Vysoké učení technické v Brně

Fakulta strojního inženýrství

Ústav Mechaniky těles, Mechatroniky a Biomechaniky

Ing. Tomáš Marada

ŘÍZENÍ PŘECHODOVÝCH STAVŮ ASYNCHRONNÍHO ELEKTROMOTORU ZALOŽENÉ NA Q-UČENÍ

CONTROL OF TRANSITION STATES OF ASYNCHRONOUS ELECTROMOTOR BASED ON Q-LEARNING

Zkrácená verze Ph.D. Thesis

Obor:	Inženýrská mechanika
Školitel:	Doc. RNDr. Ing. Tomáš Březina, CSc.
Oponenti:	Prof. Ing. Otakar Kurka, CSc.
	Doc. Ing. Čestmír Ondrůšek, CSc.
Datum obhajoby:	20.12.2004

Klíčová slova
Q-učení, asynchronní elektromotor, řízení

Key Words
Q-learning, asynchronous electromotor, control

Místo uložení disertační práce
Ústav Mechaniky těles, Mechatroniky a Biomechaniky, FSI VUT v Brně

ISBN 80-214-2928-3
ISSN 1213-4198

OBSAH

1 ÚVOD

Trojfázové asynchronní motory jsou v současné době jedny z nejrozšířenějších motorů používaných v průmyslu.

Výkonovou část elektrického pohonu, tj. elektromotor a výkonový měnič, spolu s poháněným zařízením, lze modelovat jako dynamickou soustavu vyššího řádu, zpravidla nelineární eventuelně s proměnnými parametry. Tuto soustavu je nutno řídit. Nedílnou součástí elektrického pohonu je tedy jeho řídicí systém, na jehož vstupy jsou přiváděny jednak žádané hodnoty a jednak skutečné hodnoty veličin ze zpětnovazebních snímačů. Řídicí podsystém je často realizován mikropočítačem s příslušnými přizpůsobovacími obvody, realizovanými A/D a D/A převodníky, obvody typu programovatelných logických polí a hardwarovými modulátory pro pulsní šířkovou modulaci (PWM).

Výkonné mikropočítače umožňují aplikace moderních metod řízení elektrických pohonů pomocí metod umělé inteligence (metody UI). Zejména metody strojového učení v reálném čase mohou představovat východiska návrhu nových metod řízení („inteligentních řídicích členů“), zlepšujících řízení asynchronního elektromotoru, případně vyžadujících méně složitou řídicí elektroniku. Učení se stane dokonce zásadním faktorem, jestliže se parametry elektromotoru nebo cíle jeho řízení mění v čase.

1.1 FORMULACE PROBLÉMU A CÍLE DISERTAČNÍ PRÁCE

Cílem předložené práce je navrhnout robustní regulátor pro řízení asynchronního elektromotoru s využitím metody Q-učení, spadající do oblasti metod opakovaně posilovaného učení.

Úplný matematický model, který by věrně popisoval regulovanou soustavu je možno vytvořit pouze pro relativně jednoduché, nebo dostatečně prozkoumané případy. Na základě úplného matematického modelu je možno navrhnout optimální regulátor. V případě, že máme pouze zjednodušený matematický popis modelu, můžeme navrhnout robustní regulátor, který je schopen řešit i takové stavy regulované soustavy na které nebyl primárně naučen, avšak mnohdy na úkor kvality regulace.

Řízení se skládá z fáze předučení a fáze doučení. Během fáze předučení jsou na výpočtovém modelu prováděny pokusy, které jsou zpracovávány prováděním zálohování Q-učení v reálném čase. Výpočtový model může být pouze přibližný.

V předložené práci je provedeno simulační ověření navržené metody na modelu asynchronního elektromotoru. Při řízení jsou měřeny pouze aktuální otáčky, z nichž je vypočtena aktuální regulační odchylka, její rychlost a zrychlení.

Počáteční testy mají za úkol zjistit, jeví-li se jako vhodnější z hlediska úspěšnosti učení stav prostředí definovaný jako 1-D (uvažující pouze regulační odchylku), 2-D (uvažující regulační odchylku a její rychlost) nebo jako 3-D (uvažující regulační odchylku, její rychlost a zrychlení).

Další experimenty se týkají optimalizace počtu intervalů lineárního a nelineárního rastru jednotlivých stavových veličin. Získané strategie řízení jsou posuzovány nejprve z hlediska dosažené hodnoty integrálního kriteria kvality regulace, z hlediska odolnosti dosažených strategií vůči chybám pozorování soustavy, odolnosti vůči zpoždění akčního zásahu a odezvy na skokový moment.

Experimenty prováděné v další etapě jsou provedeny pro dokreslení vlastností strategií získaných ve fázi předučení. Je zde testován vliv různých posilovacích funkcí a vliv rozšíření množiny akcí jednak na rychlost předučení, jednak na odolnost strategie vůči náhodné chybě pozorování veličin, odolnost strategie vůči zpoždění akčního zásahu a odolnost strategie vůči skokovému momentu. V této etapě experimentů je také provedeno porovnání QL-regulátoru s referenčním PID regulátorem, jehož parametry byly nastaveny pomocí Ziegler-Nicholsova pravidla.

Po ukončení experimentů s fází předučení jsou provedeny experimenty s fází doučení. Experimenty jsou zaměřeny na vyzkoušení zpřesňování a přizpůsobování již dosažené strategie získané ve fázi předučení s matematickým modelem asynchronního motoru, na změny parametrů reálné soustavy oproti simulačnímu modelu AM.

Aktuálnost problematiky je vysoká. Vyhovuje současnému trendu výzkumu nových metod řízení, které jsou založeny na využití metod UI, zejména učení. Základním rysem učení je rozvinutá schopnost adaptace, tj. schopnost automaticky zlepšovat chování řízené soustavy např. při změně provozních parametrů, apod.

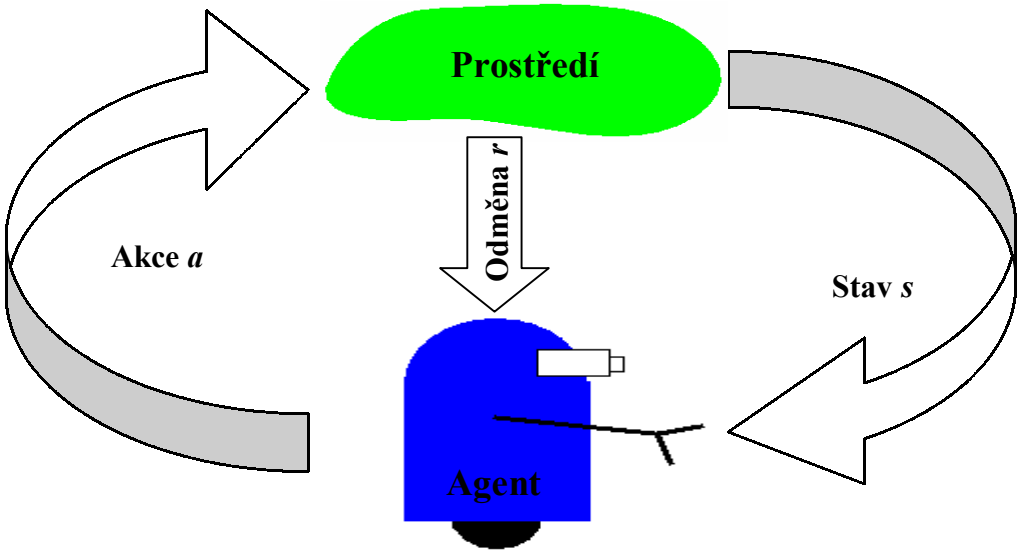
Dílčí cíle lze formulovat v následujících bodech:

- Analyzovat současný stav oboru v oblasti použití Q-učení a řízení AM.
- Vytvořit model pro simulaci dynamických charakteristik asynchronního stroje, konkrétně otáček rotoru.
- Navrhnout QL-regulátor pro řízení asynchronního stroje.
- Implementace QL-regulátoru v prostředí programovacího jazyka Borland Delphi.
- Nalezení optimálních parametrů nastavení QL-regulátoru.
- Provedení simulačních experimentů s QL-regulátorem.

2 OPAKOVANĚ POSILOVANÉ UČENÍ

Metody opakovaně posilovaného učení pracují na principu naučení agentova chování v dynamickém prostředí metodou „pokus-omyl“. Agent vnímá prostředí, na které působí akcemi. O míře okamžité úspěšnosti jednotlivých akcí je informován posílením generovaným vhodnou posilovací funkcí. Tento způsob učení má tu výhodu, že není nutné mít k dispozici fakta, která je potřeba se naučit. Získání potřebných faktů nemusí být ani jednoduché, ani přímočaré.

V modelu opakovaně posilovaného učení je agent v interakci s prostředím tak, jak je zobrazeno na obr. 2-1. V každém kroku vzájemného působení agent přijímá údaje o nynějším stavu s v prostředí. Pro vytvoření výstupu si agent vybere akci a, která představuje výstupní informaci. Tato akce změní aktuální stav prostředí a okamžitý prospěch z tohoto stavu je agentovi sdělen formou odměny r. Agent volí akce tak, aby maximalizovaly sumu celkové odměny. Soustavným opakováním těchto „pokusů-omylů“ může dojít k naučení velkého množství různých strategií.



Obr. 2-1: Základní model opakovaně posilovaného učení.

Formálně se model opakovaně posilovaného učení skládá z:

- množiny diskrétních stavů prostředí s,
- množiny diskrétních akcí agenta a,
- množiny skalárních ohodnocení r.

3 Q-UČENÍ

Metodu Q-učení navrhl Watkins [45], [46] pro řešení Markovových rozhodovacích problémů s neúplnou informací. Tato metoda přímo odhaduje optimální Q-hodnoty dvojic stavů a přípustných akcí, které nazýváme přípustné dvojice stav-akce. Tento algoritmus je velmi jednoduché implementovat. Q-učení je pravděpodobně nejpopulárnější a nejefektivnější RL procedura opakovaně posilovaného učení bez modelu.

Cílem standardního modelu Q-učení je získat strategii μ pro řízení stochastického dynamického systému s konečnou množinou stavů $S=\{1,...,n\}$ a s konečnou množinou akcí U. Strategie vybírá akce pouze na základě stavu systému, to je $\mu: S \to U$. 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 $t = 0,1,...,$ 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:

$$\mu(s_t) = \arg\max_{u \in U(s_t)} Q(s_t, u), \tag{3-1}$$

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

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

$$Q_{t+1}(i,u) = \begin{cases} (1-\alpha_t(s_t,u_t))Q_t(s_t,u_t) + \\ \quad \alpha_t(s_t,u_t)\left[r_{s_t}(u_t) + \gamma f_t(s_{t+1})\right] & \text{pro } (i,u) = (s_t,u_t), \\ Q_t(i,u) & \text{jinak} \end{cases} \tag{3-2}$$

kde $f_t(s_{t+1}) = \max_{u \in U(s_{t+1})} Q_t(s_{t+1},u)$, a $\alpha_t(s_t,u_t)$ 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, $k = 0,1,...,$ je prováděna synchronně a Q-hodnoty ostatních přípustných párů se nemění.

Posloupnost Q-hodnot generovaná Q-učením konverguje k optimálním hodnotám Q^* za následujících podmínek:

- Je-li každá přípustná akce provedena v každém stavu nekonečně mnohokrát v nekonečném počtu kroků řízení a
- snižuje-li se během kroků řízení vhodným způsobem učící poměr $\alpha_t(s_t,u_t)$.

Pokud Q-hodnoty konvergují k optimální hodnotě, potom je optimální strategie dosaženo tím, že agent v každém stavu vykonává pro každý stav vždy akci s nejvyšší Q-hodnotou. Q-učení je necitlivé na nastavení, to znamená, že Q-hodnoty budou vždy konvergovat k optimální hodnotě, budou-li všechny dvojice stav-akce vyzkoušeny v dostatečném počtu a pokud bude učící poměr α vhodně klesat. Bohužel tato konvergence může být velmi pomalá. Q-učení nepracuje příliš dobře na rozsáhlém stavovém prostoru, nebo na stavovém prostoru se skrytými stavy.

4 NÁVRH ORGANIZACE Q-UČENÍ

Pro řízení asynchronního elektromotoru byla kvůli vysoké populárnosti a z důvodu dobrých předběžných výsledků zvolena metoda Q-učení. Učení probíhá metodou učení pokusem, to znamená, že k učení jsou použity všechny stavy prostředí, které nastanou v průběhu jednotlivých pokusů postupně. Vlastní iterační proces Q-učení je rozdělen do dvou fází:

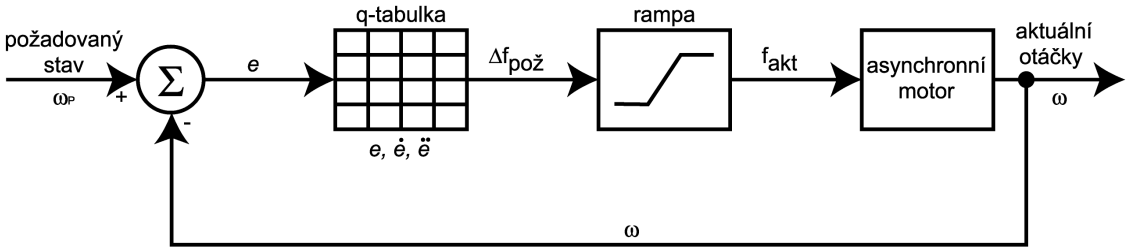
- fáze předučení a
- fáze doučování.

Fáze předučení probíhá s matematickým modelem asynchronního elektromotoru. V této fázi je úkolem dosáhnout alespoň použitelné strategie řízení. Fáze předučení představuje provádění pokusů, které začínají se zvolenými dvojicemi stav-akce (i,u), $i \in S$, $u \in U(i)$. Funkci následníka stavu i při akci u poskytuje výpočtový model soustavy. Splnění podmínky konvergence Q-učení, tj. že se během jednotlivých etap vhodným způsobem snižuje parametr učení $\alpha_k(i,u)$, má za důsledek to, že posloupnost Q-hodnot generovaná ve fázi předučení konverguje s pravděpodobností 1 k optimálním hodnotám. Protože je proces konvergentní, může být po provedení dostatečného počtu etap fáze předučení použito dosažených Q-hodnot k určení strategie, která již může být velmi blízká optimální strategii. Protože tato strategie není citlivá na malé chyby v aproximaci Q-hodnot, lze použít pouze přibližného výpočtového modelu soustavy.

Po ukončení fáze předučení, je možno provést fázi doučení s reálnou soustavou. V této fázi je strategie získaná ve fázi předučení dále zpřesňována a přizpůsobována skutečným provozním podmínkám soustavy. Jako počátečních Q-hodnot je použito Q-hodnot dosažených ve fázi předučení. Učení probíhá konvenčním způsobem, tj. provádí se zálohování paralelně s řízením. Zde se již předpokládá, že funkci následníka stavu s_t při použití akce $u_t \in S_t$ poskytuje přímo řízená soustava.

5 UVAŽOVANÝ ZPŮSOB ŘÍZENÍ ASYNCHRONNÍHO ELEKTROMOTORU

Na obr. 5-1 je zobrazeno blokové schéma QL-regulátoru při učení na minimalizaci regulační odchylky. Vstupem regulátoru jsou požadované otáčky ω_p. Na základě těchto požadovaných otáček a aktuálních otáček motoru je vypočtena aktuální hodnota regulační odchylky e. Aktuální regulační odchylka e, rychlost regulační odchylky $\dot{e}$ a zrychlení regulační odchylky $\ddot{e}$ popisují aktuální stav modelu. Pro aktuální stav prostředí se v tabulce Q-hodnot vybere akce s nejvyšší Q-hodnotou. V tomto případě jsou akcemi zvýšení či snížení řídicí frekvence o nějakou fixní hodnotu. Po provedení vybrané akce se na základě požadované a aktuální regulační odchylky modelu změna stavu vyhodnotí a dojde k aktualizaci příslušné Q-hodnoty v tabulce Q-hodnot. Z motoru vystupují aktuální otáčky ω které se na vstup regulátoru přivádějí zpětnou vazbou. Při tomto způsobu řízení stačí provést učení jednou pro libovolné otáčky.



Obr. 5-1: Blokové schéma učení na minimalizaci regulační odchylky.

6 IMPLEMENTAČNÍ PŘÍSTUPY

6.1 VÝPOČTOVÝ MODEL ASYNCHRONNÍHO STROJE

Simulace byly prováděny jednoduchým matematickým modelem asynchronního elektromotoru v transformovaných souřadnicích d, q, 0. Model přímo vychází z elektrického uspořádání a geometrie idealizovaného asynchronního stroje. Podle literatury [33] lze psát následující rovnice:

Rovnice pro statorová napětí:

$$\begin{bmatrix} u_{qs} \\ u_{ds} \\ u_{0s} \end{bmatrix} = R_s \begin{bmatrix} i_{qs} \\ i_{ds} \\ i_{0s} \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \Psi_{qs} \\ \Psi_{ds} \\ \Psi_{0s} \end{bmatrix} \tag{6.1-1}$$

Rovnice pro rotorová napětí:

$$\begin{bmatrix} u_{qr} \\ u_{dr} \\ u_{0r} \end{bmatrix} = R_r \begin{bmatrix} i_{qr} \\ i_{dr} \\ i_{0r} \end{bmatrix} - \omega_r \begin{bmatrix} \Psi_{dr} \\ -\Psi_{qr} \\ 0 \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \Psi_{qr} \\ \Psi_{dr} \\ \Psi_{0r} \end{bmatrix} \tag{6.1-2}$$

Rovnice pro spřažené magnetické toky rotoru a statoru:

$$\begin{bmatrix} \Psi_{qs} \\ \Psi_{ds} \\ \Psi_{0s} \\ \Psi^*_{qr} \\ \Psi^*_{dr} \\ \Psi^*_{0r} \end{bmatrix} = \begin{bmatrix} L_{ls}+L_m & 0 & 0 & L_m & 0 & 0 \\ 0 & L_{ls}+L_m & 0 & 0 & L_m & 0 \\ 0 & 0 & L_{ls} & 0 & 0 & 0 \\ L_m & 0 & 0 & L^*_{lr}+L_m & 0 & 0 \\ 0 & L_m & 0 & 0 & L^*_{lr}+L_m & 0 \\ 0 & 0 & 0 & 0 & 0 & L^*_{lr} \end{bmatrix} \begin{bmatrix} i_{qs} \\ i_{ds} \\ i_{0s} \\ i^*_{qr} \\ i^*_{dr} \\ i^*_{0r} \end{bmatrix} \tag{6.1-3}$$

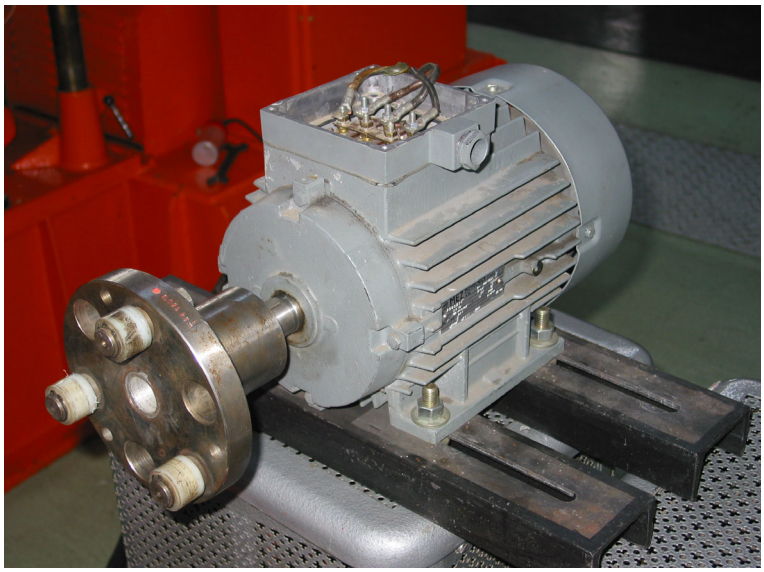
Rovnice pro elektrický moment stroje:

$$M_e = \frac{3}{2}\frac{P}{2}\left[\left(\Psi^*_{qr} i^*_{dr} - \Psi^*_{dr} i^*_{qr}\right)\right] = \frac{3}{2}\frac{P}{2}\left(\Psi_{ds} i_{qs} - \Psi_{qs} i_{ds}\right) \tag{6.1-4}$$

Implementace modelu byla provedena výpočtem rovnic 6.1-1 až 6.1-4 lichoběžníkovou metodou integrace v prostředí Borland Delphi.

Pro postupnou změnu řídicí frekvence jak při rozběhu, doběhu, tak i při vlastním řízení asynchronního stroje byla použita rozběhová, doběhová a řídicí rampa. Použití postupné změny frekvence má výhodu ve snížení hodnoty proudu tekoucího jak statorovým, tak rotorovým vinutím stroje. Použití postupné změny frekvence má za následek také snížení hodnoty překmitů otáček.

Jako fyzikální model byl použit 3-fázový asynchronní elektromotor 4AP 90S-2 jmenovitého výkonu 1.5 kW (viz. obr. 6.1-1) .



Obr. 6.1-1: Motor 4AP 90S-2.

6.2 IMPLEMENTACE Q-UČENÍ TABULKOU

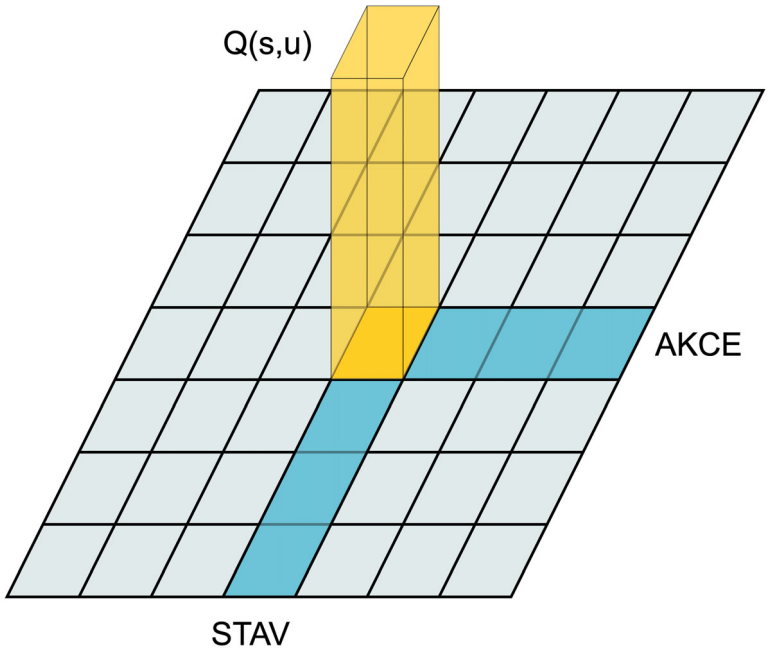
Q-funkce je v našem případě implementována tabulkou. Tato implementace je možná vždy, pokud je počet stavů přípustných akcí konečný. Existují také jiné způsoby implementace Q-funkce. Například můžeme využít metod aproximace funkcí založené na parametrizovaných modelech. Parametrické aproximace jsou vhodné zejména z toho důvodu, že mohou zobecnit tréninková data a mohou tak poskytovat odhad cen i pro stavy které ještě nebyli navštíveny. To je významné zejména pro velké množiny stavů.

Při implementaci tabulkou je stav prostředí s chápán jako aritmetický vektor jistých veličin prostředí $\mathbf{s} = \left(s^1, \ldots, s^n\right)$. Podobně akce u jako aritmetický vektor řídicích veličin $\mathbf{u} = \left(u^1, \ldots, u^m\right)$. Nechť $\left\{g_j^i\right\}_{j=0}^{p_i}$, je posloupnost uzlových bodů rastru i-té veličiny prostředí z s^i taková, že $g_j^i < g_{j'}^i \Leftrightarrow j < j'$, $j, j' = 0, \ldots, p_i$. Hodnotou indexu reprezentujícím hodnotu i-té veličiny prostředí vzhledem k rastru nazveme takový index j, pro který $g_j^i \le s^i < g_{j+1}^i$. Zcela analogicky je zavedena reprezentace akce skládající se z řídicích veličin s uzlovými body rastru $\left\{h_j^i\right\}_{j=0}^{r_i}$.

Funkce $Q(\mathbf{s},\mathbf{u})$ je potom reprezentována tabulkou

$$Q(\mathbf{s},\mathbf{u}) = q_{j^1 j^2 \ldots j^{n+m}}, \qquad (6.2\text{-}1)$$

kde j^i reprezentuje hodnotu i-té veličiny prostředí (resp. $i-n$-té řídicí veličiny). Implementace Q-učení tabulkou je znázorněna na obr. 6.2-1.



Obr. 6.2-1: Implementace Q-učení tabulkou.

Přiměřenost aproximace funkce $Q(s,u)$ tabulkou bude ovlivněna konkrétní volbou uzlových bodů rastru. Bude-li polí rastru málo, půjde zřejmě o příliš hrubou aproximaci, bude-li rastr příliš jemný, stane se neúnosným požadavek modifikací Q-hodnot nad značně vysokým počtem polí. V prvém případě je vyloučeno dosažení funkce $Q^*(s,u)$ s dostatečnou přesností, v druhém případě toto není dosažitelné bez zdlouhavého výpočtu.

Dalším faktorem je to, že "kvalita" strategie řízení bude zřejmě záviset i na velikosti a poloze (tj. uzlových bodech) jednotlivých polí rastru. Pro množinu stavů s požadavkem „jemnějších" řídicích zásahů bude třeba jemnějšího rastru, než v oblastech bez této podmínky. Pro jednoduchost byly rastry veličin konstruovány tak, aby bylo možno konkrétní rozmístění uzlových bodů rastru v-té veličiny definovat pomocí jediného koeficientu nelinearity rastru r_v. Nelinearita rastru byla nastavována hodnotou parametru $r_v \in \langle 0,1 \rangle$ úměrně členu $r_v^{|\Delta i|}$, kde $|\Delta i|$ značí vzdálenost indexu i–tého dílku rastru od indexu středového dílku. Poznamenejme, že případ $r_v = 1$ odpovídá lineárnímu rastru v-té veličiny.

Z hlediska dosažení strategie řízení s jistou hladinou úspěšnosti během určitého počtu průchodů výpočetní procedurou lze tedy očekávat existenci optimálního rastru, jak co do počtu polí, tak do jejich velikosti.

6.3 IMPLEMENTACE PROZKOUMÁVÁNÍ

K implementaci prozkoumávání bylo použito metody výběru akcí používající Boltzmanova rozložení (Sutton [41] a Watkins [46]). Metoda přiřazuje v aktuálním stavu každé přípustné akci pravděpodobnost jejího provedení:

$$P(u) = \frac{e^{Q(s_t,u)/T}}{\sum_{v \in U(s_t)} e^{Q(s_t,u)/T}},\ u \in U(s_t), \tag{6.3-1}$$

kde $T > 0$ je parametr určující, jak významně se tyto pravděpodobnosti budou podílet na výběru akce pro stav s_t. Pokud bude T vysoké, je preferováno prozkoumávání, s klesající hodnotou parametru T je náhodný výběr akce potlačován a realizuje se akce s nejvyšší Q-hodnotou pro stav s_t. Hodnota parametru T se postupně snižuje podle:

$$\left.\begin{aligned} T_0 &= T_{\max}, \\ T_{k+1} &= T_{\min} + \beta\left(T_k - T_{\min}\right), \end{aligned}\right\},\ \text{kde } k \text{ je číslo iterace.} \tag{6.3-2}$$

Q-učení navíc vyžaduje, aby učící poměr $\alpha_k(i,u)$ vhodně klesal. Bylo použito snižování učícího poměru navržené Darkenem [15]. Snižování je navrženo následovně:

$$\alpha_k(i,u) = \frac{\alpha_0 n_0}{n_0 + n_k(i,u)}, \tag{6.3-3}$$

kde $n_k(i,u)$ je počet záloh provedených na Q-hodnotě (i,u) do etapy k, α_0 je počáteční parametr učení a n_0 je parametr určený k řízení rychlosti snižování učícího poměru $\alpha_k(i,u)$.

7 SIMULAČNÍ PŘÍSTUPY

V provedených experimentech byla zjišťována kvalita regulace během 5000 kroků učení. Během každého kroku učení byl 50x proveden rozběh asynchronního elektromotoru z nulových otáček na hodnotu požadovaných otáček se zátěžným momentem $Mz = 2.5Nm$. Požadované otáčky se měnily v rozsahu $250\,\text{min}^{-1}$ až $2750\,\text{min}^{-1}$, aby byla vyzkoušena co nejširší oblast regulace. Simulace trvala 10 sekund a byla rozdělena na 2 fáze. Na fázi rozběhu o délce trvání 9s a na fázi řízení o délce trvání 1s. V obou fázích bylo vypočteno průměrné integrální kriterium kvality regulace.

V legendách následujících grafů udává první číslice počet dílků rastru regulační odchylky e, druhé číslo udává počet dílků rastru její rychlosti $\dot{e}$ a třetí číslo udává počet dílků rastru jejího zrychlení $\ddot{e}$. Poslední číslo udává počet použitých akcí.

7.1 PARAMETRY PROVEDENÝCH EXPERIMENTŮ

V jednotlivých sadách simulací byly brány jako výchozí podmínky parametry uvedené v tab. 7.1-1, vůči kterým byla měněna typicky hodnota jednoho parametru.

Množina řídicích akcí	$\{0Hz,\ 50Hz\}$	
Integrační krok	$t = 1 \cdot e^{-4}$	
Okamžité posílení	prostá pokuta	
Parametry předučení	$\alpha = 0.2$, $n_0 = 300$, $\gamma = 0.999$ $T_{\min} = 5$, $T_{\max} = 75$	viz. (6.3-3) viz. (3-2) viz. (6.3-2)
Parametry doučování	$\alpha = 0.1$, $n_0 = 100$, $\gamma = 0.999$ $T_{\min} = 5$, $T_{\max} = 75$	viz. (6.3-3) viz. (3-2) viz. (6.3-2)
Zátěžný moment	konstantní $Mz = 2.5Nm$	

Tab. 7.1-1: Standardní podmínky simulací.

7.2 HODNOCENÍ VÝSLEDKŮ SIMULACÍ

Pro porovnání kvality regulace bylo vybráno obvyklé integrální kritérium kvality regulace dané vztahem:

$$I_k(T) = \frac{1}{T}\int_0^T e^2(t)\,dt, \qquad \text{pro } T > 0, \tag{7.2-1}$$

kde I_k je hodnota integrálního kritéria a e je velikost regulační odchylky otáček. Integrální kritérium kvality regulace je vyčísleno pro každou simulaci.

8 URČENÍ STAVŮ PROSTŘEDÍ

Počáteční testy měly za úkol zjistit, jeví-li se jako vhodnější z hlediska úspěšnosti učení stav prostředí definovaný jako 1-D (uvažující pouze regulační odchylku), 2-D (uvažující regulační odchylku a její rychlost) nebo jako 3-D (uvažující regulační odchylku, její rychlost a zrychlení).

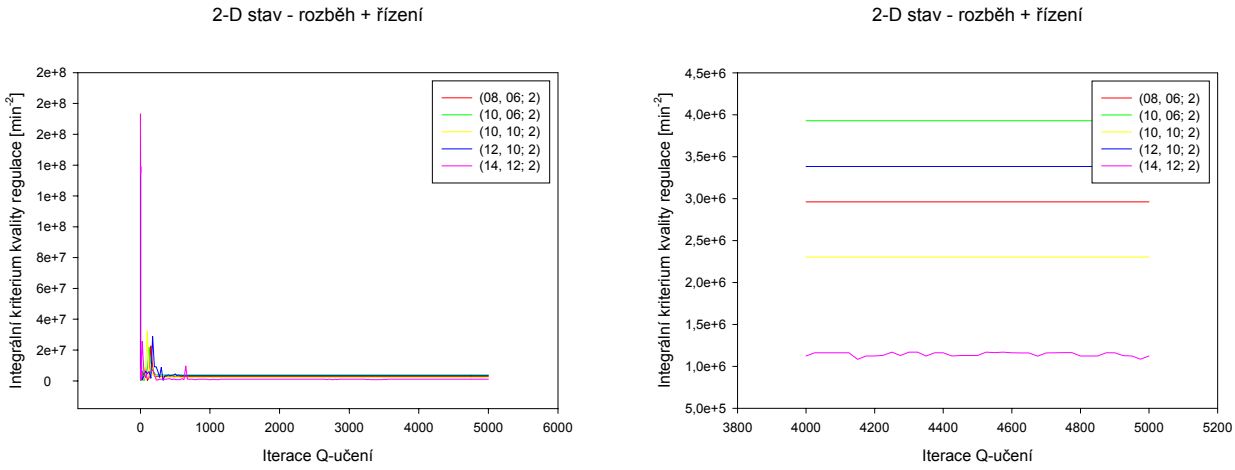
Z výsledků experimentů vyplynulo, že kvalita naučení je tím vyšší, čím větší je počet veličin popisujících aktuální stav prostředí a proto se pro další experimenty jeví jako nejnadějnější 2-D a 3-D sada prostředí. Jako naprosto nevyhovující se ukázala 1-D sada prostředí, která používá pouze regulační odchylku otáček a která ani s mřížkou definující více něž 250 buněk stavů soustavy nedosáhla použitelné strategie řízení. Proto byly 1-D stavy soustavy z dalších zkoumání vyloučeny.

9 FÁZE PŘEDUČENÍ S LINEÁRNÍMI MŘÍŽKAMI

9.1 PRŮBĚH PŘEDUČENÍ LINEÁRNÍCH 2-D SAD PROSTŘEDÍ

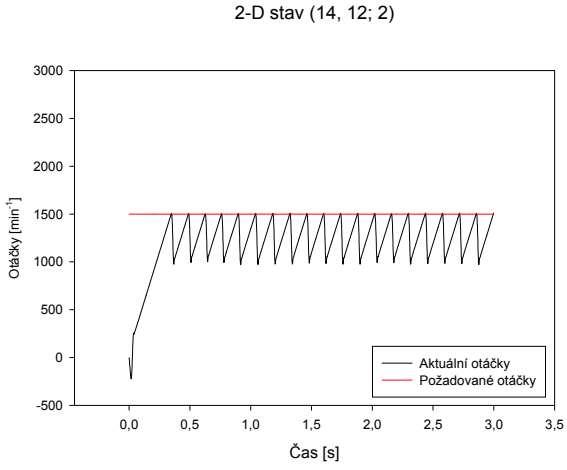
Čím více bylo polí regulační odchylky a rychlosti regulační odchylky, tím byla kvalita naučení vyšší. Můžeme proto předpokládat, že s dalším zvyšováním počtu polí regulační odchylky a rychlosti regulační odchylky se bude dále zvyšovat i kvalita regulace.

V dalších experimentech bylo zvoleno 5 různých stavových prostorů definovaných jako 2-D stav a bylo provedeno několik simulací k nalezení nejlepší mřížky z hlediska kvality regulace. Experimenty jsou na obr. 9.1-1. Nejnižší hodnoty kriteria kvality regulace bylo podle předpokladů dosaženo pro největší mřížku $(14, 12; 2)$.



Obr. 9.1-1: Vliv různých lineárních rastrů na průběh předučení, 2-D stav, rozběh + řízení AM.

Na tomto místě bych rád poznamenal, že u všech 2-D mřížek docházelo k oscilacím (viz. obr. 9.1-2), které byly zapříčiněny nedostatečným popsáním aktuálního stavu AM 2-D mřížkou.



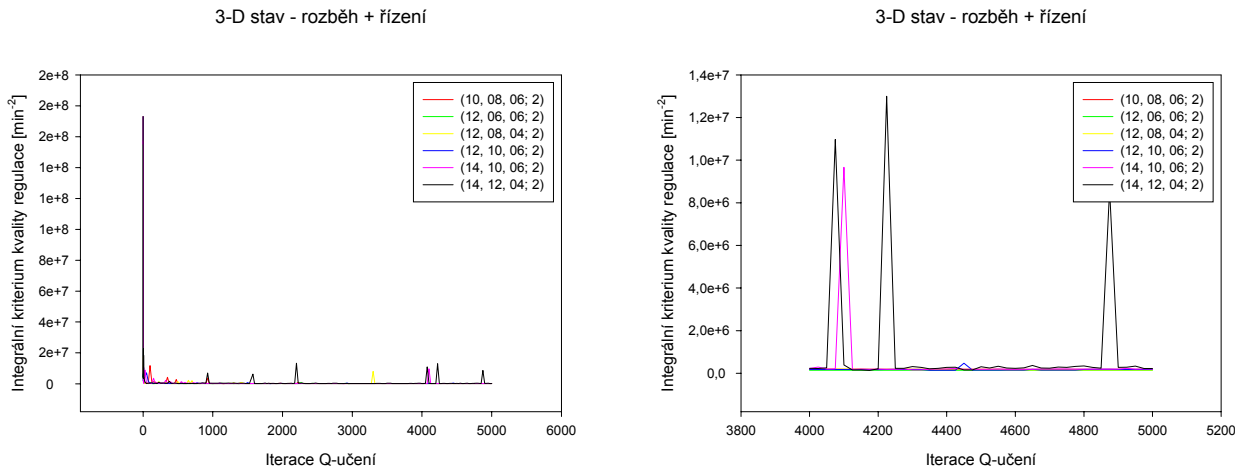
Obr. 9.1-2: Ukázka oscilací, 2-D stav, rozběh + řízení AM.

9.2 PRŮBĚH PŘEDUČENÍ LINEÁRNÍCH 3-D SAD PROSTŘEDÍ

V následujících experimentech byl zkoumán vliv lineárního rastru regulační odchylky, rychlosti regulační odchylky a zrychlení regulační odchylky na průběh kvality naučení u 3-D sad prostředí. Při zvyšování počtu polí rychlosti regulační odchylky a zrychlení regulační odchylky se kvalita dosažená učením značně zhoršuje. Pouze při zvýšení počtu polí regulační odchylky se kvalita předučení téměř nemění. Při zvyšováním počtu polí dochází ke zvýšení počtu fluktuací jednotlivých veličin. Tyto fluktuace se projevují více než u 2-D stavů prostředí.

V dalších experimentech bylo zvoleno 6 různých stavových prostorů definovaných jako 3-D a bylo provedeno několik experimentů k nalezení nejlepší mřížky z hlediska kvality regulace. Experimenty jsou zobrazeny na obr. 9.2-1. Nejnižší hodnoty integrálního kriteria kvality regulace a nejmenšího počtu fluktuací bylo dosaženo pro mřížku $(12, 8, 4; 2)$.

Použitelnost strategie získané ve fázi předučení byla dále ověřována simulacemi, během kterých byla testována odolnost strategie vůči chybám pozorování veličin soustavy, vůči zpoždění akčního zásahu a vůči odezvě na skokový moment.



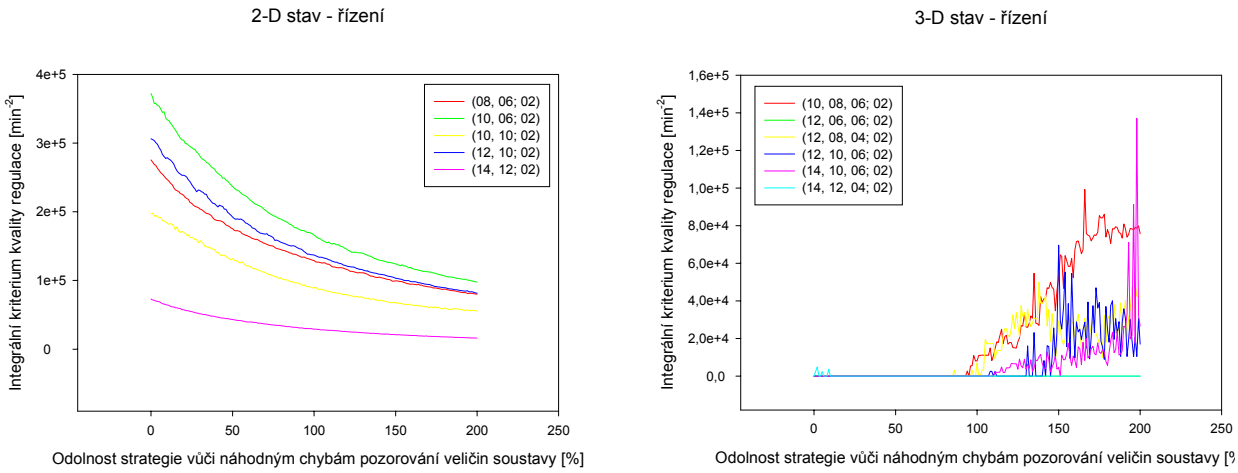
Obr. 9.2-1: Vliv různých lineárních rastrů na průběh předučení, 3-D stav, rozběh + řízení AM.

9.3 ODOLNOST STRATEGIE VŮČI NÁHODNÝM CHYBÁM POZOROVÁNÍ VELIČIN SOUSTAVY

Výsledky testů jsou shrnuty na obr. 9.3-1. Chyby pozorování veličin soustavy byly zavedeny do všech veličin stavu soustavy.

Můžeme si všimnout, že u mřížek 2-D stavů mají závislosti velmi podobný charakter. Téměř u všech mřížek docházelo se vzrůstající hodnotou chyby pozorování veličin soustavy k paradoxnímu zlepšování kvality regulace. To bylo pravděpodobně zapříčiněno narušením oscilací (viz. obr. 9.1-2) touto chybou a k následnému zlepšení strategie řízení. Nejvyšší odolnost vůči náhodné chybě pozorování veličin soustavy vykazuje mřížka $(14, 12; 2)$.

U mřížek 3-D stavů byla při řízení (9-10 sekunda simulace) zachována průměrná hodnota integrálního kriteria kvality regulace, až přibližně do 80 % úrovně chyb. Takto velká odolnost je s největší pravděpodobností zapříčiněna způsobem generování velikosti náhodných chyb. Tato procentuální velikost je totiž odvozena z aktuální hodnoty veličiny popisující stav prostředí (regulační odchylka, rychlost regulační odchylky, zrychlení regulační odchylky) a protože při řízení kolem ustálené polohy je tato velikost minimální můžeme si dovolit, až téměř 80 procentní chybu pozorování veličin soustavy. S dalším zvyšováním úrovně chyb pozorování veličin soustavy se integrální kriterium kvality regulace zvyšuje.



Obr. 9.3-1: Odolnost strategie vůči náhodné chybě pozorování veličin soustavy, řízení AM.

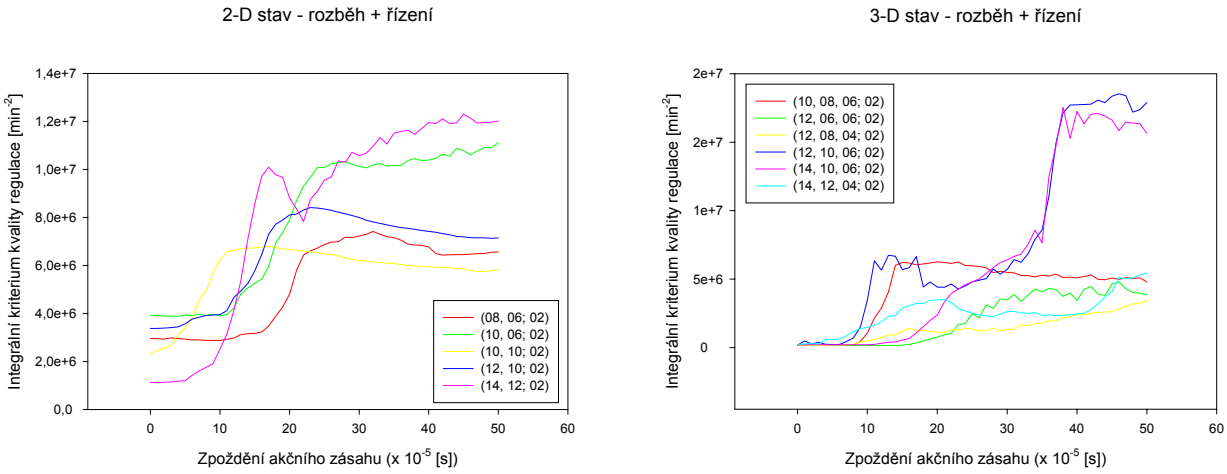
9.4 ODOLNOST STRATEGIE VŮČI ZPOŽDĚNÍ AKČNÍHO ZÁSAHU.

Výsledky jsou shrnuty na obr. 9.4-1. U všech mřížek mají závislosti velmi podobný charakter. Nad hodnotou přibližně 10×10^{-5} [s] začíná u 2-D stavu prudce stoupat hodnota integrálního kriteria kvality regulace. Nejlépe se chová mřížka $(8, 6; 2)$, u které se zlomová hodnota zpoždění akčního zásahu posouvá o něco výše, asi na hodnotu 18×10^{-5} [s]. Avšak mřížka $(14, 12; 2)$ dosahuje do hodnoty 10×10^{-5} [s] daleko nižší hodnoty integrálního kriteria kvality regulace.

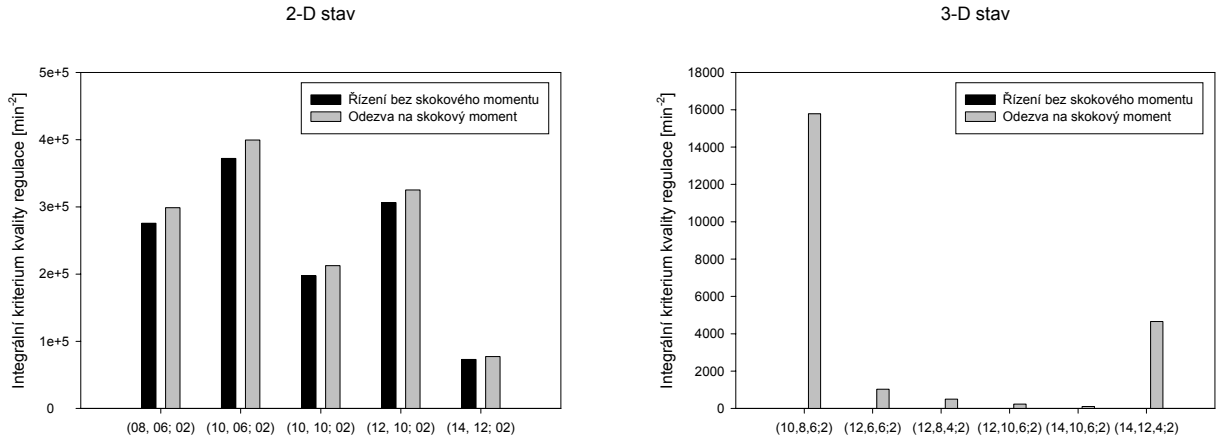
U 3-D stavu se zlomová hodnota zpoždění vyskytuje také na hodnotě 10×10^{-5} [s], přičemž strmost více závisí na volbě mřížky. Nejvíce vyniká mřížka $(12, 6, 6; 2)$ u které se zlomová hodnota zpoždění akčního zásahu posouvá přibližně na hodnotu 16×10^{-5} [s] a která průběh stoupání hodnoty integrálního kriteria kvality regulace nemá tak strmý.

9.5 ODEZVA NA SKOKOVÝ MOMENT

Výsledky testů odezvy řízení na skokový moment jsou uvedeny na obr. 9.5-1. Experimenty probíhali následovně. Nejdříve byl proveden rozběh AM bez zátěžného momentu o délce trvání 9s. Poté byl motor zatížen zátěžným momentem $Mz = 5Nm$ a bylo měřeno integrální kriterium kvality regulace. V grafu první hodnota udává velikost integrálního kriteria kvality regulace bez skokového momentu a druhá hodnota udává velikost integrálního kriteria se skokovým momentem.



Obr. 9.4-1: Odolnost strategie vůči zpoždění akčního zásahu, rozběh + řízení AM.



Obr. 9.5-1: Odezva na skokový moment.

Z testovaných 2-D mřížek se nejlépe chovala mřížka $(14, 12; 2)$. U této mřížky byl také nejmenší rozdíl mezi hodnotami integrálního kritéria kvality regulace bez skokového momentu a se skokovým momentem.

U všech testovaných 3-D mřížek byla hodnota integrálního kriteria kvality regulace bez skokového momentu rovna 0. Po zatížení skokovým momentem se nejlépe chovala mřížka $(14, 10, 6; 2)$. Nejhůře se chovala mřížka $(10, 8, 6; 2)$.

9.6 VÝBĚR VHODNÉ MŘÍŽKY

Z posuzovaných mřížek 2-D stavu je jak z hlediska odolnosti vůči chybám pozorování veličin soustavy tak z hlediska zpoždění akčního zásahu a odezvy řízení na skokový moment nejlepší mřížka $(14, 12; 2)$.

Z posuzovaných mřížek 3-D stavu byla jako nejlepší zvolena mřížka $(12, 8, 4; 2)$, protože vykazovala nejvyšší odolnost vůči náhodné chybě pozorování veličin soustavy a ještě nevykazovala nejhorší odolnost vůči zpoždění akčního zásahu a odezvě na skokový moment.

10 FÁZE PŘEDUČENÍ S NELINEÁRNÍMI MŘÍŽKAMI

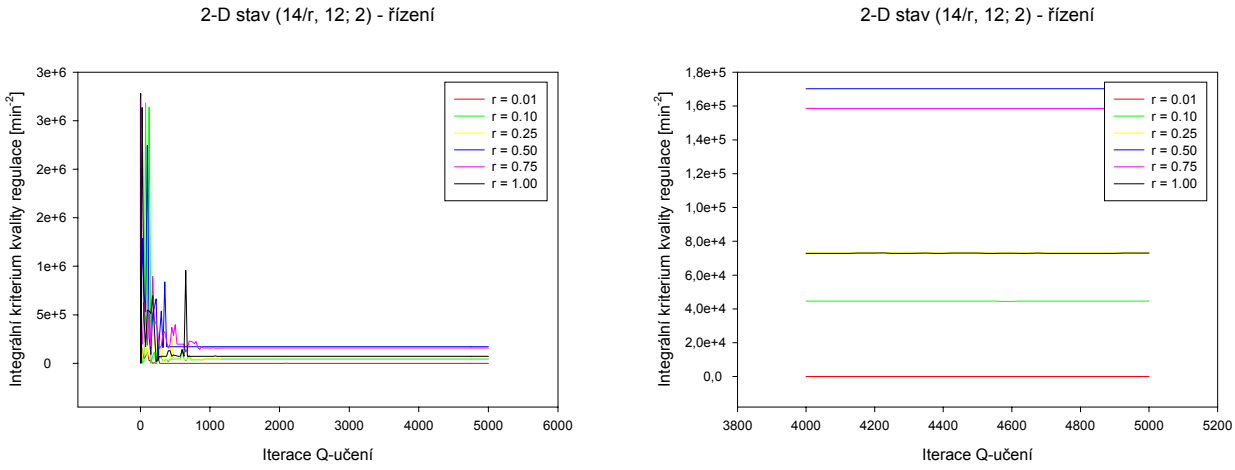
Simulace v této kapitole jsou navrženy s cílem ověřit, zda zmenšením vzdálenosti uzlových bodů nejbližších uzlovému bodu, který reprezentuje nulovou hodnotu veličiny, dojde ke zvýšení rozlišovací schopnosti strategie získané předučením v oblasti malých hodnot veličin a tím i ke zlepšení kvality regulace.

Dalším cílem bylo zjistit, jak se změní průběh procesu předučení, odolnost získané strategie vůči náhodným chybám pozorování veličin soustavy, odolnost vůči zpoždění provedení akčního zásahu a také odolnost vůči odezvě na skokový moment. K simulacím bylo použito strategií s mřížkami $(14, 12; 2)$ a $(12, 8, 4; 2)$ které byly získány experimenty v předchozích kapitolách.

10.1 PRŮBĚH PŘEDUČENÍ NELINEÁRNÍCH 2-D SAD PROSTŘEDÍ

Se vzrůstající nelinearitou regulační odchylky a rychlosti regulační odchylky se kvalita naučení zvyšuje. Toto zvýšení je nejmarkantnější pro hodnotu faktoru nelinearity $r = 0.01$.

V dalších simulacích byla nelinearizována 2-D mřížka $(14, 12; 2)$ získaná v odst. 9. Experimenty jsou zobrazeny na obr. 10.1-1. Nejnižší hodnoty integrálního kriteria kvality regulace bylo dosaženo pro faktor nelinearity $r = 0.01$.

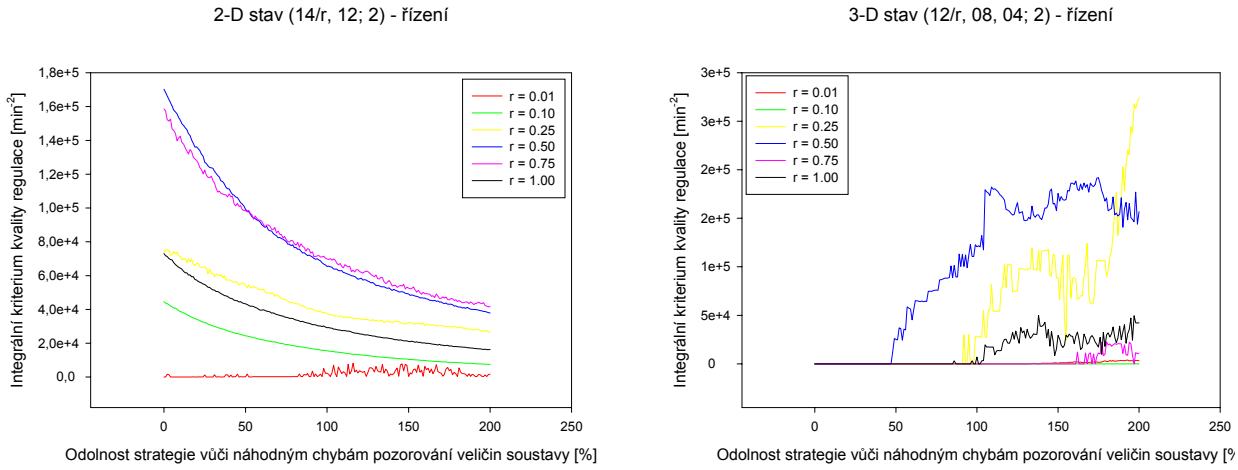


Obr. 10.1-1: Vliv různých nelineárních rastrů na průběh předučení, řízení AM.

10.2 ODOLNOST STRATEGIE VŮČI NÁHODNÝM CHYBÁM POZOROVÁNÍ VELIČIN SOUSTAVY

U mřížek 2-D stavů mají všechny závislosti velmi podobný charakter. Téměř u všech mřížek docházelo se vzrůstající hodnotou chyby pozorování veličin soustavy k paradoxnímu zlepšování kvality regulace, kromě mřížky s faktorem nelinearity $r = 0.01$. To bylo zapříčiněno narušením oscilací touto chybou a následný zlepšením strategie řízení. Nejvyšší odolnost vůči náhodné chybě pozorování veličin soustavy vykazuje mřížka $(14/0.01, 12; 2)$ u které zůstává zachována průměrná hodnota integrálního kriteria kvality regulace, až přibližně do 80 % úrovně chyb.

U mřížek 3-D stavů bylo při řízení AM dosaženo se vzrůstající hodnotou chyby pozorování veličin soustavy nejnižší hodnoty integrálního kriteria kvality regulace s mřížkou $(12/0.1, 8, 4; 2)$.

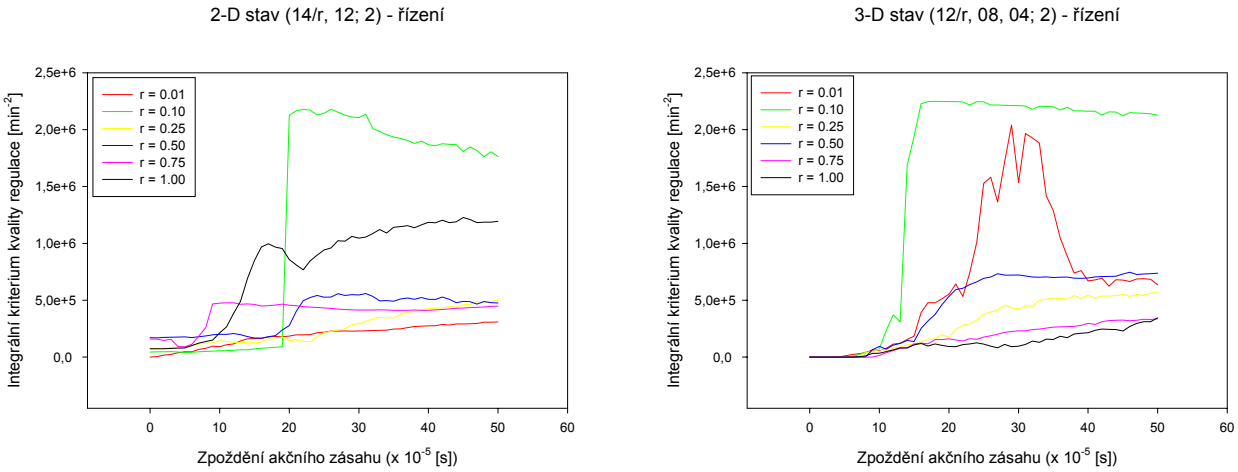


Obr. 10.2-1: Odolnost strategie vůči náhodné chybě pozorování veličin soustavy, řízení AM.

10.3 ODOLNOST STRATEGIE VŮČI ZPOŽDĚNÍ AKČNÍHO ZÁSAHU.

Na obr. 10.3-1 jsou shrnuty výsledky těchto simulací. U mřížek 2-D stavů odolnost strategie vůči zpoždění akčního zásahu velmi závisí na volbě mřížky. Nejméně strmý vzestup hodnoty integrálního kriteria kvality regulace bylo dosaženo s mřížkou $(14/0.01, 12; 2)$. Experimenty bylo zjištěno, že vzrůstající nelinearita má u 2-D mřížek positivní vliv na odolnost strategie vůči zpoždění akčního zásahu.

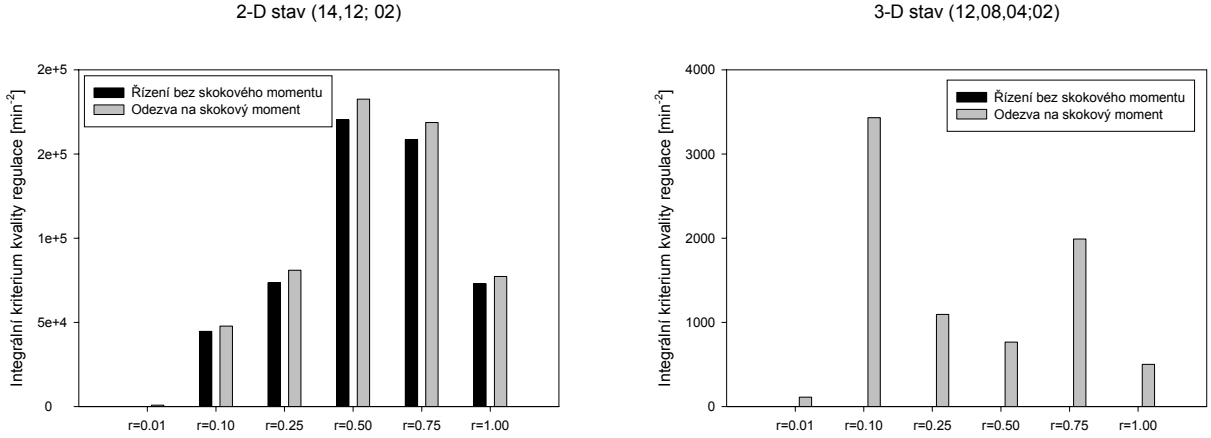
Nad hodnotou přibližně 10×10^{-5} [s] začíná u 3-D stavu prudce stoupat hodnota integrálního kriteria kvality regulace. Nejlépe se chová mřížka $(12/1, 8, 4; 2)$ u které je nejméně strmý vzestup hodnoty integrálního kriteria kvality regulace. Se vzrůstající nelinearitou se u 3-D mřížek značně zhoršuje odolnost strategie vůči zpoždění akčního zásahu.



Obr. 10.3-1: Odolnost strategie vůči zpoždění akčního zásahu, řízení AM.

10.4 ODEZVA NA SKOKOVÝ MOMENT

Výsledky testů odezvy řízení na skokový moment jsou uvedeny na obr. 10.4-1. V grafu první hodnota udává velikost integrálního kriteria kvality regulace bez skokového momentu a druhá hodnota udává velikost integrálního kriteria se skokovým momentem.



Obr. 10.4-1: Odezva na skokový moment.

Z testovaných 2-D mřížek se nejlépe chovala mřížka $(14/0.01, 12; 2)$. U této mřížky byl také nejmenší rozdíl mezi hodnotami integrálního kritéria kvality regulace bez skokového momentu a se skokovým momentem.

U všech testovaných 3-D mřížek byla hodnota integrálního kriteria kvality regulace bez skokového momentu rovna 0. Po zatížení skokovým momentem se nejlépe chovala mřížka $(12/0.01, 8, 4; 2)$ a $(12/1, 8, 4; 2)$. Nejhůře se chovala mřížka $(12/0.1, 8, 4; 2)$.

10.5 VÝBĚR VHODNÉ MŘÍŽKY

Z posuzovaných mřížek 2-D stavu je jak z hlediska odolnosti vůči chybám pozorování veličin soustavy tak z hlediska zpoždění akčního zásahu a odezvy řízení na skokový moment nejlepší nelineární mřížka $(14/0.01, 12; 2)$.

Z posuzovaných mřížek 3-D stavu byla jako nejlepší zvolena nelineární mřížka $(12/0.75, 8, 4; 2)$, protože ve všech pokusech vykazovala průměrně nejvyšší kvalitu regulace.

11 DALŠÍ EXPERIMENTY S FÁZÍ PŘEDUČENÍ

V této kapitole jsou pro úplnost uvedeny některé další výsledky, které dokreslují vlastnosti strategií získaných ve fázi předučení.

11.1 POSILOVACÍ FUNKCE

V této skupině simulací byl testován vliv posilovacích funkcí jednak na rychlost předučení, jednak na odolnost strategie vůči náhodné chybě pozorování veličin, odolnost strategie vůči zpoždění akčního zásahu a odolnost strategie vůči skokovému momentu.

Experimenty byly provedeny se strategiemi, které používali mřížek $(14, 12; 2)$, $(14/0.01, 12; 2)$, $(12, 8, 4; 2)$, $(12/0.75, 8, 4; 2)$.

Volba posilovací funkce měla značný vliv na kvalitu naučení. Posilovací funkce definované jako prostá odměna, lineární posílení a kvadratická odměna nevykazovali ani po 5000 pokusech nejmenší známky konvergence. Nejlepších výsledků bylo dosaženo s posilovacími funkcemi definovanými jako kvadratická pokuta a obecná pokuta. Je proto zřejmé, že ke správnému naučení je třeba systém vhodně pokutovat.

Nejlepší odolnosti vůči náhodným chybám pozorování soustavy bylo dosaženo s posilovacími funkcemi definovanými jako kvadratická pokuta a obecná pokuta.

Nejlepší odolnosti vůči zpoždění akčního zásahu bylo opět dosaženo pro posilovací funkce definované jako kvadratická pokuta a prostá pokuta.

V testech odezvy na skokový moment bylo s posilovací funkcí definovanou jako kvadratická pokuta dosaženo nejlepších výsledků pouze u lineární mřížky $(14, 12; 2)$. S ostatními mřížkami bylo dosaženo nejlepší odezvy na skokový moment s posilovací funkcí definovanou jako obecná pokuta.

Celkově bylo dosaženo nejlepších výsledků s posilovací funkcí ve tvaru prosté pokuty, která byla proto použita ve standardních podmínkách simulací. Jako naprosto nevyhovující se ve všech experimentech ukázaly posilovací funkce definované jako prostá odměna, lineární posílení a kvadratická odměna.

11.2 MNOŽINA AKCÍ

Tato skupina simulací byla navržena s cílem zjistit, zda bude rychlost a kvalita předučení, odolnost dosažené strategie vůči chybám pozorování veličin soustavy, odolnost vůči zpoždění akčního zásahu a vůči skokovému momentu positivně ovlivněna rozšířením množiny akcí $A = \{f_{\min}, f_{\max}\}$, tj. $\{0,\ 50\}\ [Hz]$ na množinu $\{0,\ \text{keep},\ 50\}\ [\text{Hz}]$, kde $f_{\min}$ a $f_{\max}$ je frekvence požadovaná QL-regulátorem, v našem případě 0 a 50 Hz. Hodnota *keep* znamená, že bude udržována předchozí hodnota frekvence. Použito bylo znovu mřížek $(14, 12; 2)$, $(14/0.01, 12; 2)$, $(12, 8, 4; 2)$, $(12/0.75, 8, 4; 2)$.

Vzhledem k předučení, které používalo původní dvouprvkovou množinu akcí, bylo dosaženo u všech mřížek horších výsledků. Hodnota integrálního kriteria kvality regulace byla mnohem vyšší.

11.3 POROVNÁNÍ S REFERENČNÍM PID REGULÁTOREM

Strategie s mřížkou definovanou jako $(12/0.75, 8, 4; 2)$ dosahovala až do úrovně 160 % chyb pozorování veličin nulové hodnoty regulační odchylky při řízení AM. Naproti tomu PID regulátor již od úrovně 0 % chyb pozorování veličin dosahoval vyšší hodnoty integrálního kriteria kvality regulace a tudíž horší kvalitu řízení. Tuto mírně vyšší hodnotu integrálního kriteria kvality regulace si ale PID regulátor udržel až do úrovně 200 % chyb pozorování veličin a dosáhl proto pro velké chyby pozorování veličin lepších výsledků než strategie s mřížkou $(12/0.75, 8, 4; 2)$.

V testu odolnosti jednotlivých strategií vůči zpoždění akčního zásahu dosahovala strategie s mřížkou definovanou jako $(12/0.75, 8, 4; 2)$ až do hodnoty 10×10^{-5} [s] velmi malých hodnot integrálního kriteria kvality regulace. PID regulátor nízkých hodnot integrálního kriteria dosahoval pouze do hodnoty zpoždění akčního zásahu přibližně 4×10^{-5} [s].

Bez skokového zátěžného momentu strategie definovaná lineární mřížkou $(12, 8, 4; 2)$ a nelineární mřížkou $(12/0.75, 8, 4; 2)$ dosahovala hodnoty integrálního kriteria kvality regulace rovnu nule. PID regulátor se stejnými daty dosahoval jenom mírně horších výsledků. Při zatěžování skokovým zátěžným momentem dosáhl nejlepších výsledků PID regulátor s hodnotou integrálního kriteria kvality regulace $113\ \text{min}^{-2}$. Druhého nejlepšího výsledku dosáhla strategie s lineární 3-D mřížkou $(12, 8, 4; 2)$ s hodnotou integrálního kriteria kvality regulace $501\ \text{min}^{-2}$. Na tomto místě bych rád poznamenal, že s nelineární 3-D mřížkou definovanou jako $(12/0.01, 8, 4; 2)$ bylo dosaženo hodnoty integrálního kriteria kvality regulace $112\ \text{min}^{-2}$ a dosáhla proto srovnatelného výsledku jako PID regulátor. Dá se předpokládat, že dalším doučováním QL-regulátoru by došlo ke zlepšení jeho regulačních vlastností.

12 FÁZE DOUČOVÁNÍ

Po ukončení experimentů s fází předučení byly provedeny experimenty s fází doučení. Experimenty se zaměřily na vyzkoušení zpřesňování a přizpůsobování již dosažené strategie získané ve fázi předučení s matematickým modelem asynchronního motoru, na změny parametrů reálné soustavy oproti simulačnímu modelu AM. Jako počátečních Q-hodnot je požito Q-hodnot dosažených ve fázi předučení.

Fáze doučení byla prováděna za standardních podmínek simulací použitím klasické metody Q-učení (učení pokusem). Okamžité posílení agenta bylo definováno jako integrální pokuta, kdy agent v každém kroku učení obdržel posílení o velikosti záporné hodnoty integrálního kriteria kvality regulace.

Zátěžný moment Mz byl modelován konstantním zátěžným momentem o velikosti $2.5\ Nm$ a také schodovitou funkcí s náhodnou velikostí konstantních částí z intervalu $\langle 0,5 \rangle$ [Nm]. Aby byla zajištěna srovnatelnost testů, bylo vygenerováno 1000 náhodných velikostí zátěžného momentu a uloženy do souboru. Tyto data byly použity pro každý pokus. Požadované otáčky motoru byly $1500\,\mathrm{min}^{-1}$.

Během provedených experimentů s fází doučování docházelo k dalšímu zlepšení strategií, získaných ve fázi předučení. V několika málo případech však došlo i k jejich mírnému zhoršení.

13 EXPERIMENTY S UČENÍM VYUŽÍVAJÍCÍM STOCHASTICKOU STRATEGIÍ

Jelikož při použití běžné strategie Q-učení může dojít k tomu, že se v průběhu učení vytvoří pásma, ve kterých se bude systém pohybovat nejčastěji a ostatní uzlová místa budou použita méně často případně vůbec, tak byly provedeny experimenty se stochastickou strategií procházení jednotlivých uzlových bodů.

Ve všech provedených experimentech vykazovalo Q-učení se stochastickou strategií procházení jednotlivých uzlových bodů horších výsledků než běžná metoda Q-učení, která používala učení pokusem.

Pouze v experimentech kde byla dočasně odstraněna rampa frekvenčního měniče bylo dosaženo podobných výsledků jako při učení pokusem. Z toho se dá usuzovat, že simulační model AM obsahuje explicitní čas a proto je metoda stochastického procházení jednotlivých uzlových bodů vhodná pouze v těch případech, kde explicitní čas nefiguruje. Z tohoto důvodu je běžná metoda Q-učení výhodnější.

14 ZÁVĚR

V předložené disertační práci byla navržena metoda Q-učení pro realizaci řízení asynchronního elektromotoru, která se skládá z fáze předučení a fáze doučení. Během fáze předučení jsou na výpočtovém modelu prováděny pokusy, které jsou zpracovávány prováděním zálohování Q-učení v reálném čase. Výpočtový model může být pouze přibližný.

Počáteční testy měly za úkol zjistit, jeví-li se jako vhodnější z hlediska úspěšnosti učení stav prostředí definovaný jako 1-D, 2-D nebo jako 3-D. Z výsledků experimentů vyplynulo, že kvalita naučení je tím vyšší, čím větší je počet veličin popisujících aktuální stav prostředí a proto byl v dalších experimentech uvažován pouze 2-D a 3-D stav prostředí.

Experimenty prováděné v další etapě se týkaly optimalizace počtu intervalů lineárního a nelineárního rastru jednotlivých stavových veličin. Získané strategie řízení byly posuzovány nejprve z hlediska dosažené hodnoty integrálního kriteria kvality regulace, z hlediska odolnosti dosažených strategií vůči chybám pozorování soustavy, odolnosti vůči zpoždění akčního zásahu a odezvy na skokový moment.

Tímto způsobem byly vybrány 4 mřížky Q-funkce. Jedna pro lineární 2-D stav, jedna pro nelineární 2-D stav, jedna pro lineární 3-D stav a jedna pro nelineární 3-D stav. Mřížky definují jen velmi malý počet buněk stavů a tím vytvářejí velmi jednoduchou architekturu řídicího členu. Použitelné strategie řízení tyto architektury produkují již ve fázi předučení.

Experimenty prováděné v další etapě byly provedeny pro dokreslení vlastností strategií získaných ve fázi předučení. Byl testován vliv různých posilovacích funkcí a vliv rozšíření množiny akcí jednak na rychlost předučení, jednak na odolnost strategie vůči náhodné chybě pozorování veličin, odolnost strategie vůči zpoždění akčního zásahu a odolnost strategie vůči skokovému momentu. Bylo zjištěno, že volba posilovací funkce má značný vliv na kvalitu naučení a nejlepší kvality řízení bylo dosaženo s posilovací funkcí definovanou jako kvadratická pokuta a obecná pokuta. Dále bylo zjištěno, že rozšířením množiny akcí dojde ke snížení regulačních možností strategie a proto je výhodnější dvouprvková množina akcí. V této etapě bylo také provedeno porovnání QL-regulátoru s referenčním PID regulátorem, jehož parametry byly nastaveny pomocí Ziegler-Nicholsova pravidla. QL-regulátor oproti PID regulátoru dosáhl mírně lepších výsledků.

Během experimentů s fází doučování dochází k dalšímu zlepšení strategií, získaných z procesu předučení.

Teoretický přínos práce představuje aplikaci Q-učení v oblasti automatického návrhu řídicího členu asynchronního motoru.

Praktický přínos práce je možno spatřovat v simulačním ověření metody Q-učení, v posouzení kvality řízení dosažené různými variantami Q-učení. Simulace a učení byly prováděny na tento účel vyvinutém programovém vybavení. Toto programové vybavení je dalším praktickým přínosem práce.

Výsledky práce byly získány v rámci projektů MSM 262100024 „Výzkum a vývoj mechatronických soustav“, pilotního projektu ÚT AV ČR č. 52020 „Řízení kráčivého robotu s využitím metod umělé inteligence“, projektu navazujícího č. 52022 „Realizace základních řídicích členů kráčivého robotu“. A za podpory výzkumného záměru CEZ: J22/98: 261100009 „Netradiční metody studia komplexních a neurčitých systémů“.

LITERATURA

[1] Anderson, C., W.: Strategy Learning with multilayer connectionist representations. Tech. Report TR87-509.3, GTE Laboratories, Incorporated, Waltham, MA, 1987.

[2] Barto, A., G., Sutton, R., S., Anderson, C., W.: Neuron-like adaptive elements that can solve difficult learning control problems. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13(5), 834-846, 1983.

[3] Bellman, R.: Dynamic Programming. Princeton University Press, Princeton, NJ., 1957.

[4] Berry, D., A., Fristedt, B.: Bandit Problems: Sequential Allocation of Experiments. Chapman and Hall, London, Uk. 1985.

[5] Bertsekas, D., P.: Dynamic Programming: Deterministic and Stochastic Models. Prentice Hall, Englewood Cliffs, NJ. 1987.

[6] Bertsekas, D., P.: Dynamic Programming and Optimal Control. Athena Scientific, Belmont, Massachusetts, 1995. Volumes 1 and 2.

[7] Bertsekas, D., P., Tsitsiklis, J., N.: Neuro-dynamic Programming. Athena Scientific, 1996.

[8] Březina, T., Krejsa, J.: Determination of Q-function optimum grid applied on active magnetic bearing control task, Mechatronics, Robotics and Biomechanics, Hrotovice 2003.

[9] Březina, T.: Efektivní metoda Q-učení: Simulační posouzení použitelnosti pro řízení aktivního magnetického ložiska, Habilitační práce, VUT, FSI, Brno 2003.

[10] Burghes, D., Graham, A.: Introduction to Control Theory including Optimal Control. Ellis Horwood, 1980.

[11] Champagne, R., Dessaint, L. A., Sybille, G., Khodabakhian, B.: An Approach for Real Time Simulation of Electric Drives, IEEE Transaction on Energy Conversion, Vol. 15, No. 1, March 2000.

[12] Chapman, D., Kaelbling, L., P.: Input generalization in delayed reinforcement learning: an algorithm and performance comparisons. In: Proceedings IJCAI-91, Sydney, NSW, 1991.

[13] Chee, L., Munong, T.: Dynamic Simulation of Electric Machinery Using Matlab-Simuling. Prentice Hall, 1997.

[14] Crites, R., H., Barto, A., G.: Improving elevator performance using reinforcement learning. Advances in Neural Information Processing System 8: pp. 1017-1023. MIT press, 1996.

[15] Darken, C., Moody, J.: Note on learning rate schedule for stochastic optimization. In: R.P. Lippmann, J.E. Moody and D.S. Touretzky, eds., Advances in Neural information Processing Systems 3, Morgan Kaufmann, San Mateo, CA, 1991, 832-838.

[16] Dayan, P.: The convergence of TD(λ) for general λ. Machine Learning, 8(3), 341-362, 1992.

[17] Dayan, P., Sejnowski, T., J.: TD(λ) converges with probability 1. Machine Learning, 14(3), 1994.
[18] Dolinar, D., Weerdt, R., Belmans, R., Freeman, E. M.: Calculation of Two – Axis Induction Model Parameters Using Finite Elements, IEEE Transaction on Energy Conversion, Vol. 12, No. 2, June 1997.
[19] Gastli, A.: Identification of Induction Motor Equivalent Circuit Parameters Using the Single-Phase Test, IEEE Transaction on Energy Conversion, Vol. 14, No. 1, June 1999.
[20] Hagen, S., Kröse, B,: A Short Introduction to Reinforcement Learning. Benelearn-97, 7th Belgian-Dutch Conference on Machine Learning: pp. 7-12, 1997.
[21] Howard, R., A.: Dynamic Programming and Markov Processes. The MIT Press, Cambridge, MA. 1960.
[22] Howell, M., N., Best, M., C.: On-line PID tuning for engine idle-speed control using continuous action reinforcement learning automata, 1999.
[23] Kaelbling, L., P.: Learning in Embedded Systems. The MIT Press, Cambridge, MA, 1993.
[24] Kaebling, L., P., Littmann, M., L.: Andrew W. Moore. Reinforcement learning: A survey. Journal of Artificial Intelligence Research, 4, 1996.
[25] Kalaš, V.: Technická kybernetika elektrických pohonov, Bratislava, 1978.
[26] Kudla, J.: Determination of Static and Dynamic Nonlinear Inductances of an Induction Machine, International Workshop on Electric Machines, Prague, 1999.
[27] Kudla J.: Parameter Estimation of Induction Machine Nonlinear Mathematical Model Basing on Measurements in Transient States, sborník semináře Vybrané problémy elektrických strojů a pohonů, Hustopeče, Česká republika, 28. – 29. května 2001.
[28] Kudla, J.: Rownania i schematy zastepcze nieliniowego modelu matematycznego maszyny indukcyjnej, Zeszyty naukowe politechniki Slaskiej, Elektryka z. 168, Gliwice, 1999.
[29] Kvasnička, V.: Úvod do teórie neurónových sietí, IRIS, 1997.
[30] Mařík, V.: Umělá inteligence, Academia, Praha, 1993.
[31] Měřička J., Zoubek Z.: Obecná teorie elektrického stroje, SNTL Praha, 1973.
[32] Měřička, J., Hamata, V., Voženílek, P.: Elektrické stroje, ČVUT, 1997.
[33] Ong, Ch. M.: Dynamic Simulation of Electric Machinery, Prentice Hall, New Jersey, 1998.
[34] Puterman, M., L.: Markov Decision Processes-Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., New York, 1994.
[35] Schmidhuber, J., H.: Curious model-building control systems. In Proc. International Joint Conference on Neural Networks, Singapore, volume 2, pages 1458-1463. IEEE, 1991.

[36] Schmidhuber, J.: A general method for multi-agent learning and incremental selfimprovement in unrestricted enviroments. In Yao, X. Evolution Computation: Theory and Application. Scientific Publ. Co., Singapore, 1996.
[37] Schmidhuber, J.: A general method for multi-agent learning and incremental selfimprovement in unrestricted enviroments. In Yao, X. Evolution Computation: Theory and Application. Scientific Publ. Co., Singapore, 1996.
[38] Skalický, J.: Elektrické servopohony, 1999.
[39] Stengel, R., F.: Stochastic Optimal Control. John Wiley and Sons, 1986.
[40] Sutton, R., S.: Learning to predict by the method of temporal differences. Machine Learning, 3(1), 9-44., 1988.
[41] Sutton, R., S.: Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In Proceedings of the Seventh International Conference on Machine Learning, Austin, TX, 1990. Morgan Kaufmann.
[42] Sutton, R., S., Barto, A., G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, 1998.
[43] Šubrt, J.: Elektrické stroje, Brno, 1999.
[44] Vavřín, P.: Teorie automatického řízení 1, SNTL.
[45] Watkins, C., J., C., H., Dayan, P.: Q-learning, Machine Learning 8, 1992, 279-292.
[46] Watkins, C., J., C., H.: Learning from Delayed Rewards. Ph.D. thesis, King's College, Cambridge, UK, 1989.
[47] Williams, R., J., Baird, L., C.: Tight performance bounds on greedy policies based on imperfect value functions. Tech. rep. NU-CCS-93-14, Northeastern University, College of Computer Science, Boston, MA, 1993.
[48] Williams, R., J., Baird, L., C.: Analysis of some incremental variants of policy iteration: First steps toward understanding actor-critic learning systems. Technical Report NU-CCS-93-11, Northeastern University, College of Computer Science, Boston, MA, September 1993.

SEZNAM PUBLIKACÍ AUTORA

[1] Marada, T.: Využití Q-učení pro řízení přechodových stavů asynchronního elektromotoru, Brno 2001, 137 s., Diplomová práce na Fakultě strojního inženýrství.
[2] Marada, T.: Q-learning: Control of asynchronous electric motor, Inženýrská mechanika 2002, 13.5.2002, ISBN 80-214-2109-6.
[3] Marada, T.: Q-learning: Control of asynchronous electric motor, Mendel 2002, 5.6.2002, str. 341-346, ISBN 80-214-2135-5.
[4] Marada, T., Singule, V.: Využití Q-učení pro řízení asynchronního elektromotoru, Vybrané problémy elektrických strojů a pohonů, Lomnice u Tišnova, 19.5.2003, ISBN 80-214-2400-1.

[5] Marada, T., Březina, T., Singule, V.: Determination of Q-function optimum grid applied on asynchronous electric motor control task. Inženýrská mechanika 2004, 10.5.2004, ISBN 80-85918-88-9.
[6] Marada, T., Březina, T., Singule, V.: Stanovení optimálního rastru Q-funkce pro řízení asynchronního elektromotoru. Mechatronika 2004, Račkova dolina, Slovakia, 24.5.2004, ISBN 80-227-2064-X.
[7] Marada, T., Březina, T., Singule, V.: Q-learned Controller Performance of Asynchronous Electromotor. Mechatronics 2004, Warsaw, Poland, 23.9.2004, ISSN 0033-2089.

SUMMARY

Presented PhD thesis is focused on use of Q-learning method on asynchronous electric drive control. The control consists of prelearning and tutorage phase. During prelearning the attempts which are processed by real time Q-learning backup are performed on computational model. Computational model can be approximate only.

Presented thesis show simulation verification of proposed method on asynchronous electric drive mode. Only the actual running speed was used for control; actual control error, its velocity and acceleration are calculated.

Initial tests were performed in order to find what environment state definitions are more advantageous regarding the learning succesibility: 1-D (considers control error only), 2-D (considers control error and it's velocity) or 3-D (considering control error, its velocity and acceleration).

Further experiments consider optimization of linear and nonlinear grid of particular state variables. Found control policies were evaluated with respect to control quality integral criterion value, robustness of obtained policy against noise, robustness against action delay and step torque responses.

Experiments performed in additional stage are performed to further test properties of policies found during prelearning. Tests include influence of various reinforcement functions and action set expansion on prelearning speed, robustness of obtained policy against noise, robustness against action delay and step torque responses. This stage also covers comparison of Q-learning based controller with referential PID controller, whose parameters were set by Ziegler-Nichols method.

After prelearning stage experiments the tutorage stage experiments were performed. The experiments were focused on improvement and adaptation of already obtained policy found during prelearning stage with mathematical model of asynchronous drive and on changes of real system parameters against simulation model.

The problem area is topical. It fits into current trend of research in new control methods based on artificial intelligence methods, learning particularly. The basic feature of learning is expanded ability of adaptation, meaning the ability to automatically improve the behaviour of controlled system during e.g. change of operational parameters, etc.

CURRICULUM VITAE

Ing. Tomáš Marada
Žeravice 102, PSČ 696 47
Tel: +420 604 820 702
Email: marada@uai.fme.vutbr.cz

Osobní data:

Datum a místo narození:	22. listopadu 1976, Kyjov
Rodinný stav:	Svobodný

Vzdělání:

2001 – 2004 Postgraduální doktorské studium
Vysoké učení technické v Brně
Fakulta strojního inženýrství
Obor: Inženýrská mechanika
Téma disertační práce: Využití Q-učení pro řízení AM

1996 – 2001 Vysoké učení technické v Brně
Fakulta strojního inženýrství
Obor: Inženýrská informatika a automatizace
Téma diplomové práce: Využití Q-učení pro řízení AM

1991 – 1995 Integrovaná střední škola
Obor mechanik elektronik pro výpočetní techniku
Olomoucká 61
627 00 Brno

Pedagogická činnost: V rámci doktorského studia výuka předmětů kinematika, technická mechanika II, algoritmy umělé inteligence.

Jazykové znalosti: Angličtina – slovem i písmem

Zaměstnání:

Od září 1995 do října 1996 projekční technik
Od září 2004 asistent na VUT v Brně

Zájmy:

Počítače, programování
Letecké modelářství
Stolní tenis, cyklistika

Další informace: Řidičský průkaz skupiny A, B.