat{Q}|g|}{\widehat{R}}} |gr|^{-d} S_{g,r}(\mathbf{c}) I_{g,r}(\mathbf{c}) = E_1 + E_2,$$

where

$$E_1 := \sum_{0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R}} \sum_{1 \leq |r| \leq \frac{\widehat{R}|\mathbf{c}|q^{\pi_{\mathbf{c}}}-1}{|g|}} |gr|^{-4} S_{g,r}(\mathbf{c}) I_{g,r}(\mathbf{c})$$

and

$$E_2 := \sum_{0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R}} \sum_{\frac{\widehat{R}|\mathbf{c}|q^{\pi_{\mathbf{c}}}}{|g|} \leq |r| \leq \widehat{Q}} |gr|^{-4} S_{g,r}(\mathbf{c}) I_{g,r}(\mathbf{c}).$$

Here, $\pi_{\mathbf{c}} = 0$ if \mathbf{c} satisfies $\max(|c_3|, |c_4|) > \max(|c_1|, |c_2|)$ and $\deg f$ is even; otherwise, $\pi_{\mathbf{c}} = 1$. In order to obtain the desired bound in the above lemma, it suffices to prove the desired bound for each of E_1 and E_2 separately. For simplicity, we assume for the rest of this section that g is an irreducible polynomial in $\mathbb{F}_q[t]$.

We first treat E_2 . Note that by Proposition 4.1, for $\frac{\widehat{R}|\mathbf{c}|q^{\pi_{\mathbf{c}}}}{|g|} \leq |r| \leq \widehat{Q}$, we have $I_{g,r}(\mathbf{c}) = I_{g,r}(\mathbf{0})$ or 0. Furthermore, from the definition of $I_{g,r}(\mathbf{0})$, we know that it depends on $|r|$, not r itself. Therefore, it makes sense to write $I_{g,|r|}(\mathbf{0})$ instead of $I_{g,r}(\mathbf{0})$. From this discussion, we obtain

$$|E_2| \leq \sum_{0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R}} \sum_{\frac{\widehat{R}|\mathbf{c}|q^{\pi_{\mathbf{c}}}}{|g|} \leq \widehat{T} \leq \widehat{Q}} |I_{g,\widehat{T}}(\mathbf{0})| |g|^{-4} \widehat{T}^{-4} \left| \sum_{|r|=\widehat{T}} S_{g,r}(\mathbf{c}) \right|.$$

In the following, we will use the following notation: \sum^{exc} denotes summation over those \mathbf{c} such that $S_{g,r}(\mathbf{c}) \neq 0$ and $|\mathbf{c}| \leq |g|$. We have the following accompanying lemma.

Lemma 6.4. *For every $\theta < 0$ and every $0 \leq T \leq \deg g$, we have*

$$\sum_{0 < |\mathbf{c}| \leq \widehat{T}}^{\text{exc}} |\mathbf{c}|^{\theta} \ll_{\varepsilon, F, \theta} \widehat{T}^{\theta+1}.$$

Proof. By Lemma 3.2 of [SZ20], we know that such \mathbf{c} must be polynomial multiples of $A\lambda$ modulo g . Since $0 < |\mathbf{c}| \leq |g|$, \mathbf{c} and α uniquely determine each other. By assumption, at least one coordinate of λ is relatively prime to g , from which the final inequality follows. \square

Suppose we have for every $T \geq 0$,

$$\left| \sum_{|r|=\widehat{T}} S_{g,r}(\mathbf{c}) \right| \ll_{\varepsilon, F} \widehat{Q}^{\varepsilon} |g|^{4+\varepsilon} \widehat{T}^{3+\varepsilon}. \quad (16)$$

Then it would follow from this assumption and Lemma 6.4 that

$$\begin{aligned} |E_2| &\leq \sum_{0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R}} \sum_{\substack{\widehat{R}|\mathbf{c}|q^{\pi\mathbf{c}} \\ |g| \leq \widehat{T} \leq \widehat{Q}}} |I_{g,\widehat{T}}(\mathbf{0})| |g|^{-4} \widehat{T}^{-4} \left| \sum_{|r|=\widehat{T}} S_{g,r}(\mathbf{c}) \right| \\ &\ll_{\varepsilon, F} \widehat{Q}^{4+\varepsilon} |g|^{\varepsilon} \sum_{\substack{\text{exc} \\ 0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R} \\ \widehat{R}|\mathbf{c}|q^{\pi\mathbf{c}} \\ |g| \leq \widehat{T} \leq \widehat{Q}}} \widehat{T}^{-1+\varepsilon} \\ &\ll_{\varepsilon, F} \widehat{Q}^{4+\varepsilon} |g|^{\varepsilon} \sum_{\substack{\text{exc} \\ 0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R}}} \left(\frac{\widehat{Q}|\mathbf{c}|}{|g|} \right)^{-1+\varepsilon} \\ &\ll_{\varepsilon, F} \widehat{Q}^{3+\varepsilon} |g|^{1+\varepsilon} \sum_{\substack{\text{exc} \\ 0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R}}} |\mathbf{c}|^{-1+\varepsilon} \ll_{\varepsilon, F} \widehat{Q}^{7/2+\varepsilon} |g|^{1/2+\varepsilon}. \end{aligned}$$

Note that \widehat{Q} and \widehat{R} differ by a factor of q . This is exactly the desired bound on E_2 . Therefore, we have reduced to showing inequality (16) for each integer $T \geq 0$.

By Proposition 5.1, when $\gcd(r, t-1) | c_3, c_4$, $S_{g,r}(\mathbf{c})$ is equal to

$$\begin{aligned} &\frac{|g|^4}{|m|^2} (|\gcd(r, t-1)| \tau_r \tau_{r/(r, t-1)})^2 \psi \left(\frac{-\overline{mr} \beta(\mathbf{c}) \left(\frac{f-F(\boldsymbol{\lambda})}{m} - \overline{m^2 r} \langle \boldsymbol{\lambda}, \mathbf{c} \rangle \right)}{(g/m)^2} \right) \psi \left(\frac{\langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2 r} \right) \\ &\cdot \sum_{s \in \mathcal{O}/(m)} \psi \left(\frac{-sg \overline{m} \beta(\mathbf{c})}{m} \right) \text{Kl}_{m^2 r} \left(\overline{g/m} f - mrs, \frac{1}{4} \overline{g/m}^3 F^*(\mathbf{c}) \right), \end{aligned} \quad (17)$$

where $m := (g, r^\infty)$. If $\gcd(r, t-1) \nmid c_3, c_4$, then $S_{g,r}(\mathbf{c}) = 0$. Since g is irreducible, $m = 1$ or $m = g$. For the bound on E_2 , we split the sum over r such that $m = 1$ and $m = g$, and show the desired bounds separately. First, let us show that we may assume that $m = 1$.

Lemma 6.5. *For every $0 \leq T \leq \deg g$,*

$$\sum_{\substack{|r|=\widehat{T} \\ g|r}} |S_{g,r}(\mathbf{c})| \ll_{\varepsilon, F} \widehat{Q}^{\varepsilon} |g|^{3+\varepsilon} \widehat{T}^{7/2+\varepsilon},$$

from which it follows that the contributions to E_2 from those r such that $g|r$ may be neglected.

Proof. When $g|r$, $m = g$. From the expression (17) for $S_{g,r}(\mathbf{c})$, we see that it suffices to show that

$$\sum_{\substack{|r|=\hat{T} \\ g|r}} (|\gcd(r, t-1)|\tau_r\tau_{r/(r,t-1)})^2 \left| \sum_{s \in \mathcal{O}/(g)} \psi\left(\frac{-s\beta(\mathbf{c})}{g}\right) \text{Kl}_{g^2r}\left(f - grs, \frac{1}{4}F^*(\mathbf{c})\right) \right| \\ \ll_{\varepsilon, F} \hat{Q}^\varepsilon |g|^{1+\varepsilon} \hat{T}^{7/2+\varepsilon}.$$

Note that in this case, $|\lambda| < |g|$, $|\mathbf{c}| \leq |g|$, and $g|r$, and so $\psi\left(\frac{\langle \lambda, \mathbf{c} \rangle}{g^2r}\right) = 1$. By the Weil bound on Kloosterman sums (Lemma 3.5 of [SZ20]), we have that

$$\left| \sum_{s \in \mathcal{O}/(g)} \psi\left(\frac{-s\beta(\mathbf{c})}{g}\right) \text{Kl}_{g^2r}\left(f - grs, \frac{1}{4}F^*(\mathbf{c})\right) \right| \\ \ll_{\varepsilon, F} |g|^{2+\varepsilon} |r|^{1/2+\varepsilon} |\gcd(f - grs, \frac{1}{4}F^*(\mathbf{c}), g^2r)|^{1/2}.$$

Since $\gcd(g, f)=1$ and $g|r$ by assumption, $\gcd(f - grs, \frac{1}{4}F^*(\mathbf{c}), g^2r) = \gcd(f, \frac{1}{4}F^*(\mathbf{c}), r/g)$. Therefore,

$$\sum_{\substack{|r|=\hat{T} \\ g|r}} (|\gcd(r, t-1)|\tau_r\tau_{r/(r,t-1)})^2 \left| \sum_{s \in \mathcal{O}/(g)} \psi\left(\frac{-s\beta(\mathbf{c})}{g}\right) \text{Kl}_{g^2r}\left(f - grs, \frac{1}{4}F^*(\mathbf{c})\right) \right| \\ \ll_{\varepsilon, F} |g|^{2+\varepsilon} \hat{T}^{2+1/2+\varepsilon} \sum_{\substack{|r|=\hat{T} \\ g|r \\ \gcd(r, t-1)|c_3, c_4}} |\gcd(f, r/g)|^{1/2} \ll_{\varepsilon, F} |g|^{2+\varepsilon} \hat{T}^{2+1/2+\varepsilon} |f|^\varepsilon \frac{\hat{T}}{|g|} \\ \ll_{\varepsilon, F} \hat{Q}^\varepsilon |g|^{1+\varepsilon} \hat{T}^{7/2+\varepsilon}.$$

We use this to show that the contribution to E_2 from those r such that $g|r$ already satisfies the desired bound on E_2 . Indeed, we have

$$\sum_{0 < |\mathbf{c}| < \hat{Q}|g|/\hat{R}} \sum_{\substack{\hat{R}|\mathbf{c}|g^{\pi\mathbf{c}} \\ |g| \leq \hat{Q}}} |I_{g, \hat{T}}(\mathbf{0})| \hat{T}^{-4} |g|^{-4} \sum_{\substack{|r|=\hat{T} \\ g|r}} |S_{g,r}(\mathbf{c})| \\ \ll_{\varepsilon, F} \hat{Q}^{4+\varepsilon} |g|^{-1+\varepsilon} \sum_{\substack{\text{exc} \\ 0 < |\mathbf{c}| < \hat{Q}|g|/\hat{R} \\ \hat{R}|\mathbf{c}|g^{\pi\mathbf{c}} \\ |g| \leq \hat{T} \leq \hat{Q}}} \hat{T}^{-1/2+\varepsilon} \\ \ll_{\varepsilon, F} \hat{Q}^{4+\varepsilon} |g|^{-1+\varepsilon} \sum_{\substack{\text{exc} \\ 0 < |\mathbf{c}| < \hat{Q}|g|/\hat{R}}} \left(\frac{\hat{Q}|\mathbf{c}|}{|g|}\right)^{-1/2+\varepsilon}$$

$$\ll_{\varepsilon, F} \widehat{Q}^{7/2+\varepsilon} |g|^{-1/2+\varepsilon} \sum_{\substack{\text{exc} \\ 0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R}}} |\mathbf{c}|^{-1/2+\varepsilon} \ll_{\varepsilon, F} \widehat{Q}^{7/2+\varepsilon} |g|^{\varepsilon},$$

where the final bound follows from Lemma 6.4. \square

From Lemma 6.5, we may assume that $m = 1$, that is, $g \nmid r$. Using this and the expression (17) with $m = 1$, for the inequality (16) it suffices to show that

$$\left| \sum_{\substack{|r|=\widehat{T} \\ (g,r)=1 \\ \gcd(r,t-1)|c_3, c_4}} (|\gcd(r, t-1)| \tau_r \tau_{r/(r,t-1)})^2 \psi \left(\frac{-\bar{r}(\beta(\mathbf{c}))(f - F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2} \right) \right. \\ \left. \cdot \text{Kl}_r \left(\bar{g}f, \frac{1}{4} \bar{g}^3 F^*(\mathbf{c}) \right) \right| \\ \ll_{\varepsilon, F} |g|^{\varepsilon} \widehat{T}^{3+\varepsilon}.$$

Note that since $|\boldsymbol{\lambda}| < |g|$, $|\mathbf{c}| \leq |g|$, $\psi \left(\frac{\langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2 r} \right) = 1$ unless possibly when $r = 1$, which contributes a term of norm 1 to the above sum. This is why we may suppress the $\psi \left(\frac{\langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2 r} \right)$.

We split the sum into two sums, one where $t-1|r$ and one where $t-1 \nmid r$. Summing over those r such that $t-1|r$ gives us the sum

$$q^2 \sum_{\substack{|r|=\widehat{T} \\ (g,r)=1, \ t-1|r \\ t-1|c_3, c_4}} (\tau_r \tau_{r/(r,t-1)})^2 \psi \left(\frac{-\bar{r}(\beta(\mathbf{c}))(f - F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2} \right) \text{Kl}_r \left(\bar{g}f, \frac{1}{4} \bar{g}^3 F^*(\mathbf{c}) \right).$$

Since $\tau_r^2 \tau_{r/(t-1)}^2$ only depends on $|r|$, and not r itself, we can pull it out of the sum. This term has norm \widehat{T}^2/q . The second sum, that is when $t-1 \nmid r$, is

$$\widehat{T}^2 \sum_{\substack{|r|=\widehat{T} \\ ((t-1)g, r)=1}} \psi \left(\frac{-\bar{r}(\beta(\mathbf{c}))(f - F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2} \right) \text{Kl}_r \left(\bar{g}f, \frac{1}{4} \bar{g}^3 F^*(\mathbf{c}) \right).$$

Therefore, in any case, it suffices to show that we have the following two cancellations. First, that if $t-1|c_3, c_4$,

$$\left| \sum_{\substack{|r|=\widehat{T} \\ (g,r)=1, \ t-1|r}} \psi \left(\frac{-\bar{r}(\beta(\mathbf{c}))(f - F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2} \right) \text{Kl}_r \left(\bar{g}f, \frac{1}{4} \bar{g}^3 F^*(\mathbf{c}) \right) \right| \ll_{\varepsilon, F} \widehat{Q}^{\varepsilon} |g|^{\varepsilon} \widehat{T}^{1+\varepsilon}.$$

Second, that

$$\left| \sum_{\substack{|r|=\widehat{T} \\ ((t-1)g,r)=1}} \psi \left(\frac{-\bar{r}(\beta(\mathbf{c})(f - F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle)}{g^2} \right) \text{Kl}_r \left(\bar{g}f, \frac{1}{4}\bar{g}^3 F^*(\mathbf{c}) \right) \right| \ll_{\varepsilon, F} \widehat{Q}^\varepsilon |g|^\varepsilon \widehat{T}^{1+\varepsilon}.$$

Of course, we may replace one of the above bounds, say the second one, with

$$\left| \sum_{\substack{|r|=\widehat{T} \\ (g,r)=1}} \psi \left(\frac{-\bar{r}(\beta(\mathbf{c})(f - F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle)}{g^2} \right) \text{Kl}_r \left(\bar{g}f, \frac{1}{4}\bar{g}^3 F^*(\mathbf{c}) \right) \right| \ll_{\varepsilon, F} \widehat{Q}^\varepsilon |g|^\varepsilon \widehat{T}^{1+\varepsilon}.$$

We now use the explicit computation of $I_{g,r}(\mathbf{c})$ to show that a different kind of twisted Linnik–Selberg cancellation over function fields, taking the infinite place into account as well, gives us the desired bound on E_1 . Indeed, by Proposition 4.1, for \mathbf{c} and r such that $1 \leq |r| \leq \frac{\widehat{R}|\mathbf{c}|q^{\pi_{\mathbf{c}}-1}}{|g|}$, we have $I_{g,r}(\mathbf{c}) = 0$ or

$$I_{g,r}(\mathbf{c}) = -\widehat{Q}^2 |gr|^2 |F^*(\mathbf{c})|^{-1} \text{Kl}_\infty \left(\psi, \frac{fF^*(\mathbf{c})}{4r^2 g^4} \right).$$

Therefore,

$$|E_1| \leq \widehat{Q}^2 \sum_{0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R}} |F^*(\mathbf{c})|^{-1} \left| \sum_{1 \leq |r| \leq \frac{\widehat{R}|\mathbf{c}|q^{\pi_{\mathbf{c}}-1}}{|g|}} |gr|^{-2} S_{g,r}(\mathbf{c}) \text{Kl}_\infty \left(\psi, \frac{fF^*(\mathbf{c})}{4r^2 g^4} \right) \right|.$$

We can use Proposition 5.1 to rewrite this inequality as

$$\begin{aligned} |E_1| &\leq \widehat{Q}^2 |g|^2 \sum_{m|g} \sum_{0 < |\mathbf{c}| < \widehat{Q}|g|/\widehat{R}} |F^*(\mathbf{c})|^{-1} \left| \sum_{\substack{1 \leq |r| \leq \frac{\widehat{R}|\mathbf{c}|q^{\pi_{\mathbf{c}}-1}}{|g|} \\ (g,r^\infty)=m \\ \gcd(r,t-1)|_{c_3,c_4}}} \frac{1}{|mr|^2} (|\gcd(r, t-1)| \tau_r \tau_{r/(r,t-1)})^2 \right. \\ &\quad \cdot \psi \left(\frac{-\overline{mr}\beta(\mathbf{c}) \left(\frac{f-F(\boldsymbol{\lambda})}{m} \right) - \overline{m^2 r} \langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{(g/m)^2} \right) \psi \left(\frac{\langle \boldsymbol{\lambda}, \mathbf{c} \rangle}{g^2 r} \right) \sum_{s \in \mathcal{O}/(m)} \psi \left(\frac{-sg/\overline{m}\beta(\mathbf{c})}{m} \right) \\ &\quad \left. \cdot \text{Kl}_{m^2 r} \left(\overline{g/m} f - mrs, \frac{1}{4} \overline{g/m}^3 F^*(\mathbf{c}) \right) \text{Kl}_\infty \left(\psi, \frac{fF^*(\mathbf{c})}{4r^2 g^4} \right) \right|. \end{aligned}$$

Using the fact that Morgenstern quadratic forms are anisotropic and so satisfy $|\mathbf{c}|^2 \ll_F |F^*(\mathbf{c})|$, we may reduce, as in the case of E_2 , to showing that

$$\left| \sum_{\substack{|r|=\widehat{T} \\ (g,r)=1 \\ \gcd(r,t-1)|c_3,c_4}} (|\gcd(r,t-1)|\tau_r\tau_{r/(r,t-1)})^2 \psi\left(\frac{-\bar{r}(\beta(\mathbf{c})(f-F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle)}{g^2}\right) \right. \\ \left. \cdot \text{Kl}_r\left(\bar{g}f, \frac{1}{4}\bar{g}^3 F^*(\mathbf{c})\right) \text{Kl}_\infty\left(\psi, \frac{fF^*(\mathbf{c})}{4r^2 g^4}\right) \right| \ll_{\varepsilon, F} \widehat{Q}^\varepsilon |g|^\varepsilon \widehat{T}^{3+\varepsilon}.$$

When $m := (g, r) = g$, we may argue as before and use the Weil bound $\text{Kl}_\infty(\psi, \alpha) \ll_\varepsilon |\alpha|^{1/4+\varepsilon}$ (Lemma 5.8 of [SZ20]) in addition to $|\mathbf{c}| \leq |g|$. Therefore, as in the case of E_2 , we may assume that $m = 1$, that is, $g \nmid r$.

As in the case of E_2 , we may split into two sums, one where $t-1|r$ and one where $t-1 \nmid r$. We similarly obtain that it suffices to show that

$$\left| \sum_{\substack{|r|=\widehat{T} \\ (g,r)=1, \ t-1|r}} \psi\left(\frac{-\bar{r}(\beta(\mathbf{c})(f-F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle)}{g^2}\right) \text{Kl}_r\left(\bar{g}f, \frac{1}{4}\bar{g}^3 F^*(\mathbf{c})\right) \text{Kl}_\infty\left(\psi, \frac{fF^*(\mathbf{c})}{4r^2 g^4}\right) \right| \\ \ll_{\varepsilon, F} \widehat{Q}^\varepsilon |g|^\varepsilon \widehat{T}^{1+\varepsilon}$$

and

$$\left| \sum_{\substack{|r|=\widehat{T} \\ (g,r)=1}} \psi\left(\frac{-\bar{r}(\beta(\mathbf{c})(f-F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle)}{g^2}\right) \text{Kl}_r\left(\bar{g}f, \frac{1}{4}\bar{g}^3 F^*(\mathbf{c})\right) \text{Kl}_\infty\left(\psi, \frac{fF^*(\mathbf{c})}{4r^2 g^4}\right) \right| \\ \ll_{\varepsilon, F} \widehat{Q}^\varepsilon |g|^\varepsilon \widehat{T}^{1+\varepsilon}.$$

Therefore, we have reduced proving optimal strong approximation for the Morgenstern quadratic form to proving the above square-root cancellation. These would follow from the twisted Linnik–Selberg square-root cancellations over function fields, that is, Conjecture 1.4. Indeed, we let $\alpha := \beta(\mathbf{c})(f-F(\boldsymbol{\lambda})) + \langle \boldsymbol{\lambda}, \mathbf{c} \rangle$, $a := \frac{f}{g}$, $b := \frac{F^*(\mathbf{c})}{4g^3}$, and $\delta \in \{1, t-1\}$. Note that since \mathbf{c} are such that $t-1|c_3, c_4$ (otherwise, $S_{g,r}(\mathbf{c}) = 0$), $F^*(\mathbf{c}) \in \mathbb{F}_q[t]$, and so $b \in \mathbb{F}_q[t, g^{-1}]$. Also, recall that $\psi_r(x) = \psi\left(\frac{x \bmod r}{r}\right)$, that is, we first reduce modulo r , and then divide by r . The fact that this strong approximation implies the conjectured upper bound on the diameter of Morgenstern Ramanujan graphs for odd q (see Conjecture 1.1) can be found in the introduction to the authors' paper [SZ20]; for more details, the reader is advised to look at Morgenstern's paper [Mor94].

Acknowledgments

We are grateful to the anonymous referee for many useful comments regarding a previous version of this paper. N.T. Sardari's work was supported partially by the National

Science Foundation under Grant No. DMS-2015305 and is grateful to Max Planck Institute for Mathematics in Bonn and the Institute for Advanced Study for their hospitality and financial support. M. Zargar was supported by SFB1085: Higher invariants at the University of Regensburg.

References

- [Alo86] N. Alon, Eigenvalues and expanders, vol. 6, in: *Theory of Computing*, Singer Island, Fla., 1984, 1986, pp. 83–96.
- [BKS19] T.D. Browning, V. Vinay Kumaraswamy, R.S. Steiner, Twisted Linnik implies optimal covering exponent for S^3 , *Int. Math. Res. Not.* 1 (2019) 140–164.
- [BV15] T.D. Browning, P. Vishe, Rational points on cubic hypersurfaces over $\mathbb{F}_q(t)$, *Geom. Funct. Anal.* 25 (3) (2015) 671–732.
- [Coh16] M.B. Cohen, Ramanujan graphs in polynomial time, in: *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016*, IEEE Computer Soc., Los Alamitos, CA, 2016, pp. 276–281.
- [CPS90] J.W. Cogdell, I. Piatetski-Shapiro, *The Arithmetic and Spectral Analysis of Poincaré Series, Perspectives in Mathematics*, vol. 13, Academic Press, Inc., Boston, MA, 1990.
- [HB96] D.R. Heath-Brown, A new form of the circle method, and its application to quadratic forms, *J. Reine Angew. Math.* 481 (1996) 149–206.
- [Kub74] R.M. Kubota, Waring’s problem for $\mathbb{F}_q[x]$, *Diss. Math. (Rozprawy Mat.)* 117 (1974) 60.
- [LPS88] A. Lubotzky, R. Phillips, P. Sarnak, Ramanujan graphs, *Combinatorica* 8 (3) (Sep 1988) 261–277.
- [Mar88] G.A. Margulis, Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators, *Probl. Pereda. Inf.* 24 (1) (1988) 51–60.
- [Mor94] M. Morgenstern, Existence and explicit constructions of $q + 1$ regular Ramanujan graphs for every prime power q , *J. Comb. Theory, Ser. B* 62 (1) (1994) 44–62.
- [MSS15] A.W. Marcus, D.A. Spielman, N. Srivastava, Interlacing families I: bipartite Ramanujan graphs of all degrees, *Ann. Math. (2)* 182 (1) (2015) 307–325.
- [Sar90] P. Sarnak, *Some Applications of Modular Forms*, Cambridge Tracts in Mathematics, vol. 99, Cambridge University Press, Cambridge, 1990.
- [Sar18] N.T. Sardari, Diameter of Ramanujan graphs and random Cayley graphs, *Combinatorica* (Aug 2018).
- [Sar19] N.T. Sardari, Optimal strong approximation for quadratic forms, *Duke Math. J.* 168 (10) (07 2019) 1887–1927.
- [Ste19] R.S. Steiner, On a twisted version of Linnik and Selberg’s conjecture on sums of Kloosterman sums, *Mathematika* 65 (3) (2019) 437–474.
- [SZ20] N.T. Sardari, M. Zargar, Sections of quadrics over $\mathbb{A}_{\mathbb{F}_q}^1$, arXiv e-prints, arXiv:1907.07839, Apr 2020.