Controllability and Period-Survivability of Cooperative-Reciprocal Strategies in Smart Grids

Vadim Romanuke[1,*,†] and Larysa Radzikhovska[1,†]

[1] *Vinnytsia Institute of Trade and Economics of the State University of Trade and Economics, Soborna 87, 21050, Vinnytsia, Ukraine*

Abstract

A method of stabilizing the smart grid state is presented as a response to plausible perturbations from detecting behavior, which negatively affects the grid sustainability. The method is based on modeling interactions among grid agents via Axelrod tournaments. A smart grid recovery strategic trajectory is built by iteratively optimizing the penalization level in the price change model, cooperation rewarder in the cooperation and defection payoffs, and the price deviation corrector. When the intervals of those three parameters are appropriately divided into subintervals, the solution is found within 10 iterations or fewer. The price deviation corrector has a more stable impact on both the payoff and defection rate. This is a virtual run followed the perturbation, and the recovery strategic trajectory returns the smart grid system to normal state. As the period of survivability expires, an update of the strategic trajectory may be needed.

Keywords

smart grid, controllability, adaptive reciprocity, strategy survivability, Axelrod tournament

1. Introduction

Since 2000s, smart grids have been increasingly becoming a substantial part of energy renewability and decentralization. The latter has become a valuable indicator of democratization for the past three decades. Speaking technologically, a smart grid is an extremely complex electricity network that combines traditional power infrastructure with digital technologies, communication systems, sensors, automated control, and data analytics to improve the production, transmission, distribution, and consumption of electricity [1]. So, the smart grid aims at the distributed and self-organizing control [2]. Its decentralization provides greater resilience, increased use of renewable energy, anti-monopolization, and market competition, which all align well economically and politically with democratization [1]. On the other hand, controllability of the smart grid, despite decentralization, remains fundamentally essential [2]. The reason is very simple: electric power systems obey physical laws. At every moment, power generation must be equal (at least, approximately, not exceeding some tolerable discrepancy) to power consumption, including losses, imports and exports on both sides. A significant imbalance causes frequency deviations, voltage instability, equipment damage, and even short-term blackouts [2, 3]. Electricity cannot be “stored in transmission lines” while participants (not only prosumers and consumers, but also grid operators, energy aggregators, energy storage operators, and governmental regulators) negotiate what to do. And the power balance must be maintained continuously [1, 3, 4].

2. Relevance of the Research

To provide continuous power balance in a smart grid, governmental regulators establish market rules, technical standards, reliability and cybersecurity requirements [2, 5]. The smart grid participants must follow such rules and comply with standards, and requirements to ensure grid

[1]*RS-2026: Resilient Systems: Secure Digital Technologies and Critical Infrastructure, June 30, 2026, Drohobych, Ukraine*

* Corresponding author.

† These authors contributed equally.

romanukevadimv@gmail.com (V. V. Romanuke); larirad@ukr.net (L. M. Radzikhovska)

0000-0001-9638-9572 (V. V. Romanuke); 0000-0003-0185-8036 (L. M. Radzikhovska)

overall sustainability and controllability for maintaining the grid stable [2, 6]. In particular, prosumers and consumers (also referred to as agents for further reasoning) act collectively and affect grid stability [1, 3, 6]. If even the grid becomes unstable, it usually lasts for a negligibly short period, whereupon the normal state is back. The speed of grid state normalization, however, does depend on reasonable strategic behavior of the agents.

Obviously, "reasonable behavior" in a smart-grid sense is not about being nice or cooperative in isolation. This is about keeping the system in a fast-return-to-equilibrium regime under repeated interaction and occasional noise (e. g., volatility of the energy unit price). Reasonable strategic behavior of agents consists mainly in conditional cooperation (not unconditional) and proportional response to deviation [7]. Cooperation herein implies compliance with grid-balancing requests, whereas defection, being opposite to cooperation, implies (selfish) pursuance of only own short-term benefit [7, 8]. Thus, agents cooperate when others behave cooperatively (e. g., follow demand-response signals), because this keeps the system efficient. In addition, defection or non-compliance is met with limited defection response, not excessive penalization [7, 9]. For instance, agents mostly return to cooperation once normal conditions are restored, because the common goal is correction, not escalation [5, 10, 11].

The agent decides on whether to cooperate or defect based on the energy unit price, which affects a bimatrix game between the agent and the opponent personifying either another agent or the whole community factually interacting with the agent [4, 12]. Hence, the primary task is to model the price change and the entries of the payoff matrix so it could serve as an incentive for cooperative-favoring behavior. However, earlier approaches to smart-grid stability and incentive design were based on optimization, control theory, and market-based mechanisms. Those approaches studied optimal power flow [13], stochastic and model predictive control [14], and demand-side management via centralized or distributed optimization [12], where system stability is enforced through engineered constraints rather than emergent behavior. Parallel to that, economic approaches introduced dynamic pricing schemes such as real-time pricing and locational marginal pricing, where consumption is influenced by price signals reflecting system conditions (e. g., generation cost and congestion) [15]. However, those approaches assumed either compliant agents or static behavioral responses to prices, without explicitly modeling the evolution of strategic interaction patterns over time.

Next to that, non-cooperative games were used to model energy consumption, demand response, and distributed generation as Nash equilibrium problems, where agents optimize individual benefits under pricing signals [16]. Stackelberg games are widely used to represent hierarchical interactions between system operators (leaders) and consumers or prosumers (followers), particularly in demand response and energy trading [16]. Cooperative game theory is used to further address coalition formation, cost sharing, and fair allocation of benefits among participants in microgrids and energy communities [17]. Nevertheless, these models generally remain static in structure: they characterize equilibria of a given game, rather than how behavioral strategies evolve.

Evolutionary and behavioral game-theoretic approaches extend classical models by introducing population dynamics, bounded rationality, and adaptive strategy revision over repeated interactions [18]. Such approaches tackle to model dynamic interaction in smart grids based on demand response, distributed energy trading, and cooperative behavior under replicator dynamics [19] or imitation rules [20]. These models capture how strategies spread or vanish depending on performance [10, 19]. However, most evolutionary models still rely on relatively simple strategy spaces or continuous decision variables [11, 16], and often do not explicitly incorporate rich behavioral rules with memory, such as reciprocity [7, 10], penalization [8, 10, 20], or forgiveness [7, 9], which are crucial in repeated interaction settings with heterogeneous agents. To the contrary, Axelrod Iterated Prisoner's Dilemma tournaments (AIPDTs) allows modeling smart grid interactions with reciprocity, memory, and behavioral path dependence in repeated interaction [9]. AIPDTs involve a large variety of behavioral strategies (over 200 in total), where the strategy is a rule for forming a definite sequence of complementary actions (moves) like cooperation and defection [21]. The AIPDT, if properly parametrized, organizes the bimatrix game that evolves in accordance with the change of the main grid states from the viewpoint of the agent. This is an extremely simple yet effective framework for modeling decentralized smart grid interaction process by regarding past compliance and trust.

3. Statement of the Problem

Issuing from the need for controllable cooperation among agents in smart grids, the goal is to suggest a method of determining sustainable strategies for efficient cooperation with plausible reciprocity. To achieve the goal, the energy unit price change and the payoff matrix evolution is to be modeled first with some parametrization, where the price and payoff matrix are in a feedback response to undesirable defections. The feedback should form an incentive for cooperative-favoring behavior. Second, upon conducting AIPDTs, the most influential parameter is to be determined such that its adjustment is the most tightly connected with inducing the cooperative-reciprocal strategies. Next, an algorithm of mitigating defections to a tolerable rate is to be developed, whose returns are the best parameter value and cooperative-reciprocal strategies induced by that value. In addition, given a criterion of the strategy applicability, the period of survivability of those strategies must be estimated. Finally, the results should be briefly discussed, whereupon conclusions on the findings are to be made about the efficiency of cooperative-reciprocal strategies that ought to become not only socially desirable but evolutionarily dominant, thereby supporting long-term grid stability and controllability.

4. Research Results

4.1. Energy unit price

Basically, the energy unit price consists of the standardized part $P_0 = 1$ and observably random fluctuations whose causes are non-traceable. If the defection rate at time slot t denoted by $d(t)$ is too high, then this event must be penalized by raising the price up. This is believed to serve as an incentive to defect less. So,

$$P(t) = 1 + \alpha(d(t) - d_0) + \sigma\xi \text{ for } \alpha > 0 \text{ and } \sigma > 0 \text{ by } \xi \in \mathrm{N}_{\mathrm{gen}}(0, 1), \tag{1}$$

where α is a controllable penalization level (adjusted by the governmental regulators), d_0 is a defection rate threshold, and $\mathrm{N}_{\mathrm{gen}}(0, 1)$ is a generator of values ξ of normally distributed random variable with zero mean and unit variance, to which σ (an external standard deviation, being uncontrollable) determines the strength of price fluctuations. Threshold d_0 can be both zero and nonzero. If $d_0 = 0$ then any defection within the smart grid is intolerable. If $d_0 > 0$ (but, certainly, d_0 cannot exceed 0.5 or so), then equation (1) suggests that the grid system stimulates agents to defect as less as possible because then, below threshold d_0, the price is likely to drop below 1 (depending on the price fluctuation strength or price volatility σ).

Another, stricter version of influencing the price is when it cannot drop below 1 on average (if the fluctuation is excluded) by $d < d_0$, i. e.

$$P(t) = 1 + \alpha \max\{d(t) - d_0, 0\} + \sigma\xi. \tag{2}$$

The third possibility is to reward agents by staying as close as possible to defiance rate d_0, which then becomes a desired defiance rate, rather than a threshold:

$$P(t) = 1 + \alpha|d(t) - d_0|^{\mu} + \sigma\xi \text{ by } \mu > 0. \tag{3}$$

Equation (3) suggests that the price grows up whenever $d(t)$ significantly deviates from d_0. At $\mu = 1$ the expected (or average, in practice) price growth versus the defection rate is linear. At $\mu \in (0; 1)$ the expected price growth versus the defection rate is concavely nonlinear (there are two concave curves by $d(t) \in [0; d_0]$ and $d(t) \in [d_0; 1]$), where the growth abruptness becomes more distinct as μ

is decreased off 1. At $\mu > 1$ the expected price growth is convexly nonlinear (there are two convex curves by $d(t) \in [0;\ d_0]$ and $d(t) \in [d_0;\ 1]$), where the growth weakens as μ is increased off 1.

While both the price linear-cutoff influence by (2) and price nonlinear influence by (3) are applicable, they have one substantial demerit: the incentive for mitigating defecting behavior is weaker than that by linear penalization in (1). Hence, model (1) of the price change will be utilized further. Parameter α can be called the price sensitivity (to defection deviation).

If there are N agents interacting at time slot t, the defection rate is

$$d(t) = \frac{1}{N} \sum_{n=1}^{N} \gamma_n (D,\ t), \tag{4}$$

where $\gamma_n (D,\ t) = 1$ if agent n defects at time slot t and $\gamma_n (D,\ t) = 0$ if agent n cooperates at time slot t. Occasionally, defection rate (4) can go up to 0.5 and above, but if it is mitigated in a few time slots, it does not destabilize the grid.

4.2. Payoffs

At each time slot t the agent chooses between cooperation (denoted by C) and defection (denoted by D). The payoff matrix at time slot t is

$$\mathbf{P}(t) = \begin{bmatrix} r(t) & s(t) \\ q(t) & u(t) \end{bmatrix} \text{ by } t = 0,\ 1,\ 2,\ 3,\ \ldots, \tag{5}$$

where $r(t)$ is the agent's payoff when both the agent and the opponent cooperate, $s(t)$ is the agent's payoff when the agent cooperates and the opponent defects, $q(t)$ is the agent's payoff when the agent defects and the opponent cooperates, $u(t)$ is the agent's payoff when both the agent and the opponent defect. These payoffs, measured in standardized units, are related to price change model (1). Time slot $t = 0$ corresponds to the situation, when the relationship among elements of matrix (5) turns to be

$$s(0) < u(0) < r(0) < q(0). \tag{6}$$

Inequalities (6) are the canonical Prisoner's Dilemma ordering [18], which guarantees that unilateral defection is always attractive in the short run. This actually provokes a short-term growth of defecting behavior, which can be mitigated only by achieving inequality

$$r(t) > q(t), \tag{7}$$

whenever the canonical ordering

$$s(t) < u(t) < r(t) < q(t) \tag{8}$$

occurs or just the opposite to (7) inequality

$$r(t) < q(t) \tag{9}$$

emerges. Hence, once the relationship in (8) occurs or just inequality (9) occurs, the emergent task is to return the system to normal state by minimum possible intervention, thereby preserving democratized decentralization.

Entry $r(t)$ is the cooperation payoff, which must be increased proportionally to the defection rate in order to stimulate agents for cooperative behavior. If the price grows up, the price deviation (with respect to standardized part $P_0 = 1$) must multiplicatively facilitate cooperation. So,

$$r(t+1) = r(t) + \beta(P(t) - 1)\,d(t) + \theta\zeta_r$$
$$\text{for } \beta > 0 \text{ and } \theta > 0 \text{ by } \zeta_r \in \mathrm{N}_{\mathrm{gen}}(0,\ 1)\ (t = 0,\ 1,\ 2,\ 3,\ \ldots), \tag{10}$$

where parameter β is the cooperation rewarder, and θ (an external standard deviation, being uncontrollable) determines the strength of cooperation payoff fluctuations.

To the contrary, defection payoff $q(t)$ must be decreased as the defection rate grows up. To strengthen the penalization, the price growth must be taken into account, but when the price drops below 1, defecting behavior will be a response that should stabilize the price toward 1. Hence,

$$q(t+1) = q(t) - \beta(P(t) - 1)\,d(t) - \eta(P(t) - 1) + \theta\zeta_q$$
$$\text{for } \eta > 0 \text{ by } \zeta_q \in \mathrm{N}_{\mathrm{gen}}(0,\ 1)\ (t = 0,\ 1,\ 2,\ 3,\ \ldots), \tag{11}$$

where parameter η is the defection-based price deviation corrector. Standard θ deviation determining the strength of uncontrollable defection payoff fluctuations is the same as in (10).

Entry $s(t)$ is the cooperation-versus-defection payoff, received by the cooperating agent. Obviously, it must be increased proportionally to the defection rate in order to stimulate agents for cooperative behavior, regardless of the price growth or drop. Hence,

$$s(t+1) = s(t) + \eta d(t) + \theta\zeta_s \text{ by } \zeta_s \in \mathrm{N}_{\mathrm{gen}}(0,\ 1)\ (t = 0,\ 1,\ 2,\ 3,\ \ldots), \tag{12}$$

and thus entry (12) is related to entry (11) via the price deviation corrector. To the contrary, defection-versus-cooperation payoff $u(t)$ must be symmetrically decreased as the defection rate grows up:

$$u(t+1) = u(t) - \eta d(t) + \theta\zeta_u \text{ by } \zeta_u \in \mathrm{N}_{\mathrm{gen}}(0,\ 1)\ (t = 0,\ 1,\ 2,\ 3,\ \ldots). \tag{13}$$

In this way, entries (10) — (13) of payoff matrix (5) are interrelated via defection rate $d(t)$, whereas the cooperation and defection payoffs are coupled via the cooperation rewarder, and the defection payoff includes the term with the price deviation corrector, which directly influences cooperation-versus-defection and defection-versus-cooperation payoffs. Accordingly, the governmental regulators possess three parameters inducing the cooperative-reciprocal strategies: penalization level α in price (1), cooperation rewarder β in cooperation payoff (10) and defection payoff (11), and price deviation corrector η in payoffs (11) — (13), which are the result of defection (either unilateral or mutual).

4.3. AIPDT

Sustainable strategies for efficient cooperation with plausible reciprocity are basically determined via AIPDTs, which allow also gathering and computing statistics of the best strategies (such as average payoff, rank, relative frequency, periodicity) [21]. In fact, the AIPDT is a virtual simulation of repeated strategic interactions among a set of strategies at a given time slot *t*. In an AIPDT, a set of strategies repeatedly play the bimatrix game (Prisoner's Dilemma) with payoff matrix (5) against one another (i. e., this is a round-robin tournament). Each pair of strategies forms a match, consisting of a specified number of rounds (denote this number by R_1). During each round, the competing strategies simultaneously choose either cooperation or defection, and the corresponding payoffs are assigned according to payoff matrix (5). The payoffs accumulated and averaged over R_1 rounds constitute the average payoff of the match. A complete tournament is obtained by having every strategy play against every other strategy, including a match against itself [20]. So, for K strategies, there are

$$\frac{K \cdot (K+1)}{2}$$

such matches. To get statistically stable results, the AIPDT utilizes also a number of repetitions R_2, where each repetition is an independent execution of the round-robin tournament used thereupon for averaging results [18, 19, 21]. Then, upon the AIPDT completion, the Kstrategies are ranked according to their average payoffs, and the strategy with the highest average payoff is selected. In a more general set-up, a few more strategies are additionally selected whose average payoffs are sufficiently close to the highest average payoff.

Integers $R_1$and R_2control depth of interaction and statistical averaging, respectively. For the ongoing computational simulation, they are $R_1 = 100$and $R_2 = 25$, which are quite sufficient even for severely high price volatility $\sigma$and payoff fluctuation strength θ. It is important to underline that the AIPDT is executed independently at each discrete smart-grid time slot, so integers $R_1$and $R_2$are internal characteristics of the behavioral interaction model. The time slot belongs to the external evolution of the smart-grid system, where the tournament outcomes determine subsequent updates of strategy relative frequencies, defection rates, energy prices, and payoffs in matrix (5). For the ongoing computational simulation, denote the number of time slots by T.

Another parameter of the AIPDT is the behavioral inertia parameter. Its reason is that agents do not update their strategies instantaneously. Say, if the best strategy changes at time slot t_*(according to the tournament ranking at t_*), the agent is likely to react with some delay. Such delays are particularly caused by slow decision making by the agents (the agents are humans, not machines). It is a cognitive inertia. Another causes are institutional (specificities in rules and contracts) and technological (physical delays).

Each agent must have its own inertial characteristics, although they may overlap among the whole group of Nagents. Denote the inertia set of agent $n$by

$$\Delta_n(\mathrm{D}_\delta,\ t) = \overline{\{0,\ \delta_{\max}^{(n)}(t)\}} \subset \mathbb{N} \bigcup \{0\} \ (n = \overline{1,\ N}), \tag{14}$$

where

$$\mathrm{D}_\delta \in \{\mathrm{E},\ \mathrm{U},\ \mathrm{N}\} \tag{15}$$

is a discrete distribution (chosen randomly), by which a random choice

$$\delta_n(t) \in \Delta_n(\mathrm{D}_\delta,\ t) \tag{16}$$

is made; E, U, and Nis an exponential, uniform, and normal discrete distribution, respectively; $\delta_{\max}^{(n)}(t)$ is the longest period of time slots, during which agent $n$can not update its current behavioral strategy at time slot t. Hence, if the last strategy update of agent $n$happened at time slot t_*, then at time slot t agent $n$updates its strategy only if

$$t - t_* > \delta_n(t) \ (n = \overline{1,\ N}) \tag{17}$$

upon random choices (15) and (16) for inertia set (14). The latter, being worth mentioning, can change over time slots. For instance, integer $\delta_{\max}^{(n)}(t)$may decrease over time slots due to agent $n$becomes more experienced and confident for smart grid interactions.

4.4. Algorithm

Apart from the three explicit parameters (penalization level α, cooperation rewarder β, and price deviation corrector η), there are also the length (horizon) of interactions $T$and defection rate threshold $d_0$that can be implicitly regulated. The goal is to determine an ordered set $\Omega_*$of cooperative-reciprocal strategies such that the set ensures a defection rate not exceeding $d_*$by the average payoff $\tilde{\pi}$(in the sense of mathematical expectance). The order in set Ω_*matters as, in the case if the set contains more than one element, namely the succession of elements in Ω_*leads to sustainable average payoff $\tilde{\pi}$ at a tolerable defection rate d_*, whereas any permutation may not lead to them. Meanwhile, cooperation payoff $r(t)$, evolving with regulation by (10), cannot exceed some margin $r_{\max}$. Defection payoff $q(t)$, evolving with regulation by (11), cannot drop below some margin $q_{\min}$. Likewise, cooperation-versus-defection payoff $s(t)$ by (12) cannot exceed its margin $s_{\max}$, and defection-versus-cooperation payoff $u(t)$ cannot drop below some margin $u_{\min}$. Hence, an algorithm of determining set Ω_* must comply with an operator Φ mapping the AIPDT parameters into a triplet of Ω_*, $\{d(t)\}_{t=1}^{T}$, and $\tilde{\pi}$. In addition, let set Ω_* be featured with the following two time slot periods. First, it is the period of restoration (recovery) T_*, during which the smart grid normal state is restored, starting from $t = 0$ and condition (6). Second, it is the period of survivability $T_{\max}$ of set Ω_*, upon which at least one of four inequalities

$$q(t) < q_{\min},\ r(t) > r_{\max},\ s(t) > s_{\max},\ u(t) < u_{\min} \tag{18}$$

turns true. Therefore, operator Φ maps as follows:

$$\begin{aligned} G_*(N,\ T) &= \left\{\Omega_*,\ \{d(t)\}_{t=1}^{T},\ \tilde{\pi},\ T_*,\ T_{\max}\right\} = \\ &= \Phi\left(\alpha,\ \beta,\ \eta,\ T,\ d_0;\ r_{\max},\ s_{\max},\ q_{\min},\ u_{\min};\ \sigma,\ \theta,\ \left\{\{\Delta_n(\mathrm{D}_\delta,\ t)\}_{t=1}^{T}\right\}_{n=1}^{N}\right), \end{aligned} \tag{19}$$

where set

$$\Omega_* = \left\{\omega_*^{(t)}\right\}_{t=1}^{T_{\max}}. \tag{20}$$

Obviously, set $G_*(N,\ T)$ is a random object, wherein each of its elements (set Ω_*, the defection rate, average payoff, period of recovery, period of survivability) is a random object, being it an integer number or more complex item.

If $t > T_*$ then inequality $d(t) \leqslant d_*$ should be true at least until $t = T_{\max} + 1$. For example, Figure 1 shows a defection rate $d(t)$ throughout a month of smart grid interactions (an action is made once per day) corresponding to price $P(t)$ in Figure 2 by payoff matrix

$$\mathbf{P}(0) = \begin{bmatrix} 0.11 & 0.08 \\ 0.12 & 0.09 \end{bmatrix} \tag{21}$$

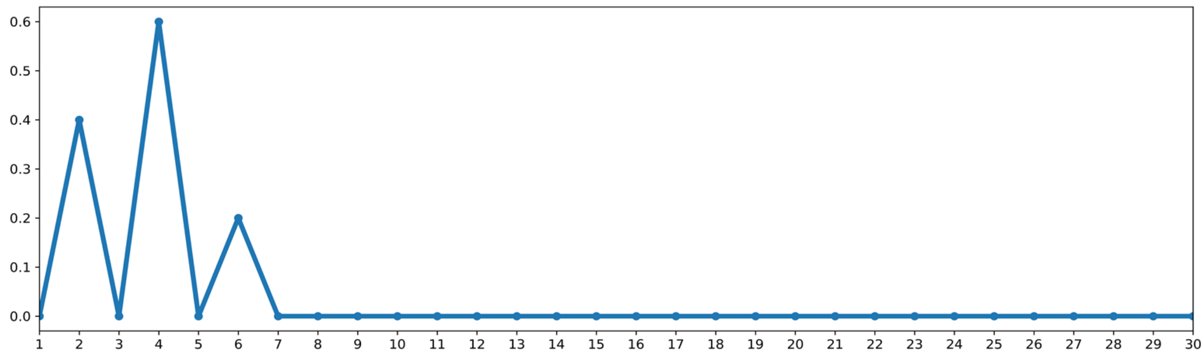

Figure 1: Defection rate evolution throughout a month of smart grid interactions among 1000 agents with a maximum weekly inertia period.

and parameters

$$\alpha = 0.25,\ \beta = 0.125,\ \eta = 0.01,\ T = 30,\ d_0 = 0.25,\ r_{\max} = 0.16,\ q_{\min} = 0.07,\ s_{\max} = 0.13,\ u_{\min} = 0.04,$$
$$\sigma = 0.05,\ \eta = 0.005,\ \delta_{\max}^{(n)}(t) = 7\ \forall\ n = \overline{1,\ 1000} \text{ and } \forall\ t = \overline{1,\ 30}, \tag{22}$$

where a disclosure of set Ω_*presented in Figure 3 reveals that cooperative-favoring behavior starts at $t = 4$with strategy "TitForTat" (it is $\omega_*^{(4)}$here). So, here $T_* = 6$, which is seen not merely from Figure 1, but also from Figure 3, where only strategies "Cooperator", "TitForTat", and "Grudger" are used since $t = 7$until $T_{\max} = T = 30$. It must be noted that strategy "Grudger" in the presented case is only conditionally cooperative, because it performs with zero tolerance (forgiveness) to defectors, where a single accidental defection may destroy cooperation chain. If the defection rate was not mitigated everywhere to 0, the presence of "Grudger" in set (20) since $t = 7$would be risky. Another peculiarity is strategy "Defector" at $t = 1$and $t = 3$: from smart-grid perspective, it is not that "Defector" is "good", but rather it characterizes the initial unstable state of the system. Likewise, strategy "Grudger" acts as a transitional enforcement mechanism that suppresses residual defection. Strategy "TitForTat" (the agent starts by cooperating and then mimics the previous action of the opponent), by which agents learn that cooperation is reciprocated and defection is immediately answered, acts as reciprocal adaptation. And finally, strategy "Cooperator" characterizes the restored stable regime once defection has been eliminated. However, the stability does not relate to the evolution of payoff matrix (Figure 4).

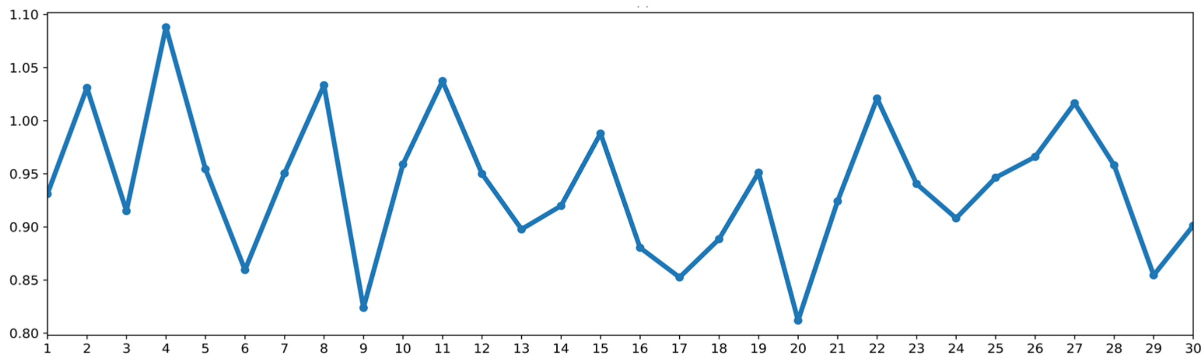

Figure 2: Energy unit price change throughout a month of smart grid interactions among 1000 agents, where the change causes defective behavior until $t = 7$(Figure 1).

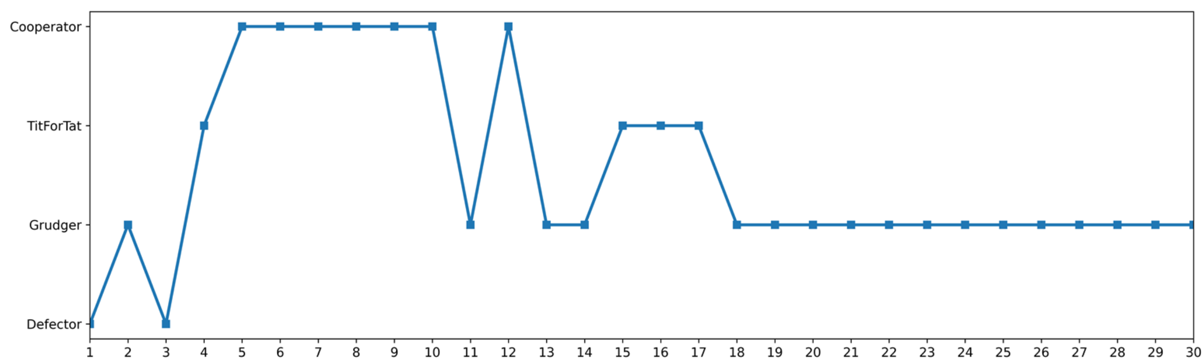

Figure 3: Disclosure of set Ω_*that mitigates the defection rate (Figure 1) down to zero level.

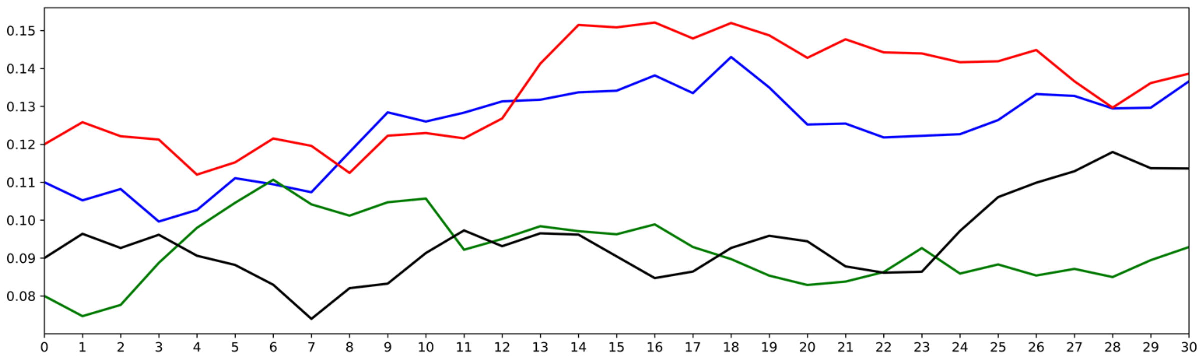

Figure 4: Evolution of payoff matrix (5) throughout a month of smart grid interactions among 1000 agents, where initial condition (5) is still satisfied in the last seven time slots (from $t = 24$ to $t = 30$).

Therefore, set Ω_* in the example above, just like set (20) in the general case, is a trajectory of dominant strategies, and this trajectory is a behavioral recovery path toward grid stability. But what is more important is that how the trajectory is actually obtained: it is the aggregating "work" of penalization level α, cooperation rewarder β, and price deviation corrector η. Nevertheless, each agent decides eventually on its own, and this along with behavioral inertia effects makes individual choices of agents slightly various throughout $T_{\max}$ time slots (Figure 5).

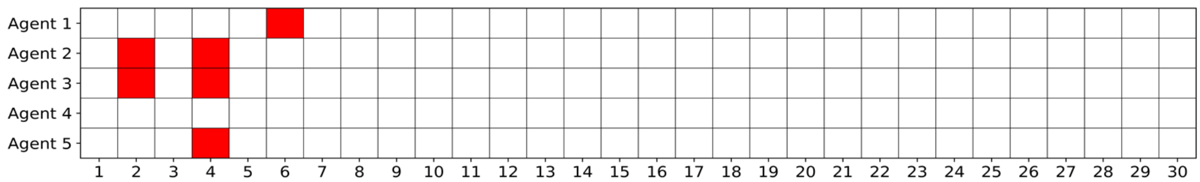

Figure 5: The factual actions of a quintet of agents throughout a month of smart grid interactions, where darker squares correspond to defections with the defection rate shown in Figure 1.

Apart from exemplified "Cooperator", "TitForTat", and conditionally cooperative strategy "Grudger", there are other cooperative-reciprocal strategies like "WinStayLoseShift", "GTFT" (a generous "TitForTat"), "TitFor2Tats" (retaliation after two defections rather than one), etc. [18, 19, 21]. Denote a set of such strategies by S_{CRS}. Hence, if $\omega_*^{(t)} \notin S_{\mathrm{CRS}}$ then the smart grid recovery has not been completed and $t < T_*$, regardless whether $d(t) \leqslant d_*$. Besides, the period of survivability $T_{\max}$ of set Ω_* (now, let it be called recovery trajectory Ω_*) is determined not only by (18), but also by the situation, when a non-cooperative strategy (e. g., like "Defector", "Cycler", "Alternator", "Random", etc.) occurs after a row of strategies from set S_{CRS}. Denote a set of non-cooperative strategies by $\bar{S}_{\mathrm{CRS}}$. Hence, if $\omega^{(t)} \in \bar{S}_{\mathrm{CRS}}$ following a sequence of strategies in S_{CRS} (that indicates a recovery period), then $\omega^{(t)} \neq \omega_*^{(t)}$ and $T_{\max} = t - 1$.

Preliminary computational simulation reveals that average payoff $\tilde{\pi}$ grows as penalization level α and cooperation rewarder β are increased, but $\tilde{\pi}$ drops as price deviation corrector η is increased. Dependence $\tilde{\pi}(\eta)$ resembles an exponential one and it is more stable than both dependences $\tilde{\pi}(\alpha)$ and $\tilde{\pi}(\beta)$. Meanwhile, the defection rate nonlinearly drops as any of parameters α, β, η is increased, where exponential-like dependences prevail also. Besides, price deviation corrector η has a seemingly stronger impact on the defection rate than the impact of both α and β (it is weak and more chaotic). Therefore, it is reasonable to approximately minimize the defection rate first by η (within the inner loop of successive ascent), and then by α and β simultaneously (it will be the outer loop of successive descent). A triple $\{\alpha^*, \beta^*, \eta^*\}$ of parameters α, β, η, which approximately minimize the defection rate (whether to tolerable rate d_* or not) and produce set a recovery trajectory Ω_*, is the best regulative configuration for smart grid recovery under the given conditions.

To start the algorithm of determining the best regulative configuration for smart grid recovery, the minimal $\alpha_{\min}$, $\beta_{\min}$, $\eta_{\min}$ and maximal values of parameters $\alpha_{\max}$, $\beta_{\max}$, $\eta_{\max}$ are input, respectively, along with the corresponding integers m_α, m_β, m_η, each of which is the number of equal-length subintervals of lengths *a*, *b*, *h*, respectively:

$$m_\alpha = \frac{\alpha_{\max} - \alpha_{\min}}{a}, \; m_\beta = \frac{\beta_{\max} - \beta_{\min}}{b}, \; m_\eta = \frac{\eta_{\max} - \eta_{\min}}{h}.$$

Another three inputs are tolerable defection rate d_*, desired level of average payoff $\tilde{\pi}$, and desired duration L^*_{recov} of recovered grid regime, where the output duration of recovered grid regime is $L_{\text{recov}} = T_{\max} - T_*$ (where T_* is the recovery duration). The algorithms returns triple $\{\alpha^*, \; \beta^*, \; \eta^*\}$ that must ensure the corresponding set $G_*\,(N, \; T)$ under the same conditions of price volatility and payoff fluctuation (Figure 6).

Iterations along α and β constitute rough adjustment, and iterations along η constitute more precise adjustment (fine-tuning). The two dot-rectangled inequalities

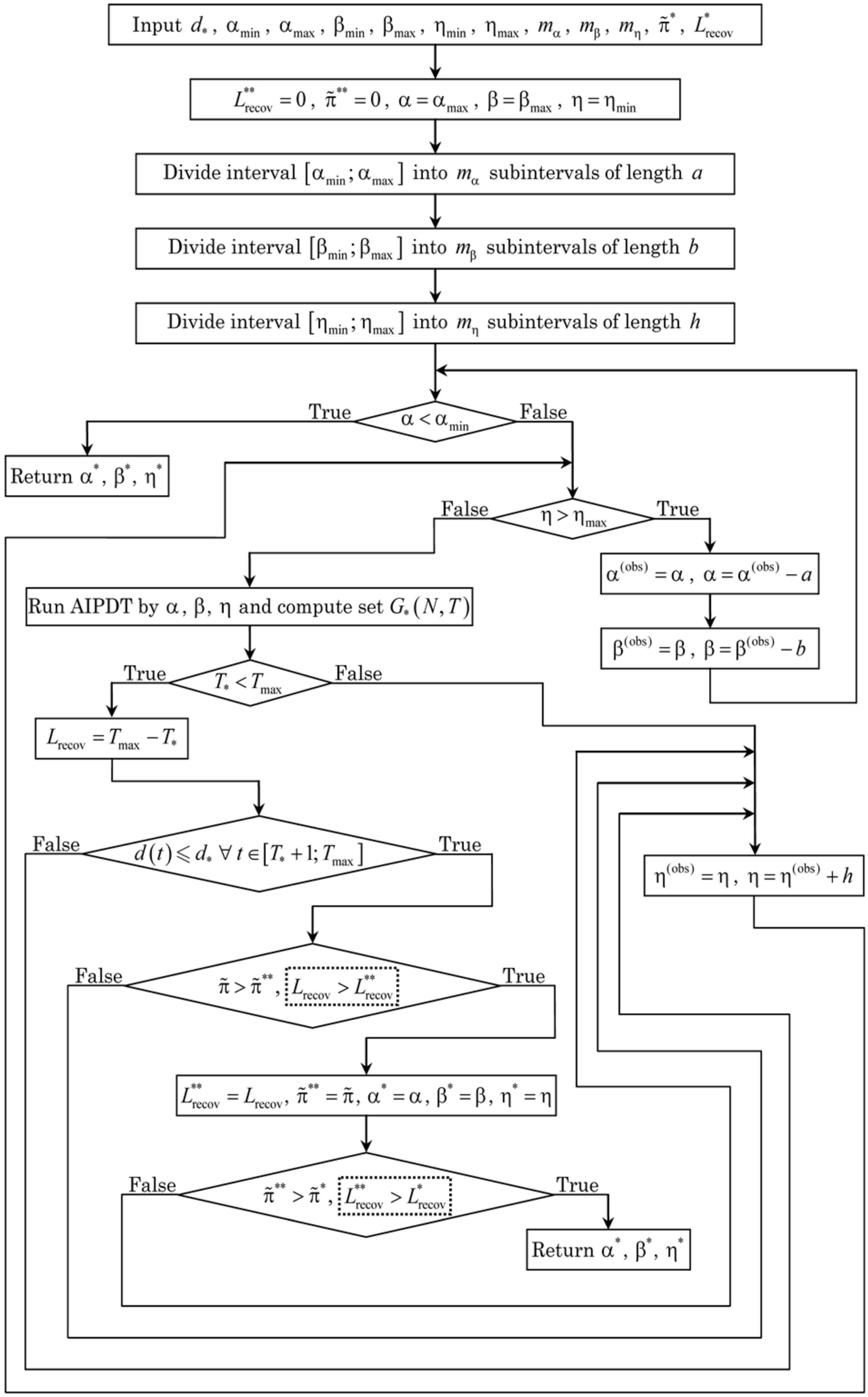

Figure 6: The algorithm of determining the best regulative configuration for smart grid recovery.

$$L_{\text{recov}} > L^{**}_{\text{recov}} \text{ and } L^{**}_{\text{recov}} > L^{*}_{\text{recov}} \tag{23}$$

next to basic conditions

$$\tilde{\pi} > \tilde{\pi}^{**} \text{ and } \tilde{\pi}^{**} > \tilde{\pi}^{*} \tag{24}$$

are optional. And if inequalities (23) are added to the criterion of the strategy (recovery trajectory) applicability, this likely either changes solution $\{\alpha^*, \beta^*, \eta^*\}$or the algorithm returns non-optimal triple $\{\alpha^*, \beta^*, \eta^*\}$. For instance, the smart grid recovery with parameters (22) and payoff matrix (21) by singly (24) is made under

$$\alpha^* = 0.25, \beta^* = 0.25, \eta^* = 0.015,$$

where

$$\tilde{\pi}^{**} \in (0.114;\ 0.119) (\text{by } \tilde{\pi}^* = 0.113), T_* = 6, L^{**}_{\text{recov}} = 24,$$

the resulting defection rate mitigation is very similar to that in Figure 1 and the resulting recovery trajectory is very similar to that in Figure 3 (obviously, the price change in Figure 2 cannot be reproducible, just like the payoffs in Figure 4, due to stochasticity). But when inequalities (23) are added to the criterion by $L^{*}_{\text{recov}} = 20$, the solution changes to

$$\alpha^* = 0.21, \beta^* = 0.21, \eta^* = 0.008,$$

where

$$\tilde{\pi}^{**} \in (0.115;\ 0.122) (\text{by } \tilde{\pi}^* = 0.113), T_* = 3, L^{**}_{\text{recov}} = 6,$$

and, despite the defection rate quickly drops to 0 (Figure 7), the price change (Figure 8) somehow impacted agents behavior (Figure 9) so that they start alternating between cooperation and defection since $t = 10$(Figure 10), let alone that two out of four inequalities (18) turn true at $t = 28$.

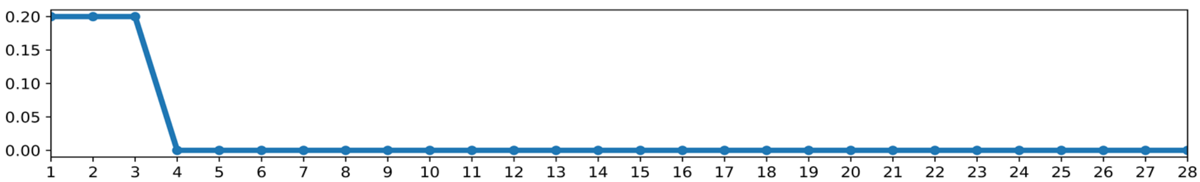

Figure 7: Defection rate drops off 0.2 rate to almost 0 level ($d_* = 0.003$) just in three time slots.

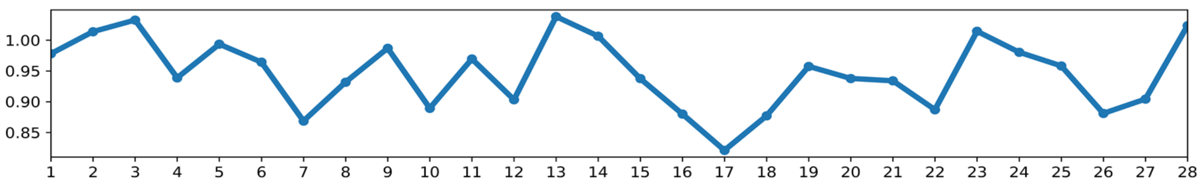

Figure 8: Energy unit price change stochasticity resembles that in Figure 2.

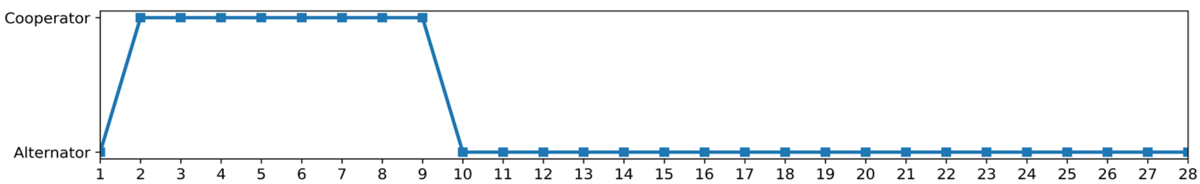

Figure 9: Recovery trajectory abruptly holds at strategy “Cooperator”, but then switches back to strategy “Alternator”, which is not in set $S_{\text{CRS}}$of cooperative-reciprocal strategies.

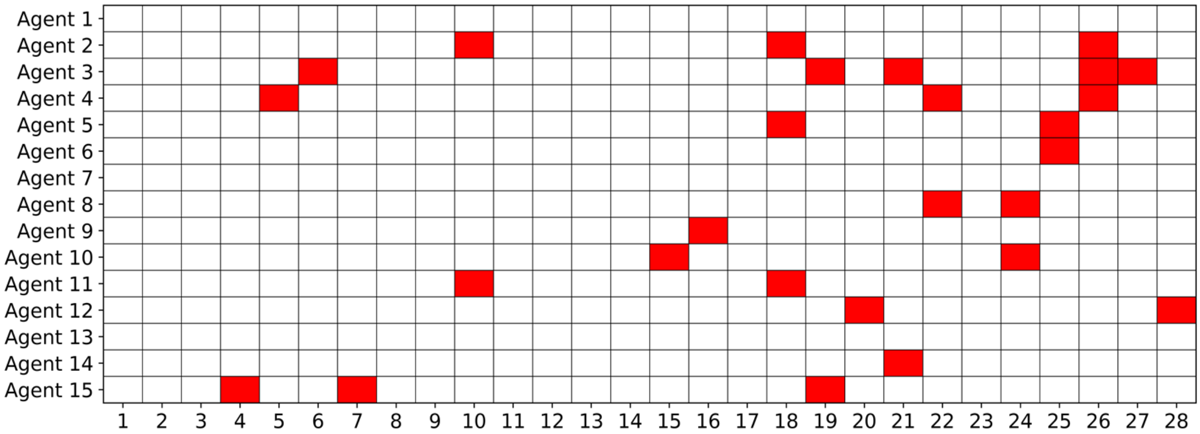

Figure 10: The factual actions of a subset of 15 agents (re-enumerated) throughout four weeks of smart grid interactions, where darker squares correspond to defections with the defection rate shown in Figure 7 (due to the maximum weekly inertia period, "Alternator" exposes sporadically).

It is noteworthy that the average payoff shown above as intervals varies more as price deviation corrector $\eta$is decreased. The impact of penalization level $\alpha$and cooperation rewarder βon the average payoff is weak, and it is even weaker with respect to the defection rate if $\alpha > 0.05$and $\beta > 0.05$or so. Therefore, it is firstly reasonable to estimate appropriate values of $\alpha$and β, and then set $\alpha_{\min} = \alpha_{\max}$, $\beta_{\min} = \beta_{\max}$before running the algorithm.

4.5. Period of survivability

The algorithm in Figure 6 even with basic conditions (24) singly tries to find the best regulative configuration for smart grid recovery by requirements

$$d(t) \leqslant d_* \ \forall \ t \in [T_* + 1; \ T_{\max}] \tag{25}$$

and

$$L_{\text{recov}} = T_{\max} - T_* > 0, \tag{26}$$

along with payoff maximization. If inequalities (23) are included, the complication builds up so that there may be no solution. In general, the algorithm is an attempt to solve a stochastic multi-objective constrained optimization problem, where satisfying inequality (25) is far easier than satisfying inequality (26), because the latter additionally requires specific set S_{CRS}in recovery trajectory Ω_*after mapping (19). Inequality (26) can be forced true by either shortening the period of recovery T_*or elongating the period of survivability $T_{\max}$, or by optimizing them both. Figure 6 shows that elongating the duration of recovered grid regime is currently optional, but it can be substituted with the option for elongating the period of survivability $T_{\max}$by inserting the two dot-rectangled inequalities

$$T_{\max} > T_{\max}^{**} \text{ and } T_{\max}^{**} > T_{\max}^{*} \tag{27}$$

instead of (23), where $T_{\max}^{**}$is currently the best value (in an algorithm run), $T_{\max}^{*}$is a desired value set to 0 at the start (instead of setting $L_{\text{recov}}^{*} = 0$). The computational simulation reveals that elongation of the period of survivability is solved far easier than shortening the period of recovery. This is explained by that the latter is usually shorter itself and more stochastic.

5. Conclusions

The developed iterative algorithm realizes the suggested method of determining a smart grid recovery strategic trajectory, which is done by the penalization level in the price change model, cooperation rewarder in the cooperation and defection payoffs, and the price deviation corrector in all payoffs but the cooperation payoff in the payoff matrix. The three parameters are iteratively optimized by taking into account that the price deviation corrector has a more stable impact on both the average payoff and defection rate. Eventually, the price deviation corrector is the most influential parameter in the method. The recovery strategic trajectory is mathematically a very complex object, whose principally random components are ordered cooperative-reciprocal strategies and the periods of recovery and sustainability. The efficiency of cooperative-reciprocal strategies in the recovery trajectory depends heavily on the initial inputs to the algorithm, but when the intervals of the penalization level, cooperation rewarder, and the price deviation corrector are appropriately divided into subintervals, the solution is found within 10 iterations or fewer. This computation, needed when the smart grid is "disturbed" by defectors, is a virtual run immediately followed the perturbation, and this run result (recovery strategic trajectory) returns the smart grid system to normal state by minimum possible intervention, thereby preserving democratized decentralization and supporting long-term grid stability and controllability. As the period of survivability expires, an update of the strategic trajectory may be needed.

Further research can be directed toward alternative approaches to building the recovery trajectory. One of such approaches is the use of genetic algorithms [22]. However, one must remember that the recovery trajectory should be built quickly as it is a real-time response to system perturbations, whereas fine-tuning a genetic algorithm and running it virtually takes significantly longer computational time than running the suggested algorithm.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] M. J. B. Kabeyi, O. A. Olanrewaju, Smart grid technologies and application in the sustainable energy transition: a review, International Journal of Sustainable Energy 42(1) (2023) 685–758. doi:10.1080/14786451.2023.2222298.

[2] H. A. Muqeet, R. Liaqat, M. Jamil, A. A. Khan, A state-of-the-art review of smart energy systems and their management in a smart grid environment, Energies 16(1) (2023) 472. doi:10.3390/en16010472.

[3] C. He, H. Geng, K. Rajashekara, A. Chandra, Analysis and control of frequency stability in low-inertia power systems: A review, IEEE/CAA Journal of Automatica Sinica 11(12) (2024) 2363–2383. doi:10.1109/JAS.2024.125013.

[4] S. Dorji, A. A. Stonier, G. Peter, R. Kuppusamy, Y. Teekaraman, An extensive critique on smart grid technologies: recent advancements, key challenges, and future directions, Technologies 11(3) (2023) 81. doi:10.3390/technologies11030081.

[5] R. Alsaigh, R. Mehmood, I. Katib, AI explainability and governance in smart energy systems: A review, Frontiers in Energy Research 11 (2023) 1071291. doi:10.3389/fenrg.2023.1071291.

[6] A. R. Kojonsaari, J. Palm, The development of social science research on smart grids: a semi-structured literature review, Energy, Sustainability and Society 13 (2023) 1. doi:10.1186/s13705-023-00381-9.

[7] R. Axelrod, The Evolution of Cooperation, Basic Books, 1984.

[8] V. V. Romanuke, Compliance-and-defiance dilemma game best strategy for three and four agents, Statistics, Optimization and Information Computing 14(5) (2025) 415–433. doi:10.19139/soic-2310-5070-2370.

[9] A. J. Stewart, J. B. Plotkin, From extortion to generosity, evolution in the Iterated Prisoner's Dilemma, Proceedings of the National Academy of Sciences 110(38) (2013) 15348–15353. doi:10.1073/pnas.1306246110.

[10] M. A. Nowak, Five rules for the evolution of cooperation, Science 314(5805) (2006) 1560–1563. doi:10.1126/science.1133755.

[11] V. V. Romanuke, Best strategy in multi-agent symmetric dilemma game with a fixed fine for total defiance, Journal of Mathematics and Applications 48 (2025) 73–89. doi:10.7862/rf.2025.4.

[12] H. Wu, Y. Gao, A real-time locational marginal pricing strategy for smart grids based on demand-side management, Open Journal of Applied Sciences 16(1) (2026) 143–167. doi:10.4236/ojapps.2026.161010

[13] S. Frank, I. Steponavice, S. Rebennack, Optimal power flow: a bibliographic survey I, Energy Systems 3 (2012) 221–258. doi:10.1007/s12667-012-0056-y.

[14] Q. Li, Y. Shi, Y. Jiang, Y. Shi, H. Wang, H. V. Poor, A distributionally robust model predictive control for static and dynamic uncertainties in smart grids, arXiv:2403.16402 [eess.SY] (2024). doi:10.48550/arXiv.2403.16402.

[15] N. V. S. J. Jami, J. Kardoš, O. Schenk, H. Köstler, AI driven near real-time locational marginal pricing method: a feasibility and robustness study, arXiv:2306.10080 [cs.CE] (2023). doi:10.48550/arXiv.2306.10080.

[16] W. Saad, Z. Han, H. V. Poor, T. Basar, Game-theoretic methods for the smart grid: an overview of microgrid systems, demand-side management, and smart grid communications, IEEE Signal Processing Magazine 29(5) (2012) 86–105. doi:10.1109/MSP.2012.2186410.

[17] C. Zhang, J. Wu, Y. Zhou, M. Cheng, C. Long, Peer-to-peer energy trading in a microgrid, Applied Energy 220 (2018) 1–12. doi:10.1016/j.apenergy.2018.03.010.

[18] J. W. Weibull, Evolutionary Game Theory, MIT Press, 1997.

[19] W. H. Sandholm, Population Games and Evolutionary Dynamics, MIT Press, 2010.

[20] K. Sigmund, The Calculus of Selfishness, Princeton University Press, 2010.

[21] https://github.com/Axelrod-Python/Axelrod

[22] B. D. Deebak, F. Al-Turjman, Sustainable Networks in Smart Grid, Elsevier, 2022.