1

[70分] 游戏的OOMD及其实现

我们考虑一个三人游戏，其中三名玩家的策略分别用 x、y 和 z 表示。该游戏重复进行 T 轮。在第 t 轮中，三名玩家同时提交策略 (x_t, y_t, z_t) 后，每名玩家的个体成本根据各自的成本函数计算。例如，x-玩家的成本函数记为 $G_x : (x, y, z) \to \mathbb{R}$，$y$-玩家的成本函数为 $G_y : (x, y, z) \to \mathbb{R}$，而 z-玩家的成本函数为 $G_z : (x, y, z) \to \mathbb{R}$。

定义三名玩家的总成本为 $G(\cdot) \triangleq G_x(\cdot) + G_y(\cdot) + G_z(\cdot)$。理想情况下，我们希望所有玩家合作以实现总成本的最小值 $G_{\text{Min}} \triangleq \min_{x,y,z} G(x, y, z)$。然而，更常见的情况是，每名玩家自私地尝试在游戏中最小化自己的成本。在本题中，我们关注一个量：所有玩家的平均总成本，即

$$G_T \triangleq \frac{1}{T} \sum_{t=1}^{T} G(x_t, y_t, z_t),$$

并研究以下问题：

什么条件可以保证 G_T 不会比 G_{Min} 差太多?

为简单起见，我们聚焦于如下定义的平滑游戏。

假设 1（平滑游戏）

对于游戏 G，若存在参数 $\lambda > 0$ 和 $\mu < 1$，使得存在一个策略组合 (x^*, y^*, z^*)，满足对任意策略 (x, y, z)：

$$G_x(x^*, y, z) + G_y(x, y^*, z) + G_z(x, y, z^*) \leq \lambda \cdot G_{\text{Min}} + \mu \cdot G(x, y, z), \tag{1.1}$$

则称该游戏为一个 (λ, μ)-平滑游戏。

直观上，在平滑游戏中，即使其他玩家的策略是任意的，只要一个玩家采用其最优策略，她仍然表现良好。

以下问题中定义：

$$\mathrm{Reg}_T^x \triangleq \max_x \sum_{t=1}^{T} \big(G_x(x_t, y_t, z_t) - G_x(x, y_t, z_t)\big),$$

$$\mathrm{Reg}_T^y \triangleq \max_y \sum_{t=1}^{T} \big(G_y(x_t, y_t, z_t) - G_y(x_t, y, z_t)\big),$$

$$\mathrm{Reg}_T^z \text{ 类似定义。}$$

问题

1. **[10分]** 在假设（1.1）的条件下，证明以下结论：

$$\frac{1}{T}\sum_{t=1}^{T} G(x_t, y_t, z_t) \le \frac{\lambda}{1-\mu} G_{\mathrm{Min}} + \frac{1}{(1-\mu)T}\left(\mathrm{Reg}_T^x + \mathrm{Reg}_T^y + \mathrm{Reg}_T^z\right),$$

这意味着，在次线性遗憾的条件下，可以保证

$$\lim_{T\to\infty} G_T \le \frac{\lambda}{1-\mu} G_{\mathrm{Min}},$$

从而回答了上述问题。

2. **[10分]** 在课程中，我们学习了乐观在线镜像下降（OOMD）可以对二人零和游戏实现快速收敛率。现在我们考虑本题中的三人游戏，并假设每名玩家从 Δ^d 中选择混合策略。每名玩家有其自己的张量表示成本函数，即 x-玩家有 $G_x \in [0,1]^{d\times d\times d}$，$y$-玩家有 $G_y \in [0,1]^{d\times d\times d}$，$z$-玩家有 $G_z \in [0,1]^{d\times d\times d}$。对于张量 $G \in \mathbb{R}^{d\times d\times d}$ 和三个向量 $x,y,z \in \mathbb{R}^d$，我们缩写

$$G[x,y,z] = \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} G_{i,j,k}x_i y_j z_k\text{。}$$

则三名玩家的成本函数分别定义为：

$$G_x(x,y,z) \triangleq G_x[x,y,z]\text{，}\ G_y(x,y,z) \triangleq G_y[x,y,z]\text{，}\ G_z(x,y,z) \triangleq G_z[x,y,z]\text{。}$$

在第 t 轮中，三名玩家提交策略 x_t, y_t, z_t 后，可以观察到各自成本函数的梯度：x-玩家观察到 $\nabla_x G_x(x_t,y_t,z_t)$，$y$-玩家观察到 $\nabla_y G_y(x_t,y_t,z_t)$，$z$-玩家观察到 $\nabla_z G_z(x_t,y_t,z_t)$。

设计一种每名玩家使用 NE-熵的 OOMD 算法，并证明以下结果：

$$\mathrm{Reg}_T^x \lesssim \frac{1}{\eta_x} + \eta_x \sum_{t=2}^{T} \|\nabla_x G_x(x_t,y_t,z_t) - \nabla_x G_x(x_{t-1},y_{t-1},z_t)\|_\infty^2 - \frac{1}{\eta_x}\sum_{t=2}^{T}\|x_t - x_{t-1}\|_1^2,$$

类似地，证明 Reg_T^y 和 Reg_T^z 的结果。

3. **[10分]** 证明以下不等式：

$$\|\nabla_x G_x(x_t,y_t,z_t) - \nabla_x G_x(x_t,y_{t-1},z_t)\|_\infty^2 \le \|y_t - y_{t-1}\|_1^2\text{。}$$

然后设计步长 η_x, η_y, η_z，并证明以下保证：

$$\mathrm{Reg}_T^x + \mathrm{Reg}_T^y + \mathrm{Reg}_T^z \le O(1)\text{。}$$

4. **[40分]** 实现 OMD 和 OOMD 算法来解决上述游戏，并附上两种算法的平均总成本曲线对比图。详细指导请参考文件 `AOpt-Lab2/AOpt-Lab2.ipynb`。提交文件 `AOpt-Lab2.ipynb` 作为作业的一部分，确保结果可验证。

2

[50分] 加速复合优化

考虑以下定义在有界域上的复合优化问题：

$$\min_{x \in X} F(x) \triangleq f(x) + h(x),$$

其中 $f(\cdot)$ 和 $h(\cdot)$ 都是凸函数，且 $f(\cdot)$ 关于 $\|\cdot\|_2$ 是 L-平滑的，而 $h(\cdot)$ 则不是。假设定义域的直径是有界的，即

$$\sup_{x,y \in X} \|x - y\|_2 \leq D。$$

在课堂中，我们学习了一种简单的基于乐观在线学习框架的加速平滑凸优化方法。类似的方法是否可以应用于复合优化？

更具体地，我们考虑以下加权的在线到批处理转换公式：

$$x_t = \frac{1}{A_t} \sum_{s=1}^{t} \alpha_s x_s, \qquad A_t = \sum_{s=1}^{t} \alpha_s, \qquad \alpha_t > 0, \ \forall t \in [T]。 \tag{2.1}$$

问题

1. **[10分]** 尝试证明公式 (2.1) 保证以下收敛：

$$F(x_T) - F(x^*) \leq \frac{\sum_{t=1}^{T} \left(\langle \alpha_t \nabla f(x_t), x_t - x^* \rangle + \alpha_t h(x_t) - \alpha_t h(x^*) \right)}{A_T}。 \tag{2.2}$$

2. **[20分]** 公式 (2.2) 允许我们将离线优化问题转化为在线优化问题。定义在线函数为

$$F_t(x) \triangleq f_t(x) + h_t(x),$$

其中 $f_t(x) \triangleq \langle \alpha_t \nabla f(x_t), x \rangle$，$h_t(x) \triangleq \alpha_t h(x)$。为此，我们设计以下乐观在线学习算法：

$$x_t = \arg\min_{x \in X} \left\{ \eta \big(\langle M_t, x \rangle + h_t(x) \big) + \frac{1}{2} \|x - \hat{x}_t\|_2^2 \right\}, \tag{2.3}$$

$$\hat{x}_{t+1} = \arg\min_{x \in X} \left\{ \eta \big(\langle \nabla f_t(x_t), x \rangle + h_t(x) \big) + \frac{1}{2} \|x - \hat{x}_t\|_2^2 \right\}。 \tag{2.4}$$

(2.i) **[10分]** 证明更新公式 (2.3) 和 (2.4) 的稳定性属性：

$$\|x_t - \hat{x}_{t+1}\|_2 \leq \eta \|\nabla f_t(x_t) - M_t\|_2。$$

(2.ii) **[10分]** 证明公式 (2.4) 的 Bregman 近端不等式：

$$\eta \langle \nabla f_t(x_t) + \nabla h_t(\hat{x}_{t+1}), \hat{x}_{t+1} - x^* \rangle \leq \frac{1}{2} \|x^* - \hat{x}_t\|_2^2 - \frac{1}{2} \|x^* - \hat{x}_{t+1}\|_2^2 - \frac{1}{2} \|\hat{x}_{t+1} - \hat{x}_t\|_2^2。$$

3. **[10分]** 证明以下结果：

$$\sum_{t=1}^{T} \big(F_t(x_t) - F_t(x^*) \big) \leq \frac{\|u - \hat{x}_1\|_2^2}{2\eta} + \eta \sum_{t=1}^{T} \|\nabla f_t(x_t) - M_t\|_2^2 - \frac{1}{4\eta} \sum_{t=2}^{T} \|x_t - x_{t-1}\|_2^2。$$

4. **[10分]** 设计权重 α_t、步长 η 和乐观值 M_t，并证明以下结果：

$$F(x_T) - F(x^*) \leq O\left(L \cdot \frac{1}{T^2} \right)。$$

3

[50分] 两点带宽凸优化

我们考虑具有两点反馈的带宽凸优化（Bandit Convex Optimization，简称 BCO）问题。在每一轮 t，在线学习者可以查询两个点 x_t^1 和 x_t^2，它们属于 $X \subseteq \mathbb{R}^d$，并观察函数值 $f_t(x_t^1)$ 和 $f_t(x_t^2)$。假设在线函数 $\{f_t\}_{t=1}^T$ 是 G-Lipschitz 连续的。目标是最小化以下期望遗憾：

$$E[\mathrm{Reg}_T] = E\left[\sum_{t=1}^{T} \frac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big)\right] - \min_{x \in X} \sum_{t=1}^{T} f_t(x)。 \tag{3.1}$$

基于两点反馈，我们希望改进课堂中介绍的带宽梯度下降（Bandit Gradient Descent）算法。在每一轮，我们利用观测信息估计梯度 $\tilde{g}_t$，然后用它执行梯度下降：

$$y_{t+1} = \Pi_{(1-\alpha)X}[y_t - \eta \tilde{g}_t],$$

其中 $\Pi_{(1-\alpha)X}$ 表示投影到缩小的集合 $(1-\alpha)X$ 上。

问题

1. **[10分]** 梯度估计的一个基本思路是：首先从单位球面 $S \triangleq \{x \in \mathbb{R}^d \mid \|x\|_2 = 1\}$ 中随机采样一个单位向量 s_t，并提交以下查询：

$$x_t^1 = y_t + \delta s_t, \qquad x_t^2 = y_t - \delta s_t。$$

然后利用观测到的值 $f_t(x_t^1)$ 和 $f_t(x_t^2)$ 构造梯度估计器：

$$\tilde{g}_t = \frac{d}{2\delta}\big(f_t(y_t + \delta s_t) - f_t(y_t - \delta s_t)\big) s_t。$$

(1.i) **[5分]** 请证明该梯度估计器满足无偏性条件：

$$\hat{f}_t(y_t) = E_{v \in B}[f_t(y_t + \delta v)], \qquad E_{s \in S}[\tilde{g}_t] = \nabla \hat{f}(y_t),$$

其中 $B = \{x \in \mathbb{R}^d \mid \|x\|_2 \le 1\}$ 是单位球。

(1.ii) **[5分]** 请证明该梯度估计器的范数有界：$\|\tilde{g}_t\|_2 \le Gd$。

2. **[15分]** 现在我们分析改进的 BGD 算法的遗憾。基于问题 (1) 的分析，我们知道梯度估计器满足 $E_{s\in S}[\tilde{g}_t] = \nabla \hat{f}_t(y_t)$。这意味着改进的 BGD 算法实际上是在函数 $\hat{f}_t$ 上（好像具有完整信息）对集合 $(1-\alpha)X$ 执行在线梯度下降（OGD）。因此，在分析遗憾 (3.1) 时，我们希望将其与 $\hat{f}_t$ 上的 OGD 遗憾联系起来，即：

$$\sum_{t=1}^{T} \hat{f}_t(x_t) - \sum_{t=1}^{T} \hat{f}_t((1-\alpha)x)。$$

我们首先考虑单轮遗憾：

$$\frac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big) - f_t(x),$$

可以分解为五个部分，每部分捕捉了算法的具体影响：

$$\frac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big) - f_t(x) = \frac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big) - f_t(y_t)\underbrace{\qquad}_{\text{(a)}} + f_t(y_t) - \hat{f}_t(y_t)\underbrace{\qquad}_{\text{(b)}} + \hat{f}_t(y_t) - \hat{f}_t((1-\alpha)x)\underbrace{\qquad}_{\text{(c)}} + \hat{f}_t((1-\alpha)x) - f_t((1-\alpha)x$$

(2.i) **[5分]** 请解释这 5 项的含义。每项分别代表什么具体影响？

(2.ii) **[10分]** 假设 $\|x\|_2 \le D$，并且 f_t 是 G-Lipschitz 连续的，利用上述分解证明以下遗憾界成立：

$$\sum_{t=1}^{T} \frac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big) - \sum_{t=1}^{T} f_t(x) \le \sum_{t=1}^{T} \hat{f}_t(y_t) - \sum_{t=1}^{T} \hat{f}_t((1-\alpha)x) + 3TG\delta + TGD\alpha。$$

3. **[10分]** 定义 $h_t(x) \triangleq \hat{f}_t(x) + \big(\tilde{g}_t - \nabla \hat{f}_t(x_t)\big)^\top x$。显然，$h_t(x)$ 是凸函数，且 $\nabla h_t(x_t) = \tilde{g}_t$，这意味着改进的 BGD 算法是在集合 $(1-\alpha)$ 对函数 h_t 执行确定性 OGD。

(3.i) **[5分]** 请证明：

$$\sum_{t=1}^{T} h_t(x_t) - \sum_{t=1}^{T} h_t((1-\alpha)x) \le \frac{D^2}{\eta T} + G^2 d^2 \sum_{t=1}^{T} \eta_t。$$

(3.ii) **[5分]** 基于 (2.ii) 和 (3.i) 的结果，进一步证明以下结论：（提示：$E[h_t(x)] = \hat{f}_t(x)$）

$$E\left[\sum_{t=1}^{T} \frac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big) - \sum_{t=1}^{T} f_t(x)\right] \le \frac{D^2}{\eta T} + G^2 d^2 \sum_{t=1}^{T} \eta_t + 3TG\delta + TGD\alpha。$$

4. **[10分]** 假设 $rB \subseteq X \subseteq DB$。请设计学习率 η_t、δ、α，确保每一步 $x_t^1, x_t^2 \in X$，并证明以下遗憾界：

$$E\left[\sum_{t=1}^{T} \frac{1}{2}\big(f_t(x_t^1) + f_t(x_t^2)\big) - \sum_{t=1}^{T} f_t(x)\right] \le 3DGd\sqrt{T} + 3G + \frac{GD}{r}。$$

5. **[5分]** 注意，在课程中，我们介绍了具有单点反馈的 BCO，其遗憾界为 $O(T^{3/4})$。请解释两点反馈如何比单点反馈更具优势，从而使我们可以实现 $O(\sqrt{T})$ 的遗憾界。

4

[50分] 高效的随机逻辑带宽问题

我们考虑随机逻辑带宽问题（Stochastic Logistic Bandits，简称 LogB）。奖励函数满足

$$r_t = \mu(X_t^\top \theta^*) + \eta_t,$$

其中 $\mu(z) = \frac{1}{1+\exp(-z)}$，噪声 η_t 遵循伯努利分布，即 $P(r_t = 1|X_t) = \mu(X_t^\top \theta^*)$ 和 $P(r_t = 0|X_t) = 1 - \mu(X_t^\top \theta^*)$。学习者的目标是最小化以下遗憾：

$$\mathrm{Reg}_T = \max_{x \in X} \sum_{t=1}^{T} \mu(x^\top \theta^*) - \sum_{t=1}^{T} \mu(X_t^\top \theta^*)。$$

为简化分析，假设可行集合 X 和未知参数 θ^* 是有界的：对于所有 $X \in X \subseteq \mathbb{R}^d$，$\|X\|_2 \leq 1$，且 $\|\theta^*\|_2 \leq S$。此外，$\mu(z)$ 在 $z \in [-S, S]$ 上是 L-Lipschitz 连续的，并且其导数满足 $\inf_{z \in (-S,S)} \mu'(z) = \kappa$。

为估计 LogB 中的未知参数 θ^*，一种常见的方法是用最大似然估计（MLE）替代 LinUCB 中的最小二乘估计器，并最小化负对数似然：

$$\hat{\theta}_t = \arg\min_{\|\theta\| \leq S} \sum_{s=1}^{t} \ell_s(\theta),$$

其中

$$-\ell_s(\theta) = r_s \log \mu(X_s^\top \theta) + (1 - r_s) \log(1 - \mu(X_s^\top \theta))。$$

然而，MLE 在这种情况下带来了显著的计算挑战，因为它不支持在线更新。因此，每一轮决策都需要重新计算所有历史数据，这导致了可扩展性问题。

为了解决这一问题，最近的研究建议将问题建模为一个在线凸优化（OCO）问题，通过逐步将损失函数 $\{\ell_s(\theta)\}_{s=1}^t$ 输入在线学习器 B。基于每一轮的输出 θ_s，我们构造一个虚拟的线性奖励 $z_s = X_s^\top \theta_s$。利用这些奖励，我们基于历史数据对 (X_s, z_s) 进行最小二乘估计，得到参数估计器 $\hat{\theta}_t$。这种方法允许在每轮中高效地进行在线更新。

接下来我们将证明此方法可以实现可靠的估计误差保证以及良好的遗憾界。

问题

1. **[10分]** 假设在线学习器 B 满足以下遗憾界：对于所有 $\|\theta\|_2 \le S$ 和 $t \ge 1$，

$$\sum_{s=1}^{t} \ell_s(\theta_s) - \ell_s(\theta) \le B_t。$$

为了为在线估计器构建置信区间，需要将参数估计误差与遗憾 B_t 联系起来。请证明以下结果：（提示：对 $\ell_s(\theta)$ 使用泰勒展开）

$$\sum_{s=1}^{t} \left(X_s^\top(\theta_s - \theta^*)\right)^2 \le \frac{2}{\kappa} B_t + \frac{2}{\kappa} \sum_{s=1}^{t} \eta_s \left(X_s^\top(\theta_s - \theta^*)\right)。 \tag{4.1}$$

2. **[15分]** 对于公式 (4.1) 中的项 $\sum_{s=1}^{t} \eta_s \left(X_s^\top(\theta_s - \theta^*)\right)$，其结构类似于课程中介绍的自归一化浓度不等式。我们希望应用自归一化浓度不等式来处理这一项。

(2.i) **[5分]** 注意，自归一化浓度不等式要求噪声是次高斯分布的。请证明噪声 $\eta_t = r_t - \mu(X_t^\top \theta^*)$ 是 R-次高斯的，其中 $R \le \frac{1}{2}$。（提示：使用 Hoeffding 引理）

(2.ii) **[5分]** 现在我们知道噪声 η_t 是 R-次高斯的，请证明：对于任意 $t \in [T]$，以下结果以至少 $1-\delta$ 的概率成立：（提示：将自归一化浓度不等式转换为一维形式）

$$\sum_{s=1}^{t} \eta_s \left(X_s^\top(\theta_s - \theta^*)\right) \le R \sqrt{\left(2 + 2\sum_{s=1}^{t} \left(X_s^\top(\theta_s - \theta^*)\right)^2\right) \cdot \log\left(\frac{1}{\delta}\sqrt{1 + \sum_{s=1}^{t} \left(X_s^\top(\theta_s - \theta^*)\right)^2}\right)}。$$

(2.iii) **[5分]** 将上述不等式代入 (4.1)，进一步证明：对于任意 $t \in [T]$，以下结果以至少 $1-\delta$ 的概率成立：（提示：令 $q=\sqrt{1+\sum_{s=1}^{t}\left(X_s^\top(\theta_s-\theta^*)\right)^2}$）

$$\sum_{s=1}^{t}\left(X_s^\top(\theta_s-\theta^*)\right)^2 \le \beta_t' \equiv 1+\frac{4}{\kappa}B_t+\frac{8R^2}{\kappa^2}\log\left(\frac{1}{\delta}\sqrt{4+\frac{8}{\kappa}B_t+\frac{64R^4}{\kappa^4}\cdot\frac{4}{\delta^2}}\right)。\tag{4.2}$$

3. **[15分]** 令 $z_s = X_s^\top \theta_s$ 为第 s 轮的虚拟奖励。然后我们使用历史数据对 (X_s, z_s) 进行最小二乘估计，得到估计器：

$$\hat{\theta}_t = \arg\min_{\theta}\left\{\frac{\lambda}{2}\|\theta\|_2^2+\sum_{s=1}^{t}\left(z_s-X_s^\top\theta\right)^2\right\}。\tag{4.3}$$

基于公式 (4.2) 和估计器 (4.3)，设 $V_t=\lambda I_d+\sum_{s=1}^{t}X_sX_s^\top$。请证明：对于任意 $t\in[T]$，以下结果以至少 $1-\delta$ 的概率成立：（提示：使用最小二乘的闭式解形式）

$$\|\theta^*-\hat{\theta}_t\|_{V_t}^2 \le \beta_t \equiv \lambda S^2+\beta_t'-\left(\lambda\|\hat{\theta}_t\|_2^2+\sum_{s=1}^{t}\left(z_s-X_s^\top\hat{\theta}_t\right)^2\right)。\tag{4.4}$$

4. **[10分]** 基于公式 (4.4) 中的 UCB β_t，设计以下 UCB 选择准则：

$$X_t=\arg\max_{x\in X}\left\{\langle x,\hat{\theta}_{t-1}\rangle+\beta_{t-1}\|x\|_{V_{t-1}^{-1}}\right\}。\tag{4.5}$$

对于参数估计器 (4.3) 和选择准则 (4.5)，请证明：以下遗憾界以至少 $1-2\delta$ 的概率成立：（提示：$\mu(x^\top\theta^*)-\mu(X_t^\top\theta^*)\le L(x-X_t)^\top\theta^*$）

$$\mathrm{Reg}_T=\max_{x\in X}\sum_{t=1}^{T}\mu(x^\top\theta^*)-\sum_{t=1}^{T}\mu(X_t^\top\theta^*)=O\left(L\sqrt{\beta_T dT\log T}\right)。$$

5

[50分] 利用辅助信息的在线回归

本题探讨如何结合可用的辅助信息来改进在线回归的遗憾界。在每一轮 $t \in [T]$ 中，在线学习者提交一个决策 $x_t \in X \subseteq \mathbb{R}^d$，在线函数定义为：

$$f_t(x_t) = \frac{1}{2}(x_t^\top \psi_t - y_t)^2,$$

其中 $\psi_t \in \Psi \subseteq \mathbb{R}^d$ 表示特征，$y_t \in Y \subseteq \mathbb{R}$ 表示对应的标签。目标是对任意 $u \in X$ 最小化以下遗憾：

$$\mathrm{Reg}_T(u) \triangleq \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(u)。$$

假设对于所有 $t \in [T]$，$y_t \in [-Y, Y]$ 成立。

问题

1. **[5分]** 证明在线函数 $f_t(x)$ 是 α-指数凹的，其中：

$$\alpha = \min_{x \in X} \left\{ \frac{1}{(x^\top \psi_t - y_t)^2} \right\}。$$

基于上述结果，当假设域的直径和梯度范数有界时，使用在线牛顿法（Online Newton Step）可以实现 $O\big(\max_{x \in X, t \in [T]} \{(x^\top \psi_t - y_t)^2\} \cdot d \log T\big)$ 的遗憾。然而，当域 X 或特征空间 Ψ 很大时（即指数凹参数 α 很小），该遗憾界可能并不理想。

下面，我们将利用该问题中的可用信息来解决这一问题。事实上，在在线回归中，特征 ψ_t 在提交决策 x_t 之前是已知的（而标签 y_t 显然是未知的）。这意味着学习者可以在更新之前部分了解 $f_t(\cdot)$ 的信息。因此，我们使用乐观在线镜像下降（Optimistic Online Mirror Descent，OOMD）来利用这些辅助信息，并将其视为"提示"：

$$x_t = \arg\min_{x \in \mathbb{R}^d} \left\{ \frac{1}{2}(x^\top \psi_t)^2 + \frac{1}{2}\|x - \hat{x}_t\|_{A_{t-1}}^2 \right\}, \tag{5.1}$$

$$\hat{x}_{t+1} = \arg\min_{x \in \mathbb{R}^d} \left\{ \frac{1}{2}(x^\top \psi_t - y_t)^2 + \frac{1}{2}\|x - \hat{x}_t\|_{A_{t-1}}^2 \right\},$$

其中 $A_{t-1} = \lambda I + \sum_{s=1}^{t-1} \psi_s \psi_s^\top$ 是正则化协方差矩阵。

在 (5.1) 中，我们考虑了 $X = \mathbb{R}^d$ 的困难场景（注意，此时 α 可能接近 0）。为简化后续的分析，我们定义 $h_t(x) = \frac{1}{2}(x^\top \psi_t)^2$，并将其视为通过假设 $y_t = 0$ 而得到的 $f_t(x)$ 的"预测"。

问题

2. **[20分]** 证明在线函数 $f_t(x)$ 的以下性质：

$$f_t(x) - f_t(y) = \langle \nabla f_t(x), x - y \rangle - \frac{1}{2}\|x - y\|_{\psi_t \psi_t^\top}^2。$$

基于该性质，尝试证明算法 (5.1) 的以下中间结果：

$$\mathrm{Reg}_T(u) \le \sum_{t=1}^T \frac{1}{2}\left(\|u - \hat{x}_t\|_{A_{t-1}}^2 - \|u - \hat{x}_{t+1}\|_{A_{t-1}}^2 - \|u - \hat{x}_{t+1}\|_{\psi_t \psi_t^\top}^2 \right) + \sum_{t=1}^T \left(f_t(x_t) - f_t(\hat{x}_{t+1}) + h_t(\hat{x}_{t+1}) - h_t(x_t) \right)$$

3. **[10分]** 证明以下技术性结果：

$$f_t(x_t) - f_t(\hat{x}_{t+1}) + h_t(\hat{x}_{t+1}) - h_t(x_t) = y_t^2 \psi_t^\top A_t^{-1} \psi_t。$$

4. **[15分]** 证明算法 (5.1) 的最终遗憾界：

$$\mathrm{Reg}_T(u) \le O\big(\lambda \|u\|_2^2 + dY^2 \log(T)\big)。$$

与 $O\big(\max_{x \in X, t \in [T]} \{(x^\top \psi_t - y_t)^2\} \cdot d \log T\big)$ 相比，改进的遗憾界依赖于 Y^2，而不是可能较大的数量 $\max_{x \in X, t \in [T]} \{(x^\top \psi_t - y_t)^2\}$，从而展示了在在线回归中利用辅助信息的价值。