Efficient Bayesian inference of systemic risk interlinkages

Veni Arakelian*[1] and Apostolos Chalkis[2,3,4]

[1]Economic Analysis and Investment Strategy, Piraeus Bank, Greece
[2]Department of Informatics & Telecommunications, National & Kapodistrian University of Athens, Greece
[3]ATHENA Research & Innovation Center, Greece
[4]GeomScale org

Abstract

We study the systematic risk interlinkages between European banks of two core countries, France and Germany, and the peripheries, Greece, Ireland, Italy, Portugal and Spain. To address this problem we introduce a novel computational and algorithmic framework to incorporate three Bayesian models to inference on the covariance matrix. The models are differentiated by the choice of prior distribution. Our main algorithmic tool is a Markov Chain Monte Carlo (MCMC) algorithm to sample from any given probability density over the correlation matrices. We also introduce a novel geometric representation for the set of the correlation matrices which allow the MCMC sampler to mix fast for matrices of size in a few hundred.

1 Introduction

The systemic risk as a function of the financial system's architecture and the size of the financial institutional participants is still under debate among

*Disclaimer: This study expresses the views of the authors and it should not be reported as representing the views of the Piraeus Bank.

Acknowledgement: This paper has received funding from the European Union's Horizon 2020 research and innovation program "FIN-TECH: A Financial supervision and Technology compliance training programme" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA). Also, it is based upon work from COST Action 19130, supported by COST (European Cooperation in Science and Technology).

regulators and researchers, especially that large banks are found to have an integral role at the center of the recent financial crisis, see Laeven et al. (2016). The authors study the significant variation in the cross-section of systemic risk measures of large banks during the recent financial crisis in a broad sample of countries, intending to identify bank-specific factors, like banks size, capital, funding, and activities, that determine systemic risk and shed light on the ongoing debate on the merits of restricting bank size, imposing capital surcharges on large banks, and/or restricting their unstable funding and risky activities. Several theories are supporting the view that large and complex banks contribute to systemic risk. According to one view, which the authors name it "the unstable banking hypothesis", large banks tend to engage more in risky activities (e.g., trading) and be financed more with short-term debt, which makes them more vulnerable to generalized liquidity shocks and market failures such as liquidity shortages and fire sales (Kashyap et al. (2002); Shleifer and Vishny (2010); Gennaioli et al. (2013). In our view, the importance and the implications of this "hypothesis" is not much addressed by the literature in particular the cross border bank lending. The cross border bank lending rose rapidly until the financial crisis in 2008. Increased lending is often seen as an unambiguously good thing, as it evens out rates of return on investment between countries,and hence raises aggregate output. However,this will depend on reasonable evaluations of the risks involved in cross border lending, and if these are inaccurate the outcome may not be so beneficial. Cross border lending is a source of capital in good times but can propagate risks from home to host and the reverse in bad times. Cross border lending is seen as a service trade activity, rather than as a portfolio decision, and stress the primacy of lender characteristics over those of borrowers, suggesting supply factors are important in the market for cross border lending. After the financial crisis in 2008 bank lending became more constrained by regulators in many countries, and after the Euro Area sovereign debt crisis starting in 2011 cross border lending within Europe fell more noticeably than elsewhere. Therefore the determination of the correlation structure among the group of countries is of main importance to identify systemic risk linkages.

The literature discussing measures of banks' systemic risk is vast and Benoit et al. (2017) provide an excellent survey of systemic risk measures. The literature either uses only market data, i.e, financial returns or credit default swaps (CDS) (e.g., Billio et al. (2012), Acharya et al. (2012), Allen et al. (2012), Adrian and Brunnermeier (2016)), or enriches the dataset with balance sheet data (among the others, Brownlees and Engle (2017)) to measure systemic risk. A combination of both microeconomic and macroeconomic data is used by Calabrese and Giudici (2015), who build a statistical model of

bank distress, based on the balance sheet of a bank, and on macroeconomic information on the country where the bank operates. Several authors discuss the systemic risk vulnerability in the context of financial networks facing the challenge to provide not only a good fit but also a good interpretation. This may result in choosing a model that has little support from the data, leading to predictions worse than could be obtained with other models. Additionally, the graphical models are essentially static, photographing a situation in a given period. This assumption seems to be restrictive in economics, in the case of variables that change over time, for example, during periods of financial stress. Giudici and Green (1999) and recently Ahelegbey et al. (2016) proposed more advanced, Bayesian, graphical models to overcome this limitation. Battiston et al. (2012) contribute to the debate on the resilience of financial networks by introducing a dynamic model for the evolution of financial robustness, showing that, in the presence of financial acceleration and persistence, the probability of default does not decrease monotonically with diversification. As a result, the financial network is most resilient for an intermediate level of connectivity. Diebold and Yilmaz (2014) propose several connectedness measures built from pieces of variance decompositions, and they argue that they provide natural and insightful measures of connectedness. They also show that variance decompositions define weighted, directed networks, so that our connectedness measures are intimately related to key measures of connectedness used in the network literature. Building on these insights, they track daily time-varying connectedness of major US financial institutions' stock return volatilities in recent years, with emphasis on the financial crisis of 2007–2008. Abedifar et al. (2017) compared the results of three different measures to gauge systemic risk, employing an application of the graphical network model described by Giudici and Spelta (2016) to identify the most interconnected banking sector. Avdjiev et al. (2019) based on a tensor decomposition method that extracts an adjacency matrix from a multi-layer network, propose a distress measure for national banking systems that incorporates not only banks' CDS spreads, but also how they interact with the rest of the global financial system via multiple linkage types, using banks' foreign exposures.

Our approach is to estimate the correlation matrix between the random variables under study. To achieve this we provide a novel computational and algorithmic framework to incorporate three Bayesian models which inference on the covariance matrix of Gaussian data. The models are differentiated from the choice of the prior distribution. Our framework provides a Markov Chain Monte Carlo (MCMC) algorithm which could be used to sample from the corresponding joint posterior distributions without requiring any modifications when we change the model we use. In particular, the MCMC algorithm can sample from any given probability density function supported on the same

set that supports the aforementioned posterior distributions.

The posterior distributions for Gaussian data involves computations with the inverse covariance $\Omega = \Sigma^{-1}$. Our approach is the same with that in Barnard et al. (2000); Wong et al. (2003); that is considering the decomposition $\Omega = T \times C \times T$, where C is a correlation matrix and T a diagonal matrix with entries $T_i = \Omega_{ii}^{1/2}$. Thus, each T_i^2 is the inverse of the corresponding partial variance. Clearly, in our model, sampling from the posterior distribution is reduced to sample from the joint distribution between C and T. Since $T_i > 0$ the latter means that the support of the posterior distribution is the Cartesian product between the set of correlation matrices and the set $[0, +\infty]$.

Undoubtedly, the sampling of correlation matrices is crucial mainly when the correlation matrix becomes large. We specialize two MCMC algorithms to sample from that set. The first –which is used by the three Bayesian models– samples according to any probability density function and the second generates uniformly distributed correlation matrices. We experimentally show that our samplers achieves superior mixing than the previous MCMC samplers, i.e. To estimate a 100×100 concentration matrix it takes $\sim$ 1hour.

We develop open source `MATLAB` software to implement the two MCMC algorithms as well as the three Bayesian models. Our implementation scales up to size of matrix equal to a few hundreds. Considering sampling uniformly distributed correlation matrices, BILLIARD_WALK achieves a PSRF < 1.1 in $\sim$ 30min when generating uniformly distributed 100×100 correlation matrices. Generating uniformly correlation matrices is crucial in several applications in biology (Numpacharoen, 2012) and in simulation studies (Barnard et al., 2000).

The structure of the paper is the following: Section 2 explains the Bayesian estimation and reviews some prior formulations and presents the prior used. Section 3 reviews the existing methods to generate correlation matrices. Section 4 develops the geometric framework we use to represent correlation matrices. Section 5 presents the MCMC algorithms to sample from the set of correlation matrices according to any given probability density function. Section 6 evaluates the performance of the model. Section 7 applies the models to analysts' recommendations and takes up the question of the existence of abnormal returns with the question of analysts' herding behavior. Section 8 concludes and presents the main direction of our future work. In Appendix we develop algorithms that perform efficient geometric operations on the set of correlation matrices and are needed for the implementation of the MCMC samplers.

2 Bayesian inference for the covariance matrix

We discuss Bayesian models to estimate the inverse covariance matrix –also known as the concentration matrix or precision matrix– from Gaussian data. More formally, the problem is to estimate the concentration matrix when k observations $y_1, \ldots, y_k$ are given such that,

$$y_t \sim \mathcal{N}(0, \Sigma),\ y_t \in \mathbb{R}^n,\ t = 1, \ldots, k. \tag{1}$$

Let the concentration matrix $\Omega = \Sigma^{-1}$. Then, From Equation (1) the likelihood of Ω is,

$$\begin{aligned} p(Y|\Omega) &= \{\det(2\pi\Omega^{-1})\}^{-k/2} exp\bigg(-\frac{1}{2}\sum_{t=1}^{k} y_t^T \Omega y_t \bigg) \\ &\propto \det(\Omega)^{k/2} exp\bigg\{ -\frac{1}{2}\mathrm{tr}(\Omega S_y) \bigg\}, \end{aligned} \tag{2}$$

where $S_y = \sum_{t=1}^{k} y_t y_t^T$. Our approach is to use the decomposition of Ω introduced in Barnard et al. (2000), that is we let

$$\Omega = T \times C \times T, \tag{3}$$

where C is a correlation matrix and T is a diagonal matrix. Each entry of matrix T is $T_i = \Omega_{ii}^{1/2},\ i = 1, \ldots, n$ so that T_i^2 corresponds to the inverse partial variance of $y_{i,t}$. Moreover, the partial correlation coefficients ρ^{ij} are given by,

$$\rho_{ij} = -\Omega_{ij}/(\Omega_{ii}\Omega_{jj})^{1/2} = -C_{ij}, \tag{4}$$

which means that C contains the negative of the partial correlation coefficients. Consequently, the likelihood in Equation (2) becomes,

$$p(Y|C, T) \propto \det(T)^k \det(C)^{k/2} exp\bigg\{ -\frac{1}{2}\mathrm{tr}(TCTS_y) \bigg\}. \tag{5}$$

Similarly to Barnard et al. (2000) we assume that in the prior the elements Ω_{ii} are independent and identically distributed, and are independent of the elements of C. As they notice in Barnard et al. (2000) the flexibility in dealing with tails of individual components is a key practical advantage of this separation strategy.

The geometrical framework to represent correlation matrices and the MCMC sampler we introduce allow us to use several priors for both the

T_i and the correlation matrix C while our sampler handles all the possible different choices of priors without requiring any modification.

Priors for the partial precisions Ω_{ii}. The goal is to set a prior such that T_i are independent and identically distributed and the prior is also uninformative. In Barnard et al. (2000) they consider T_i as a vector $T \in \mathbb{R}^n$ and they use the prior

$$T \sim \mathcal{N}(\xi, \Lambda), \tag{6}$$

where the matrix $\Lambda \in \mathbb{R}^{n \times n}$ is diagonal, that is, they choose independent log normal distributions for each of the standard deviations. They also study the choice of independent scaled inverted chi-squared distributions for each of the variances, as this is the commonly used conjugate prior for a variance.

In Wong et al. (2003) the T_i follows a gamma distribution with parameters α and β,

$$p(T_i) \propto T_i^{2\alpha-1} exp(-\beta T_i^2). \tag{7}$$

To make the prior uninformative they choose both α and β to be small, e.g. $\alpha = 10^{-10}$ and $\beta = 10^{-8}$.

Priors for the concentration matrix C. In Barnard et al. (2000) they consider two choices for the prior $p(C)$. The first is the jointly uniform prior for the partial correlations ρ_{ij} and the second is the marginally uniform prior for each ρ^{ij}. For the first one clearly

$$p(C) \propto 1. \tag{8}$$

For this choice the marginal priors on each ρ_{ij} are not uniform. They favor values close to zero over values close to ± 1. However, in several applications, it is desirable to have a prior that favors value close to zero.

To obtain a marginally uniform distribution they first consider as prior for the covariance matrix Σ the inverse-Wishart distribution, $W_n^{-1}(I, \nu)$, $\nu \geq n$. Then, they prove that marginal distribution of ρ_{ij} is

$$p(\rho^{ij}) \propto (1 - \rho_{ij})^{\frac{\nu-n-1}{2}}. \tag{9}$$

Clearly, when $\nu = n + 1$ the marginal of ρ_{ij} is uniform. Moreover, we have a family of priors for C parameterized by a single parameter ν. This prior could favor larger values of ρ_{ij} that the jointly uniform prior.

A third popular choice for the prior of Ω is the Jeffreys prior Robert et al. (2009); Geisser and Cornfield (1963); Geisser (1965) which lead to

$$p(C) \propto |C|^{(k+1)/2},\ k \in \mathbb{N}_+. \tag{10}$$

However, the use of Jeffreys prior in many cases fails to shrink the eigenvalues appropriately and suffers from estimation errors when C is close to a singular matrix.

3 Literature review on generating correlation matrices.

This is a relatively difficult problem due to three constraints imposed on a rectangular matrix: positive definiteness —*i.e.,* symmetric matrices with non-negative eigenvalues—, fixed unit diagonal elements and non-diagonal elements bounded in $[-1, 1]$. In particular the set of correlation matrices can be written as follows,

$$\mathcal{C}_R := \{R^{jk} : R^{jk} = 1\ (j = k),\ |R^{jk}| < 1\ (j \neq k) \text{ and R is pos. def.}\}. \tag{11}$$

The simplest method for constructing a correlation matrix is to use the rejection sampling method, which generates correlation coefficients using uniform random variables in the closed interval $[-1, 1]$. Subsequently, we check whether the matrix is positive definite and, if not, another correlation matrix is generated. This procedure is repeated until a valid matrix is obtained. In general, the rejection sampling method is considered inefficient for the large-scale construction of correlation matrices (typically when the number of partial correlations is larger than 4 (Rousseeuw and Molenberghs, 1994)).

Instead, for large-dimensional problems, there are several techniques for generating a correlation matrix. In (Chalmers, 1975; Bendel and Mickey, 1978; Davies and Higham, 2000; Dhillon et al., 2005; Waller, 2020) they generate correlation matrices with predetermined eigenvalues and spectrum. In (Holmes, 1991) they are based on a random Gram matrix. In (Marsaglia and Olkin, 1984) they generate correlation matrices with a given mean value, structure or eigenvalues. In (Budden Mark, 2008; Joe, 2006; Numpacharoen, 2012) they generate a correlation matrix in which each correlation coefficient is distributed within its boundaries. In (Liu and Daniels, 2006; Zhang et al., 2006; Wong et al., 2003; Pitt et al., 2006) they generate a correlation matrix with Markov Chain Monte Carlo algorithms (to sample from the posterior distribution in certain Bayesian models). Other methods and approaches can be found in (Kollo and Ruul, 2003; Emmert-Streib et al., 2019; Hardin et al., 2013; Brown and Zidek, 1994; Alvarez et al., 2016).

The main disadvantages of the current methods is that the most of them are model or problem wise oriented. That is, they are specialized methods for sampling from a certain (posterior) distribution in order to solve a specific problem or to build a specific (Bayesian) model. Moreover, previous methods are computationally expensive while their mixing time is not well studied and thus, typically, they do not scale in more than a few decades of dimensions.

4 Geometric representation of correlation matrices

We introduce a new geometric and computational framework to represent the set of correlation matrices $\mathcal{C}_R$ with a convex body $K \in \mathbb{R}^p$ where p is the number of non-diagonal elements in the upper (or lower) triangular part. Then, each point in the interior of K represents a correlation matrix. We define the body K as the intersection of the hypercube $H = [-1, 1]^p$ and a spectrahedron $S \in \mathbb{R}^p$. The first guarantees that the non-diagonal elements lie in $[-1, 1]$ and the second that the matrix has unit diagonal elements and is positive definite. Spectrahedra are the feasible region of semidefinite programs as the polytopes are the feasible region of linear programs. They are convex bodies with –in general– non linear boundary.

Then, given a probability density function (pdf) $\pi(R)$ supported on $\mathcal{C}_R$, we provide a computational framework which allows to sample from $\mathcal{C}_R$ according to π. We directly sample from K using several well known Markov Chain Monte Carlo (MCMC) algorithms.

Spectrahedra are probably the most well studied shapes after polyhedra. A spectrahedron $S \subset \mathbb{R}^p$ is the feasible set of a linear matrix inequality. That is, let

$$\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{A}_0 + x_1\boldsymbol{A}_1 + \cdots + x_p\boldsymbol{A}_p, \tag{12}$$

where $\boldsymbol{A}_i \in \mathbb{R}^{n\times n}$ are symmetric matrices; then the corresponding spectrahedron is the set,

$$S = \{\boldsymbol{x} \in \mathbb{R}^p \,|\, \boldsymbol{F}(\boldsymbol{x}) \succeq 0\}, \tag{13}$$

where $\succeq$ denotes positive semidefiniteness. We assume throughout that S is bounded of dimension p. Spectrahedra are convex sets and every polytope is a spectrahedron, but not the opposite. They are the feasible regions of semidefinite programs Ramana and Goldman (1999) in the way that polytopes are feasible regions of linear programs.

Now let,

$$\mathcal{C}_R \in \mathbb{R}^{n\times n} \text{ and } p = n(n-1)/2.$$

We let $\boldsymbol{x} \in \mathbb{R}^p$ to represent the non-diagonal elements of a correlation matrix. Thus, we first impose that the point $\boldsymbol{x}$ belongs to the interior of

$$H := [-1, 1]^p,$$

that is the p-dimensional hypercube. To impose positive definiteness we define the following ordered set

$$E_n := \Big((1,2), (1,3), \ldots, (1,n), (2,3), \ldots, (2,n), \ldots, (n-1,n) \Big). \tag{14}$$

Then, let the following map,

$$\begin{aligned} &\phi_n \ : \ \mathbb{N} \mapsto \mathbb{N} \times \mathbb{N}, \\ &\phi_n(i) := E_n[i], \ i \in [|E_n|] \end{aligned} \tag{15}$$

where $E_n[i]$ is the i-th pair in E_n. Clearly, E_n enumerates row-wise the indices of the elements in the upper triangular part of a matrix in $\mathbb{R}^{n \times n}$. In the sequel, we set the matrices $\boldsymbol{A}_i$ in Equation (12). Explicitly, let

$$\boldsymbol{A}_0 = I_n, \ \boldsymbol{A}_i = \boldsymbol{B}_i + \boldsymbol{B}_i^T, \ i \in [p], \tag{16}$$

where I_n is the identity matrix and the matrix $\boldsymbol{B}_i \in \mathbb{R}^{n \times n}$,

$$\boldsymbol{B}_i := \Big\{ B_i^{jk} : \ B_i^{jk} = 0 \ \Big(\phi_n(i) \neq (j,k) \Big), \ B_i^{jk} = 1 \ \Big(\phi_n(i) = (j,k) \Big) \Big\}. \tag{17}$$

Then, the set S stands for the set of positive semidefinite matrices with unit diagonal elements, while the intersection with the hypercube H guarantees that all the non-diagonal elements have absolute value smaller than 1. Thus, the set of correlation matrices can be represented by the interior of the following convex body,

$$K = \{ \boldsymbol{x} \in \mathbb{R}^p \mid \boldsymbol{A}_0 + x_1 \boldsymbol{A}_1 + \cdots + x_d \boldsymbol{A}_d \succeq 0 \} \cap [-1, 1]^p = S \cap H, \tag{18}$$

where the matrices $\boldsymbol{A}_i \in \mathbb{R}^{n \times n}$ are given by Equation (16). Thus, the correlation matrices can also be seen as a convex subset of the hypercube H. Moreover, the following map,

$$\boldsymbol{F}(\boldsymbol{x}) \ : \ K \leftrightarrow \mathcal{C}_R \tag{19}$$

defines a bijection from K to $\mathcal{C}_R$. For any $\boldsymbol{x} \in \mathbb{R}^p$, it is clear that,

$$\boldsymbol{F}(\boldsymbol{x}) \equiv \begin{bmatrix} 1 & x_1 & x_2 & \ldots & x_{n-1} \\ & 1 & x_n & \ldots & x_{2n-3} \\ & & \ddots & & \\ & & & 1 & x_p \\ & & & & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}, \ \boldsymbol{x} \in \mathbb{R}^p. \tag{20}$$

4.0.1 A three-dimensional example

Let us consider an example for the 3-dimensional case. Let the hypercube $H := [-1, 1]^3$ and the spectrahedron $S := \{\boldsymbol{x} = (x_1, x_2, x_3) \in \mathbb{R}^3 \mid \boldsymbol{F}(\boldsymbol{x}) \succeq 0\}$, where,

$$\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{A}_0 + x_1\boldsymbol{A}_1 + x_2\boldsymbol{A}_2 + x_3\boldsymbol{A}_3,$$

is the LMI in Equation (12) and,

$$\boldsymbol{A}_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ \boldsymbol{A}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \boldsymbol{A}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \ \boldsymbol{A}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{21}$$

Notice that for any point $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ in the interior of $K = S \cap H$ the matrix

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{bmatrix}$$

is a correlation matrix. Thus, the set of 3×3 correlation matrices can be represented by the interior of $K = S \cap H$. We use Hit-and-Run (see Section 5) to sample uniformly from K. The plots in Figure 1 illustrate the sample from two different angles.

5 MCMC algorithms to sample correlation matrices

Considering the MCMC sampling algorithms, we introduce Reflective Hamiltonian Monte Carlo (Afshar and Domke, 2015) (ReHMC) and Billiard walk (Gryazina and Polyak, 2012) (BILLIARD_WALK) for the case of correlation matrices. The first random walk can be used to sample correlation matrices from any given probability distribution. The BILLIARD_WALK can be used to sample only from the uniform distribution, while extended experimental results Chalkis et al. (2020) have shown superior efficiency compared to all the other random walks for uniform sampling for several convex bodies.

Next, we present the two random walks.

5.1 Billiard Walk

Billiard walk was first introduced in Gryazina and Polyak (2012) for uniform sampling from general convex bodies. Thus, we introduce BILLIARD_WALK (Algorithm 1) to sample from the set of correlation matrices K under the

Algorithm 1: BILLIARD_WALK($\pi,\ \boldsymbol{x},\ W,\ \rho,\ \delta$) (BiW)

Input : a point $\boldsymbol{x} \in \mathbb{R}^p$; the number of steps W; an upper bound ρ on the number of reflections; a parameter τ to control the length of the trajectory

Require: $\boldsymbol{x} \in K,\ W \in \mathbb{N},\ \rho \in \mathbb{N},\ \delta \in \mathbb{R}_+$

for $i = 1, \ldots, W$ **do**
  $L \leftarrow -\delta \ln \eta$;
  $\eta \leftarrow_R \mathcal{U}((0,1))$; // *choose length*
  $\boldsymbol{v} \leftarrow_R \mathcal{U}(\partial\mathcal{B}_d)$; // *choose direction*
  $\boldsymbol{x}_0 \leftarrow \boldsymbol{x}$;
  do
    $\tau \leftarrow$ INTERSECTION($\boldsymbol{x},\ \boldsymbol{v}$); // *get the intersection time*
    $\tilde{\boldsymbol{x}},\ \tilde{\boldsymbol{v}} \leftarrow$ REFLECTION($\boldsymbol{x},\ \tau,\ \boldsymbol{v}$);
    if$(L < \tau)$ **then** $\boldsymbol{x} \leftarrow \boldsymbol{x} + L\boldsymbol{v}$ **else** $\boldsymbol{x} \leftarrow \tilde{\boldsymbol{x}}$; $\boldsymbol{v} \leftarrow \tilde{\boldsymbol{v}}$;
    $L \leftarrow L - \tau$;
  while $L > 0$;
  if $\#\{\textit{reflections}\} > \rho$ **then** $\boldsymbol{x} \leftarrow \boldsymbol{x}_0$;
return $\boldsymbol{x}$

uniform distribution. At each step, being at the current point of the walk $\boldsymbol{x}$, it chooses uniformly a direction vector $\boldsymbol{v}$ and a number L, where $L = -\delta \ln \eta$ and $\eta \sim U(0,1)$. Then, it moves at the direction of $\boldsymbol{v}$ for distance L. If during the movement, it hits the boundary without having covered the required distance L, then it continues on the reflected trajectory. This results a new point $\boldsymbol{y}$ inside K. We repeat the procedure from $\boldsymbol{y}$. Asymptotically it converges to the uniform distribution over K. If the number of reflections exceeds an upper bound ρ, it stays at $\boldsymbol{x}$.

The implementation of BILLIARD_WALK requires to compute the intersection $\ell(t) \cap \partial K$ and to compute the reflection of $\ell(t)$ when it hits the boundary. Considering the mixing time of BILLIARD_WALK there is not a known theoretical bound.

5.2 Reflective Hamiltonian Monte Carlo

We introduce Reflective Hamiltonian Monte Carlo (REHMC), a HMC-based algorithm to sample from a truncated density to K using leapfrog dynamics with boundary reflections. There are several works that have previously examined constrained versions of the HMC algorithm. More specifically, Afshar and Domke (2015) examines the HMC variant with reflection and

Algorithm 2: ReHMC $(\pi,\ \boldsymbol{x},\ W,\ \eta,\ M)$

Input : a point $\boldsymbol{x} \in \mathbb{R}^p$; the number of steps W; the size of leapfrog step; the number of leapfrog steps

Require: $\boldsymbol{x} \in K,\ W, M \in \mathbb{N},\ \eta \in \mathbb{R}_+$

for $i = 1, \ldots, W$ **do**
  $\boldsymbol{v} \leftarrow_R \mathcal{N}(0, I_d)$;
  for $j = 1, \ldots, M$ **do**
    $\boldsymbol{v} \leftarrow \boldsymbol{v} + \frac{\eta}{2}\nabla \log(\pi(\boldsymbol{x}))$;
    $L \leftarrow \eta$;
    do
      $\tau \leftarrow$ INTERSECTION$(\boldsymbol{x},\ \boldsymbol{v})$; // *get the intersection time*
      $\tilde{\boldsymbol{x}},\ \tilde{\boldsymbol{v}} \leftarrow$ REFLECTION$(\boldsymbol{x},\ \tau,\ \boldsymbol{v})$;
      if$(L < \tau)$ **then** $\boldsymbol{x} \leftarrow \boldsymbol{x} + L\boldsymbol{v}$ **else** $\boldsymbol{x} \leftarrow \tilde{\boldsymbol{x}}$; $\boldsymbol{v} \leftarrow \tilde{\boldsymbol{v}}$;
      $L \leftarrow L - \tau$;
    while $L > 0$;
    $\boldsymbol{v} \leftarrow \boldsymbol{v} + \frac{\eta}{2}\nabla \log(\pi(\boldsymbol{x}))$;
return $\boldsymbol{x}$

refraction using a leapfrog integrator, proving a guarantee for the convergence to the target distribution. This is a specialized approach for sampling from densities, where the Hamiltonian dynamics are relatively simple and easily derived.

The Hamiltonian dynamics behind HMC operate on a p-dimensional position vector $\boldsymbol{x}$ and a p-dimensional momentum $\boldsymbol{v}$. So the full state space has $2p$ dimensions. The system is described by a function of $\boldsymbol{x}$ and $\boldsymbol{v}$ known as the Hamiltonian,

$$H(\boldsymbol{x}, \boldsymbol{v}) = U(\boldsymbol{x}) + K(\boldsymbol{v}) = -\log(\pi(\boldsymbol{x})) + \frac{1}{2}|\boldsymbol{v}|^2.$$

To sample from π, one has to solve the following system of Ordinary Differential Equations (ODE):

$$\begin{aligned} \frac{d\boldsymbol{x}}{dt} &= \frac{\partial H(\boldsymbol{x}, \boldsymbol{v})}{\partial \boldsymbol{v}} \\ \frac{d\boldsymbol{v}}{dt} &= -\frac{\partial H(\boldsymbol{x}, \boldsymbol{v})}{\partial \boldsymbol{x}} \end{aligned} \quad \Rightarrow \quad \begin{cases} \frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{v}(t) \\ \\ \frac{d\boldsymbol{v}(t)}{dt} = \nabla \log(\pi(\boldsymbol{x})) \end{cases} . \tag{22}$$

However, in general, for the solution of the ODE in Equation (22) there is not a closed form. Thus, ReHMC approximates the solution of the ODE by

discretizing it. A well known and efficient in practice way to discretize the Hamiltonian Dynamics is through the *leapfrog integrator*

$$\hat{\boldsymbol{v}}_{i+1} = \boldsymbol{v}_i + \frac{\eta}{2}\nabla \log(\pi(\boldsymbol{x}_i)) \quad \boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \eta\hat{\boldsymbol{v}}_{i+1}, \quad \boldsymbol{v}_{i+1} = \hat{\boldsymbol{v}}_{i+1} + \frac{\eta}{2}\nabla \log(\pi(\boldsymbol{x}_{i+1})) \tag{23}$$

Then, at each step ReHMC performs M leapfrog steps of length η each. In particular, it chooses a momenta $\boldsymbol{v}$ randomly from $\mathcal{N}(0, I_p)$, and updates the position $\boldsymbol{x}$ using the leapfrog trajectory. Then, to guarantee convergence to π ReHMC applies the Metropolis filter using the current position of the random walk and the proposed one. While π is truncated in K, ReHMC exploits boundary reflections. This operation guarantees that the dynamics remain volume preserving and time-reversible (Afshar and Domke, 2015). ReHMC uses INTERSECTION and REFLECTION as defined in Appendix 1.

When there is not a closed formula for the gradient $\nabla \log(\pi(\boldsymbol{x}))$ we approximate it using the multivariate Taylor expansion. To further improve the performance of ReHMC we introduce the NUTS sampler in Homan and Gelman (2014) in our framework. In particular, we use the dual averaging scheme of Nesterov (2009) to initialize the step η as also proposed in (Homan and Gelman, 2014). Then, we use the NUTS stopping criterion as the boundary reflections preserve volume and the detailed balance holds.

6 Simulations

We compare all the prior choices for several cases. Our experiments show that the efficiency of ReHMC depends on the number of Gaussian observations k (or sample size) and the ratio between the largest over the minimum eigenvalue $\lambda_{\max}/\lambda_{\min}$ of the covariance matrix Σ. The sample size k determines the estimation error, i.e. the larger the sample size the smaller the estimation error. The ratio $\lambda_{\max}/\lambda_{\min}$ determines the mixing rate of the MCMC algorithm. We perform the following experiment:

1. Generate a covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ from the Wishart distribution with the covariance matrix being the identity matrix and degrees of freedom equal to 80.

2. Modify the element Σ_{11} such that the ratio of the maximum over the minimum eigenvalue of Σ to be $\frac{\lambda_{\max}}{\lambda_{\min}} = 5$.

3. We set the sample size $k = 20000$ and generate $y_1, \ldots, y_k \in \mathbb{R}^n$, i.i.d. vectors from $\mathcal{N}(0, \Sigma)$.

4. Sample from the posterior distribution for uniform prior with ReHMC.

5. End sampling when the Potential Scale Reduction Factor (PSFR) (Gelman and Rubin (1992)) of each marginal (partial correlation) is smaller than 1.1, and then compute the sample average to estimate the covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$.

The results in Table 1 show that ReHMC achieves a PSRF < 1.1 for all marginals (partial correlations) in less than 1hr run-time for covariance size $n \leq 100$. The plots in Figure 2 illustrates the efficiency of ReHMC for various number of observations. They both show that the estimation error decreases as the number of observations increase for both $n = 30,\ 60$; 4000 sampled matrices from the posterior suffices to reach best possible estimation. Notice that for $n = 60$ ReHMC achieves better estimation errors for each k than the case of $n = 30$, which illustrates the better of mixing of ReHMC for lower dimensions –as expected. The plots in Figure 3 illustrate the efficiency of ReHMC for various values of the ratio of the maximum, $\lambda_{\max}$, over the minimum, $\lambda_{\min}$, eigenvalue of Σ. As expected the convergence to Σ slows down while the ratio $\lambda_{\max}/\lambda_{\min}$ increases. This implies that the mixing rate is worse for large values of the eigenvalue ratio. Again for $n = 60$ ReHMC achieves better estimation errors for each ratio than the case of $n = 30$.

Uniformly distributed correlation matrices. We use BILLIARD_WALK to generate uniformly distributed correlation matrices and we use both PSRF and Effective Sample Size to evaluate its efficiency. The results in Table 4 imply that BILLIARD_WALK for matrix size $n \leq 80$ achieves a PSRF < 1.1 in a few minutes. For $n = 100$ achieves the same PSRF value in 30 minutes.

7 Empirical analysis

We apply our methodology to cross border bank total claims at the two core countries of the Euro zone, France and Germany and at GIIPS, namely Greece, Ireland, Italy, Portugal and Spain. The correlation matrices at (24), (25), and (26) are the results obtained using different prior distributions. The prior distributions are the (jointly) uniform, the inverse-Wishart (marginally uniform prior on the partial correlations), and the Jeffrey prior. We find that under different prior specifications the biggest correlations are found between Germany and Ireland, enhancing the implications of the banking crisis in the country.

$$\hat{C}_1 = \begin{array}{c} \\ \textbf{France} \\ \textbf{Italy} \\ \textbf{Germany} \\ \textbf{Greece} \\ \textbf{Ireland} \\ \textbf{Spain} \\ \textbf{Portugal} \end{array} \begin{array}{c} \begin{array}{ccccccc} \textbf{France} & \textbf{Italy} & \textbf{Germany} & \textbf{Greece} & \textbf{Ireland} & \textbf{Spain} & \textbf{Portugal} \end{array} \\ \begin{pmatrix} 1 & 0.3481 & 0.5226 & 0.1416 & -0.1804 & 0.0975 & 0.1124 \\ 0.3481 & 1 & 0.3736 & -0.0086 & -0.3928 & 0.5222 & -0.3239 \\ 0.5226 & 0.3736 & 1 & -0.1345 & 0.8331 & -0.5701 & 0.5859 \\ 0.1416 & -0.0086 & -0.1345 & 1 & 0.0839 & 0.0836 & 0.0288 \\ -0.1804 & -0.3928 & 0.8331 & 0.0839 & 1 & 0.4053 & -0.5792 \\ 0.0975 & 0.5222 & -0.5701 & 0.0836 & 0.4053 & 1 & 0.7623 \\ 0.1124 & -0.3239 & 0.5859 & 0.0288 & -0.5792 & 0.7623 & 1 \end{pmatrix} \end{array} \tag{24}$$

$$\hat{C}_2 = \begin{array}{c} \\ \textbf{France} \\ \textbf{Italy} \\ \textbf{Germany} \\ \textbf{Greece} \\ \textbf{Ireland} \\ \textbf{Spain} \\ \textbf{Portugal} \end{array} \begin{array}{c} \begin{array}{ccccccc} \textbf{France} & \textbf{Italy} & \textbf{Germany} & \textbf{Greece} & \textbf{Ireland} & \textbf{Spain} & \textbf{Portugal} \end{array} \\ \begin{pmatrix} 1 & 0.3531 & 0.4626 & 0.0979 & 0.0728 & 0.0835 & 0.2717 \\ 0.3531 & 1 & 0.2170 & 0.0244 & -0.0974 & 0.2750 & 0.1016 \\ 0.4626 & 0.2170 & 1 & -0.1024 & 0.4913 & -0.1284 & 0.2364 \\ 0.0979 & 0.0244 & -0.1024 & 1 & -0.0108 & 0.2488 & 0.0614 \\ 0.0728 & -0.0974 & 0.4913 & -0.0108 & 1 & -0.0894 & -0.1657 \\ 0.0835 & 0.2750 & -0.1284 & 0.2488 & -0.0894 & 1 & 0.4183 \\ 0.2717 & 0.1016 & 0.2364 & 0.0614 & -0.1657 & 0.4183 & 1 \end{pmatrix} \end{array} \tag{25}$$

$$\hat{C}_3 = \begin{array}{c} \\ \textbf{France} \\ \textbf{Italy} \\ \textbf{Germany} \\ \textbf{Greece} \\ \textbf{Ireland} \\ \textbf{Spain} \\ \textbf{Portugal} \end{array} \begin{array}{c} \begin{array}{ccccccc} \textbf{France} & \textbf{Italy} & \textbf{Germany} & \textbf{Greece} & \textbf{Ireland} & \textbf{Spain} & \textbf{Portugal} \end{array} \\ \begin{pmatrix} 1 & 0.3786 & 0.5185 & 0.0893 & -0.0149 & 0.0734 & 0.2382 \\ 0.3786 & 1 & 0.2193 & -0.0170 & -0.1516 & 0.3516 & -0.0090 \\ 0.5185 & 0.2193 & 1 & -0.1101 & 0.6064 & -0.2466 & 0.3106 \\ 0.0893 & -0.0170 & -0.1101 & 1 & 0.0397 & 0.2218 & 0.0390 \\ -0.0149 & -0.1516 & 0.6064 & 0.0397 & 1 & -0.0231 & -0.2467 \\ 0.0734 & 0.3516 & -0.2466 & 0.2218 & -0.0231 & 1 & 0.5290 \\ 0.2382 & -0.0090 & 0.3106 & 0.0390 & -0.2467 & 0.5290 & 1 \end{pmatrix} \end{array} \tag{26}$$

8 Conclusions and future work

European integration and increased market contestability has been a major factor in increasing cross border banking. In particular, the membership of the European Union's Single Market, rather than membership of the Currency Union, has raised cross border lending significant. Under different prior specifications we obtain the same results in term of the magnitude of the

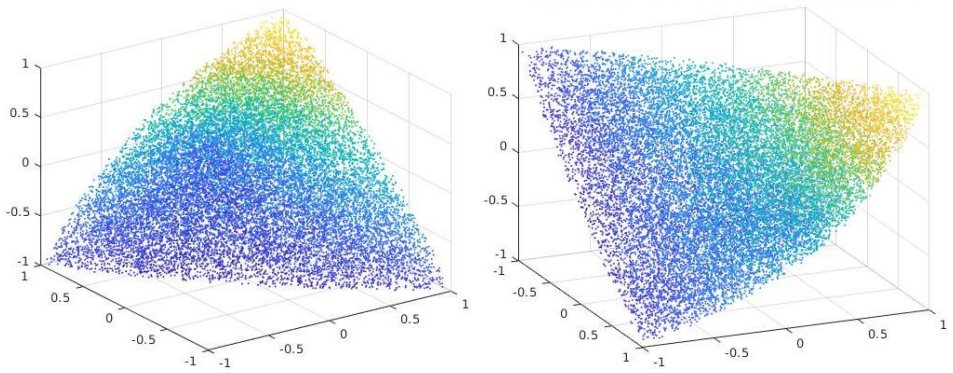
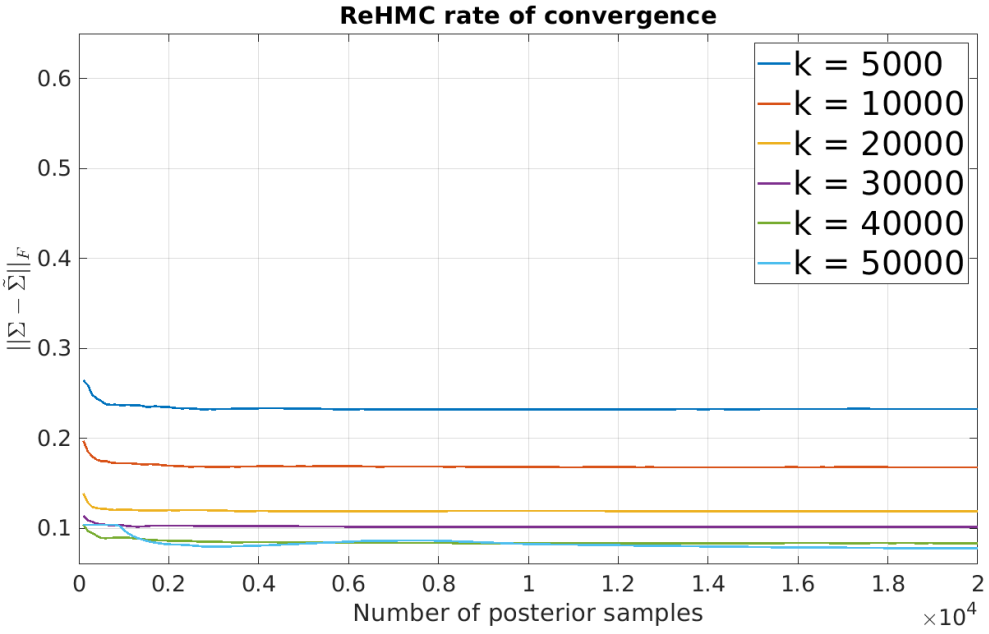
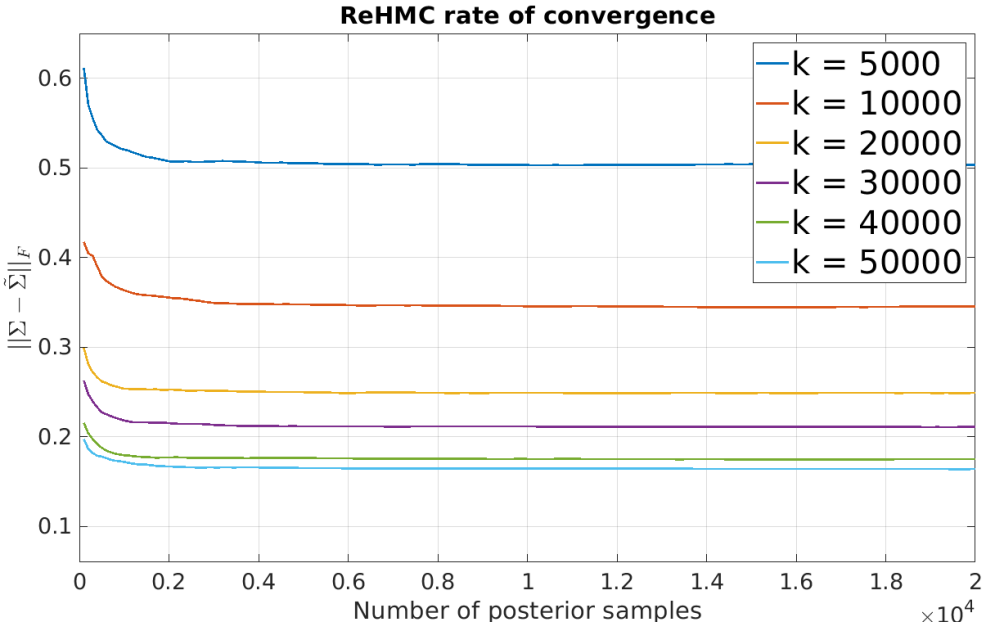


Figure 1: A uniform sample from the set of 3×3 correlation matrices, $K = S \cap H$, as defined in Section 4.0.1. Each point corresponds to a correlation matrix. We sample 20000 uniformly distributed points and plot them from two different angles.

correlations between Germany and Ireland, and Germany and France. The conclusion on the importance of the EU is strengthened by our investigation of the impacts of the 2011 Euro Area sovereign debt crisis on patterns of lending, when a significant fall in lending took place.

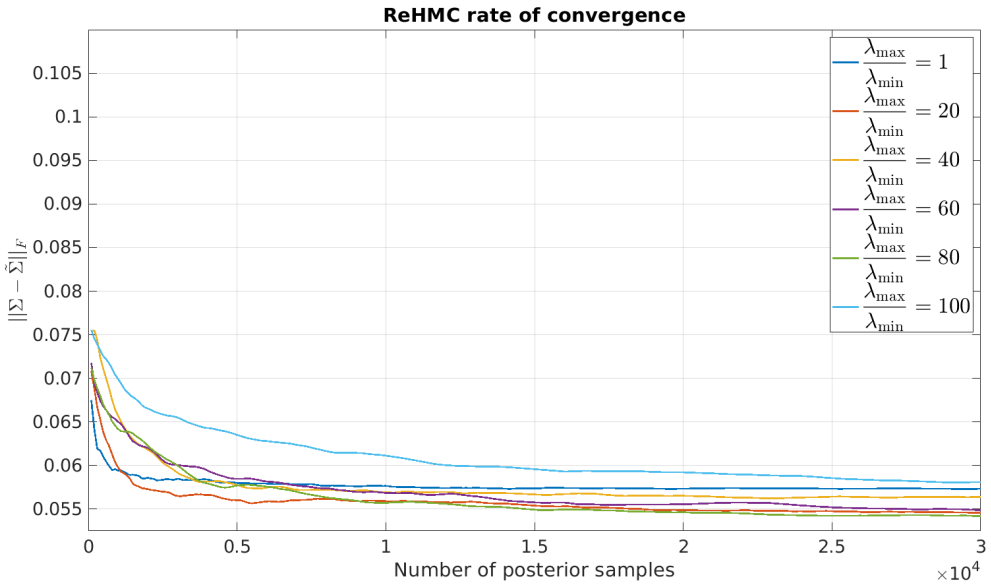


(a) $n = 30$

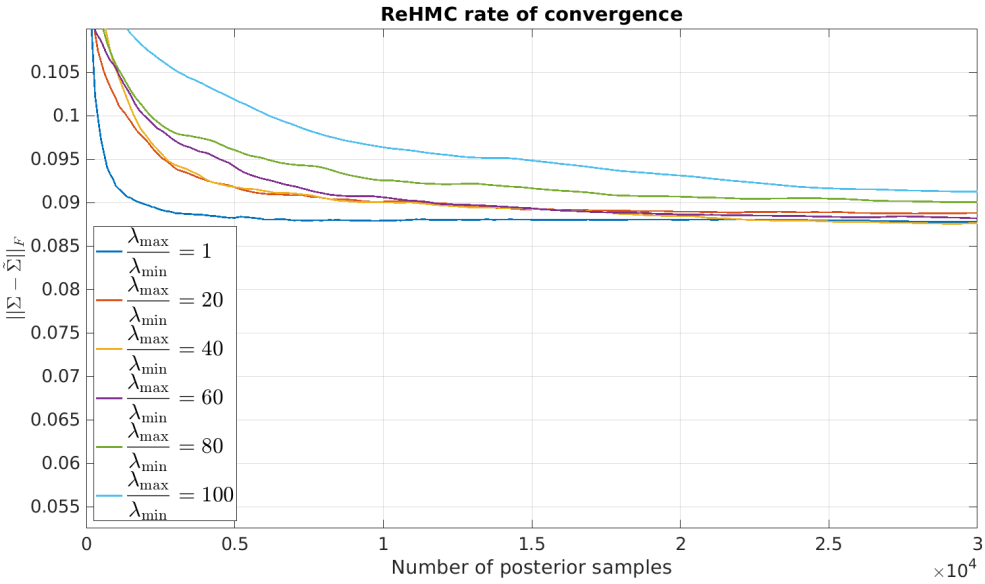


(b) $n = 60$

Figure 2: The figures show the rate of convergence when sampling from the posterior distribution for **uniform prior** with ReHMC. The ratio of the maximum, $\lambda_{\max}$, over the minimum, $\lambda_{\min}$, eigenvalue of Σ is fixed to 5. We increase the sample size k from $\mathcal{N}(0, \Sigma)$. To measure the estimation we use the Frobenius norm of the difference $||\Sigma - \tilde{\Sigma}||_F$, where $\tilde{\Sigma}$ is the estimation of the covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ obtained from our methodology.



(a) $n = 30$



(b) $n = 60$

Figure 3: The figures show the rate of convergence when sampling from the posterior distribution for **uniform prior** with ReHMC. Given a fixed sample size, $k = 20000$, drawn from $\mathcal{N}(0, \Sigma)$, the ratio of the maximum, $\lambda_{\max}$, over the minimum, $\lambda_{\min}$, eigenvalue of Σ is increased from 1 to 100. To measure the estimation error, we use the Frobenius norm of the difference $||\Sigma - \tilde{\Sigma}||_F$, where $\tilde{\Sigma}$ is the estimation of the covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ obtained from our methodology.

Table 1: We generate $k = 20000$ vectors from the $\mathcal{N}(0, \Sigma)$, $\Sigma \in \mathbb{R}^{n\times n}$ and sample N correlation matrices from the posterior distribution for the **uniform prior** to compute an estimation $\tilde{\Sigma} \in \mathbb{R}^{n\times n}$ for the covariance Σ by taking the sample average. In all cases the ratio of the maximum over the minimum eigenvalue of matrix Σ is equal to 5. (n) the dimension of the covariance matrix; (p) the total number of parameters to estimate; (N) the total number of covariance matrices that our implementation samples; (PSRF) the maximum value of PSRF among all marginals; (Ef) the average ESS of the univariate marginals over the run-time; (Time) the runtime in seconds; $\tilde{\Sigma} \in \mathbb{R}^{n\times n}$ our estimation. To measure the estimation error, we compute several norms of the matrix $\Sigma - \tilde{\Sigma}$; $||\cdot||_1$, is the maximum absolute column sum of the matrix; $||\cdot||_2$ is the maximum singular value; $||\cdot||_\infty$ is the maximum absolute row sum of the matrix; $||\cdot||_F$ is the Frobenius norm; $||\cdot||_{\mathcal{P}}$ is the Bures–Wasserstein distance between positive definite matrices.

(n)	(p)	(N)	(PSRF)	(Ef)	(Time)	$\|\|\cdot\|\|_1$	$\|\|\cdot\|\|_2$	$\|\|\cdot\|\|_\infty$	$\|\|\cdot\|\|_F$	$\|\|\cdot\|\|_{\mathbb{P}}$
10	55	2000	1.0724	32.6348	3.3	0.035705	0.021928	0.035705	0.034324	2.6718
20	210	4000	1.0577	12.7512	10.4	0.075923	0.029873	0.075923	0.071673	4.0058
30	465	6000	1.0659	4.0932	25.6	0.12342	0.045028	0.12342	0.12067	5.1406
40	820	8000	1.0959	1.9651	50.6	0.17039	0.059043	0.17039	0.16957	6.104
50	1275	10000	1.0831	0.87278	118.6	0.20963	0.060095	0.20963	0.20814	7.0111
60	1830	14000	1.0807	0.38039	302.1	0.24964	0.072358	0.24964	0.25737	7.8063
70	2485	18000	1.0895	0.15345	645.3	0.29741	0.079939	0.29741	0.30457	8.6117
80	3240	18000	1.0968	0.085758	1035	0.34885	0.088545	0.34885	0.35096	9.3502
90	4095	18000	1.0945	0.059121	1477	0.39587	0.094438	0.39587	0.40633	10.058
100	5050	28000	1.0812	0.027792	3739	0.45016	0.11067	0.45016	0.46043	10.6889

Table 2: Similar to the Table 1 we generate $k = 20000$ vectors from the $\mathcal{N}(0, \Sigma)$, $\Sigma \in \mathbb{R}^{n\times n}$ and sample N correlation matrices from the posterior distribution for the **inverse-Wishart prior** on the covariance matrix Σ (marginally uniform prior on the partial correlations). We compute an estimation $\tilde{\Sigma} \in \mathbb{R}^{n\times n}$ for the covariance Σ by taking the sample average. In all cases the ratio of the maximum over the minimum eigenvalue of matrix Σ is equal to 5. (n) the dimension of the covariance matrix; (p) the total number of parameters to estimate; (N) the total number of covariance matrices that our implementation samples; (PSRF) the maximum value of PSRF among all marginals; (Ef) the average ESS of the univariate marginals over the run-time; (Time) the runtime in seconds; $\tilde{\Sigma} \in \mathbb{R}^{n\times n}$ our estimation. To measure the estimation error, we compute several norms of the matrix $\Sigma - \tilde{\Sigma}$; $||\cdot||_1$, is the maximum absolute column sum of the matrix; $||\cdot||_2$ is the maximum singular value; $||\cdot||_\infty$ is the maximum absolute row sum of the matrix; $||\cdot||_F$ is the Frobenius norm; $||\cdot||_{\mathcal{P}}$ is the Bures–Wasserstein distance between positive definite matrices.

(n)	(p)	(N)	(PSRF)	(Ef)	(Time)	$\|\|\cdot\|\|_1$	$\|\|\cdot\|\|_2$	$\|\|\cdot\|\|_\infty$	$\|\|\cdot\|\|_F$	$\|\|\cdot\|\|_{\mathbb{P}}$
10	55	2000	1.0293	69.0955	1.7	0.04022	0.019063	0.04022	0.035261	2.6392
20	210	6000	1.0507	11.653	12.1	0.078027	0.030522	0.078027	0.072878	4.1151
30	465	6000	1.0682	3.6884	33.7	0.1277	0.040661	0.1277	0.11813	5.1847
40	820	6000	1.077	1.5175	61.1	0.18244	0.062127	0.18244	0.17631	6.1069
50	1275	10000	1.0788	0.56586	175.2	0.2752	0.089099	0.2752	0.26491	6.9469
60	1830	12000	1.0829	0.24811	362.2	0.35968	0.11203	0.35968	0.37015	7.7821
70	2485	18000	1.0836	0.11837	861.9	0.50111	0.14321	0.50111	0.50183	8.5745
80	3240	14000	1.0835	0.10233	1009	0.61395	0.16934	0.61395	0.68428	9.207
90	4095	18000	1.0822	0.064164	1855	0.78279	0.19559	0.78279	0.855	9.8984
100	5050	28000	1.0955	0.022348	4445	0.96951	0.21895	0.96951	1.0909	10.5272

Table 3: Similar to the Tables 1 2 we generate $k = 20000$ vectors from the $\mathcal{N}(0, \Sigma)$, $\Sigma \in \mathbb{R}^{n\times n}$ and sample N correlation matrices from the posterior distribution for the **Jeffreys prior** on the covariance matrix Σ. We compute an estimation $\tilde{\Sigma} \in \mathbb{R}^{n\times n}$ for the covariance Σ by taking the sample average. In all cases the ratio of the maximum over the minimum eigenvalue of matrix Σ is equal to 5. (n) the dimension of the covariance matrix; (p) the total number of parameters to estimate; (N) the total number of covariance matrices that our implementation samples; (PSRF) the maximum value of PSRF among all marginals; (Ef) the average ESS of the univariate marginals over the run-time; (Time) the runtime in seconds; $\tilde{\Sigma} \in \mathbb{R}^{n\times n}$ our estimation. To measure the estimation error, we compute several norms of the matrix $\Sigma - \tilde{\Sigma}$; $||\cdot||_1$, is the maximum absolute column sum of the matrix; $||\cdot||_2$ is the maximum singular value; $||\cdot||_\infty$ is the maximum absolute row sum of the matrix; $||\cdot||_F$ is the Frobenius norm; $||\cdot||_{\mathcal{P}}$ is the Bures–Wasserstein distance between positive definite matrices.

(n)	(p)	(N)	(PSRF)	(Ef)	(Time)	$\|\|\cdot\|\|_1$	$\|\|\cdot\|\|_2$	$\|\|\cdot\|\|_\infty$	$\|\|\cdot\|\|_F$	$\|\|\cdot\|\|_{\mathbb{P}}$
10	55	2000	1.0372	86.7837	1.4	0.10369	0.052772	0.10369	0.097158	2.6548
20	210	4000	1.034	24.022	6.1	0.21444	0.10439	0.21444	0.20084	4.0746
30	465	4000	1.0803	5.7592	14.2	0.31413	0.12456	0.31413	0.29484	5.1996
40	820	6000	1.0772	1.973	47.6	0.43008	0.16028	0.43008	0.43188	6.0951
50	1275	6000	1.07	0.89524	102.6	0.56755	0.18019	0.56755	0.5566	6.9751
60	1830	6000	1.0941	0.46302	197.9	0.61402	0.20016	0.61402	0.66329	7.7361
70	2485	10000	1.0759	0.19227	596.2	0.75874	0.20802	0.75874	0.79466	8.4541
80	3240	10000	1.0869	0.096447	999.2	0.80188	0.21287	0.80188	0.86547	9.2639
90	4095	12000	1.0917	0.054895	1870	0.8941	0.21456	0.8941	0.973	9.904
100	5050	30000	1.0874	0.025721	3589	0.87346	0.2294	0.9075	1.1034	10.8923

(n)	(p)	(N)	(PSRF)	(Ef)	(Time)
20	210	600	1.08	38	4.2
40	820	1200	1.09	21	25
60	1830	3000	1.07	57	228
80	3240	3600	1.09	51	656
100	5050	6200	1.09	42	1838

Table 4: We generate uniformly correlation matrices with BILLIARD_WALK until we achieve a PSRF < 1.1. (n) the dimension of the covariance matrix; (p) the total number of parameters to estimate; (N) the total number of correlation matrices that our implementation samples; (PSRF) the maximum value of PSRF among all marginals; (Ef) the average ESS of the univariate marginals over the run-time; (Time) the runtime in seconds;.

References

Abedifar, Pejman, Paolo Giudici, and Shatha Qamhieh Hashem (2017) 'Heterogeneous market structure and systemic risk: Evidence from dual banking systems.' *Journal of Financial Stability* 33, 96 – 119

Acharya, Viral, Robert Engle, and Matthew Richardson (2012) 'Capital shortfall: A new approach to ranking and regulating systemic risks.' *American Economic Review* 102(3), 59–64

Adrian, Tobias, and Markus K. Brunnermeier (2016) 'CoVaR.' *American Economic Review* 106(7), 1705–1741

Afshar, Hadi Mohasel, and Justin Domke (2015) 'Reflection, refraction, and Hamiltonian Monte Carlo.' In 'Advances in neural information processing systems' pp. 3007–3015

Ahelegbey, Daniel Felix, Monica Billio, and Roberto Casarin (2016) 'Bayesian graphical models for structural vector autoregressive processes.' *Journal of Applied Econometrics* 31(2), 357–386

Allen, Franklin, Ana Babus, and Elena Carletti (2012) 'Asset commonality, debt maturity and systemic risk.' *Journal of Financial Economics* 104(3), 519–534

Alvarez, Ignacio, Jarad Niemi, and Matt Simpson (2016) 'Bayesian inference for a covariance matrix'

Avdjiev, Stefan, Paolo Giudici, and Alessandro Spelta (2019) 'Measuring contagion risk in international banking.' *Journal of Financial Stability* 42, 36–51

Barnard, John, Robert McCulloch, and Xiao Li Meng (2000) 'Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage.' *Statistica Sinica* 10(4), 1281–1311

Battiston, Stefano, Domenico Delli Gatti, Mauro Gallegati, Bruce Greenwald, and Joseph E.Stiglitz (2012)) 'Liaisons dangereuses:increasing connectivity, risk sharing, and systemic risk.' *Journal of Economic Dynamics & Control* 36, 1121–1141

Bendel, Robert B., and M. Ray Mickey (1978) 'Population correlation matrices for sampling experiments.' *Communications in Statistics - Simulation and Computation* 7(2), 163–182

Benoit, Sylvain, Jean-Edouard Colliard, Christophe Hurlin, and Christophe Pérignon (2017) 'Where the risks lie: A survey on systemic risk.' *Review of Finance* 21(1), 109–152

Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon (2012) 'Econometric measures of connectedness and systemic risk in the finance and insurance sectors.' *Journal of Financial Economics* 104(3), 535–559

Brown, P., and J. Zidek (1994) 'Inference for a covariance matrix'

Brownlees, Christian T., and Robert F. Engle (2017) 'SRISK: A Conditional Capital Shortfall Measure of Systemic Risk.' *The Review of Financial Studies* 30(1), 48–79

Budden Mark, Hadavas Paul, Hoffman Lorrie (2008) 'On the generation of correlation matrices.' *Applied Mathematics E-Notes [electronic only]* 8, 279–282

Calabrese, Raffaella, and Paolo Giudici (2015) 'Estimating bank default with generalised extreme value regression models.' *Journal of the Operational Research Society* 66, 1783—1792

Chalkis, Apostolos, Ioannis Emiris, Vissarion Fisikopoulos, Panagiotis Repouskos, and Elias Tsigaridas (2020) 'Efficient sampling from feasible sets of sdps and volume approximation'

Chalmers, C.P. (1975) 'Generation of correlation matrices with a given eigen–structure.' *Journal of Statistical Computation and Simulation* 4(2), 133–139

Davies, Philip, and Nicholas Higham (2000) 'Numerically stable generation of correlation matrices and their factors.' *BIT* 40, 640–651

Dhillon, Inderjit S., Robert W. Heath, Mátyás A. Sustik, and Joel A. Tropp (2005) 'Generalized finite algorithms for constructing hermitian matrices with prescribed diagonal and spectrum.' *SIAM J. Matrix Anal. Appl.* 27(1), 61–71

Diebold, F., and K Yilmaz (2014) 'On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms.' *Journal of Econometrics* 182(1), 119–134

Emmert-Streib, Frank, Shailesh Tripathi, and Matthias Dehmer (2019) 'Constrained covariance matrices with a biologically realistic structure: Comparison of methods for generating high-dimensional gaussian graphical models.' *Frontiers in Applied Mathematics and Statistics* 5, 17

Geisser, Seymour (1965) 'Bayesian estimation in multivariate analysis.' *The Annals of Mathematical Statistics* 36(1), 150–159

Geisser, Seymour, and Jerome Cornfield (1963) 'Posterior distributions for multivariate normal parameters.' *Journal of the Royal Statistical Society. Series B (Methodological)* 25(2), 368–376

Gelman, Andrew, and Donald B. Rubin (1992) 'Inference from iterative simulation using multiple sequences.' *Statistical Science* 7(4), 457–472

Gennaioli, Nicola, Andrei Shliefer, and Robert W. Vishny (2013) 'A model of shadow banking.' *Journal of Finance* 68(4), 1331–1363

Giudici, P., and A. Spelta (2016) 'Graphical network models for international financial flows.' *Journal of Business & Economic Statistics* 34(1), 128–138

Giudici, Paolo, and Peter J. Green (1999) 'Decomposable graphical gaussian model determination.' *Biometrika* 86, 785—801

Gryazina, Elena, and Boris Polyak (2012) 'Random sampling: Billiard walk algorithm.' *European Journal Operational Research* 238, 497–504

Hardin, Johanna, Stephan Ramon Garcia, and David Golan (2013) 'A method for generating realistic correlation matrices.' *The Annals of Applied Statistics* 7(3), 1733–1762

Holmes, R. B. (1991) 'On random correlation matrices.' *SIAM Journal on Matrix Analysis and Applications* 12(2), 239–272

Homan, Matthew D., and Andrew Gelman (2014) 'The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo.' *J. Mach. Learn. Res.* 15(1), 1593–1623

Joe, Harry (2006) 'Generating random correlation matrices based on partial correlations.' *Journal of Multivariate Analysis* 97(10), 2177 – 2189

Kashyap, Anil K., Raghuram Rajan, and Jeremy C. Stein (2002) 'Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking.' *The Journal of Finance* 57(1), 33–74

Kollo, Tõnu, and Kaire Ruul (2003) 'Approximations to the distribution of the sample correlation matrix.' *Journal of Multivariate Analysis* 85(2), 318 – 334

Laeven, Luc, Lev Ratnovski, and Hui Tong (2016) 'Bank size, capital, and systemic risk: Some international evidence.' *Journal of Banking & Finance* 69, S25–S34

Liu, Xuefeng, and Michael J. Daniels (2006) 'A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization.' *Journal of Computational and Graphical Statistics* 15(4), 897–914

Marsaglia, George, and Ingram Olkin (1984) 'Generating correlation matrices.' *SIAM Journal on Scientific and Statistical Computing* 5(2), 470–475

Nesterov, Yurii (2009) 'Primal-dual subgradient methods for convex problems.' *Mathematical Programming* 120, 221–259

Numpacharoen, Kawee, Amporn Atsawarungruangkit (2012) 'Generating correlation matrices based on the boundaries of their coefficients.' *PloS one* 7, 11

Pitt, Michael, David Chan, and Robert Kohn (2006) 'Efficient bayesian inference for gaussian copula regression models.' *Biometrika* 93(3), 537–554

Ramana, M., and A. Goldman (1999) 'Some geometric results in semidefinite programming.' *J. Global Optimization*

Robert, Christian P., Nicolas Chopin, and Judith Rousseau (2009) 'Harold Jeffreys's Theory of Probability Revisited.' *Statistical Science* 24(2), 141 – 172

Rousseeuw, Peter J., and Geert Molenberghs (1994) 'The shape of correlation matrices.' *The American Statistician* 48(4), 276–279

Shleifer, Andrei, and Robert W. Vishny (2010) 'Unstable banking.' *Journal of Financial Economics* 97(3), 306–318

Waller, Niels G. (2020) 'Generating correlation matrices with specified eigenvalues using the method of alternating projections.' *The American Statistician* 74(1), 21–28

Wong, Frederick, Christopher K. Carter, and Robert Kohn (2003) 'Efficient estimation of covariance selection models.' *Biometrika* 90(4), 809–830

Zhang, Xiao, W. John Boscardin, and Thomas R Belin (2006) 'Sampling correlation matrices in bayesian models with correlated latent variables.' *Journal of Computational and Graphical Statistics* 15(4), 880–896

Appendix

1 Geometrical operations on the set of correlation matrices

The majority of geometric random walks are defined for general convex bodies and the implementation of their step is based on an oracle; usually the membership oracle. To specialize a random walk for a family or representation of convex bodies one has to come up with efficient algorithms for the basic geometric operations to realize the (various) oracles.

Thus, to use the geometric random walks we introduce we need to define certain fundamental geometrical operations that are necessary for their implementation. In particular, we develop three operations: MEMBERSHIP, INTERSECTION and REFLECTION. Then, each algorithm will employ some or all of these operations in order to sample from a density $\pi(\boldsymbol{x})$ supported in the set of correlation matrices K of Equation (18).

Algorithm 3: MEMBERSHIP($\boldsymbol{x}$)

Input : A point $\boldsymbol{x} \in \mathbb{R}^p$
Output : answer whether $\boldsymbol{x} \in K$ or not

1 $is\_in \leftarrow$ true;
2 **if** the Equation (27) does not hold **then** $is\_in \leftarrow$ false;
3 **if** the Equation (28) does not hold **then** $is\_in \leftarrow$ false;
4 **return** is_in;

1.1 Membership

The operation MEMBERSHIP decides whether a point $\boldsymbol{x} \in \mathbb{R}^p$ lies in the interior of K. Since $K = S \cap H$ it checks whether $\boldsymbol{x} \in H$ and $\boldsymbol{x} \in S$. The first holds if and only if

$$-1 \leq x_i \leq 1,\ i \in [p]. \tag{27}$$

To check whether $\boldsymbol{x} \in S$ we check if the matrix

$$F(\boldsymbol{x}) \succ 0, \tag{28}$$

that is whether matrix $\boldsymbol{F}(\boldsymbol{x})$ is a positive definite matrix. If both the Equations (27) and (28) holds then MEMBERSHIP answers positively, otherwise it answers negatively.

Algorithm 4: INTERSECTION$(\boldsymbol{x}, \boldsymbol{v})$

Input : A point $\boldsymbol{x} \in \mathbb{R}^p$ and a direction vector $\boldsymbol{v} \in \mathbb{R}^p$
Require : $\boldsymbol{x} \in K$
Output : the intersection time τ of $\ell(t) \cap \partial K$, where $\ell(t) = \{t \in \mathbb{R}_+ \mid \boldsymbol{x} + t\boldsymbol{v}\}$

compute the intersection time t_1 of $\ell(t) \cap H$ in Equation (30);
compute the intersection time t_2 of $\ell(t) \cap S$ in Equation (32);
$\tau \leftarrow \min\{t_1, t_2\}$;
return τ;

1.2 Intersection

The operation INTERSECTION computes the intersection of a line $\ell(t) = \{t \in \mathbb{R}_+ \mid \boldsymbol{x} + t\boldsymbol{v}\}$ with the boundary of K (∂K), where $\boldsymbol{x},\ \boldsymbol{v} \in \mathbb{R}^p$. Due to convexity and since $K = S \cap H$ to compute the intersection time of $\ell(t) \cap \partial K$ one has to compute the intersection times for both $\ell(t) \cap \partial H$ and $\ell(t) \cap \partial S$ and keep the smallest one; that is,

$$\tau = \arg\min_{t \in \mathbb{R}_+} \{\ell(t) \cap \partial H,\ \ell(t) \cap \partial S\}. \tag{29}$$

Then, the intersection point is $\ell(\tau)$.

To compute the intersection time of $\ell(t) \cap \partial H$ we compute the intersection time for each facet of the hypercube,

$$t_i = \frac{1 \mp x_i}{\pm v_i},\ i \in [p], \tag{30}$$

and we keep the smallest positive value among the $2p$ values.

To compute the intersection time of $\ell(t) \cap \partial S$ one has to compute,

$$\tau = \arg\sup_{t \in \mathbb{R}_+} \boldsymbol{F}(\boldsymbol{x} + t\boldsymbol{v}) \succ 0. \tag{31}$$

Since $\boldsymbol{F}(\boldsymbol{x} + t\boldsymbol{v}) = \boldsymbol{F}(\boldsymbol{x}) + t(\boldsymbol{F}(\boldsymbol{v}) - I_n)$ the intersection time in Equation (31) corresponds to the smallest positive eigenvalue of the following generalized eigenvalue problem,

$$\boldsymbol{F}(\boldsymbol{x}) + t(\boldsymbol{F}(\boldsymbol{v}) - I_n) = 0. \tag{32}$$

1.3 Reflection

The operation REFLECTION computes the reflection of a line $\ell(t) = \{t \in \mathbb{R}_+ \mid \boldsymbol{x} + t\boldsymbol{v}\}$ when it hits the ∂K at the boundary point $\tilde{\boldsymbol{x}}$. Let, the

Algorithm 5: REFLECTION($\boldsymbol{x},\ \tau,\ \boldsymbol{v}$)

Input : A point $\boldsymbol{x} \in \mathbb{R}^p$, a scalar $\tau > 0$ and a direction vector $\boldsymbol{v} \in \mathbb{R}^p$

Require: $\boldsymbol{x} \in K$ and $\boldsymbol{x} + \tau\boldsymbol{v} \in \partial K$

Output : a point $\tilde{\boldsymbol{x}} \in \partial K$ and a direction vector $\tilde{\boldsymbol{v}} \in \mathbb{R}^p$ that define the line $\tilde{\ell}(t) = \{t > \tau \mid \tilde{\boldsymbol{x}} + t\tilde{\boldsymbol{v}}\}$.

$\tilde{\boldsymbol{x}} \leftarrow \boldsymbol{x} + \tau\boldsymbol{v}$;
if $\tilde{\boldsymbol{x}} \in \partial H$ **then**
  let j the index computed in Equation (30);
  $j \leftarrow j \mod d$;
  $\tilde{\boldsymbol{v}} \leftarrow \boldsymbol{v}$;
  $\tilde{v}_j \leftarrow -\tilde{v}_j$;
endif;
if $\tilde{\boldsymbol{x}} \in \partial S$ **then**
  $\boldsymbol{s} \leftarrow \nabla \boldsymbol{F}(\tilde{\boldsymbol{x}})$; // the inner vector of the tangent hyperlpane
  $\tilde{\boldsymbol{v}} \leftarrow \boldsymbol{v} - \frac{2}{||\boldsymbol{s}||}(\boldsymbol{v}^T\boldsymbol{s})\boldsymbol{s}$;
endif;
return $\tilde{\boldsymbol{x}},\ \tilde{\boldsymbol{v}}$;

normalized inner vector $\boldsymbol{s}$ of the tangent hyperplane at $\tilde{\boldsymbol{x}}$. Then, the reflected line $\tilde{\ell}(t) = \{t > \tau \mid \tilde{\boldsymbol{x}} + t\tilde{\boldsymbol{v}}\}$, where τ is the intersection time and

$$\tilde{\boldsymbol{v}} = \boldsymbol{v} - 2(\boldsymbol{v}^T\boldsymbol{s})\boldsymbol{s} \tag{33}$$

is the reflected direction vector. Thus, REFLECTION takes as input the line $\ell(t)$ and the intersection time τ and then computes the inner vector $\boldsymbol{s}$ and returns the reflected line.

When $\tilde{\boldsymbol{x}} \in \partial H$ the intersection time has been derived by the Equation (30). Let $\ell(t)$ to hit the j-th facet of H. Then, the inner vector of the tangent hyperplane is

$$s_j = \pm 1,\ s_i = 0,\ i \neq j. \tag{34}$$

When the line $\ell(t)$ hits ∂S at $\tilde{\boldsymbol{x}}$ then in Chalkis et al. (2020) they prove that,

$$\boldsymbol{s} \sim \nabla \boldsymbol{F}(\tilde{\boldsymbol{x}}) = c \cdot (\boldsymbol{w}^T\boldsymbol{A}_1\boldsymbol{w}, \ldots, \boldsymbol{w}^T\boldsymbol{A}_p\boldsymbol{w}) \tag{35}$$

where the constant $c = \frac{\mu(\boldsymbol{F}(\tilde{\boldsymbol{x}}))}{|\boldsymbol{w}|^2}$, $\mu(\boldsymbol{F}(\tilde{\boldsymbol{x}}))$ the product of the non-zero eigenvalues of $\boldsymbol{F}(\tilde{\boldsymbol{x}})$ and $\boldsymbol{w}$ is a non-trivial vector in the kernel of $\boldsymbol{F}(\tilde{\boldsymbol{x}})$. Thus, $\boldsymbol{w}$ corresponds to the eigenvector of the smallest positive eigenvalue we compute in Equation (32).