Meeting Minutes

CMPT 409 F2025 – Dual SPMA Project
Final Report Planning

Project Team

December 13, 2025

Meeting Overview

Course:	CMPT 409 F2025
Project:	Dual SPMA (Convex MDP)
Agenda:	Final Report Planning & Task Division
Report Limit:	8 pages max + Appendix

---

1 Final Report Structure

The agreed-upon report structure is as follows:

1. **Abstract**
2. **Introduction** + Contribution + Motivation
3. **Problem Formulation**
4. **Related Work**
5. **Method**
6. **Experiments + Results** + Practical Tips
7. **Discussion / Conclusion** + Future Work
8. **Acknowledgement** (optional)
9. **References**
10. **Author Contribution**
11. **Appendix** (ablation results, d_π estimation details, etc.)

2 Task Division

Note: Pegah will write boilerplate for Discussion/Conclusion so Ahmed and Shervin can fill in experimental insights.

Section	Assigned To	Page Allocation
Abstract → Method	Danie	∼3.5 pages
Experiments + Results + Practical Tips	Ahmed + Shervin	∼2.5 pages
Discussion + Conclusion	Pegah	∼1.5 pages

Table 1: Writing assignments for the final report

3 Implementation Tasks

3.1 Current Status

- ✓ SPMA implementation – **Working**
- ✓ PPO implementation – **Working**
- Natural Policy Gradient (NPG) – **Not yet implemented**

3.2 Pending Implementation

1. **Implement PPO + SPMA integration**
2. **Implement ground-truth d_π**
   - Reference: Reza's GitHub repository
   - Use known transition matrix for Frozen Lake environment
3. **Implement NPG-PD**
   - Ahmed has pseudo-code ready for implementation

3.3 SPMA Usage Note

To use Reza's SPMA: utilize the **Decision-Aware Actor-Critic** approach. The existing codebase implements SPMA but uses **SMDPO** (not standard SPMA).

4 Experimental Design

4.1 Ahmed's Experiment

Objective: Investigate agent behavior with termination states.

- Make agent navigate toward the hole (termination state)
- Observe: If termination exists, does the agent avoid both holes *and* the goal?

4.2 Occupancy Measure (d_π) Analysis

Current approach: Estimating d_π
Proposed experiment: Compare estimated d_π vs. ground-truth occupancy measure

- Ground-truth available via known transition matrix (Frozen Lake)
- Implementation exists in Reza's GitHub repository

Note: A comprehensive list of experiments will be finalized after this meeting.

5 Method Section Notes

5.1 Convex MDP Framework

CMDP Objective:

$$\min_{d_\pi \in \mathcal{K}} f(d_\pi) = \min_{d_\pi \in \mathcal{K}} \max_{\lambda} [d_\pi \cdot \lambda - f^*(\lambda)] =: L(d_\pi, \lambda)$$

Meta-Algorithm maintains:

- A sequence of policies: $\pi_1, \ldots, \pi_k$
- Their corresponding occupancy measures: $d^1, \ldots, d^k$
- Cost vectors: $\lambda_1, \ldots, \lambda_k$

5.2 Follow-The-Leader (FTL)

Key advantage: No need to compute the Fenchel conjugate.

$$\lambda_k = \arg\max_{\lambda} \sum_{j=1}^{k-1} L(d^j, \lambda) = \arg\max_{\lambda} \left[\lambda \cdot \bar{d}_{k-1} - f^*(\lambda)\right]$$

where $\bar{d}_{k-1} = \frac{1}{k-1} \sum_{j=1}^{k-1} d^j$.
Finding optimal λ: Set gradient to zero:

$$\nabla_\lambda \left[\lambda \cdot \bar{d}_{k-1} - f^*(\lambda)\right] = 0 \quad \Rightarrow \quad \bar{d}_{k-1} - \nabla f^*(\lambda) = 0$$

Important points for the report:

- We use FTL because it eliminates the need to compute the Fenchel conjugate
- We are **not** using Online Mirror Descent
- FTL gives the optimal exact solution in **one step**
- Online convex optimization framework: algorithm chooses W_k

5.3 Entropy-Based Objective

Project motivation (from Sharan): Plug in entropy as the convex function f to see if the agent optimizes entropy (explores the environment) while solving the task.
Method section should include:

- How we obtain optimal λ using the gradient of the objective function
- Derivation for entropy-based objective (reference: slide 27 of presentation)

5.4 Beta Parameter – Exploration vs. Exploitation

Critical insight:

- $\beta = 1$: Pure exploration
- $\beta = 0$: Entropy increases (exploration encouraged)

This is the most important part of the method – the role of β in balancing exploration and task performance.

Person	Task
Danie	Write Abstract through Method sections (∼3.5 pages)
Ahmed	Implement NPG-PD from pseudo-code; run experiments
Shervin	Implement ground-truth d_π; assist with experiments
Ahmed + Shervin	Write Experiments + Results (∼2.5 pages)
Pegah	Draft Discussion/Conclusion boilerplate (∼1.5 pages)

Table 2: Action items summary

6 Action Items

7 References to Include

- Project presentation slides (especially pp. 26–27 for FTL and entropy derivations)
- Reza's GitHub repository (for SPMA and ground-truth d_π implementation)
- Decision-Aware Actor-Critic paper

End of Meeting Minutes