MIT 6.S890 Topics in Multiagent Learning — Fall 2024

Homework 2 - Reference Solution

Released on: Oct. 8, 2024 — **Due on: Oct. 22, 2024, 11:59pm ET**

Problem 1

1.1

Note that ψ is twice differentiable on $\mathcal{X}$ with hessian given by

$$\nabla^2\psi(x) = \operatorname{diag}\Big(\Big[\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n}\Big]\Big).$$

We have that

$$u^\top \nabla^2 f(x) u = \sum_{i=1}^n u_i^2 \cdot \frac{1}{x_i} \geq \sum_{i=1}^n u_i^2 = \|u\|_2^2.$$

where the inequality holds since $x_i \in (0,1)$. Since $\mathcal{X} = \operatorname{relint} \Delta^n$ is convex, together with Lemma 1, we have that ψ is strongly convex with respect to the Euclidean norm on $\operatorname{relint} \Delta^n$.

1.2

Following from Problem 1.1, note that

$$\begin{aligned}
u^\top \nabla^2 f(x) u &= \sum_{i=1}^n u_i^2 \cdot \frac{1}{x_i} \\
&= \Big(\sum_{i=1}^n \frac{u_i^2}{x_i}\Big) \cdot \Big(\sum_{i=1}^n x_i\Big) \\
&\geq \Big(\sum_{i=1}^n \sqrt{\frac{u_i^2}{x_i}} \cdot \sqrt{x_i}\Big)^2 \\
&= \Big(\sum_{i=1}^n |u_i|\Big)^2 \\
&= \|u\|_1^2,
\end{aligned}$$

where the second equality holds since x is a probabilistic distribution and the inequality follows from the Cauchy-Schwarz inequality. Together with Lemma 1, we have that ψ is strongly convex with respect to the ℓ_1 norm on $\operatorname{relint} \Delta^n$.

1.3

By direct calculation, it is easy to see $\nabla\psi(x) = (1 + \log x_1, \ldots, 1 + \log x_n)$.

1.4

Plugging in the results in Problem 1.3, we have that $g_i - (1 + \log x_i^*) = \alpha$ for all $i \in \{1, \dots n\}$. This directly leads to $x_i^* = e^{-1-\alpha} \cdot e^{g_i}$ for all $i \in \{1, \dots n\}$, following from the Lagrange multiplier theorem.

1.5

From Problem 1.4, we now that $x_i^* \propto e^{g_i}$. Since x^* is a probabilistic distribution, we directly have that

$$x_i^* = \frac{e^{g_i}}{\sum_{j=1}^n e^{g_j}}.$$

1.6

Recalling the update rule of predictive FTRL:

$$x^{(t)} \leftarrow \arg\max_{x \in \mathcal{X}} \left\{ \left\langle m^{(t)} + \sum_{\tau=1}^{t-1} g^{(\tau)}, x \right\rangle - \frac{1}{\eta}\psi(x) \right\}.$$

This can be written into the following optimization problem:

$$x^{(t)} \leftarrow \arg\max_{x \in \mathcal{X}} \left\{ \left\langle \eta\Big(m^{(t)} + \sum_{\tau=1}^{t-1} g^{(\tau)}\Big), x \right\rangle - \psi(x) \right\}.$$

Applying Problem 1.5, with $g = \eta\Big(m^{(t)} + \sum_{\tau=1}^{t-1} g^{(\tau)}\Big)$, we have that

$$x^{(t)} \propto \exp\left(\eta\Big(m^{(t)} + \sum_{\tau=1}^{t-1} g^{(\tau)}\Big)\right)$$

Comparing $x^{(t)}$ and $x^{(t-1)}$, we have that

$$x^{(t)} \propto x^{(t-1)} \exp\left(\eta(g_i^{(t-1)} + m_i^{(t)} - m_i^{(t-1)})\right)$$

Since $x^{(t)}$ is a probabilistic distribution, we conclude that

$$x_i^{(t)} = \frac{x_i^{(t-1)} \exp\left(\eta(g_i^{(t-1)} + m_i^{(t)} - m_i^{(t-1)})\right)}{\sum_{j=1}^n x_j^{(t-1)} \exp\left(\eta(g_j^{(t-1)} + m_j^{(t)} - m_j^{(t-1)})\right)}.$$

1.7

The ℓ_1 norm will lead to a better result. This is because, when we analyze the problem under the Euclidean norm, both the dual norm $\|\cdot\|_*$ and the primal norm $\|\cdot\|$ are ℓ_2 norms. However, when we analyze the problem under the ℓ_1 norm, the dual norm $\|\cdot\|_*$ becomes the ℓ_∞ norm, and the primal norm $\|\cdot\|$ is the ℓ_1 norm.

Since we have $\|g\|_2 \geq \|g\|_\infty$ and $-\|x\|_2 \geq -\|x\|_1$, the regret induced by the Euclidean norm is always no less than that induced by the ℓ_1 norm. As a result, the ℓ_1 norm will lead to a better outcome.

1.8

Since negative entropy is always non-positive, we have $\varphi(x^*) \le 0$. Furthermore, it is easy to compute that for $x_i^{(1)} = 1/n$, it holds that $\psi(x^{(1)}) = -\log n$. Combining both statements gives

$$\max_{x^* \in \text{relint } \Delta} \psi(x^*) - \psi(x^{(1)}) \le \log n.$$

By choosing $\|\cdot\|$ as the primal norm and $\|\cdot\|_\infty$ as the dual norm, and substituting the above relation into Proposition 1, with strong convexity given in Problem 1.2, we directly obtain the following result:

$$\text{Reg}^{(T)} \le \frac{\log n}{\eta} + \eta \sum_{t=1}^{T} \|g^{(t)} - m^{(t)}\|_\infty^2 - \frac{1}{8\eta} \sum_{t=2}^{T} \|x^{(t)} - x^{(t-1)}\|_1^2.$$

1.9

Recall the definition of Bregman divergence:

$$D_\psi(z \| z^{(t-1)}) = \psi(z) - \psi(z^{(t-1)}) - \langle \nabla\psi(z^{(t-1)}), z - z^{(t-1)} \rangle.$$

The updating rule for $z^{(t)}$ can be written as:

$$\begin{aligned}
z^{(t)} &= \arg\max_{z \in \mathcal{X}} \left\{ \eta \langle g^{(t-1)}, z \rangle - D_\psi(z \| z^{(t-1)}) \right\} \\
&= \arg\max_{z \in \mathcal{X}} \left\{ \langle \eta g^{(t-1)}, z \rangle - \left(\psi(z) - \psi(z^{(t-1)}) - \langle \nabla\psi(z^{(t-1)}), z - z^{(t-1)} \rangle \right) \right\} \\
&= \arg\max_{z \in \mathcal{X}} \left\{ \left\langle \eta g^{(t-1)} + \nabla\psi(z^{(t-1)}), z \right\rangle - \psi(z) \right\}.
\end{aligned}$$

Plugging in the results from Problem 1.5, we have:

$$z^{(t)} \propto \exp\left(\eta g^{(t-1)} + \nabla\psi(z^{(t-1)}) \right),$$

which simplifies to:

$$z^{(t)} = \exp\left(\eta g^{(t-1)} + 1 + \log z^{(t-1)} \right),$$

according to Problem 1.3. Therefore:

$$z^{(t)} \propto z^{(t-1)} \exp\left(\eta g^{(t-1)} \right),$$

Similarly, it holds that:

$$x^{(t)} \propto z^{(t)} \exp\left(\eta m^{(t)} \right).$$

By combining all the results above, we conclude that the updating rule follows OMWU:

$$x^{(t)} \propto x^{(t-1)} \exp\left(\eta \left(g^{(t-1)} + m^{(t)} - m^{(t-1)} \right) \right),$$

Problem 2

2.1

Notice that the gradient descent step $x^{(t)} \leftarrow x^{(t-1)} - \eta g^{(t)}$ coincides with

$$x^{(t)} \leftarrow \arg\max_{x \in \mathbb{R}^n} \left\{ \langle g^{(t)}, x \rangle - \frac{1}{\eta} D_\psi(x \| x^{(t-1)}) \right\}$$

when choosing $\psi(x) = \frac{x^2}{2}$. This indicates that the OGDA steps align with Predictive OMD with $\psi(x) = \frac{x^2}{2}$. This allows us to apply Proposition 1.

Furthermore, from $\nabla^2 \psi(x) = I$, we see that ψ is strongly convex with respect to $\| \cdot \|_2$. Applying Proposition 1 with the ℓ_2-norm gives

$$\mathrm{Reg}_1^{(T)} \leq \frac{\|\mathcal{X}\|_2^2}{2\eta} + \eta \sum_{t=2}^{T} \|Ay^{(t)} - Ay^{(t-1)}\|_2^2 - \frac{1}{8\eta} \sum_{t=2}^{T} \|x^{(t)} - x^{(t-1)}\|_2^2.$$

Similarly, we can get

$$\mathrm{Reg}_2^{(T)} \leq \frac{\|\mathcal{Y}\|_2^2}{2\eta} + \eta \sum_{t=2}^{T} \|A^\top x^{(t)} - A^\top x^{(t-1)}\|_2^2 - \frac{1}{8\eta} \sum_{t=2}^{T} \|y^{(t)} - y^{(t-1)}\|_2^2.$$

2.2

Denote $\overline{x} = \frac{1}{T} \sum_{t=1}^{T} x^{(t)}$ and similarly for $\overline{y}$. Notice that

$$\mathrm{Reg}_1^{(T)} + \mathrm{Reg}_2^{(T)} = \max_{x^* \in \mathcal{X}} \sum_{t=1}^{T} (x^*)^\top A y^{(t)} - \sum_{t=1}^{T} (x^{(t)})^\top A y^{(t)} + \sum_{t=1}^{T} (x^{(t)})^\top A y^{(t)} - \min_{y^* \in \mathcal{Y}} \sum_{t=1}^{T} (x^{(t)})^\top A y^*$$

$$= \max_{x^* \in \mathcal{X}} \sum_{t=1}^{T} (x^*)^\top A y^{(t)} - \min_{y^* \in \mathcal{Y}} \sum_{t=1}^{T} (x^{(t)})^\top A y^*$$

$$\geq \sum_{t=1}^{T} (\overline{x})^\top A y^{(t)} - \sum_{t=1}^{T} (x^{(t)})^\top A \overline{y}$$

$$= 0.$$

Plugging in the results from Problem 2.1, it holds that

$$\frac{\|\mathcal{X}\|_2^2}{2\eta} + \eta \sum_{t=2}^{T} \|Ay^{(t)} - Ay^{(t-1)}\|_2^2 - \frac{1}{8\eta} \sum_{t=2}^{T} \|x^{(t)} - x^{(t-1)}\|_2^2$$

$$+ \frac{\|\mathcal{Y}\|_2^2}{2\eta} + \eta \sum_{t=2}^{T} \|A^\top x^{(t)} - A^\top x^{(t-1)}\|_2^2 - \frac{1}{8\eta} \sum_{t=2}^{T} \|y^{(t)} - y^{(t-1)}\|_2^2 \geq 0.$$

Using $\|Ay^{(t)} - Ay^{(t-1)}\|_2 \geq \|A\|_2\|y^{(t)} - y^{(t-1)}\|_2$ from the definition, rearranging the above inequality gives

$$\frac{\|\mathcal{X}\|_2^2 + \|\mathcal{Y}\|_2^2}{2\eta} + \Big(\eta\|A\|_2^2 - \frac{1}{8\eta}\Big)\Big(\sum_{t=2}^{T}\|x^{(t)} - x^{(t-1)}\|_2^2 + \|y^{(t)} - y^{(t-1)}\|_2^2\Big) \geq 0.$$

As long as $\eta \leq 1/(4\|A\|_2)$, it holds that $\eta\|A\|_2^2 - 1/(8\eta) \leq -1/(16\eta)$. In this case, it holds that

$$\sum_{t=2}^{T}\big(\|x^{(t)} - x^{(t-1)}\|_2^2 + \|y^{(t)} - y^{(t-1)}\|_2^2\big) \leq 8\big(\|\mathcal{X}\|_2^2 + \|\mathcal{Y}\|_2^2\big)^2.$$

According to the Cauchy-Schwarz inequality, it holds that

$$\Big(\sum_{t=2}^{T}\big(\|x^{(t)} - x^{(t-1)}\|_2 + \|y^{(t)} - y^{(t-1)}\|_2\big)\Big)^2 \leq \sum_{t=2}^{T}(1^2 + 1^2)\cdot\sum_{t=2}^{T}\big(\|x^{(t)} - x^{(t-1)}\|_2^2 + \|y^{(t)} - y^{(t-1)}\|_2^2\big).$$

Rearranging the terms concludes that

$$\sum_{t=2}^{T}\big(\|x^{(t)} - x^{(t-1)}\|_2 + \|y^{(t)} - y^{(t-1)}\|_2\big) \leq 4\sqrt{T}\big(\|\mathcal{X}\|_2 + \|\mathcal{Y}\|_2\big).$$

2.3

By inspecting the proximal descent inequality for $x^{(t)} \leftarrow \Pi_{\mathcal{X}}\big(x^{(t-1)} + 2\eta Ay^{(t-1)} - \eta Ay^{(t-2)}\big)$ at $x = x^*$, we have

$$\langle x^{(t-1)} + 2\eta Ay^{(t-1)} - \eta Ay^{(t-2)} - x^{(t)}, x^* - x^{(t)}\rangle \leq 0.$$

Notice that

$$\begin{aligned}\langle x^{(t-1)} + 2\eta Ay^{(t-1)} - \eta Ay^{(t-2)} - x^{(t)}, x^* - x^{(t)}\rangle &= \langle \eta Ay^{(t-1)}, x^* - x^{(t-1)}\rangle + \langle \eta Ay^{(t-1)}, x^{(t-1)} - x^{(t)}\rangle \\ &\quad + \langle x^{(t-1)} - x^{(t)}, x^* - x^{(t)}\rangle + \langle \eta Ay^{(t-1)} - \eta Ay^{(t-2)}, x^* - x^{(t)}\rangle \leq 0.\end{aligned} \tag{1}$$

According to the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}|\langle \eta Ay^{(t-1)}, x^{(t-1)} - x^{(t)}\rangle| &\leq \eta\|A\|_2\|y^{(t-1)}\|_2\|x^{(t-1)} - x^{(t)}\|_2 \leq \|x^{(t-1)} - x^{(t)}\|_2\|\mathcal{Y}\|_2 \\ |\langle x^{(t-1)} - x^{(t)}, x^* - x^{(t)}\rangle| &\leq \|x^{(t-1)} - x^{(t)}\|_2\|x^* - x^{(t)}\|_2 \leq \|x^{(t-1)} - x^{(t)}\|_2\|\mathcal{X}\|_2, \\ |\langle \eta Ay^{(t-1)} - \eta Ay^{(t-2)}, x^* - x^{(t)}\rangle| &\leq \eta\|A\|_2\|y^{(t-1)} - y^{(t-2)}\|_2\|x^* - x^{(t)}\|_2 \leq \|y^{(t-1)} - y^{(t-2)}\|_2\|\mathcal{X}\|_2,\end{aligned}$$

where the second inequalities are given by $\eta \leq 1/(4\|A\|_2)$ and the definitions of $\|\mathcal{X}\|_2$ and $\|\mathcal{Y}\|_2$. When $\eta \leq 1/(4\|A\|_2)$, summing up the above inequalities gives

$$\begin{aligned}&|\langle \eta Ay^{(t-1)}, x^{(t-1)} - x^{(t)}\rangle + \langle x^{(t-1)} - x^{(t)}, x^* - x^{(t)}\rangle + \langle \eta Ay^{(t-1)} - \eta Ay^{(t-2)}, x^* - x^{(t)}\rangle| \\ &\leq \big(\|x^{(t-1)} - x^{(t)}\|_2 + \|y^{(t-1)} - y^{(t-2)}\|_2\big)\big(\|\mathcal{X}\|_2 + \|\mathcal{Y}\|_2\big).\end{aligned}$$

Plugging the above results into (1) gives

$$\langle \eta Ay^{(t-1)}, x^* - x^{(t-1)}\rangle \leq \big(\|x^{(t-1)} - x^{(t)}\|_2 + \|y^{(t-1)} - y^{(t-2)}\|_2\big)\big(\|\mathcal{X}\|_2 + \|\mathcal{Y}\|_2\big).$$

Summing over t yields

$$\begin{aligned}\max_{x^*\in\mathcal{X}}\sum_{t=1}^{T}(x^*)^\top Ay^{(t)} - \sum_{t=1}^{T}(x^{(t)})^\top Ay^{(t)} &\leq \sum_{t=2}^{T+1}\frac{1}{\eta}\big(\|x^{(t-1)} - x^{(t)}\|_2 + \|y^{(t-1)} - y^{(t-2)}\|_2\big)\big(\|\mathcal{X}\|_2 + \|\mathcal{Y}\|_2\big) \\ &\leq \frac{4\sqrt{T+1}}{\eta}\big(\|\mathcal{X}\|_2 + \|\mathcal{Y}\|_2\big)^2,\end{aligned}$$

where the last inequality follows from Problem 2.2. Similarly, we have

$$\sum_{t=1}^{T}(x^{(t)})^\top Ay^{(t)} - \min_{y^*\in\mathcal{X}}\sum_{t=1}^{T}(x^{(t)})^\top Ay^* \le \frac{4\sqrt{T+1}}{\eta}\big(\|\mathcal{X}\|_2 + \|\mathcal{Y}\|_2\big)^2.$$

Adding both terms gives

$$\sum_{t=1}^{T}\Big((x^*)^\top Ay^{(t)} - (x^{(t)})^\top Ay^*\Big) \le \frac{8\sqrt{T+1}}{\eta}\big(\|\mathcal{X}\|_2 + \|\mathcal{Y}\|_2\big)^2.$$

This indicates that there must be some t such that

$$(x^*)^\top Ay^{(t)} - (x^{(t)})^\top Ay^* \le \mathcal{O}\Big(\frac{1}{\sqrt{T}}(\|\mathcal{X}\|_2 + \|\mathcal{Y}\|_2)^2\|A\|_2\Big).$$