随机博弈在 P2P 交易中的 Nash 均衡存在性证明

本文主要利用λ-Lipschitz 连续条件与 Brouwer 不动点定理严格证明了在随机博弈框架下 P2P 能源交易中 Nash 均衡的存在性。

在基于随机博弈优化 P2P 能源交易的过程中，为更全面地刻画产消者之间交易行为的随机性与交互性，将论文正文中的公式(6)通过 Bellman 等价公式重新表述为以下形式：

$$V_{\pi_n}(s)=\sum_{a_n^t\in\boldsymbol{A}_n}\pi_n(a_n^t\mid s)(R_s^{(a_n^t,\pi_{-n})}+\gamma\sum_{s^{'}\in\boldsymbol{S}_n}P(s^{'}\mid s,a_n^t)V_{\pi_n}(s^{'})) \tag{A1}$$

$$Q_{\pi_n}(a_n^t,\boldsymbol{\pi}_{-n})=R_s^{(a_n^t,\pi_{-n})}+\gamma(\sum_{s^{'}\in\boldsymbol{S}_n}P(s^{'}\mid s,a_n^t)V_{\pi_n}(s^{'})) \tag{A2}$$

式中：$V_{\pi_n}(s)$、$Q_{\pi_n}(a_n^t,\boldsymbol{\pi}_{-n})$分别表示论文正文公式(8)中的状态值函数和动作值函数；$R_s^{(a_n^t,\pi_{-n})}$为产消者 n 当前状态下的奖励，即交易收益；$V_{\pi_n}(s^{'})$为表示产消者 n 在状态 $s^{'}$ 下的即时奖励；γ 为折扣因子，本文设置为$\gamma=1$。

之后，定义如下函数：

$$\varphi_n(\pi)=\max\left\{0,Q_{\pi_n}(\tilde{a}_n^t,\pi_{-n})-Q_{\pi_n}(\pi_n,\pi_{-n})\right\} \tag{A3}$$

上式表明，产消者 n 改变自身交易行为会影响其自身的经济效益。当对于所有产消者总有$\varphi_n(\pi)\equiv 0$，即可证明随机博弈在 P2P 交易中存在 Nash 均衡，且π为最佳策略，能够最大化产消者的经济效益。

进一步地，基于公式(A3)，构建一个辅助函数如下：

$$f_n(\pi)=\frac{\pi_n(s_n^t,\tilde{a}_n^t)+\varphi_n(\pi)}{1+\sum_{b_n^t\in\boldsymbol{A}_n}\varphi_n(\pi)} \tag{A4}$$

从上述辅助函数可以看出，该函数将原始交易策略映射到了一个新的策略空间中。若产消者 n 在新策略空间中的某个动作能够提升交易效益，则该动作的选择概率也将会相应增加。这一机制反映了产消者在随机博弈框架下对收益改进的动态响应，有助于逐步逼近全局最优策略，同时保证 Nash 均衡解的渐进性收敛。

本文首先证明公式(A4)满足 *λ-Lipschitz* 连续性。在 P2P 交易过程中，由于所有产消者的交易策略存在交互性，因此由所有混合策略构成的策略集$\boldsymbol{\pi}$可以表示为所有产消者策略的笛卡尔积，即：

$$\boldsymbol{\pi}=\boldsymbol{\pi}_1\times\boldsymbol{\pi}_2\times\cdots\times\boldsymbol{\pi}_n\times\cdots\times\boldsymbol{\pi}_N \tag{A5}$$

相应地，所有产消者的动作集合可以表示为：

$$\boldsymbol{A}=\boldsymbol{A}_1\times\boldsymbol{A}_2\times\cdots\times\boldsymbol{A}_n\times\cdots\times\boldsymbol{A}_N \tag{A6}$$

在 P2P 交易过程中，由于每个产消者的交易策略集合均为有限集，因此$\boldsymbol{\pi}$中任意两个策略的无穷范数距离可以表示为：

$$\left\|\boldsymbol{\pi}^1-\boldsymbol{\pi}^2\right\|_\infty\leq\max\left|\boldsymbol{\pi}^1(s_n,a_n)-\boldsymbol{\pi}^2(s_n,a_n)\right|=\delta \tag{A7}$$

式中：δ为所有策略中无穷范数的最大距离度量。

为了证明所构建的辅助函数$f_n(\pi)$在$\boldsymbol{\pi}$中满足 *λ-Lipschitz* 连续性，只需要任选两个策略$\boldsymbol{\pi}^1$、$\boldsymbol{\pi}^2$，并证明其满足$\left\|f(\boldsymbol{\Psi}^1)-f(\boldsymbol{\Psi}^2)\right\|_\infty\leq\lambda\delta$即可。为支持这一证明，本文参考了文献[30]中的引理：

引理 1：$f_n(\pi)$具备 *λ-Lipschitz* 连续性，因为对任意两个交易策略$\boldsymbol{\pi}^1,\boldsymbol{\pi}^2\in\boldsymbol{\pi}$，存在以下结论：

$$\left\|f(\boldsymbol{\pi}^1)-f(\boldsymbol{\pi}^2)\right\|_\infty\leq\frac{11NS^2A_{\max}^2r_{\max}}{(1-\gamma)^2}\delta \tag{A8}$$

式中：S为产消者 n 的状态维度；$A_{\max}$为产消者的动作空间的最大维度；$r_{\max}$为产消者在所有时段可收获的最大交易收益。

引理 1 表明，在交易策略空间$\boldsymbol{\pi}$内，对于任意两个策略，所构建的辅助函数的变化不会超过策略间最大距离度量的$\frac{11NS^2A_{\max}^2r_{\max}}{(1-\gamma)^2}$倍，从而确保了辅助函数在该空间内具有 *λ-Lipschitz* 连续性。这一性质对后续的 Nash 均衡存在性证明至关重要。

基于引理 1，结合$\pi_n\in[0,1]$和$f_n(\pi)\in[0,1]$，可知$f_n(\pi)$是从[0,1]映射[0,1]的函数。因此，可以进一步通过 *Brouwer* 不动点定理证明随机博弈框架下 P2P 能源交易中 Nash 均衡的存在性。在本文中，只需证明当且仅当 P2P 交易策略π_n是$f_n(\pi)$的一个不动点，即可证明该策略也是随机博弈的 Nash 均衡点。接下来本文将从必要性和充分性两个方面对随机博弈 Nash 均衡点满足 *Brouwer* 不动点定理进行证明。

1）必要性证明

首先假设π_n^*为随机博弈的 Nash 均衡点，基于公式(A3)可知：

$$\varphi_n(\pi_n^*)=\max\left\{0,Q_{\psi_n}(\tilde{a}_n^t,\boldsymbol{\pi}_{-n}^*)-Q_{\Psi_n}(\pi_n^*,\boldsymbol{\pi}_{-n}^*)\right\} \tag{A9}$$

从(A9)中可知，π_n^*作为随机博弈的Nash均衡点，无论产消者n的策略如何改变始终存在$Q_{\psi_n}(\tilde{a}_n^t,\boldsymbol{\pi}_{-n}^*)\le Q_{\Psi_n}(\pi_n^*,\boldsymbol{\pi}_{-n}^*)$，即$\varphi_n(\pi_n^*)\equiv 0$。因此，公式(A4)可以表示为：

$$f_n(\pi_n^*)=\frac{\pi_n^*(s_n^t,\tilde{a}_n^t)+\varphi_n(\pi_n^*)}{1+\sum_{b_n^t\in A_n}\varphi_n(\pi_n^*)}=\pi_n^* \tag{A10}$$

所以，随机博弈的Nash均衡点也是$f_n(\pi)$的不动点。

2）充分性证明

假设π_n为辅助函数$f_n(\pi)$的不动点，则此时有：

$$f_n(\pi_n)=\frac{\pi_n(s_n^t,a_n^t)+\varphi_n(\pi_n)}{1+\sum_{b_n^t\in A_n}\varphi_n(\pi_n)}=\pi_n \tag{A11}$$

在不动点π_n处，对产消者n的任意动作a_n^t，做出以下假设：

$$Q_{\psi_n}(a_n^t,\boldsymbol{\pi}_{-n})\le Q_{\psi_n}(\pi_n,\boldsymbol{\pi}_{-n}) \tag{A12}$$

为了证明公式(A12)假设成立，采用反证法进行证明。首先假设存在某个动作$c_n\in\boldsymbol{A}$使其满足$Q_{\psi_n}(c_n,\boldsymbol{\pi}_{-n})>Q_{\psi_n}(\pi_n,\boldsymbol{\pi}_{-n})$。

之后，利用*Bellman*策略方程可以将效益函数(A1)进一步等价为下式：

$$\begin{aligned}V_{\pi_n}(s)&=\sum_{a_n^t\in\boldsymbol{A}_n}\pi_n(s,a_n^t)(R_s^{(a_n^t,\pi_{-n})}+\gamma\sum_{s'\in\boldsymbol{S}_n}P(s'\mid s,a_n^t)V_{\pi_n}(s')\\&=\sum_{a_n^{t'}\in\boldsymbol{A}_n^+}\pi_n(s,a_n^{t'})(Q_{\pi_n}(a_n^{t'},\boldsymbol{\pi}_{-n}))\end{aligned} \tag{A13}$$

式中：$\boldsymbol{A}_n^+$为满足$\pi_n(s,a_n^{t'})>0$的动作的集合。

又因为$\sum_{a_n^{t'}\in\boldsymbol{A}_n^+}\pi_n(s,a_n^{t'})=1$，所以一定存在至少一个动作$d_n\in\boldsymbol{A}_n^+$满足$Q_{\pi_n}(d_n,\boldsymbol{\pi}_{-n})<Q_{\pi_n}(\pi_n,\boldsymbol{\pi}_{-n})$。将其带入公式(A11)可得：

$$\begin{aligned}f_n(\pi_n(s,d_n))&=\frac{\pi_n(s_n^t,d_n)+\varphi_n(\pi_n)}{1+\sum_{b_n^t\in A_n}\varphi_n(\pi_n)}\\&=\frac{\pi_n(s_n^t,d_n)}{1+\sum_{b_n^t\in A_n}\varphi_n(\pi_n)}\\&\le\frac{\pi_n(s_n^t,d_n)}{1+Q_{\pi_n}(c_n,\boldsymbol{\pi}_{-n})-Q_{\pi_n}(\pi_n,\boldsymbol{\pi}_{-n})}\\&\ne\pi_n(s,d_n)\end{aligned} \tag{A14}$$

基于上式，可以发现，如果在不动点π_n处存在$Q_{\pi_n}(c_n,\boldsymbol{\pi}_{-n})>Q_{\pi_n}(\pi_n,\boldsymbol{\pi}_{-n})$，则与*Brouwer*不动点定理相悖，因此对于$Q_{\psi_n}(a_n^t,\boldsymbol{\pi}_{-n})\le Q_{\psi_n}(\pi_n,\boldsymbol{\pi}_{-n})$必定存在不动点，公式(A12)假设也得以成立。进而公式(A11)可以表示为：

$$\pi_n = \frac{\pi_n(s_n^t, a_n^t)}{1 + \sum_{b_n^t \in \boldsymbol{A}_n} \varphi_n(\pi_n)} \tag{A15}$$

进一步可知上式等号两侧相等，即右侧分母等于 1。这也表明，对产消者 n 和动作 $b_n^t \in \boldsymbol{A}_n$，有 $\sum_{b_n^t \in \boldsymbol{A}_n} \varphi_n(\pi_n) = 0$。这意味着没有产消者能够通过改变自身策略来增加 P2P 交易的经济效益。因此 π_n 也就是产消者 n 的最优策略，即随机博弈框架下 P2P 能源交易的 Nash 均衡点。

至此，基于随机博弈的 P2P 交易中 Nash 均衡的存在性得以证明。