Hamiltonian Monte Carlo for Bayesian Inference on Privatized Data

Arin Chang
Purdue University
chan1074@purdue.edu

Jordan Awan
University of Pittsburgh
jaa557@pitt.edu

Vinayak Rao
Purdue University
varao@purdue.edu

Abstract

Bayesian inference provides a robust framework for quantifying uncertainty in complex models, yet its application to sensitive datasets remains a significant challenge under modern privacy constraints. In particular, when performing Bayesian inference with only an observed privatized statistic, the likelihood of the observed private statistic is intractable, making existing Markov Chain Monte Carlo methods inapplicable. One promising line of work is based on data augmentation, where latent confidential data X are introduced as auxiliary variables and the joint posterior distribution over parameters and confidential data is targeted using Gibbs sampling, alternately updating one given the other. We find that this method struggles with mixing in the regime where there is coupling between the data and underlying parameters. In this work, we propose using Hamiltonian Monte Carlo to target the joint posterior, which is to our knowledge the first gradient-based posterior sampler for Bayesian inference on privatized data. We show that it drastically outperforms the Gibbs sampler in the regimes where Gibbs fails, and that it scales well with the dimension of the dataset. Finally, we discuss the limitations of Hamiltonian Monte Carlo and current ongoing work to address them.

CCS Concepts

• **Security and privacy → Privacy protections; Usability in security and privacy; Social aspects of security and privacy**;

Keywords

Differential Privacy, Bayesian Inference, Hamiltonian Monte Carlo

1 Introduction

Differential privacy (DP) has emerged as the gold standard for privacy-preserving data analysis, providing rigorous mathematical guarantees that limit information leakage about individuals in a dataset [5, 6]. At its core, DP ensures that the presence or absence of any single individual's data has minimal impact on the analysis output, typically achieved by adding carefully calibrated noise to query responses or summary statistics. This framework has been widely adopted in practice, from the U.S. Census Bureau's disclosure avoidance system to technology companies' data collection practices.

In the Bayesian setting, DP introduces unique challenges for posterior inference. Consider a standard Bayesian problem where we wish to estimate parameters θ based on observed data $X = (x_1, \ldots, x_n)$. In the non-private setting, we compute the posterior $p(\theta \mid X)$ directly from the observed data using Bayes' theorem. Under DP, however, we do not observe X itself. Instead, a trusted curator applies a privacy mechanism $g(\cdot \mid X)$ to produce a privatized statistic s_{dp} that satisfies DP guarantees. Common mechanisms include adding Gaussian or Laplace noise to sufficient statistics, with the noise magnitude calibrated to the sensitivity of the statistic and the desired privacy level ϵ. The fundamental challenge of private Bayesian inference is to compute the posterior distribution $p(\theta \mid s_{dp})$ given only this noisy observation, while properly accounting for the uncertainty introduced by the privacy mechanism. This problem has attracted significant attention [10, 15], as Bayesian methods provide a principled framework for quantifying uncertainty—crucial when privacy noise may substantially affect inference quality.

A natural approach is to directly sample from the marginal posterior of the parameters given the privatized statistic:

$$p(\theta \mid s_{DP}) = \frac{\int p(\theta)p(X \mid \theta)g(s_{DP} \mid X)dX}{p(s_{DP})} \tag{1}$$

However, this posterior is doubly intractable: the numerator requires integrating over the high-dimensional latent data $X \in \mathbb{R}^n$, and the denominator requires a further integration over θ. This motivates the data augmentation approach of [10], which treats the true (unprivatized) data X as latent variables and samples from the joint posterior $p(\theta, X \mid s_{DP})$, thereby avoiding the intractable normalizing constants. [10] introduces a Gibbs sampling framework that uses this data augmentation strategy for Bayesian inference on privatized data, alternating between updating $x|\theta, s_{dp}$ and $\theta|x$.

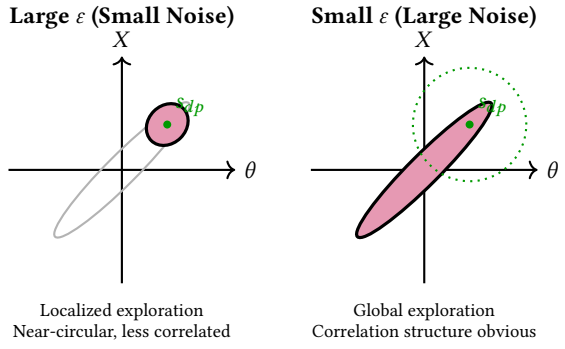

Figure 1: Posterior concentration under different privacy levels. Left: Large ϵ (small noise) localizes exploration near s_{dp} where posterior appears nearly circular. The gray ellipse represents the true correlation structure. Right: Small ϵ (large noise) requires exploring the full correlated structure. The green dotted circle represents the uncertainity around s_{dp} due to privacy noise.

Although data augmentation avoids the doubly intractable posterior, it introduces a fundamental challenge: when ϵ is small (large privacy noise), the observed statistic s_{dp} provides little constraint

on the posterior, forcing the sampler to explore the full joint distribution $p(\theta, X \mid s_{dp})$ including its highly correlated global structure (Figure 1). While [10] show that the Gibbs sampler maintains good local acceptance rates bounded by $\exp(-\epsilon)$ for Metropolis-within-Gibbs updates, the axis-aligned nature of these updates makes global exploration of the narrow correlated ridge inefficient. In contrast, when ϵ is large (small privacy noise), s_{dp} tightly constrains the posterior to a localized region where coordinates appear nearly uncorrelated. Empirically, we find that Gibbs struggles to mix at low ϵ values, particularly when θ and X exhibit strong correlation.

1.1 Our Contributions

To address this issue, we propose using gradient based joint updates of X and θ to propel the chain through the joint posterior to accelerate mixing. To the best of our knowledge, this is the first gradient-based DP posterior sampler for Bayesian inference on privatized data. Specifically, we adapt Hamiltonian Monte Carlo (HMC) to this problem setting and show that

(1) In a Gaussian simulation setting, HMC strongly outperforms Gibbs in the problematic region we discussed earlier and gives robust performance across many different settings of privacy noise and dependence between parameter and data despite the likelihood discontinuity caused by data clamping
(2) HMC scales well with the data dimension n

We also discuss the limitations of HMC and our current ongoing work to address these limitations. We ultimately aim to propose a novel, general-purpose gradient based MCMC methodology with improved mixing for Bayesian inference on privatized data.

2 Background

2.1 Differential Privacy

Differential Privacy (DP) provides a rigorous mathematical framework for quantifying and bounding the privacy risk associated with data analysis [5]. A randomized algorithm satisfies DP if its output is relatively insensitive to the presence or absence of any single individual's data in the dataset.

Definition 1 ((ϵ, δ)-Differential Privacy)**.** A randomized algorithm $\mathcal{A}$ satisfies (ϵ, δ)-differential privacy if for all neighboring datasets D, D' differing by a single entry, and all measurable subsets of outcomes $S \subseteq \text{Range}(\mathcal{A})$:

$$P(\mathcal{A}(D) \in S) \leq e^{\epsilon} P(\mathcal{A}(D') \in S) + \delta \tag{2}$$

In this formulation, ϵ (the privacy budget) represents the maximum allowable log-odds ratio of any outcome between neighboring datasets, while δ accounts for a small probability of the privacy guarantee failing.

2.2 Hamiltonian Monte Carlo

Standard Markov Chain Monte Carlo (MCMC) methods, such as the Random Walk Metropolis-Hastings algorithm, often struggle in high-dimensional settings due to the curse of dimensionality. In these spaces, random walks tend to explore the target distribution inefficiently, resulting in high autocorrelation and slow convergence to the stationary distribution [2].

Hamiltonian Monte Carlo (HMC) [1, 12] addresses these limitations by augmenting the parameter space with auxiliary momentum variables and utilizing the gradient of the target density to guide the Markov chain [4]. Let $\eta \in \mathbb{R}^d$ be the parameters of interest with a target distribution $p(\eta)$. HMC introduces a momentum vector $r \in \mathbb{R}^d$, typically sampled from $r \sim \mathcal{N}(0, M)$, where M is a mass matrix. Intuitively, the mass matrix M acts as a "tuner" for the geometry of the parameter space and determines how much "momentum" is required to move in that direction. The system is then governed by the Hamiltonian function:

$$H(\eta, r) = U(\eta) + K(r) \tag{3}$$

where $U(\eta) = -\log p(\eta)$ represents the potential energy and $K(r) = \frac{1}{2} r^\top M^{-1} r$ represents the kinetic energy. The evolution of the state (η, r) over time t is determined by Hamilton's equations:

$$\frac{d\eta}{dt} = \frac{\partial H}{\partial r} = M^{-1} r, \quad \frac{dr}{dt} = -\frac{\partial H}{\partial \eta} = -\nabla U(\eta) \tag{4}$$

In practice, these continuous dynamics are discretized using the *Leapfrog Integrator*, a volume-preserving and reversible numerical scheme. A single Leapfrog step with step size ϵ is defined as:

$$r_{t+\epsilon/2} = r_t - \frac{\epsilon}{2} \nabla U(\eta_t) \tag{5}$$

$$\eta_{t+\epsilon} = \eta_t + \epsilon M^{-1} r_{t+\epsilon/2} \tag{6}$$

$$r_{t+\epsilon} = r_{t+\epsilon/2} - \frac{\epsilon}{2} \nabla U(\eta_{t+\epsilon}) \tag{7}$$

By alternating these coherent trajectories with a Metropolis acceptance step, HMC can perform long-distance moves in the parameter space while maintaining a high acceptance probability, significantly outperforming random walk methods in complex, high-dimensional posteriors [1, 12].

2.3 Sequential and Particle-Based Methods

Other approaches to Bayesian inference on privatized data include sequential and particle-based methods. Parallel to MCMC developments, sequential methods have been proposed to handle private data. [8] provides a theoretical foundation for exact inference on privatized data by framing the differential privacy mechanism as a stochastic perturbation of the underlying data. This approach demonstrates that by treating the privacy noise as a part of the hierarchical model, one can achieve exact posterior distributions despite the "approximate" nature of the sanitized measurements. [3] introduced a Particle Filter approach for private Bayesian inference, which uses importance sampling and tempering to navigate the privacy-constrained landscape. While effective for sequential tasks, particle filters often suffer from weight degeneracy in high-dimensional parameter spaces—a limitation that gradient-based HMC methods are specifically designed to overcome by leveraging the geometry of the log-posterior.

3 Methodology

In this work, we adapt HMC to the specific setting of targetting the joint distribution $p(\theta, X \mid s_{DP})$. We require the densities of the prior, likelihood, and privacy mechanism to all be able to be evaluated efficiently, as well as the gradient and hessian. Algorithm 1 describes

the HMC methodology in detail. As part of our methodology, we set the mass matrix to be the negative hessian of the log-likelihood, evaluated at an estimate of the mode of the distribution. This is a simple way to automatically select the mass matrix hyperparameter that takes into account the correlation structure between the data and parameter, effectively standardizing the differently-scaled dimensions of the joint posterior.

Algorithm 1 Hamiltonian Monte Carlo for Private Bayesian Inference

Input: Privatized statistic s_{dp}, prior s_0^2, likelihood s_x^2, step size η, leapfrog steps L, clamping threshold C.
Initial State: $\Theta^{(0)} = (\theta^{(0)}, \mathbf{x}^{(0)})$.
Metric Construction:
  Compute Hessian $H = \nabla^2 \log \pi(\hat{\theta}, \hat{\mathbf{x}}|s_{dp})$.
  Set Mass Matrix $M = -H$
for $t = 1$ **to** T **do**
  Sample momentum $p^{(t)} \sim \mathcal{N}(0, M)$.
  Set $q_0 = \Theta^{(t-1)}$ and $p_0 = p^{(t)}$.
  Leapfrog Integration:
  $p \leftarrow p_0 + \frac{\eta}{2} \nabla \log \pi(q_0|s_{dp})$
  for $i = 1$ **to** $L - 1$ **do**
    $q \leftarrow q + \eta M^{-1} p$ {Preconditioned position update}
    $p \leftarrow p + \eta \nabla \log \pi(q|s_{dp})$
  end for
  $q \leftarrow q + \eta M^{-1} p$
  $p \leftarrow p + \frac{\eta}{2} \nabla \log \pi(q|s_{dp})$
  Metropolis-Hastings Acceptance:
  Calculate Hamiltonian $\mathcal{H}(q, p) = -\log \pi(q|s_{dp}) + \frac{1}{2} p^T M^{-1} p$.
  $\alpha = \min(1, \exp(\mathcal{H}(q_0, p_0) - \mathcal{H}(q, p)))$.
  $\Theta^{(t)} = q$ with probability α, else $\Theta^{(t)} = \Theta^{(t-1)}$.
end for

3.1 Complexity Analysis

To evaluate the computational efficiency of our methodology, we analyze the scaling of the per-iteration cost with respect to the number of latent variables n. Each iteration involves L leapfrog steps. Each step requires evaluating the gradient of the log-posterior $\nabla \log \pi$, which is a summation over n likelihood terms. While matrix inversion of the Hessian $H \in \mathbb{R}^{(n+1)\times(n+1)}$ is generally $O(n^3)$, it can be precomputed only once in the entire algorithm. Thus, the cost per HMC iteration is $O(L \cdot n)$. This is the same scaling in n as the Gibbs sampler; moreover, we do not require record-additivity [10]. We also note that while one iteration of HMC is more expensive than Gibbs, the additional overhead pays off through better mixing.

4 Simulation Results

We give simulation results and show HMC is robust to decreasing values of ϵ compared to the Gibbs sampler. In our simulation setup, we consider a simple Gaussian model where $\theta \sim N(0, \sigma_0^2)$, $x_i \sim N(\theta, \sigma_x^2)$, and $s_{DP} = \frac{1}{n} \sum_{i=1}^{n} \text{clamp}(x_i, C) + Z$ for $Z \sim N(0, \sigma_{dp}^2)$ where $\text{clamp}(x, C) = \max(-C, \min(C, x))$. The privacy noise scale is $\sigma_{dp} = \frac{2C}{n\epsilon}\sqrt{2\ln(1.25/\delta)}$. For our parameter settings, we set $C = 1.0$, $\sigma_0 = 1.0$, and we vary ϵ in $\{0.1, 0.3, 1, 3, 10\}$, n in $\{100, 500, 1000\}$ and the ratio of σ_x^2/σ_0^2 in $\{0.01, 0.1, 0.5, 1, 2, 5, 10\}$. We choose the negative Hessian evaluated at estimates of the log-joint likelihood mode for the mass matrix M, step size of 0.03 and leapfrog step $L = 20$. We perform 10 MCMC runs for each such parameter setting combination for 5000 iterations per run and record the average Effective Sample Size (ESS)/second for each setting [11]. We track the ESS/Second because it is an estimate of the number of independent samples produced per second of wall-clock time.

For this model, the log-joint density for the parameters $(\theta, \mathbf{x})$ given the observation s_{dp} is:

$$\log p(\theta, \mathbf{x}|s_{dp}) \propto -\frac{\theta^2}{2\sigma_0^2} - \sum_{i=1}^{n} \frac{(x_i - \theta)^2}{2\sigma_x^2} - \frac{(s_{dp} - \frac{1}{n}\sum \text{clamp}(x_i))^2}{2\sigma_{dp}^2} \tag{8}$$

. The gradient with respect to x_i depends on an indicator function $\mathbb{I}_i = \mathbb{I}(|x_i| < C)$:

$$\frac{\partial \log p}{\partial \theta} = -\frac{\theta}{\sigma_0^2} + \frac{n(\bar{x} - \theta)}{\sigma_x^2} \tag{9}$$

$$\frac{\partial \log p}{\partial x_i} = -\frac{x_i - \theta}{\sigma_x^2} + \frac{s_{dp} - \frac{1}{n}\sum [x_j]_C}{n\sigma_{dp}^2} \cdot \mathbb{I}_i \tag{10}$$

. The hessian is calculated similarly. In both the gradient and hessian, the terms with respect to x_i have a component due to the data likelihood and a component due to the likelihood of the observed private statistic s_{dp}. Thus, when evaluated outside the clamped region C, only the component due to the data likelihood will be active.

We can see in Figure 2 that the Gibbs sampler struggles to mix as ϵ and s_x^2/s_0^2 decreases, while HMC signficantly outperforms Gibbs in this regime and maintains robust performance across all different settings. We also note that while for Gibbs the top left region corresponding to small ϵ and s_x^2/s_0^2 is difficult, it is easiest for HMC.

While choosing the step size hyperparameter for HMC, we encountered some numerical stability issues in the bottom right corner of the plots, which made it so that we chose a slightly smaller step size than what would have worked better for values in the top left. We highlight that despite the hyperparameter selection here not being necessarily optimal, HMC still achieves robust mixing across all settings.

5 Limitations of HMC

Although HMC gives good mixing performance in this simple simulation study and appears to resolve the coupling between x and θ that the Gibbs sampler suffered from, we encountered some numerical stability issues as s_x^2/s_0^2 increased while performing these experiments. In particular, for the settings corresponding to the bottom right of the plots with $\epsilon = 10$ and s_x^2/x_0^2 larger than 1, we saw that they were much more susceptible to poor step size tuning than the top left corner and required smaller step sizes. We can understand this based on the fact that as s_x^2/s_0^2 increases, the proportion of data generated by the likelihood that lie outside the clamped region increases, thus making it more likley for the gradient and mass matrix $M = -H$ to be evaluated outside the clamped region.

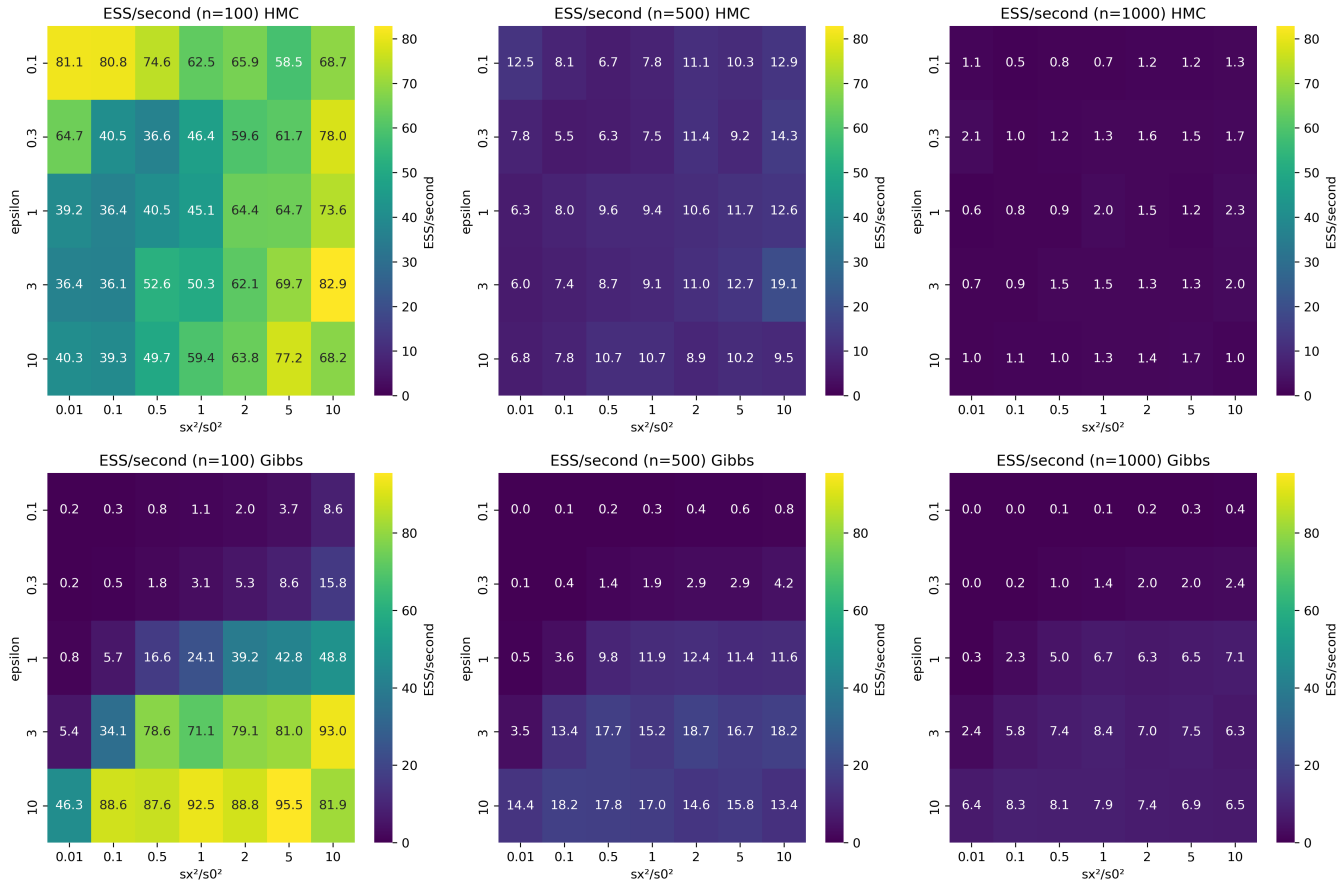

Figure 2: ESS/second for HMC sampler compared to Gibbs under the Gaussian data model, with varying ϵ, s_x^2/s_0^2, and n.

6 Ongoing and Future Work

In our current ongoing work, we are actively working to address the above limitation. In addition, we plan to handle settings where the data are discrete valued by re-parameterization or surrogate model as well as expand to more substantial simulation settings. We also plan to study principled ways of choosing the step size and leap frog steps hyperparameters for HMC. We also note that our method directly applies to the Metropolis Adjusted Langevin Algorithm (MALA) [13] preconditioned with the negative hessian and we are actively working to understand how MALA behaves in the same simulation settings as well.

We also aim to theoretically quantify how HMC interacts with privacy. In [7], they find that in a simple Gaussian toy model, the Bayesian fraction of missing information for the Gibbs sampler goes to 1 as $\epsilon \to 0$, indicating that the chain mixing worsens with more privacy. We are working on analyzing a similar quantity for HMC with the goal of ultimately showing how the convergence rate is more robust to privacy parameter ϵ than Gibbs, as well as how privacy noise can help HMC in certain settings like in the top left region of Figure 2.

References

[1] M. Betancourt. "A conceptual introduction to Hamiltonian Monte Carlo." *arXiv preprint arXiv:1701.02434*, 2017.

[2] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.

[3] Y.-W. Chen, P. Sanghi, and J. Awan. "Particle Filter for Bayesian Inference on Privatized Data." *arXiv preprint arXiv:2505.00877*, 2025.

[4] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. "Hybrid Monte Carlo." *Physics Letters B*, 195(2), pp. 216–222, 1987.

[5] C. Dwork, F. McSherry, K. Nissim, and A. Smith. "Calibrating noise to sensitivity in private data analysis." In *Theory of Cryptography Conference*, pp. 265–284. Springer, 2006.

[6] C. Dwork and A. Roth. "The algorithmic foundations of differential privacy." *Foundations and Trends in Theoretical Computer Science*, 9(3-4), pp. 211–407, 2014.

[7] K. Eng, J. A. Awan, N. P. Ju, V. A. Rao, and R. Gong. "DAPPER: Data Augmentation for Private Posterior Estimation in R." *arXiv preprint arXiv:2403.04100*, 2024.

[8] R. Gong. "Exact inference with approximate computation for differentially private data via perturbations." *Journal of Privacy and Confidentiality*, 12(2), 2022.

[9] U. Grenander and M. I. Miller. "Representations of knowledge in complex systems." *Journal of the Royal Statistical Society: Series B*, 56(4), pp. 549–581, 1994.

[10] N. Ju, J. Awan, R. Gong, and V. Rao. "Data augmentation MCMC for Bayesian inference from privatized data." *Advances in Neural Information Processing Systems*, 35, pp. 10846–10857, 2022.

[11] A. Kong. "A note on importance sampling using standardized weights." *Technical Report 348*, Department of Statistics, University of Chicago, 1992.

[12] R. M. Neal. "MCMC using Hamiltonian dynamics." In *Handbook of Markov Chain Monte Carlo*, pp. 113–162, 2011.

[13] G. O. Roberts and R. L. Tweedie. "Exponential convergence of Langevin distributions and their discrete approximations." *Bernoulli*, 2(4), pp. 341–363, 1996.

[14] G. O. Roberts and J. S. Rosenthal. "Optimal scaling of discrete approximations to Langevin diffusions." *Journal of the Royal Statistical Society: Series B*, 60(1), pp. 255–268, 1998.

[15] O. Williams and F. McSherry. "Probabilistic Inference and Differential Privacy." *Advances in Neural Information Processing Systems 23 (NIPS 2010)*, pp. 2451–2459, 2010.