Weekly Report 3 (B)

Further literature review about stochastic optimization and recent works

Yuchen Jin

Dept. of ECE, University of Houston

February 15, 2019

Outline

Outline

Introduction

- **Recent works**:
  - Complete an new article
    1. note20190129: A collection of researches about inverse problem..
  - Work with an new article
    1. note20190215sp: The first topic about stochastic optimization: from Monte-Carlo methods to Gibbs sampling..
  - Read the lecture notes about Monte-Carlo methods. [1]
  - Read 4 more papers in this week.
    1. About an improved Metropolis-Hastings algorithm. [2]
    2. About using Gibbs sampling to solve inverse problem. [3]
    3. About using ant colony algorithm to solve inverse problem [4].
    4. About using bat algorithm to solve general optimization [5].
  - Setup and preparing for migrating to Tensorflow 1.12 API (tf-keras).

Outline

Read lecture notes

Markov Chain Monte Carlo and Gibbs Sampling [1]

- Mainly learn Monte-Carlo theory, Metropolis-Hastings algorithm theory and Gibbs sampling theory. In the future, I would post my notes to my website.
- Learn to prove that
  - If the average is $\hat{I}(y) = \frac{1}{n}\sum_i f(y|x_i)$, the standard error is $\mathrm{SE}^2\left(\hat{I}(y)\right) = \frac{1}{n}\left(\frac{1}{n-1}\left(\sum_i \left(f(y|x_i) - \hat{I}(y)\right)^2\right)\right)$.
- Why stationary condition $p(x,y)\pi(x) = p(y,x)\pi(y)$ ensures the convergence of a Markov chain.
- The stationary quality (convergence) of Metropolis-Hastings algorithm and Gibbs sampling.

Read lecture notes

Markov Chain Monte Carlo and Gibbs Sampling [1]

- The approximate standard error of a firstorder autoregressive process (AR1) is $\mathrm{SE}(\bar{\theta}) = \frac{\sigma}{n}\sqrt{1+\rho}1-\rho$, where σ is SE for a normal distribution and ρ is the first order covariance of AR1.
- The approximate Gibbs standard error: $\mathrm{SE}^2(\hat{h}) = \frac{1}{n}(\hat{\gamma}(0)\sum_i 2\hat{\gamma}(i))$, where $\hat{\gamma}(i)$ is the lag-i auto-covariance.

Outline

Recent reviews

Bayesian calibration of a largescale geothermal reservoir ... [2]

- For a non-linear problem $\mathbf{y} \sim \mathbf{f}(\mathbf{x})$, we have

$$p(\mathbf{x}|\mathbf{y}) \propto e^{-\frac{1}{2}(\mathbf{y}-\mathbf{f}(\mathbf{x}))^T \Sigma_e^{-1} (\mathbf{y}-\mathbf{f}(\mathbf{x}))} p(\mathbf{x}), \tag{1}$$

- Model $\mathbf{f}(\cdot)$ may be very complicated. So we may use the reduced order model $\mathbf{f}_d(\cdot)$ as a surrogate, hence the error is:

$$\mathbf{e} = \mathbf{y} - \mathbf{f}_d(\mathbf{x}) + (\mathbf{f}_d(\mathbf{x}) - \mathbf{f}(\mathbf{x})) \tag{2}$$

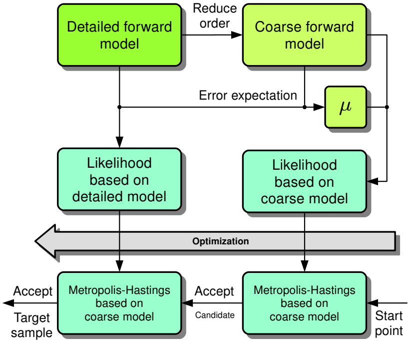
  - Denote that $\mathbf{d} = \mathbf{f}_d(\mathbf{x}) - \mathbf{f}(\mathbf{x})$, we assume that $\mathbf{d} \sim \mathcal{N}(\mu_d, \Sigma_d)$, where we could calculate the expectation of the mean error: μ_d.

- We get the coarse version of the likelihood function:

$$p_d(\mathbf{x}|\mathbf{y}) \propto e^{-\frac{1}{2}(\mathbf{y}_d-\mathbf{f}_d(\mathbf{x})-\mu_d)^T (\Sigma_e+\Sigma_d)^{-1} (\mathbf{y}_d-\mathbf{f}_d(\mathbf{x})-\mu_d)} p(\mathbf{x}). \tag{3}$$

Recent reviews

Bayesian calibration of a largescale geothermal reservoir ... [2]



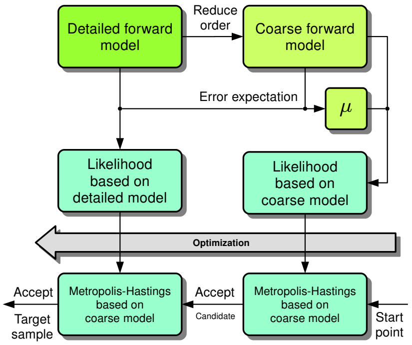
- Suppose that we have a proposal probability $q(\cdot, \cdot)$ so that we could sample $\mathbf{x}^\dagger$ from $\mathbf{x}_k$.
- First we use Metropolis-Hastings algorithm with the coarse model $\alpha(\mathbf{x}_k, \mathbf{x}^\dagger)$.

Figure 1: 2-level architecture of improved Metropolis-Hastings algorithm.

$$\alpha(\mathbf{x}_k, \mathbf{x}^\dagger) = \min\left(1, \frac{p_d(\mathbf{x}^\dagger|\mathbf{y})q(\mathbf{x}^\dagger \to \mathbf{x}_k)}{p_d(\mathbf{x}_k|\mathbf{y})q(\mathbf{x}_k \to \mathbf{x}^\dagger)}\right). \quad (4)$$

Recent reviews

Bayesian calibration of a largescale geothermal reservoir ... [2]



- If the sample is rejected, get back to step 1; otherwise, use MH algorithm with the detailed model: $\beta(\mathbf{x}_k, \mathbf{x}^\dagger)$.
- This 2-level algorithm could helps us calculate detail model for less times if the samples are easy to be rejected.

Figure 1: 2-level architecture of improved Metropolis-Hastings algorithm.

$$\beta(\mathbf{x}_k, \mathbf{x}^\dagger) = \min\left(1, \frac{p(\mathbf{x}^\dagger|\mathbf{y})\alpha(\mathbf{x}^\dagger, \mathbf{x}_k)q(\mathbf{x}^\dagger \to \mathbf{x}_k)}{p(\mathbf{x}_k|\mathbf{y})\alpha(\mathbf{x}_k, \mathbf{x}^\dagger)q(\mathbf{x}_k \to \mathbf{x}^\dagger)}\right). \quad (4)$$

Recent reviews

A Bayesian inference approach to the inverse heat ... [3]

- This article just aims at applying Gibbs sampling to solve inverse problem. According to Bayesian method, the likelihood is

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto e^{-\frac{1}{2\sigma^2}(\mathbf{f}(\boldsymbol{\theta})-\mathbf{y})^T(\mathbf{f}(\boldsymbol{\theta})-\mathbf{y})} p(\boldsymbol{\theta}). \tag{5}$$

- According to random field, the prior is defined as

$$p(\boldsymbol{\theta}) \propto e^{-\sum_{i,\ j} W_{ij}(\theta_i-\theta_j)^2} = \lambda^{\frac{m}{2}} e^{-\frac{1}{2}\boldsymbol{\theta}^T\mathbf{W}\boldsymbol{\theta}} \tag{6}$$

Recent reviews

A Bayesian inference approach to the inverse heat ... [3]

- Suppose that we have $\boldsymbol{\theta}$ and its conditional distribution $p(\boldsymbol{\theta}|\mathbf{y})$.
- Although $\mathbf{y}$ is known, it is difficult to sample a vector with such a high-order joint distribution.

Algorithm 1 Gibbs sampling algorithm

Input: A known sample $\boldsymbol{\theta}^{(k)}$, the conditional distribution $p(\theta_i|\boldsymbol{\theta}\backslash\theta_i)$.
Output: The next sample $\boldsymbol{\theta}^{(k+1)}$.

1: $\theta_1^{(k+1)} \sim p\left(\theta_1 \middle| \theta_2 = \theta_2^{(k)},\ \theta_3 = \theta_3^{(k)},\ \theta_n = \theta_n^{(k)}\right);$
2: $\theta_2^{(k+1)} \sim p\left(\theta_2 \middle| \theta_1 = \theta_1^{(k+1)},\ \theta_3 = \theta_3^{(k)},\ \theta_n = \theta_n^{(k)}\right);$
3: $\theta_3^{(k+1)} \sim p\left(\theta_3 \middle| \theta_1 = \theta_2^{(k+1)},\ \theta_2 = \theta_2^{(k+1)},\ \theta_n = \theta_n^{(k)}\right);$
4: ...
5: $\theta_n^{(k+1)} \sim p\left(\theta_n \middle| \theta_1 = \theta_1^{(k+1)},\ \theta_2 = \theta_2^{(k+1)},\ \theta_n = \theta_{n-1}^{(k+1)}\right).$

Recent reviews

Application of homogenous continuous Ant Colony ... [4]

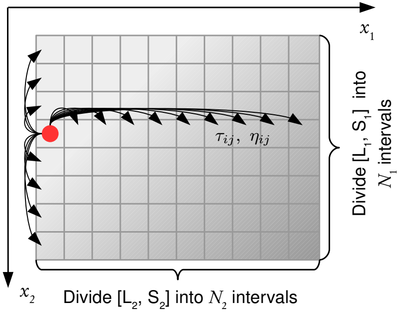


Figure 2: Applying ACO to solve inverse problem.

- Ant colony algorithm is originally used to solve Travel Salesman Problem (TSP). To apply it to solve an inverse problem in continuous space, we need to divide the parameter space into a series of meshes.

Recent reviews

Application of homogenous continuous Ant Colony ... [4]

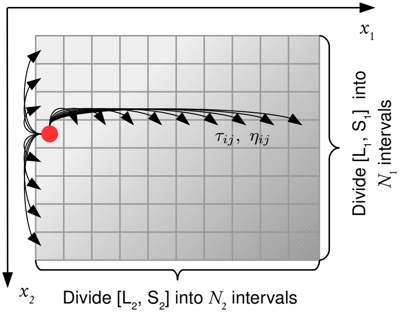


Figure 2: Applying ACO to solve inverse problem.

- Consider a problem $\arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_2^2$:
  - Initialize solutions (ants) $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(M)}$, we have $\mathscr{L}(\mathbf{x}) = \frac{1}{N}\left\|\mathbf{y} - \mathbf{f}\left(\mathbf{x}^{(k)}\right)\right\|_2$.
  - For each parameter in a solution $\mathbf{x}^{(k)} = \begin{bmatrix} x_1^{(k)} & x_2^{(k)} & \cdots & x_n^{(1)} \end{bmatrix}$, set initial scope $x_i^{(k)} \in [L_i, S_i]$.
  - Divide each scope into N_i intervals.
- Initialize that $\tau_{ij} = Q$ and $\eta_{ij} = \frac{1}{N}$.
- Hence we could calculate the transfer probability $p_{ij}^{(k)}$.

Recent reviews

Application of homogenous continuous Ant Colony ... [4]

1. Consider $\zeta \in [0,\ 1]$, $\alpha,\ \beta > 0$. We have the probability that $x_i^{(k)}$ would convert to $x_j^{(k)}$.

$$p_{ij}^{(k)} = \frac{\zeta}{N_i} + (1-\zeta)\frac{\tau_{ij}^{\alpha}\eta_{ij}^{\beta}}{\sum_{j=1}^{N_i}\tau_{ij}^{\alpha}\eta_{ij}^{\beta}}. \tag{7}$$

2. Generate new solutions according to $p_{ij}^{(k)}$. Then use searching method or random vector to find a local minimum, i.e. find $\boldsymbol{\varepsilon}$ that $\mathscr{L}(\mathbf{x}^{(k)} + \boldsymbol{\varepsilon}^{(k)}) < \mathscr{L}(\mathbf{x}^{(k)})$ if possible. Use local minimum to update solutions $x_i^{(k)}$.
3. Find the best 3 solutions $\mathbf{x}^{(s_1)}$, $\mathbf{x}^{(s_2)}$, $\mathbf{x}^{(s_3)}$ in this iteration, and update the optimal solution $\mathbf{x}^*$ in the record.
4. For any $x_i^{(k)}$ that changes to $x_j^{(k)}$ in step 1, update the information parameter increments $\Delta\tau_{ij}^{(k)} = Q$ and $\Delta\eta_{ij}^{(k)} = \frac{1}{N\mathscr{L}(\mathbf{x}_j)}$.

Recent reviews

Application of homogenous continuous Ant Colony ... [4]

4. Update the information parameters.

$$\tau_{ij} = (1-\rho)\tau_{ij} + \sum_{k=1}^{M} \Delta\tau_{ij}^{(k)}, \tag{8}$$

$$\eta_{ij} = \max\left(\eta_{ij}, \sum_{k=1}^{M} \Delta\eta_{ij}^{(k)}\right). \tag{9}$$

5. Consider $\gamma \in (0,1)$, reset the searching scope

$$L_i = \max\left(\min\left(x_i^*, x_i^{(s_1)}, x_i^{(s_2)}, x_i^{(s_3)}\right) - \gamma\frac{S_i - L_i}{2}, L_i\right), \tag{10}$$

$$S_i = \min\left(\max\left(x_i^*, x_i^{(s_1)}, x_i^{(s_2)}, x_i^{(s_3)}\right) + \gamma\frac{S_i - L_i}{2}, S_i\right). \tag{11}$$

6. Resampling the new N_i intervals according to new scope $x_i^{(k)} \in [L_i, S_i]$. Note that we need to interpolate τ_{ij}, η_{ij} for new intervals due to the resampling. Then return to step 1.

Recent reviews

A New Metaheuristic Bat-Inspired Algorithm [4]

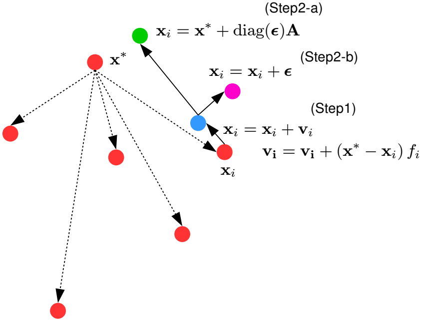


Figure 3: Applying ACO to solve inverse problem.

- Bat algorithm is like particle swarm algorithm (PSO). Suppose we have loss function $\mathscr{L}(\mathbf{x})$, where solution $\mathbf{x}_i$ is a "bat".
- For each bat, it has a velocity $\mathbf{v}_i$, a frequency f_i, a pulse rate $r_i = 0$, a loudness $A_i = 0$. Set $\alpha,\ \beta \in (0, 1)$.

Recent reviews

Application of homogenous continuous Ant Colony ... [4]

Algorithm 2 Bat algorithm

Input: $\mathbf{x}$, $\mathscr{L}(\cdot)$, $\mathbf{v}_i$, f_i, $r_i = 0$, $A_i = 0$ and α, $\beta \in (0,1)$.

Output: Optimal solution $\mathbf{x}^*$.

1: **for** $t \in [1, ^T]$ **do**
2: **for all** i **do**
3: Generate new solution $\mathbf{x}_i'$. Update frequency $f_i \sim U(0, \alpha)$;
4: $\mathbf{v}_i' = \mathbf{v}_i(1 - f_i) + (\mathbf{x}^* - \mathbf{x}_i) f_i$; Then let $\mathbf{x}_i' = \mathbf{x}_i + \mathbf{v}_i'$;
5: Generate $r \sim U(0,1)$. If $r > r_i$, select one of the best solutions, then $\mathbf{x}_i' = \mathbf{x}_{\text{selected}} + \text{diag}(\boldsymbol{\varepsilon})\mathbf{A}$;
6: Adjust $\mathbf{x}_i'$ slightly and randomly;
7: Generate $a \sim U(0,1)$. If $a < A_i$ and $\mathscr{L}(\mathbf{x}_i') < \mathscr{L}(\mathbf{x}_i)$, then let $\mathbf{x}_i = \mathbf{x}_i'$, $r_i = 1 - e^{\lambda t}$, $A_i = \beta A_i$.
8: **end for**
9: **end for**

Outline

About coding
Tensorflow has been updated

- The new Tensorflow 1.12 has been quite different from the former versions. It is necessary to catch up with the new API and coding skills.
- Most of the APIs are migrated into keras module. In this year, Tensorflow 2.0 will be released. TF 2.0 would force users to use keras style because `tf.layers` and `tf.contrib` and some other modules would be removed.
- Now I am still inspecting on the new APIs. In the next week, this work should be almost finished.

About coding

Tensorflow has been updated

- Training a model in old API is like this:

```
@tf.function
def train(model, dataset, optimizer):
  for x, y in dataset:
    with tf.GradientTape() as tape:
      prediction = model(x)
      loss = loss_fn(prediction, y)
    gradients = tape.gradients(loss, model.trainable_variables)
    optimizer.apply_gradients(gradients, model.trainable_variables)
```

- Now it becomes:

```
model.compile(optimizer=optimizer, loss=loss_fn)
model.fit(dataset)
```

Outline

Reference I

- B. Walsh, “Markov chain monte carlo and gibbs sampling,” http://nitro.biosci.arizona.edu/courses/EEB596/handouts/Gibbs.pdf, 2002.
- T. Cui, C. Fox, and M. J. O’Sullivan, “Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance metropolis hastings algorithm,” *Water Resources Research*, vol. 47, no. 10. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR010352
- J. Wang and N. Zabaras, “A bayesian inference approach to the inverse heat conduction problem,” *International Journal of Heat and Mass Transfer*, vol. 47, no. 17, pp. 3927 – 3941, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0017931004000985
- B. Zhang, H. Qi, Y.-T. Ren, S.-C. Sun, and L.-M. Ruan, “Application of homogenous continuous ant colony optimization algorithm to inverse problem of one-dimensional coupled radiation and conduction heat transfer,” *International Journal of Heat and Mass Transfer*, vol. 66, pp. 507 – 516, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S001793101300608X

Reference II

X.-S. Yang, “A New Metaheuristic Bat-Inspired Algorithm,” *arXiv e-prints*, p. arXiv:1004.4170, Apr 2010.

Thank you for Listening

It’s time for Q & A