

Compte-rendu de Réunion de Lancement

Rencontre Tuteur & Présentation du Projet Statistique

Sujet : Transport Optimal Généralisé pour l'analyse Single Cell avec quantification d'incertitudes

Date : [8/12/2025]

Participants :

- Tuteur : Stephane CHRETIEN,
 - Cheikna Amala YATABARE, cheikna-amala.yatabare@eleve.ensai.fr
 - Yarbana CHEIKH MOHAMED VADEL, yarbana.cheikh-mohamed-vadel@eleve.ensai.fr
 - Steeve Rodrigue TAKOUGOUUM, steeve-rodrigue.takougom@eleve.ensai.fr
-

1 Prise de contact (18h00 – 18h15)

2 Présentation du contexte du sujet (18h15 – 18h50)

- **Introduction de plusieurs notions essentielles :**
 - Le piège des « Big Data » et les risques de biais.
 - La problématique de la similarité et du clustering pour analyser l'hétérogénéité cellulaire.
 - Deux approches méthodologiques discutées :
 - **Feature Engineering** : création manuelle de variables pertinentes.
 - **Deep Learning** : extraction automatique de caractéristiques via les réseaux de neurones.
- **Objectif** : Analyser des données génétiques afin d'étudier l'hétérogénéité des cellules.
- **Méthode** : Utilisation du **transport optimal** pour comparer des distributions complexes et mieux comprendre les structures internes des populations cellulaires.

3 Le Projet : Transport Optimal & Analyse Single Cell

L'objectif du projet est d'exploiter le Transport Optimal (OT) pour mesurer la similarité entre cellules à partir de leurs profils génétiques, tout en intégrant une quantification de l'incertitude afin d'évaluer la robustesse des distances obtenues et la stabilité des structures cellulaires identifiées.

Défis spécifiques au Single Cell

- **Volatilité des données** : L'expression génétique varie fortement dans le temps ; deux mesures réalisées à quelques heures d'intervalle peuvent produire des profils très différents.
- **Bruit et robustesse** : Les données single-cell sont fortement bruitées. Il est donc essentiel de tester la stabilité des résultats en introduisant des perturbations contrôlées dans les données.

Une présentation de l'article de référence *Optimal Transport improves cell-cell similarity inference in single-cell omics data* (Huizing & Peyré) a été faite, illustrant comment l'OT permettrait

de mieux capturer la géométrie intrinsèque des données biologiques pour améliorer l'inférence de similarité entre cellules.

4 Feuille de route (Roadmap)

Avant la prochaine séance : Lecture du document de référence partagé, compréhension du code, et appropriation de quelques notions fondamentales du Transport Optimal.

Durant tout le projet :

1. **Compréhension théorique** : Assimiler les concepts mathématiques essentiels du Transport Optimal.
2. **Reproduction de l'algorithme proposé par Huizing & Peyré, en utilisant les techniques présentées durant le projet.**
3. **Expérimentation** : Tester l'algorithme, notamment sa robustesse face aux perturbations.

5 Questions / Réponses (18h50 – 19h10)

La séance s'est terminée par une discussion générale permettant de clarifier certains points techniques et applicatifs, notamment :

- La justification du choix du Transport Optimal comme métrique de similarité préférentielle.
- L'utilité biologique et les applications concrètes de l'analyse Single Cell.

6 Conclusion

Cette réunion de lancement a permis de poser les bases théoriques et méthodologiques du projet. Il en ressort que la réussite de l'analyse ne dépendra pas de la quantité de données, mais de la pertinence de la mesure de similarité choisie.

L'objectif à court terme est désormais clair : passer de la théorie à la pratique en reproduisant les résultats de l'article de Huizing & Peyré. Cette étape est indispensable pour valider notre compréhension avant d'aborder la quantification d'incertitudes dans un contexte « Single Cell » bruité.