RC24206 (W0703-023) March 6, 2007
Mathematics

IBM Research Report

On the Computational Complexity of MCMC-Based Estimators in Large Samples

Alexandre Belloni
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Victor Chernozhukov
Department of Economics and Operations Research Center
MIT

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

ON THE COMPUTATIONAL COMPLEXITY OF MCMC-BASED ESTIMATORS IN LARGE SAMPLES

By Alexandre Belloni*

IBM T.J. Watson Research Center and MIT Operations Research Center

and

By Victor Chernozhukov†

MIT Department of Economics and Operations Research Center

This paper studies the computational complexity of Bayesian and quasi-Bayesian estimation in large samples carried out using a basic Metropolis random walk. The framework covers cases where the underlying likelihood or extremum criterion function is possibly nonconcave, discontinuous, and of increasing dimension. Using a central limit framework to provide structural restrictions for the problem, it is shown that the algorithm is computationally efficient. Specifically, it is shown that the running time of the algorithm in large samples is bounded in probability by a polynomial in the parameter dimension d, and in particular is of stochastic order d^2 in the leading cases after the burn-in period. The reason is that, in large samples, a central limit theorem implies that the posterior or quasi-posterior approaches a normal density, which restricts the deviations from continuity and concavity in a specific manner, so that the computational complexity is polynomial. An application to exponential and curved exponential families of increasing dimension is given.

1. Introduction. Markov Chain Monte Carlo (MCMC) algorithms have dramatically increased the use of Bayesian and quasi-Bayesian methods for practical estimation and inference.(See e.g. books of Casella and Robert [6], Chib [9], Geweke [16], Liu [31] for detailed treatments of the MCMC methods and their applications in various areas of statistics, econometrics, and biometrics.) Bayesian methods rely on a likelihood formulation, while quasi-Bayesian methods replace likelihood by other criterion functions. This paper studies the computational complexity of a basic MCMC algorithm as both

*Research support from a National Science Foundation grant is gratefully acknowledged. IBM Herman Goldstein Fellowship is also gratefully acknowledged.

†Research support from a National Science Foundation grant is gratefully acknowledged. Sloan Foundation Research Fellowship and Castle Krob Chair are also gratefully acknowledged.

AMS 2000 subject classifications: Primary, 65C05; secondary 65C60

Keywords and phrases: Monte Carlo, Computational Complexity, Curved Exponential

the sample and parameter dimension grow to infinity at appropriate rates. The paper shows how and when the large sample asymptotics places sufficient restrictions on the likelihood and criterion functions that guarantee the efficient – that is, polynomial time – computational complexity of these algorithms. These results suggest that at least in large samples, Bayesian and Quasi-Bayesian estimators can be computationally efficient alternatives to maximum likelihood and extremum estimators, most of all in cases where likelihoods and criterion functions are non-concave and possibly non-smooth in parameters of interest.

To motivate our analysis, consider the M-estimation problem, which is a common method of estimating various kinds of regression models. The idea behind this approach is to maximize some criterion function:

$$Q_n(\theta) = -\sum_{i=1}^{n} m(Y_i - q_i(X_i, \theta)), \quad \theta \in \Theta \subset \mathbb{R}^d, \tag{1.1}$$

where Y_i is the response variable, X_i is a vector of regressors, and q_i is a regression function. In many examples, the problem is nonlinear and non-concave, implying that the argmax estimator may be difficult or impossible to obtain. For instance, in risk management a major problem is that of constructing the estimates of Conditional Value-at-Risk. In particular, the problem is to predict the α-quantile of a portfolio's return Y_i tomorrow, given today's and past available information $(X_i, X_{i-1}, \ldots)$. This problem fits in the M-estimation framework by taking function $m(\cdot)$ to be the asymmetric absolute deviation function, see Koenker and Bassett [28],

$$m(u) = (\alpha - 1(u < 0))u.$$

To reflect dependence on all past data and accurately capture GARCH-like dependencies, leading research in this area (see Engle and Maganelli [13]) considers recursive models of the form $q_i = f(X_i, q_{i-1}, q_{i-2}, ...; \theta)$, for instance, $f(X_i, q_{i-1}, q_{i-2}, ...; \theta) = X_i'\gamma + \rho_1 q_{i-1} + \rho_2 q_{i-2}$. This implies a highly non-linear, recursive specification for the regression function $q_i(\cdot; \theta)$, which in turn implies that the criterion function used in M-estimation defined in (1.1) is generally non-concave. Furthermore, in this example, the function $Q_n(\theta)$ is non-smooth. As a consequence the argmax estimator

$$\widetilde{\theta} \in \arg\max_{\theta \in \Theta} Q_n(\theta) \tag{1.2}$$

may be very hard to obtain. Figure 1 in Section 2 illustrates other kinds of examples where the argmax computation becomes intractable.

As an alternative to argmax estimation, consider the Quasi-Bayesian estimator obtained by integration in place of optimization:

$$\widehat{\theta} = \frac{\int_\Theta \theta \exp\{Q_n(\theta)\} d\theta}{\int_\Theta \exp\{Q_n(\theta')\} d\theta'}. \tag{1.3}$$

This estimator may be recognized as a quasi-posterior mean of the quasi-posterior density $\pi_n(\theta) \propto \exp Q_n(\theta)$. (Of course, when Q_n is a log-likelihood, the term "quasi" becomes redundant.) This estimator is not affected by local discontinuities and non-concavities and is often much easier to compute in practice than the argmax estimator; see, for example, the discussion in Liu, Tian, and Wei [30] and Chernozhukov and Hong [8].

This paper will show that if the sample size n grows to infinity and the dimension of the problem d does not grow too quickly relative to the sample size, the quasi-posterior

$$\frac{\exp\{Q_n(\theta)\}}{\int_\Theta \exp\{Q_n(\theta')\} d\theta'} \tag{1.4}$$

will be approximately normal. This result in turn leads to the main claim: the estimator (1.3) can be computed using Markov Chain Monte Carlo in polynomial time, provided the starting point is drawn from the approximate support of the quasi-posterior (1.4). As is standard in the literature, we measure running time in the number of evaluations of the numerator of the quasi-posterior function (1.4) since this accounts for most of the computational burden.

In other words, when the central limit theorem (CLT) for the quasi-posterior holds, the above estimator is computationally tractable. The reason is that the CLT, in addition to implying the approximate normality and attractive estimation properties of the estimator $\widehat{\theta}$, bounds non-concavities and discontinuities of $Q_n(\theta)$ in a specific manner that implies that the computational time is polynomial in the parameter dimension d. In particular, the bound on the running time of the algorithm is $O_p(d^2)$ in the leading cases after the so-called burn-in period. Thus, our main insight is to bring the structure implied by the CLT into the computational complexity analysis of the MCMC algorithm for computation of (1.3) and sampling from (1.4).

Our analysis of computational complexity builds on several fundamental papers studying the computational complexity of Metropolis procedures, especially Applegate and Kannan [1], Frieze, Kannan and Polson [15], Polson

[36], Kannan, Lovász and Simonovits [25], Kannan and Li [24], Lovász and Simonovits [32], and Lovász and Vempala [33, 34, 35]. Many of our results and proofs rely upon and extend the mathematical tools previously developed in these works. We extend the complexity analysis of the previous literature, which has focused on the case of an arbitrary concave log-likelihood function, to nonconcave and nonsmooth cases. The motivation is that from a statistical point of view, in concave settings it is typically easier to compute a maximum likelihood or extremum estimate than a Bayesian or quasi-Bayesian estimate, so the latter do not necessarily have practical appeal. In contrast, when the log-likelihood or quasi-likelihood is either nonsmooth, nonconcave, or both, Bayesian and quasi-Bayesian estimates defined by integration are relatively attractive computationally, compared to maximum likelihood or extremum estimators defined by optimization.

Our analysis also relies on statistical large sample theory. We invoke limit theorems for posteriors and quasi-posteriors for large samples as $n \to \infty$. These theorems are necessary to support our principal task – the analysis of computational complexity under the restrictions of the CLT. As a preliminary step of our computational analysis, we obtain a new CLT for quasi-posteriors and posteriors which generalizes the CLT previously derived in the literature for posteriors and quasi-posteriors for fixed dimension. In particular, Laplace c. 1809, Bickel and Yahav [4], Ibragimov and Hasminskii [20], and Bunke and Milhaud [5] provided CLTs theorems for posteriors. Chernozhukov and Hong [8] and Liu, Tian, and Wei [30] provided CLTs for quasi-posteriors formed using various non-likelihood criterion functions. In contrast to these previous results, we allow for increasing dimensions. Ghosal [18] also previously derived a CLT for posteriors with increasing dimension, but only for concave exponential families. We go beyond such canonical setup and establish the CLT for non-concave and discontinuous cases. We also allow for general criterion functions in place of likelihood functions. The paper also illustrates the plausibility of the approach using exponential and curved exponential families. The curved families arise for example when the data must satisfy additional moment restrictions, as e.g. in Hansen and Singleton [19], Chamberlain [7], and Imbens [21]. The curved families fall outside the log-concave framework.

The rest of the paper is organized as follows. In Section 2, we establish a generalized version of the Central Limit Theorem for Bayesian and Quasi-Bayesian estimators. This result may be seen as a generalization of the classical Bernstein-Von-Mises theorem, in that it allows the parameter dimension to grow as the sample size grows, i.e. $d \to \infty$ as $n \to \infty$. In Section 2, we also formulate the main problem, which is to characterize the complexity of MCMC sampling and integration as a function of the key parameters that describe the deviations of the quasi-posterior from the normal density. Section 3 explores the structure set forth in Section 2 to find bounds on conductance and mixing time of the MCMC algorithm. Section 4 derives bounds on the integration time of the MCMC algorithm. Section 5 considers an application to a broad class of curved exponential families, which are possibly non-concave and discontinuous, and verifies that our results apply to this class of statistical models. We verify that high-level conditions of Section 2 follow from primitive conditions for these models.

2. The Setup and The Problem. Our analysis is motivated by the problems of estimation and inference in large samples. We consider a "reduced-form" setup formulated in terms of parameters that characterize local deviations from the true statistical parameter. [1] The local parameter λ describes contiguous deviations from the true parameter and we shift it by a first order approximation of the extremum estimator s. That is, for θ denoting a parameter vector, θ_0 the true value, and $s = \sqrt{n}(\tilde{\theta} - \theta_0)$ the normalized extremum estimator (or a first order approximation to it), we have the local parameter λ defined as

$$\lambda = \sqrt{n}(\theta - \theta_0) - s.$$

The parameter space for θ is Θ, and the parameter space for λ is therefore $\Lambda = \sqrt{n}(\Theta - \theta_0) - s$.

The corresponding localized likelihood (or localized criterion) function is denoted by $\ell(\lambda)$. For example, suppose $L_n(\theta)$ is the original likelihood function in the likelihood framework or, more generally, $L_n(\theta)$ is $\exp\{nQ_n(\theta)\}$ where $Q_n(\theta)$ is the criterion function in extremum framework, then

$$\ell(\lambda) = L_n(\theta_0 + (\lambda + s)/\sqrt{n})/L_n(\theta_0).$$

The assumptions below will be stated directly in terms of $\ell(\lambda)$. (Section 5 provides more primitive conditions within the exponential and curved exponential family framework.)

Then, the posterior or quasi-posterior density for λ takes the form (implicitly indexed by the sample size n)

$$f(\lambda) = \frac{\ell(\lambda)}{\int_\Lambda \ell(\omega)d\omega}, \tag{2.5}$$

[1] Examples in Section 5 further illustrate the connection between the localized set-up and the non-localized set-ups.

and we impose conditions that force the posterior to satisfy a CLT in the sense of approaching the normal density

$$\phi(\lambda) = \frac{1}{(2\pi)^{d/2} \det\left(J^{-1}\right)^{1/2}} \exp\left(-\frac{1}{2}\lambda' J \lambda\right).$$

More formally, the following conditions are assumed to hold for $\ell(\lambda)$ as the sample size $n \to \infty$. These conditions, which in the following we will call the "CLT conditions," explicitly allow for an increasing parameter dimension d ($d \to \infty$):

C1. The local parameter λ belongs to the local parameter space $\lambda \in \Lambda \subset \mathbb{R}^d$. The vector s is a zero mean vector with variance Ω, whose eigenvalues are bounded above as $n \to \infty$, and $\Lambda = K \cup K^c$, where K is a closed ball $B(0, \|K\|)$ with $\|K\| = C\sqrt{d}$ such that $\int_K f(\lambda)d\lambda \geq 1 - o_p(1)$ and $\int_K \phi(\lambda)d\lambda \geq 1 - o(1)$.[2]

C2. The lower semi-continuous posterior or quasi-posterior function $\ell(\lambda)$ approaches a quadratic form in logs, uniformly in K, i.e., there exist positive approximation errors ϵ_1 and ϵ_2 such that for every $\lambda \in K$,

$$\left|\ln \ell(\lambda) - \left(-\frac{1}{2}\lambda' J \lambda\right)\right| \leq \epsilon_1 + \epsilon_2 \cdot \lambda' J \lambda / 2, \tag{2.6}$$

where J is a symmetric positive definite matrix with eigenvalues bounded away from zero and from above. Also, we denote the ellipsoidal norm induced by J as $\|v\|_J := \|J^{1/2}v\|$.

C3. The approximation errors ϵ_1 and ϵ_2 satisfy $\epsilon_1 = o_p(1)$, and $\epsilon_2 \cdot \|K\|_J^2 = o_p(1)$.

These conditions imply that

$$\ell(\lambda) = g(\lambda) \cdot m(\lambda)$$

over the approximate support set K where

$$\ln g(\lambda) = -\frac{1}{2}\lambda' J \lambda, \tag{2.7}$$

$$-\epsilon_1 - \epsilon_2 \lambda' J \lambda / 2 \leq \ln m(\lambda) \leq \epsilon_1 + \epsilon_2 \lambda' J \lambda / 2. \tag{2.8}$$

Figure 1 illustrates the kinds of deviations of $\ln \ell(\lambda)$ from the quadratic curve captured by the parameters ϵ_1 and ϵ_2, and also shows the types of discontinuities and non-convexities permitted in our framework. Parameter

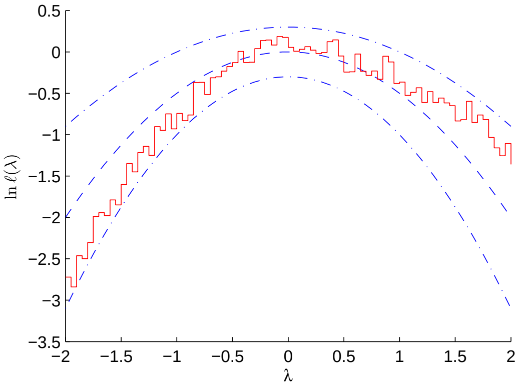

FIG 1. *This figure illustrates how* $\ln \ell(\lambda)$ *can deviate from* $\ln g(\lambda)$ *including possible discontinuities on* $\ln \ell(\lambda)$.

ϵ_1 controls the size of local discontinuities and parameter ϵ_2 controls the global tilting away from the quadratic shape of the normal log-density.

Theorem 1 *[Generalized CLT for Quasi-Posteriors] Under the conditions stated above, the density of interest*

$$f(\lambda) = \frac{\ell(\lambda)}{\int_\Lambda \ell(\omega)d\omega} \tag{2.9}$$

approaches a normal density $\phi(\lambda)$ *with variance matrix* J *in the following sense:*

$$\int_\Lambda |f(\lambda) - \phi(\lambda)|d\lambda = \int_K |f(\lambda) - \phi(\lambda)|d\lambda + o_p(1) = o_p(1). \tag{2.10}$$

Proof. See Appendix A. ■

Theorem 1 is a simple preliminary result. However, the result is essential for defining the environment in which main results of this paper – the computational complexity results – will be developed. The theorem shows that in large samples, provided some regularity conditions hold, Bayesian and Quasi-Bayesian inference has good large sample properties. The main part of the paper, namely Section 3, develops the *computational implications* of the CLT conditions. In particular, Section 3 shows that polynomial time computing of Bayesian and Quasi-Bayesian estimators by MCMC is in fact implied by the CLT conditions.

[2]Note that $\|K\| := \sup\{\|a\| : a \in K\}$. The constant C need not grow due to the phenomenon of concentration of measure under $d \to \infty$ asymptotics.

By allowing increasing dimension ($d \to \infty$) Theorem 1 extends the CLT previously derived in the literature for posteriors in the likelihood framework (Bickel and Yahav [4], Ibragimov and Hasminskii [20], Bunke and Milhadu [5], Ghosal [18]) and for quasi-posteriors in the general extremum framework, when the likelihood is replaced by general criterion functions (Chernozhukov and Hong [8], Liu, Tian, and Wei [30]). The theorem is more general than the results in Ghosal [18], who also considered increasing dimensions but limited his analysis to the exponential likelihood family framework. In contrast, Theorem 1 allows for non-exponential families and allows quasi-posteriors in place of posteriors. Recall that quasi-posteriors result from using quasi-likelihoods and other criterion functions in place of the likelihood. This expands substantially the scope of the applications of the result. Importantly, Theorem 1 allows for non-smoothness and even discontinuities in the likelihood and criterion functions, which are pertinent in a number of applications listed in the introduction.

The Problem of the Paper. Our problem is to characterize the complexity of obtaining draws from $f(\lambda)$ and of Monte Carlo integration

$$\int g(\lambda)f(\lambda)d\lambda,$$

where $f(\lambda)$ is restricted to the approximate support K. The procedure used to obtain the basic draws as well as to carry out Monte Carlo integration is a Metropolis (Gaussian) random walk, which is a standard MCMC algorithm used in practice. The tasks are thus:

- **I**. Characterize the complexity of sampling from $f(\lambda)$ as a function of $(d, n, \epsilon_1, \epsilon_2, K)$;
- **II**. Characterize the complexity of calculating $\int g(\lambda)f(\lambda)d\lambda$ as a function of $(d, n, \epsilon_1, \epsilon_2, K)$;
- **III**. Characterize the complexity of sampling from $f(\lambda)$ and performing integrations with $f(\lambda)$ in large samples as $d, n \to \infty$ by invoking the bounds on $(d, n, \epsilon_1, \epsilon_2, K)$ imposed by the CLT;
- **IV**. Verify that the CLT conditions are applicable in a variety of statistical problems.

This paper formulates and answers this problem. Thus, the paper brings the CLT restrictions into the complexity analysis and develops complexity bounds for sampling and integrating from $f(\lambda)$ under these restrictions. These CLT restrictions, arising by using large sample theory and imposing certain regularity conditions, limit the behavior of $f(\lambda)$ over the approximate support set K in a specific manner that allows us to establish polynomial computing time for sampling and integration. Because the conditions for the CLT do not provide strong restrictions on the tail behavior of $f(\lambda)$ outside K other than C1, our analysis of complexity is limited entirely to the approximate support set K defined in C1-C3.

By solving the above problem, this paper contributes to the recent literature on the computational complexity of Metropolis procedures. Early work was primary concerned with the question of approximating the volume of high dimensional convex sets where uniform densities play a fundamental role (Lovász and Simonovits [32], Kannan, Lovász and Simonovits [25, 26]). Later the approach was generalized for the cases where the log-likelihood is concave (Frieze, Kannan and Polson [15], Polson [36], and Lovász and Vempala [33, 34, 35]). However, under log-concavity the maximum likelihood estimators are usually preferred over Bayesian or quasi-Bayesian estimator from a computational point of view. In the absence of concavity, exactly the settings where there is a great practical appeal for using Bayesian and quasi-Bayesian estimates, there has been relatively less, if any, analysis. One important exception is the paper of Applegate and Kannan [1], which covers nearly-concave but smooth densities using a discrete Metropolis algorithm. [3] In contrast to Applegate and Kannan [1], our approach allows for both discontinuous and non-concave densities that are permitted to deviate from the normal density (not from an arbitrary log-concave density, like in Applegate and Kannan [1]) in a specific manner. The manner in which they deviate from the normal is motivated by the CLT and controlled by parameters ϵ_1 and ϵ_2, which are in turn restricted by the CLT conditions. The CLT restrictions provide a congenial analytical framework that allows us to treat a basic non-discrete sampling algorithm frequently used in practice. In fact, it is known that the basic Metropolis walk analyzed here does not have good complexity properties (rapidly mixing) for arbitrary log-concave density functions, in opposition to other random walks like hit-and-run[4]. Nonetheless, the CLT conditions imply enough structure that the random walk is in fact rapidly mixing. Moreover, only Subsection 3.2 depends on the particular form of the walk while all the remaining results are valid in a considerably more general setting. This suggests that the same CLT approach can be used to establish polynomial bounds for more sophisticated schemes. As is standard in the literature, we assume that the starting point of the algorithm is in the approximate support of the posterior. Indeed, the polynomial time bound that we derive applies only in this case because this is the domain where the CLT provides enough structure on the problem.

[3]The discrete Metropolis algorithm facilitates the analysis, but they are are not frequently used in practice.

[4]See Lovász and Vempala [35] for a discussion.

Our analysis does not apply outside this domain.

3. Sampling from f using a Random Walk.

3.1. *Set-Up and Main Result.* In this section we bound the computational complexity of obtaining a draw from a random variable approximately distributed according to a density function f as defined in (2.5). (Section 4 builds upon these results to study the associated integration problem.) By invoking Assumption C1, we restrict our attention entirely to the approximate support set K and the accuracy of sampling will be defined over this set. Consider a measurable space $(K, \mathcal{A})$. Our task is to draw a random variable according to a measurable density function f restricted to K (this density induces a probability distribution on K denoted by Q, i.e., $Q(A) = \int_A f(x)dx / \int_K f(x)dx$ for all $A \in \mathcal{A}$). Asymptotically, it is well-known that random walks combined with a Metropolis filter are capable of performing such task. In order to prove our complexity bounds, we will concentrate on a commonly used random walk induced by a Gaussian distribution. Such random walk is completely characterized by an initial point u_0 and a fixed standard deviation $\sigma > 0$, and its *one-step* move. The latter is defined as the procedure of drawing a point y according to a Gaussian distribution centered on the current point u with covariance matrix $\sigma^2 I$ and then, with probability $\min\{f(y)/f(u), 1\} = \min\{\ell(y)/\ell(u), 1\}$ move to y; otherwise stay at u (see Casella and Robert [6] and Vempala [41] for details). In the complexity analysis of this algorithm we are interested in bounding the number of steps of the random walk required to draw a random variable from f with a given precision. Equivalently, we are interested in bounding the number of evaluations of the local likelihood function ℓ required for this purpose.

Next we review definitions of important concepts relevant for our analysis. The definitions of these concepts follow Lovász and Simonovits [32] and Vempala [41]. Let $q(x|u)$ denote the density function associated with a random variable $N(u, \sigma^2 I)$, and $1_u(A)$ be the indicator function of the set A. For each $u \in K$ the one-step distribution P_u, the probability distribution after one step of the random walk starting from u, is defined as

$$P_u(A) = \int_{K \cap A} \min\left\{ \frac{f(x)}{f(u)}, 1 \right\} q(x|u)dx + \theta 1_u(A) \tag{3.11}$$

where

$$\theta = 1 - \int_K \min\left\{ \frac{f(x)}{f(u)}, 1 \right\} q(x|u)dx \tag{3.12}$$

is the probability of staying at u after one step of the ball walk from u. A step of the random walk is said to be proper if the next point is different from the current point (which happens with probability $1-\theta$).

The triple $(K, \mathcal{A}, \{P_u : u \in K\})$, along with a starting distribution Q_0, defines a Markov chain in K. We denote by Q_t the probability distribution obtained after t steps of the random walk. A distribution Q is called stationary on $(K, \mathcal{A})$ if for any $A \in \mathcal{A}$,

$$\int_K P_u(A) dQ(u) = Q(A). \tag{3.13}$$

Given the random walk described earlier, the unique stationary probability distribution Q is induced by the function f, $Q(A) = \int_A f(x)dx / \int_K f(x)dx$ for all $A \in \mathcal{A}$, see e.g. Casella and Roberts [6]. This is the main motivation for most of the MCMC studies found in the literature since it provides an asymptotic method to approximate the density of interest. As mentioned before, our goal is to properly quantify this convergence and for that we need to review additional concepts.

The ergodic flow of a set A with respect to a distribution Q is defined as

$$\Phi(A) = \int_A P_u(K \backslash A) dQ(u).$$

It measures the probability of the event $\{u \in A, u' \notin A\}$ where u is distributed according to Q and u' is obtained after one step of the random walk starting from u; it captures the average flow of points leaving A in one step of the random walk. It follows that Q is a stationary measure if and only if $\Phi(A) = \Phi(K \backslash A)$ for all $A \in \mathcal{A}$ since

$$\begin{aligned}
\Phi(A) &= \int_A P_u(K \setminus A) dQ(u) = \int_A (1 - P_u(A))\, dQ(u) \\
&= Q(A) - \int_A P_u(A) dQ(u) = \int_K P_u(A) dQ(u) - \int_A P_u(A) dQ(u) \\
&= \Phi(K \setminus A).
\end{aligned}$$

A Markov chain is said to be ergodic if $\Phi(A) > 0$ for every A with $0 < Q(A) < 1$, which is the case for the Markov chain induced by the random walk described earlier due to the assumptions on f.

In order to compare two probability distributions P and Q we use the total variation distance[5]

$$\|P - Q\|_{TV} = \sup_{A \subseteq K} |P(A) - Q(A)|. \tag{3.14}$$

Moreover, P is said to be a M-warm start with respect to Q if

$$\sup_{A \in \mathcal{A}: Q(A) > 0} \frac{P(A)}{Q(A)} \leq M. \tag{3.15}$$

The key concepts in the analysis are the conductance of a set A, which is defined as

$$\phi(A) = \frac{\Phi(A)}{\min\{Q(A), Q(K \backslash A)\}},$$

and the global conductance, defined as

$$\phi = \min_A \phi(A) = \min_{0 < Q(A) \leq 1/2} \frac{\Phi(A)}{Q(A)} = \min_{0 < Q(A) \leq 1/2} \frac{\int_A P_u(K \backslash A) dQ(u)}{Q(A)}.$$

Lovász and Simonovits [32] proved the connection between conductance and convergence for the continuous space setting. This result extended an earlier result of Jerome and Sinclair [22, 23], who connected convergence and conductance for discrete state spaces. Lovász and Simonovits' result can be re-stated as follows.

Theorem 2 *Let Q_0 be a M-warm start with respect to the stationary distribution Q. Then,*

$$\|Q_t - Q\|_{TV} \leq \sqrt{M} \left(1 - \frac{\phi^2}{2}\right)^t$$

Proof. See Lovász and Simonovits [32]. ■

The main result of this paper provides a lower bound for the global conductance of the Markov chain ϕ under the CLT conditions. In particular, we show that $1/\phi$ is bounded by a fixed polynomial in the dimension of the parameter space.

Theorem 3 *(Main Result) Under Assumptions C1, C2, C3, and setting the parameter σ for the random walk as defined in (3.17), the global conductance of the induced Markov chain satisfies*

$$1/\phi = O\left(d \; e^{4\epsilon_1 + 4\epsilon_2 \|K\|_J^2}\right).$$

[5] This distance is equivalent to the $L^1(K)$ distance between the density functions associated with P and Q since $\sup_{A \subseteq K} |P(A) - Q(A)| = \frac{1}{2} \int_K |dP - dQ|$.

In particular, the random walk requires at most

$$N_\varepsilon = O_p\left(e^{8\epsilon_1+8\epsilon_2\|K\|_J^2}\ \left(\frac{1}{\sigma}\right)^2\ \ln(M/\varepsilon)\right)$$

steps to achieve $\|Q_{N_\varepsilon} - Q\|_{TV} \leq \varepsilon$. *Invoking the CLT restrictions,* $\epsilon_1 = o(1)$, $\epsilon_2 \cdot \|K\|_J = o(1)$, *we have that* $1/\phi = O_p(d)$ *and the number of steps* N_ε *is bounded by*

$$O_p\left(d^2\ \ln(M/\varepsilon)\right).$$

Proof. See Section 3.2. ■

Comment 3.1 *In general, the dependence on* ϵ_1 *and* ϵ_2 *is exponential and this bound does not imply polynomial time ("efficient") computing. However, the CLT framework implies that* $\epsilon_1 = o(1)$ *and* $\epsilon_2 \cdot \|K\|_J = o(1)$, *which by Theorem 3 in turn implies polynomial time computing.*

Next we discuss and bound the dependence on M, the "distance" of the initial distribution Q_0 from the stationary distribution Q as defined in (3.15). A natural candidate for a starting distribution Q_0 is the one-step distribution conditional on a proper move from an arbitrary point $u \in K$. We emphasize that, in general, such choice of Q_0 could lead to values of M that are arbitrary large. In fact, this could happen even in the case of the stationary density being a uniform distribution on a convex set (see [35]). Fortunately, this is not the case under the CLT framework as shown by the following lemma.

Lemma 1 *Let* $u \in K$ *and* P_u *be the associated one-step distribution. With probability at least* $1/3$ *the random walk makes a proper move. Conditioned on performing a proper move, the one-step distribution is a* M*-warm start start for* f, *where*

$$\ln M = O(d\ln(\sqrt{d}\|K\|) + \|K\|_J^2 + \epsilon_1 + \epsilon_2\|K\|_J^2).$$

Under the CLT restrictions, $\epsilon_2\|K\|_J = o(1)$ *and* $\|K\|_J = O(\sqrt{d})$, *so that*

$$\ln M = O(d \ln d).$$

Proof. See Appendix A. ■

Combining this result with Theorem 3 yields the overall (burn-in plus post burn-in) running time

$$O_p(d^3 \ln d).$$

3.2. *Proof of the Main Result.* The proof of Theorem 3 uses two auxiliary results: an iso-perimetric inequality and a geometric property of the particular random walk. The first is an analytical result and is of independent mathematical interest. After the connection between the iso-perimetric inequality and the ergodic flow is established, the second result allows us to use the first result to bound the conductance from below. In what follows we provide an outline of the proof, auxiliary results, and, finally, the formal proof.

3.2.1. *Outline of the Proof.* The proof follows the arguments in Lovász and Vempala [33]. In order to bound the ergodic flow of $A \in \mathcal{A}$, consider the particular disjoint partition $K = \widetilde{S}_1 \cup \widetilde{S}_2 \cup \widetilde{S}_3$ where $\widetilde{S}_1 \subset A$, $\widetilde{S}_2 \subset K \setminus A$, and $\widetilde{S}_3$ consists of points in A or $K \setminus A$ for which the one-step probability of going to the other set is at least a fixed constant c (to be defined later). Therefore we have

$$\begin{aligned}\Phi(A) &= \int_A P_u(K \setminus A)dQ(u) = \frac{1}{2}\int_A P_u(K \setminus A)dQ(u) + \frac{1}{2}\int_{K\setminus A} P_u(A)dQ(u) \\ &\geq \frac{1}{2}\int_{\widetilde{S}_1} P_u(K \setminus A)dQ(u) + \frac{1}{2}\int_{\widetilde{S}_2} P_u(A)dQ(u) + \frac{c}{2}Q(\widetilde{S}_3).\end{aligned}$$

where the second equality holds because $\Phi(A) = \Phi(K \setminus A)$.

Since the first two terms could be arbitrarily small, the result will follow by bounding the last term from below. This will be achieved by an iso-perimetric inequality which is tailored to the CLT framework and is derived in Section 3.2.2. This result will provide a lower bound on $Q(\widetilde{S}_3)$, which is decreasing in the distance between $\widetilde{S}_1$ and $\widetilde{S}_2$. Therefore one still needs to bound the distance between these sets.

Given two points $u \in \widetilde{S}_1$ and $v \in \widetilde{S}_2$, we have $P_u(K\setminus A) \leq c$ and $P_v(A) \leq c$. Therefore, the total variation distance between their one-step distributions $\|P_u - P_v\| \geq |P_u(A) - P_v(A)| \geq 1 - 2c$. The geometric properties of the random walk are used to ensure that this condition implies that $\|u - v\|$ is also bounded from below (see Section 3.2.3). Since u and v were arbitrary points, the sets $\widetilde{S}_1$ and $\widetilde{S}_2$ are "far" apart. Therefore $\widetilde{S}_3$ cannot be arbitrarily small, i.e., $Q(\widetilde{S}_3)$ is bounded from below.

This leads to a lower bound for the global conductance. After bounding the global conductance from below, Theorem 3 follows by invoking CLT conditions and Theorem 2.

3.2.2. *An Iso-perimetric Inequality.* We start by defining a notion of approximate log-concavity. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be log-β-concave if for every $\alpha \in [0, 1]$, $x, y \in \mathbb{R}^n$, we have

$$f(\alpha x + (1 - \alpha)y) \geq \beta f(x)^\alpha f(y)^{1-\alpha}$$

for some $\beta \in (0,1]$. f is said to be logconcave if β can be taken equal to one. The class of log-β-concave functions is rather broad, for example, including various non-smooth and discontinuous functions.

Together, the relations (2.7) and (2.8) imply that we can write the functions f and ℓ as the product of $e^{-\frac{1}{2}\lambda' J\lambda}$ and a log-β-concave function:

Lemma 2 *Over the set K the functions $f(\lambda) := \ell(\lambda)/\int_\Lambda \ell(\lambda)d\lambda$ and $\ell(\lambda)$ are the product of a Gaussian function, $e^{-\frac{1}{2}\lambda' J\lambda}$, and a β-log-concave function whose parameter β satisfies*

$$\ln \beta \geq 2 \cdot \left(-\epsilon_1 - \epsilon_2 \cdot \|K\|_J^2\right).$$

Proof. It follows from (2.8). ■

In our case, the larger is the support set K, the larger is the deviation from log-concavity. That is appropriate since the CLT does not impose strong restrictions on the tail of the probability densities. Nonetheless, this gives a convenient structure to prove an iso-perimetric inequality which covers even non-continuous cases permitted in the framework described in the previous sections.

Lemma 3 *Consider any measurable partition of the form $K = S_1 \cup S_2 \cup S_3$ such that the distance between S_1 and S_2 is at least t, i.e. $d(S_1, S_2) \geq t$. Let $Q(S) = \int_S f dx / \int_K f dx$. Then for any lower semi-continuous function $f(x) = e^{-\|x\|^2} m(x)$, where m is a log-β-concave function, we have*

$$Q(S_3) \geq \beta \frac{2te^{-t^2/4}}{\sqrt{\pi}} \min\{Q(S_1),\ Q(S_2)\}.$$

Proof. See Appendix A. ■

Comment 3.2 *This new iso-perimetric inequality extends the iso-perimetric inequality in Kannan and Li [24], Theorem 2.1. The proof builds on their proof as well as on the ideas in Applegate and Kannan [1]. Unlike the inequality in [24], Lemma 3 removes smoothness assumptions on f, for example, covering both non-log-concave and discontinuous cases.*

The iso-perimetric inequality of Lemma 3 states that, under suitable conditions, if two subsets of K are far apart, the measure of the remaining subset should be comparable to the measure of at least one of the original subsets. The following corollary extends the previous theorem to cover cases with an arbitrary covariance matrix J.

Corollary 1 *Consider any measurable partition of the form $K = S_1 \cup S_3 \cup S_2$ such that $d(S_1, S_2) \geq t$, and let $Q(S) = \int_S f dx / \int_K f dx$. Then for any lower* *semi-continuous function* $f(x) = e^{-\frac{1}{2}x'Jx}m(x)$, *where* m *is a log-*β*-concave function, we have*

$$Q(S_3) \geq \beta\ te^{-\lambda_{min}t^2/8}\ \sqrt{\frac{2\lambda_{min}}{\pi}}\min\{Q(S_1),\ Q(S_2)\},$$

where λ_{min} *denotes the minimum eigenvalue of the positive definite matrix* J.

Proof. See Appendix A. ■

3.2.3. *Bounds on the Difference of One-step Distributions.* Next we relate the total variation distance between two one-step distributions with the Euclidean distance between the points that induce them. Although this approach follows the one in Lovász and Vempala [33, 34, 35] there are two important differences which call for a new proof. First, we no longer rely on log-concavity of f. Second, we use a different random walk. We start with the following auxiliary result.

Lemma 4 *Let* $g : \mathbb{R}^n \to \mathbb{R}$ *be a function such that* $\ln g$ *is Lipschitz with constant* L *over compact set* K. *Then, for every* $x \in K$ *and* $r > 0$,

$$\inf_{y\in B(x,r)\cap K}[g(y)/g(x)] \geq e^{-Lr}.$$

Proof. The result is obvious. ■

Given a compact set K, we can bound the Lipschitz constant of the concave function $\ln g$ defined in (2.7) by

$$L \leq \sup_{\lambda\in K}\|\nabla \ln g(\lambda)\| \leq \sup_{\lambda\in K}\|J\lambda\| \leq \lambda_{max}\|K\| = O\left(\sqrt{d}\right). \quad (3.16)$$

Lemma 5 *Let* $u, v \in K := B(0, \|K\|)$, $\sigma^2 \leq \frac{1}{16dL^2}$, *and suppose that* $\frac{\sigma}{\|K\|} \leq \frac{1}{120d}$ *and* $\|u - v\| < \frac{\sigma}{8}$ *where* L *is the Lipschitz constant of* $\ln g$ *on the set* K. *Under our assumptions on* f *as defined in (2.5), we have*

$$\|P_u - P_v\|_{TV} \leq 1 - \frac{\beta}{3e}.$$

Proof. See Appendix A. ■

The converse of Lemma 5 states that if two points induce one-step probability distributions that are far apart in the total variation norm, these points must also be far apart in the Euclidean norm. This geometric result provides a key ingredient in the application of the iso-perimetric inequality as discussed earlier.

3.2.4. **Proof of Theorem 3.** . Consider the compact support for f as $K = B(0, \|K\|)$, where $\|K\| = O(\sqrt{d/\lambda_{min}})$ from Assumption C1. We define

$$\sigma = \min\left\{1/4\sqrt{d}L, \|K\|/120d\right\}. \tag{3.17}$$

Therefore the assumptions of Lemma 5 are satisfied. Moreover, under the assumptions of the theorem, using (3.16) it follows that

$$\sigma \geq \frac{1}{120\lambda_{max}\sqrt{d}\|K\|}. \tag{3.18}$$

Fix an arbitrary set $A \in \mathcal{A}$ and denote by $A^c = K \setminus A$ the complement of A with respect to K. We will prove that

$$\Phi(A) = \int_A P_u(A^c)dQ(u) \geq \frac{\beta^2}{600}\sigma\sqrt{\lambda_{min}}\min\{Q(A), Q(A^c)\}, \tag{3.19}$$

which implies the desired bound on the global conductance ϕ. Note that this is equivalent to bounding $\Phi(A^c)$ since Q is stationary on $(K, \mathcal{A})$.

Consider the following auxiliary definitions:

$$\widetilde{S}_1 = \left\{u \in A : P_u(A^c) < \frac{\beta}{6e}\right\}, \widetilde{S}_2 = \left\{v \in A^c : P_v(A) < \frac{\beta}{6e}\right\}, \widetilde{S}_3 = K \backslash (\widetilde{S}_1 \cup \widetilde{S}_2).$$

First assume that $Q(\widetilde{S}_1) \leq Q(A)/2$ (a similar argument can be made for $\widetilde{S}_2$ and A^c). In this case, we have

$$\begin{aligned} \Phi(A) &= \int_A P_u(A^c)dQ(u) \geq \int_{A\backslash\widetilde{S}_1} P_u(A^c)dQ(u) \geq \int_{A\backslash\widetilde{S}_1} \frac{\beta}{6e}dQ(u) \\ &\geq \frac{\beta}{6e}Q(A\backslash\widetilde{S}_1) \geq \frac{\beta}{12e}Q(A), \end{aligned}$$

and the inequality (3.19) follows.

Next assume that $Q(\widetilde{S}_1) \geq Q(A)/2$ and $Q(\widetilde{S}_2) \geq Q(A^c)/2$. Since $\Phi(A) = \Phi(A^c)$ we have that

$$\begin{aligned} \Phi(A) = \int_A P_u(A^c)dQ(u) &= \tfrac{1}{2}\int_A P_u(A^c)dQ(u) + \tfrac{1}{2}\int_{A^c} P_v(A)dQ(v) \\ &\geq \tfrac{1}{2}\int_{A\backslash\widetilde{S}_1} P_u(A^c)dQ(u) + \tfrac{1}{2}\int_{A^c\backslash\widetilde{S}_2} P_v(A)dQ(v) \\ &\geq \tfrac{1}{2}\int_{\widetilde{S}_3} \tfrac{\beta}{6e}dQ(u) = \tfrac{\beta}{12e}Q(\widetilde{S}_3), \end{aligned}$$

where we used that $\widetilde{S}_3 = K \setminus (\widetilde{S}_1 \cup \widetilde{S}_2) = (A \setminus \widetilde{S}_1) \cup (A^c \setminus \widetilde{S}_2)$. Given the definitions of the sets $\widetilde{S}_1$ and $\widetilde{S}_2$, for every $u \in \widetilde{S}_1$ and $v \in \widetilde{S}_2$ we have

$$\|P_u - P_v\|_{TV} \geq P_u(A) - P_v(A) = 1 - P_u(A^c) - P_v(A) \geq 1 - \frac{\beta}{3e}.$$

In such case, by Lemma 5, we have that $\|u-v\| > \frac{\sigma}{8}$ for every $u \in \widetilde{S}_1$ and $v \in \widetilde{S}_2$. Thus, we can apply the iso-perimetric inequality of Corollary 1, with $d(\widetilde{S}_1, \widetilde{S}_2) \geq \sigma/8$, to bound $Q(\widetilde{S}_3)$. We obtain

$$\begin{aligned} \int_A P_u(A^c)dQ(u) &\geq \frac{\beta^2}{12e}\frac{\sigma}{8}e^{-\frac{1}{8}\lambda_{min}\sigma^2/64}\sqrt{\frac{2\lambda_{min}}{\pi}}\min\{Q(\widetilde{S}_1), Q(\widetilde{S}_2)\} \\ &\geq \frac{\beta^2\sigma\sqrt{\lambda_{min}}}{600}\min\{Q(A), Q(A^c)\}. \end{aligned}$$

where the second inequality also used that $\lambda_{min}\sigma^2 \leq \lambda_{min}\frac{\|K\|^2}{(120d)^2} \leq 1/d$ under our definitions. Therefore, using relation (3.18) and $\|K\| = O(\sqrt{d/\lambda_{min}})$, we obtain

$$1/\phi = O\left(\beta^{-2}\frac{\lambda_{max}}{\lambda_{min}}d\right) = O\left(d\ e^{4\epsilon_1 + 4\epsilon_2\|K\|_J}\right)$$

since the eigenvalues are assumed to be uniformly bounded from above and away from zero.

The remaining results in Theorem 3 follow by invoking the CLT conditions and applying Theorem 2 with the above bound on the conductance.■

4. Complexity of Monte Carlo Integration. This section considers our second problem of interest – that of computing a high dimensional integral of a bounded real valued function g:

$$\mu_g = \int_K g(\lambda)f(\lambda)d\lambda. \tag{4.20}$$

The integral is computed by simulating a dependent (Markovian) sequence of random points λ^1, λ^2, ..., which has f as the stationary distribution, and taking

$$\widehat{\mu}_g = \frac{1}{N}\sum_{i=1}^{N} g(\lambda^i) \tag{4.21}$$

as an approximation to (4.20). The dependent nature of the sample increases the sample size needed to achieve a desired precision compared to the (infeasible) case of independent draws from f. It turns out that as in the preceding analysis, the global conductance of the the Markov chain sample will be crucial in determining the appropriate sample size.

The starting point of our analysis is a central limit theorem for reversible Markov chains due to Kipnis and Varadhan [27] which is restated here for convenience. Consider a reversible Markov chain on K with stationary distribution f. The lag k autocovariance of the stationary time series $\{g(\lambda^i)\}_{i=1}^{\infty}$, obtained by starting the Markov chain with the stationary distribution f is defined as

$$\gamma_k = \text{Cov}_f \left(g(\lambda^i), g(\lambda^{i+k}) \right).$$

Let us recall a characterization of γ_k via spectral theory following Kipnis and Varadhan [27]: Let T denote the transition operator of the Markov chain induced by the random walk. In our case, since the chain is reversible, T is a linear bounded self-adjoint operator in the Hilbert space $L^2(K, \mathcal{A}, Q)$, see [32]. Let E_g denote the measure on Borel sets of $(-1, 1)$ induced by the spectral measure of T applied to g, as in [17]. With these definitions, one has that for any k

$$\gamma_k = \int w^{|k|} dE_g(w).$$

We are prepared to restate the central limit theorem of Kipnis and Varadhan [27] needed for our analysis.

Theorem 4 *For a stationary, irreducible, reversible Markov chain with $\widehat{\mu}_g$ and μ_g defined as (4.21) and (4.20),*

$$N\text{Var}(\widehat{\mu}_g) \to \sigma_g^2 = \sum_{k=-\infty}^{+\infty} \gamma_k = \int \frac{1+w}{1-w} dE_g(w)$$

almost surely. If σ_g^2 is finite, then $\sqrt{N}(\widehat{\mu}_g - \mu_g)$ converges in distribution to $N(0, \sigma_g^2)$.

Proof. See Kipnis and Varadhan [27]. ■

In our case, γ_0 is finite since g is bounded. The next result, which builds upon Theorem 2, states that σ_g^2 can be bounded using the global conductance of the Markov chain.

Corollary 2 *Let g be a square integrable function with respect to the stationary measure Q. Under the assumptions of Theorem 4, we have that*

$$\gamma_k \leq \left(1 - \frac{\phi^2}{2}\right)^{|k|} \gamma_0 \quad \text{and} \quad \sigma_g^2 \leq \gamma_0 \left(\frac{4}{\phi^2}\right).$$

Proof. See Lovaász and Simonovits [32]. ■

We will use the mean square error as the measure of closeness for a consistent estimator:

$$MSE(\widehat{\mu}_g) = E[\widehat{\mu}_g - \mu_g]^2.$$

Many approaches are possible for constructing the sequence of draws in (4.21); we refer to [17] for a detailed discussion. Here, we will analyze three common schemes:

- long run (lr),
- subsample (ss),
- multi-start (ms).

Denote the sample sizes corresponding to each method as N_{lr}, N_{ss}, and N_{ms}. The long run scheme consists of generating the first point using the starting distribution and, after the burn-in period, selecting the N_{lr} subsequent points to compute the sample average (4.21). The subsample method also uses only one sample path, but the N_{ss} draws used in the sample average (4.21) are spaced out by S steps of the chain. Finally, the multi-start scheme uses N_{ms} different sample paths, initializing each one independently from the starting distribution f_0 and picking the last draw in each sample path after the burn-in period to be used in (4.21).

There is another issue that must be addressed. All schemes require that the initial points are drawn from the stationary distribution f. We therefore need to compute the so called "burn-in" period B, that is, the number of iterations required to reach the stationary distribution f with a desired precision, beginning at the starting distribution f_0.

Theorem 5 *Let f_0 be a M-warm start with respect to f, and let $\bar{g} := \sup_{\lambda \in K} |g(\lambda)|$. Using the notation introduced in this section, to obtain*

$$MSE(\widehat{\mu}_g) < \varepsilon$$

it is sufficient to use the following lengths of the burn-in sample, B, and post-burn in samples, N_{lr}, N_{ss}, M_{ms}:

$$B = \left(\frac{2}{\phi^2}\right) \ln \left(\frac{6\sqrt{M}\bar{g}^2}{\varepsilon}\right)$$

and

$$N_{lr} = \frac{\gamma_0}{\varepsilon} \frac{6}{\phi^2}, \quad N_{ss} = \frac{3\gamma_0}{\varepsilon} \ (\textit{with } S = (2/\phi^2) \ln (6\gamma_0/\varepsilon)), \quad N_{ms} = \frac{2\gamma_0}{3\varepsilon}.$$

The overall complexities of lr, ss, and ms methods are thus $B+N_{lr}$, $B+SN_{ss}$, and $B \times N_{ms}$.

Proof. See Appendix A. ■

For convenience Table 1 tabulates the bounds for the three different schemes. Note that the dependence on M and $\bar{g}$ is only via log terms. Although the optimal choice of the method depends on the particular values of the constants, when $\varepsilon \searrow 0$, the long-run algorithm has the smallest (best) bound, while the the multi-start algorithm has the largest (worst) bound

TABLE 1
Burn-in and Post Burn-in Complexities

Method	Quantities	Complexity
Long Run	$B + N_{lr}$	$\frac{2}{\phi^2}\left(\ln\left(\frac{6\sqrt{M}\bar{g}^2}{\varepsilon}\right)\right) + \frac{2}{\phi^2}\left(\frac{3\gamma_0}{\varepsilon}\right)$
Subsample	$B + N_{ss} \cdot S$	$\frac{2}{\phi^2}\left(\ln\left(\frac{6\sqrt{M}\bar{g}^2}{\varepsilon}\right)\right) + \frac{2}{\phi^2}\left(\frac{3\gamma_0}{\varepsilon}\ln\left(\frac{6\gamma_0}{\varepsilon}\right)\right)$
Multi-start	$B \times N_{ms}$	$\frac{2}{\phi^2}\ln\left(\left(\frac{6\sqrt{M}\bar{g}^2}{\varepsilon}\right)\right) \times \frac{2\gamma_0}{3\varepsilon}$

TABLE 2
Burn-in and Post Burn-in Complexities under the CLT.

Method	Burn-in Complexity	Post-burn-in Complexity
Long Run	$O_p(d^3 \ln d \cdot \ln \varepsilon^{-1})$	$+\ O_p(d^2 \cdot \varepsilon^{-1})$
Subsample	$O_p(d^3 \ln d \cdot \ln \varepsilon^{-1})$	$+\ O_p(d^2 \cdot \varepsilon^{-1} \cdot \ln \varepsilon^{-1})$
Multi-start	$O_p(d^3 \ln d \cdot \ln \varepsilon^{-1})$	$\times\ O_p(\varepsilon^{-1})$

on the number of iterations. Table 2 presents complexities implied by the CLT conditions, namely $\|K\| = O(\sqrt{d})$, $\epsilon_1 \to 0$, and $\epsilon_2\|K\|^2 \to 0$. The table assumes γ_0 and $\bar{g}$ are constant, though it is straightforward to tabulate the results for the case where γ_0 and $\bar{g}$ grow at polynomial speed with d. Finally, note that the bounds apply under a slightly weaker condition than the CLT requires, namely that $\epsilon_1 = O_p(1)$ and $\epsilon_2\|K\|^2 = O_p(1)$.

5. Application to Exponential and Curved-Exponential Families. In this section we verify that our conditions and analysis apply to a variety of statistical problems. We begin the discussion with the canonical log-concave cases within the exponential family. Then we drop the concavity and smoothness assumptions to illustrate the full applicability of the approach developed in this paper.

5.1. *Concave Cases.* Exponential families play a very important role in statistical estimation, cf. Lehmann and Casella [29], especially in high-dimensional contexts, cf. Portnoy [37], Ghosal [18], and Stone et al. [39]. For example, the high-dimensional situations arise in modern data sets in technometric and econometric applications. Moreover, exponential familes have excellent approximation properties and are useful for approximation of densities that are not necessarily of the exponential form, cf. Stone et al. [39].

Our discussion is based on the asymptotic analysis of Ghosal [18]. In order to simplify exposition, we invoke the more canonical assumptions similar to those given in Portnoy [37].

> **E1**. Let $x_1, \ldots, x_n$ be iid observations from a d-dimensional canonical exponential family with density
>
> $$f(x;\theta) = \exp\left(x'\theta - \psi_n(\theta)\right),$$
>
> where $\theta \in \Theta$ is an open subset of $\mathbb{R}^d$, and $d \to \infty$ as $n \to \infty$. Fix a sequence of parameter points $\theta_0 \in \Theta$. Set $\mu = \psi'(\theta_0)$ and $F = \psi''(\theta_0)$, the mean and covariance of the observations, respectively. Following Portnoy [37], we implicitly re-parameterize the problem, so that the Fisher information matrix $F = I$.

For a given prior π on Θ, the posterior density of θ over Θ conditioned on the data takes the form

$$\pi_n(\theta) \propto \pi(\theta) \cdot \prod_{i=1}^{n} f(x_i;\theta) = \pi(\theta) \cdot \exp\left(n\bar{x}'\theta - n\psi(\theta)\right).$$

The local parameter space is $\sqrt{n}(\Theta - \theta_0)$. It will be convenient to associate every point θ in the parameter space Θ with an element of Λ, a translation of the local parameter space,

$$\lambda = \sqrt{n}(\theta - \theta_0) - s,$$

where $s = \sqrt{n}(\bar{x} - \mu)$ is a first order approximation to the normalized maximum likelihood/extremum estimate. By design, we have that $E[s] = 0$ and $E[ss'] = I_d$. Moreover, by Chebyshev's inequality, the norm of s can be bounded in probability, $\|s\| = O_p(\sqrt{d})$. Finally, the posterior density of λ over $\Lambda = \sqrt{n}(\Theta - \theta_0) - s$ is given by $f(\lambda) = \frac{\ell(\lambda)}{\int_\Lambda \ell(\lambda)d\lambda}$, where, for $\bar{x} = \sum_{i=1}^n x_i/n$,

$$\ell(\lambda) = \exp\left(\bar{x}'(\sqrt{n}(\lambda + s)) + n\left(\psi\left(\theta_0 + \frac{\lambda + s}{\sqrt{n}}\right) - \psi(\theta_0)\right)\right) \cdot \pi\left(\theta_0 + \frac{\lambda + s}{\sqrt{n}}\right).$$

We impose the following regularity conditions, following Ghosal [18] and Portnoy [37]:

E2. Consider the following quantities associated with higher moments in a neighborhood of the true parameter θ_0:

$$B_{1n}(c) := \sup_{\theta,a}\{E_\theta|a'(x_i-\mu)|^3 : a \in \mathbb{R}^d, \|a\| = 1, \|\theta-\theta_0\|^2 \leq cd/n\},$$
$$B_{2n}(c) := \sup_{\theta,a}\{E_\theta|a'(x_i-\mu)|^4 : a \in \mathbb{R}^d, \|a\| = 1, \|\theta-\theta_0\|^2 \leq cd/n\}.$$

For any $c > 0$ and all n there are $p > 0$ and $c_0 > 0$ such that

$$B_{1n}(c) < c_0 + c^p \text{ and } B_{2n}(c) < c_0 + c^p.$$

E3. The prior density π is proper and satisfies a positivity requirement at the true parameter

$$\sup_{\theta\in\Theta} \ln\left[\pi(\theta)/\pi(\theta_0)\right] = O(d)$$

where θ_0 is the true parameter. Moreover, the prior π also satisfies the following local Lipschitz condition

$$|\ln\pi(\theta) - \ln\pi(\theta_0)| \leq V(c)\sqrt{d}\|\theta-\theta_0\|$$

for all θ such that $\|\theta-\theta_0\|^2 \leq cd/n$, and some $V(c)$ such that $V(c) < c_0 + c^p$, the latter holding for all $c > 0$.

E4 The following condition on the growth rate of the dimension of the parameter space is assumed to hold:

$$d^3/n \to 0.$$

Comment 5.1 *Condition E2 strengthens an analogous assumption of Ghosal [18]. Both assumption are implied by the analogous assumption made by Portnoy [37]. Condition E3 is similar to the assumption on the prior in Ghosal [18]. For further discussion of this assumption, see [3]. Condition E4 states that the parameter dimension should not grow too quickly relative to the sample size.*

Theorem 6 *Conditions E1-E4 imply conditions C1-C3.*

Proof. See Appendix A. ■

Combining Theorems 1 and 6, we have the asymptotic normality of the posterior,

$$\int_\Lambda |f(\lambda)-\phi(\lambda)|d\lambda = \int_K |f(\lambda)-\phi(\lambda)|d\lambda + o_p(1) = o_p(1).$$

Furthermore, we can apply Theorem 3 to the posterior density f to bound the convergence time (number of steps) of the Metropolis walk needed to obtain a draw from f (with a fixed level of accuracy): The convergence time is at most

$$O_p(d^2)$$

after the burn-in period; together with the burn-in, the convergence time is

$$O_p(d^3 \ln d).$$

Finally, the integration bounds stated in the previous section also apply to the posterior f.

5.2. *Non-Concave and Discontinuous Cases.* Next we consider the case of a d-dimensional curved exponential family. Being a generalization of the canonical exponential family, its analysis has many similarities with the previous example. Nonetheless, it is general enough to allow for non-concavities and even various kinds of non-smoothness in the log-likelihood function.

NE1. Let $x_1, \ldots, x_n$ be iid observations from a d-dimensional curved exponential family with density

$$f(x;\theta) = \exp\left(x'\theta(\eta) - \psi_n(\theta(\eta))\right),$$

where $\theta \in \Theta$, an open subset of $\mathbb{R}^d$, and $d \to \infty$ as $n \to \infty$.
NE2. The parameter of interest is η, whose true value η_0 lies in the interior of a convex compact set $\Psi \subset \mathbb{R}^{d_1}$. The true value of θ induced by η_0 is given by $\theta_0 = \theta(\eta_0)$. The mapping $\eta \mapsto \theta(\eta)$ takes values from $\mathbb{R}^{d_1}$ to $\mathbb{R}^d$ where $c \cdot d \leq d_1 \leq d$, for some $c > 0$. Moreover, assume that η_0 is the unique solution to the system $\theta(\eta) = \theta_0$ and that $\|\theta(\eta) - \eta(\theta_0)\| \geq \epsilon_0 \|\eta - \eta_0\|$ for some $\epsilon_0 > 0$ and all $\eta \in \Psi$.

Thus, the parameter θ corresponds to a high-dimensional linear parametrization of the log-density, and η describes the lower-dimensional parametrization of the density of interest. There are many classical examples of curved exponential families; see for example Efron [12], Lehmann and Casella [29], and Bandorff-Nielsen [2]. An example of the condition that puts a curved structure onto an exponential family is a moment restriction of the type:

$$\int m(x,\alpha) f(x,\theta) dx = 0.$$

This condition restricts θ to lie on a curve that can be parameterized as $\{\theta(\eta), \eta \in \Psi\}$, where component $\eta = (\alpha, \beta)$ contains α as well as other

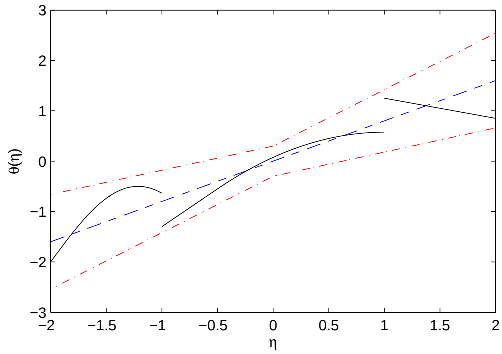

FIG 2. *This figure illustrates the mapping $\theta(\cdot)$. The (discontinuous) solid line is the mapping while the dash line represents the linear map induced by G. The dash-dot line represents the deviation band controlled by r_{1n} and r_{2n}.*

parameters β. In econometric applications, often moment restrictions represent Euler equations that result from the data x being an outcome of an optimization by rational decision-makers; see e.g. Hansen and Singleton [19], Chamberlain [7], Imbens [21], and Donald, Imbens and Newey [10]. Thus, the curved exponential framework is a fundamental complement of the exponential framework, at least in certain fields of data analysis.

We require the following additional regularity conditions on the mapping $\theta(\cdot)$.

NE3. For every κ, and uniformly in $\gamma \in B(0, \kappa\sqrt{d})$, there exists a linear operator $G : \mathbb{R}^{d_1} \to \mathbb{R}^d$ such that $G'G$ has eigenvalues bounded from above and away from zero, and for every n

$$\sqrt{n}\,(\theta(\eta_0 + \gamma/\sqrt{n}) - \theta(\eta_0)) = r_{1n} + (I_d + R_{2n})G\gamma,$$

where $\|r_{1n}\| \leq \delta_{1n}$ and $\|R_{2n}\| \leq \delta_{2n}$. Moreover, those coefficients are such that

$$\delta_{1n}\sqrt{d} \to 0 \quad \text{and} \quad \delta_{2n}d \to 0.$$

Thus the mapping $\eta \mapsto \theta(\eta)$ is allowed to be nonlinear and discontinuous. For example, the additional condition of $\delta_{1n} = 0$ implies the continuity of the mapping in a neighborhood of η_0. More generally, condition NE3 does impose that the map admits an approximate linearization in the neighborhood of η_0 whose quality is controlled by the errors δ_{1n} and δ_{2n}. An example of a kind of map allowed in this framework is given in the figure.

Again, given a prior π on Θ, the posterior of η given the data is denoted by

$$\pi_n(\eta) \propto \pi(\theta(\eta)) \cdot \prod_{i=1}^{n} f(x_i;\eta) = \pi(\theta(\eta)) \cdot \exp\left(n\bar{x}'\theta(\eta) - n\psi(\theta(\eta))\right).$$

In this framework, we also define the local parameters to describe contiguous deviations from the true parameter as

$$\gamma = \sqrt{n}(\eta - \eta_0) - s, \quad s = (G'G)^{-1}G'\sqrt{n}(\bar{x} - \mu),$$

where s is a first order approximation to the normalized maximum likelihood/extremum estimate. Again, similar bounds hold for s: $E[s] = 0$, $E[ss'] = (G'G)^{-1}$, and $\|s\| = O_p(\sqrt{d})$. The posterior density of γ over Γ, where $\Gamma = \sqrt{n}(\Psi - \eta_0) - s$, is $f(\gamma) = \frac{\ell(\gamma)}{\int_\Gamma \ell(\gamma)d\gamma}$, where

$$\begin{aligned} \ell(\gamma) &= \exp\left(n\bar{x}'(\theta(\eta_0 + (\gamma + s)/\sqrt{n}) - \theta(\eta_0))\right) \\ &\quad \times \exp\left(n\psi(\theta(\eta_0 + (\gamma + s)/\sqrt{n})) - n\psi(\theta(\eta_0))\right) \\ &\quad \times \pi(\theta(\eta_0 + (\gamma + s)/\sqrt{n})). \end{aligned} \tag{5.22}$$

The condition on the prior is the following:

NE4 The prior $\pi(\eta) \propto \pi(\theta(\eta))$, where $\pi(\theta)$ satisfies condition E3.

Theorem 7 *Conditions E2-E4 and NE1-NE4 imply conditions C1-C3.*

Proof. See Appendix A. ∎

As before, Theorems 1 and 7 prove the asymptotic normality of the posterior,

$$\int_\Lambda |f(\gamma) - \phi(\gamma)|d\gamma = \int_K |f(\gamma) - \phi(\gamma)|d\gamma + o_p(1) = o_p(1),$$

where

$$\phi(\gamma) = \frac{1}{(2\pi)^{d/2}\det\left((G'G)^{-1}\right)^{1/2}} \exp\left(-\frac{1}{2}\gamma'(G'G)\gamma\right).$$

Theorem 3 implies further that the main results of the paper on the polynomial time sampling and integration apply to this curved exponential family.

6. Conclusion. This paper studies the computational complexity of Bayesian and quasi-Bayesian estimation in large samples carried out using a basic Metropolis random walk. Our framework permits the parameter dimension of the problem to grow to infinity and allows the underlying log-likelihood or extremum criterion function to be discontinuous and/or non-concave. We establish polynomial complexity by exploiting a central limit theorem framework which provides structural restrictions for the problem, i.e., the posterior or quasi-posterior density approaches a normal density in large samples.

The analysis of this paper focused on a basic random walk. Although it is widely used for its simplicity, it is not the most sophisticated algorithm available. Thus, in principle further improvements could be obtained by considering different kinds of random walks (or variance reduction schemes). As mentioned before, essentially only one lemma of our analysis relies on the particular choice of the random walk. This suggests that most of the analysis is applicable to a variety of different implementations.

Acknowledgement. We would like to thank Ron Gallant, Lars Hansen, Jerry Hausman, James Heckman, Nick Polson, Santosh Vempala and participants of seminars at the University of Chicago, MIT, Duke, and the INFORMS Conference, for useful comments. We also thank Raymond Guiteras, Alp Simsek, and Theophane Weber for thorough readings of the paper and their valuable comments that have considerably improved this paper.

APPENDIX A: PROOFS OF OTHER RESULTS

Proof of Theorem 1. From **C1** it follows that

$$\begin{aligned}\int_\Lambda |f(\lambda)-\phi(\lambda)|d\lambda &\leq \int_K |f(\lambda)-\phi(\lambda)|d\lambda + \int_{K^c}\left(f(\lambda)+\phi(\lambda)\right)d\lambda \\ &= \int_K |f(\lambda)-\phi(\lambda)|d\lambda + o_p(1)\end{aligned}$$

where the last equality follows from Assumption C1.[6]

Now, denote $C_n = \dfrac{(2\pi)^{d/2}\det\left(J^{-1}\right)^{1/2}}{\int_K \ell(\omega)d\omega}$ and write

$$\int_K \left|\frac{f(\lambda)}{\phi(\lambda)}-1\right|\phi(\lambda)d\lambda \;=\; \int_K \left|C_n\cdot\exp\left(\ln\ell(\lambda)-\left(-\frac{1}{2}\lambda' J\lambda\right)\right)-1\right|\phi(\lambda)d\lambda$$

[6]For the case of ϕ, it follows from the standard concentration of measure arguments for Gaussian densities, see Lovász and Vempala [33].

Combining the expansion in **C2** with conditions imposed in **C3**,

$$\begin{aligned}
\int_{\Lambda}\left|\frac{f(\lambda)}{\phi(\lambda)}-1\right|\phi(\lambda)d\lambda &\leq \textstyle\int_K \left|C_n\cdot\exp\left(\epsilon_1+\epsilon_2\lambda' J\lambda\right)-1\right|\phi(\lambda)d\lambda \\
&\quad +\textstyle\int_K \left|C_n\cdot\exp\left(-\epsilon_1-\epsilon_2\lambda' J\lambda\right)-1\right|\phi(\lambda)d\lambda \\
&\leq 2\int_K\left|C_n\cdot e^{o_p(1)}-1\right|\phi(\lambda)d\lambda \\
&\leq 2|C_n e^{o_p(1)}-1|
\end{aligned}$$

The proof then follows by showing that $C_n \to 1$. We have that $R = \|K\| = O(\sqrt{d})$, and by assumption C1

$$\begin{aligned}
\frac{1}{C_n} &\geq \frac{\displaystyle\int_{\|\lambda\|\leq R}\ell(\lambda)d\lambda}{(1+o(1))\displaystyle\int_{\|\lambda\|\leq R}g(\lambda)d\lambda} \geq \frac{\displaystyle\int_{\|\lambda\|\leq R}e^{-\frac{1}{2}\lambda' J\lambda}e^{-\epsilon_1-\frac{\epsilon_2}{2}(\lambda' J\lambda)}d\lambda}{(1+o(1))\displaystyle\int_{\|\lambda\|\leq R}e^{-\frac{1}{2}\lambda' J\lambda}d\lambda} \\
&= \frac{e^{-2\epsilon_1}}{(1+o(1))}\sqrt{\frac{\det(J)}{\det(J+\epsilon_2 J)}}\;\frac{\displaystyle\int_{\|\lambda\|\leq R}\frac{e^{-\frac{1}{2}\lambda'(J+\epsilon_2 J)\lambda}}{(2\pi)^{d/2}\det((J+\epsilon_2 J)^{-1})^{1/2}}d\lambda}{\displaystyle\int_{\|\lambda\|\leq R}\frac{e^{-\frac{1}{2}\lambda' J\lambda}}{(2\pi)^{d/2}\det(J^{-1})^{1/2}}d\lambda}
\end{aligned}$$

Since $\epsilon_2 < 1/2$, we can define $W \sim N(0,(1+\epsilon_2)^{-1}J^{-1})$ and $V \sim N(0,J^{-1})$ and rewrite our bounds as

$$\begin{aligned}
\frac{1}{1+o(1)}\frac{\int_{\|\lambda\|\leq R}\ell(\lambda)d\lambda}{\int_{\|\lambda\|\leq R}g(\lambda)d\lambda} &\geq \frac{e^{-2\epsilon_1}}{(1+o(1))}\left(\frac{1}{1+\epsilon_2}\right)^{d/2}\frac{P(\|W\|\leq R)}{P(\|V\|\leq R)} \\
&\geq \frac{e^{-2\epsilon_1}}{(1+o(1))}\left(\frac{1}{1+\epsilon_2}\right)^{d/2}
\end{aligned}$$

where the last inequality follows from $P(\|W\| \leq R) \geq P(\|\sqrt{1+\epsilon_2}W\| \leq R) = P(\|V\| \leq R)$. Likewise,

$$\frac{1}{C_n} \leq \frac{\int_{\|\lambda\|\leq R}\ell(\lambda)d\lambda}{\int_{\|\lambda\|\leq R}g(\lambda)d\lambda} \leq e^{2\epsilon_1}\left(\frac{1}{1-\epsilon_2}\right)^{d/2}$$

Therefore $C_n \to 1$ since $\epsilon_1 \to 0$, $\epsilon_2 \cdot d \to 0$. ■

Proof of Lemma 2. The result follows immediately from equations (2.7)-(2.8). ■

Proof of Lemma 3. Let $M := \beta\frac{2te^{-t^2/4}}{\sqrt{\pi}}$. We will prove the lemma by contradiction. Assume that there exists a partition of $K = S_1 \cup S_2 \cup S_3$, with $d(S_1, S_2) \geq t$ such that

$$\int \left(M 1_{S_i}(x) - 1_{S_3}(x)\right) f(x)dx > 0, \text{ for } i = 1, 2.$$

We will use the Localization Lemma of Kannan, Lovász, and Simonovits [25] in order to reduce a high-dimensional integral to a low-dimensional integral.

Lemma 6 (Localization Lemma) *Let g and h be two lower semi-continuous Lebesgue integrable functions on $\mathbb{R}^d$ such that*

$$\int_{\mathbb{R}^d} g(x)dx > 0 \quad \textit{and} \quad \int_{\mathbb{R}^d} h(x)dx > 0.$$

Then there exist two points $a, b \in \mathbb{R}^d$, and a linear function $\tilde{\gamma} : [0,1] \to \mathbb{R}_+$ such that

$$\int_0^1 \tilde{\gamma}^{d-1}(t) g((1-t)a + tb)dt > 0 \quad \textit{and} \quad \int_0^1 \tilde{\gamma}^{d-1}(t) h((1-t)a + tb)dt > 0,$$

where $([a,b], \tilde{\gamma})$ is said to form a needle.

Proof. See Kannan, Lovász, and Simonovits [25]. ■

By the Localization Lemma, there exists a needle $(a, b, \tilde{\gamma})$ such that

$$\int_0^1 \tilde{\gamma}^{d-1}(u) f((1-u)a+ub) \left(M 1_{S_i}((1-u)a+ub) - 1_{S_3}((1-u)a+ub) \right) du > 0,$$

for $i = 1, 2$. Equivalently, using $\gamma(u) = \tilde{\gamma}(u/\|b-a\|)$ and $v := (b-a)/\|b-a\|$ where $\|b-a\| \geq t$, we have

$$\int_0^{\|b-a\|} \gamma^{d-1}(u) f(a+uv) \left(M 1_{S_i}(a+uv) - 1_{S_3}(a+uv) \right) du > 0,$$

for $i = 1, 2$. In turn, this last expression can be rewritten as, for $i = 1, 2$,

$$M \int_0^{\|b-a\|} \gamma^{d-1}(u) f(a+uv) 1_{S_i}(a+uv) du > \int \gamma^{d-1}(u) 1_{S_3}(a+uv) f(a+uv) du. \tag{A.23}$$

In order for the left hand side of (A.23) be positive for $i = 1$ and $i = 2$, the line segment $[a, b]$ must contain points in S_1 and S_2. Since $d(S_1, S_2) \geq t$, we have that $S_3 \cap [a, b]$ contains an interval whose length is at least t. We will prove that for every $w \in \mathbb{R}$

$$\begin{aligned} \int_w^{w+t} \gamma^{d-1}(u) f(a+uv) du \quad &\geq M \min \left\{ \int_0^w \gamma^{d-1}(u) f(a+uv) du, \right. \\ &\left. \int_{w+t}^{\|b-a\|} \gamma^{d-1}(u) f(a+uv) du \right\} \end{aligned} \tag{A.24}$$

which contradicts relation (A.23) and proves the lemma.

First, note that $f(a+uv) = e^{-\|a+uv\|^2} m(a+uv) = e^{-u^2 + r_1 u + r_0} m(a+uv)$ where $r_1 := 2a'v$ and $r_0 := -\|a\|^2$.

Next, recall that $m(a+uv)\gamma^{d-1}(u)$ is still a unidimensional log-β-concave function on u. By Lemma 7 presented in Appendix B, there exists a unidimensional logconcave function $\widehat{m}$ such that $\beta\widehat{m}(u) \leq m(a+uv)\gamma^{d-1}(u) \leq \widehat{m}(u)$ for every u. Moreover, there exists numbers s_0 and s_1 such that $\widehat{m}(w) = s_0 e^{s_1 w}$ and $\widehat{m}(w+t) = s_0 e^{s_1(w+t)}$. Due to the log-concavity of $\widehat{m}$, this implies that

$$\widehat{m}(u) \geq s_0 e^{s_1 u} \text{ for } u \in (w, w+t) \text{ and } \widehat{m}(u) \leq s_0 e^{s_1 u} \text{ otherwise.}$$

Thus, we can replace $m(a+uv)\gamma^{d-1}(u)$ by $s_0 e^{s_1 u}$ on the right hand side of (A.24) and replace $m(a+uv)\gamma^{d-1}(u)$ by $\beta s_0 e^{s_1 u}$ on the left hand side of (A.24). After defining $\widehat{r}_1 = r_1 + s_1$ and $\widehat{r}_0 := r_0 + \ln s_0$, we have

$$\beta \int_w^{w+t} e^{-u^2+\widehat{r}_1 u+\widehat{r}_0} du \geq M \min\left\{ \int_0^w e^{-u^2+\widehat{r}_1 u+\widehat{r}_0} du, \int_{w+t}^{\|b-a\|} e^{-u^2+\widehat{r}_1 u+\widehat{r}_0} du \right\}$$

which is equivalent to

$$\text{(A.25)} \quad \begin{aligned} \beta \int_w^{w+t} e^{-(u-\frac{\widehat{r}_1}{2})^2+\widehat{r}_0+\frac{\widehat{r}_1^2}{4}} du \quad &\geq M \min\left\{ \int_0^w e^{-(u-\frac{\widehat{r}_1}{2})^2+\widehat{r}_0+\frac{\widehat{r}_1^2}{4}} du, \right. \\ &\left. \int_{w+t}^{\|b-a\|} e^{-(u-\frac{\widehat{r}_1}{2})^2+\widehat{r}_0+\frac{\widehat{r}_1^2}{4}} du \right\}. \end{aligned}$$

Now, cancel the term $e^{\widehat{r}_0+\widehat{r}_1^2/4}$ on both sides and; since we want the inequality (A.25) holding for any w, (A.25) is implied by

$$\text{(A.26)} \qquad \int_w^{w+t} e^{-u^2} du \geq \frac{2te^{-t^2/4}}{\sqrt{\pi}} \min\left\{ \int_{-\infty}^w e^{-u^2} du, \int_{w+t}^{\infty} e^{-u^2} du \right\}$$

holding for any w. This inequality is Lemma 2.2 in Kannan and Li [24]. For brevity, we will not reproduce the proof. ■

Proof of Corollary 2. Consider the change of variables $\tilde{x} = \frac{J^{1/2}x}{\sqrt{2}}$. Then, in $\tilde{x}$ coordinates, $f(\tilde{x}) = e^{\tilde{x}'\tilde{x}} m(\sqrt{2}J^{-1/2}\tilde{x})$ satisfies the assumption of Lemma 3 and $d(S_1, S_2) \geq t\sqrt{\lambda_{min}/2}$. The result follows by applying Lemma 3 with $\tilde{x}$ coordinates. ■

Proof of Lemma 5. Define $K := B(0, R)$, so that R is the radius of K; also let $r := 4\sqrt{d}\sigma$ (where $\sigma^2 \leq \frac{1}{16dL^2}$), and let $q(x|u)$ denote the normal density function centered at u with covariance matrix $\sigma^2 I$. We use the following notation: $B_u = B(u, r)$, $B_v = B(v, r)$, and $A_{u,v} = B_u \cap B_v \cap K$. By definition of r, we have that $\int_{B_u} q(x|u)dx = \int_{B_v} q(x|v)dx > 1 - \frac{1}{e^3}$.

Define the direction $w = (v-u)/\|v-u\|$. Let $H_1 = \{x \in B_u \cap B_v : w'(x-u) \geq \|v-u\|/2\}$, $H_2 = \{x \in B_u \cap B_v : w'(x-u) \leq \|v-u\|/2\}$.

Consider the one-step distributions from u and v. We have that

$$\begin{aligned}
\|P_u - P_v\|_{TV} &\leq 1 - \int_{A_{u,v}} \min\{dP_u, dP_v\} \\
&= 1 - \int_{A_{u,v}} \min\left\{ q(x|u) \min\left\{\frac{f(x)}{f(u)}, 1\right\}, q(x|v) \min\left\{\frac{f(x)}{f(v)}, 1\right\}\right\} dx \\
&\leq 1 - \beta e^{-Lr} \int_{A_{u,v}} \min\{q(x|u), q(x|v)\} dx \\
&\leq 1 - \beta e^{-Lr} \left(\int_{H_1 \cap K} q(x|u)dx + \int_{H_2 \cap K} q(x|v)dx \right)
\end{aligned}$$

where $\|u - v\| < \sigma/8$. Next we will bound from below the last sum of integrals for an arbitrary $u \in K$.

We first bound the integrals over the possibly larger sets, respectively H_1 and H_2. Let h denote the density function of a univariate random variable distributed as $N(0, \sigma^2)$. It is easy to see that $h(t) = \int_{w'(x-u)=t} q(x|u)dx$, i.e. h is the marginal density of $q(.|u)$ along the direction w (up to a translation). Let $H_3 = \{x : -\|u - v\|/2 < w'(x - u) < \|v - u\|/2\}$. Note that $B_u \subset H_1 \cup (H_2 - \|u - v\|w) \cup H_3$ where the union is disjoint. Armed with these observations, we have

$$\begin{aligned}
\int_{H_1} q(x|u)dx + \int_{H_2} q(x|v)dx &= \int_{H_1} q(x|u)dx + \int_{H_2 - \|u-v\|w} q(x|u)dx \\
&\geq \int_{B_u} q(x|u)dx - \int_{H_3} q(x|u)dx \\
&= \int_{B_u} q(x|u)dx - \int_{-\|u-v\|/2}^{\|u-v\|/2} h(t)dt \\
&\geq 1 - \frac{1}{e^3} - \int_{-\|u-v\|/2}^{\|u-v\|/2} \frac{e^{-t^2/2\sigma^2}}{\sqrt{2\pi}\sigma} dt \\
&\geq 1 - \frac{1}{e^3} - \|u - v\| \frac{1}{\sqrt{2\pi}\sigma} \\
&\geq 1 - \frac{1}{e^3} - \frac{1}{8\sqrt{2\pi}} \geq \frac{9}{10}, \qquad \text{(A.27)}
\end{aligned}$$

where we used that $\|u - v\| < \sigma/8$ by the hypothesis of the lemma.

In order to take the support K into account, we can assume that $u, v \in \partial K$, i.e. $\|u\| = \|v\| = R$ (otherwise the integral will be larger). Let $z = (v + u)/2$ and define the half space $H_z = \{x : z'x \leq z'z\}$ whose boundary passes through u and v (Using $\|u\| = \|v\| = R$ it follows that $z'v = z'u = z'z/2$).

By the symmetry of the normal density, we have

$$\int_{H_1 \cap H_z} q(x|u)dx = \frac{1}{2} \int_{H_1} q(x|u)dx.$$

Although $H_1 \cap H_z$ does not lie in K in general, simple arithmetic shows that $H_1 \cap \left(H_z - \frac{r^2 z}{R\|z\|}\right) \subseteq K$.[7]

Using that $\int_{H_z \setminus (H_z - \frac{r^2 z}{R\|z\|})} q(x|u) = \int_0^{r^2/R} h(t)dt$, we have

$$\begin{aligned}\int_{H_1\cap K} q(x|u)dx &\geq \int_{H_1\cap\left(H_z - \frac{r^2 z}{R\|z\|}\right)} q(x|u)dx \geq \int_{H_1\cap H_z} q(x|u)dx - \int_0^{r^2/R} h(t)dt\\ &\geq \frac{1}{2}\int_{H_1} q(x|u)dx - \int_0^{r^2/R} \frac{e^{-t^2/2\sigma^2}}{\sqrt{2\pi}\sigma}dt\\ &\geq \frac{1}{2}\int_{H_1} q(x|u)dx - 4\sqrt{d}\sigma\frac{1}{30\sqrt{d}}\frac{1}{\sqrt{2\pi}\sigma},\end{aligned}$$

where we used that $\frac{r}{R} < \frac{1}{30\sqrt{d}}$ since $r = 4\sqrt{d}\sigma$ and $\frac{\sigma}{R} < \frac{1}{120d}$.

By symmetry, the same inequality holds when u and H_1 are replaced by v and H_2 respectively. Adding these inequalities and using (A.27), we have

$$\left(\int_{H_1\cap K} q(x|u)dx + \int_{H_2\cap K} q(x|v)dx\right) \geq \frac{9}{20} - \frac{4}{15\sqrt{2\pi}} \geq 1/3.$$

Thus, we have

$$\|P_u - P_v\| < 1 - \frac{\beta}{3}e^{-Lr}$$

and the result follows since $Lr \leq 1$. ■

Proof of Lemma 1. Starting from an arbitrary point in K, assume that the random walk makes a proper move. If this is the case note that

$$\begin{aligned}\max_{A:Q(A)>0, A\in\mathcal{A}} \frac{P(A)}{Q(A)} &\leq \max_{x\in K} \frac{e^{-\frac{1}{2\sigma^2}\|x\|^2}}{(2\pi)^{d/2}\sigma^d}\ (2\pi)^{d/2}\det(J^{-1})e^{\frac{1}{2}x'Jx}e^{2\epsilon_1 + 2\epsilon_2 x'Jx}\\ &\leq O\left((4\sqrt{d}\lambda_{max}\|K\|/\lambda_{min})^d e^{\lambda_{max}\|K\|^2 + 2\epsilon_1 + 2\epsilon_2\|K\|^2\lambda_{max}}\right).\end{aligned}$$

The result follows by invoking the CLT restrictions.

Next we show that the probability p of making a proper move is at least a positive constant. We will use the notation defined in the proof of Lemma 5. Let u be an arbitrary point in K. We have that

$$\begin{aligned} p &= \int_K \min\left\{\frac{f(x)}{f(u)},\ 1\right\} q(x|u)dx \geq \beta e^{-Lr}\int_{B_u\cap K} q(x|u)dx\\ &\geq \beta e^{-Lr}\int_{B_u\cap H_u} q(x|u)dx - \int_0^{r^2/R} h(t)dt \geq \frac{1}{3}.\end{aligned}$$

[7]Indeed, take $y \in H_1 \cap \left(H_z - \frac{r^2}{R}\frac{z}{\|z\|}\right)$. We can write $y = \frac{z}{\|z\|}\left(\frac{y'z}{\|z\|}\right) + s$, where $\|s\| \leq r$ (since $\left\|y - \frac{z}{\|z\|}\left(\frac{y'z}{\|z\|}\right)\right\| \leq \|y - z\| = \|y - \frac{u+v}{2}\| \leq \frac{1}{2}\|y-u\| + \frac{1}{2}\|y-v\| \leq r$) and s is also orthogonal to z. Since $y \in \left(H_z - \frac{r^2}{R}\frac{z}{\|z\|}\right)$, we have $\frac{y'z}{\|z\|} \leq \frac{z'z}{\|z\|} - \frac{r^2}{R} = \|z\| - \frac{r^2}{R} \leq R - \frac{r^2}{R}$. Therefore, $\|y\| = \sqrt{\left(\frac{y'z}{\|z\|}\right)^2 + \|s\|^2} \leq \sqrt{(R - \frac{r^2}{R})^2 + r^2} = \sqrt{R^2 - 2R\frac{r^2}{R} + \frac{r^4}{R^2} + r^2} \leq R$.

■

Proof of Theorem 5. Consider the sample mean defined by

$$\widehat{\mu}_{B,N} = \frac{1}{N}\sum_{i=1}^{N} g(\lambda^{i,B})$$

with the underlying sequence $(\lambda^{1,B}, \lambda^{2,B}, ..., \lambda^{N,B})$ produced by one of the schemes (lr, ss, ms) as follows:

- for lr, $\lambda^{i,B} = \lambda^{i+B}$, where λ^{i+B} is produced by iterating the chain $B+i$ times starting with an initial draw from f_0. Define the density after B steps of the chain starting with f_0 by $T^B f_0$. Thus λ^B has the distribution $T^B f_0$, and λ^{i+B} has the distribution $T^{i+B} f_0$.
- for ss, $\lambda^{i,B} = \lambda^{iS+B}$, where S is the number of draws that are "skipped".
- for ms, $\lambda^{i,B}$ are i.i.d. draws from $T^B f_0$, i.e. each i-th draw is obtained by sampling an initial point from f_0 and iterating the chain B times.

We have that

$$\begin{aligned}
MSE(\widehat{\mu}_{B,N}) &= E_{T^B f_0}\left[MSE(\widehat{\mu}_{B,N}|\lambda^B = \lambda)\right] \\
&= E_f\left[MSE(\widehat{\mu}_{B,N}|\lambda^B = \lambda)\frac{T^B f_0(\lambda)}{f(\lambda)}\right] \\
&= E_f\left[MSE(\widehat{\mu}_{B,N}|\lambda^B = \lambda)\right] + \\
&\quad +E_f\left[MSE(\widehat{\mu}_{B,N}|\lambda^B = \lambda)\left(\tfrac{T^B f_0(\lambda)}{f(\lambda)} - 1\right)\right] \\
&\leq E_f\left[MSE(\widehat{\mu}_{B,N}|\lambda^B = \lambda)\right] + \bar{g}^2 E_f\left[\left|\frac{T^B f_0(\lambda)}{f(\lambda)} - 1\right|\right] \\
&= (\sigma^2_{g,N}/N) + 2\bar{g}^2\|T^B f_0 - f\|_{TV},
\end{aligned}$$

where $\sigma^2_{g,N}$ is the variance of the sample average under the assumption that λ^B is distributed exactly according to f. (We also used the fact that $\|T^B f_0 - f\|_{TV} = \frac{1}{2}\|T^B f_0 - f\|_{L_1}$.)

The bound on $\sigma^2_{g,N}$ will depend on the particular scheme, as discussed below.

We require that the second term is smaller than $\varepsilon/3$, which is equivalent to imposing that $\|T^B f_0 - f\|_{TV} < \frac{\varepsilon}{6\bar{g}^2}$. Using Theorem 2, since f_0 is a M-warm start for f,

$$\begin{aligned}
\sqrt{M}\left(1 - \tfrac{\phi^2}{2}\right)^B &< \sqrt{M}e^{-B\frac{\phi^2}{2}} < \frac{\varepsilon}{6\bar{g}^2} \\
-B\tfrac{\phi^2}{2} &< \ln\left(\frac{\varepsilon}{6\sqrt{M}\bar{g}^2}\right) \\
B &\geq \left(\frac{2}{\phi^2}\right)\ln\left(\frac{6\sqrt{M}\bar{g}^2}{\varepsilon}\right).
\end{aligned}$$

Next we bound $\sigma^2_{g,N}$. Specifically, we determine the number of post-burn iterations N_{lr}, N_{ss}, or N_{ms} needed to set the overall mean square error less than ε.

To bound N_{lr}, note that $\sigma^2_{g,N} \leq \sigma^2_g \leq \gamma_0 \frac{4}{\phi^2}$ where the last inequality follows from Corollary 2. Thus, $N_{lr} = \frac{\gamma_0}{\varepsilon}\frac{6}{\phi^2}$ and B set above suffice to obtain $MSE(\widehat{\mu}_{B,N}) \leq \varepsilon$.

To bound N_{ss}, we first must choose a spacing S to ensure that the autocovariances γ_k are sufficiently small. We start by bounding $\sigma^2_{g,N}$,

$$\sigma^2_{g,N} \leq \gamma_0 + 2N\gamma_1 \leq \gamma_0 + 2N\gamma_0\left(1 - \frac{\phi^2}{2}\right)^S$$

where we used Corollary 2 and that $\lambda^{i,B}$ and $\lambda^{i+1,B}$ are spaced by S steps of the chain. By choosing the spacing S as

$$\left(1 - \frac{\phi^2}{2}\right)^S \leq e^{-S\frac{\phi^2}{2}} \leq \frac{\varepsilon}{6\gamma_0}, \text{ i.e. } S \geq \frac{2}{\phi^2}\ln\left(\frac{6\gamma_0}{\varepsilon}\right),$$

and using $N_{ss} = \dfrac{3\gamma_0}{\varepsilon}$, the mean square error for the ss method can be bounded as

$$\begin{aligned}
MSE(\widehat{\mu}_{B,N}) &\leq \frac{1}{N_{ss}}(\gamma_0 + 2N_{ss}\gamma_1) + 2\bar{g}^2\|T^B f_0 - f\|_{TV} \\
&\leq \frac{\varepsilon}{3\gamma_0}\left(\gamma_0 + 2\frac{3\gamma_0}{\varepsilon}\gamma_0\frac{\varepsilon}{6\gamma_0}\right) + \bar{g}^2\frac{\varepsilon}{3\bar{g}^2} \\
&\leq \frac{\varepsilon}{3\gamma_0}(\gamma_0 + \gamma_0) + \frac{\varepsilon}{3} \leq \varepsilon
\end{aligned}$$

To bound N_{ms}, we observed that $\gamma_k = 0$ for all $k \neq 0$ implying that $MSE(\widehat{\mu}_{B,N}) \leq \frac{\gamma_0}{N_{ms}} + \varepsilon/3 \leq \varepsilon$ provided that $N_{ms} \geq 2\gamma_0/(3\varepsilon)$. ■

Proof of Theorem 6. Given

$$K = B(0, \|K\|) \text{ where } \|K\|^2 = cd,$$

our condition C1 is satisfied by the argument given in proof of Ghosal's Lemma 4. Further, our condition C2 is satisfied by the argument given in the proof of Ghosal's Lemma 1 with $\epsilon_1 = 0$ and

$$\epsilon_2 = \frac{1}{3}\left(\sqrt{\frac{cd}{n}}B_{1n}(0) + \frac{cd}{n}B_{2n}(c)\right),$$

and our condition C3 is satisfied since by E3 and E4

$$\epsilon_2\|K\|^2 \to 0. ■$$

Comment A.1 *Ghosal [18] proves his results for the support set* $K' = B(0, C\sqrt{d}\log d)$. *His arguments actually go through for the support set* $K = B(0, C\sqrt{d})$ *due to the concentration of normal measure under* $d \to \infty$ *asymptotics. For details, see [3].*

Proof of Theorem 7. Take $K = B(0, \|K\|)$, where $\|K\|^2 = Cd_1$ for some C sufficiently large independent of d (see [3] for details). Then condition C1 is satisfied by the argument given in the proof of Ghosal's Lemma 4 and NE3. Further, condition C2 is satisfied by the argument given in the proof of Ghosal's Lemma 1 and NE3 with

$$\epsilon_1 = O_p\left(\delta_{1n}^2 + (1+\delta_{2n})\delta_{1n}\sqrt{d}\right),$$
$$\epsilon_2 = O_p\left(\delta_{2n} + \delta_{2n}^2 + \left(\sqrt{\frac{cd}{n}}B_{1n}(0) + \frac{Cd}{n}B_{2n}(C)\right)\right),$$

and condition C3 is satisfied since by E3, E4, NE3, and NE4,

$$\epsilon_2\|K\|^2 \to 0. \blacksquare$$

Comment A.2 *For further details and discussion, see [3].*

APPENDIX B: BOUNDING LOG-β-CONCAVE FUNCTIONS

Lemma 7 *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a unidimensional log-*β*-concave function. Then there exists a logconcave function* $g : \mathbb{R} \to \mathbb{R}$ *such that*

$$\beta g(x) \leq f(x) \leq g(x) \quad \textit{for every} \quad x \in \mathbb{R}.$$

Proof. Consider $h(x) = \ln f(x)$ a $(\ln\beta)$-concave function. Now, let m be the smallest concave function greater than $h(x)$ for every x, that is,

$$m(x) \quad = \quad \sup\left\{\sum_{i=1}^{k}\lambda_i h(y_i) : k \in \mathbf{N}, \lambda \in \mathbb{R}^k, \lambda \geq 0, \sum_{i=1}^{k}\lambda_i = 1, \sum_{i=1}^{k}\lambda_i y_i = x\right\}.$$

Recall that the epigraph of a function w is defined as $epi_w = \{(x,t) : t \leq w(x)\}$. Using our definitions, we have that $epi_m = \text{conv}(epi_h)$ (the convex hull of epi_h), where both sets lie in $\mathbb{R}^2$. In fact, the values of m are defined only by points in the boundary of $\text{conv}(epi_h)$. Consider $(x, m(x)) \in epi_m$, since the epigraph is convex and this point is on the boundary, there exists a supporting hyperplane H on $(x, m(x))$. Moreover, $(x, m(x)) \in \text{conv}(epi_h \cap H)$. Since H is one dimensional, $(x, m(x))$ can be written as convex combination of at most 2 points of epi_h.

Furthermore, by definition of log-β-concavity, we have that

$$\ln 1/\beta \geq \sup_{\lambda\in[0,1],y,z} \lambda h(y) + (1-\lambda)h(z) - h\left(\lambda y + (1-\lambda)z\right).$$

Thus, $h(x) \leq m(x) \leq h(x) + \ln(1/\beta)$. Exponentiating gives $f(x) \leq g(x) \leq \frac{1}{\beta}f(x)$, where $g(x) = e^{m(x)}$ is a logconcave function. ■

REFERENCES

[1] D. Applegate and R. Kannan (1993). Sampling and Integration of Near Logconcave Functions, Proceedings 23th ACM STOC, 156-163.

[2] O. Barndorff-Nielsen (1978). Information and exponential families in statistical theory. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester.

[3] A. Belloni and V. Chernozhukov (2007). Posterior Inference in Curved Exponential Families under Increasing Dimensions, IBM Research Report, http://web.mit.edu/belloni/www/curvexp.pdf.

[4] P. J. Bickel and J. A. Yahav (1969). Some contributions to the asymptotic theory of Bayes solutions, Z. Wahrsch. Verw. Geb 11, 257-276.

[5] Bunke, O., Milhaud, X. (1998) Asymptotic behavior of Bayes estimates under possibly incorrect models. The Annals of Statistics 26 (2), 617–644.

[6] G. Casella and C. P. Robert (1999). Monte Carlo Statistical Methods, Springer Texts in Statistics.

[7] G. Chamberlain (1987). Asymptotic efficiency in estimation with conditional moment restrictions, Journal of Econometrics 34, no. 3, 305–334

[8] V. Chernozhukov and H. Hong (2003). An MCMC approach to classical estimation, Journal of Econometrics 115 293-346.

[9] S. Chib (2001). Markov Chain Monte Carlo Methods: Computation and Inference, Handbook of Econometrics, Volume 5, by J.J. Heckman and E. Leamer, Elsvier Science.

[10] S. G. Donald, G. W. Imbens, W. K. Newey (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. Journal of Econometrics, 117, no. 1, 55–93.

[11] R. Dudley (2000). Uniform Cental Limit Theorems, Cambridge Studies in advanced mathematics.

[12] B. Efron (1978). The geometry of exponential families, Annals of Statistics, 6, no. 2, 362–376.

[13] R. F. Engle, S. Manganelli (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, Journal of Business & Economic Statistics, Vol. 22.

[14] G. S. Fishman (1994). Choosing sample path length and number of sample paths when starting at steady state, Operations Research Letters, Vol. 16, No. 4, November, pp. 209-220.

[15] A. Frieze, R. Kannan and N. Polson (1994). Sampling from log-concave functions, Annals of Applied Probability 4, pp. 812-834.

[16] J. Geweke and M. Keane (2001). Computationally Intensive Methods for Integration in Econometrics, Handbook of Econometrics, Volume 5, by J.J. Heckman and E. Leamer, Elsvier Science.

[17] C. J. Geyer (1992). Practical Markov Chain Monte Carlo, Statistical Science, Vol. 7, No. 4, 473-511.

[18] S. Ghosal (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity, Journal of Multivariate Analysis, vol 73, 49-68.

[19] L. P. Hansen and K. J. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models, Econometrica 50, no. 5, 1269–1286.

[20] I. Ibragimov and R. Has'minskii (1981). Statistical Estimation: Asymptotic Theory, Springer, Berlin.

[21] G. W. Imbens (1997). One-step estimators for over-identified generalized method of moments models. Review of Economics Studies 64, no. 3, 359–383.

[22] M. Jerrum and A. Sinclair (1988). Conductance and the rapid mixing property for Markov chains: the approximation of permanent resolved, Proceedings of the 20th Annual ACM symposium on Theory of Computing, pp. 235-244.

[23] M. Jerrum and A. Sinclair (1989). Approximating the permanent, SIAM Journal on Computing, 18:6, 1149-1178.

[24] R. Kannan and G. Li (1996). Sampling according to the multivariate normal density, 37th Annual Symposium on Foundations of Computer Science (FOCS '96), pp. 204.

[25] R. Kannan, L. Lovász, and M. Simonovits (1995). Isoperimetric Problems for Convex Bodies and a Localization Lemma, J. Discr. Comput. Goem., volume 13, pp. 541-559.

[26] R. Kannan, L. Lovász, and M. Simonovits (1997). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies, Random Structures and Algorithms, volume 11, pp. 1–50.

[27] C. Kipnis and S. R. S. Varadhan (1988). Central limit theorem for additive functionals of reversible processes and applications to simple exclusions, Comm. Math. Phys., 104, 1-19.

[28] R. Koenker and G. Bassett (1978).Regression Quantiles, Econometrica, Volume 46, No. 1, pp. 33-50

[29] E. L. Lehmann and G. Casella, Theory of point estimation. Second edition. Springer Texts in Statistics. Springer-Verlag, New York, 1998.

[30] J.S. Liu, L. Tian and L.J. Wei. Implementation of estimating-function based inference procedures with MCMC samplers. Journal of American Statistical Association, to appear.

[31] J.S. Liu (2001) Monte Carlo Strategies in Scientific Computing, Springer-Verlag, New York.

[32] L. Lovász and M. Simonovits (1993). Random Walks in Convex Bodies and an Improved Volume Algorithm, Random Structures and Algorithm, 4:4, 359-412.

[33] L. Lovász and S. Vempala, The Geometry of Logconcave Functions and Sampling Algorithms, To appear in Random Structures and Algorithms.

[34] L. Lovász and S. Vempala (2003). Hit-and-Run is Fast and Fun, Technical report MSR-TR-2003-05. Available at http://www-math.mit.edu/ vempala/papers/logcon-hitrun.ps.

[35] L. Lovász and S. Vempala (2004). Hit-and-Run from a Corner, Proc. of the 36th ACM Symp. on the Theory of Computation (STOC'04), pp. 310-314.Available at http://www-math.mit.edu/ vempala/papers/start.ps.

[36] N. Polson (1996). Convergence of Markov Chain Monte Carlo Algorithms, Bayesian Statistics 5, pp. 297-321 .

[37] S. Portnoy (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. Annals of Statistics 16, no. 1, 356–366.

[38] G. O. Roberts and R. L. Tweedie (1996). Exponential Convergence of Langevin Diffusions and Their Discrete Approximations Bernoulli, 2, pp. 341-364.

[39] C. J. Stone, M. H. Hansen, C. Kooperberg, and Y. K. Truong (1997). Polynomial splines and their tensor products in extended linear modeling. With discussion and a rejoinder by the authors and Jianhua Z. Huang. Annals of Statistics, 25, no. 4, 1371–1470.

[40] A. W. van der Vaart and J. A. Wellner (1996) Weak Convergence and Empirical Processes Spring Series in Statistics (1996).

[41] S. Vempala (2005). Geometric Random Walks: A Survey, Combinatorial and Computational Geometry, MSRI Publications Volume 52.

E-mail: belloni@mit.edu
1101 Kitchawan Road,
Building 800 Office 32-221,
Yorktown Heights, NY 10598
http://web.mit.edu/belloni/www

E-mail: vchern@mit.edu
50 Memorial Drive
Room E52-262f
Cambridge , MA 02142
http://web.mit.edu/vchern/www