KYBERNETIKA ČÍSLO 5, ROČNÍK 4/1968

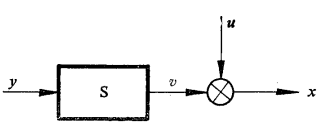
Nový přístup k identifikaci diskrétních dynamických soustav

VÁCLAV PETERKA

V článku se uvádí rozbor současného stavu problému identifikace dynamických soustav z naměřených průběhů vstupního a výstupního signálu a navrhuje se nový přístup k řešení této úlohy.

1. ÚVOD

Charakteristickou vlastností většiny dynamických soustav, se kterými se regulační technik setkává při automatizaci průmyslových technologických procesů je, že výstupní signál soustavy není jednoznačně určován pouze průběhem vstupního signálu (a event. počátečním stavem soustavy), ale že podléhá dalším neměřitelným náhodným vlivům. Transformace mezi vstupním a výstupním signálem, kterou soustava zprostředkuje, není tedy deterministická, nýbrž stochastická. V současné době nejčastěji užívaný (nikoliv však jediný možný) způsob popisu takovýchto soustav je



Obr. 1.

znázorněn na obr. 1. Blok S zde představuje ideální deterministickou soustavu, na jejímž výstupu je myšlený (virtuální) signál v, k němuž se přičítá šum u. Součet

$$x = v + u \tag{1}$$

je pak skutečně zjistitelným výstupním signálem soustavy. Takovýto způsob popisu má své oprávnění především v případě lineárních soustav, kde platí zákon superpozice a kde vnitřní šum může být bez potíží transformován na výstup soustavy. V šumu u jsou zahrnuty i případné náhodné chyby měření výstupního signálu.

Úlohou, kterou se budeme zabývat, je odhad neznámých parametrů deterministické části S soustavy, je-li znám konečný počet diskrétních hodnot vstupního signálu y a výstupního signálu x. Dynamické vlastnosti časově neproměnné lineární diskrétní soustavy lze popsat buď průběhem impulsní odezvy nebo diferenční rovnicí. V prvém případě jsou neznámými parametry soustavy diskrétní hodnoty impulsní odezvy $g(i)$, $(i = 0, 1, 2, \ldots)$, v druhém případě koeficienty diferenční rovnice, které jsou shodné s koeficienty diskrétního přenosu soustavy.

2. SOUČASNÝ STAV PROBLÉMU

V tomto odstavci popíšeme stručně nejužívanější metody identifikace diskrétních soustav a ukážeme důvody, které daly podnět k vypracování nového přístupu.

Pro výpočet impulsní odezvy se často užívá známé Wienerovy-Hopfovy rovnice, jejíž diskrétní tvar je

$$\varphi_{yx}(j) - \sum_{i=0}^{\infty} g(i)\, \varphi_{yy}(j - i) = 0 \quad \text{pro} \quad j \geqq 0 . \tag{2}$$

Místo vzájemně korelační funkce $\varphi_{yx}(j)$ a autokorelační funkce $\varphi_{yy}(j)$ se do rovnice dosazují jejich odhady

$$\hat{\varphi}_{yx}(j) = \frac{1}{L - j} \sum_{k=j+1}^{L} [x(k) - \bar{x}]\, [y(k - j) - \bar{y}] ,$$

$$\hat{\varphi}_{yy}(j) = \frac{1}{L - j} \sum_{k=j+1}^{L} [y(k) - \bar{y}]\, [y(k - j) - \bar{y}] , \tag{3}$$

kde $\bar{x}$ a $\bar{y}$ jsou střední hodnoty resp. jejich odhady.

Impulsní odezva dynamické soustavy může trvat teoreticky nekonečně dlouho, prakticky se však vždy předpokládá, že odezní v konečné době a určuje se pouze konečný počet jejích hodnot. O dalších hodnotách impulsní odezvy se předpokládá, že jsou nulové. Wienerova-Hopfova rovnice (2) tím přejde na soustavu lineárních rovnic pro odhady $\hat{g}(i)$, $(i = 0, 1, 2, \ldots, \lambda - 1)$

$$\hat{\varphi}_{yx}(j) - \sum_{i=0}^{\lambda-1} \hat{g}(i)\, \hat{\varphi}_{yy}(j - i) = 0 , \quad j = 0, 1, 2, \ldots, \lambda . \tag{4}$$

Wienerova-Hopfova rovnice (2) platí za předpokladu, že

a) šum u je stacionární ergodický,

b) rovněž y je stacionární a ergodický náhodný proces,

c) šum u není korelován se vstupním signálem y.

Předpoklad b) je možno vypustit, použijeme-li regresní model [2] a odhady pořadnic impulsní charakteristiky určíme tak, aby ztrátová funkce

$$Q = \sum_{k=1}^{L} [x(k) - \sum_{i=0}^{\lambda-1} g(i)\, y(k-i)]^2 \tag{5}$$

nabyla své minimální hodnoty. Při použití maticového zápisu lze (5) vyjádřit následovně

$$Q = (\boldsymbol{x} - \boldsymbol{Y}_\lambda \hat{\boldsymbol{g}})^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{Y}_\lambda \hat{\boldsymbol{g}}), \tag{6}$$

kde

$$\boldsymbol{Y}_\lambda = \begin{bmatrix} y(1) & y(0) & y(-1) & \dots & y(-\lambda+2) \\ y(2) & y(1) & y(0) & \dots & y(-\lambda+3) \\ y(3) & y(2) & y(1) & \dots & y(-\lambda+4) \\ \vdots & \vdots & \vdots & & \vdots \\ y(L) & y(L-1) & y(L-2) & \dots & y(L-\lambda+1) \end{bmatrix}, \tag{7}$$

$$\boldsymbol{x} = \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(L) \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} \hat{g}(0) \\ \hat{g}(1) \\ \vdots \\ \hat{g}(\lambda-1) \end{bmatrix}. \tag{8}$$

Minimalizací (6) plyne

$$\hat{\boldsymbol{g}} = (\boldsymbol{Y}_\lambda^{\mathrm{T}} \boldsymbol{Y}_\lambda)^{-1}\, \boldsymbol{Y}_\lambda^{\mathrm{T}} \boldsymbol{x}. \tag{9}$$

Získaný odhad bude nestranný, bude-li

$$E\{u(k) \mid y(k), y(k-1), \dots\} = E\{u(k)\} = 0 \quad \text{pro} \quad k = 1, 2, \dots, L, \tag{10}$$

tj. bude-li střední hodnota šumu nezávislá na předcházejících hodnotách vstupního signálu a rovna nule pro každý okamžik k v uvažovaném intervalu.

Platí totiž

$$E\{\hat{\boldsymbol{g}} - \boldsymbol{g}\} = E\{(\boldsymbol{Y}_\lambda^{\mathrm{T}} \boldsymbol{Y}_\lambda)^{-1}\, \boldsymbol{Y}_\lambda^{\mathrm{T}} \boldsymbol{x}\} - \boldsymbol{g} =$$
$$= E\{(\boldsymbol{Y}_\lambda^{\mathrm{T}} \boldsymbol{Y}_\lambda)^{-1}\, \boldsymbol{Y}_\lambda^{\mathrm{T}} (\boldsymbol{v} + \boldsymbol{u})\} - \boldsymbol{g} = E\{\boldsymbol{Y}_\lambda^{\mathrm{T}} \boldsymbol{Y}_\lambda)^{-1}\, \boldsymbol{Y}_\lambda^{\mathrm{T}} \boldsymbol{u}\},$$

což dává nulu právě za předpokladu (10).

Aby řešení (9) existovalo, musí být matice $(\boldsymbol{Y}_\lambda^{\mathrm{T}} \boldsymbol{Y}_\lambda)$ regulární. Vstupní signál nemůže být tedy jakýkoliv.

Jsou známé i jiné odhady pořadnic impulsní charakteristiky, které mají příznivější statistické vlastnosti. Nebudeme je zde však uvádět, protože vesměs vyžadují podrobnější apriorní znalosti statistických charakteristik šumu, které zpravidla nebývají

k dispozici. Dobrý přehled o metodách statistické identifikace soustav dávají přehledové referáty [3, 4], kde lze nalézti i další literaturu.

V praxi se často stává, že systém rovnic (4), na který jsme převedli řešení Wienerovy-Hopfovy rovnice (2), je špatně podmíněný, což značně ztěžuje a někdy i znemožňuje jeho řešení. Podaří-li se v takovém případě překonat potíže numerického řešení, bývá výsledek velmi nejistý, vzhledem k tomu, že i malé chyby v odhadech korelačních funkcí φ_{yx} a φ_{yy} mohou způsobit velké chyby ve výsledku. Podobné potíže se objevují i při inverzi matice $(\mathbf{Y}_\lambda^T \mathbf{Y}_\lambda)$ ve výrazu (9) a jejich hlavní příčiny jsou

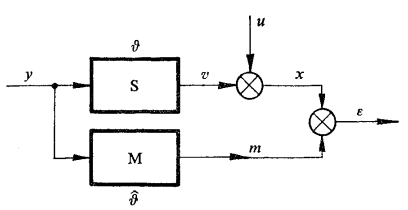
1. nevhodný vstupní signál (autokorelační funkce $\varphi_{yy}(j)$ nemá výrazný a dostatečně ostrý vrchol),
2. příliš velký počet odhadovaných parametrů.

První příčinu lze odstranit pouze vhodnější volbou vstupního signálu, je-li to možné. Dobré výsledky dává např. pseudonáhodný binární signál. Použití tohoto signálu pro identifikaci je popsáno např. v [5, 6, 7], kde lze nalézt též další literaturu.

Druhá příčina potíží může být odstraněna volbou jiné struktury deterministické části modelu soustavy. Při odhadu pořadnic impulsní charakteristiky jsme předpokládali jejich konečný počet. Nahrazovali jsme tak dynamickou soustavu soustavou s konečnou pamětí. Jak z hlediska dalšího použití, tak i z hlediska snížení počtu odhadovaných parametrů soustavy může být výhodnější předpokládat vazbu mezi vstupem y a výstupem v (obr. 1) ve tvaru diferenční rovnice a odhadovat přímo koeficienty této rovnice.

Uvažujme diferenční rovnici soustavy S ve tvaru

$$v(k) + \sum_{i=1}^{n} a_i v(k-i) = \sum_{i=0}^{n} b_i y(k-i)\,. \tag{11}$$



Obr. 2.

Při odhadování koeficientů diferenční rovnice soustavy je situace podstatně složitější než v případě odhadu pořadnic impulsní charakteristiky. Představme si, že na vstup myšleného modelu M (obr. 2) se přivádí tentýž signál jako na vstup originální soustavy. Přitom parametry modelu $\hat{\vartheta}$ jsou odhady parametrů ϑ skutečné soustavy S. Signál ε je rozdílem mezi výstupem reálné soustavy a jejího modelu.

Budou-li odhadovanými parametry ϑ diskrétní hodnoty impulsní charakteristiky g, bude ztrátová funkce (5) totožná se součtem čtverců

$$Q = \sum_{k=1}^{L} \varepsilon^2(k) . \tag{12}$$

Při daném vstupu model M s impulsní charakteristikou vypočtenou podle (9) reprodukuje tedy výstupní signál skutečné soustavy optimálně ve smyslu nejmenších čtverců.

Podobně Wienerovu-Hopfovu rovnici lze odvodit minimalizací střední kvadratické odchylky [8]

$$\bar{\varepsilon}^2 = \lim_{T=\infty} \frac{1}{2T+1} \sum_{k=-T}^{T} \varepsilon^2(k) . \tag{13}$$

Odhad koeficientů diferenční rovnice (11) lze rovněž založit na minimalizaci ztrátové funkce (12) nebo (13), vznikají však přitom značné obtíže. Systém rovnic pro odhady koeficientů $\hat{a}_i$ a $\hat{b}_i$ takto získaný je nelineární, jeho řešení je velmi nesnadné a může být i víceznačné. Navíc přistupuje otázka počátečních podmínek. Při zpracování dlouhých měření lze zpravidla vliv počátečních podmínek zanedbat a předpokládat např., že jsou nulové. V případě relativně krátkých záznamů by bylo třeba uvažovat počáteční podmínky jako další neznámé parametry, které je třeba odhadnout. Uvážíme-li dále, že získané odhady koeficientů diferenční rovnice nemají (až na konsistenci) žádné výjimečné statistické vlastnosti, dojdeme k závěru, že tento přístup není pro explicitní numerické řešení vhodný. Pro uvedené obtíže lze ho použít pouze ve spojení s technikou adaptivních modelů, která nabyla v poslední době široké uplatnění. Pojednává o ní celá řada referátů v [1].

Uvedeme ještě jeden užívaný přístup k řešení diskutované úlohy. Dosadíme-li do diferenční rovnice (11) vztah $v = x - u$, dostaneme

$$x(k) + \sum_{i=1}^{n} a_i x(k-i) - \sum_{i=0}^{n} b_i y(k-i) = \delta(k) , \tag{14}$$

kde $\delta(k)$ je tzv. chyba rovnice

$$\delta(k) = u(k) + \sum_{i=1}^{n} a_i u(k-i) . \tag{15}$$

Máme-li k dispozici $L + n$ naměřených hodnot vstupního a výstupního signálu $y(k)$, $x(k)$, $(k = -n+1, -n+2, \ldots, L)$, máme možnost sestavit z těchto výsledků měření L rovnic typu (14). Tuto soustavu rovnic můžeme zapsat v maticovém tvaru

$$\boldsymbol{x} + \boldsymbol{X}\boldsymbol{a} - \boldsymbol{Y}\boldsymbol{b} = \boldsymbol{\delta} , \tag{16}$$

kde

$$\boldsymbol{x} = \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(L) \end{bmatrix}, \quad \boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}, \quad \boldsymbol{\delta} = \begin{bmatrix} \delta(1) \\ \delta(2) \\ \vdots \\ \delta(L) \end{bmatrix}, \tag{17}$$

$$\boldsymbol{X} = \begin{bmatrix} x(0) & x(-1) & x(-2) & \dots & x(-n+1) \\ x(1) & x(0) & x(-1) & \dots & x(-n+2) \\ x(2) & x(1) & x(0) & \dots & x(-n+3) \\ \vdots & \vdots & \vdots & & \vdots \\ x(L-1) & x(L-2) & x(L-3) & \dots & x(L-n) \end{bmatrix}, \tag{18}$$

$$\boldsymbol{Y} = \begin{bmatrix} y(1) & y(0) & y(-1) & \dots & y(-n+1) \\ y(2) & y(1) & y(0) & \dots & y(-n+2) \\ y(3) & y(2) & y(1) & \dots & y(-n+3) \\ \vdots & \vdots & \vdots & & \vdots \\ y(L) & y(L-1) & y(L-2) & \dots & y(L-n) \end{bmatrix}, \tag{19}$$

a nebo ještě stručněji

$$\boldsymbol{x} - \boldsymbol{Z}\boldsymbol{\vartheta} = \boldsymbol{\delta}\,, \tag{20}$$

kde

$$\boldsymbol{\vartheta} = \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}, \quad \boldsymbol{Z} = [-\boldsymbol{X}, \boldsymbol{Y}]\,. \tag{21}$$

Předpokládejme, že $L > 2n + 1$, tj. že máme k dispozici více rovnic než je neznámých, jejich pravé strany (15) nejsou ovšem známy a představují v rovnicích náhodný člen.

Přesvědčme se, že často užívaná metoda nejmenších čtverců aplikovaná na systém rovnic (20) dává asymptoticky jednostranné odhady a není proto vhodná pro tento případ. Minimalizací ztrátové funkce

$$Q = \boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\delta} \tag{22}$$

dostaneme následující odhad vektoru neznámých koeficientů

$$\hat{\boldsymbol{\vartheta}}_L = (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^{-1}\,\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{x}\,. \tag{23}$$

Vyjádřeme $\boldsymbol{x}$ v (23) pomocí (20). Dostaneme tak

$$\hat{\boldsymbol{\vartheta}}_L - \boldsymbol{\vartheta} = (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^{-1}\,\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\delta} \tag{24}$$

a po rozepsání podle (21)

$$\begin{bmatrix} \hat{\boldsymbol{a}}_L - \boldsymbol{a} \\ \hat{\boldsymbol{b}}_L - \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X}, & -\boldsymbol{X}^T\boldsymbol{Y} \\ -\boldsymbol{Y}^T\boldsymbol{X}, & \boldsymbol{Y}^T\boldsymbol{Y} \end{bmatrix}^{-1} \begin{bmatrix} -\boldsymbol{X}^T \\ \boldsymbol{Y}^T \end{bmatrix} \boldsymbol{\delta} . \tag{25}$$

Označme analogicky s (18)

$$\boldsymbol{V} = \begin{bmatrix} v(0) & v(-1) & v(-2) & \dots & v(-n+1) \\ v(1) & v(0) & v(-1) & \dots & v(-n+2) \\ v(2) & v(1) & v(0) & \dots & v(-n+3) \\ \vdots & \vdots & \vdots & & \vdots \\ v(L-1) & v(L-2) & v(L-3) & \dots & v(L-n) \end{bmatrix}, \tag{26}$$

$$\boldsymbol{U} = \begin{bmatrix} u(0) & u(-1) & u(-2) & \dots & u(-n+1) \\ u(1) & u(0) & u(-1) & \dots & u(-n+2) \\ u(2) & u(1) & u(0) & \dots & u(-n+3) \\ \vdots & \vdots & \vdots & & \vdots \\ u(L-1) & u(L-2) & u(L-3) & \dots & u(L-n) \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} u(1) \\ u(2) \\ u(3) \\ \vdots \\ u(L) \end{bmatrix}. \tag{27}$$

Zřejmě platí $\boldsymbol{X} = \boldsymbol{V} + \boldsymbol{U}$ a podle (14)

$$\boldsymbol{\delta} = \boldsymbol{u} + \boldsymbol{U}\boldsymbol{a} . \tag{28}$$

Po dosazení do (25) a zavedení faktoru $1/L$ dostáváme

$$\begin{bmatrix} \hat{\boldsymbol{a}}_L - \boldsymbol{a} \\ \hat{\boldsymbol{b}}_L - \boldsymbol{b} \end{bmatrix} = \tag{29}$$

$$= \begin{bmatrix} \frac{1}{L}(\boldsymbol{V}^T + \boldsymbol{U}^T)(\boldsymbol{V} + \boldsymbol{U}), & -\frac{1}{L}(\boldsymbol{V}^T + \boldsymbol{U}^T)\boldsymbol{Y} \\ -\frac{1}{L}\boldsymbol{Y}^T(\boldsymbol{V} + \boldsymbol{U}), & \frac{1}{L}\boldsymbol{Y}^T\boldsymbol{Y} \end{bmatrix}^{-1} \begin{bmatrix} -\frac{1}{L}(\boldsymbol{V}^T + \boldsymbol{U}^T)(\boldsymbol{u} + \boldsymbol{U}\boldsymbol{a}) \\ \frac{1}{L}\boldsymbol{Y}^T(\boldsymbol{u} + \boldsymbol{U}\boldsymbol{a}) \end{bmatrix}.$$

Za předpokladu, že šum u a vstupní signál y jsou stacionární ergodické náhodné procesy s nulovou střední hodnotou a že šum u není korelován se vstupním signálem, platí pro $L \to \infty$ skoro jistě

$$\begin{bmatrix} \hat{\boldsymbol{a}}_\infty - \boldsymbol{a} \\ \hat{\boldsymbol{b}}_\infty - \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_{vv} + \boldsymbol{\Phi}_{vv}, & -\boldsymbol{\Phi}_{vy} \\ -\boldsymbol{\Phi}_{vy}^T, & \boldsymbol{\Phi}_{yy} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\varphi}_{uu} + \boldsymbol{\Phi}_{uu}\boldsymbol{a} \\ \boldsymbol{0} \end{bmatrix}, \tag{30}$$

kde $\boldsymbol{\Phi}_{uu}$, $\boldsymbol{\varphi}_{uu}$, $\boldsymbol{\Phi}_{vy}$, $\boldsymbol{\Phi}_{yy}$, $\boldsymbol{\Phi}_{vv}$ jsou matice sestavené z hodnot korelačních funkcí

$$\boldsymbol{\Phi}_{uu} = \begin{bmatrix} \varphi_{uu}(0) & \varphi_{uu}(1) & \dots \varphi_{uu}(n-1) \\ \varphi_{uu}(1) & \varphi_{uu}(0) & \dots \varphi_{uu}(n-2) \\ \vdots & \vdots & \vdots \\ \varphi_{uu}(n-1) & \varphi_{uu}(n-2) & \dots \varphi_{uu}(0) \end{bmatrix}, \tag{31}$$

$$\boldsymbol{\varphi}_{uu} = \begin{bmatrix} \varphi_{uu}(1) \\ \varphi_{uu}(2) \\ \vdots \\ \varphi_{uu}(n) \end{bmatrix},$$

$$\boldsymbol{\Phi}_{vy} = \begin{bmatrix} \varphi_{vy}(1) & \varphi_{vy}(0) & \dots & \varphi_{vy}(-n+1) \\ \varphi_{vy}(2) & \varphi_{vy}(1) & \dots & \varphi_{vy}(-n+2) \\ \vdots & \vdots & & \vdots \\ \varphi_{vy}(n) & \varphi_{vy}(n-1) & \dots & \varphi_{vy}(0) \end{bmatrix},$$

$$\boldsymbol{\Phi}_{yy} = \begin{bmatrix} \varphi_{yy}(0) & \varphi_{yy}(1) & \dots & \varphi_{yy}(n) \\ \varphi_{yy}(1) & \varphi_{yy}(0) & \dots & \varphi_{yy}(n-1) \\ \vdots & \vdots & & \vdots \\ \varphi_{yy}(n) & \varphi_{yy}(n-1) & \dots & \varphi_{yy}(0) \end{bmatrix},$$

$$\boldsymbol{\Phi}_{vv} = \begin{bmatrix} \varphi_{vv}(0) & \varphi_{vv}(1) & \dots & \varphi_{vv}(n-1) \\ \varphi_{vv}(1) & \varphi_{vv}(0) & \dots & \varphi_{vv}(n-2) \\ \vdots & \vdots & & \vdots \\ \varphi_{vv}(n-1) & \varphi_{vv}(n-2) & \dots & \varphi_{vv}(0) \end{bmatrix}.$$

Z výrazu (30) je patrno, že odhad koeficientů získaný metodou nejmenších čtverců je asymptoticky jednostranný, pokud rozptyl šumu u není roven nule. Systematická chyba odhadu závisí nejen na vlastnostech vstupního signálu, nýbrž i na neznámé korelační funkci šumu a dosti složitým způsobem (přes korelační funkce φ_{vy} a φ_{vv}) i na dynamických vlastnostech soustavy. Nelze proto udat jednoduchou korekci odhadu. Tato korekce může být získána pouze iterací a jednu takovou možnost uvádí Clarke [9].

Metodu pro odhad koeficientů diferenční rovnice z konečného počtu naměřených hodnot vstupního a výstupního signálu podal též Levin [10]. Jeho přístup vyžaduje však znalost kovarianční matice šumu a má nedostatky, které byly kritizovány McBridem [11].

Podle experimentů provedených v ÚTIA ČSAV dává velmi dobré výsledky pro soustavy nízkého řádu metoda Aströma a Bohlina [12, 13, 14]. Je založena na maximální věrohodnosti a odhadují se při ní současně jak koeficienty soustavy, tak i statistické charakteristiky šumu. Předpokládá se, že šum je gaussovský a že jeho spektrální hustota může být vyjádřena racionální funkcí lomenou. Metoda je rovněž iterační.* Je třeba souhlasit s autory metody, že určení koeficientů, které udávají statistické vlastnosti šumu, je nedílnou součástí identifikace soustavy, pokud má být korekční člen stanoven tak, aby optimálně kompenzoval šum existující v soustavě. Na druhé straně autoru tohoto článku není zřejmé, zda je rozumné určovat koeficienty soustavy současně s koeficienty, které charakterizují šum. Je možné postupovat

* Metoda konverguje spolehlivě pouze pro řád soustavy $n \leqq 2$.

též tak, že se nejprve identifikuje samotná soustava a teprve potom – je-li to vzhledem k účelu identifikace třeba – se odhadnou statistické charakteristiky šumu. Jako určitou nevýhodu metody Aströma a Bohlina lze uvést, že se při ní často určuje více koeficientů, než je nezbytně třeba.

V následujícím odstavci popíšeme přístup, který umožňuje konstrukci odhadu koeficientů diferenční rovnice nezávisle na statistických charakteristikách šumu. Přitom odhad je lineární, tj. lze ho nalézt řešením systému lineárních rovnic. Tím odpadají potíže s lokálními extrémy a s konvergencí iterací, které se vyskytují u výše zmíněných metod. Předností navrhovaného postupu je i to, že v případě bez šumu dostáváme přesné výsledky i při minimální délce pozorování.

3. NOVÝ PŘÍSTUP

V reálných podmínkách lze při identifikaci průmyslových soustav předpokládat jen minimální znalosti o statistických vlastnostech šumu u (obr. 1). Náš přístup založíme na předpokladu, že střední hodnota šumu je v uvažovaném intervalu pozorování nezávislá na předchozích hodnotách vstupního signálu a rovna nule

$$E\{u(k)|y(k), y(k-1), y(k-2), \ldots\} = E\{u(k)\} = \bar{u} = 0 \tag{32}$$

pro $k = 1, 2, 3, \ldots, L$.

Naším cílem je konstrukce lineárního a konsistentního odhadu koeficientů diferenční rovnice soustavy. Princip přístupu si však vysvětlíme nejprve na jednodušším případě, kdy se dynamické vlastnosti vyšetřované soustavy popisují konečným počtem pořadnic impulsní charakteristiky.

3.1 Odhad pořadnic impulsní charakteristiky

Při konstrukci odhadu neznámých parametrů soustavy použijeme následujícího postupu, který sestává ze tří kroků:

1. S použitím hypotézy o ergodičnosti šumu sestavíme odhad pro tu statistickou charakteristiku, jejíž znalost se předpokládá.

V případě, že hledáme pořadnice impulsní charakteristiky soustavy, bude výhodné vyjít z odhadu koeficientů lineárního regresního modelu

$$E\{u(k) \mid y(k), y(k-1), \ldots, y(k-N+1)\} = \sum_{i=0}^{N-1} r_i y(k-i)\,, \tag{33}$$

přičemž N je dostatečně velké číslo, o jehož volbě se zmíníme později. r_i jsou koeficienty regrese.

Za předpokladu, že náhodný šum má ergodické vlastnosti, můžeme odhad regresních koeficientů r_i nalézt minimalizací ztrátové funkce

$$Q = \sum_{k=1}^{L} q^2(k), \tag{34}$$

kde

$$q(k) = u(k) - \sum_{i=0}^{N-1} r_i y(k-i), \quad k = 1, 2, \ldots, L. \tag{35}$$

Zavedeme-li matice

$$\boldsymbol{q} = \begin{bmatrix} q(1) \\ q(2) \\ \vdots \\ q(L) \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} u(1) \\ u(2) \\ \vdots \\ u(L) \end{bmatrix}, \quad \boldsymbol{r} = \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_{N-1} \end{bmatrix}, \tag{36}$$

$$\boldsymbol{Y}_N = \begin{bmatrix} y(1) & y(0) & y(-1) & \ldots & y(-N+2) \\ y(2) & y(1) & y(0) & \ldots & y(-N+3) \\ y(3) & y(2) & y(1) & \ldots & y(-N+4) \\ \vdots & \vdots & \vdots & & \vdots \\ y(L) & y(L-1) & y(L-N+1) & \ldots & y(L-N+1) \end{bmatrix},$$

můžeme systém rovnic (35) zapsat následovně

$$\boldsymbol{q} = \boldsymbol{u} - \boldsymbol{Y}_N \boldsymbol{r}, \tag{37}$$

a (34) pak bude

$$Q = \boldsymbol{q}^{\mathrm{T}} \boldsymbol{q}. \tag{38}$$

Z podmínky pro extrém

$$\frac{\partial Q}{\partial \boldsymbol{r}} = 2 \frac{\partial \boldsymbol{q}^{\mathrm{T}}}{\partial \boldsymbol{r}} \boldsymbol{q} = -2 \boldsymbol{Y}_N^{\mathrm{T}} (\boldsymbol{u} - \boldsymbol{Y}_N \boldsymbol{r})\big|_{\boldsymbol{r} = \hat{\boldsymbol{r}}} = \boldsymbol{0} \tag{39}$$

dostaneme odhad regresních koeficientů.

2. Výraz pro odhad vektoru regresních koeficientů obsahuje neznámé diskrétní hodnoty šumu u. Vyjádříme je pomocí naměřených hodnot vstupu y a výstupu x a pomocí hledaných pořadnic impulsní charakteristiky g

$$u(k) = x(k) - \sum_{i=0}^{\lambda-1} g(i)\, y(k-i), \quad k = 1, 2, \ldots, L. \tag{40}$$

Použijeme-li označení (7) a (8) z odstavce 2.1, můžeme místo (40) psát

$$\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{Y}_\lambda \boldsymbol{g} \tag{41}$$

a po dosazení do (39) dostáváme

$$\hat{\boldsymbol{r}} = (\boldsymbol{Y}_N^{\mathrm{T}}\boldsymbol{Y}_N)^{-1}\,\boldsymbol{Y}_N^{\mathrm{T}}\,(\boldsymbol{x} - \boldsymbol{Y}_\lambda \boldsymbol{g})\,. \tag{42}$$

3. Skutečné hodnoty regresních koeficientů $r_i (i = 0, 1, \ldots, N-1)$ jsou známé a podle předpokladu (32) jsou rovny nule. Protože odhad těchto regresních koeficientů (39) resp. (42) je nestranný, je možno předpokládat, že experiment dopadl tak, že odhad je „co nejbližší" skutečným hodnotám $\boldsymbol{r} = \boldsymbol{0}$. Význam slovního obratu „co nejbližší" je třeba blíže specifikovat. Podle toho, jaký je počet odhadovaných pořadnic impulsní charakteristiky $g(i)$, $i = 0, 1, \ldots, \lambda - 1$ vzhledem k počtu N regresních koeficientů r_i, $i = 0, 1, \ldots, N-1$, mohou nastat tři případy:

a) Nejjednodušší a nejvýznamnější je případ, kdy $\lambda = N$. Potom $\boldsymbol{Y}_\lambda = \boldsymbol{Y}_N$ a jako nejpravděpodobnější výsledek pokusu lze jednoduše volit

$$\hat{\boldsymbol{r}} = \boldsymbol{r} = \boldsymbol{0}\,. \tag{43}$$

Dosadíme-li do této podmínky vztah (42), dostaneme pro pořadnice impulsní charakteristiky odhad

$$\hat{\boldsymbol{g}} = (\boldsymbol{Y}_\lambda^{\mathrm{T}}\boldsymbol{Y}_\lambda)^{-1}\,\boldsymbol{Y}_\lambda^{\mathrm{T}}\boldsymbol{x} \tag{44}$$

shodný s odhadem (9), který jsme jiným postupem získali již dříve.

b) Je-li $\lambda > N$, dostáváme se do situace, že máme více neznámých, než je počet podmínek, které je třeba splnit, a vztah (42) nelze použít pro určení odhadů hodnot impulsní charakteristiky g.

Při identifikaci dynamických soustav – až na zcela výjimečné případy není obvykle a priori přesně známa „skutečná délka" impulsní charakteristiky λ. Je proto rozumné volit ji maximální možnou, tj. $\lambda = N$ nebo naopak odhadovat potřebné N podle předpokládané délky impulsní charakteristiky. Největší praktický význam má tedy případ a). Přesto s ohledem na další uvažujme i případ, kdy počet nenulových pořadnic impulsních charakteristik je předem znám a $\lambda < N$.

c) Je-li $\lambda < N$, nelze volbou $\hat{g}(i)$, $i = 0, 1, \ldots, \lambda - 1$ splnit podmínku (43) beze zbytku a požadavku, že odhady $\hat{\boldsymbol{r}}$ mají být co nejbližší skutečným hodnotám $\boldsymbol{r}$, je třeba dát kvantitativní formu. Lze k tomu např. užít vážené čtverce

$$S = (\hat{\boldsymbol{r}} - \boldsymbol{r})^{\mathrm{T}}\,\boldsymbol{W}(\hat{\boldsymbol{r}} - \boldsymbol{r})\,, \tag{45}$$

kde $\boldsymbol{W}$ je matice váhových koeficientů. Odhady $\hat{\boldsymbol{r}}$ můžeme pak najít minimalizací (45). Blíže se budeme tímto případem zabývat v následujícím odstavci při odhadu koeficientů diferenční rovnice.

Přístup popsaný v předcházejícím odstavci použijeme nyní pro konstrukci odhadu koeficientů diferenční rovnice. Vyjdeme přitom z rovnice (14)

$$x(k) + \sum_{i=1}^{n} a_i x(k-i) - \sum_{i=0}^{n} b_i y(k-i) = \delta(k)\,, \tag{46}$$

kde v souhlase s (15) náhodná chyba rovnice je

$$\delta(k) = u(k) + \sum_{i=1}^{n} a_i u(k-i)\,. \tag{47}$$

O statistických vlastnostech posloupnosti náhodných veličin $\delta(k)$ nevíme nic víc ani nic méně než o statistických vlastnostech šumu $u(k)$. Je-li podle předpokladu (32) střední hodnota šumu nezávislá na předcházejících hodnotách vstupního signálu y, bude totéž platit o náhodné chybě rovnice

$$E\{\delta(k) \mid y(k), y(k-1), \ldots\} = E\{\delta(k)\} = \left(1 + \sum_{i=1}^{n} a_i\right) \bar{u} = d\,. \tag{48}$$

Předpokládejme pro jednoduchost, že střední hodnota šumu a tím i konstanta d na pravé straně vztahu (48) je nula.*

Konstrukci odhadu neznámých koeficientů rozdělíme – podobně jako v odstavci 3.1 – do třech kroků:

1. V prvém kroku sestavíme odhady pro koeficienty regresního modelu

$$E\{\delta(k) \mid y(k), y(k-1), \ldots, y(k-N+1)\} = \sum_{i=0}^{N-1} \varrho_i y(k-i)\,. \tag{49}$$

Použijeme-li označení (36), můžeme tyto odhady nalézt minimalizací výrazu

$$Q = \boldsymbol{q}^{\mathrm{T}} \boldsymbol{q}\,, \tag{50}$$

kde v tomto případě

$$\boldsymbol{q} = \boldsymbol{\delta} - \boldsymbol{Y}_N \boldsymbol{\varrho}\,. \tag{51}$$

Vektor $\boldsymbol{\delta}$ je definován posledním z výrazů (17) a platí pro něj vztahy (16) resp. (20).

Minimalizací kvadratické ztrátové funkce (50) dostaneme

$$\boldsymbol{Y}_N^{\mathrm{T}} \boldsymbol{\delta} - \boldsymbol{Y}_N^{\mathrm{T}} \boldsymbol{Y}_N \hat{\boldsymbol{\varrho}} = \boldsymbol{0}\,. \tag{52}$$

* Tento předpoklad není pro další postup podstatný a bez velkých potíží by bylo možno považovat d za další neznámý parametr [15].

Za předpokladu, že průběh vstupního signálu splňuje požadavek, aby existovala inversní matice

$$\boldsymbol{R} = (\boldsymbol{Y}_N^{\mathrm{T}}\boldsymbol{Y}_N)^{-1}, \tag{53}$$

můžeme z (52) vypočíst

$$\hat{\boldsymbol{\varrho}} = \boldsymbol{R}\boldsymbol{Y}_N^{\mathrm{T}}\boldsymbol{\delta}. \tag{54}$$

2. V rovnici (54) vyjádříme neznámý vektor $\boldsymbol{\delta}$ pomocí naměřených hodnot vstupního a výstupního signálu a hledaných koeficientů. Použijeme k tomu vztahu (20)

$$\hat{\boldsymbol{\varrho}} = \boldsymbol{R}\boldsymbol{Y}_N^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{Z}\boldsymbol{\vartheta}). \tag{55}$$

3. Podle předpokladu (32) resp. (51) jsou skutečné regresní koeficienty ϱ_i rovny nule. Pokud je počet odhadovaných koeficientů stejný jako počet odhadů regresních koeficientů $\hat{\varrho}_i$ tj. $N = 2n + 1$, jsme oprávněni předpokládat, že výsledek experimentu byl takový, že

$$\hat{\boldsymbol{\varrho}} = \boldsymbol{\varrho} = \boldsymbol{0}; \tag{56}$$

z této podmínky můžeme určit $\hat{\vartheta}_L$.

Pokud $N > 2n + 1$, lze volbou ϑ požadavek (56) splnit pouze v určitém přiblížení. Jako míru tohoto přiblížení zvolíme vážený součet čtverců

$$S = \hat{\boldsymbol{\varrho}}^{\mathrm{T}}\boldsymbol{W}\hat{\boldsymbol{\varrho}} \tag{57}$$

a jako odhad $\hat{\vartheta}_L$ vezmeme hodnotu, která tento vážený součet minimalizuje. Jako výsledek této minimalizace obdržíme konečně

$$\hat{\vartheta}_L = (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Y}_N\boldsymbol{P}\boldsymbol{Y}_N^{\mathrm{T}}\boldsymbol{Z})^{-1}\,\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Y}_N\boldsymbol{P}\boldsymbol{Y}_N^{\mathrm{T}}\boldsymbol{x}, \tag{58}$$

kde

$$\boldsymbol{P} = \boldsymbol{R}\boldsymbol{W}\boldsymbol{R} = (\boldsymbol{Y}_N^{\mathrm{T}}\boldsymbol{Y}_N)^{-1}\,\boldsymbol{W}(\boldsymbol{Y}_N^{\mathrm{T}}\boldsymbol{Y}_N)^{-1}. \tag{59}$$

To je hledaný odhad koeficientů diferenční rovnice soustavy zapsaný v maticové symbolice. Obdélníkové matice $\boldsymbol{Z}$, $\boldsymbol{Y}_N$ a $\boldsymbol{x}$ mají L řádků a toto číslo může být velmi vysoké, např. stovky nebo tisíce. Vyčíslení odhadu $\hat{\vartheta}_L$ podle vzorců (58) a (59) klasickými metodami maticové algebry by bylo velmi zdlouhavé a kladlo by příliš vysoké nároky na paměť použitého číslicového počítače. Proto mají vzorce (58) a (59) jen malou praktickou cenu pokud není znám algoritmus na jejich vyčíslení, který

nemá uvedené nedostatky. Takový algoritmus byl skutečně nalezen, prakticky vyzkoušen a bude podrobně popsán v jiném článku [15].

3.3 Asymptotické vlastnosti odhadu

V tomto odstavci ukážeme, že odhad (58) je silně konsistentní, tj. že platí

$$P\{\lim_{L\to\infty}(\hat{\vartheta}_L - \vartheta) = \mathbf{0}\} = 1\,. \tag{60}$$

Dosadíme-li do (58) z (20)

$$\boldsymbol{x} = \boldsymbol{\delta} + \boldsymbol{Z}\vartheta\,, \tag{61}$$

nalezneme, že

$$\hat{\vartheta}_L - \vartheta = (\boldsymbol{Z}^T\boldsymbol{Y}_N\boldsymbol{P}\boldsymbol{Y}_N^T\boldsymbol{Z})^{-1}\,\boldsymbol{Z}^T\boldsymbol{Y}_N\boldsymbol{P}\boldsymbol{Y}_N^T\boldsymbol{\delta}\,. \tag{62}$$

Označíme-li

$$\boldsymbol{\Phi}_L = \frac{1}{L}\,\boldsymbol{Y}_N^T\boldsymbol{Z}\,, \tag{63}$$

$$\boldsymbol{\varphi}_L = \frac{1}{L}\,\boldsymbol{Y}_N^T\boldsymbol{\delta}\,, \tag{64}$$

$$\boldsymbol{\psi}_L = \frac{1}{L}\,\boldsymbol{Y}_N^T\boldsymbol{Y}_N\,, \tag{65}$$

bude zřejmě

$$\hat{\vartheta}_L - \vartheta = \left(\boldsymbol{\Phi}_N^T\frac{1}{L^2}\,\boldsymbol{P}\boldsymbol{\Phi}_L\right)^{-1}\boldsymbol{\Phi}_N^T\frac{1}{L^2}\,\boldsymbol{P}\boldsymbol{\varphi}_L\,, \tag{66}$$

a podle (59)

$$\frac{1}{L^2}\,\boldsymbol{P} = \boldsymbol{\Psi}_L\boldsymbol{W}\boldsymbol{\Psi}_L\,. \tag{67}$$

Za předpokladu, že vstupní signál je „trvale vybuzující" [12], tj. že existují limity

$$\bar{y} = \lim_{L\to\infty}\frac{1}{L}\sum_{k=1}^{L} y(k)\,,\quad \varphi_{yy}(i) = \lim_{L\to\infty}\frac{1}{L}\sum_{k=1}^{L} y(k)\,y(k+i)\,,$$

přejdou v limitě pro $L \to \infty$ prvky matic $\boldsymbol{\Phi}_\infty$, $\boldsymbol{\varphi}_\infty$, $\boldsymbol{\psi}_\infty$, skoro jistě v hodnoty korelačních funkcí

$$\boldsymbol{\Phi}_\infty = \begin{bmatrix} -\varphi_{yx}(-1) & -\varphi_{yx}(-2) & \dots & -\varphi_{yx}(-n) \\ -\varphi_{yx}(0) & -\varphi_{yx}(-1) & \dots & -\varphi_{yx}(-n+1) \\ \vdots & \vdots & & \vdots \\ -\varphi_{yx}(N-2) & -\varphi_{yx}(N-3) & \dots & -\varphi_{yx}(N-n-1) \end{bmatrix} \tag{68}$$

$$\begin{bmatrix} \varphi_{yy}(0) & \varphi_{yy}(1) & \dots \varphi_{yy}(n) \\ \varphi_{yy}(1) & \varphi_{yy}(0) & \dots \varphi_{yy}(n-1) \\ \vdots & \vdots & \vdots \\ \varphi_{yy}(N-1) & \varphi_{yy}(N-2) & \dots \varphi_{yy}(0) \end{bmatrix},$$

$$\boldsymbol{\varphi}_\infty = \begin{bmatrix} \varphi_{y\delta}(0) \\ \varphi_{y\delta}(1) \\ \vdots \\ \varphi_{y\delta}(N-1) \end{bmatrix}, \tag{69}$$

$$\boldsymbol{\Psi}_\infty = \begin{bmatrix} \varphi_{yy}(0) & \varphi_{yy}(1) & \dots \varphi_{yy}(N-1) \\ \varphi_{yy}(1) & \varphi_{yy}(0) & \dots \varphi_{yy}(N-2) \\ \vdots & \vdots & \vdots \\ \varphi_{yy}(N-1) & \varphi_{yy}(N-2) & \dots \varphi_{yy}(0) \end{bmatrix}. \tag{70}$$

Podle předpokladu není šum u a tudíž ani veličina δ korelován se vstupním signálem. Platí tedy že, $\varphi_\infty = \boldsymbol{0}$ a dosazením do (66) pro $L \to \infty$ plyne (60).

(Došlo dne 1. března 1968.)

LITERATURA

[1] Identification in Automatic Control Systems. Preprints of IFAC Symposium, Prague June 1967. Academia, Prague 1967.

[2] Levin M. J.: Optimum Estimation of Impulse Response in the Presence of Noise. IRE Trans. on Circuit Th. *CT-7* (March 1960), 1.

[3] Eykhoff P., Van der Grinten P. M. E. M., Kwakernaak H., Veltman B. P. Th.: Systems Modeling and Identification. Survey paper, Third Congress IFAC, London, June 1966.

[4] Eykhoff P.: Process Parameter and State Estimation. V [1].

[5] Briggs P. A. N., Godfrey K. R., Hammond P. H.: Estimation of Process Dynamic Characteristics by Correlation Methods using Pseudo—Random Signals. V [1].

[6] Izawa K., Furuta K.: Process Identification Using Correlation Coefficient. V [1].

[7] Krtolica R.: Some Applications of Pseudorandom Maximum Length Binary Sequences in the Statistical Identification. V [1].

[8] Fořtová A., Kryže J., Peterka V.: Numerické řešení Wienerovy-Hopfovy rovnice při statistické identifikaci lineární dynamické soustavy. Kybernetika *2* (1966), 4, 331—346.

[9] Clarke D. W.: Generalised-Least-Squares Estimation of the Parameters of a Dynamic Model. V [1].

[10] Levin M. J.: Estimation of System Pulse Transfer Function in the Presence of Noise. IEE Trans. on Aut. Control *AC-9* (July 1964), 3, 229–235.
[11] Discussion of „Estimation of a System Pulse Transfer Function in the Presence of Noise". IEE Trans. on Aut. Control *AC-10* (April 1965).
[12] Aström K. J., Bohlin T.: Numerical Identification of Linear Dynamic Systems from Normal Operating Records. Proc. IFAC Symposium on Theory of Self-Adaptive Control Systems, Teddington Sept. 1965.
[13] Aström K. J., Bohlin T., Wensmark S.: Automatic Construction of Linear Stochastic Dynamic Models for Stationary Industrial Processes with Random Disturbances Using Operating Records. IBM Nordic Laboratory, Sweden, June 1965 (Report TP 18.150).
[14] Vališ J.: Experimentální ověření Aströmovy-Bohlinovy metody numerické identifikace lineárních dynamických soustav z provozních záznamů vstupů a výstupů. Zpráva ÚTIA č. 210.
[15] Peterka V., Šmuk K.: On-Line Estimation of Dynamic Model Parameters from Experimental Data. Referát pro IV. kongres IFAC, Varšava 1969.

SUMMARY

A New Approach to the Identification of Discrete Dynamic Systems

VÁCLAV PETERKA

In the paper the present state of the problem of dynamic systems identification is analysed and on the basis of this analysis a new approach to the solution of the given problem is proposed. The approach is applied to the estimation of the parameters of a single variable discrete dynamic system from experimental input-output data. The advantages of the estimates are:

a) linearity, i.e. the estimates of the unknown coefficients are obtained as a solution of a system of linear equations (no difficulties with local extrema and convergence of iterations),

b) only the minimal prior knowledge of the statistical characteristics of the noise is required,

c) when the noise is not present exact results are obtained even for a minimal length of observation.

The solution is given in a general matrix form and it is proved that the estimates are strongly consistent.

Ing. Václav Peterka, CSc., Ústav teorie informace a automatizace ČSAV, Vyšehradská 49, Praha 2.