

THE CONTROL OF ACTIVE MAGNETIC BEARING USING TWO-PHASE Q-LEARNING

T. Březina*, J. Krejsa⁺

Summary: *Q-learning method proved to be usable in active magnetic bearing (AMB) control task, however the learning speed remains the main problem. Two-phase variant of the Q-learning speeds up the learning process by dividing the method into two stages – prelearning stage which uses mathematical model of AMB and is highly efficient and tutoring stage which uses conventional Q-learning and is used for further fining of the Q-function. Comparison with PID controller is included to demonstrate algorithms control abilities. As table (step function) is used as Q-function implementation, the optimum configuration of the table (number and relative positions of node points) was also investigated.*

1. Úvod

Aktivní magnetické ložisko (AML) eliminuje přímý mechanický dotyk levitací rotoru v řízeném magnetickém poli. AML je nestabilní a je nutné je stabilizovat zpětnovazební regulační smyčkou. Pro návrh metod zlepšujících řízení AML, případně snižujících nároky na řídicí elektroniku mohou sloužit metody umělé inteligence využívající strojového učení. Vzhledem k tomu že je obtížné získat množinu vzorů požadovaného chování regulátoru, což je nezbytná podmínka při použití učení s učitelem, nabízí se použití metody opakovaně posilovaného učení (RL-učení), která množinu vzorů nevyžaduje. Jednou z možných variant RL-učení je tzv. Q-učení. Proces učení konvenčních architektur Q-učení je pro reálné úlohy řízení příliš pomalý. Použitelnou metodou na zrychlení učení je rozdělení učení do dvou fází – fáze předučení a fáze doučování [2]. Fáze předučení vyžaduje výpočtový model, ale je velmi efektivní, fáze doučování používá konvenčního Q-učení a předpokládá interakci se soustavou.

V tomto článku je nejprve stručně uvedena podstata metody Q-učení a její aplikace na řízení aktivního magnetického ložiska. Je zhodnocen vliv různých konfigurací implementace Q-funkce a je vybrána optimální implementace. Pro určení kvality vyvinutých řídicích členů je provedeno porovnání chování modelu AML řízeného tímto členem s chováním modelu AML řízeného referenčním PID regulátorem.

2. Q-učení

Cílem standardního modelu Q-učení je získat strategii μ řízení stochastického dynamického systému s konečnou množinou stavů $S = \{1, \dots, n\}$ a konečnou množinou akcí

* RNDr. Ing. Tomáš Březina, CSc, Ústav automatizace a informatiky, Fakulta strojního inženýrství, VUT v Brně, Technická 2, 616 69, Brno, tel: 54114-2885, email: brezina@uai.fme.vutbr.cz

⁺ Ing. Jiří Krejsa, PhD., Ústav termomechaniky – pobočka Brno, Akademie věd ČR, Technická 2, 616 69, Brno, tel: 54114-2885, email: jkrejsa@umt.fme.vutbr.cz

U [5]. Strategie vybírá akce pouze na základě stavu systému, tj. $\mu: S \rightarrow U$. Pro stav $i \in S$ je provedena akce $\mu(i) \in U$. Optimální strategie je obsažena v optimální Q-funkci (odkud ji získává řídicí člen). Optimální funkce je odhadována pomocí pozorování přechodů stavů dynamického systému.

Časové závislosti jsou vyjádřeny pomocí indexů časových úseků $t = 0, 1, \dots$ ve kterých řídicí člen provádí akce; stav pozorovaný řídicím členem až do okamžiku t včetně je označen s_t . Řídicí člen má k dispozici odhad optimálních Q-hodnot z předešlých kroků Q-učení. Označme tento odhad $Q_t(i, u)$ pro všechny přípustné dvojice stav-akce (i, u) (kde $i \in S$ je stav systému a $u \in U(i)$ je přípustná akce v tomto stavu. Řídicí člen vybírá přípustnou akci $u_t \in U(s_t)$ která odpovídá dosažené strategii μ podle vztahu:

$$\mu(s_t) = \arg \max_{u \in U(s_t)} Q(s_t, u), \quad (1)$$

přičemž je prováděno také prozkoumávání (to znamená, že je použit i jiný mechanismus výběru akce než podle (1), čímž je zaručeno že řídicí člen může vybrat i akce které podle aktuálního stavu Q-funkce nejsou optimální).

Provedení akce u_t uvede systém do následného stavu s_{t+1} a během toho přechodu stavu je získáno okamžité posílení $r_{s_t}(u_t)$. Na základě tohoto posílení řídicí člen provede aktualizaci příslušné Q-hodnoty podle vztahu:

$$Q_{t+1}(i, u) = \begin{cases} (1 - \alpha_t(s_t, u_t))Q_t(s_t, u_t) + \\ \alpha_t(s_t, u_t)[r_{s_t}(u_t) + \gamma f_t(s_{t+1})], & \text{pro } (i, u) = (s_t, u_t), \\ Q_t(i, u), & \text{jinak,} \end{cases} \quad (2)$$

kde $f_t(s_{t+1}) = \max_{u \in U(s_{t+1})} Q_t(s_{t+1}, u)$, $\alpha_t(s_t, u_t)$ je parametr učení v časovém kroku t (tento parametr je v čase proměnný) a γ , $0 < \gamma < 1$ je srážkový faktor. Tento proces je opakován pro každý časový krok t .

Aktualizace Q-hodnot přípustných párů stav-akce v kroku k , $k = 0, 1, \dots$, je prováděna synchronně a Q-hodnoty ostatních přípustných párů se nemění. Aktualizace Q-hodnot je prováděna pro každý přípustný pár stav-akce nekonečně mnohokrát a pokud vhodně klesá parametr učení, potom posloupnost $\{Q_k(i, u)\}$ generovaná off-line učením konverguje pro $k \rightarrow \infty$ s pravděpodobností 1 k optimálním hodnotám Q-funkce $Q^*(i, u)$ pro všechny přípustné páry (i, u) . Jeden z možných předpisů pro změnu parametru učení zaručujících konvergenci uvádí např. Darken [3]:

$$\alpha_k(i, u) = \frac{\alpha_0 n_0}{n_0 + n_k(i, u)}, \quad (3)$$

kde $n_k(i, u)$ je počet aktualizací provedených na Q-hodnotě (i, u) ve všech krocích až do kroku k , α_0 je počáteční krok učení a parametr n_0 určuje rychlost snižování $\alpha_k(i, u)$.

3. Aplikace

Jako modelu prostředí je použito jednoduchého jednohmotového modelu AML (pohyb v rovině kolmé na osu rotace) jehož autorem je Půst [4]. Úkolem řídicího členu je to, aby svým generováním akcí udržel výchylku hmotného bodu v intervalu $\langle -x_r, x_r \rangle$. Pro popis prostředí byl použit třírozměrný stavový prostor s proměnnými $\mathbf{y} = (x, \dot{x}, \ddot{x})$ (výchylka, rychlost a zrychlení v rovině kolmé na osu rotace), dále označený jako 3D, a dvourozměrný (2D) s proměnnými $\mathbf{y} = (x, \dot{x})$. Jako akce bylo zvoleno napájecí napětí u . Q-funkce byla implementována tabulkou o různých rozměrech a uspořádání mřížky (počtu a poloze uzlových bodů rastru). Jelikož konfigurace tabulky ovlivňuje zásadním způsobem chování řídicího členu, byla provedena řada numerických experimentů s cílem získat optimální konfiguraci pro obě základní uspořádání (2D a 3D popis stavu AMB).

Konfigurace tabulky je dále popsána jako $(|g_1|/r_1, \dots, |g_{\dim Y}|/r_{\dim Y}; |g_{\dim Y+1}|/r_{\dim Y+1})$, kde $\dim Y$ je počet proměnných stavu (2 pro 2D stavů 3 pro 3D stav), $|g_i|$ je počet uzlů rastru, r_i je koeficient nelinearity rastru, který představuje relativní vzdálenost uzlu nejbližšího ke středovému uzlu vzhledem k případu lineárního rastru. V případě lineárního rastru není pro zjednodušení tento koeficient uváděn. Když je například stav řízené soustavy určen proměnnými x a \dot{x} , je použita jediná proměnná pro akci a mřížka tabulky má nelineární rastr s koeficientem nelinearity $r = 0.1$ pouze pro polohu proměnné stavu, potom mřížka tabulky s 11 uzlovými body pro proměnnou x , sedmi uzlovými body pro proměnnou \dot{x} a třemi možnými hodnotami řídicí proměnné je označena jako $(11/0.1, 7/0.1; 3)$.

Pro zrychlení konvergence Q-učení byl proces učení rozdělen do dvou fází – fáze předučení a fáze doučování [2]. Fáze předučení využívá výpočtového modelu AML a v každém kroku učení jsou použity všechny kombinace stavů a akcí právě jednou s využitím pouze jednoho přechodu jednotlivých stavů. Kombinace stavů jsou předkládány při učení sekvenčně postupným průchodem všech uzlů rastru tabulky podle jednotlivých rozměrů. Po ukončení fáze předučení bylo provedeno doučování, kdy je již předpokládána interakce s reálnou soustavou a je využíván konvenční model Q-učení. Vzhledem k tomu že reálná soustava nebyla v době provádění experimentů k dispozici je i fáze doučování provedena s využitím modelu soustavy.

Jako kritéria pro posouzení kvality řídicího členu bylo použito kvadratického kritéria kvality řízení, robustnost řízení byla posuzována vzhledem k odolnosti řídicího členu na zpoždění akčního zásahu a na chybu pozorování veličin soustavy. Strategie dosažené na optimálních mřížkách byly porovnány s referenčním PID regulátorem na datech s náhodnými počátečními hodnotami stavových veličin a náhodným průběhem zátěžné síly, která byla použita i pro ostatní numerické experimenty.

4. Numerické experimenty

4.1 Simulační detaily

Posilovací funkce

V prostoru stavů soustavy byla definována množina cílových stavů $C_0 \subset Y$, ve kterých je hřídel pokládána za stabilizovanou

$$C_0 = \{ \mathbf{y}; |x| \leq x_{tol}, |\dot{x}| \leq \dot{x}_{tol}, |I| \leq I_{max} \},$$

kde $x_{tol} = 50 \times 10^{-6}$ [m], $\dot{x}_{tol} = 20 \times 10^{-3}$ [ms⁻¹], $I_{max} = 2$ [A] jsou po řadě maximální povolená velikost výchylky, axiální rychlosti hřídele a proudu protékajícího elektromagnetem ložiska. Bylo použito posílení řídicího členu r ve tvaru prosté omezené pokuty

$$r(\mathbf{y}) = \begin{cases} 0, & \text{pro } \mathbf{y} \in C_0, \\ -1, & \text{jinak.} \end{cases}$$

Mřížka tabulky Q-funkce

Rozsahy rastrů jednotlivých veličin stavů (viz. tab. 1) byly stanoveny z technologických omezení hodnot veličin soustavy, pro rychlost \dot{x} a zrychlení \ddot{x} z kontrolních simulačních výpočtů. Pro akci (napájecí napětí) byly použity tři hodnoty: $(-100, 0, 100)$ [V].

Výchylka hřídele	$\pm x_{max}$	$\pm 500 \times 10^{-6}$	[m]
Rychlost hřídele	$\pm \dot{x}_{max}$	$\pm 100 \times 10^{-3}$	[ms ⁻¹]
Zrychlení hřídele	$\pm \ddot{x}_{max}$	± 450	[ms ⁻²]
Budící napětí	$\pm u_{max}$	± 100	[V]

Tab. 1. Rozsahy rastrů veličin

Při použití lineární mřížky byl počet uzlových bodů určen především na základě předchozích zkušeností. Konkrétní konfigurace mřížek jsou uvedeny v tabulce 2. Pro určení vlivu nelineární mřížky byly vybrány nejlepší lineární mřížky pro obě základní konfigurace (2D a 3D) a nelineární rastr byl aplikován pouze na veličinu x (cílem použití nelineární mřížky je zlepšení kritéria kvality řízení, čehož může být dosaženo především zjemněním rastru v okolí cílových stavů). Koefficient nelinearity byl nastaven na $r = 0.5$ a $r = 0.1$. S ohledem na změnu rastru mřížky byla modifikována také posilovací funkce změnou cílových stavů C_0 přepočítáním hranic těchto cílových stavů x_{tol} a \dot{x}_{tol} (viz. Tab. 3).

2D stav	(7,5;3)	(9,5;3)	(9,9;3)	(11,9;3)	(13,11;3)
3D stav	(9,7,5;3)	(11,7,3;3)	(11,9,5;3)	(13,9,5;3)	(13,11,3;3)

Tab. 2. Konfigurace lineárních mřížek

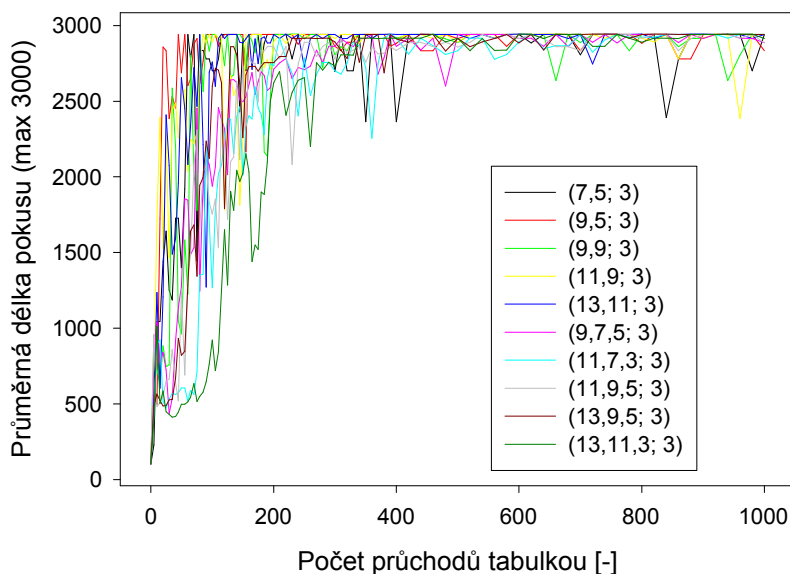
	Původní hodnota	Přepočtené hodnoty, 3D stav				Přepočtené hodnoty, 2D stav		
r	1	1	0,5	0,1	1	0,5	0,1	
x_{tol} [m]	50×10^{-6}	83×10^{-6}	41×10^{-6}	8×10^{-6}	100×10^{-6}	50×10^{-6}	10×10^{-6}	
\dot{x}_{tol} [ms ⁻¹]	20×10^{-3}	25×10^{-3}	25×10^{-3}	25×10^{-3}	25×10^{-3}	25×10^{-3}	25×10^{-3}	

Tab.3. Přepočet hranic cílových stavů pro nelineární mřížku

4.2 Předučení – lineární mřížka

Během předučení byl maximální počet průchodů tabulkou omezen na 1000. Parametr učení $\alpha_t(i, u)$ byl snižován podle vztahu (3), s parametry $\alpha_0 = 0.2$ a $n_0 = 300$. Počáteční stav AML byl volen náhodně s omezením, že během 10×10^{-6} [s] nedošlo k přechodu AML do stavu $|x| = x_{max}$ s použitím libovolné akce. Zátěžná síla $F_e(t)$ byla modelována jako schodovitá funkce s velikostí konstantních úseků branou náhodně z intervalu $\langle -200, 200 \rangle$ [N]. Délka těchto úseků odpovídala během předučení délce přechodu stavu, zatímco během testování byla brána náhodně z intervalu $\langle 0, 1 \times 10^{-3} \rangle$ [s]. Pro zjištění průběhu učení byl systém v každém kroku testován na 100 pokusech, lišících se v počátečních podmínkách a průběhu zátěžné síly. Pro jednotlivé pokusy byl zaznamenán počet kroků řízení ve kterých byla výchylka rotoru v požadovaných mezích. Jako kritérium úspěšnosti učení byla potom použita průměrná délka pokusu (suma kroků řízení pro všechny pokusy / počet pokusů).

Průběh předučení se pro různé konfigurace mřížek příliš neliší, viz obr. 1. Je zřejmé že 3D stav obecně vyžaduje k dosažení použitelné strategie více průchodů tabulkou, ale rozdíly jsou nepodstatné. K dosažení použitelné strategie dojde u většiny konfigurací mřížek Q-funkce již po 500 průchodech tabulkou.

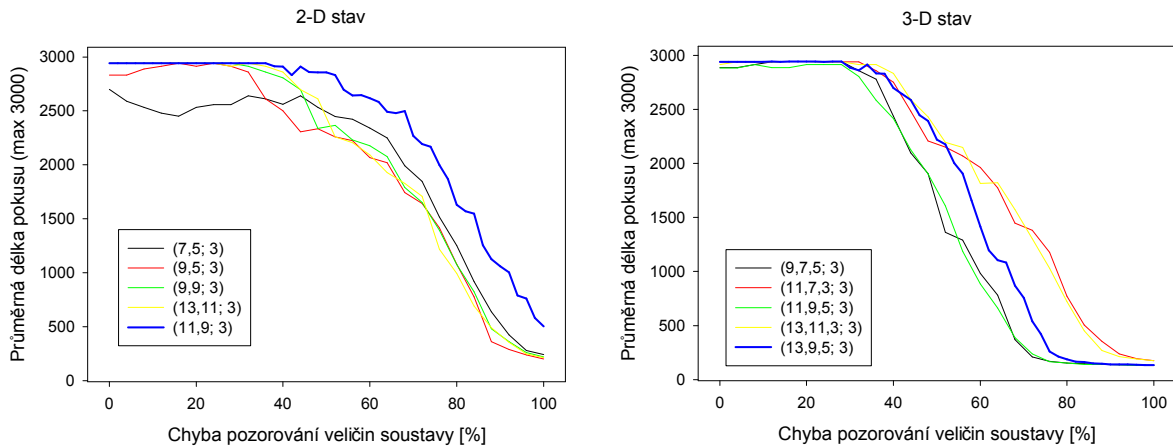


Obr. 1. Průběh předučení – lineární mřížka

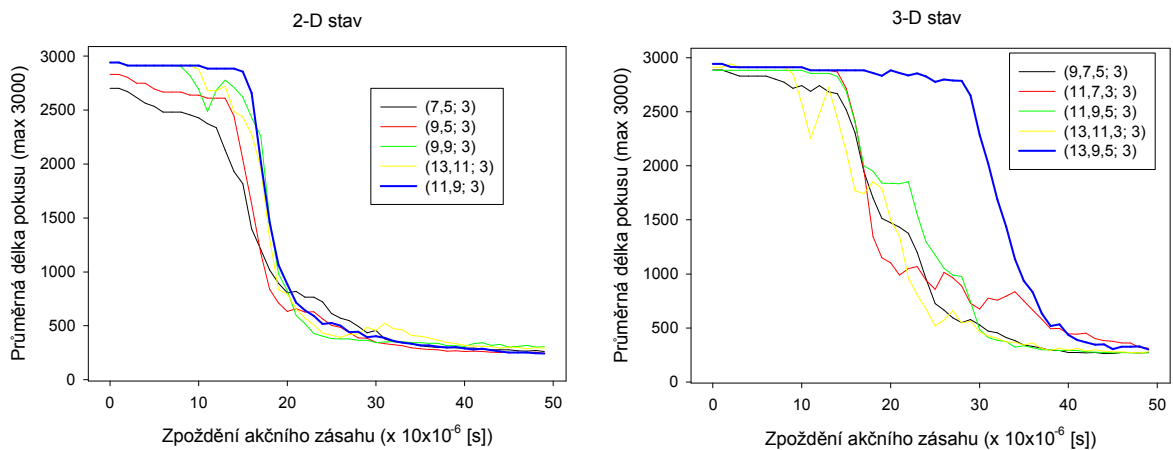
Vliv konfigurací mřížky (především mezi 2D a 3D stavy) je patrnější při sledování robustnosti naučených strategií vůči chybám pozorování veličin soustavy a vůči zpoždění akčního zásahu. Výsledky testů jsou shrnuty na obr. 2 a 3.

Chyby pozorování veličin soustavy byly zavedeny do všech veličin stavu soustavy. S výjimkou mřížek 2D stavů, které definují méně nebo 32 buněk stavů soustavy, mají závislosti podobný charakter: do přibližně 35 % úrovně chyb pozorování veličin soustavy je zachována průměrná úspěšnost pokusů dosažená při simulacích prováděných bez chyb pozorování. S dalším zvyšováním úrovně chyb pozorování veličin soustavy se průměrná délka pokusů

snižuje. Z mřížek 2D stavu vykazuje nejméně strmý sestup mřížka (11, 9; 3) se zlomovou hodnotou úrovně chyb pozorování asi 40 %. Pro 3D stav vykazují prakticky stejný sestup mřížky (11, 7, 3; 3) a (13, 11, 3; 3), které mají zlomovou hladinu úrovně chyb pozorování asi na 35 %, zatímco mřížka (13, 9, 5; 3) vykazuje průměrně strmý sestup.



Obr. 2. Odolnost vůči chybám pozorování veličin soustavy – lineární mřížka



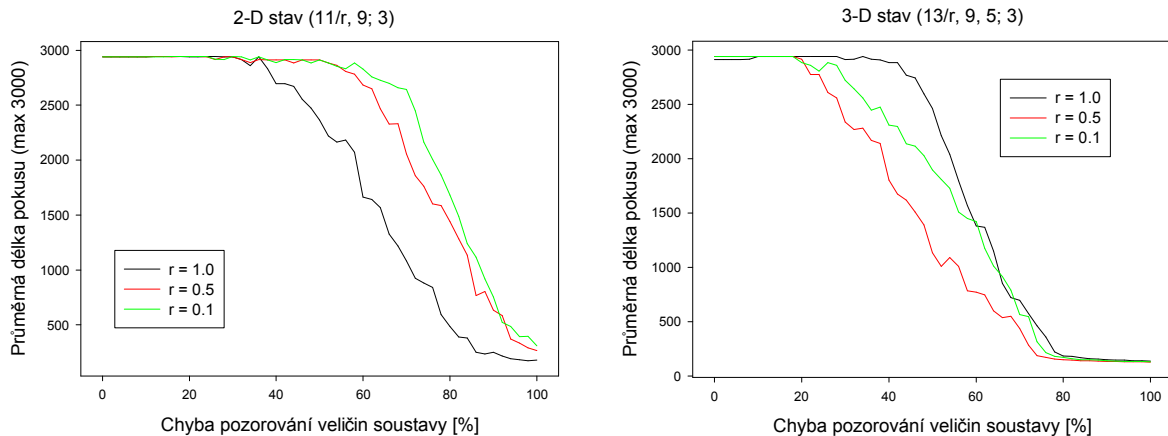
Obr. 3. Odolnost vůči zpoždění akčního zásahu – lineární mřížka

Co se týká zpoždění akčního zásahu, je zřejmé že nad hodnotou zpoždění přibližně 100×10^{-6} [s] začíná u 2D stavu prudce klesat průměrná délka pokusů. S výjimkou mřížek, které definují 32 buněk stavů soustavy nebo méně, mají závislosti opět podobný charakter. Strmost sestupu není volbou mřížek příliš ovlivněna. Nejlépe se chová mřížka (11, 9; 3), u které se zlomová hodnota zpoždění posouvá o něco výše, asi na 130×10^{-6} [s]. U 3D stavu se zlomová hodnota zpoždění vyskytuje také na 130×10^{-6} [s], přičemž průběh poklesu je o něco méně strmý a více závisí na volbě mřížky. Vyniká mřížka (13, 9, 5; 3), u které se posouvá zlomová hodnota zpoždění až na 280×10^{-6} [s].

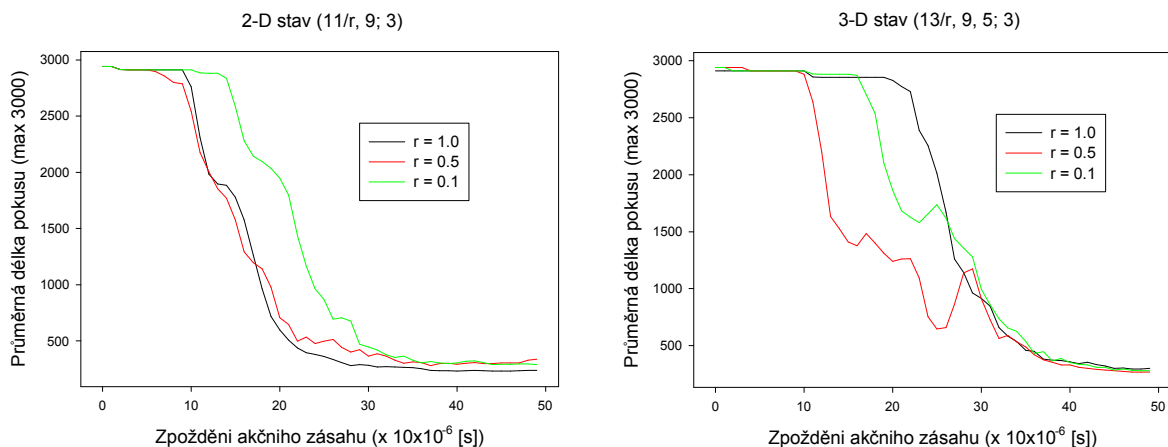
4.3 Předučení – nelineární mřížka

Na základě výsledků pro lineární mřížku byla jako reprezentant pro 3D stav vybrána mřížka (13, 9, 5; 3) a pro 2D stav pak mřížka (11, 9; 3). Tyto vybrané mřížky byly modifikovány použitím nelineárního rastru pro proměnnou x , výsledné nelineární mřížky pak mají tvar: (11/0.5, 9; 3), (11/0.1, 9; 3) pro 2D stav a (13/0.5, 9, 5; 3), (13/0.1, 9, 5; 3) pro 3D stav. Průběh učení pro nelineární mřížky se příliš nemění a odpovídá průběhu učení s mřížkami lineárními; použitelné strategie je dosaženo po přibližně 400 krocích.

Odolnost dosažené strategie vůči chybám pozorování veličin soustavy se při 2D stavu pro nelineární variantu vybrané mřížky s prohlubováním nelinearity mřížky zvyšuje (viz. obr. 4). Zlomová úroveň chyb pozorování veličin soustavy se pro $r = 0.5$ i $r = 0.1$ posouvá z původních 40 % až na téměř 61 %, přičemž pro $r = 0.5$ je sestup poněkud méně strmý. Diametrálně odlišná je situace u vybrané mřížky 3D stavu, kdy se odolnost dosažené strategie vůči chybám pozorování zhorší. Největší zhoršení nastane při koeficientu nelinearity $r = 0.5$, kdy se zlomová hodnota úrovně chyb pozorování, která u lineární mřížky činí 35 %, sníží až na 20 %. Pro $r = 0.1$ je sice zlomová hodnota úrovně chyb pozorování také přibližně 20 %, avšak sestup je méně strmý, než v případě $r = 0.5$.



Obr. 4. Odolnost vůči chybám pozorování veličin soustavy – nelineární mřížka



Obr. 5. Odolnost vůči zpoždění akčního zásahu – nelineární mřížka

Výsledky pro odolnost vůči zpoždění akčního zásahu jsou uvedeny na obr. 5. U mřížky 2D stavu dochází ke zlepšení, zlomová hodnota se při koeficientu nelinearity $r = 0.1$ posouvá z původních 130×10^{-6} [s] na 170×10^{-6} [s] a také se snižuje strmost poklesu. Pro $r = 0.5$ se zlomová hodnota zhoršuje, stejně jako pro lineární mřížku s přepočtenou množinou cílových stavů. U mřížky 3D stavu dochází ke zhoršení, přičemž největší zhoršení nastane, stejně jako u odolnosti strategie vůči náhodným chybám pozorování veličin soustavy, při koeficientu nelinearity $r = 0.5$, kdy se zlomová hodnota posouvá z 280×10^{-6} [s] na 110×10^{-6} [s].

4.4 Srovnání s PID regulátorem

Pro porovnání jednotlivých řídicích členů mezi sebou a také s referenčním PID regulátorem bylo použito dvou různých sad testovacích dat, které se liší celkovou délkou doby řízení (30×10^{-3} [s] pro data A a 1[s] pro data B). Obě sady dat obsahují 100 pokusů, lišících se počátečními podmínkami a průběhem zátěžné síly, která byla modelována schodovitou funkcí s náhodnou velikostí konstantních částí z intervalu $\langle -200, 200 \rangle$ [N]. Jako kritéria bylo použito běžné kvadratické kritérium kvality řízení a dále procento úspěšných pokusů (za úspěšný pokus je považován takový, kdy po celou dobu řízení nedojde k vychýlení rotoru na výchylku větší než je přípustná).

Řídicí člen	Úspěšné pokusy [%]		Kritérium kvality řízení [m^2]	
	Data A	Data B	Data A	Data B
PID regulátor	94	91	28×10^{-9}	23×10^{-9}
2-D stav (11.1,9;3)	97	94	41×10^{-9}	55×10^{-9}
2-D stav (11/0.5,9;3)	98	94	16×10^{-9}	17×10^{-9}
2-D stav (11/0.1,9;3)	98	94	10×10^{-9}	16×10^{-9}
3-D stav (13,9,5;3)	97	94	21×10^{-9}	29×10^{-9}
3-D stav (13/0.5,9,5;3)	97	94	14×10^{-9}	15×10^{-9}
3-D stav (13/0.1,9,5;3)	98	94	9×10^{-9}	15×10^{-9}

Tab. 4. Srovnání řídicích členů – fáze předučení

Z výsledků je zřejmé že řídicí členy založené na Q-učení překonaly PID regulátor z hlediska procenta úspěšných pokusů. Výsledky jednotlivých konfigurací mřížek jsou z tohoto hlediska prakticky totožné. Odlišná je situace z hlediska kritéria kvality řízení. Zde dosahuje PID regulátor lepších výsledků než nejlepší řídicí členy s lineární mřížkou. Při použití nelineární mřížky se jak pro 2D tak pro 3D stav kritérium kvality řízení se vzrůstající nelinearitou mřížky výrazně zlepšuje, přičemž již pro hodnotu $r = 0.5$ výsledky překonávají hodnoty dosažené PID regulátorem.

Při hodnocení dosažených strategií je dáována přednost robustnosti před kvadratickým kritériem kvality řízení. Jako nejlepší mřížka pro 2D stav byla proto vyhodnocena nelineární mřížka (11/0.1, 9; 3); jako optimální mřížka pro 3D stav potom lineární mřížka (13, 9, 5; 3).

4.5 Doučování

Strategie dosažené s implementací Q-funkce pomocí tabulky s konfigurací která byla vyhodnocena jako optimální byly dále použity pro fázi doučování. Výsledky testů jsou uvedeny v tabulce 5.

	Před doučováním		Po doučování		Změna [%]
	STP [%]	$q(T)$ [m ²]	STP [%]	$q(T)$ [m ²]	
2-D stav (11/0.1, 9; 3)	98	$10,0 \times 10^{-9}$	98	$10,0 \times 10^{-9}$	0
3-D stav (13, 9, 5; 3)	97	$20,9 \times 10^{-9}$	97	$20,3 \times 10^{-9}$	3

Tab. 5. Doučování řídicích členů používajících vybraných mřížek

Jak je zřejmé z výsledků, doučování má na vlastnosti řídicích členů prakticky zanedbatelný vliv. Určitého zlepšení dosahuje pouze lineární mřížka pro 3D stav, u nelineární mřížky 2D stavu je změna nepodstatná. Zlepšení je ovšem pro vybrané (ideální) mřížky nevýrazné. Pokud použijeme doučování na ostatní mřížky, může doučování jejich „nedokonalost“ napravit, například při pokusech s lineární mřížkou 2D stavu bylo při doučování dosaženo výrazného zlepšení 37%.

Slabý vliv doučování na výsledky řídicích členů je zřejmě způsoben použitím téhož modelu při doučování, jakého bylo použito ve fázi předučení. Bohužel v současné době není dosud k dispozici reálná soustava, na které by bylo možno ověřit vliv doučování na řídicí člen.

5. Závěr

Použití dvoufázové varianty Q-učení se vyznačuje především vysokou rychlostí konvergence ve fázi předučení. Již po 1000 průchodech tabulkou Q-funkce ve této fázi bylo dosaženo použitelné strategie řízení. Při použití lineární mřížky u implementace Q-funkce tabulkou dosahují strategie získané pomocí Q-učení lepších výsledků než referenční PID regulátor z hlediska procenta úspěšných pokusů, z hlediska kritéria kvality řízení má strategie využívající 3D stavu výsledky srovnatelné, pro 2D stav jsou výsledky horší, ovšem dojde k jejich zlepšení ve fázi doučování.

Při určování optimální konfigurace mřížky tabulky implementující Q-funkci byl prokázán pozitivní vliv nelineárního rastru mřížky na „jemnost“ řízení (zlepšení kvadratického kritéria kvality řízení), robustnost řídicího členu se nelineární mřížkou zlepšila pro 2D stav a mírně zhoršila pro 3D stav. Protože robustnost řídicího členu je brána jako primární kritérium, byla jako optimální konfigurace mřížky vybrána nelineární mřížka pro 2D stav a ponechána původní lineární mřížka pro 3D stav.

Použití nelineární mřížky u 2D stavu zlepšuje kvalitu řídicího členu natolik, že jeho dalším doučováním se jeho vlastnosti již nemění. U lineární mřížky 3D stavu dojde při doučování k mírnému zlepšení kritéria kvality řízení při zachování robustnosti.

6. Poděkování

Tato práce je podporována projektem GAČR č. 101/00/1471 a výzkumnými záměry MSM 262100024 a CEZ: J22/98: 261100009.

Literatura

- [1] Barto A.G, Bradtke S.J., Singh S.P.: Learning to act real time dynamic programming. *Artificial Intelligence* 72 (1995), pp.81-138
- [2] Březina T., Krejsa J.: Efficient Q-learning Modification Applied on Active Magnetic Bearing Control, *Engineering Mechanics*, in print
- [3] Darken C., and Moody J.: Note on learning rate schedule for stochastic optimization, in: *Advances in Neural information Processing Systems* 3, Morgan Kaufmann, San Mateo, CA, 1991, pp.832-838
- [4] Půst L.: Dynamic stability of magnetic bearing. *Engineering Mechanics'97*, Svratka 1997, pp.57-62
- [5] Watkins C.J.C.H. and Dayan, P.: Q-learning, *Mach. Learn.* 8 (1992), pp.279-292