Lagrange Optimization

Presenter: Zhe Wang

https://qdata.github.io/deep2Read

Zhe Wang

201909

primal and dual problem

Primal optimization problem:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leqslant 0, \quad i = 1, \cdots, m \end{aligned} \tag{1}$$

Equivalent form:

$$\min_{x} \max_{\lambda_i \geqslant 0} \quad f_0(x) + \lambda_i f_i(x) \tag{2}$$

Dual problem:

$$\max_{\lambda_i \geqslant 0} \min_{x} \quad f_0(x) + \lambda_i f_i(x) \tag{3}$$

We have:

$$p* = \min_{x} \max_{\lambda_i \geqslant 0} \quad f_0(x) + \lambda_i f_i(x) \geqslant \max_{\lambda_i \geqslant 0} \min_{x} \quad f_0(x) + \lambda_i f_i(x) = d^*$$

- The dual problem is always convex regardless of the convexity of the primal.
- If the equality holds for (x^*, λ_i^*), the problem satisfies strong duality, (x^*, λ_i^*) are called saddle points.

Complementary Slackness

Suppose the strong duality holds for (x^*, λ_i^*), we have:

$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$$

The property is known as complementary slackness.

KKT optimality conditions

Assume all functions $f_0, \cdots, f_m$ are differentiable.

Karush-Kuhn-Tucker conditions

Suppose the strong duality holds for (x^*, λ_i^*), we have the following conditions:

- $f_i(x^*) \leqslant 0, i = 1, \cdots, m$
- $\lambda_i^* \geqslant 0, i = 1, \cdots, m$
- $\lambda_i^* f_i(x^*) = 0, i = 1, \cdots, m$
- $\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) = 0$

KKT conditions for convex problems

Main conclusion

If $f_i(x)$ are convex, $\hat{x}, \hat{\lambda}$ are points satisfy the KKT conditions, then the strong duality holds, and $(\hat{x}, \hat{\lambda})$ is a pair of saddle point.

Primal

$$\max_{p,d\in U} < p, r > -\frac{1}{\eta}D(p||p_0) - \frac{1}{\alpha}H(d||d_0)$$
$$s.t. \quad E^T d = \gamma P^T p + (1-\gamma)\nu_0$$
$$\Phi^T d = \Phi^T p$$

Lagrange form:

$$\begin{aligned}\mathcal{L}(p,d;V,\theta,\rho) &= \langle p, r\rangle + \langle V, \gamma P^\top p + (1-\gamma)\nu_0 - E^\top d\rangle + \langle \theta, \Phi^\top d - \Phi^\top p\rangle + \rho\,(1 - \langle p, \mathbf{1}\rangle) - \frac{1}{\eta}D(p\|p_0) - \frac{1}{\alpha}H(d\|d_0)\\ &= \langle p, r + \gamma PV - \Phi\theta - \rho\mathbf{1}\rangle + \langle d, \Phi\theta - EV\rangle + (1-\gamma)\,\langle \nu_0, V\rangle + \rho - \frac{1}{\eta}D(p\|p_0) - \frac{1}{\alpha}H(d\|d_0)\\ &= \langle p, \Delta_{\theta,V} - \rho\mathbf{1}\rangle + \langle d, Q_\theta - EV\rangle + (1-\gamma)\,\langle \nu_0, V\rangle + \rho - \frac{1}{\eta}D(p\|p_0) - \frac{1}{\alpha}H(d\|d_0), \qquad (13)\end{aligned}$$

where $Q_\theta = \Phi\theta, \triangle_{\theta,V} = r + \gamma PV - Q_\theta$, the Lagrange form is a concave function of p and d.

Take the derivative w.r.t. p and d, we have:

$$p^*(x,a) = p_0(x,a)\exp^{\eta(\triangle_{\theta,V}(x,a)-\rho)}$$

$$\pi_d^*(x,a) = \pi_0(x,a)\exp^{\alpha(Q_\theta(x,a)-V(x))}$$

$$\rho^* = \log(\sum_{x,a} p_0(x,a)e^{\eta\triangle_{\theta,V}(x,a)})$$

Take the optimal values into the Lagrangian, we have:

$$L(p^*,d^*;V_\theta,\theta,\rho^*) = (1-\gamma) < \nu_0, V > + \frac{1}{\eta}\log(\sum_{x,a} p_0(x,a)e^{\eta\triangle_{\theta,V}(x,a)})$$

$$\begin{aligned}\mathcal{L}(p,d;V,\theta,\rho) &= \langle p,r\rangle + \langle V,\gamma P^\intercal p + (1-\gamma)\nu_0 - E^\intercal d\rangle + \langle \theta, \Phi^\intercal d - \Phi^\intercal p\rangle + \rho\,(1-\langle p,\mathbf{1}\rangle) - \frac{1}{\eta}D(p\|p_0) - \frac{1}{\alpha}H(d\|d_0)\\ &= \langle p, r+\gamma PV - \Phi\theta - \rho\mathbf{1}\rangle + \langle d, \Phi\theta - EV\rangle + (1-\gamma)\,\langle \nu_0, V\rangle + \rho - \frac{1}{\eta}D(p\|p_0) - \frac{1}{\alpha}H(d\|d_0)\\ &= \langle p, \Delta_{\theta,V} - \rho\mathbf{1}\rangle + \langle d, Q_\theta - EV\rangle + (1-\gamma)\,\langle \nu_0, V\rangle + \rho - \frac{1}{\eta}D(p\|p_0) - \frac{1}{\alpha}H(d\|d_0), \qquad (13)\end{aligned}$$

Primal

$$\max_{p,d \in U} < p, r > -\frac{1}{\eta} D(p||p_0) - \frac{1}{\alpha} H(d||d_0)$$
$$s.t. \quad E^T d = \gamma P^T p + (1-\gamma)\nu_0$$
$$\Phi^T d = \Phi^T p$$

Dual form:

Define the Q-function $Q_\theta = \Phi\theta$, the value function:

$$V_\theta(x) = \frac{1}{\alpha} \log(\sum_a \pi_i(x,a) e^{a_\theta(x,a)}),$$

and the Bellman error function $\triangle_\theta = r + \gamma P V_\theta - Q_\theta$. Then the optimal solution for the primal takes the form:

$$p^*(x,a) \propto p_0(x,a) e^{\eta \triangle_{\theta *}(x,a)}$$

$$\pi_{d*}(a|x) = \pi_0(a|x) e^{\alpha(Q_{\theta *}(x,a) - V_{\theta *}(x))}$$

where θ^* is the minimizer of the convex function.

$$\mathcal{G}(\theta) = \frac{1}{\eta}\log(\sum_{x,a} p_0(x,a)e^{\eta\triangle_\theta(x,a)}) + (1-\gamma) < \nu_0, V_\theta >$$

By analogy with the classic logistic loss, the loss function is called logistic Bellman error, its solutions Q_θ and V_θ the logistic value functions.
Two advantages:

- The $\mathcal{G}$ is convex.
- The $\mathcal{G}$ satisfies $||\nabla_Q \mathcal{G}(Q)||_1 \leqslant 2$

Q-REPS: a mirror-descent algorithm.

Suppose the feasible region is M, then the iterative optimization algorithm is :

$$(p_{k+1}, d_{k+1}) = \arg\max_{p,d \in M} < p, r > -\frac{1}{\eta} D(p||p_K) - \frac{1}{\alpha} H(d||d_k)$$

Implementing requires the minimum θ_k^* of the logistic Bellman error function

$$\mathcal{G}(\theta) = \frac{1}{\eta} \log(\sum_{x,a} p_k(x,a) e^{\eta \triangle_\theta(x,a)}) + (1-\gamma) < \nu_0, V_\theta >$$

In practice, exact minimization can be often infeasible due to the lack of knowledge of the transition function P and limited access to computation and data.

To use the Q-REPS, the optimization should directly work with the sample transitions obtained through interaction with the env.

In each epoch k, we execute policy π_k and obtain a batch of N sample transitions $\{\epsilon_{k,n}\}_{n=1}^{N}$, with $\epsilon_{k,n} = (X_{k,n}, A_{k,n}, X'_{k,n})$ drawn from the occupancy measure p_k induced by π_k.

Furthermore, defining the empirical Bellman error for any (x, a, x) as:

$$\hat{\triangle}_\theta(x, a, x') = r(x, a) + \gamma V_\theta(x') - Q_\theta(x, a)$$

Then the empirical logistic Bellman error (ELBE) is defined as:

$$\hat{\mathcal{G}}_k(\theta) = \frac{1}{\eta} \log(\frac{1}{N} \sum_{n=1}^{N} e^{\eta \hat{\triangle}_\theta(\epsilon_{k,n})}) + (1-\gamma) < \nu_0, V_\theta >$$

Variational method can be used to transform the distribution based target to sample based target.

Let D_N be the set of all probability distributions over $[N]$ and define

$$S_k(\theta, z) = \sum_z z(n)(\hat{\triangle}_\theta(\epsilon_{k,n}) - \frac{1}{\eta}\log(Nz(n))) + (1-\gamma) < \nu_0, V_\theta >$$

for each $z \in D_N$. we have :

$$\min_\theta \hat{\mathcal{G}}_k(\theta) = \min_\theta \max_{z \in D_N} S_k(\theta, z)$$

in each round $\tau = 1, 2, \cdots, T$, the sampler proposes a distributionz $Z_{k,\tau} \in D_N$ over sample transtions and the learner updates the parameters $\theta_{K,T}$

To optimize the θ,

- sample an index I from the distribution $z_{k,\tau}$
- let $(X, A, X) = (X_{k,I}, A_{k,I}, X_{k,I})$
- sample a state $\bar{X}$ and two actions $A', \bar{A}$

then, $\hat{g}_{k,t}(\theta) = \gamma\phi(X', A') - \phi(X, A) + (1-\gamma)\phi(\bar{X}, \bar{A})$ is an unbiased estimation for the $\frac{\partial S}{\partial \theta}$. the introduced variable Z can also be optimized through gradient based method.