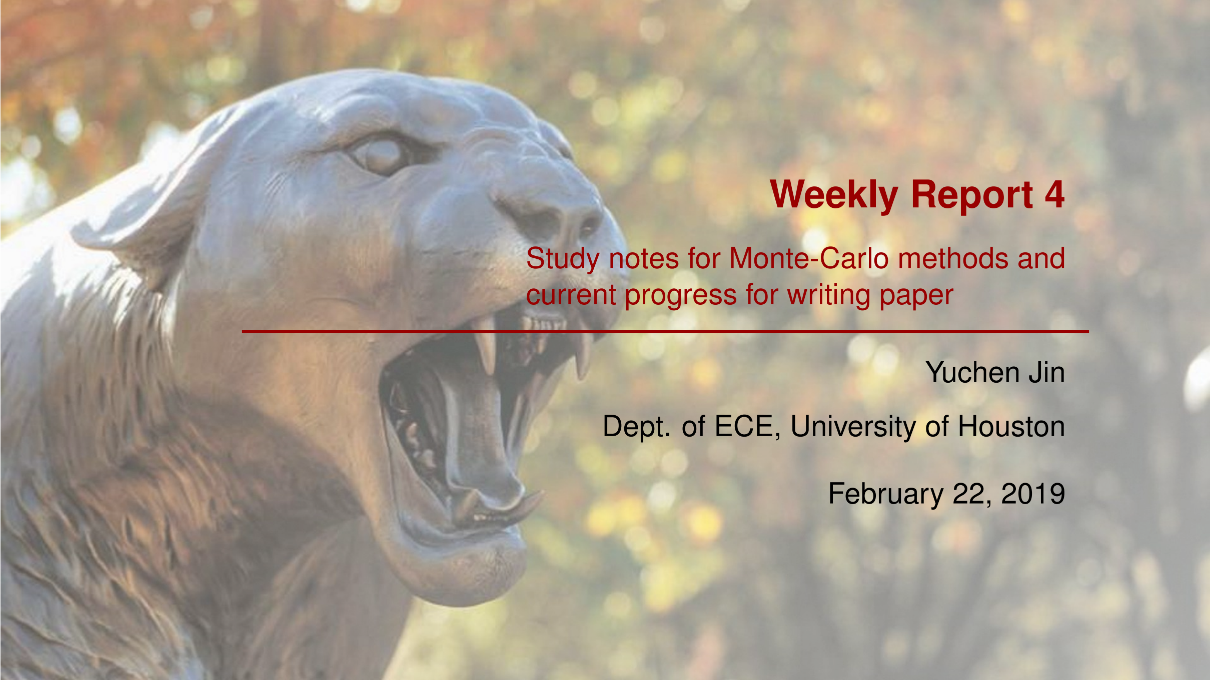

Weekly Report 4

Study notes for Monte-Carlo methods and current progress for writing paper

Yuchen Jin

Dept. of ECE, University of Houston

February 22, 2019

Outline

Outline

Introduction

- **Recent works**:
  - Complete an new article
    1. note20190215sp: The first topic about stochastic optimization: from Monte-Carlo methods to Gibbs sampling [1].
  - Progress on writing paper
    1. Have finished the framework of what we are planning for writing. (notes are attached to this slice).
    2. Generate more detail testing estimations.
    3. Propose a possible and slight revision for current work (Need to get more testing results).

Outline

Monte-Carlo study
Importance sampling

Unbiased estimator for importance sampling

$$\overline{wf(x)} := \frac{1}{N}\sum_{i=1}^{N} w_i f(x_i), \tag{1}$$

$$s^2 := \frac{1}{N-1}\sum_{i=1}^{N}\left(w_i f(x_i) - \overline{wf(x)}\right)^2, \tag{2}$$

$$\mathrm{SE}^2 := \frac{1}{N(N-1)}\sum_{i=1}^{N}\left(w_i f(x_i) - \overline{wf(x)}\right)^2, \tag{3}$$

- We use $w_i = \frac{q(x_i)}{p(x_i)}$ in importance sampling.
- We prove that (1) is an unbiased estimator for $\mathbb{E}_q[f(x)]$, however, (2) and (3) are unbiased estimators for larger standard deviation and stand error respectively.

Monte-Carlo study
Importance sampling

Unbiased estimator for importance sampling

$$\widetilde{wf(x)} = \frac{\sum_{i=1}^{N} w_i f(x_i)}{\sum_{i=1}^{N} w_i}, \tag{4}$$

$$\widetilde{\mathrm{SE}}^2 = \frac{\sum_{i=1}^{N} w_i^2 \left(f(x_i) - \widetilde{wf(x)}\right)^2}{\left(\sum_{i=1}^{N} w_i\right)^2}. \tag{5}$$

- Prove that both of these two estimators are biased.
- $\widetilde{wf(x)}$ is an asymptotically unbiased estimator.

Monte-Carlo study
Importance sampling

Figure 1: Monte-Carlo simulation for importance sampling

Importance sampling

- When we use a distribution which is close to the target distribution ($p \to q$), importance sampling is effective.
- When the distribution of the sampling is extremely different from the target distribution, the prediction would be in-precise.

Monte-Carlo study

Metropolis-Hastings algorithm

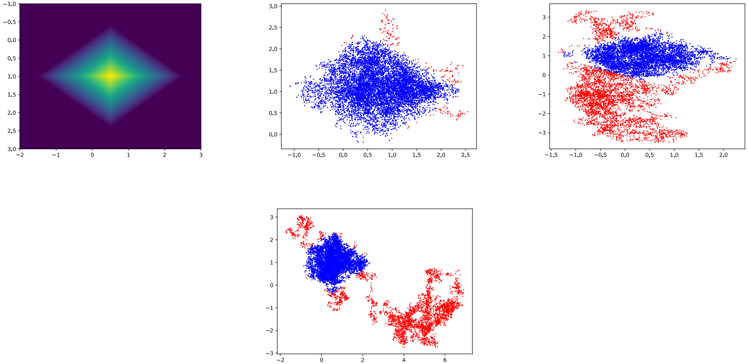

Figure 2: Distribution an using Metropolis sampling.

- Metropolis algorithm get bad performance if $p(x_k) = 0$ for current sample.

Monte-Carlo study
Metropolis-Hastings algorithm

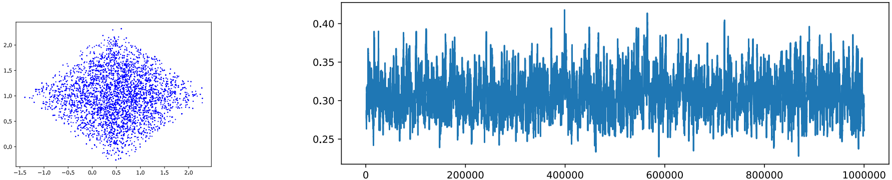

Figure 3: Improved Metropolis algorithm.

- Give a constraint that all cases that $p(x_{k+1}) = 0$ are not allowed. We could ensure that the sampling is restricted in the target domain.

Monte-Carlo study

Gibbs sampling

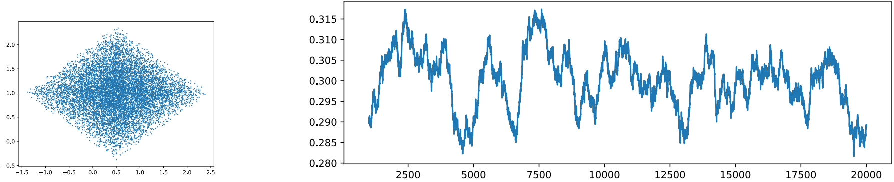

Figure 4: Improved Metropolis algorithm.

- Gibbs sampling is generally better than Metropolis-Hastings algorithm.
- Gibbs sampling could be use only when sampling under marginal distribution is easy to do.

Monte-Carlo study
Gibbs sampling

- Prove that if view MH algorithm and Gibbs sampling as AR_1 process, the standard error could be approximated by $\mathrm{SE} \approx \frac{\sigma}{\sqrt{N}}\sqrt{\frac{1+\alpha}{1-\alpha}}$.
- The exact standard error should be represented as $\mathrm{SE}^2 = \frac{1}{N}\left(\gamma(0) + 2\sum_{i=1}^{N-1}\left(1-\frac{i}{N}\right)\gamma(i)\right)$, where $\gamma(k)$ is the lag-k covariance.

Outline

Paper progress

- The paper now follows like this structure: (this article will not be published on Github until the paper is reviewed.)
  - Introduction
  - Related works
  - Problem background
    - Industrial background
    - Data
    - Geosteering inverse problem
    - Challenges
  - Proposed method
    - Bayesian view of our problem
    - Proposed network
      - Train
      - Test
  - Testing results

Paper progress

- Reformulate our problem by Bayesian regularization.
- Bayesian regularization term is $||\mathbf{x}-\mathbf{x}_0||_2^2$, which could introduce the initial guess.

Reformulate the loss function by likelihood

$$p(\mathbf{x}|\mathbf{y}) \sim p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = \frac{1}{2\pi\sigma\sigma_0} e^{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathscr{F}(\mathbf{x})\|_2^2-\frac{1}{2\sigma_0^2}\|\mathbf{x}-\mathbf{x}_0\|_2^2}. \tag{6}$$

$$\mathscr{L}(\mathbf{x}|\mathbf{y}) = \|\mathbf{y}-\mathscr{F}(\mathbf{x})\|_2^2 + \lambda\,\|\mathbf{x}-\mathbf{x}_0\|_2^2 + C. \tag{7}$$

Paper progress

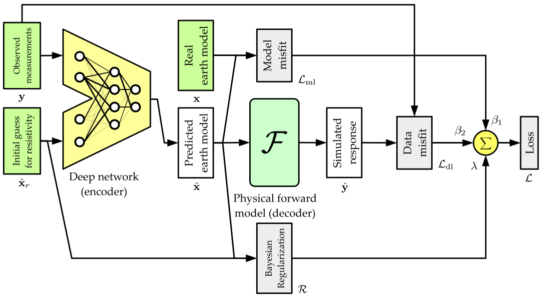

Figure 5: Improved Metropolis algorithm.

- Now we could add the Bayesian regularization to our network.
- This is a minor change, but may help us overcome the problem caused by undetermined solutions.
- Still need further tests to verify the effectiveness of this idea.

Paper progress

- Suggestions for adding more components to the paper.
  1. Use all predictions as the initial guess, to record the LMA optimization curves respectively. Use a fixed initial guess as a comparison. Then we could compare the two kinds of optimization curves and confirm that our network could help us accelerate the convergence of traditional methods.
  2. Predict some data with random noise.
  3. Show two more graphs to compare the model misfits and data misfits sample by sample.

Outline

Reference I

B. Walsh, "Markov chain monte carlo and gibbs sampling," http://nitro.biosci.arizona.edu/courses/EEB596/handouts/Gibbs.pdf, 2002.

Thank you for Listening

It’s time for Q & A