Exploring Densities of Gaussian Quadratic Forms

Max Li

18.338 - Eigenvalues of Random Matrices
Fall 2020

Abstract

Quadratic forms are commonplace in many applications, such as in the cost function of the linear-quadratic regulator for a continuous- or discrete-time linear system. We are interested in exploring the characteristics of random quadratic forms, specifically Gaussian quadratic forms, where the coefficients come from some positive semi-definite matrix and the arguments are random multivariate Gaussian variables. We survey some results regarding the moments, densities, and generating functions of these Gaussian quadratic forms, and implement numerical approximations of the densities via saddlepoint approximations.

1 Introduction

Quadratic forms appear in many engineering applications, and random quadratic forms are rich in terms of their applications as well as analyses. To motivate our exploration of these objects, we provide an example from stochastic control, with that of a stochastic linear-quadratic regulator:

Example 1 (Linear quadratic stochastic control)**.** *Suppose we have the following finite horizon, discrete-time linear dynamical system, with dynamics*

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + \mathbf{w}(t), \qquad t = 0, \dots, T-1, \tag{1}$$

where the initial condition $\mathbf{x}(0)$ *is drawn from* $\mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{x}})$*, and the system is perturbed by additive noise* $\mathbf{w}(t) \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{w}})$*. The system matrix* A

may also be random, along with the matrix B modulating control inputs $\mathbf{u}(t)$, which are assumed to be deterministic. For stability, we set $\max_i |\lambda_i(A)| = 1$.

The objective is to obtain a state-feedback controller $\mathbf{u}(t) \triangleq \psi_t(\mathbf{x}(t))$ *for all time steps, where* ψ_t *is the* control policy *at time* t. *We do this by solving the* linear quadratic stochastic control problem*, which seeks to minimize the cost, given by*

$$\mathcal{J} = \mathbb{E}\left[\mathbf{x}^\intercal(T)Q_f\mathbf{x}(T) + \sum_{t=0}^{T-1}\mathbf{x}^\intercal(t)Q\mathbf{x}(t) + \mathbf{u}^\intercal(t)R\mathbf{u}(t)\right]. \tag{2}$$

This problem can be solved via dynamic programming, and the optimal policies $\psi_t^\star(\mathbf{x}(t))$ *will have the form*

$$\psi_t^\star(\mathbf{x}(t)) = \underset{\mathbf{v}}{\operatorname{argmin}}\left\{\mathbf{v}^\intercal R\mathbf{v} + \mathbb{E}\left[\min_{\psi_{t+1},\ldots,\psi_{T-1}}\mathcal{J}\right]\right\}. \tag{3}$$

These random quadratic forms appear in the cost function (2) for the linear quadratic stochastic control problem, and the value functions that dictate the optimal policy are also quadratic forms of the state (random vector) and a symmetric (random) matrix that is a function of A, B, Q, and R (specifically, in the case of Example 1, this matrix takes the form $P_t = A^\intercal P_{t+1}A - A^\intercal P_{t+1}B\left(B^\intercal P_{t+1}B + R\right)^{-1}B^\intercal P_{t+1}A + Q$, and is known as the discrete-time algebraic Riccati equation).

We are interested in quadratic forms for their applications in signal processing. Specifically, the notion of signal processing on graph-supported signals is central to *graph signal processing* [SM13], which extends the notion of a Fourier decomposition to a graph setting, and provides a toolkit to analyze signals on graphs. The identification of unexpected spatial delay distributions is related to the broader problem of outlier detection in multivariate data sets. One particular characteristic, *Total Variation (TV)*, is a measure of the smoothness of a signal supported on a graph. TV can be written as a quadratic form between graph-supported signals and a *shift operator* defined on the graph. This operator could be the graph adjacency matrix, or the (combinatorial) graph Laplacian. Such an outlier detection framework using the TV of a graph signal has some key advantageous: (1) The ability to identify outliers based on spatial distributions; (2) Decouple the detection of outliers based on magnitude versus spatial distribution; and (3) Provide

an interpretable explanation as to why a particular point was classified as an outlier. From an applications perspective, we are interested in interpretability to provide insights for a variety of stakeholders, ranging from air traffic flow managers to airline flight dispatchers and network analysts [LGPB20].

For the rest of this project report, we first discuss some basic notions behind graph signal processing, and motivate the total variation both as a measure of signal smoothness, as well as a prototypical example of the kinds of quadratic forms we are interested in exploring. We then survey some results for the eigenvalue spectrum of the graph Laplacian, and moments of Gaussian quadratic forms. We observe that the densities for general Gaussian quadratic forms tend to be "messy", in the sense that they can only be written as infinite expansions of other unwieldy densities or functions. Finally, we explore using saddlepoint approximations to obtain a closed-form density for Gaussian quadratic forms in a limited setting.

2 Fourier transform on graphs

Consider a signal $\mathbf{x} \in \mathbb{R}^{m \times 1}$ drawn from a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, supported on a graph $G = (V, E)$, where V is the set of $|V| = m$ nodes and $E \subseteq V \times V$ is the set of edges. We take $\mathbf{x}$ to be supported on V, i.e., there is a mapping $f : V \to \mathbb{R}$ from $i \in V$ to the i^{th} element, x_i, of $\mathbf{x}$. There is a weight map $w : E \to \mathbb{R}$ that assigns a weight w_{ij} to edge $(i, j) \in E$. These weights can be represented using an adjacency matrix $A(G) \in \mathbb{S}^{m \times m}$, where $[A]_{ij} = w_{ij}$. We restrict ourselves to undirected graphs, where $w_{ij} = w_{ji}$ and $A(G) = A(G)^\intercal$.

Let $\mathbf{X} \in \mathbb{R}^{n \times m}$ be the data matrix with n rows $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}, \ldots, \mathbf{x}^{(n)}\}$ and m columns. Denote by $\hat{\mu}_i = \frac{1}{n} \sum_{k=1}^{n} x_i^{(k)}$ the empirical mean of the graph signal at node i. We can also compute the sample Pearson correlation coefficient $r_{ij|\mathbf{X}}$ with which to weigh edge (i, j):

$$r_{ij|\mathbf{X}} = \frac{\sum_{k=1}^{n} \left(x_i^{(k)} - \hat{\mu}_i\right)\left(x_j^{(k)} - \hat{\mu}_j\right)}{\sqrt{\sum_{k=1}^{n} \left(x_i^{(k)} - \hat{\mu}_i\right)^2} \sqrt{\sum_{k=1}^{n} \left(x_j^{(k)} - \hat{\mu}_j\right)^2}}. \tag{4}$$

We now define some objects from spectral graph theory that we will use to motivate an analogue of the discrete Fourier transform on graphs.

Definition 1 (Graph Laplacian)**.** *The (combinatorial) graph Laplacian* $\mathcal{L}(G)$ *with respect to a graph with adjacency matrix* $A(G)$ *is* $\mathcal{L}(G) = D(G) - A(G)$*, where* $D(G) = [d_{ij}] \in \mathbb{R}^{m \times m}$ *is the diagonal degree matrix of* G *with* $d_{ii} = \sum_{j=1}^{m} w_{ij}$*.*

The graph Laplacian $\mathcal{L}(G)$ is a real symmetric matrix with a full set of orthogonal eigenvectors. The normalized eigenvectors are denoted by $v_i \in \mathbb{R}^{m \times 1}$, $i \in \{1, \ldots, m\}$, with $v_i^{\intercal} v_j = \delta_{ij}$, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

We sort the eigenvalues such that $\lambda_1(\mathcal{L}) \leq \lambda_2(\mathcal{L}) \leq \ldots \leq \lambda_m(\mathcal{L})$. Since the graph Laplacian has row sums of 0 (Definition 1), $v_1 = \mathbf{1}$ is the *constant* eigenvector corresponding to the eigenvalue $\lambda_1(\mathcal{L}) = 0$. Furthermore, the multiplicity of eigenvalues equal to 0 is the number of connected components in the underlying graph. Thus, if the correlation network is fully connected, then $0 = \lambda_1(\mathcal{L}) < \lambda_2(\mathcal{L}) \leq \ldots \leq \lambda_m(\mathcal{L})$, and $\text{span}\left(\{v_1, ..., v_m\}\right) \cong \mathbb{R}^{m \times 1}$. The set of eigenvectors of $\mathcal{L}(G)$ forms an orthogonal basis for $\mathbb{R}^{m \times 1}$. Thus, any vector $\mathbf{x} \in \mathbb{R}^{m \times 1}$ can be written as a linear combination of $\{v_1, ..., v_m\}$, i.e., there exist scalars α_i such that $\mathbf{x} = \sum_{i=1}^{m} \alpha_i v_i$.

Definition 2 (Graph Fourier Transform)**.** *The Graph Fourier Transform (GFT) of a graph signal vector* $\mathbf{x}$ *is the set of scalars* $\{\alpha_1, ..., \alpha_m\}$ *where* $\alpha_i = v_i^{\intercal} \mathbf{x}$*.*

To draw an analogy to the classical Fourier transform, the eigenvectors are equivalent to sinusoids on graphs, and the eigenvalues correspond to discrete frequencies. The scalar $\alpha_i \in \{\alpha_1, ..., \alpha_m\}$ represents the magnitude of contribution of the i^{th} eigenvector of "frequency" $\lambda_i(\mathcal{L})$. Similar to the notion of spectral energies for the classical Fourier transform, larger eigenvalues are associated with eigenvectors having higher graph spectral energies, as follows:

Definition 3 (Spectral and total energy)**.** *The spectral energy of* $\mathbf{x}$ *corresponding the* i^{th} *eigenvector is* α_i^2*, and the total energy of* $\mathbf{x}$ *is given by* $\|\mathbf{x}\|_2^2 = \sum_{i=1}^{m} \alpha_i^2$*.*

Note that the total energy of $\mathbf{x}$ is also equal to $\|\mathbf{x}\|_2^2$.

The graph Laplacian can also be used to compute a measure of the "smoothness" of a graph signal $\mathbf{x}$. This measure adopts a specific quadratic form involving $\mathcal{L}(G)$ and $\mathbf{x}$, called the *total variation*:

Definition 4 (Total variation)**.** *The total variation (TV) of a graph signal* $\mathbf{x}$ *with respect to the graph Laplacian* $\mathcal{L}(G)$ *is defined as:*

$$\mathrm{TV}(\mathcal{L}(G), \mathbf{x}) = \frac{1}{2} \sum_{i \neq j} w_{ij}(x_i - x_j)^2 = \mathbf{x}^\intercal \mathcal{L}(G)\mathbf{x}. \tag{6}$$

When $\mathbf{x}$ is a multivariate Gaussian, we will refer to $\mathrm{TV}(\mathcal{L}(G), \mathbf{x})$ as a *Gaussian quadratic form.* In general, $\mathcal{L}(G)$ may be replaced by any real, symmetric matrix.

Loosely speaking, the total variation measures the "smoothness" of a graph signal. A low value of TV corresponds to a graph signal that is said to be *smooth* relative to one with a higher TV. This notion of smoothness is related to how much the graph signal varies across adjacent nodes. When the difference in nodal signal values is large across a given edge, then its contribution to the TV is larger. The following proposition helps to interpret the GFT in terms of the TV and signal smoothness:

Proposition 1. *Suppose we examine one observation* $\mathbf{x}^{(k)}$ *of* $\mathbf{X}$ *and compute its GFT* $\{\alpha_1^{(k)}, ..., \alpha_m^{(k)}\}$*. Then, the following two statements are equivalent:*

(i) $\mathrm{TV}\left(\mathcal{L}(G), \mathbf{x}^{(k)}\right) = \left(\mathbf{x}^{(k)}\right)^\intercal \mathcal{L}(G)\mathbf{x}^{(k)}$.

(ii) $\mathrm{TV}\left(\mathcal{L}(G), \mathbf{x}^{(k)}\right) = \sum_{i=1}^{N} \left(\alpha_i^{(k)}\right)^2 \lambda_i(\mathcal{L})$.

Proof. The proof is straightforward through applying the definition of total variation and orthogonal eigenvectors of $\mathcal{L}(G)$. Starting with the definition for $\mathrm{TV}\left(\mathcal{L}(G), \mathbf{x}^{(k)}\right)$, we show that it is equivalent to $\sum_{i=1}^{m} \left(\alpha_i^{(k)}\right)^2 \lambda_i(\mathcal{L})$:

$$\begin{aligned} \mathrm{TV}\left(\mathcal{L}(G), \mathbf{x}^{(k)}\right) &= \sum_{j=1}^{m} \alpha_j^{(k)} v_j^\intercal \mathcal{L}(G) \sum_{i=1}^{m} \alpha_i^{(k)} v_i \\ &= \sum_{i,j} \alpha_i^{(k)} \alpha_j^{(k)} \lambda_i v_j^\intercal v_i = \sum_{i=1}^{N} \left(\alpha_i^{(k)}\right)^2 \lambda_i(\mathcal{L}). \end{aligned} \tag{7}$$

■

Proposition 1 formalizes the relationship between the GFT, the TV of a graph signal, and the spectral and total energies of a graph signal. Larger contributions of the eigenvector v_i to the GFT of $\mathbf{x}$ (i.e., larger values of α_i) result in a higher TV, translating to a less smooth graph signal. Similarly, the more energetic eigenvectors (i.e., larger values of $\lambda_i(\mathcal{L})$) contribute to a higher TV, resulting in a less smooth graph signal. Since the eigenvalues are sorted in ascending order with respect to index $i \in \{1, ..., m\}$, we compare eigenvalue magnitudes using the index i.

3 Graph topology and spectrum

As we will see in the later sections of this report, we will be interested in the density of the derived random variable TV $\left(\mathcal{L}(G), \mathbf{x}^{(k)}\right)$ when restricted to the case where $\mathbf{x}$ is distributed as a multivariate Gaussian random variable. These densities will all involve eigenvalues of the graph Laplacian $\mathcal{L}(G)$, or the *spectrum* of the graph Laplacian. In the simpler case where $\mathbf{x}$ has an identity covariance matrix, then the eigenvalues in the quadratic form densities will depend exactly on the graph spectrum.

Lemma 1. *Let G be an undirected cycle on m nodes, and let G be regular with degree d. Denote by $\{\lambda_1(A), \ldots, \lambda_m(A)\}$ the spectrum of the adjacency matrix $A(G)$. Then, the spectrum of the graph Laplacian $\mathcal{L}(G)$ is $\{d - \lambda_1(A), \ldots, d - \lambda_m(A)\}$, with*

$$\lambda_k(A) = 2\cos\left(\frac{2\pi k}{m}\right), \qquad k = \{1, ..., m\}. \tag{8}$$

Proof. Lemma 13.1.2 of [GR01]. ■

4 Quadratic forms of random variables

In our previous work with the TV of a graph signal vector, we were interested in constructing bounds that could be used to detect outlying values of TV with respect to its mean. In order to construct these bounds, or leverage any concentration inequalities that might be useful, we require expressions for the moments of these Gaussian quadratic forms.

Theorem 1. *For $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with valid $m \times m$ covariance matrix Σ, denote by $Q = \mathbf{x}^\intercal A \mathbf{x}$ where A is a $m \times m$ symmetric, real matrix. The r^{th} moment of Q is given by*

$$\mathbb{E}\left[Q\right]^r = \sum_{r_1=0}^{r-1} \binom{r-1}{r_1} g\left(r-1-r_1\right) \sum_{r_2=0}^{r_1-1} \binom{r_1-1}{r_2} g\left(r_1-1-r_2\right) \cdots \tag{9}$$

where $g(k) = 2^k k! \left(\operatorname{tr}(A\Sigma)^{k+1} + (k+1)\boldsymbol{\mu}^\intercal (A\Sigma)^k A\boldsymbol{\mu}\right)$ for $k \in \mathbb{N}_{\geq 0}$.

Proof. Theorem 3.2b.2 of [MP92] ■

From Theorem 1, the interesting moments of $Q = \mathbf{x}^\intercal A \mathbf{x}$ for applications, i.e., the mean and the variance, can be written explicitly:

$$\begin{aligned} \mathbb{E}\left[Q\right] &= \operatorname{tr}(A\Sigma) + \boldsymbol{\mu}^\intercal A \boldsymbol{\mu}, \\ \operatorname{Var}\left[Q\right] &= \mathbb{E}\left[Q\right]^2 - g(0)^2 \\ &= 2\operatorname{tr}(A\Sigma)^2 + 4\boldsymbol{\mu}^\intercal A\Sigma A\boldsymbol{\mu}. \end{aligned} \tag{10}$$

Unfortunately, unlike for the moments of $Q = \mathbf{x}^\intercal A \mathbf{x}$, no convenient, closed-form expression exists for a general Gaussian quadratic form when there are no assumptions on centrality and identity covariance. In particular, [MP92] provides several convergent expressions of the density $f_Q(q)$ for Q as infinite series expansions. These expansions can be written in terms of power series expansions, Laguerre series expansions, expansions in central χ^2 densities, confluent Hypergeometric functions, zonal polynomials, and densities of Gamma variates. We detail the Laguerre series expansion below, given the many connections explored in this course between Laguerre polynomials and Wishart matrices.

Let $\mathbf{x} \in \mathbb{R}^{m \times 1}$ be a multivariate Gaussian drawn from $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with possibly non-zero mean $\boldsymbol{\mu} \in \mathbb{R}^{m \times 1}$ and positive definite covariance matrix $\Sigma \in \mathbb{S}_{\succ 0}^{m \times m}$. We are interested in distribution of the derived random variable $Q = \mathbf{x}^\intercal A \mathbf{x}$, where A is a $m \times m$ symmetric, real, and positive semi-definite matrix. We will refer to $Q = \mathbf{x}^\intercal A \mathbf{x}$ as a Gaussian quadratic form. From [MP92], the density of Q is surprisingly complicated, and can be written in terms of a series expansion using generalized Laguerre polynomials $L_k^{(\alpha)}(x)$, expressed via the Rodrigues formula as

$$L_k^{(\alpha)}(x) = \frac{1}{x!} e^x x^{-\alpha} \left[\frac{d^k}{dx^k}\left(e^{-x} x^{k+\alpha}\right)\right], \qquad \alpha > -1, \quad k = 0, 1, ... \tag{11}$$

Specifically, by way of Theorem 4.2c.1 from [MP92], the density of Q using the Laguerre series expansion is

$$f_Q\left(\boldsymbol{\lambda};\mathbf{b};q\right)=\sum_{k=0}^{\infty}c_k\frac{k!}{2\beta\Gamma\left(\frac{m}{2}+k\right)}\left(\frac{q}{2\beta}\right)^{\frac{m}{2}-1}e^{-\frac{q}{2\beta}}L_k^{\left(\frac{m}{2}-1\right)}\left(\frac{q}{2\beta}\right), \tag{12}$$

for $q\in(0,\infty)$, where β is an arbitrary positive constant, c_k are power series expansion coefficients with $c_0=1$ and are dependent on $\boldsymbol{\lambda}=(\lambda_1,...,\lambda_m)$ and $\mathbf{b}=(b_1,...,b_m)$. The eigenvalues in $\boldsymbol{\lambda}$ belong to $\Sigma^{1/2}A\Sigma^{1/2}$ diagonalized by an orthogonal $m\times m$ matrix P, i.e., $P^{\intercal}\left(\Sigma^{1/2}A\Sigma^{1/2}\right)P=\mathrm{diag}(\boldsymbol{\lambda})$. The vector $\mathbf{b}$ is dependent on the mean and covariance of $\mathbf{x}$, and is computed as $\mathbf{b}^{\intercal}=P^{\intercal}\Sigma^{-1/2}\boldsymbol{\mu}$.

5 Numerical experiments

5.1 Saddlepoint approximations

Saddlepoint approximation provides an accurate, closed-form expression of an unknown probability density (mass) function for a continuous (discrete) random variable, given that the mass generating function and cumulant generating function for the random variable are known [Dan54, But07]. Compared to the three popular alternatives of approximations – enumerating exact probabilities, approximation via Gaussian densities, and brute force simulation with kernel density estimators – saddlepoint approximation oftentimes overcome issues of intractability, inaccuracy, and convergence issues.

Let X be an univariate, continuous random variable supported on $\mathcal{X}\subseteq\mathbb{R}$ with an unknown probability density function $f_X(x)$ that is non-zero on $\mathcal{X}$. Let $M_X(s)=\mathbb{E}\left[e^{sX}\right]$ be the moment generating function of X, and $K_X(s)=\ln M_X(s)$ be the cumulant generating function of X. The saddlepoint approximation $\widehat{f}_X(x)$ to $f_X(x)$ is given by

$$\widehat{f}_X(x)=\left(2\pi\frac{d^2}{ds^2}K_X(\hat{s})\right)^{-1/2}\exp\left(K(\hat{s})-\hat{s}x\right), \tag{13}$$

$\widehat{f}_X(x)$ is known as the *saddlepoint equation*, and $\hat{s}\overset{\Delta}{=}\hat{s}(x)$ is the *saddlepoint associated with* x, where $\hat{s}(x)$ is the solution to the following differential

equation in the cumulant generating function:

$$\frac{d}{ds}K(\hat{s}) = x. \tag{14}$$

We note that the saddlepoint equation serves as a meaningful approximation for $f_X(x)$ only for $x \in \mathcal{X}$, and that in general, the saddlepoint equation is not a true density, as $\int_{\mathcal{X}} \widehat{f}_X(x)\,dx \neq 1$. However, if we set a normalizing constant $c = \int_{\mathcal{X}} \widehat{f}_X(x)\,dx$, then $\widetilde{f}_X(x) \stackrel{\Delta}{=} c^{-1}\widehat{f}(x)$ is a valid density over the support $\mathcal{X}$. Finally, the discrete version of the saddlepoint approximation mirrors the continuous version, with the additional caveat that the (discrete) saddlepoint equation $\widehat{p}_X(x)$ is meaningful as an approximation to some probability mass function $p_X(x)$ on integer-valued arguments.

5.2 Approximating quadratic forms via saddlepoints

To construct the saddlepoint approximation, we refer to the following theorem from [MP92] that provides the moment generating function for Gaussian quadratic forms that we are interested in.

Theorem 2. *Let A be a real, symmetric $m \times m$ matrix, and $\mathbf{x} \in \mathbb{R}^{m\times 1}$ with $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Then, the moment generating function $M_Q(s)$ of $Q = \mathbf{x}^\intercal A\mathbf{x}$ is*

$$M_Q(s) = \det\left(I - 2sA\Sigma\right)^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\mu}^\intercal\left(I - \left(I - 2sA\Sigma\right)^{-1}\right)\Sigma^{-1}\boldsymbol{\mu}\right), \tag{15}$$

where I is the $m \times m$ identity. In the central case where $\boldsymbol{\mu} = \mathbf{0}$, the moment generating function reduces to

$$M_Q(s) = \det\left(I - 2sA\Sigma\right)^{-1/2}. \tag{16}$$

Furthermore, $M_Q(s)$ can be written in a scalar form involving the eigenvalues of $\Sigma^{1/2}A\Sigma^{1/2}$ and constants that depend on the mean $\boldsymbol{\mu}$ and precision Σ^{-1}. Specifically, let $\lambda_1, ..., \lambda_m$ be eigenvalues of $\Sigma^{1/2}A\Sigma^{1/2}$, and define the vector of constants $\mathbf{b} = (b_1, ..., b_m)^\intercal = P^\intercal\Sigma^{-1/2}\boldsymbol{\mu}$, where P is any $m \times m$ orthogonal matrix that diagonalizes $\Sigma^{1/2}A\Sigma^{1/2}$. Then, $M_Q(s)$ can be rewritten as follows:

$$M_Q(s) = \begin{cases} \exp\left(s\sum_{j=1}^{m}\frac{b_j^2\lambda_j}{1-2s\lambda_j}\right)\prod_{j=1}^{m}(1-2s\lambda_j)^{-\frac{1}{2}}, & \text{if } \boldsymbol{\mu}\neq\mathbf{0}, \\ \prod_{j=1}^{m}(1-2s\lambda_j)^{-\frac{1}{2}}, & \text{if } \boldsymbol{\mu}=\mathbf{0}. \end{cases} \tag{17}$$

Proof. Theorems 3.2a.1, 3.2a.2, and Corollary 3.2a.1 of [MP92]. ■

In order to deploy saddlepoint approximation, we first compute the cumulant generating function, as well as its first and second derivatives with respect to s. Starting with the simpler case of $\boldsymbol{\mu} = \mathbf{0}$, we have that

$$\begin{aligned} K_Q(s) &= \ln\left(\prod_{j=1}^{m}(1-2s\lambda_j)^{-\frac{1}{2}}\right) = \sum_{j=1}^{m}\ln\left((1-2s\lambda_j)^{-\frac{1}{2}}\right), \\ \frac{d}{ds}K_Q(s) &= \sum_{j=1}^{m}\frac{\lambda_j}{1-2s\lambda_j}, \\ \frac{d^2}{ds^2}K_Q(s) &= \sum_{j=1}^{m}\frac{2\lambda_j^2}{(1-2s\lambda_j)^2}. \end{aligned} \tag{18}$$

For the non-central case where $\boldsymbol{\mu} \neq \mathbf{0}$, the cumulant generating function and its first two s derivatives are

$$\begin{aligned} K_Q(s) &= \sum_{j=1}^{m}\left(\frac{b_j^2\lambda_j s}{1-2s\lambda_j} + \ln\left((1-2s\lambda_j)^{-\frac{1}{2}}\right)\right), \\ \frac{d}{ds}K_Q(s) &= \sum_{j=1}^{m}\frac{\lambda_j\left(1+b_j^2-2s\lambda_j\right)}{(1-2s\lambda_j)^2}, \\ \frac{d^2}{ds^2}K_Q(s) &= 2\sum_{j=1}^{m}\frac{\lambda_j^2\left(-1-2b_j^2+2s\lambda_j\right)}{(-1+2s\lambda_j)^3}. \end{aligned} \tag{19}$$

The difficulty with saddlepoint approximations arises from the fact that solving for the saddlepoint requires one to solve successively higher-degree polynomials. For the central case, the saddlepoint for the case of $m = 1$ is

$$\hat{s} = \frac{1}{2}\left(\frac{1}{\lambda_1} - \frac{1}{x}\right), \tag{20}$$

and for $m = 2$ we have the saddlepoints given by:

$$\hat{s} = \frac{1}{4\lambda_1\lambda_2 x}\left(\pm\sqrt{\lambda_1^2x^2 - 2\lambda_1\lambda_2x^2 + \lambda_2^2\left(4\lambda_1^2 + x^2\right)} + \lambda_2 x + \lambda_1\left(x - 2\lambda_2\right)\right). \tag{21}$$

For the central case with $m = 3$, a "solution" is obtainable (e.g., via the `Reduce` function in `Mathematica` specifying solutions to cubic polynomials), but it is not convenient for computation. For the non-central case, the one-node (i.e., $m = 1$) case has saddlepoints given by

$$\hat{s} = \frac{\pm\sqrt{4b_1^2\lambda_1^3x + \lambda_1^4} - \lambda_1^2 + 2\lambda_1 x}{4\lambda_1^2 x}, \tag{22}$$

with larger values of m being analogously difficult to obtain.

We plot in Figure 1 the saddlepoint approximation of the Gaussian quadratic form $Q = \mathbf{x}^\intercal A\mathbf{x}$ for a 2-node graph in the central case, and observe that the (non-normalize) saddlepoint approximation seems to follow the distribution of Q quite well. The approximation is close for both an identity covariance matrix, as well as for any arbitrary valid covariance matrix. We also explored a randomized version of this approximation by repeatedly sampling covariance matrices, and fitting a single saddle approximation using the average of the eigenvalues of all sampled covariance matrices. This fit can be seen in Figure 2.

6 Concluding remarks

Motivated by the *total variation* of a graph signal vector with respect to its underlying graph, we explored the moments, densities, and density approximations of Gaussian quadratic forms $Q = \mathbf{x}^\intercal A\mathbf{x}$. In particular, we take $\mathbf{x}$ to be a multivariate Gaussian vector, and A to be any positive semi-definite matrix, with an eye towards adjacency and graph Laplacian matrices of unweighted graphs. Some future questions of interest relate to the various expansions of Q, in particular how to amenably implement them to approximate densities of non-central, non-identity covariance signals $\mathbf{x}$. Other ways to approximate Q, such as using Gamma distributions, or transforming the saddlepoint problem to a rank one eigenvalue update problem, may also lead to some interesting applications.

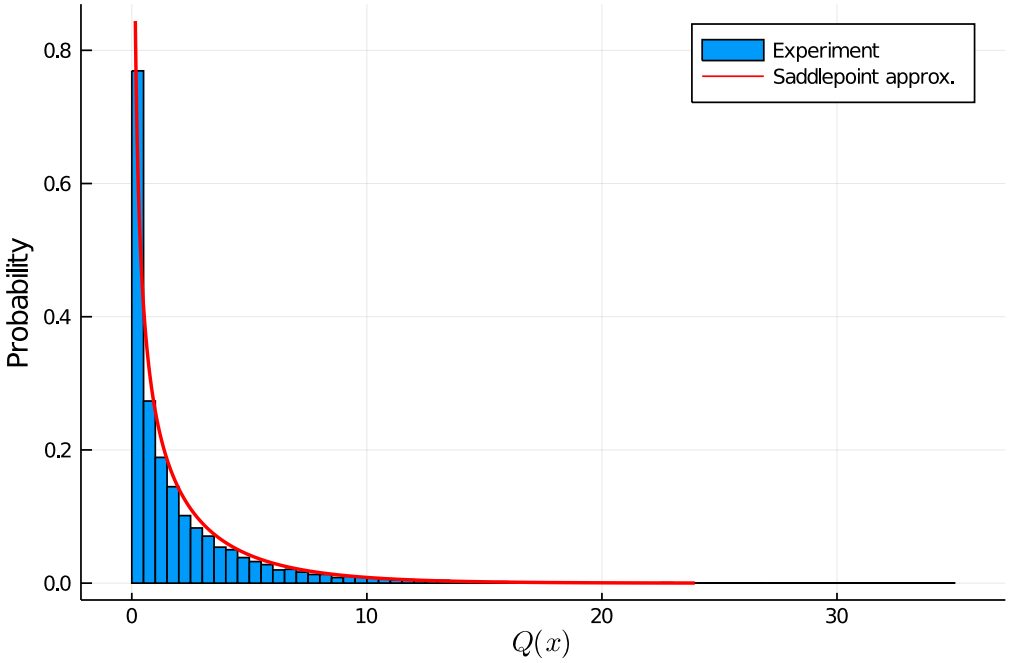

Figure 1: Saddlepoint approximation to Q in the central case for $m = 2$.

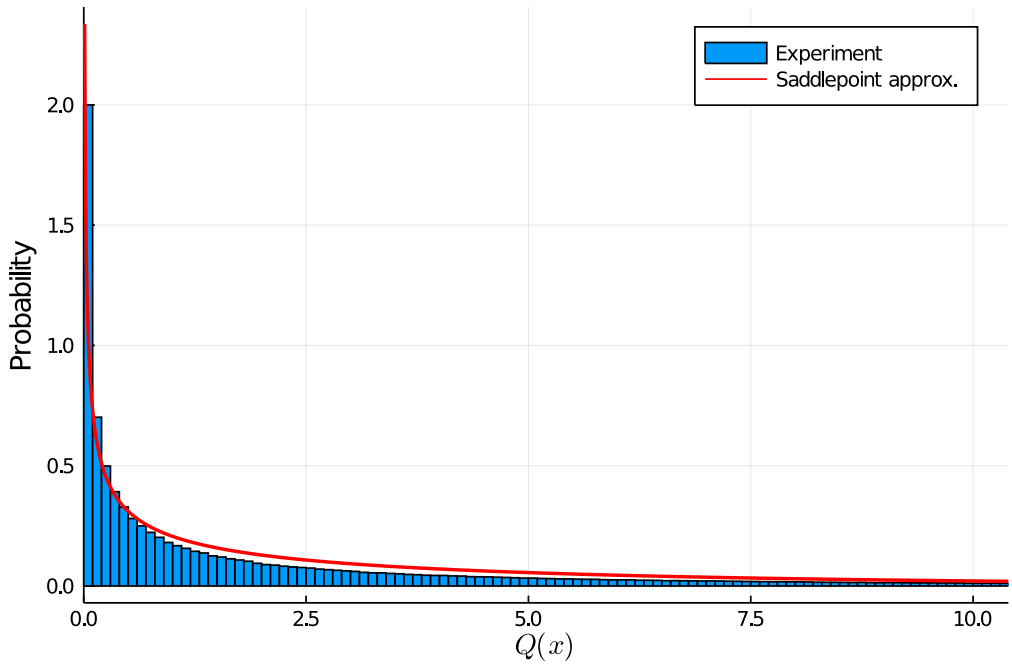

Figure 2: Saddlepoint approximation to Q in the central case for $m = 2$, with randomized covariance matrices.

References

[But07] Ronald W. Butler. *Saddlepoint Approximations with Applications*. Cambridge University Press, 2007.

[Dan54] H. E. Daniels. Saddlepoint approximations in statistics. *Ann. Math. Statist.*, 25(4):631–650, 12 1954.

[GR01] C. Godsil and G. F. Royle. *Algebraic Graph Theory*. Springer, 2001.

[LGPB20] M. Z. Li, Karthik Gopalakrishnan, Kristyn Pantoja, and Hamsa Balakrishnan. Graph signal processing techniques for analyzing aviation disruptions. *Transportation Science*, 2020.

[MP92] A. M. Mathai and Serge B. Provost. *Quadratic Forms in Random Variables*. Marcel Dekker, Inc., 1992.

[SM13] A. Sandryhaila and J. M. F. Moura. Discrete signal processing on graphs. *IEEE Transactions on Signal Processing*, 61(7):1644–1656, April 2013.