**MM Science Journal | www.mmscience.eu**
ISSN 1803-1269 (Print) | ISSN 1805-0476 (On-line)
**Special Issue | HSM 2025**
**18th International Conference on High Speed Machining**
October 15-16, 2025, Metz, France
DOI: 10.17973/MMSJ.2025_12_2025164

**HSM2025-44933**

# STATE-DEPENDENT REGENERATIVE STABILITY-CONSTRAINED REINFORCEMENT LEARNING OPTIMIZATION FOR MACHINING EFFICIENCY IN ROBOTIC MILLING

Si Hao Mao[1], Songtao Ye[1,2], Yanru Jiang[1], Xiaojian Zhang[1]*, Sijie Yan[1], Han Ding[1]

[1]State Key Laboratory of Intelligent Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Luo Yu Road No.1037, Wuhan, 430074, Hubei, China

[2]HUST-Wuxi Research Institute, Wuxi, 214174, Jiangsu, China

*Corresponding author; e-mail: xjzhang@hust.edu.cn;

**Abstract**

Improving robotic milling efficiency enhances productivity and reduces costs. While feed rate and spindle speed critically influence efficiency, chatter instability complicates their optimization. Existing stability constraints ignore state-dependent regenerative mechanisms, while the nonlinear effects of feed further complicate optimization. Although adjusting the robotic configuration improves stability, operational constraints such as joint singularity should be considered. This work proposes a reinforcement learning (RL) method to jointly optimize feed rate, spindle speed, and robotic configuration. RL dynamically maximizes efficiency in high-dimensional space using a reward function integrating stability and operability. Simulation results validate the method's superior performance.

**Keywords:**

State-dependent regenerative stability, Machining efficiency, DDPG, Robotic milling

## 1 INTRODUCTION

Feed rate and spindle speed optimization play a pivotal role in improving machining efficiency by maximizing material removal rates (MRR) subject to varying constraints [Chiang 1995]. Traditional approaches, such as constant cutting force strategies [Oh 2023], constant feed-per-tooth methods [Vavruska 2023], fixed cutting width planning [García 2021], and servo-drive actuation limit constraints [Sencer 2008] [Erkorkmaz 2013], have demonstrated effectiveness in process parameter optimization under quasi-static conditions. However, few studies address constrained optimization under dynamic machining conditions.

Recent advances highlight the need to account for dynamic forces—including inertia and Coriolis effects—to plan conservative feed strategies [Chu 2025]. Nevertheless, a critical gap remains: the absence of stability-constrained optimization methods in dynamic systems. The spindle speed and feed rate joint govern the nominal feed per tooth, e.i. static chip thickness. While conventional milling stability models (e.g., constant delay differential equations [Zhao 2001]) treat the static chip thickness as an independent factor governing stability, forced vibrations induced by cutting forces dynamically modulate system stability [Bachrathy 2011]. This state-dependent stability imposes coupled constraints on feed planning, fundamentally limiting the traditional methods' applicability.

In robotic milling, these challenges are exacerbated by two configuration-dependent factors: 1) robot operational performance, such as joint singularity [Wang 2022]; and 2) the configuration-dependent tool tip dynamics significantly affect the state-dependent stability [Chen 2022]. Although redundancy angle optimization has been proposed to enhance operational performance, critical state-dependent stability remains overlooked in such methodologies.

This paper proposes the RL-based efficiency optimization framework that explicitly integrates the state-regenerative stability constraint. The key innovations include:

1. Leveraging the deterministic policy gradient (DDPG) algorithm, our method generates continuous feed-spindle speed trajectories without requiring pre-defined interpolation patterns, overcoming the discontinuities through inertial filtering analogous to Finite Impulse Response (FIR) filtering [Tajima 2018].
2. We introduce the state-dependent regenerative stability constraint into the efficiency optimization, where the stability improvement effect imposes a new lower bound on feed parameters.
3. By simultaneously optimizing redundancy angle, speed rate, and spindle speed, the proposed method achieves chatter-free machining efficiency improvement compared to the fixed redundancy angle strategy, as validated in our simulation.

The remainder of this paper is organized as follows: Section 2 details the state-dependent stability modeling framework, Section 3 presents the RL-based optimization architecture, Section 4 provides simulation validation, and Section 5 concludes with future research directions.

## 2 STATE-DEPENDENT STABILITY

In conventional milling stability models, the influence of the residual surface from the previous tooth pass on the cutting process is regarded as a single regenerative effect (red), as shown in Fig. 1a, where the feed rate is treated as a static component that does not affect the system stability. However, under high-feed conditions, the low stiffness of the robotic system may induce nonlinear phenomena such as contact loss (green) and multiple regenerative effects (orange), as shown in Fig. 1b. This results in a state-dependent delay system, which complicates the stability analysis.

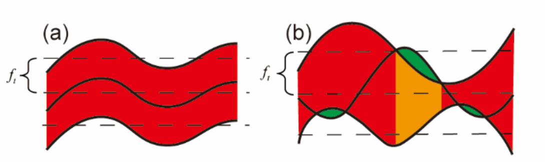

*Fig.1: (a) Small amplitude vibration with single regenerative phenomena. (b) Large amplitude vibration with contact loss and multiple regenerative phenomena.*

Robotic milling is modeled as a two-degree-of-freedom oscillator-feedback system, as illustrated in Fig. 2, where the tool feeds along the negative X-direction, and the vibrational displacements of the tool center point (TCP) in the X-Y frame are expressed as $\boldsymbol{o}(t)=[x(t)\quad y(t)]^T$ . To analyze the state-dependent regenerative stability, we follow the equilibrium-linearization methodology in [Bachrathy 2011].

The equilibrium $\tilde{\boldsymbol{o}}\in C[0,T;\square^2]$ with the dominant period $T = 60/(\Omega\cdot N_t)$ associated with the spindle speed $\Omega$ and the number of teeth $N_t$ , can first be derived from the governing dynamics equation (Eq. (1)).

$$\mathbf{M}\ddot{\boldsymbol{o}}(t)+\mathbf{C}\dot{\boldsymbol{o}}(t)+\mathbf{K}\boldsymbol{o}(t)=\mathbf{F}\left(h\left(\boldsymbol{o}(t),\boldsymbol{o}\left(t-t\left(t,\boldsymbol{o}_t\right)\right)\right)\right), \quad (1)$$

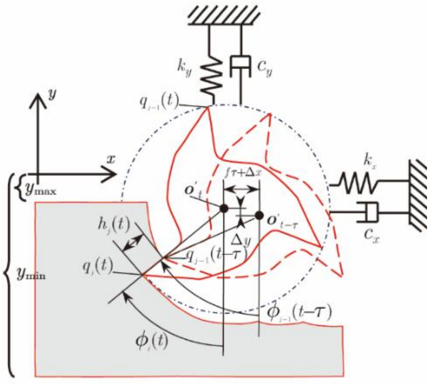

*Fig. 2: Schematic of the vibration-dependent robotic milling system.*

where the left-hand term represents the second-order dynamic characteristics of the system, while the right-hand term represents the cutting forces associated with state-dependent delay $t\left(t,\boldsymbol{o}_t\right),\boldsymbol{o}_t(s)=\boldsymbol{o}(t+s),s\in[-r,0],r\in\square^+$ , chip thickness $h$ , and tool vibrations $x(t),y(t)$ , among whom an implicit constraint can be formulated as follows:

$$\begin{bmatrix} x(t-t_j^m)-R\sin(f_{j-m}(t-t_j^m))+ft_j^m \\ y(t-t_j^m)-R\cos(f_{j-m}(t-t_j^m)) \end{bmatrix}=\begin{bmatrix} x(t)-(R-h_j^m)\sin(f_j(t)) \\ y(t)-(R-h_j^m)\cos(f_j(t)) \end{bmatrix},$$

(2)

where $R$ denotes the nominal radius of the tool, $f$ denotes the feed rate, and the immersion angle $f_{j\text{-}m}\left(t-t_j^m\right)$ implies that the $m$ regenerative delay $t_j^m$ corresponding to the $j$th tooth depends on the vibrations generated by the $(j\text{-}m)$th tooth.

Due to the computational demand of iterative stability evaluation in robotic milling planning, which involves root-finding for Eq. (2), we transform the root-finding problem of state-dependent delays into the numerical integration process under the periodic solution assumption to alleviate computational costs.

Specifically, eq. (2) is recast into the following form :

$$R\sin\left(t_j^m\Omega-2mp/N_t\right)+\left(ft_j^m+x\left(t-t_j^m\right)-x(t)\right)\cos\left(f_j(t)\right)$$
$$-\left(y\left(t-t_j^m\right)-y(t)\right)\sin\left(f_j(t)\right)=0,$$

(3)

$$h_j^m=R\left(1-\cos\left(t_j^m\Omega-2pm/N_t\right)\right)$$
$$+\left(ft_j^m+x\left(t-t_j^m\right)-x(t)\right)\sin\left(f_j(t)\right)$$
$$+\left(y\left(t-t_j^m\right)-y(t)\right)\cos\left(f_j(t)\right)+\left(y\left(t-t_j^m\right)-y(t)\right)\cos\left(f_j(t)\right).$$

(4)

By differentiating both sides of Eq. (3) with respect to time $t$ , we obtain

$$\begin{aligned}&R\Omega\cos\left(t_j^m\Omega-2pm/N_t\right)\dot{t}_j^m\\&+\left(f\dot{t}_j^m+\dot{x}\left(t-t_j^m\right)\left(1-\dot{t}_j^m\right)-\dot{x}(t)\right)\cos\left(f_j(t)\right)\\&-\left(ft_j^m+x\left(t-t_j^m\right)-x(t)\right)\sin\left(f_j(t)\right)\Omega\\&-\left(\dot{y}\left(t-t_j^m\right)\left(1-\dot{t}_j^m\right)-\dot{y}(t)\right)\sin\left(f_j(t)\right)\\&-\left(y\left(t-t_j^m\right)-y(t)\cos\left(f_j(t)\right)\Omega\right)=0.\end{aligned} \quad (5)$$

Rewriting and simplifying the Eq. (5) yields:

$$\dot{t}_j^m=\frac{G_1+G_2}{G_3+G_4}, \quad (6)$$

where

$$\begin{aligned}G_1&=(\dot{x}(t)-\dot{x}\left(t-t_j^m\right)+y\left(t-t_j^m\right)\Omega-y(t)\Omega)\cos\left(f_j(t)\right),\\G_2&=\left(ft_j^m\Omega+x\left(t-t_j^m\right)\Omega-x(t)\Omega+\dot{y}\left(t-t_j^m\right)-\dot{y}(t)\right)\sin\left(f_j(t)\right),\\G_3&=R\Omega\cos\left(t_j^m\Omega-2pm/N_t\right)+f\cos\left(f_j(t)\right),\\G_4&=-\dot{x}\left(t-t_j^m\right)\cos\left(f_j(t)\right)+\dot{y}\left(t-t_j^m\right)\sin\left(f_j(t)\right).\end{aligned} \quad (7)$$

Once the equilibrium at the initial time $t_0$ is obtained via root-finding, the state-dependent delay over subsequent time intervals can be efficiently computed using the explicit fourth-order Runge-Kutta method in eq. (6).

By integrating eq. (1), eq. (4), and eq. (6), the equilibrium $\tilde{\boldsymbol{o}}(t)$ can be rapidly obtained. This approach thereby facilitates efficient agent-environment interaction in the subsequent RL framework by decoupling state-dependent delays from real-time control actions.

For stability analysis, the variational periodic linear system is derived as the linearization of periodic state-dependent delay differential equations around the equilibrium $\tilde{\boldsymbol{o}}(t)$ as follows:

$$\begin{aligned}&\mathbf{M}\Delta\ddot{\boldsymbol{o}}(t)+\mathbf{C}\Delta\dot{\boldsymbol{o}}(t)+\mathbf{K}\Delta\boldsymbol{o}(t)=\\&\sum_{j=1}^{N_t}\boldsymbol{W}_j\left(t,\tilde{\boldsymbol{o}}_t\right)\left(\Delta\boldsymbol{o}(t)\text{-}\ \Delta\boldsymbol{o}\left(t-t_j\left(t,\tilde{\boldsymbol{o}}_t\right)\right)\right),\end{aligned} \quad (8)$$

In constant delay models, periodic directional cutting coefficients $\boldsymbol{W}$ depend only on the immersion angle and cutting stiffness. In contrast, state-dependent directional coefficients exhibit additional dependencies on equilibrium. In [Bachrathy 2011], this effect is modeled as an additional term to the nominal coefficients, which has been reported to be negligible compared to the reference values under the small amplitude vibrations. However, Fig. 3 demonstrates significant deviations in directional cutting coefficients during different feed rates due to the large amplitude vibration (with parameters consistent with [Xin 2022]).

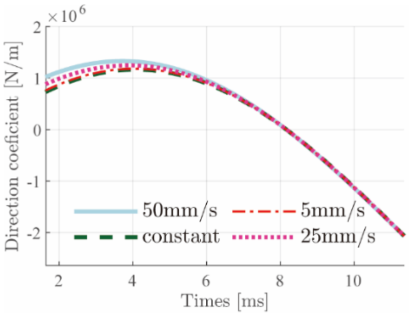

*Fig. 3: X-Y direction coefficient at spindle speed 900 RPM and depth of axial 0.9 mm during different feed rates.*

Finally, analogous to time-varying delay systems, the state transition matrix $\Phi$ of the discretized approximation is constructed using the semi-discretization method or numerical integration techniques [Ding 2011]. Following Floquet theory, the max modulus of characteristic multipliers $|m|_{\max}$ of the $\Phi$ serves as chatter indicators: the system becomes unstable when $|m|_{\max} > 1$.

Notably, the nonlinear phenomena including loss contact may interrupt the regenerative mechanism to enhance system stability. As illustrated in Fig. 4, which plots feed rate against chatter indicators, this phenomenon provides an additional incentive to increase feed rates beyond traditional efficiency-driven optimizations.

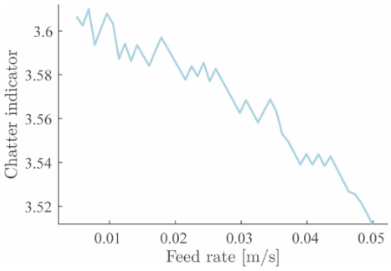

*Fig. 4: Chatter vs Feed Rate Interaction Characteristics.*

Increasing feed rates necessitates a proportional rise in spindle speed $\Omega$ to suppress forced excitation. This defines a set of planning parameters $(\Omega, f)$ constrained by stability indicators. Such synergy was unattainable under conventional stability constrains, where feed rates negligibly impact stability margins. In contrast, traditional parameter optimization focused on $(\Omega, a_p)$ (speed vs. depth of cut), but varying $a_p$ induces bottom-edge cutting effects that complicate layer-by-layer toolpath planning.

Moreover, given the adverse effects of high-impact transients during high-speed material engagement—particularly on surface integrity and machine tool longevity—parameter planning should initiate from regimes with low dynamic. Fixing a low-value initial parameter set transforms the optimization task into an initial value problem (IVP) with fixed boundary conditions. This reformulation significantly simplifies the RL architecture proposed in Section 3, enabling it to achieve enhanced convergence efficiency while maintaining physical feasibility.

## 3 DDPG FOR MACHINING EFFICIENCY

The Deep Deterministic Policy Gradient (DDPG) algorithm [Sumiea 2024] is employed to address the continuous control challenges in machining parameter optimization. In contrast to conventional gradient-based methods that struggle with non-smooth optimization landscapes, DDPG's actor-critic architecture enables direct policy learning in high-dimensional parameter spaces through offline policy updates. This framework inherently respects three critical constraints:

1. Physical realizability: Inertially filtered actuation commands enforce smooth parameter transitions with parameter limits;
2. Stability guarantees: Nonlinear constraint derived from indicator (Section 2) is embedded in the reward shaping process;
3. Machining efficiency: The MRR is optimized through adaptive exploration-exploitation.

In DDPG updates, it requires the definition of state, action, and reward. Consider a uniform arc-length sampling along the specified path $\mathrm{L} = \{l_1, l_2, \cdots, l_t, \cdots, l_n\}$. The segment index $id \in Z^+$ forms part of the state space, i.e., $s_{pa} = id,\ id = 1, 2, \cdots, n$. This index, combined with the redundant angle $g$ (which originates from the industrial robot's six degrees of freedom exceeding the five required for milling tasks), collectively modifies the tooltip dynamics. Their stability effects are captured in the left-hand side of Eq. (1). The prediction of dynamic parameters (refer to [Rasmussen 2005]) is achieved via multi-task Gaussian process regression (MTPGR). This aspect is not elaborated here as it lies beyond the scope of this paper. Therefore, the state space is defined as $S = (s_{pa}, \Omega, f, g)$.

Constrained by physical realizability, the action space isomorphic to the state space is defined as:

$$A = (a_{pa}, a_{pr}) = (\Delta s_{pa}, \Delta\Omega, \Delta f, \Delta g) \tag{9}$$

For the segment index, the state transition is deterministic since the path progression follows a predefined sequence:

$$\begin{aligned} & p(s_{pa,t+1} = id + a_{pa,t} \,|\, s_{pa,t} = id) \\ & = p(s_{pa,t+1} = id + 1 \,|\, s_{pa,t} = id) = 1. \end{aligned} \tag{10}$$

For other controllable parameters $s_{pr} = (\Omega, f, g)$, the state transition is governed by inertial filtering and parameter limits. Specifically :

$$\begin{aligned} & p\left(s_{pr,t+1} = \mathrm{clip}\left(s_{pr,t} + \hat{a}_{pr,t+1}, s_{pr,t}^{\max}, s_{pr,t}^{\min}\right) \,|\, s_{pr,t}\right) = 1, \\ & \Delta\hat{a}_{pr,t+1} = a a_{pr,t+1} + (1 - a) a_{pr,t}. \end{aligned} \tag{11}$$

where inertia coefficient $a \in (0,1)$, and $s_{pr,t}^{\max}, s_{pr,t}^{\min}$ denotes the lower and upper bounds, respectively. The clip operator guarantees that the constrained optimization problem can attain global extrema within bounds.

To balance robotic operational performance and machining efficiency, the reward function is formulated as a multi-objective combination :

$$r_t = w_o r_{o,t} + w_e r_{e,t} + w_p r_{p,t} \tag{12}$$

where the weight coefficients satisfy $w_o + w_e + w_p = 1$, and each sub-reward is defined as follows:

**Operational Performance Reward:**

Industrial robots may lose maneuverability when approaching kinematic singularities, where the Jacobian matrix becomes rank-deficient. Proximity to singular configurations also degrades stiffness, adversely affecting machining accuracy. The kinematic redundancy allows singularity avoidance by adjusting the redundant angle $g_t \in [p, p]$. The mapping from $g_t$ to joint angle $\boldsymbol{q}_t$ is resolved through the inverse kinematics formulation:

$$\boldsymbol{q}_t = \mathrm{IK}(g_t) \tag{13}$$

The singularity proximity is quantified using the normalized condition number  of the Jacobian:

$$K\left(\boldsymbol{J}_N\left(\boldsymbol{q}_t\right)\right)=\frac{1}{n}\sqrt{\operatorname{tr}\left(\boldsymbol{J}_N\boldsymbol{J}_N^T\right)\operatorname{tr}\left(\left(\boldsymbol{J}_N\boldsymbol{J}_N^T\right)^{-1}\right)},\ \boldsymbol{J}_N\in\square^{n\times n} \quad (14)$$

where $n$ is the joint space dimension and $\boldsymbol{J}_N$ denotes the characteristic-length normalized Jacobian. The singularity-avoidance reward is inversely proportional to this metric:

$$r_{o,t}=\frac{1}{K\left(\boldsymbol{J}_N\left(\boldsymbol{q}_t\right)\right)+\text{ň}} \quad (15)$$

with $\text{ň}=10^{-6}$ preventing numerical instability near singularities.

**Efficiency Reward:**

Given the uniform arc-length segmentation $l_t$ along the toolpath, the traversal time for each segment is inversely proportional to the feed rate $f_t$. To incentivize high-speed machining while respecting actuator limits, the efficiency reward $r_{e,t}$ is designed as a piecewise linear function:

$$r_{e,t}=\begin{cases}0.5\dfrac{f_t}{f_{t,\max}}\\ 3\left(\dfrac{f_t}{f_{t,\max}}-0.7\right)+0.35\end{cases} \quad (16)$$

This implies that, under no constraints, a feed rate closer to the maximum allowable value yields more optimal performance. Therefore, the system should be guided to prioritize optimization within the high-speed region to enhance machining efficiency.

**Stability Reward:**

The max modulus of characteristic multipliers $|m|_{\max}$ from Section 2 is utilized as the stability metric. When the system remains stable, its stability margin need not be considered. However, when chatter occurs, the planning parameter adjustments should be adaptively tuned based on the modulus. When the modulus significantly exceeds unity, more aggressive parameter adjustments can be implemented to expedite the identification of stable parameter space. Conversely, when the modulus approaches unity, conservative adjustments are adopted to enhance both convergence and stability. Consequently, the reward function is formulated as:

$$r_{s,t}=-\exp\left(\operatorname{clip}\left(|m|_{\max}-1,0,5\right)\right) \quad (17)$$

where the clip operation limiting modulus to $[0,5]$ prevents gradient explosion during optimization and implemented using Python NumPy.

The transition tuple $(s_t,a_t,r_t,s_{t+1})$ constructed from above definitions serves as the input to the DDPG framework, which maximizes average reward $J_r\left(w^a,w^q\right)=\mathrm{E}_{s,a\sim A}\left[r_Q\left(s\right)\right]$ through optimizing the policy (actor network) $A\left(s_t\,|w^a\right)$ and Q-function (critic network) $Q\left(s_t,a_t\,|w^q\right)$. Algorithm 1 formalizes the DDPG learning procedure where:

```
Algorithm 1: Deep Deterministic Policy Gradient (DDPG)
Require: Actor network $A(s_t|w^a)$, critic network $Q(s_t,a_t|w^q)$, target network $A',Q'$, replay buffer $\mathrm{B}$, soft update rate $t_u$, discount factor $\Gamma>0$  $a_{w^a},a_{w^q}>0$.
Initialization:
- Actor and Critic networks with random weights $w_0^a,w_0^q$ ;
- Target networks : $w_0^{a'}\leftarrow w_0^a, w_0^{q'}\leftarrow w_0^q$ ;
- Initial point : $s_0=(1,\Omega_0,f_0,g_0)$ .
Goal : Learn an opimal policy to maximize $J_r(w^a,w^q)$ .
For time step $t$ in each episode, do
    Generate $a_t$ following $A'(s_t|w^a)$ and the observe $r_{t+1}$ , $s_{t+1}$ stored in $\mathrm{B}$ .
    Sample batth $(s_t,a_t,r_t,s_{t+1})$ from $\mathrm{B}$
    TD error :
        $d_t=r_{t+1}+\Gamma Q(s_{t+1},A(s_{t+1}|w_t^a))-Q(s_t,a_t|w_t^q)$
    Actor network update :
        $w_{t+1}^a=w_t^a+a_{w^a}\nabla_{w^a}A(s_t|w^a)(\nabla_a Q(s_t,a|w_t^q))|_{a=A(s_t)}$
    Critic network update :
        $w_{t+1}^q=w_t^q+a_{w^q}d_t\nabla_{w^q}Q(s_t,a_t|w_t^q)$
    Soft target network update :
        $w^{a'}\leftarrow t_u w^a+(1-t_u)w^{a'}$
        $w^{q'}\leftarrow t_u w^q+(1-t_u)w^{q'}$
End for
```

## 4 SIMULATION

The primary objective of this simulation study is to elucidate how RL facilitates machining efficiency optimization under state-dependent stability constraints. Experiments were conducted on a virtual ABB IRB-6660 industrial robotic manipulator, whose kinematic chain was modeled using the Modified Denavit-Hartenberg (MDH) convention with inertial parameters derived from CAD models. The MTGPR training dataset was generated through a transformation framework that bridges robotic joint-space dynamics and  tooltip vibration dynamics.

The milling trajectory was parametrized as a cubic B-spline curve interpolating four coplanar control points in the base frame, ensuring C² continuity for smooth motion generation. Environmental configurations including cutting stiffness, parameter limits, and tool properties are systematically listed in Tab. 1. Training hyperparameters such as policy network architecture (2×256 ReLU), exploration noise decay schedule, and learning rate are specified in Tab. 2.

The efficiency of the trained agent's policy was validated through trajectory reward mapping within the constrained action space. Fig. 5 illustrates the evolution of training rewards, actor loss throughout the learning process. The actor loss demonstrates an initial decrease before 300th epoch followed by gradual increase, and converges after 600th epoch, reflecting the exploration-exploitation trade-off in RL. The total average reward converges and stabilizes at approximately $\mathrm{E}_{s,a\sim A}\left[r_Q\left(s\right)\right]=4.2$ after 800th epoch.

As depicted in Fig. 6(d)-(e), the planning parameter trajectories in terms of step spindle speed, feed rate, and redundancy angle evolve from the initialization point ($\Omega_0=2000$ RPM, $f_0=0.04$ m/s, $g_0=0$ rad) defined as the IVP. The feed rate rapidly approaches the maximum velocity ($f_{\max}=0.06$ m/s) under acceleration limits but is dynamically adjusted due to state-dependent stability constraints. Simultaneously, spindle speed continuously increases to reduce feed per tooth, thereby mitigating forced excitation. The redundancy angle transformation fluctuates between 0 and 2 to improve system stability. The total machining time is $16.66$ s.

As shown in Fig. 6(a)-(c), the step reward stabilizes at $r_t=0.05\pm0.01$, the chatter indicator consistently remains below the threshold, e.i., $|m|_{\max}=0.998<1$, and the minimum singularity indicator $K\left(\boldsymbol{J}_N\left(\boldsymbol{q}_t\right)\right)=0.48$, validating the absence of kinematic singularity.

To elucidate the feed rate lower bounds imposed by state-dependent regenerative stability constraints, the heatmap of stability reward distribution across the parameter space was generated (Fig. 7). The stability-critical boundary (zero-magnitude reward contour) is delineated by red dashed lines, revealing that each spindle speed induces feed rate lower bound.

To decouple the contribution of kinematic redundancy, we conducted ablation studies with fixed redundancy angle $g=0$ rad. The planning parameter trajectories are depicted as Fig. 8(a)-(e), the constrained policy exhibits: a) The total machining time $17.53$ s in terms of $5.22\%$ longer machining cycles; b) Chatter occurrence in the latter half of the machining path. These results demonstrate that redundancy optimization synergistically improves machining efficiency and stability.

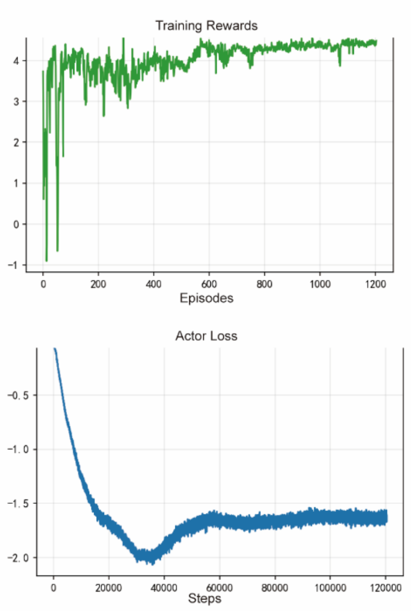

*Fig. 5: Training rewards, actor loss throughout the learning process.*

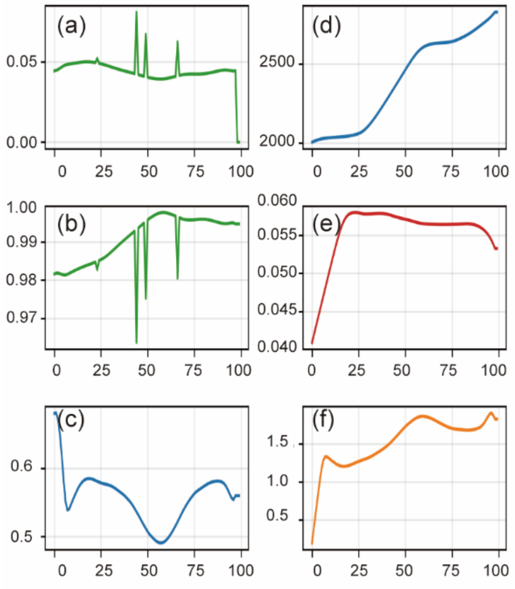

*Fig. 6: Planning parameter trajectories including (a) step reward*, *(b) chatter indicator*, *(c) singularity indicator, (d) spindle speed, (e)* step feed rate, *and (f) redundancy angle.*

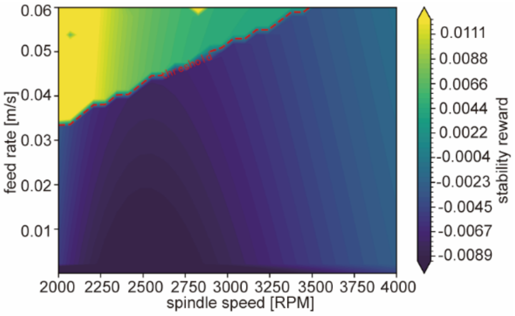

*Fig. 7: heatmap of stability reward distribution across the parameter space.*

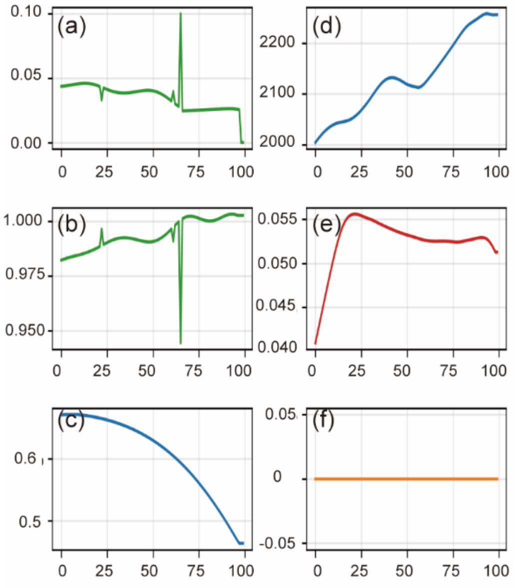

*Fig. 8: Planning parameter trajectories* with *fixed redundancy angles including (a) step reward*, *(b) chatter indicator*, *(c) singularity indicator, (d) spindle speed, (e)* step feed rate, *and (f) redundancy angle is set to 0 rad.*

## 5 SUMMARY

This paper proposes a machining efficiency optimization framework that integrates DDPG-based RL framework. The optimization simultaneously addresses constraints on robotic operational performance and state-dependent regenerative stability. Simulation results demonstrate the method's effectiveness, with kinematic redundancy optimization shown to extend achievable performance limits.

While this study employs the DDPG to ensure repeatable parameter trajectories during iterative planning, future work could explore stochastic reinforcement learning approaches to enhance generalization across variable machining conditions. For instance, Soft Actor-Critic (SAC) algorithms, which maximize policy entropy, may improve robustness to environmental perturbations. However, such methods would require advanced techniques like adaptive entropy regularization or prioritized experience replay to maintain guaranteed iterative convergence while balancing exploration-exploitation trade-offs.

## 6 ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (Grant No. 2024YFB3310602); the National Natural Science Foundation of China (52375495, 52188102);

## 7 REFERENCES

[Chiang 1995] Chiang, S.-T., et al. Adaptive control optimization in end milling using neural networks. International Journal of Machine Tools and Manufacture, 1995, Vol.35, pp. 637–660. ISSN 08906955.

[Oh 2023] Oh, J. Y., et al. Model-based feed rate optimization for cycle time reduction in milling. Journal of Manufacturing Processes, 2023, Vol.94, pp. 289–296. ISSN 15266125.

[Vavruska 2023] Vavruska, P., et al. Effective feed rate control to maintain constant feed per tooth along toolpaths for milling complex–shaped parts. The International Journal of Advanced Manufacturing Technology, 2023, Vol.128, pp. 3215–3232. ISSN 02683768.

[García-Hernández 2021] García-Hernández, C., et al. Trochoidal Milling Path with Variable Feed. Application to the Machining of a Ti-6Al-4V Part. Mathematics, 2021, Vol.9, pp. 2701. ISSN 22277390.

[Sencer 2008] Sencer, B., et al. Feed optimization for five-axis CNC machine tools with drive constraints. International Journal of Machine Tools and Manufacture, 2008, Vol.48, pp. 733–745. ISSN 08906955.

[Erkorkmaz 2013] Erkorkmaz, K., et al. Feedrate optimization for freeform milling considering constraints from the feed drive system and process mechanics. CIRP Annals, 2013, Vol.62, pp. 395–398. ISSN 00078506.

[Chu 2025] Chu, A.-M., et al. Dynamic modelling for the family of 5-axis CNC milling machines with application to feed-rate optimization. Engineering Science and Technology, an International Journal, 2025, Vol.65, pp. 102015. ISSN 22150986.

[Zhao 2001] Zhao, M. X., and Balachandran, B. Dynamics and stability of milling process. International Journal of Solids and Structures, 2001, Vol.38, pp. 2233–2248. ISSN 00207683.

[Bachrathy 2011] Bachrathy, et al. State Dependent Regenerative Effect in Milling Processes. Journal of Computational and Nonlinear Dynamics, 2011, Vol.6, pp. 041002. ISSN 15551415.

[Wang 2022] Wang, G., et al. Trajectory Planning and Optimization for Robotic Machining Based On Measured Point Cloud. IEEE Transactions on Robotics, 2022, Vol.38, pp. 1621–1637. ISSN 15523098.

[Chen 2022] Chen, H., and Ahmadi, K. Estimating pose-dependent FRF in machining robots using multibody dynamics and Gaussian Process Regression. Robotics and Computer-Integrated Manufacturing, 2022, Vol.77, pp. 102354. ISSN 07365845.

[Tajima 2018] Tajima, S., et al. Accurate interpolation of machining tool-paths based on FIR filtering. Precision Engineering, 2018, Vol.52, pp. 332–344. ISSN 01416359.

[Xin 2022] Xin, S., et al. Research on the influence of robot structural mode on regenerative chatter in milling and analysis of stability boundary improvement domain. International Journal of Machine Tools and Manufacture, 2022, Vol.179, pp. 103918. ISSN 08906955.

[Ding 2011] Ding, Y., et al. Numerical Integration Method for Prediction of Milling Stability. Journal of Manufacturing Science and Engineering, 2011, Vol.133, pp. 031005. ISSN 10871357.

[Sumiea 2024] Sumiea, E. H., et al. Deep deterministic policy gradient algorithm: A systematic review. Heliyon, 2024, Vol.10, pp. e30697. ISSN 24058440.

[Rasmussen 2005] Rasmussen, C. E., and Williams, C. K. I. Gaussian processes for machine learning. The MIT Press 2005. ISBN 9780262256834.

*Tab. 1: Environmental configurations.*

| Control points of path L (m) | Parameters of MDH (m, m, rad) | State range $s_t$ (m/s, RPM, rad) | Action range $a_t$ (m/s, RPM, rad) | Depth of axial (m) |
|---|---|---|---|---|
| $p_1(1.4,-0.3,1.7), p_3(1.8,0.1,1.7)$<br>$p_2(1.6,-0.1,1.7), p_4(2.1,0.3,1.7)$ | $\begin{bmatrix} a \\ d \\ a \end{bmatrix} = \begin{bmatrix} 0.8145 & 0 & 0 & 1.5055 & 0 & 0.2 \\ 0 & 0.3 & 1.28 & 0.28 & 0 & 0 \\ 0 & 3p/2 & 0 & 3p/2 & 3p/2 & p/2 \end{bmatrix}$ | $\begin{bmatrix} f \\ \Omega \\ g \end{bmatrix} \in \begin{bmatrix} 0,6e^{-2} \\ 2e^{3},4e^{3} \\ -p,p \end{bmatrix}$ | $\begin{bmatrix} \Delta f \\ \Delta\Omega \\ \Delta g \end{bmatrix} \in \begin{bmatrix} -1e^{-3},1e^{-3} \\ -2e^{1},2e^{1} \\ -2e^{-1},2e^{-1} \end{bmatrix}$ | $3e^{-3}$ |
| Shear force coefficient (N/m^2) | Engagement range (rad) | Number $n$ of segments | Number $N_t$ of teeth | Tool radius $R$ (m) |
| $[6.44e^{8},2.37e^{8}]$ | $[0.75,1]p$ | 100 | 2 | 0.04 |

*Tab. 2: DDPG training configurations.*

| Total episodes $N_{ep}$ | Discount factor $\Gamma$ | Size of Buffer B | Actor learning rate | Critic learning rate |
|---|---|---|---|---|
| 1000 | 0.99 | 10000 | 0.0001 | 0.001 |
| Initial noise deviation | Noise decay rate | Noise decay episodes | Batch size | Soft update rate $a_u$ |
| 0.1 | 0.999 | 50 | 64 | 0.005 |