Comité de suivi Individuel

Réinscription en 3ème année de thèse

Victor Thuot[†]

encadrement: Nicolas Verzelen[†] et Alexandra Carpentier[*]

8 juillet 2025

[†] INRAE, Mistea, Institut Agro, Univ Montpellier, Montpellier, France

[*] Institut für Mathematik, Universität Potsdam, Potsdam, Germany

Bilan scientifique

   Project 1 – Clustering with bandit feedback

   Project 2 – Clustering with Bandit Feedback (entry by entry)

   Project 3 – Condorcet Winner Identification

Bilan globale

   Séminaires et conférences

   Formation à l'ED

   Enseignement

Bilan scientifique

Pure exploration over a matrix (with bandit feedback)

- **matrix environment** – M ba $n \times d$ matrix,

$$
M = \begin{bmatrix}
M_{1,1} & \cdots & M_{1,j} & \cdots & M_{1,d} \\
\vdots & & \vdots & & \vdots \\
M_{i,1} & \cdots & M_{i,j} & \cdots & M_{i,d} \\
\vdots & & \vdots & & \vdots \\
M_{n,1} & \cdots & M_{n,j} & \cdots & M_{n,d}
\end{bmatrix}
$$

- **learning protocol** – a learner observes[1] sequentially and adaptively a subset of the entries of M

- **unknown structure** – unknown structure over the matrix

- **objective** – (a) recover the unknown structure w.h.p, while (b) minimize the budget[2]

1. with some sub-Gaussian noise
2. total number of observations – chosen by the learner

Pure exploration over a matrix (with bandit feedback)

- **matrix environment** – M ba $n \times d$ matrix,

$$M = \begin{bmatrix} M_{1,1} & \cdots & M_{1,j} & \cdots & M_{1,d} \\ \vdots & & \vdots & & \vdots \\ M_{i,1} & \cdots & M_{i,j} & \cdots & M_{i,d} \\ \vdots & & \vdots & & \vdots \\ M_{n,1} & \cdots & M_{n,j} & \cdots & M_{n,d} \end{bmatrix}$$

- **learning protocol** – a learner observes[1] sequentially and adaptively a subset of the entries of M

- **unknown structure** – unknown structure over the matrix

- **objective** – (a) recover the unknown structure w.h.p, while (b) minimize the budget[2]

    1. with some sub-Gaussian noise
    2. total number of observations – chosen by the learner

Problem

- **Observations** – one entire row (dimension d) at a time
- **Unknown structure** – there exists a partition of the rows G^* into K groups, so that, two rows μ_i and μ_j are in the same group, iff $\mu_i = \mu_j$.

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 & 5 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{cases} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_2 \\ \mu_1 \end{cases}$$

- **Objective** – recover G^* with probability larger than $1 - \delta$

Problem

- **Observations** – one entire row (dimension d) at a time
- **Unknown structure** – there exists a partition of the rows G^* into K groups, so that, two rows μ_i and μ_j are in the same group, iff $\mu_i = \mu_j$.

$$
M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 & 5 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \left\{ \begin{array}{l} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_2 \\ \mu_1 \end{array} \right.
$$

- **Objective** – recover G^* with probability larger than $1 - \delta$

Problem

- **Observations** – one entire row (dimension d) at a time

- **Unknown structure** – there exists a partition of the rows G^* into K groups, so that, two rows μ_i and μ_j are in the same group, iff $\mu_i = \mu_j$.

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 & 5 \\ \boxed{0} & \boxed{0} & \boxed{1} & \boxed{1} & \boxed{0} & \boxed{0} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \left\{ \begin{array}{l} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_2 \\ \mu_1 \end{array} \right.$$

- **Objective** – recover G^* with probability larger than $1 - \delta$

Problem

- **Observations** – one entire row (dimension d) at a time
- **Unknown structure** – there exists a partition of the rows G^* into K groups, so that, two rows μ_i and μ_j are in the same group, iff $\mu_i = \mu_j$.

$$
M = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
\boxed{5} & \boxed{5} & \boxed{5} & \boxed{5} & \boxed{5} & \boxed{5} \\
5 & 5 & 5 & 5 & 5 & 5 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\quad
\left\{
\begin{array}{l}
\mu_1 \\
\mu_2 \\
\mu_3 \\
\mu_3 \\
\mu_2 \\
\mu_1
\end{array}
\right.
$$

- **Objective** – recover G^* with probability larger than $1 - \delta$

Problem

- **Observations** – one entire row (dimension d) at a time
- **Unknown structure** – there exists a partition of the rows G^* into K groups, so that, two rows μ_i and μ_j are in the same group, iff $\mu_i = \mu_j$.

$$
M = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 5 & 5 & 5 \\
5 & 5 & 5 & 5 & 5 & 5 \\
0 & 0 & 1 & 1 & 0 & 0 \\
\boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0}
\end{bmatrix}
\left\{
\begin{array}{l}
\mu_1 \\
\mu_2 \\
\mu_3 \\
\mu_3 \\
\mu_2 \\
\mu_1
\end{array}
\right.
$$

- **Objective** – recover G^* with probability larger than $1 - \delta$

Project 1 – Clustering with bandit feedback

Advancement ⌛
- (23/24) bibliography ⌛– algorithms ⌛(construction and analysis)– lower bound, optimality ⌛– numerical experiment ⌛– redaction and submission ICML,Neurips 📵 (rejected)
- (24/25-s1) revision of the paper ⌛– publication[3] for ALT (February, Milan)📵
- (25/26) improve lower bound ⌛– extension to nonparametric setting (in RKHS) ⌛? (joint work with Sebastian Vogt, and Debarghya Ghoshdastidar)

---

3. Victor Thuot et al. **"Clustering with bandit feedback: breaking down the computation/information gap"**. In: *Proceedings of The 36th International Conference on Algorithmic Learning Theory*. PMLR.

Project 1 – Clustering with bandit feedback

Advancement ⏳

- (23/24) bibliography ⏳ – algorithms ⏳ (construction and analysis)– lower bound, optimality ⏳ – numerical experiment ⏳ – redaction and submission ICML,Neurips 🚫 (rejected)

- (24/25-s1) revision of the paper ⏳ – publication[3] for ALT (February, Milan)👆

- (25/26) improve lower bound ⏳ – extension to nonparametric setting (in RKHS) ⏳❓ (joint work with Sebastian Vogt, and Debarghya Ghoshdastidar)

---

3. Victor Thuot et al. **"Clustering with bandit feedback: breaking down the computation/information gap"**. In: *Proceedings of The 36th International Conference on Algorithmic Learning Theory*. PMLR.

*joint work with Maximilian Graf– PhD student in Potsdam

Problem

- **Observations** – one entry $(I_j, J_t) \in [n] \times [d]$ at a time
- **Unknown structure** – there exists a partition of the rows G^*, so that, two rows μ_i and μ_j are in the same group, iff $\mu_i = \mu_j$, restriction to two groups

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & \boxed{0} \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \end{bmatrix} \begin{cases} \mu_0 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_0 \\ \mu_1 \end{cases}$$

- **Objective** – recover G^* with probability larger than $1 - \delta$

*joint work with Maximilian Graf– PhD student in Potsdam

Problem

- **Observations** – one entry $(I_j, J_t) \in [n] \times [d]$ at a time
- **Unknown structure** – there exists a partition of the rows G^*, so that, two rows μ_i and μ_j are in the same group, iff $\mu_i = \mu_j$, restriction to two groups

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 1 & 0 & 0.5 & \boxed{0.05} & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \end{bmatrix} \left\{ \begin{array}{l} \mu_0 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_0 \\ \mu_1 \end{array} \right.$$

- **Objective** – recover G^* with probability larger than $1 - \delta$

*joint work with Maximilian Graf– PhD student in Potsdam

Problem

- **Observations** – one entry $(I_j, J_t) \in [n] \times [d]$ at a time
- **Unknown structure** – there exists a partition of the rows G^*, so that, two rows μ_i and μ_j are in the same group, iff $\mu_i = \mu_j$, restriction to two groups

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \end{bmatrix} \begin{cases} \mu_0 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_0 \\ \mu_1 \end{cases}$$

- **Objective** – recover G^* with probability larger than $1 - \delta$

*joint work with Maximilian Graf– PhD student in Potsdam

Problem

- **Observations** – one entry $(I_j, J_t) \in [n] \times [d]$ at a time
- **Unknown structure** – there exists a partition of the rows G^*, so that, two rows μ_i and μ_j are in the same group, iff $\mu_i = \mu_j$, restriction to two groups

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 1 & 0 & 0.5 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \boxed{0} & 0 & 0.5 & 0.05 & 0 \end{bmatrix} \left\{ \begin{array}{l} \mu_0 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_0 \\ \mu_1 \end{array} \right.$$

- **Objective** – recover G^* with probability larger than $1 - \delta$

Project 2 – Clustering with Bandit Feedback (entry by entry)

*joint work with Maximilian Graf– PhD student in Potsdam

Advancement

- (2023/2024–s2) bibliography on pure exploration ⌛ – algorithm, upper bound ($K = 2$) ⌛ – lower bound ($K = 2$) ⌛

- (2024/2025) algorithm, upper bound ($K = 2$) ⌛ – numerical experiments (Max) ⌛ – lower bound, optimality ($K = 2$) ⌛ (but room for improvement) – redaction, submission, and correction ⌛ – publication[4] ICML (Vancouver July) 👏

- (2025/2026) improved lower bound ⌛ – extension to $K > 2$ groups ⌛**?**

---

4. Maximilian Graf, Victor Thuot, and Nicolas Verzelen. *Clustering Items through Bandit Feedback: Finding the Right Feature out of Many.* Accepted at the 42nd International Conference on Machine Learning. 2025. arXiv: 2503.11209 [stat.ML].

*joint work with Maximilian Graf– PhD student in Potsdam

Advancement

- (2023/2024–s2) bibliography on pure exploration ⏳ – algorithm, upper bound ($K = 2$) ⏳ – lower bound ($K = 2$) ⏳

- (2024/2025) algorithm, upper bound ($K = 2$) ⏳ – numerical experiments (Max) ⏳ – lower bound, optimality ($K = 2$) ⏳ (but room for improvement) – redaction, submission, and correction ⏳ – publication[4] ICML (Vancouver July) 🤝

- (2025/2026) improved lower bound ⏳ – extension to $K > 2$ groups ⏳❓

4. Maximilian Graf, Victor Thuot, and Nicolas Verzelen. **Clustering Items through Bandit Feedback: Finding the Right Feature out of Many.** Accepted at the 42nd International Conference on Machine Learning. 2025. arXiv: 2503.11209 [stat.ML].

Project 3 – Condorcet Winner Identification

*joint work with El Mehdi Saad – Kaust

Problem

- **Observations** – one comparison between two items $(I_j, J_t) \times [d]$
- **Unknown structure** – $M - \frac{1}{2}I$ antisymmetric, there exists a Condorcet Winner such that $M_{i_*,j} > 1/2$ for each j

$$M = \begin{bmatrix} 0.8 & 0.9 & \boxed{0.7} & 0.85 \\ 0.1 & 0.5 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0.5 & 0.4 \\ 0.3 & 0.4 & 0.7 & 0.5 \end{bmatrix}$$

- **Objective** – recover the Condorcet winner w.h.p

Project 3 – Condorcet Winner Identification

*joint work with El Mehdi Saad – Kaust

Problem

- **Observations** – one comparison between two items
  $(I_j, J_t) \times [d]$
- **Unknown structure** – $M - \frac{1}{2}I$ antisymmetric, there exists a
  Condorcet Winner such that $M_{i_*,j} > 1/2$ for each j

$$
M = \begin{bmatrix}
0.8 & 0.9 & 0.7 & 0.85 \\
0.1 & 0.5 & 0.3 & 0.6 \\
0.2 & \boxed{0.6} & 0.5 & 0.4 \\
0.3 & 0.4 & 0.7 & 0.5
\end{bmatrix}
$$

- **Objective** – recover the Condorcet winner w.h.p

Project 3 – Condorcet Winner Identification

*joint work with El Mehdi Saad – Kaust

Problem

- **Observations** – one comparison between two items $(I_j, J_t) \times [d]$
- **Unknown structure** – $M - \frac{1}{2}I$ antisymmetric, there exists a Condorcet Winner such that $M_{i_*, j} > 1/2$ for each j

$$
M = \begin{bmatrix}
0.8 & 0.9 & 0.7 & 0.85 \\
0.1 & 0.5 & 0.3 & 0.6 \\
0.2 & 0.6 & 0.5 & 0.4 \\
0.3 & 0.4 & 0.7 & 0.5
\end{bmatrix}
$$

- **Objective** – recover the Condorcet winner w.h.p

*joint work with El Mehdi Saad – Kaust

Problem

- **Observations** – one comparison between two items
  $(I_j, J_t) \times [d]$
- **Unknown structure** – $M - \frac{1}{2}I$ antisymmetric, there exists a
  Condorcet Winner such that $M_{i_*,j} > 1/2$ for each j

$$M = \begin{bmatrix} \boxed{0.8} & 0.9 & 0.7 & 0.85 \\ 0.1 & 0.5 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0.5 & 0.4 \\ 0.3 & 0.4 & 0.7 & 0.5 \end{bmatrix}$$

- **Objective** – recover the Condorcet winner w.h.p

Project 3 – Condorcet Winner Identification

*joint work with El Mehdi Saad – Kaust

Advancement

- (2023/2024-s2) bibliography on dueling bandit ⌛– algo ⌛
- (2024/2025- s2) lower bound ⌛
- (2025/2026) submission to a conference ❓

Bilan globale

Conférences et séminaires en 1ère année

- Conférence StatMatAppli 2023
  Fréjus, France – 17 au 22 sept 2023

- Workshop on Bandits and Statistical Tests
  Potsdam, Allemagne – 24 au 25 nov 2023

- 3rd ASCAI Workshop in Potsdam
  Potsdam, Alemagne – 20 au 21 fev 2024
    * Présentation projet [1]

- Amazon StatML Workshop 2024
  Berlin, Allemagne – 15 au 17 avril 2024

- Séminaires ML-MTP
  Montpellier, France – 02 mai 2024
    * Présentation projet [1]

Conférences et séminaires en 2ème année

- **36th International Conference on Algorithmic Learning Theory**
  Milan, Italie – 24 au 27 février 2025
    * Présentation orale et poster – Publication du papier [1]
- German Probability and Statistics Days 2025 (GPSD)
  Dresde, Allemagne – 11 au 14 mars 2025
    * Présentation projet [2]
- 4th ASCAI Workshop in Orsay
  Orsay,France – 18 au 20 juin 2025
    * Présentation projet [2]
- Semaine du pôle MIPS *Université de Montpellier –*
  03/06/2025
    * Présentation courte du projet [2]
- Forty-Second International Conference on Machine Learning
  Vancouver, Canada– 13-19 juillet 2025
    * Poster – Publication du papier sur projet [2]

Implication locale

- Encadrement d'un projet **"Apprentis Chercheurs"** 2024 – 10 séances de 3h – encadrement avec David Metivier d'un binôme collègien-lycéen

- Référent scientifique pour **Éloquensciences 2025** – échanges de mails – accompagnement et conseil à des collégiens/lycéens préparant un concours d'éloquence scientifique

- **Représentant des doctorants** en 2024/2025 – environ 10h – participation à certains conseils d'unité – coordination du séminaire mensuel des (post)-doctorant.e.s – organisation journée des doctorant.e.s

- **Présentation tutoriel** au semdoc de l'IMAG et au semdoc de MISTEA – introduction aux problèmes de bandits

- **Présentation semaine MIPS**– présentation du projet [2]

Conférences prévues en troisième année

- Conférence StatMatAppli 2025
  Fréjus, France – 1 au 5 septembre 2025
  présentation projet [2]

- Rencontres de Statistique Mathématique
  CIRM, France – 15 au 19 décembre 2025

- Journées de Statistique de la SFdS

Formations 1ère année

- 8.10 - Édition scientifique, Open Access et Science ouverte : des clés pour publier
  - 20 mars 2024 - 26 mars 2024 (6 heures)
- 8.7 - MOOC Intégrité scientifique dans les métiers de la recherche
  - juin 2024 (15 heures)
- 4.6 - MOOC Rédiger et publier un article scientifique
  - juin 2024 (20 heures)
- Médiation scientifique dispositif apprentis chercheurs
  - 04 décembre 2023- juin 2024 (30 heures)
- réunion rentrée
  - nov 2023 (2 heures)

Total : 74 heures / 5 modules

Formations 2ème année

- Publication et accès ouvert en mathématiques
  - 03 avril 2025 (2h)
- Premiers secours en santé mentale
  - 23/24 septembre 2024 (14 heures)
- MOOC Machine learning in Python with scikit-learn
  - 01 août 2024-20 août 2024 (30 heures)
- MOOC FORMENSUP Se former pour enseigner dans le supérieur
  - 07 janvier 2025-29 avril 2025 (24 heures)
- 4.1 Zotero
  - 01 octobre 2024-30 novembre 2024 (3 heures)

Total : 146 heures / 10 modules

Enseignements

MCE à l'Université de Montpellier

- 2023/2024 – 63.5h TD (pour le stage agrégation)
  - **Algèbre I** : TD en L1 maths (S1) – systèmes linéaires – 25.5h
  - **Analyse II**: TD en L1 maths (S2) – suites, séries, et DL – 30h
  - **Prépa Agrég** : M2 Maths agreg.– prépa. écrits et oraux– 8h
- 2024/2025 – 63 h TD (pour le stage agrégation)
  - **Mesure et intégration, Fourier**: TD en L3 maths (S1) – théorie de l'intégration selon Lebesgue, transformée de Fourier — 36 h
  - **Remédiation en mathématiques**: TD en L1 bio (S1) – proportionnalité, dilution, mathématiques pour la biologie — 27 h