# Research Idea

Xu Yunpeng

December 9, 2024

**Uniformly Quantized State Variables as Training Samples** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a discrete-time filtered probability space with a finite discrete index set $N = \{0, 1, ..., T\}$.

Consider a vector of $n$ controlled Markovian processes $X_{i,t}$ with Gaussian noise $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ live in $\mathbb{R}^n$, where the control variable is denoted by $u$ live in the set $\mathbb{U}$, and the control at a time $t$ is $u_t \in \mathbb{U}_t$, $\mathbb{U}_t \subseteq \mathbb{U}$:

$$\boldsymbol{Y}_t = (X_{1,t}, ..., X_{n,t})^\intercal.$$

$$\boldsymbol{Y}_{t+1} = g(\boldsymbol{Y}_t, u_t, \boldsymbol{\epsilon}_{t+1})$$

Here $g : \mathbb{R}^n \times \mathbb{U} \times \mathbb{R}^n \to \mathbb{R}^n$ establishes the dynamics of $\boldsymbol{Y}_t$.

Denote the realisation of $\boldsymbol{Y}_t$ as $\boldsymbol{y}$, and consider functions that do the following mappings, reward functions $f_1 : \mathbb{R}^n \times \mathbb{U} \to \mathbb{R}$, $f_2 : \mathbb{R}^n \to \mathbb{R}$, value function $J : N \times \mathbb{R}^n \times \mathbb{U} \to \mathbb{R}$, and optimal value function $V : N \times \mathbb{R}^n \to \mathbb{R}$,

$$J_t(\boldsymbol{y}, u) = \mathbb{E}\left(\sum_{s=t}^{T-1} f_1(\boldsymbol{Y}_s, u) + f_2(\boldsymbol{Y}_T) \middle| \boldsymbol{Y}_t = \boldsymbol{y}\right)$$

$$\begin{aligned} V_t(\boldsymbol{y}) &= \sup_{\forall u \in \mathbb{U}} \big\{ J_t(\boldsymbol{y}, u) \big\} && (1) \\ &= \sup_{\forall u \in \mathbb{U}_t} \left\{ f_1(\boldsymbol{y}, u) + \mathbb{E}\left( V_{t+1}\Big(g(\boldsymbol{y}, u, \boldsymbol{\epsilon}_{t+1})\Big) \middle| \boldsymbol{Y}_t = \boldsymbol{y} \right) \right\}, t \leq T-1 && (2) \\ V_T(\boldsymbol{y}) &= f_2(\boldsymbol{y}) && (3) \end{aligned}$$

where the closed-form expressions of $f_1, f_2, g$ are known. The aim is to find the surrogate of $V_t$ using the above dynamic programming set up, so that given a sample path, one can use the surrogate to find the optimal control according to Equation 2. The closed-form expression of $V_{T-1}$ can be very challenging to find, but we can input a training sample $(\hat{\boldsymbol{y}}_1, ..., \hat{\boldsymbol{y}}_p)$ generated by random uniform with suitable ranges, optimize $f_1(\hat{\boldsymbol{y}}_i, u) + \mathbb{E}\big(V_T(g(\hat{\boldsymbol{y}}_i, u, \boldsymbol{\epsilon}_T))\big)$ using existing algorithms, denote the optimized value of it as $\hat{v}_i$, and gather $p$ number of pairs $(\hat{\boldsymbol{y}}_i, \hat{v}_i), i = 1, ..., p$. We can then do a feedforward neural network training to find an approximated mapping, i.e. a surrogate, of $V_{T-1}$, denoted as $\hat{V}_{T-1}$, and we can repeat such procedure to find $V_t, t \leq T-2$ by replacing $V_{t+1}$ in Equation 2 by $\hat{V}_{t+1}$.

The problem is whether if we can find the optimal training sample such that the surrogate is consistent for $V$. Standing at time $T-1$, if one can find a $m$ level optimal uniform quantization of $\boldsymbol{Y}_{T-1}$

$$\tilde{\boldsymbol{y}}_{T-1}(m) = \begin{pmatrix} \tilde{\boldsymbol{y}}_{1,T-1} \\ \tilde{\boldsymbol{y}}_{2,T-1} \\ \vdots \\ \tilde{\boldsymbol{y}}_{m,T-1} \end{pmatrix} \text{ with } \begin{pmatrix} \frac{1}{m} \\ \frac{1}{m} \\ \vdots \\ \frac{1}{m} \end{pmatrix}, \tag{4}$$

such that $\tilde{\boldsymbol{y}}_{T-1}(m) \to \boldsymbol{Y}_{T-1}$ as $m \to \infty$, then we shall treat the $\tilde{\boldsymbol{y}}_{T-1}(m)$ as training samples, and at the same time a discrete uniform random vector.

Assuming there exist an optimal feedforwork neural network architecture such that the MSE can equal 0, standing at $T-1$, we can use a similar method to optimize $f_1(\tilde{\boldsymbol{y}}_{i,T-1}, u) + \mathbb{E}\big(V_T(g(\tilde{\boldsymbol{y}}_{i,T-1}, u, \boldsymbol{\epsilon}_T))\big)$ using existing algorithms, denote the optimized value of it as $\tilde{v}_i$, and gather $m$ number of pairs $(\tilde{\boldsymbol{y}}_{i,T-1}, \tilde{v}_i), i = 1, ..., m$. We can then do a feedforward neural network training to find a surrogate of $V_{T-1}$, denoted as $\tilde{V}_{T-1}$. Standing at time 0, we can treat the $\tilde{V}_{T-1}(\tilde{\boldsymbol{y}}_{T-1}(m))$ as a discrete uniform random variable at time $T-1$. If we can apply some existing theorems and managed to prove the following,

$$\lim_{m\to\infty} \mathbb{E}\Bigg(\Big(\tilde{V}_{T-1}\big(\tilde{\boldsymbol{y}}_{T-1}(m)\big) - V_{T-1}\big(\tilde{\boldsymbol{y}}_{T-1}(m)\big)\Big)^2\Bigg) \to \mathbb{E}\Bigg(\big(\tilde{V}_{T-1}(\boldsymbol{Y}_{T-1}) - V_{T-1}(\boldsymbol{Y}_{T-1})\big)^2\Bigg),$$

then based on the below

$$\mathbb{E}\Bigg(\big(\tilde{V}_{T-1}(\tilde{\boldsymbol{y}}_{T-1}(m)) - V_{T-1}(\tilde{\boldsymbol{y}}_{T-1}(m))\big)^2\Bigg) = \sum_{j=1}^{m} \frac{1}{m} \cdot \big(\tilde{V}_{T-1}(\tilde{\boldsymbol{y}}_{i,T-1}) - \tilde{v}_i\big)^2 = 0$$

we can conclude $\mathrm{Var}\big(\tilde{V}_{T-1}(\boldsymbol{Y}_{T-1}) - V_{T-1}(\boldsymbol{Y}_{T-1})\big)$ and $\mathbb{E}\big(\tilde{V}_{T-1}(\boldsymbol{Y}_{T-1}) - V_{T-1}(\boldsymbol{Y}_{T-1})\big)$ are " asymptotically " 0, therefore standing at time 0, we can say $\hat{V}_{T-1}$ is an " asymptotically " consistent esitmate of $V_{T-1}$. The procedure can then be repeated to find $V_t, t \leq T-2$ by replacing $V_{t+1}$ in Equation 2 by $\tilde{V}_{t+1}$.