

---

# UPDATES - MITACS PROJECT

---

## 1 Week 1

### 1.1 Deciding on the project

We talked about three project ideas

1. Explicit block couplings on "symmetric", "big"  $\Omega$ 's.  
Eg. The Ising model on  $\mathbb{Z}_n^2$
2. Existence of MMC's Look at prof Aaron's notes on MMC.
3. Computing couplings for stats. (Eg. work done by Jacob Pierre)  
Read: Pierre's notes on Couplings

I decided to try the third one first.

### 1.2 To proceed on the third project

There were three tasks to be done.

1. Set up dropbox to write notes  
**Update:** Done.
2. Find Pierre's notes on Couplings and read a bit.  
**Update:** I found the notes. I'm reading them up, and will be done with 3 chapters of it by our next meeting (I guess on Tuesday)
3. Try a simple example (Suggestion : MH algorithm on  $N(0, 1)$  with mirror and maximal)

## 2 Week 2

- Created a github repo for the whole project.
- Coded up MH on  $N(0, 1)$  using the maximal couplings.  
The algorithm is as follows:  
Let  $K$  be the kernel for the MH algo. At any point  $X_0, Y_0$  of the chain, define the point-wise maxima as  $\Pi_{X_0, Y_0}(\cdot) = \min(K(X_0, \cdot), K(Y_0, \cdot))$ .  
Define the residuals as  $r_{X_0}(\cdot) = K(X_0, \cdot) - \Pi_{X_0, Y_0}(\cdot)$  and  $r_{Y_0}(\cdot) = K(Y_0, \cdot) - \Pi_{X_0, Y_0}(\cdot)$ .

Algorithm:

w.p.  $\Pi_{X_0, Y_0}(\mathbb{K})$ , choose

$$X_1 = Y_1 \sim \Pi_{X_0, Y_0}(\cdot) / \Pi_{X_0, Y_0}(\mathbb{K})$$

Otherwise

$$X_1 \sim r_{X_0}(\cdot) / r_{X_0}(\mathbb{K})$$

$$Y_1 \sim r_{Y_0}(\cdot) / r_{Y_0}(\mathbb{K})$$

Repeat the above steps to get the chain  $(X_n)$  and  $(Y_n)$

- Studied maximal coupling, optimal transport coupling, total variation norm, weisserstein distances.

- Meeting #2 with Aaron:
  - Discussed about when a coupling is faithful, coupling inequality and mixing time in our context. A coupling is said to be "good/faithful/stichy" if :

$$\forall s > \Gamma_{\text{hit}}, X_s = Y_s,$$

$$\text{where, } \Gamma_{\text{hit}}(x, y) = \min\{s : X_s = Y_s, X_0 = x, Y_0 = y\}$$

**Theorem 2.1.**

$$\|K^s(x, \cdot) - K^s(y, \cdot)\|_{TV} \leq \Pr(\Gamma_{\text{hit}}(x, y) > s)$$

**Corollary 2.1.1.**

$$\Gamma_{\text{mix}}(\epsilon) \leq \min\{s : \max_{x,y} \Pr(\Gamma_{\text{hit}}(x, y) > s) \leq \epsilon\}$$

- Discussed the main idea of Broderick et al's paper. The setting of the problem is basically we have a lot of vertices with +1 and -1 labellings. We want to connect them using edges to make a graph such that the energy of the graph is as minimum as possible (energy is lower if we minimize the number of edges between opposite signed vertices) This is kind of like a clustering problem. The problem is to draw from configurations like these.  
Energy of a particular configuration is of the form  $H(\sigma) = -\sum_{(i,j) \in E} \sigma(i)\sigma(j)$ , where  $\sigma(i)$  denotes the label of the vertex of a particular configuration.  
The probability density is then of the form  $p(\sigma) \propto \exp(-\beta H(\sigma))$  (doubt: what exactly are we drawing from this distribution, how to do it?)  
As there is no difference if we switch the labels in our case, we take the space to be  $\bar{\Sigma} = \Sigma/(\text{flip})$
- The paper doesn't contribute much theoretically but is very well written. The work done is on a computer and some new things and ideas are presented. The aim of our project is to take one of the below presented problems, try to make some progress in them and write a Broderick style paper on it.
- We then discussed the following three problems:
  1. NNMF : The problem is that we have a very large matrix and we need to decompose it into product of smaller dimension matrices such that the energy is minimized.  
 $A \approx C^t R$ ,  $H(C, R) \approx \|C\| + \|R\| + \|A - C^t R\|$  (**TODO**: different norms are used, fill it up after reading the paper)  
Then draw from the defined distribution.  $p(C, R) \propto \exp(-\beta H(C, R))$   
This is also similar to the classification problem, just not very obvious. (**TODO**: read the paper to understand why this problem is also similar to a classification problem)
  2. Ranking : Observe games  $g_1(h_1, a_1) \dots g_N(h_N, a_N) \in \{-1, 1\}$ ,  $h_i, a_i \in \{1, \dots, n\}/\Delta$ , n: #teams, N: #games  
Based on the win/lose data, we try to rank them so that the energy is minimized. Energy is as follows:  
 $H(\sigma) = \sum_{i=1}^N \mathbb{1}_{\sigma(h_i) > \sigma(a_i)} \cdot \mathbb{1}_{g_i(h_i, a_i) = 1}$  (**TODO**: get a symmetric formula for energy (HINT: use the bradley-terry model)).  
Then draw from the probability distribution defined as  $p(\sigma) \propto \exp(-\beta H(\sigma))$ .  
This problem for  $\beta = 0$  is called the random transposition walk (like glauher dynamics)
  3. Q: What happens when  $\beta$  is small. Problem is a similar phenomenon to Gibbs algo on simplexes.
- Studied convergence of Markov chains and application of coupling techniques to prove convergence.

### 3 Week 3

Tasks for the week

- Complete Jacob's notes and lecture videos
- Broderick's paper
- Study more about these problems, try to work out some details

Decided to work on NNMF

## 4 Week 4

- Meeting #3 with Aaron:
  - Read the following stuff:
    - \* hit and run as a unifying device (also see geyer notes on hit and run algorithms)
    - \* "sandwich algorithm" by Kshitij Khare
    - \* Hamiltonian Monte Carlo (simplest concrete example for our problem)
  - NNMF is analogous to mixture models, and we can visualize mixture model in the 1-dim case. Problem is first to find an analogous version of mixture models in the discrete case so that visualization is easier.
  - Paper idea #1 (easier): Deal with the label-switching problem in the case of 3 modes by projecting state space.  
 $M = [v_1, \dots, v_1, v_2, \dots, v_2, v_3, \dots, v_3]$ , modes are  $C = [v_1, v_2]$  or  $C = [v_1, v_3]$  or  $C = [v_3, v_1]$ , all possible label switchings, all convex combinations.  
 These modes are good candidates in the case of  $\|\cdot\|_0$ , bad for  $\|\cdot\|_\infty$ . Need to test for all the other distances in between.  
 The problem is to deal with all these label switching cases by projecting the state space to partition space of some sort as done in Broderick et. al's paper.
  - Paper idea #2 (tougher but more interesting): Work in the discrete state space setting, find couplings that hit in here, and try to tweak it so that the same thing kind of works in the continuous state as well.  
 Our approach is the following:
    1. From the general matrix  $M'$  in consideration, we somehow get  $M$  which has finite number of distinct columns less than the number of columns of  $M'$  (this is a clustering problem on its own).
    2. Now from columns of  $M$ , we get all the possible modes and consider a graph with those modes as our vertices.
    3. Q: How does likelihood look on paths between nodes.
    4. We find couplings in this graph that hits in finite time (maybe we can use Broderick's ideas).
    5. Then using the coupling in the discrete case, by some "tweaks" we can try to find a coupling in the continuous case.
  - Another interesting question: in higher dimension, can we get situations where mixing is quick but meeting is slower? (for eg. in the case of random transposition walk, mixing is  $O(n \log n)$  and meeting is  $O(n^2)$ )  
 Idea: construct some examples where there are "good paths" between vertices of our graph.
- Plan of action: First I'll start working on the easier idea.
  - Hard code a 3 distinct vector matrix of choice.
  - Create a suitable penalty function and probability distribution using that function.
  - Samples nodes from this distribution using MCMC (MH-algo for starters).