



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Daniel Schwarz

# Lineární a adaptivní zpracování dat

Leden 2012



Příprava a vydání této publikace byly podporovány projektem ESF č. CZ.1.07/2.2.00/07.0318 „Víceborová inovace studia Matematické biologie“ a státním rozpočtem České republiky.

## Předmluva

Učební text *Lineární a adaptivní zpracování dat* vznikl v souvislosti s řešením projektu ESF č. CZ.1.07/2.2.00/07.0318 „Víceoborová inovace studia matematické biologie“. Cíle této publikace jsou dva. Prvním cílem je ukázat studentům matematické biologie a dalším laskavým čtenářům analýzu a zpracování dat z pohledu teorie lineárních systémů, které zde vystupují jednak v roli generátorů dat a dále také v roli prostředků pro analýzu a zpracování dat. Druhým cílem je pak ukázat, jak lze efektivně řešit vybrané úlohy z této oblasti pomocí výpočetních nástrojů v prostředí MATLAB.

Data, o jejichž analýzu a zpracování se ve všech kapitolách tohoto učebního textu usiluje, mají povahu jednorozměrných diskrétních signálů a pro účely nejvíce možného přiblížení se studentům inženýrských disciplín jsou zde tato data označována jako časové řady. Toto pojmenování však nemění nic na skutečnosti, že většina z uvedených metod a postupů je využitelná také jako základ pro zpracování a analýzu obecných diskrétních signálů a posloupností, jako jsou obrazová, prostorová nebo sekvenční data. Zcela záměrně je zde vynechána problematika zpracování a analýzy v čase nebo prostoru spojitých dat včetně jejich vzorkování – zájemci o tuto oblast necht' využijí jiné dostupné literární prameny. Systémy nebo také filtry, o nichž je napříč tímto textem zhusta vedena řeč, si tak může student – díky zaměření výhradně na diskrétní případy – představit jako algoritmy realizované pomocí programů na počítači.

Témata, kterým se tento učební text věnuje, jsou strukturována tak, že čtenáři na samém začátku neujdou základní vlastnosti systémů, aby následně rozeznal lineární časově invariantní systémy a způsoby jejich popisu v časové i frekvenční doméně. Ve druhé kapitole je pak představen základní koncept lineární filtrace, jsou objasněny rozdíly mezi různými přístupy ke klasifikaci typů filtrů a jsou také ukázány některé návrhové metody. Třetí kapitola se věnuje kumulačním technikám pro zvýrazňování užitečné složky dat a potlačení složky rušivé. Někdy bývají tyto techniky označovány jako metody průměrování nebo obecněji metody pro zlepšování poměru signálu k šumu. Následující dvě kapitoly jsou pak věnovány modelování vzniku náhodných časových řad a jejich predikce, a to nejdříve s využitím teorie podle statistiků George Boxe a Gwilyma Jenkinse a v poslední kapitole pak pomocí základních technik optimální a adaptivní filtrace. Na konci každé kapitoly jsou kromě řešených úloh také stručně shrnuty klíčové znalosti, které by si měl čtenář při studiu tohoto textu osvojit. Kéž je jich dost a jsou čtenářům k užítku.

Na tomto místě bych rád poděkoval prof. Ing. Jiřímu Holčíkovi, CSc. za cenné rady a postřehy ke mně vyslané během práce na učebním textu a dále oběma recenzentům prof. Ing. Ivo Provazníkovi, Ph.D. a Ing. Zoltánovi Szabó, Ph.D. za důkladné pročtení, odstranění produktů autorské slepoty a podnětné připomínky.

V Brně 15. ledna 2012

Daniel Schwarz

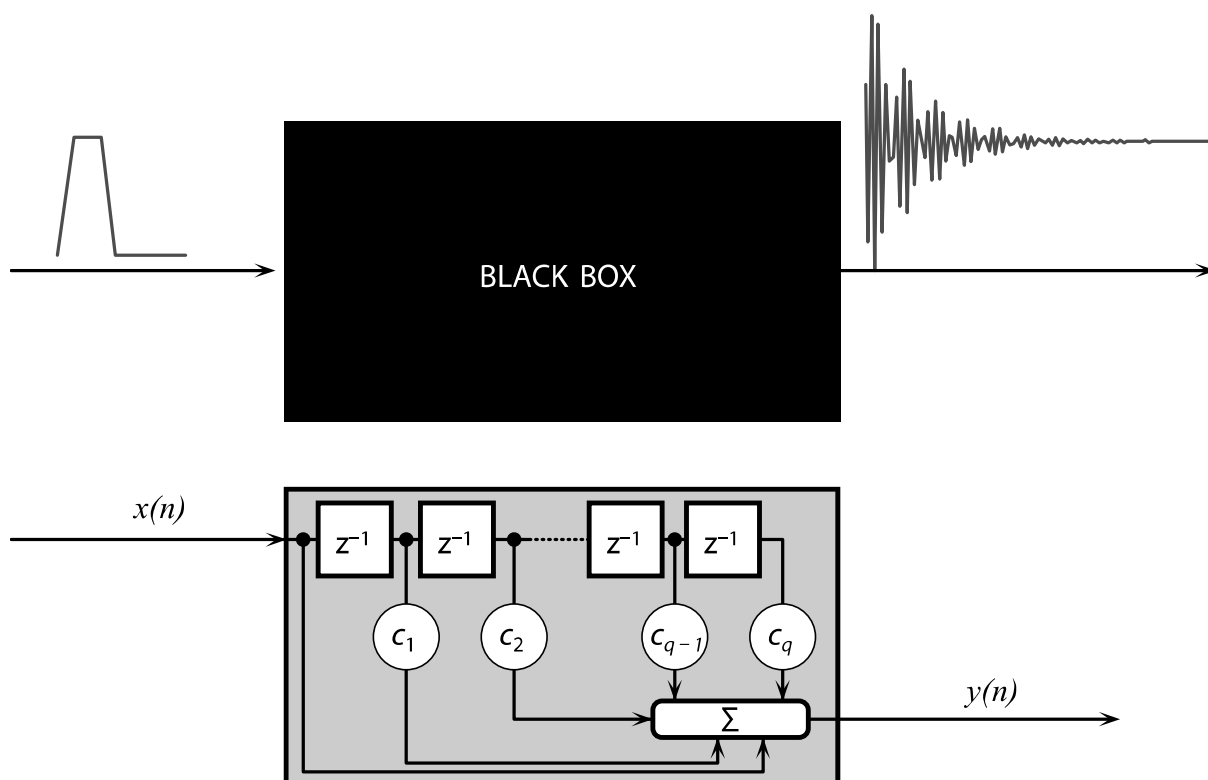
© Daniel Schwarz, 2012

ISBN 978-80-7204-779-6

# 1 Systémy, signály a časové řady

Za systém se obvykle považuje soustava entit a jejich vzájemné vztahy. V reálném světě jsou systémy příliš složité, a ty živé obzvlášť, na to, aby je bylo možné jednoznačně popsat nebo analyzovat v celé jejich komplexnosti. Využívají se proto často matematické modely, které vznikají abstrakcí reálných systémů. Model by měl být jednoduchý a měl by vykazovat takové chování, jaké je pozorováno u systému, jenž je předmětem zájmu. Abstrakcí je myšleno vymezení důležitých veličin reálného systému a jejich vzájemných vztahů, přičemž ostatní veličiny tvoří okolí systému a mohou být ignorovány, či mohou ovlivňovat vstupy systému nebo mohou být ovlivňovány jeho výstupy.

Systém mnohdy vysílá spontánně či vynuceně různé signály a mění svoje stavy – mňouká, štěká, přibírá na váze, stárne, umírá, nebo v případě populace třeba i vymírá či se nestabilně množí. Pokud pozorovatel tyto signály a stavy měří, tj. neznámý systém pozoruje, pak vlastně získává o neznámém systému experimentální data, viz obr. 1.1.



**Obr. 1.1** Identifikace systémů je disciplína zabývající se tvorbou matematických modelů systémů. Identifikace začíná pozorováním či měřením modelovaného systému – jsou tak získána experimentální data, která se pak využívají pro identifikaci struktury a pro odhad parametrů modelu. Data mohou být mj. reprezentována časovými řadami neboli 1-D diskrétními signály. Systémy nebo procesy mohou signály vysílat jednak spontánně nebo jako odezvu na buzení vstupním signálem. Popisem systému se vysvětlují souvislosti mezi signály či časovými řadami na výstupu a na vstupu. Na ilustračním schématu je znázorněn diskrétní systém, který je realizován pomocí zpožďovacích členů  $z^{-1}$ , dále pomocí operací násobení koeficienty  $c_{1..q}$  a také pomocí operace sčítání. Systém zpracovává vstupní časovou řadu  $x(n)$  a na výstup vrací časovou řadu  $y(n)$ .

Pro pojem signál existuje v literatuře celá řada definic. Pro účely tohoto textu bude podle [10] signál definován jako jev fyzikální, chemické, biologické, ekonomické či jiné materiální povahy, který nese nehmotnou informaci o stavu a dynamice systému či procesu, který jej generuje. Je-li zdrojem informace živý organismus, pak hovoříme o biosignálech, a to bez ohledu na podstatu nosiče informace. Z pohledu matematika je možno se dívat na signál jako na funkci v čase nebo v prostoru proměnných a měřitelných veličin.

Signály se dělí podle nezávislých veličin na spojité nebo diskrétní. Rozlišují se také podle počtu dimenzí nezávislých veličin na jednorozměrné 1-D, dvourozměrné 2-D, třírozměrné 3-D, čtyřrozměrné 4-D a vícerozměrné  $N$ -D. U vícerozměrných signálů jde většinou o statické nebo dynamické obrazy definované pomocí dvou až čtyř časoprostorových souřadnic, zatímco jednorozměrné signály zpravidla reprezentují časové průběhy veličin. V případě jednorozměrných diskrétních signálů s ekvidistantním vzorkováním se pak hovoří o časových řadách.

Identifikací systému se rozumí sestavení modelu reálného systému a odhadnutí jeho parametrů tak, aby chování modelu odpovídalo tomu, co je v reálném světě pozorováno. Častou aplikací je konstrukce modelu za účelem předpovědi budoucího chování systému, který je modelován. Systémy nebo jejich modely se často popisují soustavami diferenciálních nebo diferenčních rovnic. V teorii systémů a signálů se velmi často používají také neparаметrické způsoby popisu, a to prostřednictvím charakteristik v originální časové či prostorové doméně, nebo po vhodných transformacích ve frekvenční doméně.

Z hlediska zpracování a analýzy časových řad je možno definovat systém jako jakoukoli sadu procesů, které náhodně nebo systematicky ovlivňují povahu časové řady, viz příklad na obr. 1.2.

## 1.1 Vlastnosti systémů

Existují určité důležité vlastnosti systémů, které má smysl vždy vyšetřovat, neboť zjištění, že systém oplývá danou vlastností, může usnadnit jeho modelování, identifikaci či popis. U systémů tedy sledujeme následující vlastnosti.

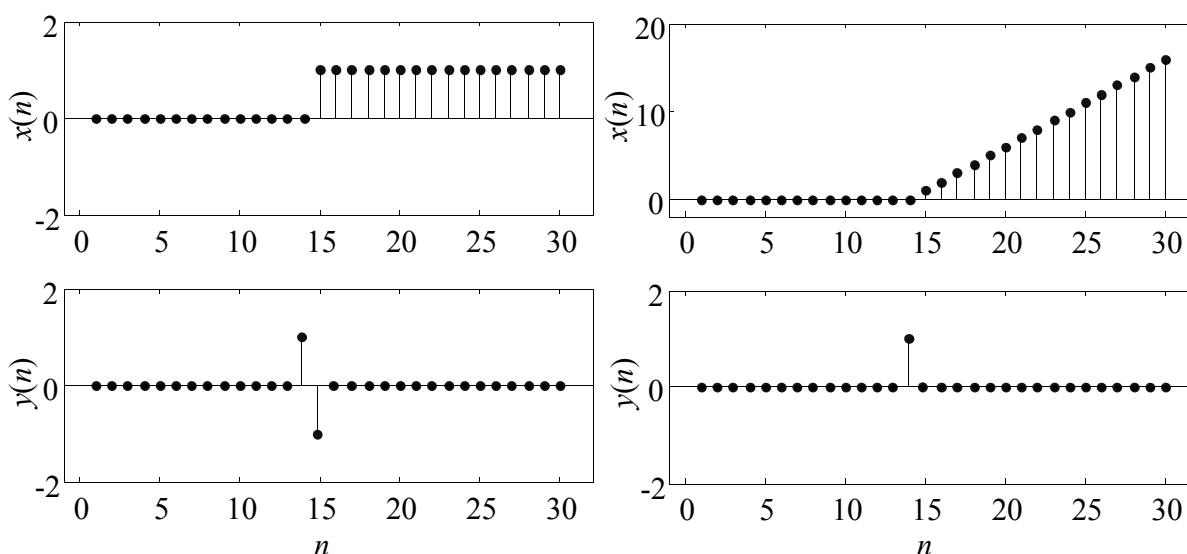
- **Kauzální / nekauzální.** Systém je kauzální, pokud jeho výstup závisí pouze na minulých a současných vstupních hodnotách – v jazyce českém se hovoří také o příčinnosti a následku. Kauzalitu nevyšetřujeme u systémů s prostorově závislými proměnnými – např. zpracování 2-D obrazu probíhá v daném systému s použitím vzorků obrazu ve všech směrech souřadného systému. Naproti tomu v reálném čase jsou všechny systémy, které má smysl uvažovat, kauzální, neboť „čas běží pouze dopředu“ – např. představa nekauzálního finančního systému, jehož výkonnost závisí na zítřejší hodnotě cenných papírů, je nesmyslná. Z uvedených příkladů se zdá, že se potencionální nekauzalitou není ani třeba zabývat, ovšem při analýze systémů na základě časových řad se často využívá nahraných dat do paměti, a zavádí se tak přirozeně nekauzální operace, jako je např. výpočet difference časové řady, viz příklad systému pro detekci bodů zlomu na obr. 1.2.
- **Časově invariantní / časově proměnné.** Systém je časově invariantní, pokud jeho odezva  $y(t)$  na vstupní časovou řadu  $x(t)$  je vždy stejná bez závislosti na prodlevách mezi buzeními. Jinými slovy je odezva časově invariantního systému na posunutou v čase  $x(t-t_0)$  odezvou v čase stejně posunutou:  $y(t-t_0)$ . Biologické systémy nebo procesy nejsou ovšem kvůli své adaptabilitě či stárnutí časově invariantní. Použití časově invariantního systému pro zpracování časových řad generovaných biologickými procesy je možné jen tehdy, lze-li v omezeném časovém intervalu, který je dostatečný pro danou aplikaci, zavést omezující předpoklady. Robustnějším řešením pro zpracování

časových řad generovaných z biologických procesů je nasazení systémů adaptivních, tj. časově proměnných, o kterých je v této učebnici skriptech rovněž pojednáno.

- **Lineární / nelineární.** Lineární systémy jsou popsány lineárními diferenciálními nebo diferenčními rovnicemi. Vzhledem k tomu, že takové systémy lze poměrně snadno řešit, je často snahou vyšetřovat reálné nelineární systémy v určitých oblastech jako systémy lineární, a to pokud lze dosáhnout dostatečné přesnosti takovéto aproximace. Lineární systém je takový systém, ve kterém lze uplatnit princip superpozice. Ten se matematicky popisuje takto:

$$\begin{aligned} x_k(n) &\rightarrow y_k(n); \\ \sum_k a_k x_k(n) &\rightarrow \sum_k a_k y_k(n), \end{aligned} \quad (1.1)$$

což je rovnice vyjadřující odezvu na buzení pomocí kombinace  $k$  časových řad  $x_k(n)$ , která odpovídá kombinaci výstupních časových řad  $y_k(n)$  získaných při individuálním, nezávislém působení oněch  $k$  vstupních časových řad na lineární systém. Symbolem  $n$  je v rovnicích vyjádřen bezrozměrný diskretní čas.



**Obr. 1.2** Příklad systému pracujícího s diskretními 1-D signály neboli časovými řadami. Vstupní časové řady  $x(n)$  jsou zpracovány hranovým detektorem tak, že lze výstupní časové řady  $y(n)$  použít pro detekci bodu zlomu v původních časových řadách. Hranový detektor lze popsat následující diferenciální rovnicí:  $y(n) = x(n+1) - 2x(n) + x(n-1)$ . Jedná se tedy o nekauzální systém.

## 1.2 Lineární časově invariantní systémy, konvoluce

Lineární časově invariantní (LTI – *linear time invariant*) systémy disponují matematickými vztahy mezi jejich vstupními a výstupními signály či časovými řadami. Pomocí nich lze určit výstupní odezvu LTI systému na jakýkoli vstup a lze také určit vstup systému při pozorování jeho výstupu. Operaci, která určuje vztah mezi časovými řadami na

vstupu a výstupu LTI systému, se říká konvoluce. Pro její odvození pomocí základních matematických operací je třeba přijmout následující poučku, kterou lze selským rozumem vyvodit z vlastností LTI systémů.

**PRVNÍ POUČKA ODVOZENÁ SELSKÝM ROZUMEM:**

**„Známe-li odezvu LTI systému na velmi krátkou časovou řadu, můžeme pomocí těchto velmi krátkých řad seskládat libovolně dlouhou časovou řadu, a odezvu LTI systému na ni pak seskládat pomocí známé odezvy na velmi krátkou řadu.“**



Nejkratší časovou řadou, kterou je možno pro tento účel použít, je jednotkový impuls  $\delta(n)$ , kterému ve spojitě časové doméně odpovídá Diracova distribuce. Kombinací posunutých jednotkových impulsů lze sestavit libovolnou časovou řadu, viz obr. 1.3, což je možno také vyjádřit následující rovnicí:

$$\begin{aligned} x(n) &= \dots + x(-1)\delta(n+1) + x(0)\delta(n) + x(1)\delta(n-1) + x(2)\delta(n-2) \dots = \\ &= \sum_{k=-\infty}^{\infty} x(k)\delta(n-k). \end{aligned} \quad (1.2)$$

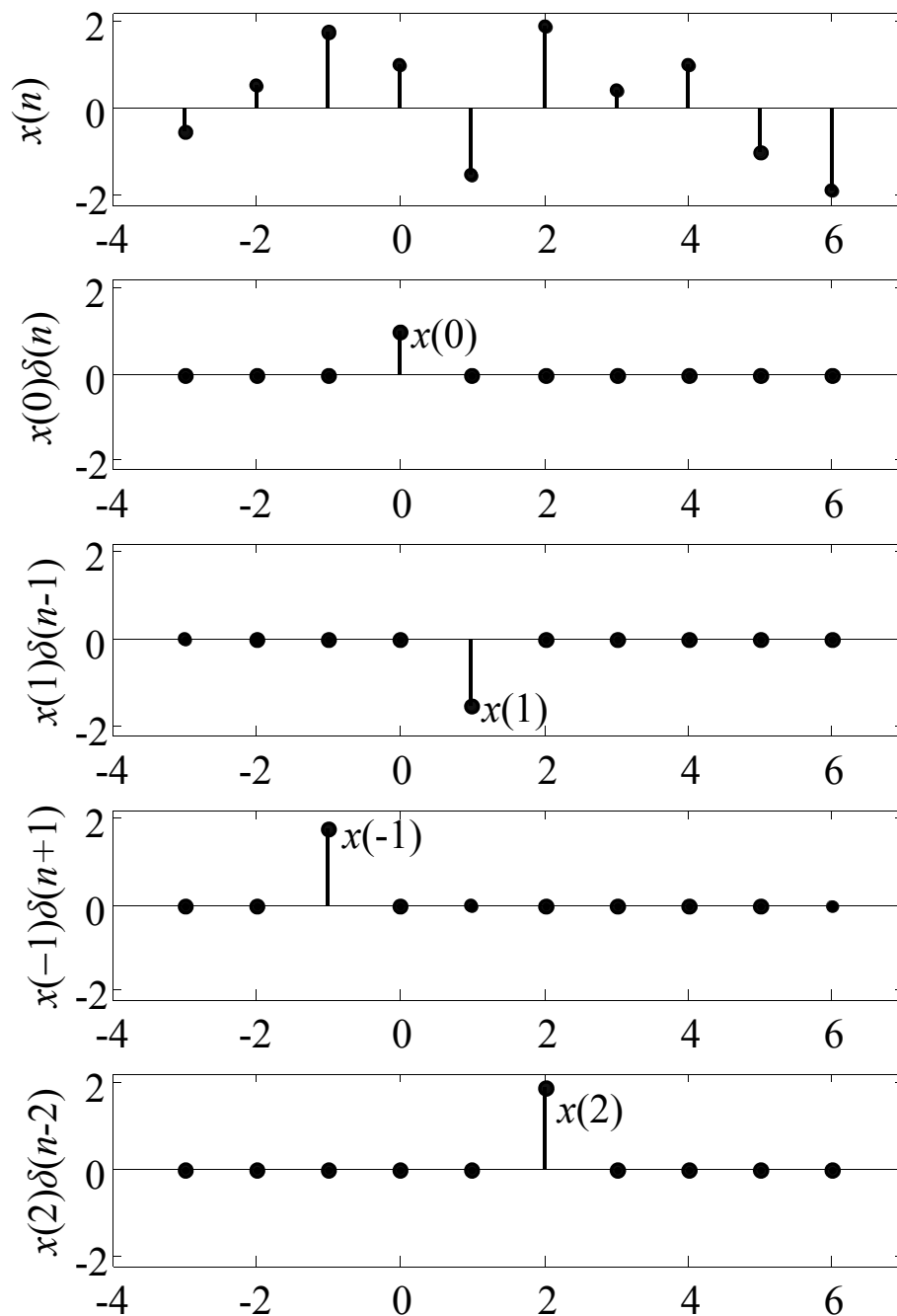
Za předpokladu, že odezvou LTI systému na jednotkový impuls  $\delta(n-k)$  je posloupnost  $h_k(n)$ , lze při aplikaci výše uvedené poučky, odvozené selským rozumem, vyjádřit odezvu LTI systému na časovou řadu  $x(n)$  takto:

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)\delta(n-k) \rightarrow y(n) = \sum_{k=-\infty}^{\infty} x(k)h_k(n). \quad (1.3)$$

Protože pro časově invariantní systém můžeme jistě říci, že jeho odezvy na časově posunuté jednotkové impulsy budou mít všechny stejný průběh a budou se lišit pouze právě časovým posunem. Můžeme tedy psát, že:

$$\begin{aligned} \delta(n) &\rightarrow h(n) \Rightarrow \delta(n-k) \rightarrow h(n-k) \\ x(n) &= \sum_{k=-\infty}^{\infty} x(k)\delta(n-k) \rightarrow y(n) = x(n) * h(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k). \end{aligned} \quad (1.4)$$

Poslední výraz v rovnici (1.4) se označuje jako konvoluční suma a operátor konvoluce  $*$  představuje matematický způsob výpočtu odezvy LTI systému  $y(n)$  na libovolnou časovou řadu  $x(n)$  při znalosti odezvy  $h(n)$  tohoto systému na jednotkový impuls  $\delta(n)$ . Tato odezva  $h(n)$  se nazývá **impulsní charakteristika** systému a představuje jeden ze základních způsobů neparametrického popisu systému v časové nebo prostorové doméně.



**Obr. 1.3** Sestrojení časové řady  $x(n)$  pomocí jednotkových impulsů  $\delta(n)$ .

### 1.3 Lineární časově invariantní systémy a periodické signály, Fourierova řada

Při zkoumání odezvy LTI systému na jednotkový impuls byl v předchozí podkapitole odvozen operátor konvoluce a definován pojem impulsní charakteristika. Tato podkapitola se zabývá odezvou LTI systému na harmonickou časovou řadu, kterou lze vyjádřit v goniometrickém a exponenciálním tvaru následujícími funkcemi:

$$\begin{aligned} f(n) &= A \cdot \sin(\omega_f n + \varphi), \\ f(n) &= A \cdot e^{j(\omega_f n + \varphi)}, \end{aligned} \quad (1.5)$$

kde  $A$  je amplituda signálu,  $\omega_f$  je normalizovaná úhlová frekvence<sup>1</sup> a  $\varphi$  je počáteční fázový posuv. Průběh harmonické časové řady si lze představit jako rotující fázor v komplexní rovině, přičemž rychlost jeho rotace je dána právě normalizovanou úhlovou frekvencí  $\omega_f = 2\pi/N_f$ , kde  $N_f$  představuje počet vzorků harmonické časové řady v její jedné periodě. Harmonická časová řada je nekonečně dlouhá a ve frekvenční oblasti je její amplitudové spektrum reprezentováno v jedné půlperiodě čarou na jediné frekvenci, viz obr. 1.4.

Zatímco v rovnicích (1.2) až (1.4) byla libovolná časová řada vyjádřena pomocí kombinace jednotkových impulsů, je v následující rovnici (1.6) provedeno vyjádření libovolné periodické časové řady pomocí harmonických složek.

$$x(n) = \sum_{k=\langle N \rangle} a_k e^{jk\omega_0 n} \rightarrow y(n) = \sum_{k=\langle N \rangle} H(e^{jk\omega_0}) a_k e^{jk\omega_0 n} \quad (1.6)$$

Takovýto rozklad se označuje jako Fourierova řada a kromě jejího exponenciálního tvaru (1.6) je Fourierova řada známa v matematické analýze také ve tvaru goniometrickém, kde jde o rozklad periodické funkce do kombinace goniometrických funkcí. V rovnici (1.6) je vyjádřena libovolná časová řada  $x(n)$  jako lineární kombinace harmonických časových řad, přičemž úhlová frekvence jednotlivých harmonických složek je vyjádřena jako  $k$ -násobek základní harmonické složky s úhlovou frekvencí  $\omega_0 = 2\pi/N$ . Počet vzorků  $N$  v jedné periodě časové řady  $x(n)$  zároveň určuje maximální počet harmonických složek ve Fourierově řadě. Je zřejmé, že Fourierova řada z diskrétního signálu je řada s konečným počtem koeficientů  $a_k$ . Je-li směs harmonických složek  $x(n)$  přivedena na vstup LTI systému, pak jeho výstup je podle rovnice (1.6) opět směsí harmonických složek se stejnými úhlovými frekvencemi. Jednotlivé harmonické složky jsou násobeny komplexními čísly  $H(e^{jk\omega_0})$ , čímž dojde buď k zesílení, nebo k zeslabení dané složky a dále může vlivem tohoto násobení dojít také k posunu fáze dané složky. Na základě rovnice (1.6) lze tedy usuzovat, že LTI systém nevytváří nové frekvenční složky, ale pouze zesiluje nebo potlačuje frekvenční komponenty již existující ve zpracovávané časové řadě.

Zesilování, zeslabování a fázový posuv harmonických složek zpracovávaných časových řad lze pro systém vyjádřit pomocí **frekvenční charakteristiky**  $G(\omega)$ , která je dána jako:

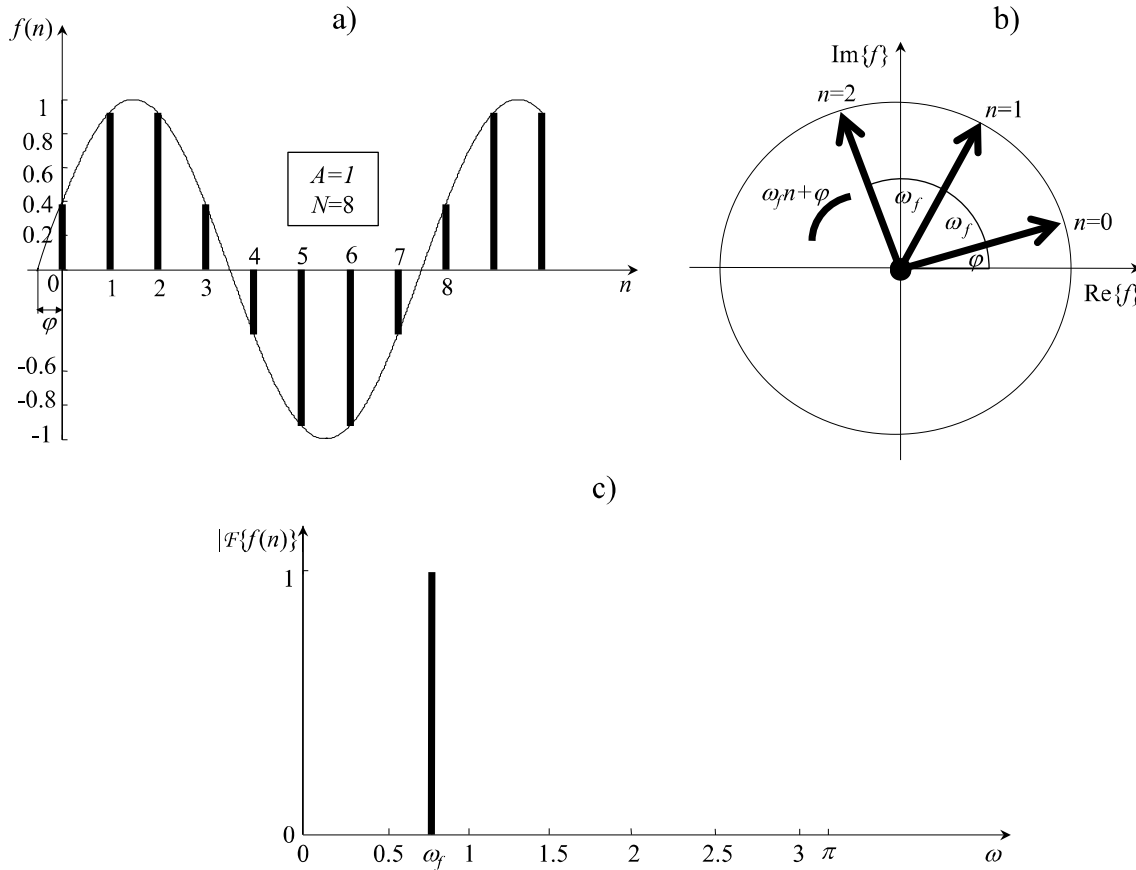
---

<sup>1</sup> Normalizovaná úhlová frekvence  $\omega$  je mírou rychlosti změny v časové řadě. Zatímco běžná frekvence udává rychlost změny pomocí počtu cyklů za jednotku času (1 Hertz = jeden cyklus za sekundu), úhlová frekvence udává rychlost změny pomocí počtu radiánů za jednotku času. Normalizovaná frekvence  $f$  udává rychlost změny pomocí počtů cyklů za vzorek a v případě normalizované úhlové frekvence  $\omega = 2\pi f$  jde o počet radiánů za vzorek. Výpočet normalizované frekvence (cykly za vzorek) se provádí prostým vydělením frekvence udávané v Hertzích (cykly za sekundu) vzorkovací frekvencí (vzorky za sekundu). Normalizovaná frekvence  $f$  a normalizovaná úhlová frekvence  $\omega$  jsou tedy bezrozměrné veličiny a jejich hodnoty jsou v intervalu  $f \in \langle 0; 1 \rangle$ , respektive  $\omega \in \langle 0; 2\pi \rangle$ . Vzhledem k zaměření tohoto učebního textu výhradně na časové řady a nikoli na spojité signály, jsou zde rychlosti změny vyjadřované pouze pomocí normalizovaných bezrozměrných frekvencí a čas  $n$  zde vystupuje pouze ve své diskrétní podobě - rovněž bezrozměrně.



$$G(\omega) = H(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h(n)e^{-j\omega n}. \quad (1.7)$$

Z rovnice (1.7) lze usuzovat, že frekvenční charakteristika diskrétního systému je spojitou periodickou funkcí úhlové frekvence  $\omega$ . Tato funkce je dána Fourierovou řadou s koeficienty, které odpovídají vzorkům impulsní charakteristiky  $h(n)$  popisovaného systému.



**Obr. 1.4** Grafická znázornění harmonické časové řady a veličin, které určují její průběh.

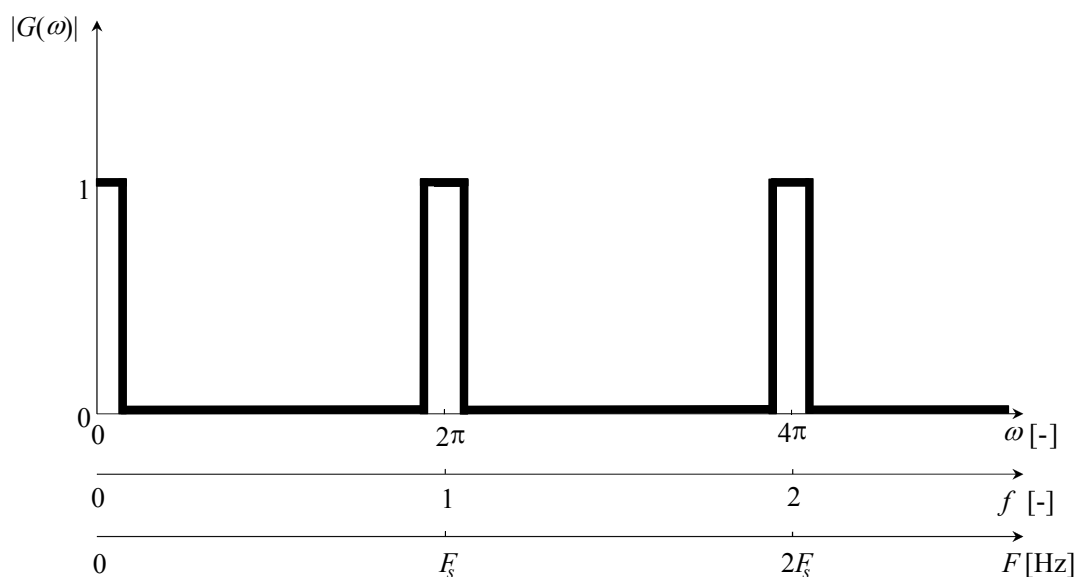
a) Harmonická časová řada. b) Rotující fázor v komplexní rovině. c) První půlperioada amplitudového frekvenčního spektra harmonické časové řady.

Perioda frekvenční charakteristiky s nezávislou bezrozměrnou proměnnou  $\omega$  je  $2\pi$ , pro bezrozměrnou normalizovanou frekvenci  $f$  je perioda 1 a pro nezávislou proměnnou  $F$ , kterou je frekvence v Hertzích, odpovídá perioda frekvenční charakteristiky hodnotě vzorkovací frekvence signálu  $F_s$ , který je systémem zpracováván, viz příklad idealizované dolní propusti na obr. 1.5.

Způsob, jakým systém preferenčně zesiluje či zeslabuje určité frekvenční komponenty zpracovávané časové řady a jakým způsobem tyto složky fázově zpožďuje, lze jistě považovat za důležitou součást popisu systému. Frekvenční charakteristika tedy vlastně tvoří popis systému ve frekvenční oblasti.

Pro doplnění je vhodné dodat, že Fourierova řada diskrétní posloupnosti je totéž co Fourierova transformace s diskrétním časem (DTFT – *Discrete Time Fourier Transform*) této

posloupnosti, a je tedy možné říci, že frekvenční charakteristika systému odpovídá impulsní charakteristice tohoto systému transformované do frekvenční domény pomocí DTFT.



**Obr. 1.5** Frekvenční charakteristika systému – modulová část. Jedná se zřejmě o systém, který představuje ideální dolní propust. Systém propouští harmonické komponenty v dolní části frekvenčního spektra (zesílení je 1) a utlumuje frekvenční komponenty v horní části frekvenčního spektra (zesílení je 0) – viz také obr. 2.2. Frekvenční charakteristika je symetrická okolo normalizované úhlové frekvence  $\pi$  a periodická s periodou  $2\pi$ . Obrázek zároveň ukazuje vztah mezi normalizovanou úhlovou frekvencí  $\omega$  [-], normalizovanou frekvencí  $f = \omega/2\pi = F/F_s$  [-] a mezi frekvencí  $F$  [Hz].

Fourierova transformace s diskretním časem DTFT by se však neměla zaměňovat nebo plést s diskretní Fourierovou transformací (DFT – *Discrete Fourier Transform*). Zatímco DTFT produkuje z diskretních posloupností spojité periodické funkce frekvence, viz (1.8), výsledkem DFT je posloupnost konečného počtu vzorků DTFT, viz definice  $N$ -bodové DFT (1.9).

$$DTFT\{h(n)\} = H(\omega) = \sum_{n=-\infty}^{\infty} h(n)e^{-j\omega n}, \quad (1.8)$$

$$DFT\{h(n)\} = H(\omega_k) = \sum_{n=0}^{N-1} h(n) \cdot e^{-\frac{j2\pi kn}{N}}, \quad \omega_k = \frac{k}{N}, \quad k = 0, 1, 2, \dots, N-1. \quad (1.9)$$

## 1.4 Z-transformace, přenosová funkce diskretního systému

Transformace  $Z$  je důležitým matematickým nástrojem pro reprezentaci a manipulaci s diskretními posloupnostmi<sup>2</sup>. Můžeme ji považovat za zveřejnění Fourierovy transformace pro diskretní systémy a signály.

Obraz diskretní posloupnosti v  $Z$ -transformaci je dán mocninou řadou:

$$\mathcal{Z}\{x(n)\} = X(z) = \sum_{n=-\infty}^{\infty} x_n z^{-n}, \quad (1.10)$$

přičemž nejčastěji uvažujeme tzv. jednostrannou  $Z$ -transformaci pro nezáporné indexy  $n$ . Jak ukazuje rovnice (1.8), není samoučelné se  $Z$ -transformací zabývat, neboť obraz konvoluce vykazuje zajímavé vlastnosti:

$$\begin{aligned} \mathcal{Z}\{y(n)\} &= \mathcal{Z}\{x(n) * h(n)\} = \mathcal{Z}\left\{\sum_{k=-\infty}^{\infty} x(k) \cdot h(n-k)\right\} = \\ &= \sum_{n=-\infty}^{\infty} \left[ \sum_{k=-\infty}^{\infty} x(k) \cdot h(n-k) \right] \cdot z^{-n} = \sum_{k=-\infty}^{\infty} x(k) \cdot \left[ \sum_{n=-\infty}^{\infty} h(n-k) \right] \cdot z^{-n}, \end{aligned} \quad (1.11)$$

*subst* :  $m = n - k$ ,

$$\mathcal{Z}\{y(n)\} = \sum_{k=-\infty}^{\infty} x(k) \cdot \left[ \sum_{m=-\infty}^{\infty} h(m) \right] \cdot z^{-m} = X(z) \cdot H(z).$$

Z rovnice (1.11) vyplývá, že násobení obrazů v  $Z$ -transformaci odpovídá konvoluce původních posloupností v časové nebo prostorové doméně. Tato vlastnost  $Z$ -transformace se označuje jako **konvoluční teorém**<sup>3</sup> a obraz impulsní charakteristiky diskretního systému  $H(z)$  se označuje jako **přenosová funkce** diskretního systému:

$$\begin{aligned} H(z) &= \frac{Y(z)}{X(z)} = \mathcal{Z}\{h(n)\} = \sum_{n=0}^{\infty} h(n) \cdot z^{-n}, \\ \text{subst : } z &= e^{j\omega}, \\ H(z) &= H(e^{j\omega}) = G(\omega). \end{aligned} \quad (1.12)$$

Z rovnice (1.10) vyplývá, že přenosová funkce vyjadřuje na jednotkové kružnici  $|z|=1$  frekvenční charakteristiku diskretního systému, což může lépe osvětlit periodicitu frekvenční charakteristiky, a to zejména při srovnání obr. 1.5 a obr. 1.6.

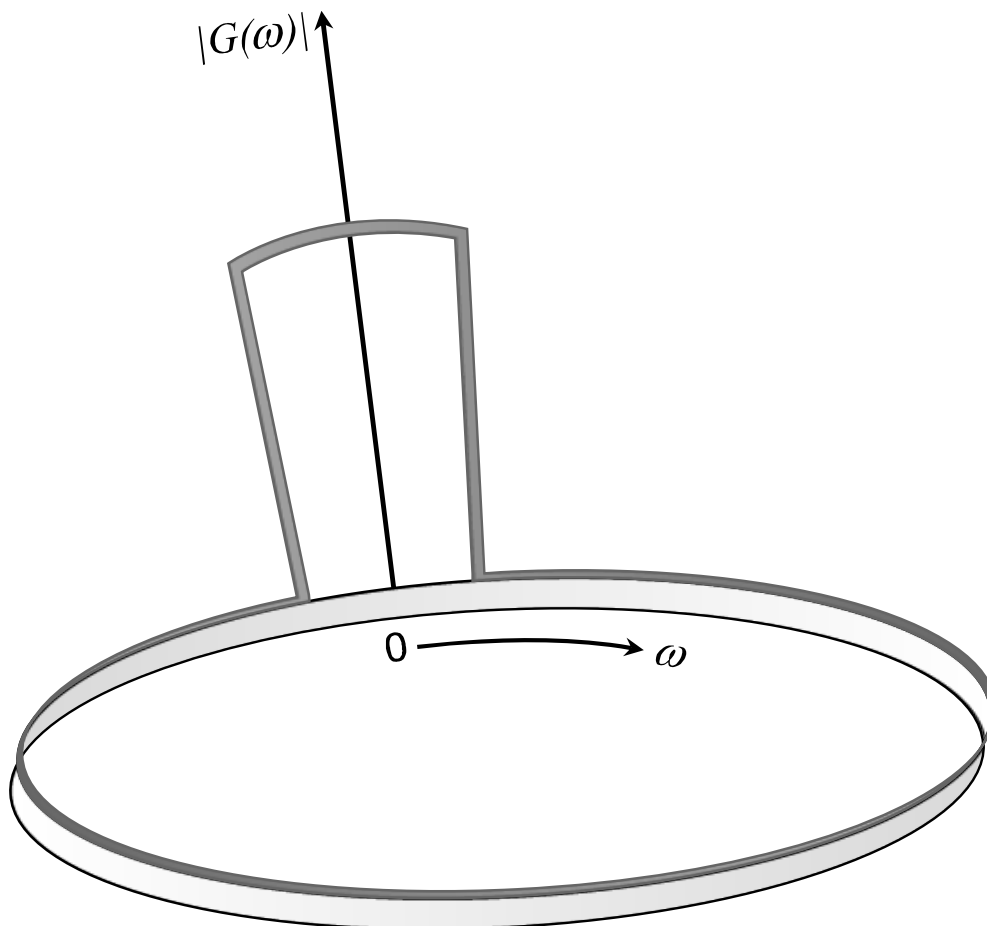
Pokud se vyjádří přenosová funkce ve tvaru racionálně lomené funkce a provedou se úpravy obou polynomů v čitateli a ve jmenovateli tak, aby se jednalo o polynomy s kladnými mocninami  $z$ , je pak možné vyjádřit čítec a jmenovatel přenosové funkce  $H(z)$  pomocí součinu kořenových činitelů:

<sup>2</sup>  $Z$ -transformace nemá nic společného s tzv.  $z$ -score transformací, která se aplikuje ve statistické matematice pro převod libovolné normální distribuce na standardní normální distribuci.

<sup>3</sup> Konvoluční teorém platí také pro speciální případy  $Z$  transformace, a sice pro DTFT a DFT.

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^M b_i \cdot z^{-i}}{\sum_{i=0}^L a_i \cdot z^{-i}} = A \cdot z^{L-M} \cdot \frac{\prod_{i=1}^M (z - n_i)}{\prod_{i=1}^L (z - p_i)}. \quad (1.13)$$

Kořeny polynomu v čitateli se označují jako **nulové body** přenosové funkce a kořeny polynomu ve jmenovateli se označují jako **póly** přenosové funkce.



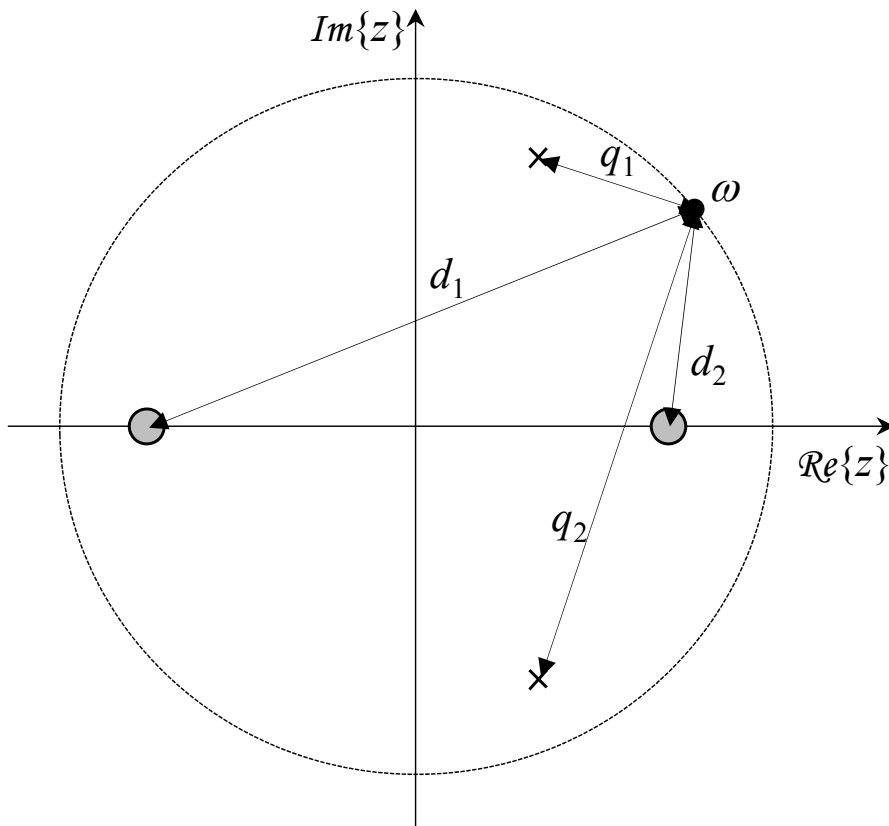
**Obr. 1.6** Přenosová funkce na jednotkové kružnici  $|z|=1$  vyjadřuje frekvenční charakteristiku systému – zde se jedná o idealizovanou dolní propust z obr. 1.5.

Rozložení nul a pólů v rovině  $z$  je dalším typem popisu diskrétního systému. Využívá se zejména v situacích, kdy je cílem u systému nebo jeho modelu vyšetřit jeho **stabilitu**<sup>4</sup>. Jedno ze dvou kritérií stability stanovuje, že diskrétní systém je stabilní, pokud jsou všechny póly jeho přenosové funkce umístěny uvnitř jednotkové kružnice. Toto kritérium se označuje také jako kritérium stability v obrazové doméně. Druhé kritérium stability diskrétního systému je definováno v časové doméně. Podle něj je systém stabilní, pokud součet absolutních hodnot vzorků jeho impulsní charakteristiky je konečné číslo.

Přímou souvislost konfigurace nulových bodů a pólů přenosové funkce systému s frekvenční charakteristikou systému ukazují následující rovnice a obr. 1.7.

<sup>4</sup> Stabilitu systému je možno vysvětlit jako jeho tendenci přiměřeně reagovat na trvající podnět a vracet se do výchozího stavu po té, co podnět zanikne.

$$\begin{aligned}
G(\omega) &= H(e^{j\omega}) = A \cdot e^{j\omega(L-M)} \cdot \frac{\prod_{i=1}^M (e^{j\omega} - n_i)}{\prod_{i=1}^L (e^{j\omega} - p_i)}, \\
|G(\omega)| &= |A| \cdot |e^{j\omega(L-M)}| \cdot \frac{\prod_{i=1}^M |e^{j\omega} - n_i|}{\prod_{i=1}^L |e^{j\omega} - p_i|} = |A| \cdot 1 \cdot \frac{d_1(\omega) \cdot d_2(\omega) \dots d_M(\omega)}{q_1(\omega) \cdot q_2(\omega) \dots q_L(\omega)}, \\
\arg(G(\omega)) &= \sum_{i=1}^M \arg(e^{j\omega} - n_i) - \sum_{i=1}^L \arg(e^{j\omega} - p_i) + \arg(A) + (L-M) \cdot \omega
\end{aligned} \tag{1.14}$$



**Obr. 1.7** Rozložení nulových bodů ○ a pólů × přenosové funkce systému v rovině  $z$  může být využito pro sestavení odhadu průběhu frekvenční charakteristiky, viz rovnice (1.14).

Průběh modulové frekvenční charakteristiky lze z konfigurace nulových bodů a pólů přenosové funkce odhadnout tak, že se pro každou úhlovou frekvenci  $\omega$ , které odpovídá bod na jednotkové kružnici  $|z|=1$ , určí vzdálenosti tohoto bodu od všech nulových bodů  $d_{1..M}$  a dále vzdálenosti tohoto bodu od všech pólů  $q_{1..L}$ . Podíl součinů těchto vzdáleností, viz rovnice (1.14), pak určuje hodnotu modulové frekvenční charakteristiky pro danou úhlovou frekvenci  $\omega$ . Obdobně se postupuje také u odhadu průběhu fázové frekvenční charakteristiky, přičemž se zde pracuje nikoli se vzdálenostmi, ale s úhly, které svírají průvodce vyšetřovaných bodů s reálnou osou.

V úvodu této kapitoly bylo uvedeno, že jedním ze základních způsobů popisu diskrétního systému je diferenční rovnice. Souvislost diferenční rovnice s popisem systému pomocí přenosové funkce sice není na první pohled zřejmá, ale pokud roznásobíme zlomky v rovnici (1.13):

$$Y(z) \cdot \sum_{i=0}^L a_i \cdot z^{-i} = X(z) \cdot \sum_{i=0}^M b_i \cdot z^{-i}, \quad (1.15)$$

koeficienty  $a_i$  a  $b_i$  se normalizují tak, aby  $a_0=1$ , a aplikuje se inverzní z-transformace včetně vět o linearitě a posunu (zde bez důkazu, podrobně viz např. [12]), získá se diferenční rovnice systému:

$$y(n) = \sum_{i=0}^M b_i \cdot x(n-i) - \sum_{i=1}^L a_i \cdot y(n-i). \quad (1.16)$$

## 1.5 Úlohy



**ÚLOHA 1.1** Vyšetřete kauzalitu, linearitu a časovou invariantnost systémů uvedených v tabulce 1.1.

*Tab. 1.1 Příkladové rovnice různých systémů a jejich vlastnosti.*

Rovnice systému	Kauzální?	Časově invariantní?	Lineární?
$y(n) = x^2(n-2)$	ANO	ANO	NE
$y(n) = x(n+1)$	NE	ANO	ANO
$y(n) = \left(\frac{1}{2}\right)^{n+1} x^3(n-19)$	ANO	NE	NE
$y(n) = n \cdot x(2n)$	NE	NE	ANO



**ÚLOHA 1.2** Určete diferenční rovnici systému s přenosovou funkcí:

$$H(z) = 0,9z / (z-0,9)$$

Po dlouhém vydělení polynomu polynomem dojdeme k výrazu se zápornými mocninami  $z$  a inverzní Z transformací získáme požadovanou diferenční rovnici.

$$H(z) = \frac{0,9z}{z-0,9} = 0,9 + 0,81z^{-1} + 0,729z^{-2} + \dots = \sum_{i=0}^{\infty} 0,9^{i+1} z^{-i} = \frac{Y(z)}{X(z)}$$

$$Y(z) = 0,9X(z) + 0,81z^{-1}X(z) + 0,729z^{-2}X(z) + \dots$$

$$y(n) = 0,9x(n) + 0,81x(n-1) + 0,729x(n-2) + \dots = \sum_{i=0}^{\infty} 0,9^{i+1} x(n-i)$$

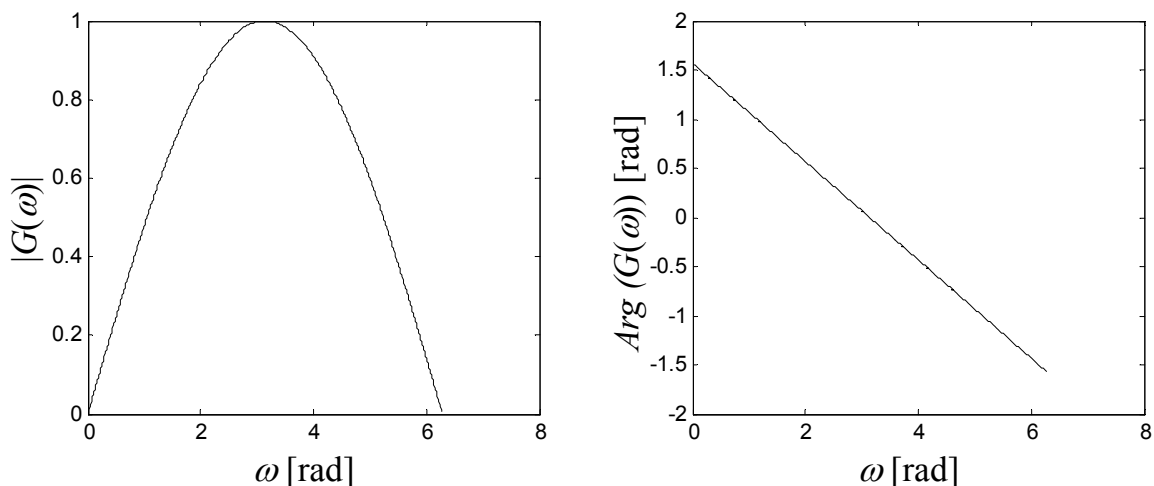
Jiný způsob řešení, který nevyžaduje obtížné dělení polynomu polynomem:

$$\begin{aligned}
 H(z) &= \frac{0,9z}{z-0,9} = \frac{Y(z)}{X(z)} \\
 Y(z)(z-0,9) &= 0,9X(z)z \quad \Big| \cdot z^{-1} \\
 Y(z)(1-0,9z^{-1}) &= 0,9X(z) \\
 y(n)-0,9y(n-1) &= 0,9x(n)
 \end{aligned}$$

**ÚLOHA 1.3** Vypočtete a nakreslete frekvenční charakteristiku jednoduchého systému, který počítá difference dvou posledních vzorků časové řady a má rovnici:  
 $y(n)=0,5x(n)-0,5x(n-1)$

Frekvenční charakteristiku je možno vypočítat např. DTFT impulsní charakteristiky a využít Eulerových vztahů mezi komplexními exponenciálami a goniometrickými funkcemi. Amplitudovou frekvenční charakteristiku ukazuje obr. 1.8.

$$\begin{aligned}
 h(n) &= \frac{1}{2}\{+1, -1\} \\
 G(\omega) &= DTFT\{h(n)\} = \frac{1}{2}e^{-j\omega \cdot 0} - \frac{1}{2}e^{-j\omega} = \frac{1}{2} - \frac{1}{2}e^{-j\omega} = \\
 &= \frac{je^{-j\frac{\omega}{2}}}{2j} \left( e^{j\frac{\omega}{2}} - e^{-j\frac{\omega}{2}} \right) = je^{-j\frac{\omega}{2}} \sin\left(\frac{\omega}{2}\right)
 \end{aligned}$$



**Obr. 1.8** Úloha 1.3 – Modulová (vlevo) a fázová (vpravo) frekvenční charakteristika systému, který počítá difference z posledních dvou vzorků časové řady na vstupu.

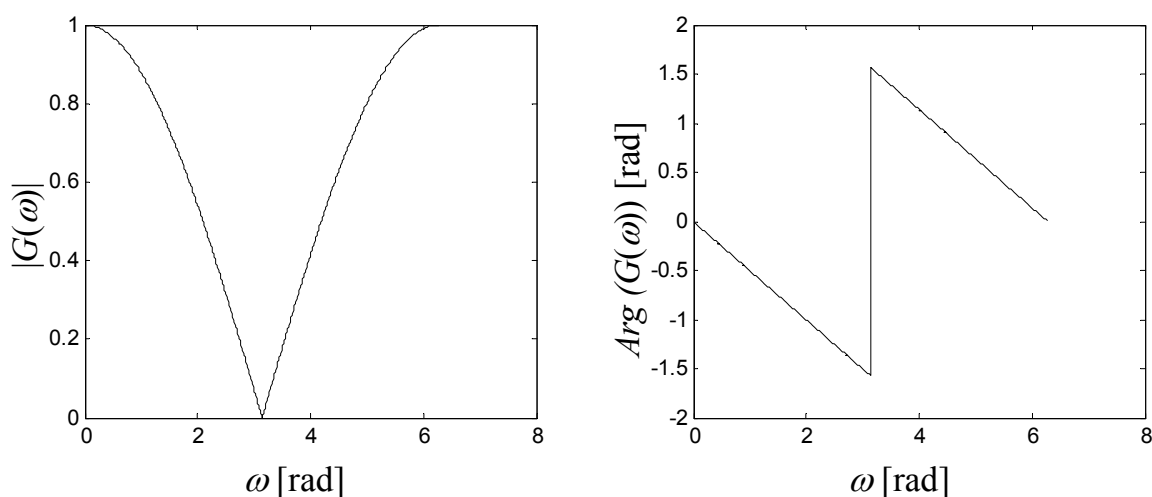
**ÚLOHA 1.4** Vypočtete a nakreslete frekvenční charakteristiku jednoduchého systému, který počítá průměr dvou posledních vzorků časové řady a má rovnici:  
 $y(n)=0,5x(n)+0,5x(n-1)$

Frekvenční charakteristiku je možno vypočítat např. DTFT impulsní charakteristiky a využít Eulerových vztahů mezi komplexními exponenciálami a goniometrickými funkcemi. Amplitudovou frekvenční charakteristiku ukazuje obr. 1.9.

$$h(n) = \frac{1}{2} \{+1, +1\}$$

$$G(\omega) = DTFT\{h(n)\} = \frac{1}{2}e^{-j\omega \cdot 0} + \frac{1}{2}e^{-j\omega} = \frac{1}{2} + \frac{1}{2}e^{-j\omega} =$$

$$= \frac{e^{-j\frac{\omega}{2}}}{2} \left( e^{j\frac{\omega}{2}} + e^{-j\frac{\omega}{2}} \right) = e^{-j\frac{\omega}{2}} \cos\left(\frac{\omega}{2}\right)$$



**Obr. 1.9** Úloha 1.4 – Modulová (vlevo) a fázová (vpravo) frekvenční charakteristika systému, který počítá průměry z posledních dvou vzorků časové řady na vstupu.

## 1.6 Shrnutí

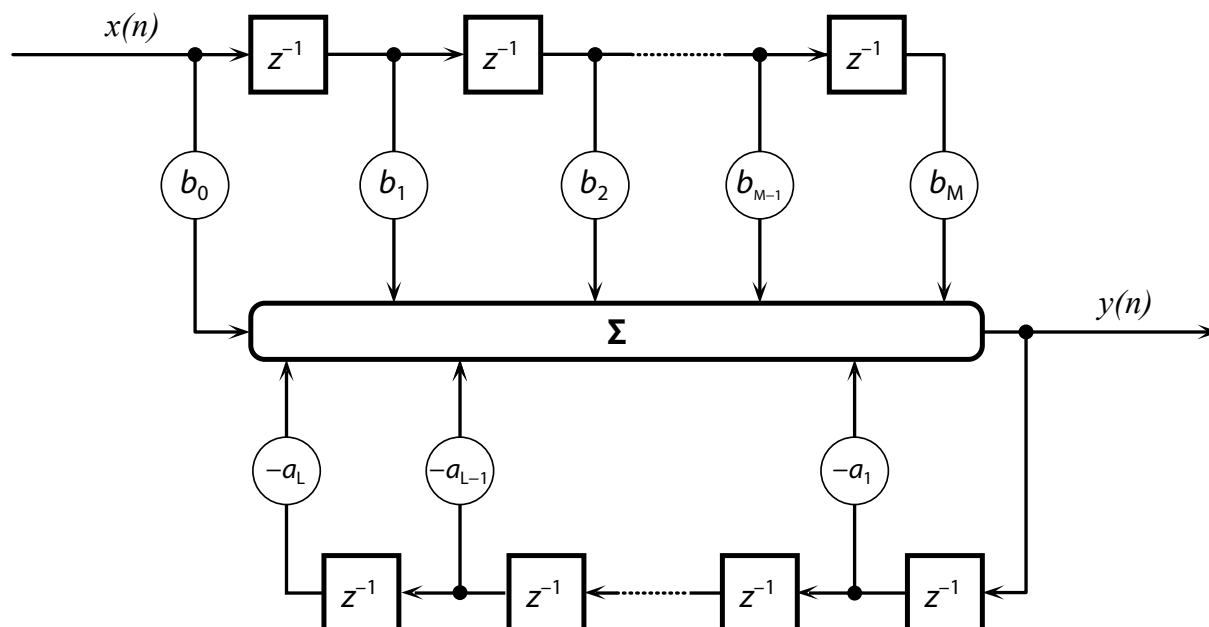
V úvodní první kapitole bylo ukázáno, že diskrétní systém lze jednoznačně popsat nejen diferenční rovnicí, ale také jeho impulsní charakteristikou  $h(n)$ , jeho frekvenční charakteristikou  $G(\omega)$ , jeho přenosovou funkcí  $H(z)$  nebo konfigurací nulových bodů  $n_i$  a pólů  $p_i$  jeho přenosové funkce. Dále byly také ukázány souvislosti a transformace mezi jednotlivými způsoby popisu. Ze znalosti diferenční rovnice by tedy nemělo čtenáři činit problém sestavit přenosovou funkci systému, k ní určit rozložení nulových bodů a pólů v komplexní rovině  $z$  a zjistit, zda je systém stabilní.



## 2 Lineární filtrace

Filtrací se označují takové operace, které vedou k oddělení užitečné a rušivé složky nebo také ke zvýraznění důležitých složek časových řad, jež jsou předmětem zpracování či analýzy. Termín filtrace sám o sobě upozorňuje na fakt, že cílem je určité složky propustit či zesílit a zároveň utlumit ty ostatní.

V předchozí kapitole bylo uvedeno, že LTI systém může požadovaným způsobem měnit frekvenční složení zpracovávaných časových řad. Takovýto systém, algoritmus nebo program, může být označen také jako **filtr** a jeden z mnoha způsobů jeho realizace pomocí operací sčítání, násobení a pomocí zpožďovacích členů  $z^{-1}$  je ukázán na obr. 2.1. Je vidět, že koeficienty realizace diskrétního systému souvisejí přímo s koeficienty v jeho diferenční rovnici (1.16). Vyšší z čísel  $M$  a  $L$  udává vyšší stupeň z obou polynomů v rovnici (1.13) a označuje se jako **řád filtru**.

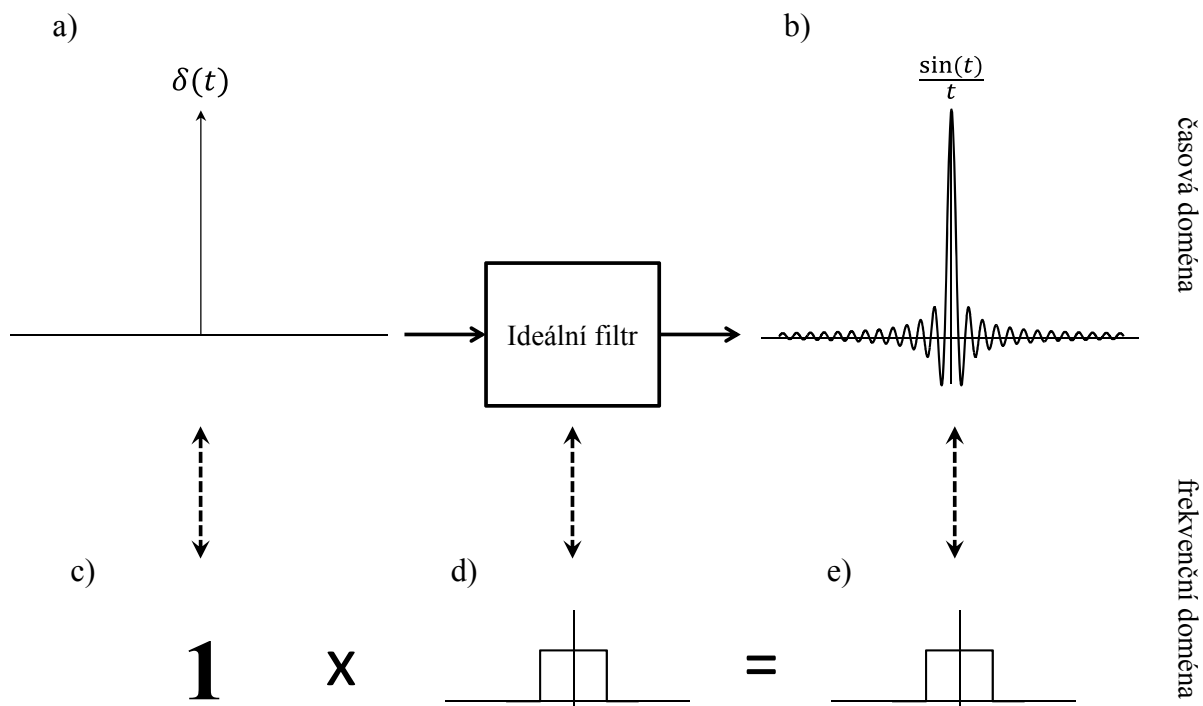


**Obr. 2.1** Realizace filtru/programu/systému přímou formou. Realizační schéma i rovnici (1.16) je možno interpretovat tak, že LTI systém (nebo LTI filtr) uchovává v paměti starší vzorky časových řad na vstupu i na výstupu a jejich lineární kombinací počítá vzorky časové řady na výstupu.

Mohlo by se na první pohled zdát, že filtrování ve frekvenční doméně umožňuje libovolně tlumit nežádoucí harmonické složky časových řad. Pro tyto operace ve frekvenční doméně však existují určitá omezení, protože mohou jednak způsobovat podstatné oscilace v přenosové funkci filtru a dále mohou být příčinou nežádoucích přechodových dějů v časové doméně. Filtrování harmonických složek proto nelze ani v digitálním světě řešit **ideálními filtry**. Stanovení typu ideálního filtru je však výchozím bodem pro sestavení reálného filtru, jehož frekvenční charakteristika se obvykle blíží ideálnímu tvaru.

Na obr. 1.5 a obr. 2.2 jsou frekvenční charakteristiky ideálního filtru typu dolní propust, jehož účelem je zbavit časové řady rušivých složek o vysokých frekvencích. Filtrace horní propust se využívá, pokud rušivé složky časových řad vykazují nižší frekvence než složky užitečné. Pokud užitečné složky časové řady leží v určitém frekvenčním pásmu, použije se

pro jejich výběr filtr typu pásmová propust. Pro odstranění úzkopásmového rušení pak slouží filtr typu pásmová zadrž, viz obr. 2.3.



**Obr. 2.2** Ideální filtr typu dolní propust by v případě své existence kompletně odstraňoval vysokofrekvenční složky a ponechával by beze změny nízkofrekvenční složky zpracovávaných časových řad. Takové chování odpovídá obdélníkové frekvenční charakteristice d), a to včetně záporných frekvencí. Výstup filtru ve frekvenční oblasti je dán součinem obrazu časové řady a frekvenční charakteristiky filtru e). Odezva takového systému v časové oblasti – impulsní charakteristika b) – na jednotkový impuls a) brání v existenci takového systému, neboť odezva je nenulová pro časy  $t < 0$ . V analogovém světě nelze ideální filtr realizovat, poněvadž jde o nekauzální systém. V digitálním světě nelze ideální filtr realizovat, poněvadž impulsní charakteristika pro svůj rozsah od  $-\infty$  do  $+\infty$  vyžaduje nekonečnou paměť a frekvenční charakteristika obsahuje nekonečně strmé náběžné a sestupné hrany. Na obrázku jsou frekvenční charakteristiky a spektra znázorněna jen ilustrativně, bez fázových částí.

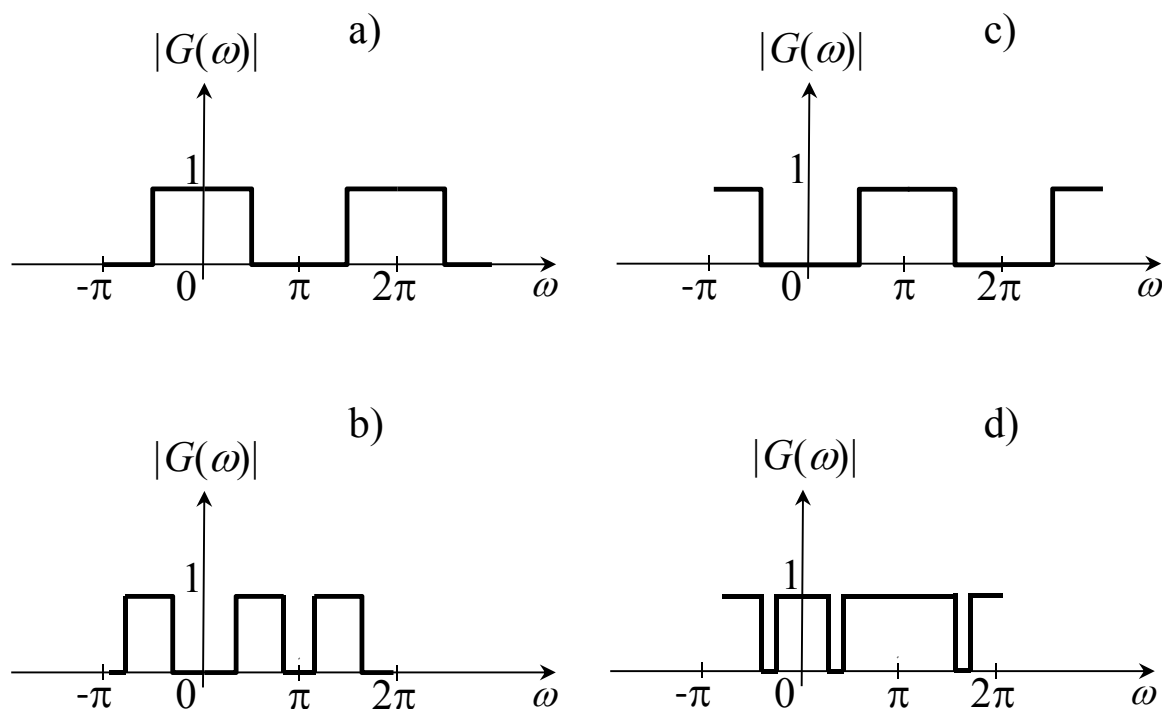
## 2.1 Klasifikace filtrů podle algoritmu (AR, MA, ARMA)

Horní část schématu na obr. 2.1 představuje **nerekurzivní** část realizace filtru, tj. část, ve které je pro výpočet vzorku časové řady  $y(n)$  na výstupu potřeba pouze zpožděných hodnot vstupní časové řady  $x(n)$ . Na opačné straně realizačního schématu se ve výpočtu výstupní časové řady  $y(n)$  uplatňují její předchozí vzorky – jedná se tedy o **rekurzivní** část filtru se zpětnými vazbami.

Budou-li koeficienty  $a_i$  v rovnici (1.16) nulové pro všechna  $i \geq 1$ , filtr bude nerekurzivní a bude reprezentován pouze koeficienty  $b_i$ . Takový filtr bez zpětné vazby dává na výstup

vážený průměr z konečného počtu  $M+1$  posledních vzorků časové řady na vstupu, a proto se často označuje jako **MA** filtr (*moving average*)<sup>5</sup>.

Budou-li koeficienty  $b_i$  nulové pro všechna  $i \geq 1$ , filtr bude rekurzivní a bude reprezentován pouze koeficienty  $b_0$  a  $a_i$ . Takový filtr se označuje jako **AR** filtr (*autoregressive*). Výstup autoregresního filtru závisí pouze na aktuální hodnotě vstupní časové řady a dále na konečném počtu  $L$  starších vzorků časové řady na výstupu.



**Obr. 2.3** Amplitudové frekvenční charakteristiky a) ideální dolní propusti, b) ideální pásmové propusti, c) ideální horní propusti a d) ideální pásmové zádrže.

Pokud filtr obsahuje ve svém algoritmu rekurzivní i nerekurzivní část, tj. alespoň jeden z koeficientů  $b_i$  pro  $i \geq 1$  je nenulový a současně alespoň jeden z koeficientů  $a_i$  pro  $i \geq 1$  je nenulový, pak se hovoří o filtru **ARMA** (*autoregressive moving average*). Pro úplnost je nutné dodat, že i ARMA filtr je vždy rekurzivní – obsahuje zpětnou vazbu.

Přenosová funkce filtru ARMA má tvar racionálně lomené funkce, viz (1.13). Lze-li tuto přenosovou funkci zjednodušit krácením zlomků tak, že výsledkem dělení je přenosová funkce ve tvaru s polynomy nižšího řádu, pak lze říci, že obě verze přenosové funkce před krácením a po krácení popisují dva filtry s podobnými vlastnostmi, avšak s různými realizačními algoritmy. Prostým vydělením racionálně lomené přenosové funkce filtru ARMA lze ve vybraných případech dospět k přenosové funkci filtru AR nebo i k přenosové funkci filtru MA. Z tohoto lze usuzovat, že filtry s velmi podobnými vlastnostmi lze realizovat zcela rozdílnými algoritmy – např. filtr ARMA se zpětnými vazbami a filtr MA bez zpětných vazeb v úloze 2.2.

<sup>5</sup> Přestože MA filtr pracuje s  $M+1$  vzorky, jeho řád je  $M$ , neboť řád neurčuje počet koeficientů filtru, ale počet jeho zpožďovacích členů.

## 2.2 Klasifikace filtrů podle impulsní charakteristiky (FIR, IIR)

V literatuře zabývající se zpracováním a analýzou diskrétních signálů se obvykle využívá jiné třídící terminologie než AR, MA, ARMA. Systémy nebo filtry se zde nedělí podle toho, zda využívají rekurzivní či nerekurzivní algoritmus, nýbrž podle charakteru jejich impulsní charakteristiky. Rozeznávají se jednak filtry typu **FIR** (*finite impulse response*) s konečnou impulsní charakteristikou a dále filtry typu **IIR** (*infinite impulse response*), jejichž impulsní charakteristika je nekonečná, tj. odezva filtru na jednotkový impuls je nenulová pro  $n \rightarrow \infty$ . Pro FIR filtry platí stejná pravidla jako pro výše uvedené systémy MA. Pod označení IIR se pak řadí většinou systémy AR i systémy ARMA, neboť pokud časová řada na výstupu filtru závisí na svých předcházejících vzorcích, bude zřejmě odezva na jednotkový impuls nekonečná, i když tomu tak ve zvláštních případech být nemusí.

Přestože je zde patrný překryv klasifikací AR, MA, ARMA a IIR, FIR, není tento překryv úplný, neboť lze navrhnout filtr či systém, který sice bude mít svou rekurzivní část, a přesto bude mít jeho impulsní charakteristika konečný počet vzorků<sup>6</sup>. V literatuře věnující se oblastem identifikace, návrhu i realizace filtrů či jiných diskrétních systémů převažuje klasifikace filtrů na FIR a IIR, zatímco pro modely procesů generujících data náhodné povahy je v literatuře zavedené využití klasifikace na AR, MA, ARMA, která je spojována se jmény statistiků George Boxe a Gwilyma Jenkinse.

## 2.3 Filtry s konečnou impulsní charakteristikou

Filtry s konečnou impulsní charakteristikou se realizují obvykle nerekurzivním algoritmem. Tyto filtry svou diferenční rovnicí odpovídají MA systémům a bývají často označovány také jako nerekurzivní filtry nebo konvoluční filtry. Pojem *moving average* vcelku jasně osvětluje funkci MA nebo FIR filtru, neboť na výstup dává vážený průměr  $M+1$  posledních vzorků časové řady na vstupu. Rovnice (2.1) ukazuje, že realizační koeficienty  $b_i$  FIR filtrů, které vlastně tvoří váhy výsledného váženého průměru, odpovídají přímo hodnotám impulsní charakteristiky  $h(n)$ .

$$\begin{aligned} y(n) &= b_0 \cdot x(n) + b_1 \cdot x(n-1) + b_2 \cdot x(n-2) + \dots + b_{M-1} \cdot x(n-M) = \\ &= \sum_{k=0}^M b_k \cdot x(n-k) = \text{subst: } b_k = h(n) \mid = h(n) * x(n). \end{aligned} \quad (2.1)$$

Přenosová funkce FIR systémů má polynomiální tvar, viz rovnice (2.2), z čehož lze usuzovat, že  $M$  nulových bodů bude rozmístěno kdekoli v rovině  $z$  (obvykle přímo na jednotkové kružnici  $|z|=1$ ) a  $M$ -násobný pól bude v počátku roviny  $z$ , tj. ve středu jednotkové kružnice. FIR filtry jsou tedy vždy stabilní.

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{k=0}^M b_k \cdot z^{-k} = A \cdot \frac{\prod_{k=1}^M (z - n_k)}{z^M}. \quad (2.2)$$

---

<sup>6</sup> Filtr s rekurzivním algoritmem může ve zvláštních případech vykazovat konečný počet vzorků impulsní charakteristiky, a to tehdy, je-li zajištěno, aby se v určitých taktech vzájemně vyrušily příspěvky zpětnovazební smyčky s příspěvky z nerekurzivní části filtru.

Kromě zaručené stability je další výhodou FIR filtrů také možnost dosažení přesně lineární fázové frekvenční charakteristiky, což je často požadovaná vlastnost. Vykazuje-li totiž filtr lineární fázovou charakteristiku, dochází při zpracování časové řady takovým filtrem ke shodnému tzv. **skupinovému zpoždění** jednotlivých harmonických komponent, a tvar časové řady tak není zkreslen. Lineární fázová charakteristika je zaručena v případě symetrické nebo antisymetrické impulsní charakteristiky, tj. pokud je splněna podmínka osové nebo bodové souměrnosti podle rovnice (2.3).

$$h(n) = \pm h(M - n), \quad n = 0, 1, 2, \dots, M. \quad (2.3)$$

Filtry s lineární fázovou charakteristikou vykazují specifické rozložení nulových bodů, neboť je-li  $H(n_k)=0$ , pak i  $H(1/n_k)=0$ . Pokud se jedná navíc o systém s reálnými koeficienty, pak platí také, že komplexně sdružené  $n_k^*$  a  $1/n_k^*$  jsou rovněž nulovými body přenosové funkce, a tedy nulové body takového filtru se vyskytují hned ve čtveřicích, viz obr. 2.4.

Pro návrhy filtrů FIR se užívá celá řada metod, jejichž přehled včetně detailních postupů lze nalézt např. v [12] a [14]. Základní a přímočarou metodou je návrh zadáním průběhu modulové frekvenční charakteristiky. Po ekvidistantním vzorkování požadované frekvenční charakteristiky  $G_d(\omega_k)$  v  $N$  bodech<sup>8</sup> se aplikuje inverzní DFT, kterou je vypočtena impulsní charakteristika  $h(n)$ :

$$h(n) = \frac{1}{N} \sum_{k=0}^{N-1} G_d\left(\frac{2\pi}{N}k\right) e^{j2\pi nk/N}. \quad (2.4)$$

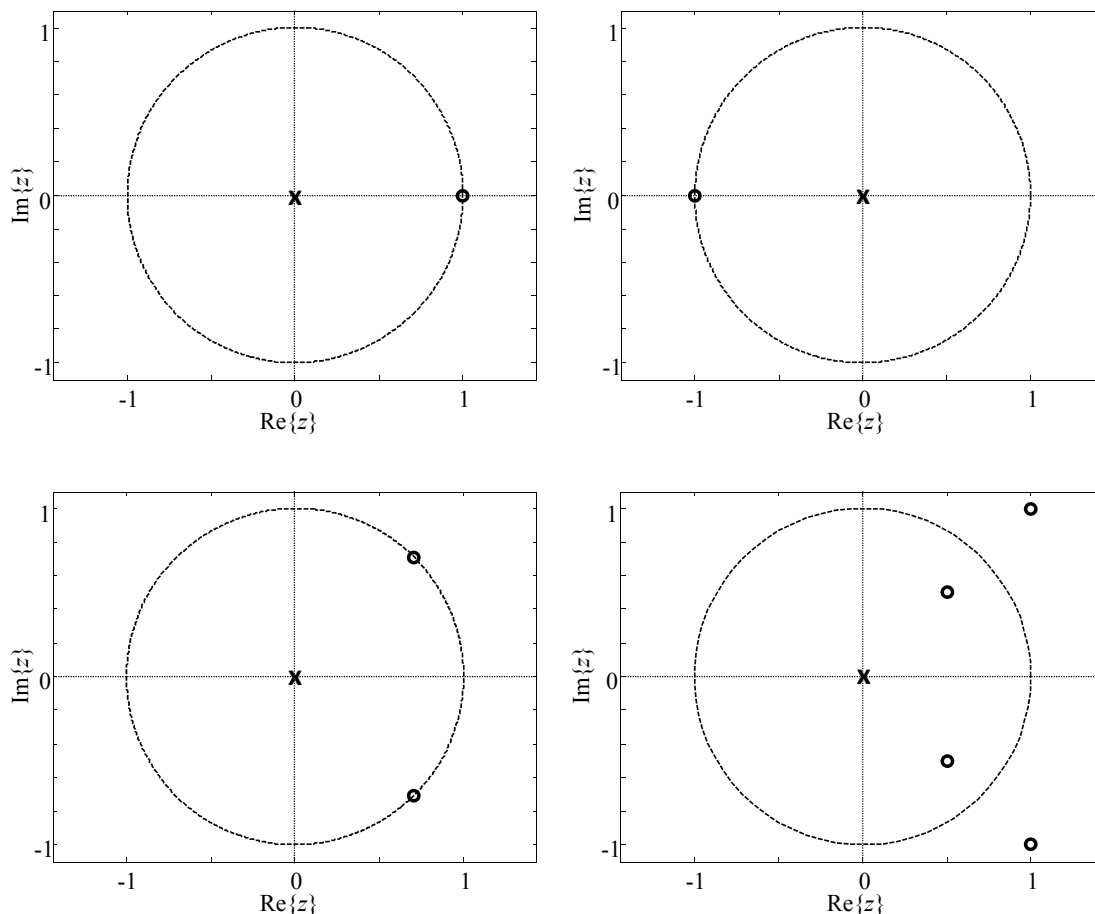
Modulová frekvenční charakteristika výsledného filtru  $G(\omega)$  interpoluje přesně všech  $N$  zadaných vzorků, nicméně její průběh mezi nimi nelze nijak ovlivnit. Jsou-li požadovány strmé přechody mezi propustným a nepropustným pásmem, jsou ve výsledné frekvenční charakteristice patrné významné oscilace, viz obr. 2.5. Právě těmito oscilacemi se odlišuje výsledná frekvenční charakteristika od jejího požadovaného ideálního průběhu.

Aby byla impulsní charakteristika navrhovaného filtru  $h(n)$  reálná a kauzální, je třeba při vzorkování pamatovat na symetrii amplitudové i fázové frekvenční charakteristiky. Pro sudé a liché počty vzorků a dále pro symetrickou a antisymetrickou  $h(n)$  je možno z rovnice (2.4) odvodit čtyři typy FIR filtrů s lineární fází, viz tab. 2.1.

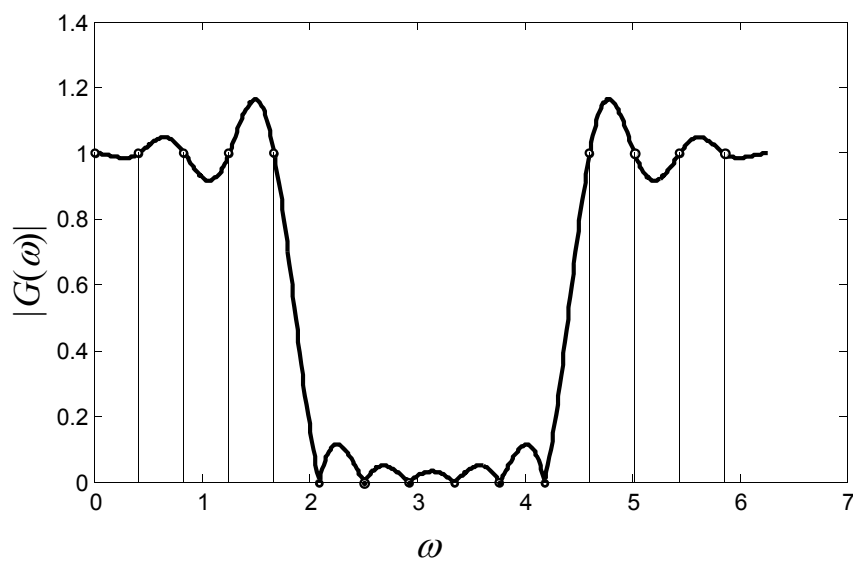
Alternativně lze při návrhu zadat výpočetnímu programu (např. Matlab, Signal processing toolbox) pouze vzorky reálné modulové frekvenční charakteristiky a fázovou charakteristiku ponechat nulovou ve všech vzorcích. Výslednou impulsní charakteristiku je pak nutno kauzalizovat, tj. přeuspořádat pořadí jejich vzorků, jak bude ukázáno v řešené úloze na konci kapitoly.

<sup>7</sup> Záporná derivace fázové frekvenční charakteristiky filtru podle úhlové frekvence se nazývá **skupinové zpoždění** filtru. Při lineární fázové charakteristice je skupinové zpoždění konstantní – nezávislé na frekvenci jednotlivých harmonických složek. FIR filtr řádu  $M$  s lineární fází má skupinové zpoždění  $M/2$ , a zpozdí tedy filtrovanou časovou řadu o  $M/2$  vzorků.

<sup>8</sup> Vzhledem k periodicitě frekvenční charakteristiky jsou hodnoty  $G_d(\omega_k)$  totožné pro  $k=0$  a pro  $k=N$ . Řád výsledného FIR filtru získaného po  $N$ -bodové inverzní DFT bude  $M=N-1$ .



**Obr. 2.4** Příklady rozložení nulových bodů a pólů čtyř různých reálných FIR filtrů. U nulových bodů lze pozorovat specifické rozložení, neboť se vyskytují v komplexně sdružených a recipročních párech.



**Obr. 2.5** Modulová frekvenční charakteristika FIR filtru typu dolní propust metodou vzorkování frekvenční charakteristiky. Čáry a tečky v grafu představují 15 vzorků ideální horní propusti a tlustá čára představuje amplitudovou charakteristiku výsledného filtru.

**Tab. 2.1** Čtyři typy FIR filtrů s lineární fázovou frekvenční charakteristikou.

Impulsní charakteristika	počet vzorků $N$	$G(\omega)$	Typ
$h(n) = + h(N-1-n)$	Lichý	$e^{-j\frac{\omega(N-1)}{2}} \left( h\left(\frac{N-1}{2}\right) + 2 \sum_{k=1}^{\frac{N-3}{2}} h\left(\frac{N-1}{2}-k\right) \cos(\omega k) \right)$	1
$h(n) = + h(N-1-n)$	Sudý	$e^{-j\frac{\omega(N-1)}{2}} 2 \sum_{k=1}^{\frac{N-3}{2}} h\left(\frac{N}{2}-k\right) \cos\left(\omega\left(k-\frac{1}{2}\right)\right)$	2
$h(n) = - h(N-1-n)$	Lichý	$e^{-j\left[\frac{\omega(N-1)}{2} \frac{\pi}{2}\right]} 2 \sum_{k=1}^{\frac{N-1}{2}} h\left(\frac{N-1}{2}-k\right) \sin(\omega k)$	3
$h(n) = - h(N-1-n)$	Sudý	$e^{-j\left[\frac{\omega(N-1)}{2} \frac{\pi}{2}\right]} 2 \sum_{k=1}^{\frac{N-1}{2}} h\left(\frac{N}{2}-k\right) \sin\left(\omega\left(k-\frac{1}{2}\right)\right)$	4

## 2.4 Filtry s nekonečnou impulsní charakteristikou

Filtry s nekonečnou impulsní charakteristikou se realizují vždy rekurzivním algoritmem. Tyto filtry svou diferenční rovnicí odpovídají AR nebo ARMA systémům. Přenosová funkce IIR filtrů má tvar racionální lomené funkce, viz rovnice (1.13), z čehož lze usuzovat, že u těchto systémů bude vždy nutné vyšetřovat jejich stabilitu. Filtry IIR nemají lineární průběh fázové frekvenční charakteristiky, a tak mohou zkreslovat zpracovávané časové řady i v propustných pásmech.

Návrh IIR filtrů je méně intuitivní a složitější než v případě návrhu FIR filtrů. Provádí se buď interaktivním rozmísťováním nulových bodů a pólů, nebo se používají optimalizační metody podle požadované frekvenční charakteristiky, které vedou na řešení soustavy nelineárních rovnic. Při dostupnosti potřebného softwarového vybavení (např. Matlab a Signal processing toolbox) lze pro nejrychlejší návrh IIR filtru využít přístupu založeného na podobnosti s analogovými filtry<sup>9</sup>, pro které existují tabulkové definice normovaných dolních propustí. Výhodou IIR filtrů oproti FIR filtrům je možnost dosažení mnohem strmějších přechodů mezi propustným a nepropustným pásmem při stejném řádu filtru, viz obr. 2.11. Tuto výhodu je možné vyjádřit také jako schopnost dosahovat menších zpoždění IIR filtry než FIR filtry při stejných požadavcích na tvar frekvenční charakteristiky.

<sup>9</sup> V minulosti, kdy se většina součástek realizovala analogovými obvody, bylo standardizováno několik typů filtrů pojmenovaných podle svých objevitelů: Čebyševův filtr, Butterworthův filtr, Besselův filtr, eliptický Cauerův filtr aj. Převod analogového filtru na digitální může být založen na a) nalezení ekvivalence mezi diferenciálem a konečně malým přírůstkem, b) impulzně invariantní transformací, c) bilineární transformací (Skalický, 1995). Dále se ještě aplikují frekvenční transformace, aby mohly být z normovaných dolních propustí odvozeny i filtry s jinou konfigurací propustného a nepropustného pásma.

## 2.5 Úlohy

**ÚLOHA 2.1** Pan Josef Vdoleček se přistěhoval do obce, jejíž zastupitelstvo před časem schválilo budoucím sousedům pana Josefa vystavět a provozovat autolakovnu přímo v uliční zástavbě (realitní kancelář na tento fakt pana Josefa pro jistotu neupozornila). Odbor životního prostředí obce s vyšší působností na četné stížnosti kvůli zápachu linoucímu se pravidelně z lakovny sice reagoval provedením místního šetření, avšak opožděnými kontrolami nebyly nikdy zjištěny nadlimitní koncentrace zapáchajících látek v okolním ovzduší, neboť provozovatelé lakovny stihli vždy v pravý čas uvést veškerou nákladnou filtraci do chodu. Pan Josef Vdoleček se proto rozhodl vzít vše do svých rukou, pořídil si certifikovaný chemický plynový analyzátor a ovzduší na svém dvorku průběžně monitoroval a to tak, že každou hodinu změřil koncentrace zapáchajících látek a zapsal si je do sešitu v programu MS Excel pro pozdější analýzu. Jev, který sledoval, se však skládal z více komponent, neboť zápach vypouštěla v krátkých intervalech s průměrnou periodou 4 hodiny (opravy karosářských dílů) jednak lakovna a dále ke koncentracím zapáchajících látek přispívaly i pozvolné děje s průměrnou periodou 12 hodin (provoz továren) od vzdálenějších producentů zápachu.

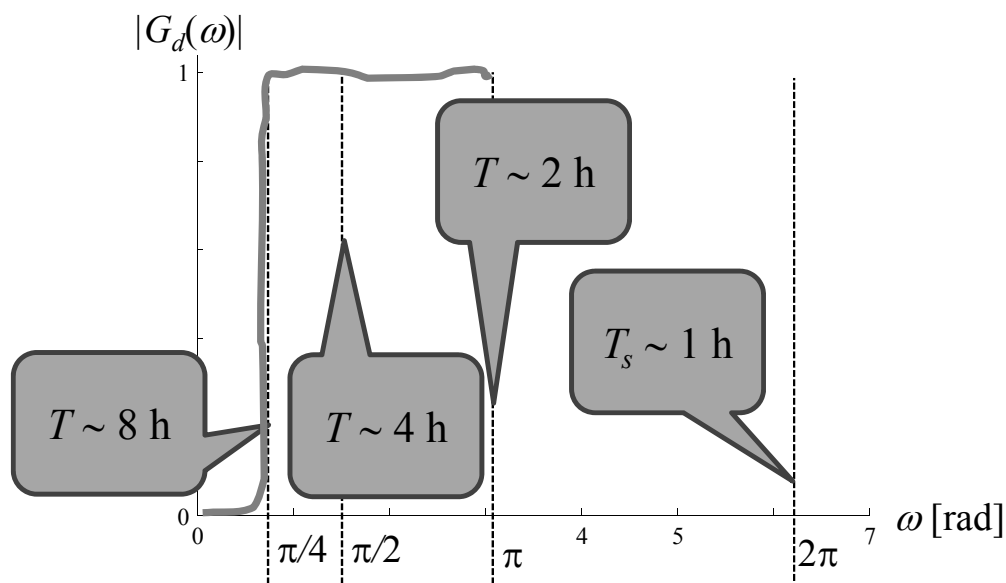
Úkoly:

- Navrhněte FIR filtr pro odstranění rušivé složky v časové řadě reprezentující monitoring údajů o koncentraci zapáchajících látek na dvorku pana Josefa. Volte filtr s polynomem takového řádu, aby zpoždění filtru bylo maximálně 10 hodin.
- Navržený filtr popište impulsní charakteristikou a diferenční rovnicí.
- Vykreslete frekvenční charakteristiku navrženého filtru.
- Zkontrolujte linearitu fázové charakteristiky a určete skupinové zpoždění filtru.
- Je nutné u tohoto typu filtru kontrolovat stabilitu?

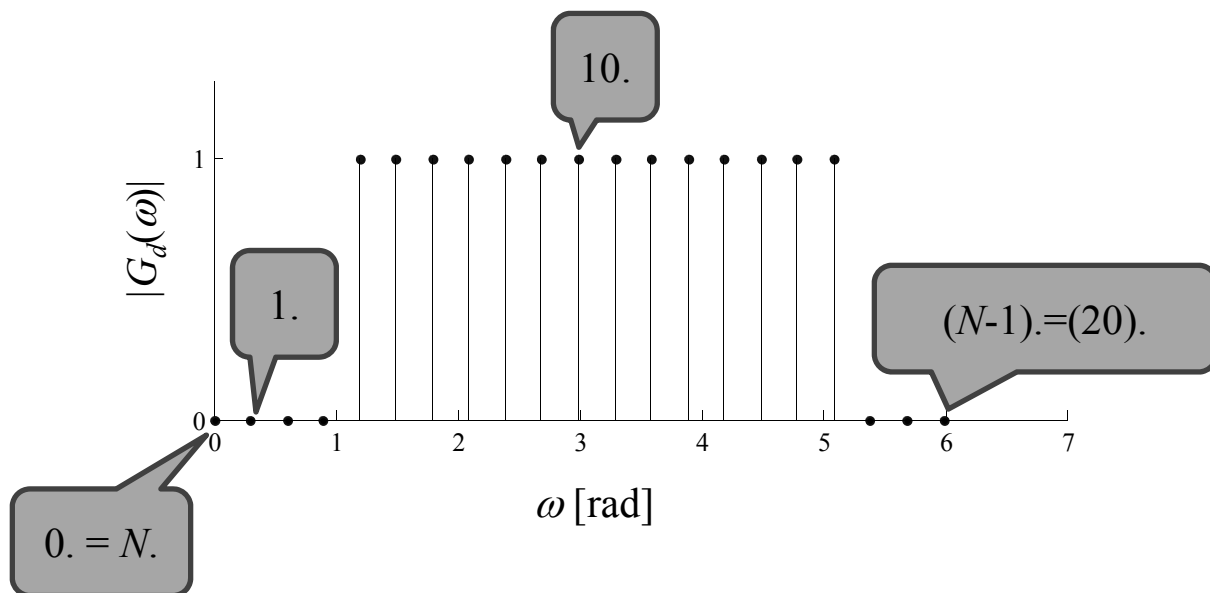
Příslušný program pro Matlab je na obr. 2.11.

- Volíme filtr s nejvyšším možným řádem, který nám ještě umožní požadované skupinové zpoždění:  $M = \tau \cdot 2 = 20$ . Počet vzorků frekvenční charakteristiky:  $N = M + 1 = 21$ . Rušivá složka časové řady určené k filtraci má delší periodu než složka užitečná – filtr tedy bude typu horní propust. Rozvahu ve frekvenční doméně a návrh filtru ukazují komentované obrázky 2.6 a 2.7.
- Viz obr. 2.8.
- Viz obr. 2.9.
- Fázová charakteristika je dle očekávání lineární – impulsní charakteristika vykazuje sudou/lichou symetrii. Body zlomu odpovídají nulovým bodům přenosové funkce. Odečtení skupinového zpoždění z grafu je na obr. 2.9.
- FIR filtry jsou vždy stabilní, neboť obsahují pouze násobné póly v počátku. Pro zajímavost je rozložení nulových bodů ukázáno na obr. 2.10.

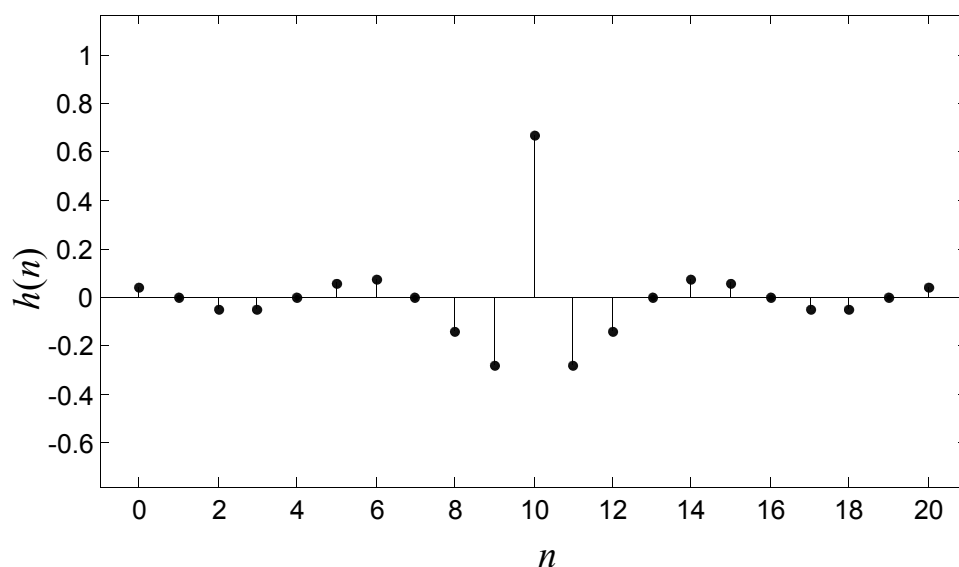




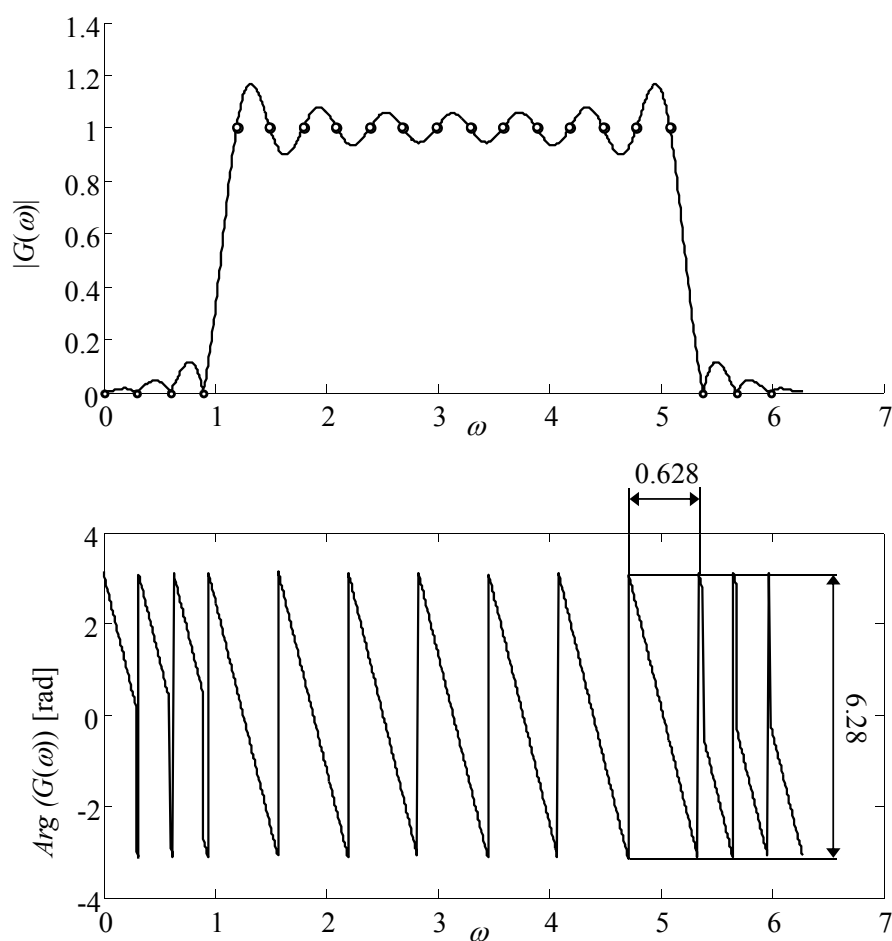
**Obr. 2.6** Úloha 2.1 – rozvaha ve frekvenční doméně. Normované úhlové frekvenci  $2\pi$  odpovídá vzorkovací frekvence – zde vyjádřeno pomocí vzorkovací periody. Dalším dělením frekvenční osy (přerušované čáry) jsou orientačně vyznačeny periody jevů a nakonec tenkou tlustou čarou naznačen ideální průběh amplitudové frekvenční charakteristiky.



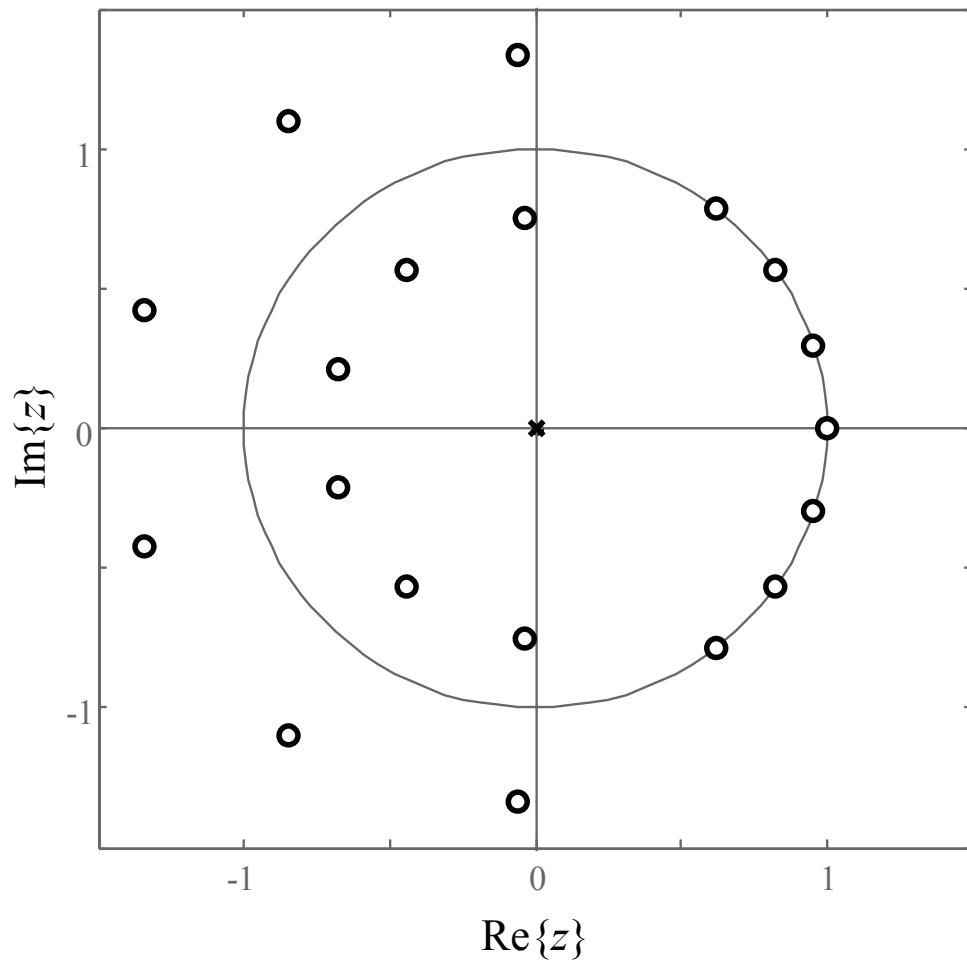
**Obr. 2.7** Úloha 2.1 – návrh horní propusti vzorkováním frekvenční charakteristiky. Požadovaná modulová frekvenční charakteristika je symetrická podle 10. vzorku.



**Obr. 2.8** Úloha 2.1 – impulsní charakteristika navrženého FIR filtru. Jeho diferenční rovnice:  
 $y(n)=h_0x(0)+h_1x(1)+\dots+h_{21}x(21),$



**Obr. 2.9** Úloha 2.1 – modulová (nahore) a fázová (dole) frekvenční charakteristika navrženého FIR filtru. Z grafu fázové charakteristiky lze odečíst její směrnici, a určit tak skupinové zpoždění, které je v tomto případě deset vzorkovacích intervalů, tj. 10 hodin.



**Obr. 2.10** Úloha 2.1 – rozložení nulových bodů a pólů navrženého filtru FIR.

```

%%% ULOHA 2.1 %%%

Ts = 1;          % vzorkovací perioda je jedna hodina
                  % nejkratší perioda ve sledovaném ději
                  % má tedy 2 hodiny
                  % průměrná perioda sledovaného děje je 4 h
                  % průměrná perioda rušivého děje je 12 h

Gd = zeros(1,21); % rad filtru M=20, délka char-stik N=21
                  % F(10) odpovídá π rad/s - místo symetrie
                  % a děje s periodou 2 h
                  % F(5) odpovídá dějum s periodou 4 h
                  % F(5/2) odpovídá dějum s periodou 8 h
                  % F(5/3) odpovídá dějum s periodou 12 h

Gd(5:18) = ones(1,14);
                  % vynulované jsou: (0=21), 1,2,3 a 20,19,18
                  % tj. 4 vzorky z každé strany jedné periody
                  % frekvencní char-ky

h = ifft(Gd);    % inverzní DFT

```

```

% kauzalizace imp. char-ky
% prehoz. leve a prave poloviny posloupnosti
h = [h(ceil(length(h)/2)+1:end) h(1:ceil(length(h)/2))];

figure; % vykreslení impulsní char-ky
stem([0:length(h)-1],h);
xlabel('n'); ylabel('h(n)');

[G,w]=freqz(h,1,'whole',1024);
% vypočet „spojité“ frekv. charakteristiky
% komplexní frekv. char-ka G bude definována
% na 1024 vzorcích a intervalu omega=0..2pi

figure, % vykreslení frekvenční charakteristiky
plot(w,abs(G),'k');

wk = linspace(0,2*pi-(2*pi)/21,21);
% uhlové frekvence, na kterých byla G vzorkována
% neboli uhlové frekvence,
% na kterých je definována Gd

hold on; % vykreslení vzorku Gd do stejného grafu jako G
stem(wk,Gd);
xlabel('w'); ylabel('|G(w)|');

figure; % vykreslení fázové frekv. char-ky
plot(w,angle(G),'k');
xlabel('w'); ylabel('arg(G(w))');

sys = firlt(h,1); % definice objektu typu LTI filtr
figure, pzmap(sys); % vykreslení nulových bodů a polů

```

**Obr. 2.11** Úloha 2.1 – řešení v Matlabu.

**ÚLOHA 2.2** Je dán filtr s přenosovou funkcí  $H(z) = (z^{-2} - z^{-6}) / (1 - z^{-2})$ .

Otázky:

- Jedná se o filtr FIR nebo IIR?
- Je tento filtr stabilní?

$$H(z) = \frac{z^{-2} - z^{-6}}{1 - z^{-2}} = \frac{z^4 - 1}{z^6 - z^4} = \frac{(z^2 + 1)(z^2 - 1)}{z^4(z^2 - 1)} = \frac{(z + j)(z - j)}{z^4}.$$

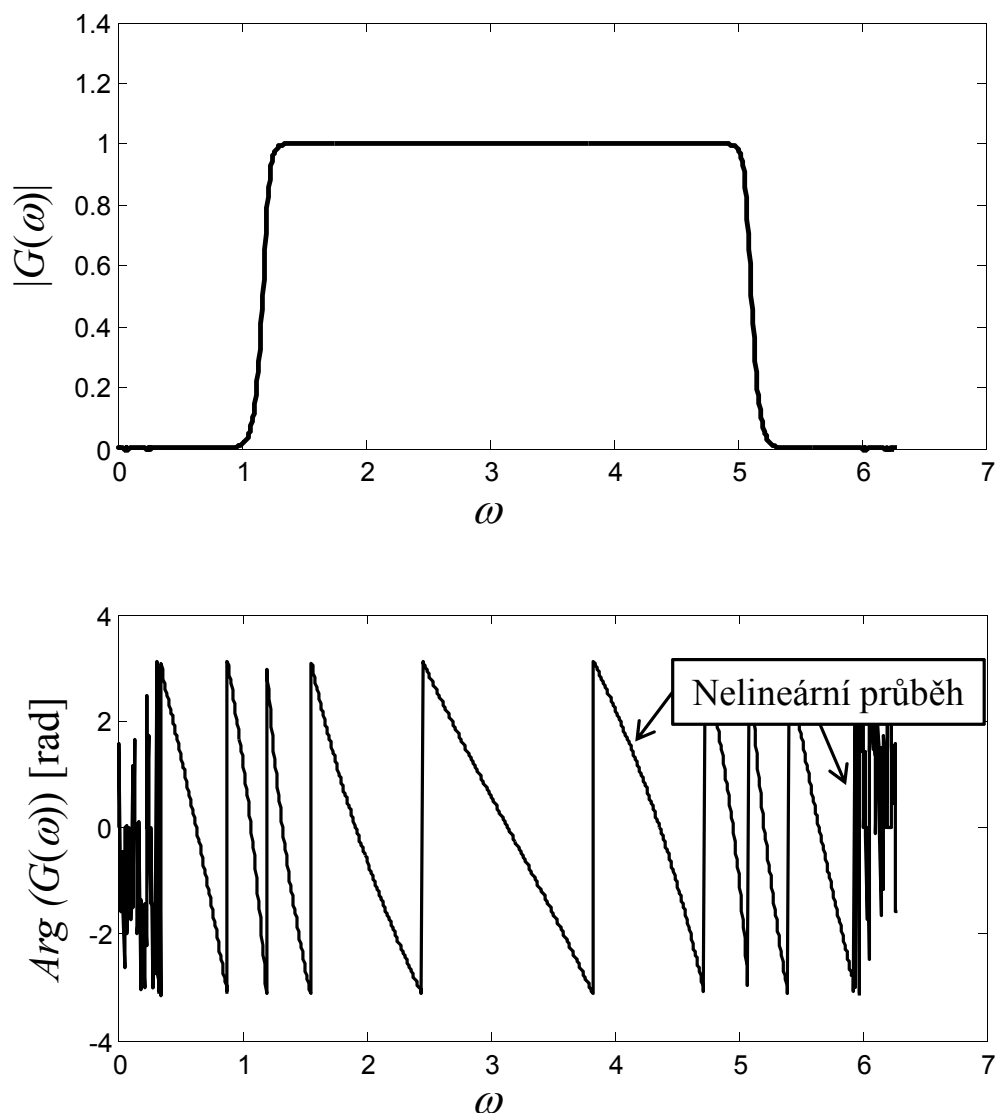
$$H(z) = \frac{z^{-2} - z^{-6}}{1 - z^{-2}} = \frac{z^{-2}(1 + z^{-2})(1 - z^{-2})}{1 - z^{-2}} = z^{-2} + z^{-4}.$$

Po krácení polynomů byla získána přenosová funkce ve tvaru jediného polynomu. Jedná se tedy o filtr FIR, který je vždy stabilní. Zaručenou stabilitu lze odvodit také z krácené přenosové funkce ve tvaru součinu kořenových činitelů, která má dva nulové body ( $n_1 = -j$ ,  $n_2 = +j$ ) a čtyřnásobný pól v počátku ( $p_{1..4} = 0$ ).

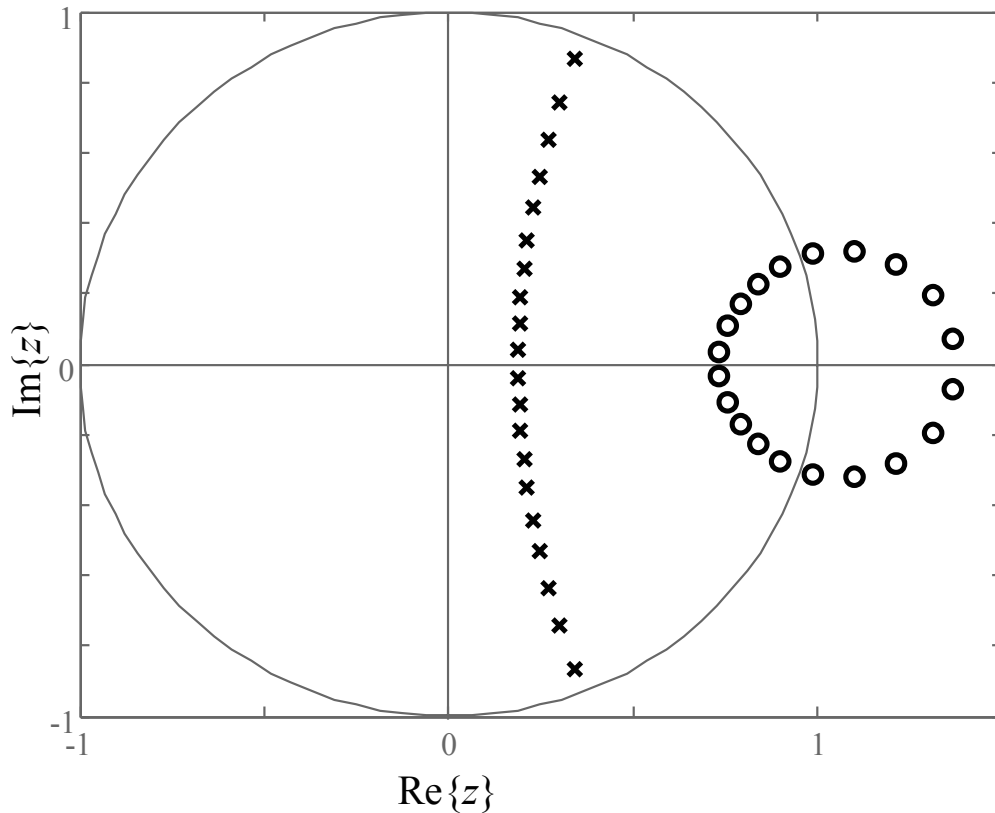
Pokud by byl filtr realizován podle původní nekrácené přenosové funkce, tj. s algoritmem ARMA, nešlo by zaručit stabilitu takového filtru, neboť vzhledem k pólům ( $p_5 = 1$ ,  $p_6 = -1$ ) by se takto realizovaný filtr nacházel na mezi stability.

**ÚLOHA 2.3** Navrhněte pro pana Josefa Vdolečka, se kterým jste se seznámili v úloze 2.1, IIR filtr 20. řádu a porovnejte jeho vlastnosti s FIR filtrem. Upozorněte pana Josefa na eventuální obtíže vyplývající z použití IIR filtru.

Příslušný program pro Matlab je na obr. 2.13. Frekvenční charakteristika Butterworthova IIR filtru 20. řádu je na obr. 2.11 a rozložení nulových bodů a pólů je na obr. 2.12. Pana Josefa Vdolečka je třeba upozornit na nelinearitu fázové frekvenční charakteristiky, díky které se jednotlivé harmonické složky zpracovávané časové řady dostanou na výstup s různým skupinovým zpožděním, a dojde tak ke tvarovému zkreslení výstupní časové řady.



**Obr. 2.11** Úloha 2.3 – modulová (nahore) a fázová (dole) frekvenční charakteristika IIR Butterworthova filtru typu horní propust 20. řádu. V grafu fázové charakteristiky lze sledovat její nelineární průběh.



**Obr. 2.12** Úloha 2.3 – rozložení nulových bodů a pólů IIR Butterworthova filtru typu horní propust 20. řádu.

```
%%% ULOHA 2.3 %%%%%%%%% IIR BUTTERWORTHUV FILTR - HORNI PROPUST

wcutoff = 1.1968/pi; % cut-off se zadava v intervalu 0..1,
                    % kde 1 odpovida polovina vzorkovací frekvence

[b,a] = butter(20,wcutoff,'high'); % koeficienty Butterworthova
                                   % filtru 20. radu, horni propust

sys = filt(b,a); % definice objektu typu LTI filtr
figure, pzmap(sys); % vykreslení nulových bodu a polu

[G,w]=freqz(b,a,'whole',1024); % komplexni frekvencni char-ka

figure, plot(w,abs(G),'k'); xlabel('w'); ylabel('|G(w)|');
figure, plot(w,angle(G),'k'); xlabel('w'); ylabel('arg(G(w))');
```

**Obr. 2.13** Úloha 2.4 – řešení v Matlabu.

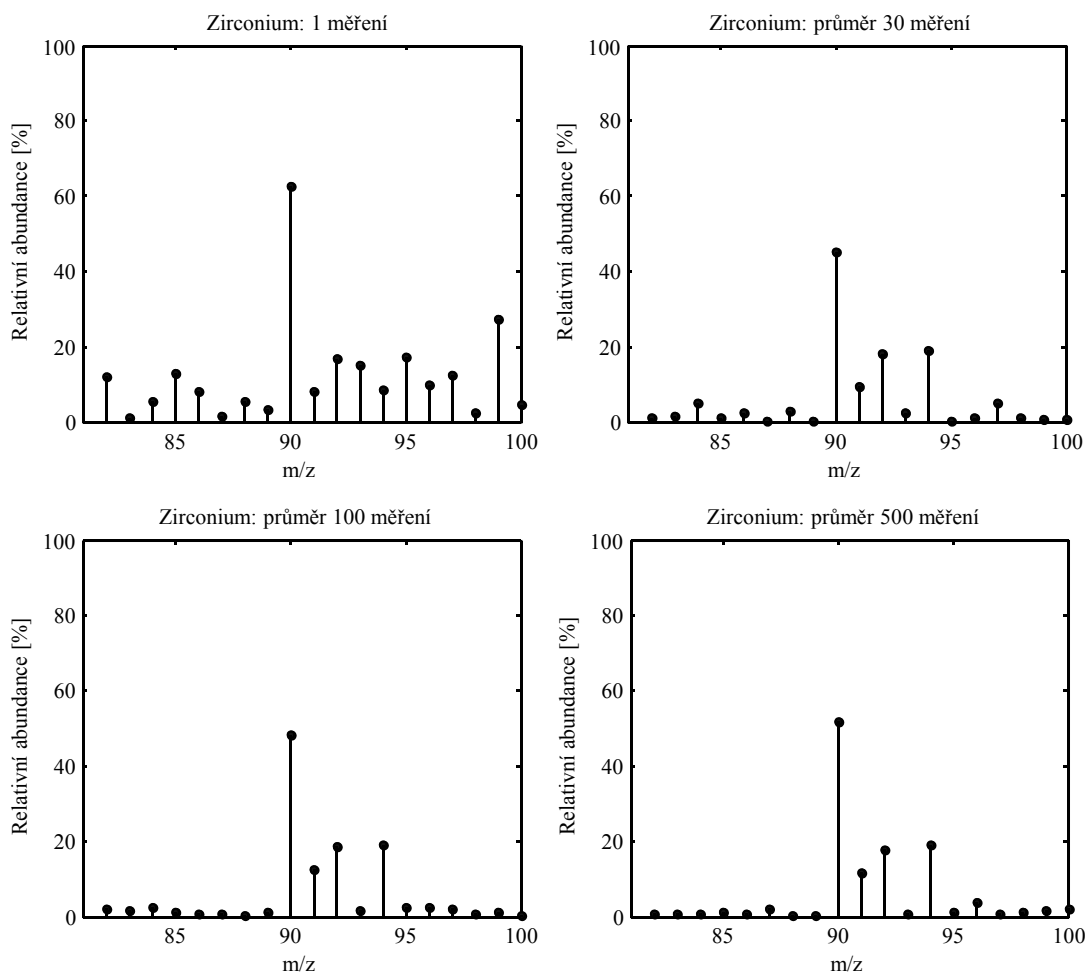
## 2.6 Shrnutí

Ve druhé kapitole byl představen koncept lineární filtrace pomocí LTI systémů, které se označují jako filtry, neboť selektivně vybírají ze zpracovávaných časových řad požadované frekvenční složky a jiné tlumí. Každý filtr je jednoznačně popsán svou impulsní charakteristikou  $h(n)$ . Odezvu filtru na libovolnou časovou řadu  $x(n)$  je možno vypočítat pomocí diskrétní konvoluce:  $y(n) = x(n) * h(n)$  a dále také pomocí diskrétní Fourierovy transformace:  $y(n) = \text{DFT}^{-1}\{X(\omega) \cdot H(\omega)\} = \text{DFT}^{-1}\{\text{DFT}\{x(n)\} \cdot \text{DFT}\{h(n)\}\}$ . Pokud má impulsní charakteristika filtru konečný počet vzorků, označuje se filtr zkratkou FIR, je vždy stabilní a může být navržen tak, aby jeho fázová frekvenční charakteristika byla lineární. Filtry s nekonečnou impulsní charakteristikou pak označujeme jako filtry IIR a tyto nemají lineární fázovou frekvenční charakteristiku, avšak poskytují strmější přechod mezi propustným a nepropustným pásmem, než je tomu v případě FIR filtrů stejného řádu. Podle realizačního algoritmu dělíme filtry na MA, AR a ARMA. Filtry MA neobsahují žádné zpětnovazební smyčky se zpožďovacími členy - na rozdíl od filtrů AR a ARMA. Zatímco přenosová funkce filtru MA má ryze polynomiální tvar, u filtrů AR a ARMA je přenosová funkce popsána racionální lomenou funkcí. Obsahuje-li realizační algoritmus jakéhokoli filtru zpětné vazby se zpožďovacími členy, je vždy nutné kontrolovat jejich stabilitu.

### 3 Kumulační zvýrazňování užitečné složky časových řad v šumu

V této kapitole jsou probrány techniky pro analýzu časových řad v časové doméně, které slouží pro zvýrazňování užitečné složky dat skryté v šumu. Tyto techniky se označují také jako zvyšování poměru signálu k šumu, metoda kumulačních součtů nebo jednoduše kumulace či průměrování (*signal averaging*). Jejich použití je obvykle svázáno s následujícími předpoklady:

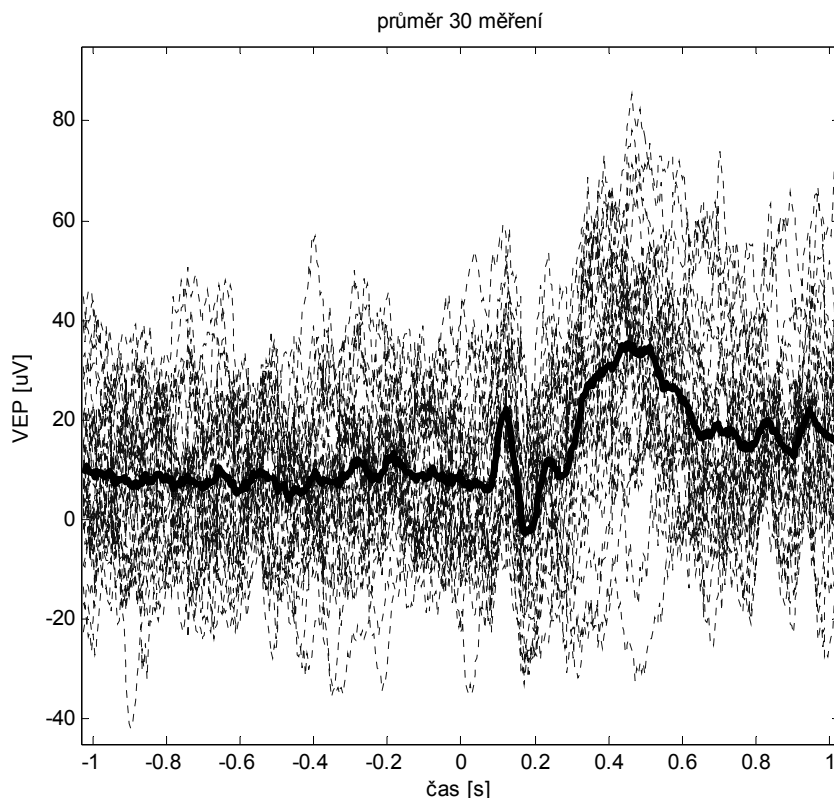
1. Užitečná a rušivá složka (šum) nejsou navzájem korelovány.
2. Časové řady mají buď repetiční povahu, anebo se jedná o opakovaná měření s konzistentní užitečnou složkou.
3. Rušivá složka je aditivní šum.
4. Šum je generován náhodným procesem s nulovou střední hodnotou.



**Obr. 3.1** Příkladová aplikace kumulačního zvýrazňování užitečné složky časové řady skryté v šumu – hmotnostní spektrometrie při identifikaci neznámých látek v analytické chemii. Proměření hmotnostního spektra iontů se opakuje obvykle 100 – 1000 krát a výsledná posloupnost spektrálních čar je počítána jako průměr ze všech získaných měření.



Každý z těchto předpokladů bývá v reálných aplikacích do jisté míry nesplněn, avšak i v takových případech se kumulační techniky pro zvýraznění užitečné složky časových řad v šumu díky své robustnosti obvykle osvědčují. Příkladové aplikace jsou ukázány na obr. 3.1 a na obr. 3.2.



**Obr. 3.2** Příkladová aplikace kumulačního zvýrazňování užitečné složky časové řady skryté v šumu – zrakové evokované potenciály v neurologii. Podíl amplitudy evokované odpovědi k amplitudě bazální aktivity centrálního nervového systému se odhaduje na desetiny až jednotky procent. Mnohočetným vybavením stejného zrakového, sluchového, motorického nebo somatosenzorického podnětu je možno díky kumulačním technikám oddělit evokované odpovědi od náhodné EEG a EMG aktivity. Jako doporučený počet repetice se uvádí čísla řádově od desítek do tisíců [2]. V příkladu na obrázku odpovídá záporný čas době před stimulem a kladný čas odpovídá době po stimulu.

### 3.1 Poměr signálu k šumu SNR

Kvantifikované hodnocení poměru užitečné složky časových řad (signálu) a složky rušivé (šumu) se obvykle označuje jako poměr signálu k šumu (*SNR – signal-to-noise ratio*) a vychází se přitom z amplitudy nebo výkonu každé z těchto složek. **Průměrný výkon** časové řady lze vyjádřit jako časový průměr čtverců amplitud (*ms – mean squared amplitude*):

$$ms_x = \frac{1}{N} \sum_{i=1}^N x(n)^2. \quad (3.1)$$

Po odmocnění výkonu získáme tzv. **efektivní amplitudu** (*rms – root of the mean squared amplitude*):

$$rms_x = \sqrt{\frac{1}{N} \sum_{i=1}^N x(n)^2} . \quad (3.2)$$

Poměr signálu k šumu *SNR* se obvykle uvádí pomocí logaritmické decibelové stupnice:

$$SNR = 10 \log_{10} \frac{ms_{signál}}{ms_{šum}} \text{ dB} \quad (3.3)$$

a alternativně lze pomocí substituce  $ms = rms^2$  používat také tento vztah:

$$SNR = 10 \log_{10} \left[ \frac{rms_{signál}}{rms_{šum}} \right]^2 = 20 \log_{10} \frac{rms_{signál}}{rms_{šum}} \text{ dB} . \quad (3.4)$$

## 3.2 Repetiční časové řady

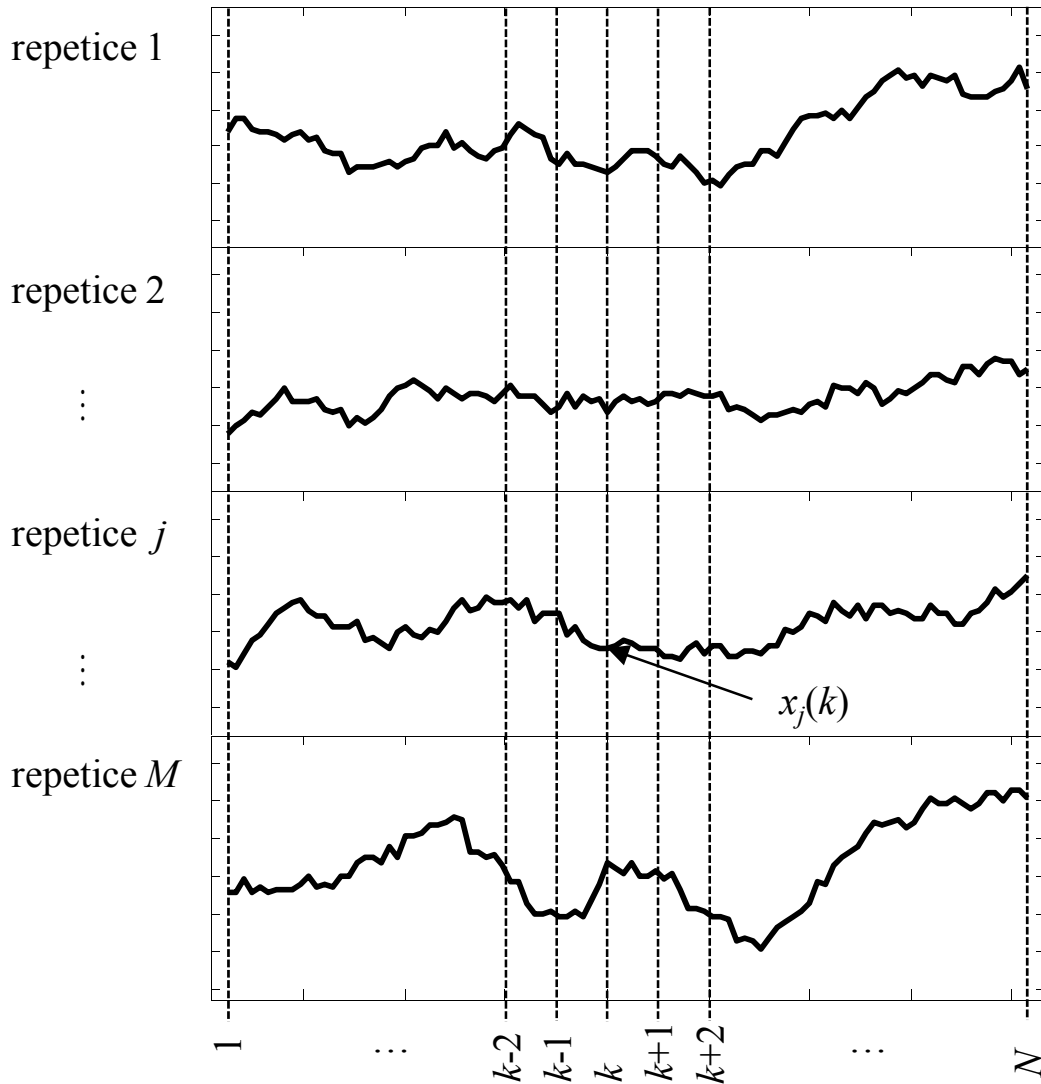
Kumulační techniky se nejčastěji uplatňují u tzv. **repetičních časových řad**. Ty jsou charakteristické tím, že se v nich po nekonstantní době opakuje stejná, časově omezená posloupnost hodnot, která se označuje jako jedna repetice. Prakticky se lze setkat se dvěma různými typy repetičních časových řad:

1. časové řady s repetiční povahou, které nelze považovat za perfektně periodické kvůli mezirepetiční variabilitě a nekonstantnímu zpoždění mezi jednotlivými repeticemi – např. diskrétní signály získané při měření aktivity srdce (signály EKG);
2. časové řady získané opakovaným měřením stejného jevu – např. evokované potenciály v neurologii, viz obr. 3.2.

U repetičních časových řad prvního typu, které se někdy označují jako quasi-periodické, je nutno zajistit koherenci jednotlivých repetic neboli „lícování“ navzájem si odpovídajících vzorků jednotlivých repetic, viz obr. 3.3. Stanovení referenčních souřadnic na časové ose lze provést např. jedním z těchto postupů (nebo pomocí jejich kombinace).

- A. **Prahování.** Za referenční časovou souřadnici se stanoví ta, na které hodnota posloupnosti v repetici poprvé přesáhne předdefinovaný práh.
- B. **Detekce podle strmosti.** Za referenční časovou souřadnici se stanoví ta, která je ve středu intervalu, přes který dochází v posloupnosti k požadované změně. Vyhledávají se obvykle souřadnice s maximální strmostí (náběžné nebo sestupné hrany impulsů či vln – rychlé změny) nebo s minimální strmostí (plató vln – velmi pomalé nebo žádné změny).
- C. **Detekce podle vzoru.** Za referenční časovou souřadnici se stanoví ta, od které je posloupnost svými hodnotami nejbližší hodnotám předdefinované vzorové posloupnosti. Vzorem může být např. vybraná reprezentativní repetice nebo i uměle sestavená posloupnost, která má svým průběhem odpovídat užitečné složce časové řady v jedné repetici. Pro výpočet podobnosti lze použít celou řadu metrik – v nejjednodušším případě

půjde zřejmě o nalezení minima sumy čtverců rozdílů mezi vzorovou posloupností a odpovídajícími vzorky ze zpracovávané časové řady.



**Obr. 3.3** Sada  $M$  repetič časové řady obsahující užitečnou a rušivou složku. Každá repetice je dána posloupností navzájem si odpovídajících vzorků  $x_j(k)$ , kde  $k=1,2,\dots,N$ .

### 3.3 Repetiční časové řady s náhodným rušením

Repetiční časovou řadu lze v případě aditivního rušení vyjádřit jako součet užitečné složky  $s$  a rušivé složky  $v$ , které jsou navzájem nezávislé (takový byl předpoklad stanovený na začátku této kapitoly). Při splnění podmínky časové koherence repetič lze  $k$ -tý vzorek  $j$ -té repetice vyjádřit součtem (viz také obr. 3.3):

$$x_j(k) = s_j(k) + v_j(k), \quad k = 1, 2, \dots, N \quad (3.5)$$

a posloupnost kumulovaných (či váženě zprůměrovaných) vzorků v jedné repetici se vyjádří jako lineární kombinace  $M$  repetic:

$$x(k)_M = \sum_{j=1}^M a_j x_j(k) = \sum_{j=1}^M a_j [s_j(k) + v_j(k)], \quad (3.6)$$

kde  $a_1, a_2, \dots, a_M$  jsou váhy jednotlivých repetic – např. pro prosté průměrování <sup>10</sup>platí, že  $a_j = 1/M$  pro  $j=1, 2, \dots, M$ . Pro stanovení, jak se zlepší poměr užitečné složky ke složce rušivé, tj. zlepšení  $SNR$ , lze rovnici (3.6) přepsat do tvaru:

$$x(k)_M = s_j(k) \sum_{j=1}^M a_j + \sum_{j=1}^M a_j v_j(k). \quad (3.7)$$

Z prvního členu pravé strany rovnice (3.7) lze vyčíst změnu užitečné složky  $s_j(k)$ , a sice na  $\sum_{j=1}^M a_j$  - násobky původních hodnot jejich vzorků. Druhý člen v rovnici (3.7) pak vyjadřuje lineární kombinaci  $M$  realizací náhodného šumu – konkrétní změnu v zastoupení rušivé složky v kumulované časové řadě tedy nelze z rovnice (3.7) přímo vyjádřit. Za předpokladu, že šum je generován náhodným procesem s nulovou střední hodnotou, lze očekávat, že pro nekonečně velký počet navzájem nekorelovaných repetic  $v_j(k), j=1, 2, \dots, \infty$  bude rušivá složka ve výsledné kumulované repetici redukována tak že  $x(k)_M \rightarrow s(k)_M$ . V reálných podmínkách je však  $M < \infty$  a zbytková rušivá složka může být nejlépe charakterizována svým rozptylem. Pro rozptyl lineární kombinace náhodných veličin z druhého členu rovnice (3.7) platí:

$$Var\left(\sum_{j=1}^M a_j v_j\right) = \sum_{j=1}^M a_j^2 \sigma_j^2 + \sum_i \sum_j a_i a_j \sigma_{ij}, \quad (3.8)$$

kde  $\sigma_j^2$  jsou rozptyly jednotlivých náhodných veličin a  $\sigma_{ij}$  jsou kovariance dvojic členů kombinace [12]. Za předpokladu, že je šum generován stacionárním procesem, bude jeho rozptyl stejný ve všech repeticích  $\sigma_j^2 = \sigma^2, j=1, 2, \dots, M$ . V takovém případě lze pro rozptyl zbytkové rušivé složky  $\sigma_M^2$  po kumulaci přepsat vztah (3.8) na:

$$\sigma_M^2 = \sigma^2 \sum_{j=1}^M a_j^2 + \sum_i \sum_j a_i a_j \sigma_{ij}. \quad (3.9)$$

Je-li splněn úvodní předpoklad, že realizace šumu v jednotlivých repeticích jsou na sobě nezávislé, pak druhý člen v rovnici (3.9) bude nulový a po kumulaci se změní rozptyl rušivé složky na  $\sum_{j=1}^M a_j^2$  - násobek původního rozptylu. Pokud tento předpoklad splněný není, rozptyl

---

<sup>10</sup> Kumulace s rovnoměrnými vahami.

zbytkové rušivé složky je vyšší, a zhoršuje se tak výsledek kumulačního zvýrazňování užitečné složky časové řady.

Z rovnic (3.1), (3.2), (3.7) a (3.9) lze odvodit výsledné výkonové zlepšení poměru signálu k šumu  $K_{ms}$ :

$$K_{ms} = \frac{\left( \sum_{j=1}^M a_j \right)^2}{\sum_{j=1}^M a_j^2}, \quad (3.10)$$

a amplitudové zlepšení poměru signálu k šumu  $K_{rms}$ :

$$K_{rms} = \frac{\sum_{j=1}^M a_j}{\sqrt{\sum_{j=1}^M a_j^2}}. \quad (3.11)$$

Z rovnic (3.10) a (3.11) je evidentní, že zlepšení poměru signálu k šumu s využitím kumulačních technik je závislé pouze na počtu repetice  $M$  a na konfiguraci váhových koeficientů  $a_j$ . Je důležité poznamenat, že zlepšení  $K_{ms}$  a  $K_{rms}$  jsou zlepšení průměrná, která budou dosažitelná pouze při velkém počtu repetice, tj. při velkém počtu realizací náhodného šumu. Pro nízký počet repetice je nutné počítat s tím, že některé vzorky ve výsledné repetici po kumulaci mohou být zatíženy třeba i větší chybou než před aplikací kumulačního zvýrazňování užitečné složky časových řad [12].

### 3.3 Odhad zbytkové rušivé složky $\pm$ průměrováním

Základním cílem kumulačního zpracování je sice zvýraznění užitečné složky časových řad ztracených v šumu, ale v některých případech může být hlavní idea této metody využita i pro odhad zbytkové rušivé složky a její oddělení od složky užitečné – např. za účelem stanovení statistických vlastností náhodného šumu. Lze k tomu využít tzv.  $\pm$  průměrování, což je zvláštní kumulační metoda, při které jsou váhové koeficienty  $a_j$  v (3.6) nastaveny na aritmetické průměrování a hodnoty vzorků každé druhé repetice jsou do kumulačního součtu přičítány s opačným znaménkem:

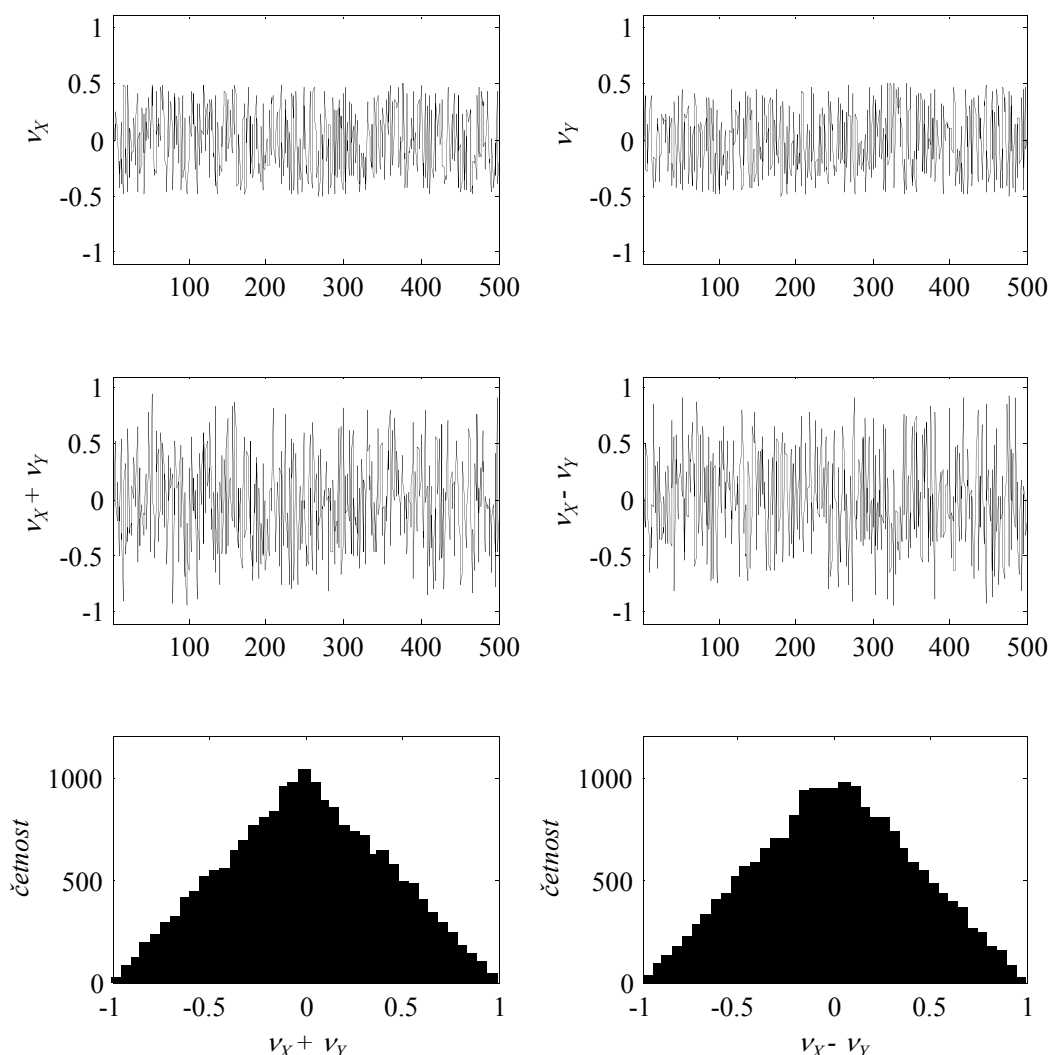
$$a_j = \begin{cases} +\frac{1}{M}, & j = 2, 4, 6 \dots M \\ -\frac{1}{M}, & j = 1, 3, 5, \dots M-1 \end{cases} \quad \text{pro sudé } M, \quad (3.12)$$

$$a_j = \begin{cases} +\frac{1}{M}, & j = 2, 4, 6 \dots M-1 \\ -\frac{1}{M}, & j = 1, 3, 5, \dots M \end{cases} \quad \text{pro liché } M.$$

Zatímco konzistentní užitečná složka bude střídavým přičítáním a odečítáním potlačena, zbytková rušivá složka bude mít stejnou efektivní amplitudu  $rms$ , jako by měla po běžném zprůměrování, poněvadž náhodný šum má v invertovaných repetičích stejné rozdělení jako v repetičích neinvertovaných, viz ukázka (nikoli důkaz) na obr. 3.4 [6].

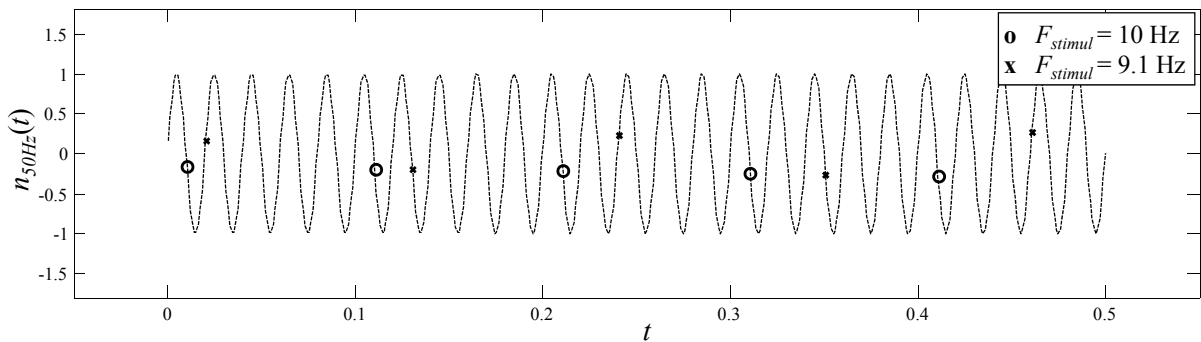
### 3.4 Repetiční časové řady s nenáhodným rušením

Zlepšení poměru signálu k šumu odvozené v předchozí kapitole bylo založeno na předpokladech, že rušivá složka zpracovávaných časových řad je realizací náhodného procesu s nulovou střední hodnotou a je nezávislá na užitečné složce. V některých případech, které však nejsou nijak neobvyklé, nemusí být rušivá složka realizací náhodného procesu. Nejčastějším příkladem takového nenáhodného rušení je nežádoucí výskyt tzv. síťového rušení, což je harmonický signál o frekvenci 50 Hz nebo 60 Hz (v závislosti na regionu) pronikající do naměřených dat ze zdrojů napájení při měření.



**Obr. 3.4** Realizace náhodného šumu  $v_X$  a  $v_Y$ , jejich součet ( $v_X + v_Y$ ) a rozdíl ( $v_X - v_Y$ ). Histogramy dole jsou pro součet a pro rozdíl podobné, což může ukazovat na to, že jsou obě tyto náhodné časové řady charakterizovány stejným rozdělením pravděpodobnosti. V případě náhodného šumu sice operace sčítání a odečítání vytvářejí rozdílné časové řady (co do amplitud), ale nezpůsobují změnu statistických vlastností šumu [6].

U časových řad získaných opakovaným měřením stejného jevu, kterým je odpověď na stimul určitého senzorického systému, je důležité si uvědomit, že kumulováním naměřených odpovědí z jednotlivých repetic bude zvýrazněna nejen užitečná složka, ale také jakékoli periodické komponenty s přímým vztahem k frekvenci stimulů [6]. Je-li např. experiment nastaven na frekvenci stimulů 50 Hz, pak bude jistě aplikací kumulačního zpracování naměřených repetic zvýrazněna rušivá složka o frekvenci 50 Hz, i když třeba byla v původních repeticích zanedbatelná. Problém s frekvencí stimulů je však ještě horší, neboť jakákoli frekvence, která beze zbytku dělí 50, vyústí při kumulačním zpracování ve zvýraznění rušivé složky o frekvenci 50 Hz (viz např. 10 Hz frekvence stimulů na obr. 3.5). Tento nepříjemný jev se bude uplatňovat přirozeně i u jiných harmonických rušivých složek a je možné mu zabránit buď použitím neceločíselné frekvence stimulů, viz např. 9.1 Hz na obr. 3.5, nebo randomizací intervalů mezi stimuly. Náhodné mezirepetiční intervaly vedou k náhodné fázi rušivé složky vzhledem k počátkům repetic, a tak i k potlačení harmonických rušivých složek kumulačním zpracováním.



**Obr. 3.5** Nevhodně zvolená frekvence stimulů může při výskytu periodické rušivé složky (např. 50 Hz síťové rušení) způsobit při kumulačním zpracování nechtěný efekt. Při kumulaci repetic získaných s frekvencí stimulů 10 Hz bude výsledek obsahovat výraznou rušivou komponentu odvozenou z 50 Hz síťového rušení a při kumulaci repetic s frekvencí stimulů 9.1 Hz tento nechtěný jev nenastane. Rozdíl tkví v tom, že při aplikaci frekvence stimulů 10 Hz jsou začátky rušivých složek v jednotlivých repeticích ve stejné fázi, zatímco při použití frekvence stimulů 9.1 Hz se fáze rušivé složky pro každý stimul liší [6].

### 3.5 Kumulační zvýrazňování s rovnoměrnými váhami

Zlepšení poměru signálu k šumu je podle rovnic (3.10) a (3.11) závislé na počtu repetic  $M$  a na nastavení váhových koeficientů  $a_j$ . V případě kumulačního zvýrazňování s rovnoměrnými váhami bývají koeficienty  $a_j$  nastaveny tak, aby kumulační zpracování nevedlo ke změně efektivní amplitudy  $rms$ :

$$a_j = \frac{1}{M}, \quad j = 1, 2, \dots, M \quad (3.13)$$

Po dosažení rovnoměrných vah do vztahu pro výpočet zlepšení poměru signálu k šumu (3.10) a (3.11):

$$K_{ms} = \frac{\left( \sum_{j=1}^M \frac{1}{M} \right)^2}{\sum_{j=1}^M \left( \frac{1}{M} \right)^2} = M, \quad (3.14)$$

$$K_{rms} = \frac{\sum_{j=1}^M \frac{1}{M}}{\sqrt{\sum_{j=1}^M \left( \frac{1}{M} \right)^2}} = \sqrt{M} \quad (3.15)$$

je zřejmé, že výkonové zlepšení  $K_{ms}$  roste lineárně s počtem repetíc a amplitudové zlepšení  $K_{rms}$  roste s odmocninou počtu repetíc.

Dynamicke vlastnosti kumulačního zpracování s rovnoměrnými vahami se liší podle způsobu implementace, přičemž v literatuře [12] se lze setkat s kumulačními algoritmy využívajícími pevného nebo klouzavého okna. Algoritmus kumulace s pevným oknem se hodí pro případy, kdy je třeba jednorázově zvýraznit užitečnou složku časových řad ztracených v šumu. Celkem  $N$  akumulacích proměnných  $x(k)_M$  se díky pevně nastavenému oknu vynuluje vždy po  $M$  repeticích a kumulační zpracování začíná jakoby znovu.

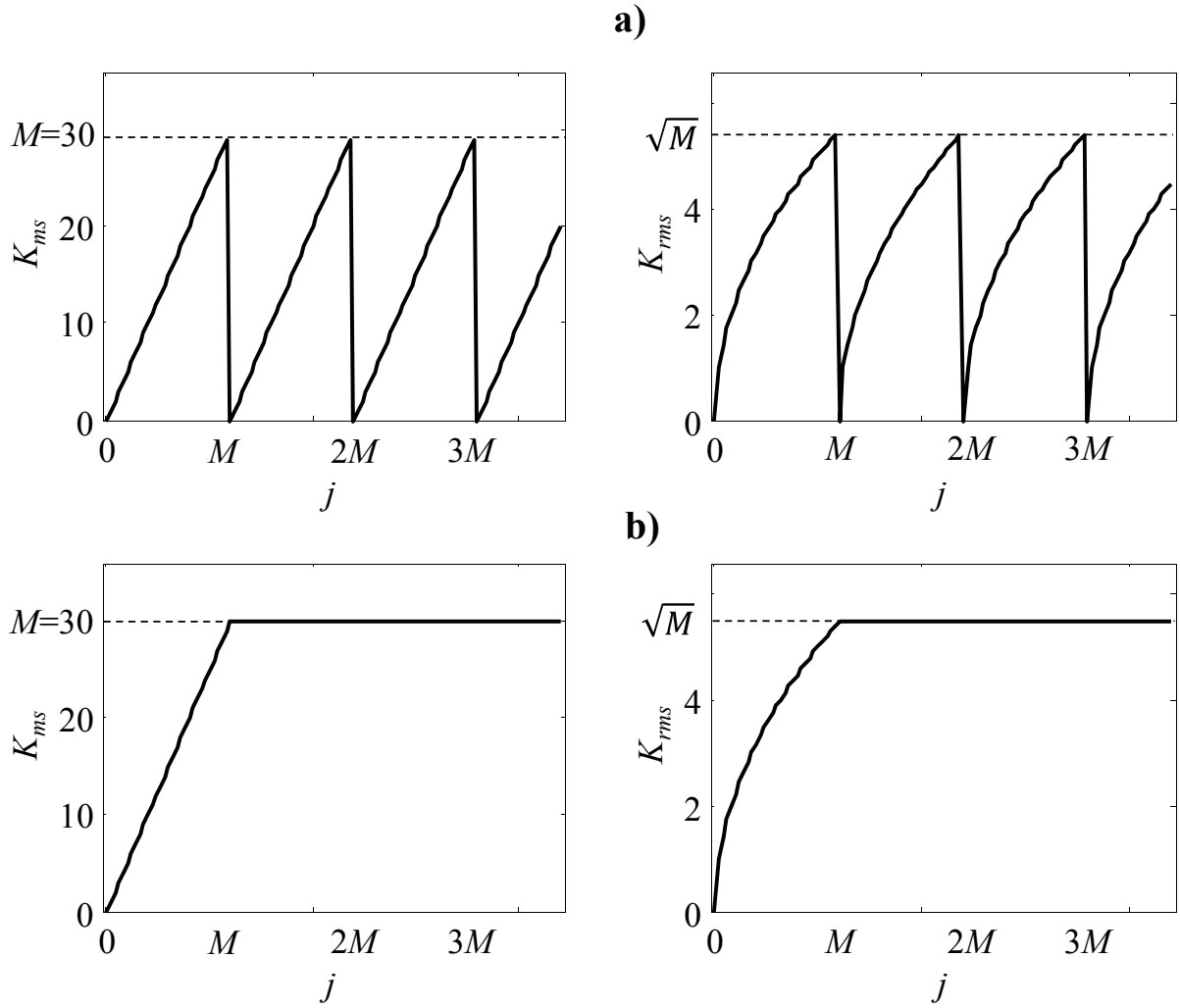
Pro kontinuální sledování repetičních časových řad je lepší použít algoritmus s klouzavým oknem, který je sice náročnější na paměť, avšak tento fakt v době počítačového zpracování dat nepředstavuje výraznější bariéru. Je vhodné si uvědomit, že u algoritmu s klouzavým oknem se musí v paměti uchovávat  $M$  repetíc, což představuje paměťové pole o  $M \times N$  proměnných (pro  $M$  posloupností o délce  $N$ ). Po té, co se během kumulačního zpracování naplní paměťové pole pro  $M$  repetíc, dochází s každou další repeticí k odstranění nejstarší repetice z pole, k posunu všech zbylých  $M-1$  repetíc v poli směrem dozadu a k nahrazení repetice na prvním místě v paměťovém poli právě repeticí nejaktuálnější.

### 3.6 Kumulační zvýrazňování s exponenciálními vahami

Oba diskutované algoritmy kumulačního zvýrazňování s rovnoměrnými vahami pracují tak, že se ve výsledné kumulované repetici uplatňuje každá zpracovávaná repetice se stejnou vahou po dobu trvání  $M$  repetíc, po té je zcela zapomenuta a více se s ní již nepracuje. Při kumulačním zvýrazňování s exponenciálními vahami se ve výsledku uplatňuje každá zpracovaná repetice, a to s klesající vahou do minulosti [12]:

$$a_j = \alpha^{M-j}, \quad 0 < \alpha < 1, \quad j = 1, 2, \dots, M \quad (3.16)$$





**Obr. 3.6** Dynamické vlastnosti kumulačního zvýrazňování s rovnoměrnými vahami pro  $M=30$ :  
a) s pevným oknem, b) s klouzavým oknem.

Po dosažení exponenciálních vah do vztahu pro výpočet zlepšení poměru signálu k šumu (3.10) a (3.11) lze při znalosti vzorce pro výpočet součtu geometrické řady dospět k:

$$\begin{aligned}
 K_{ms} &= \frac{\left( \sum_{j=1}^M a_j \right)^2}{\sum_{j=1}^M a_j^2} = \frac{\left( \sum_{j=1}^M \alpha^{M-j} \right)^2}{\sum_{j=1}^M (\alpha^{M-j})^2} = \frac{\left( \sum_{j=0}^{M-1} \alpha^j \right)^2}{\sum_{j=0}^{M-1} \alpha^{2j}} = \frac{\left( \frac{1-\alpha^M}{1-\alpha} \right)^2}{\frac{1-\alpha^{2M}}{1-\alpha^2}} = \\
 &= \frac{(1-\alpha^M)^2 (1+\alpha)(1-\alpha)}{(1-\alpha)^2 (1-\alpha^M)(1+\alpha^M)} = \frac{1-\alpha^M}{1+\alpha^M} \frac{1+\alpha}{1-\alpha}, \\
 \text{pro } M \gg 1 \Rightarrow \alpha^M \rightarrow 0 \Rightarrow K_{ms} &\cong \frac{1+\alpha}{1-\alpha},
 \end{aligned} \tag{3.17}$$

$$K_{rms} = \sqrt{\frac{1-\alpha^M}{1+\alpha^M}} \sqrt{\frac{1+\alpha}{1-\alpha}}, \text{ pro } M \gg 1 \Rightarrow \alpha^M \rightarrow 0 \Rightarrow K_{rms} \cong \sqrt{\frac{1+\alpha}{1-\alpha}}, \tag{3.18}$$

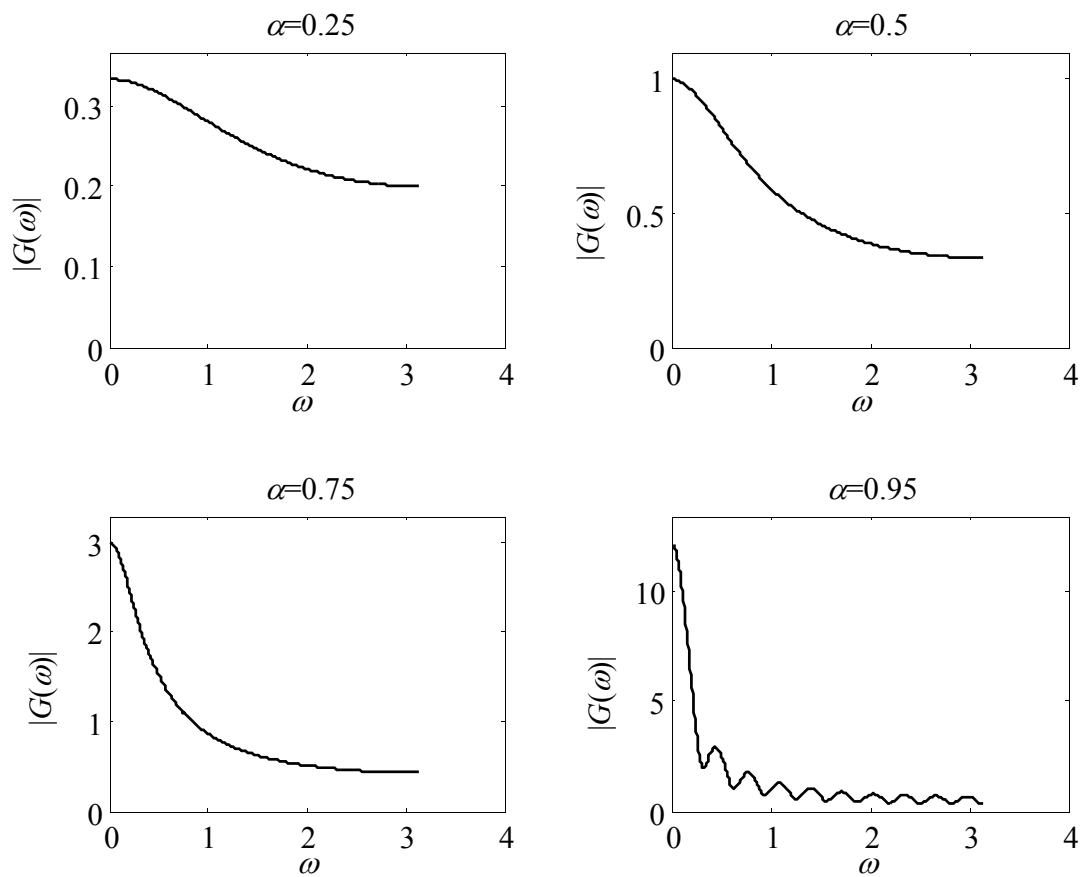
kde zjednodušené vztahy pro výkonové zlepšení poměru signálu k šumu  $K_{ms}$  a pro amplitudové zlepšení poměru signálu k šumu  $K_{rms}$  lze aplikovat při vysokých počtech zpracovaných repetíc  $M$ . Lze tedy usoudit, že pro velké počty repetíc  $M$  bude výsledek exponenciálního kumulačního zvýrazňování závislý pouze na koeficientu  $\alpha$ . Za účelem dosažení co největšího průměrného zlepšení poměru signálu k šumu bude snahou nastavit koeficient  $\alpha$  na číslo blízké se k 1, avšak je nutné si uvědomit, že v takových případech se na výsledné kumulaci nezanedbatelně podílí mnohem více repetíc, a tedy i mnohem více starších repetíc, než v případě, kdy je  $\alpha$  nižší, a kdy tedy vliv starších repetíc vlivem geometrické řady v (3.16) rychle klesá. Volba koeficientu  $\alpha$  bude proto vždy kompromisem mezi dosažitelným zlepšením poměru signálu k šumu a schopností exponenciálního kumulačního zvýrazňování reagovat na eventuální změny<sup>11</sup> probíhající v užitečné složce zpracovávaných časových řad. Volbu hodnoty koeficientu  $\alpha$  lze např. podmínit požadavkem na stejné průměrné zlepšení poměru signálu k šumu, jako je dosažitelné v případě kumulační metody s rovnoměrnými vahami a předvoleným počtem repetíc  $M$ :

$$\sqrt{M} = \sqrt{\frac{1+\alpha}{1-\alpha}} \quad \Rightarrow \quad \alpha = \frac{M-1}{M+1}. \quad (3.19)$$

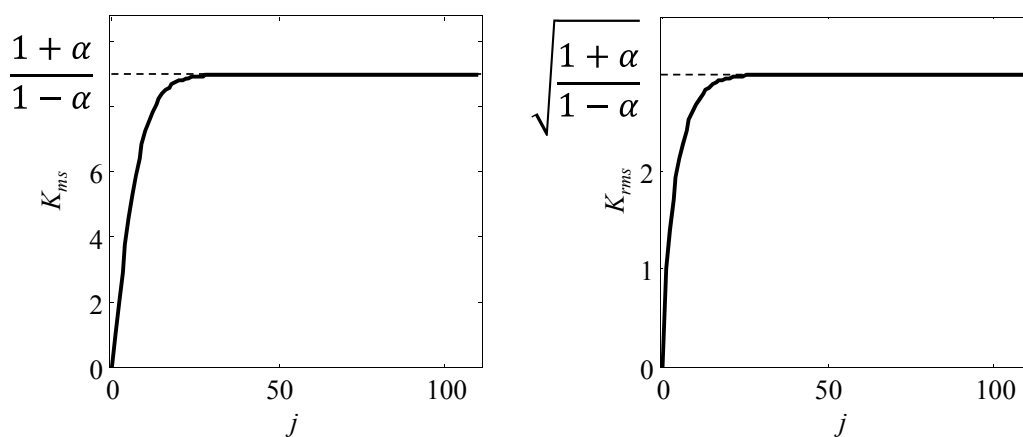
Jiným východiskem pro volbu hodnoty koeficientu  $\alpha$  může být i posouzení jeho vlivu na vlastnosti kumulačního zpracování ve frekvenční doméně. Pro konečný počet  $M$  repetíc odpovídá posloupnost vah  $a_j$  podle (3.16) hodnotám impulsní charakteristiky FIR filtru řádu  $M$ , kterým je v rámci kumulačního zpracování vlastně filtrováno  $N$  posloupností  $k$ -tých vzorků  $x_j(k)$  z jednotlivých repetíc. Obr. 3.7 ukazuje vliv hodnoty  $\alpha$  na tvar modulové frekvenční charakteristiky takového filtru. Z obrázku je patrné, že uvedený filtr vykazuje vlastnosti dolní propusti, u které hodnota  $\alpha$  ovlivňuje šířku propustného pásma. Při zvyšujícím se počtu repetíc  $M$  se prodlužuje impulsní charakteristika filtru. Pokud roste počet repetíc do nekonečna, je možné filtraci  $N$  posloupností  $k$ -tých vzorků  $x_j(k)$  z jednotlivých repetíc zajistit rekurzivním algoritmem, kterému odpovídá filtr IIR s přenosovou funkcí  $H(z) = \alpha/(1-\alpha z^{-1})$ . Frekvenční charakteristiky takového filtru jsou zobrazeny pro různé hodnoty koeficientu  $\alpha$  v další kapitole na obr. 4.9. Pro vyšší hodnoty  $\alpha$  filtr lépe potlačuje složky, které představují mezirepetiční variabilitu a má tendenci propustit pouze ty složky, které se mezi repeticemi nemění vůbec nebo jen velmi pozvolna.

Na obr. 3.8 jsou znázorněny dynamické vlastnosti kumulačního zvýrazňování s exponenciálními vahami – přímé srovnání s dynamickými vlastnostmi kumulačního zvýrazňování s rovnoměrnými vahami je předmětem řešené úlohy 3.1. Porovnáním obr. 3.6 a obr. 3.8 lze dojít k závěru, že se exponenciální kumulace svými vlastnostmi více podobá rovnoměrné kumulaci s klouzavým oknem a má přitom velmi jednoduchý algoritmus, který nevyžaduje složité operace posunu v paměťovém poli repetíc, jako je tomu u algoritmu rovnoměrné kumulace s klouzavým oknem a přitom umožňuje, stejně jako rovnoměrná kumulace s klouzavým oknem, sledování pomalých změn v užitečné složce zpracovávaných časových řad.

<sup>11</sup> Při kumulačním zvýrazňování užitečné složky repetičních časových řad se obvykle předpokládá, že se užitečná složka mezi repeticemi nemění, nicméně v reálných aplikacích lze pomalé změny v užitečné složce očekávat, a na tyto trendy by mělo proto reagovat i kumulační zpracování.



**Obr. 3.7** Modulové frekvenční charakteristiky FIR filtru řádu  $M=20$  s impulsní charakteristikou  $h_j = \alpha^{M-j+1}$  pro různé hodnoty  $\alpha$ .



**Obr. 3.8** Dynamické vlastnosti kumulačního zvýrazňování s exponenciálními vahami pro  $\alpha=0,8$ .

### 3.7 Úlohy

**ÚLOHA 3.1** Simulujte měření 1000 časových řad představujících směsi užitečné a rušivé složky. Naměřené hodnoty následně použijte při kumulačním zpracování s rovnoměrnými i exponenciálními vahami. Pro generátor repetitivních měřených časových řad vytvořte funkci, jejímž výstupem bude směs užitečné složky a náhodné rušivé složky a dále i obě složky separátně. Užitečná složka má sinusový průběh s lineárně narůstající frekvencí – tzv. chirp signál. Rušivou složkou bude šum s uniformním rozdělením pravděpodobnosti a nulovou střední hodnotou. Poměr šumu ve směsi definujte pomocí  $SNR$  v decibelech. Simulovaná data využijte pro ověření teoretických poznatků z 3. kapitoly v následujících dílčích úkolech:

- Vykreslete výslednou repetitivní kumulaci s rovnoměrnými vahami a porovnejte s průběhy repetitivních před kumulací.
- Vykreslete výslednou repetitivní kumulaci s exponenciálními vahami nastavenými tak, aby výsledné zlepšení poměru signál šum bylo stejné jako v případě rovnoměrné kumulace.
- Vykreslete dynamické vlastnosti obou kumulačních metod a popište, jak se obě metody liší. U rovnoměrné kumulace zvažujte obě varianty – s pevným a s plovoucím oknem.
- Pro případ rovnoměrné kumulace odhadněte zbytkovou rušivou složku a zjistěte, zda vypočtené průměrné zlepšení poměru signálu k šumu odpovídá zlepšení vypočtenému z experimentálních dat.

Kód funkce pro generování časových řad s lineárně rostoucí frekvencí je na obr. 3.9. Amplituda šumu generovaného funkcí  $rand()$  je odvozena z rovnice (3.3) pro  $SNR$ :

$$ms_{sum} = \frac{mssignál}{10^{\frac{SNR[dB]}{10}}}, \quad rms_{sum} = \sqrt{ms_{sum}}.$$

```
%%% ULOHA 3.1 %%%%%%%%% GENERÁTOR CHIRP SIGNÁLU

function [x,t,s,n]=chirpandnoise(tstart,tend,fstart,fend,Fs,As,snr)
%
% [x,t,s,n]=chirpandnoise(tstart,tend,fstart,fend,Fs,A)
%
% generator sinusove casove rady s linearne narustajici
% frekvenci a s nahodnym aditivnim rusenim
% tstart - zacatek casove rady [s]
% tend - konec casove rady [s]
% fstart - okamzita frekvence v case tstart
% fend - okamzita frekvence v case tend
% Fs - vzorkovaci frekvence
% As - amplituda uzitecne sinusovky
% snr - pomer signalu k sumu v dB
% x - vysledna smes uzitecne slozky a ruseni
% t - casova osa
% s - uzitecna slozka
% n - ruseni

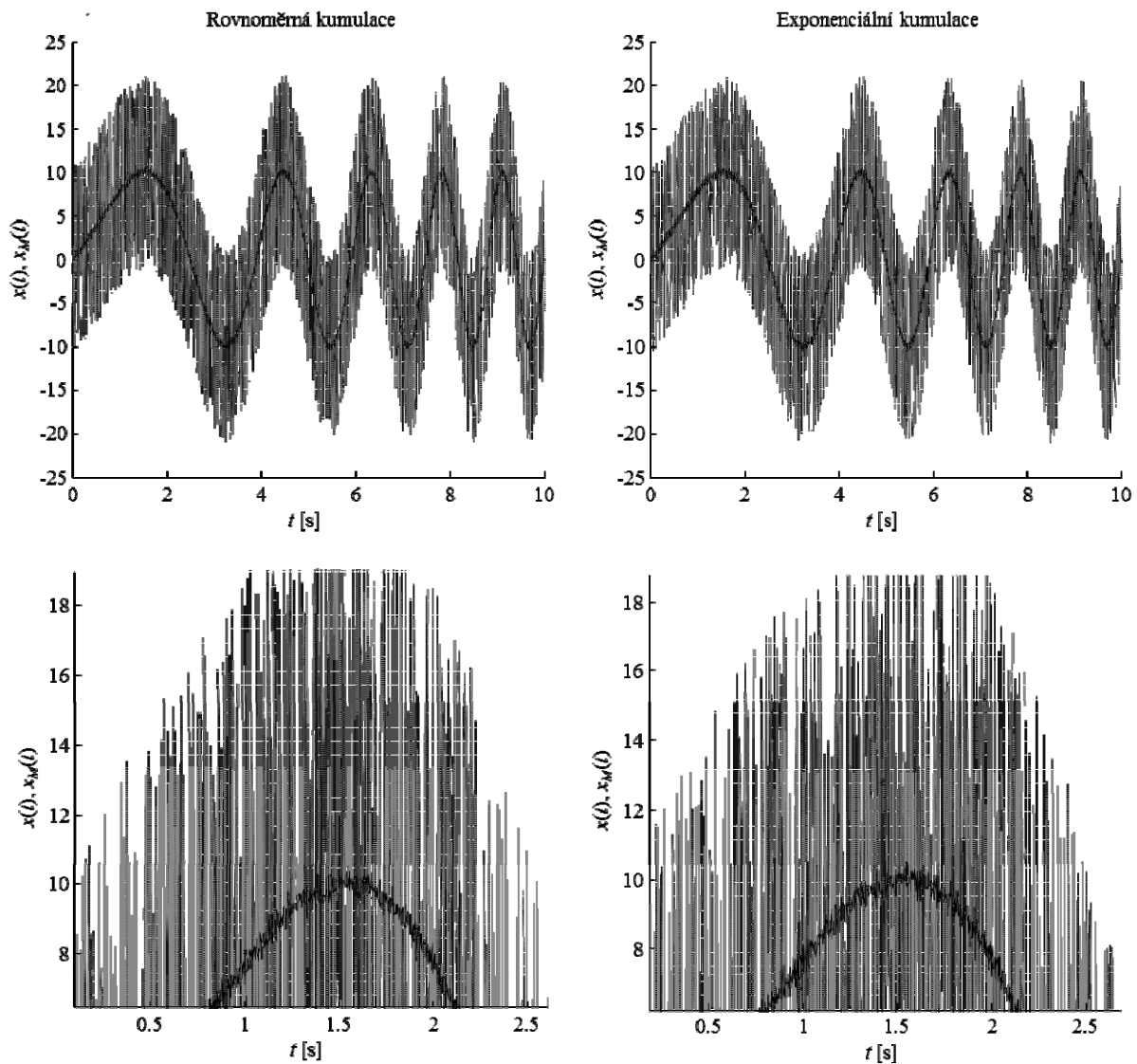
t = tstart:1/Fs:tend; N = length(t);
f = linspace(fstart,fend,N); w = 2*pi*f;
s = As*sin(w.*t);
```

```

ms_s = sum(s.*s,2)/N;
ms_n = ms_s/(10^(snr/10));
n = sqrt(ms_n)*(rand(1,N)-0.5);
x = s+n;

```

**Obr. 3.9** Úloha 3.1 – funkce pro generování časových řad s lineárně rostoucí frekvencí a aditivním šumem s nulovou střední hodnotou a uniformním rozdělením pravděpodobnosti.

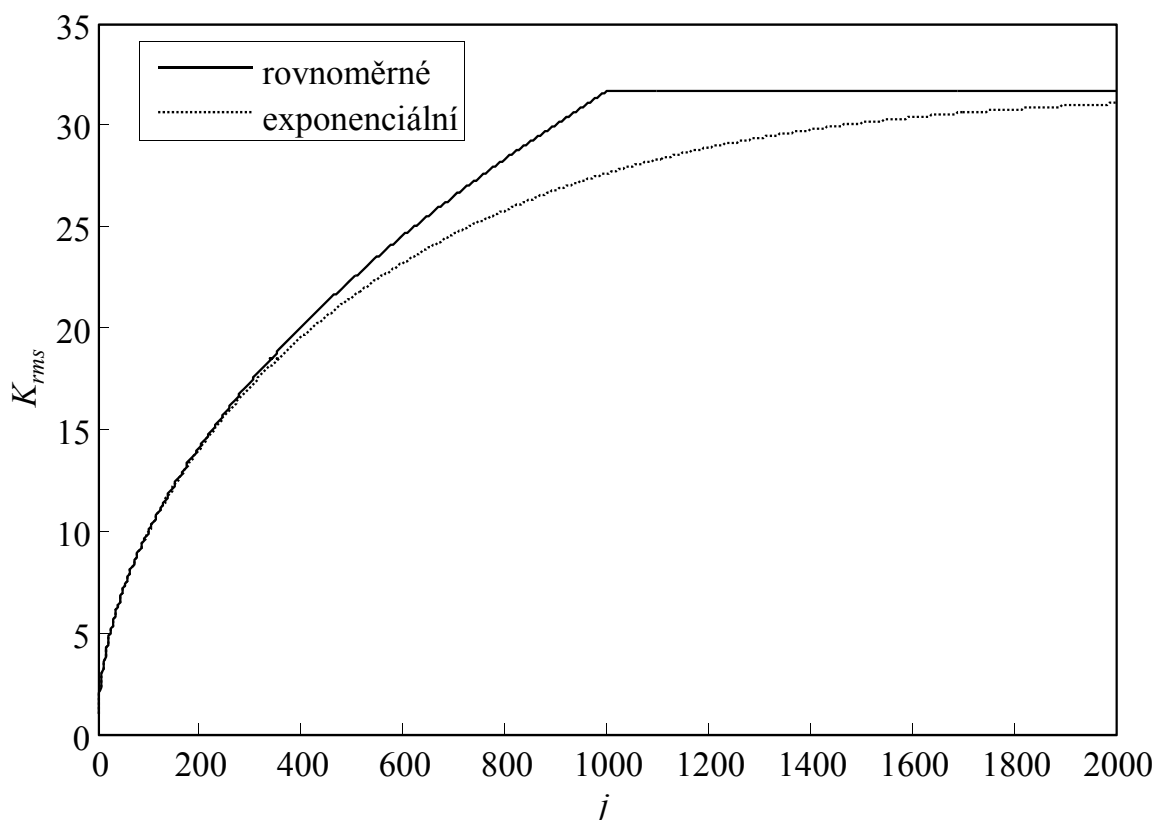


**Obr. 3.10** Úloha 3.1 – slabou tečkovanou čarou jsou vykresleny vybrané jednotlivé repetice před kumulací a silnou plnou černou čarou je vykreslena zvýrazněná užitečná složka: vlevo pomocí kumulace s rovnoměrnými vahami,  $M=1000$  a vpravo pomocí kumulace s exponenciálními vahami,  $M=1000$ ,  $\alpha=(M-1)/(M+1)$ ). Nahoře jsou průběhy celé repetice, dole jsou pak vybrané detaily pro ilustraci, že výsledky obou metod se okometricky neliší.

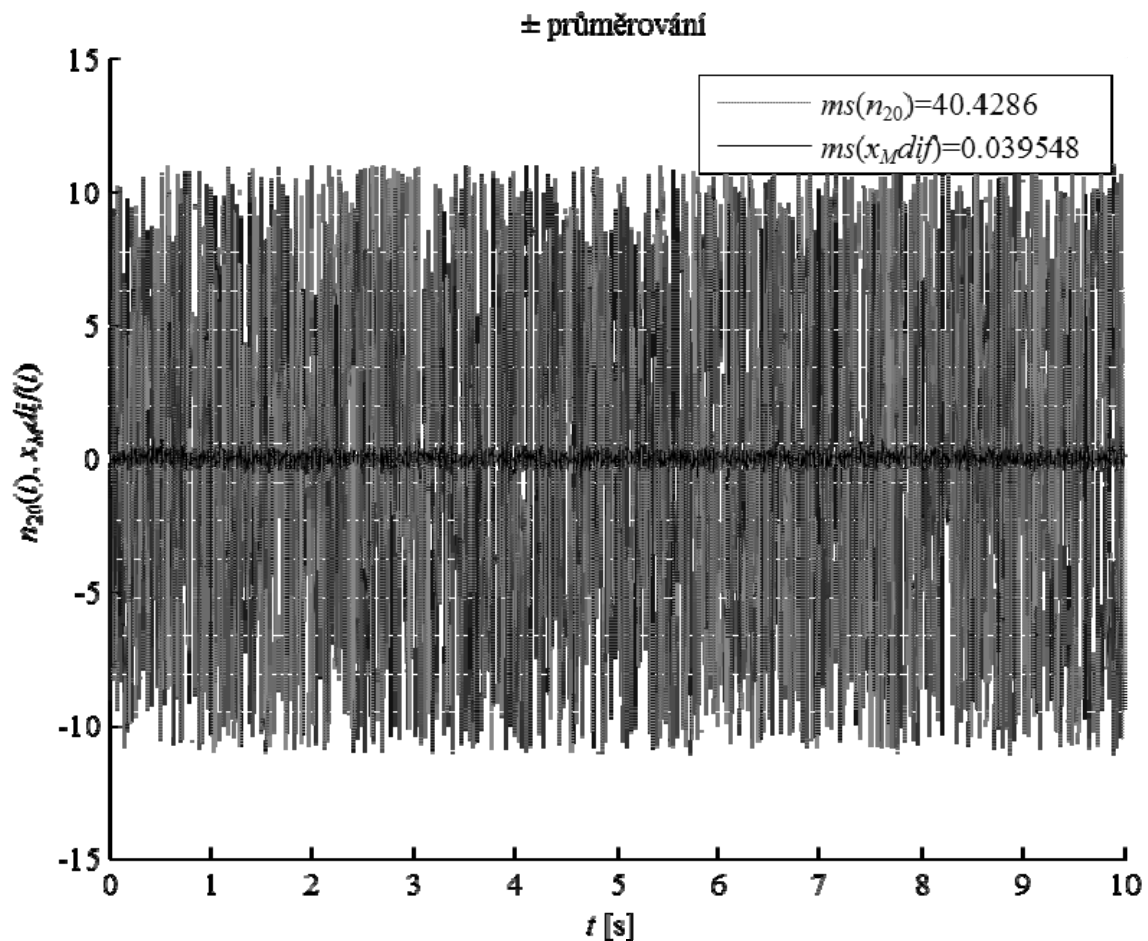
Kód programu v Matlabu, realizující řešení všech dílčích úkolů, je na obr. 3.13. Simulovaná data představují opakovaná měření stejného jevu, a tak není třeba se zabývat stanovováním referenčních časových souřadnic – všechny repetice vygenerované funkcí *chirpandnoise()* jsou v časové koherenci. Obr. 3.10 ukazuje jednak vybrané repetice, u nichž

lze pozorovat pro zvolené  $SNR = -10$  dB překrytí užitečné složky šumem a dále ukazuje zvýrazněnou užitečnou složku pomocí rovnoměrné i exponenciální kumulace. Dynamické vlastnosti jsou na obr. 3.11 vykresleny pro  $2M$  repetice, a to pro případ, že by se u kumulace s rovnoměrnými vahami pokračovalo od 1001. repetice s klouzavým oknem. U rovnoměrné kumulace lze pozorovat rychlejší dosažení nejvyššího možného průměrného zlepšení poměru signálu k šumu, než je tomu u exponenciální kumulace.

Výsledek  $\pm$  průměrování pro odhad zbytkové rušivé složky je na obr. 3.12, a to společně s výsledkem výpočtu průměrných výkonů rušivé složky v jedné vybrané repetici před kumulací  $ms(n_{20}(t))$  a reziduální rušivé složky po kumulaci. Z výsledků je patrné, že vlivem kumulace došlo k 1000-násobnému utlumení průměrného výkonu rušivé složky, což při zachování výkonu užitečné složky repetiční časové řady vede na 1000-násobné zlepšení výkonového poměru signálu k šumu a cca 32-násobné zlepšení amplitudového poměru signálu k šumu, což je v souladu s výsledkem na obr. 3.11 a s teoretickými poznatky třetí kapitoly.



**Obr. 3.11** Úloha 3.1 – porovnání dynamických vlastností kumulace s rovnoměrnými a exponenciálními vahami. V případě rovnoměrné kumulace je od 1001. repetice použit algoritmus s klouzajícím oknem.



**Obr. 3.12** Úloha 3.1 – ukázka jedné realizace rušivé složky (slabou tečkovanou čarou) a výstup z  $\pm$  průměrování (silnou plnou čarou), jehož výsledkem je zbytková rušivá složka po rovnoměrné kumulaci. Dále jsou zde vypočítané průměrné výkony obou náhodných časových řad, ze kterých je vidět, že vlivem kumulace 1000 směrů užitečné a náhodné rušivé složky došlo ke snížení průměrného výkonu rušivé složky přibližně 1000-krát.

```

%%% ULOHA 3.1 %%%%%%%%% ROVNOMĚRNÉ A EXPONENCIÁLNÍ KUMULACE

[x,t,s,n] = chirpandnoise(0,10,0.1,0.5,500,10,-10);
M = 1000;          % 1000 repetice
N = length(t); % delka jedne repetice
XX = zeros(M,N); SS = zeros(M,N); NN = zeros(M,N);

for j = 1:M        % vygenerovani 1000 repetice do radku matic pro smesi,
                    % pro uzitecnou a pro rusivou slozku
    [XX(j,:),t,SS(j,:),NN(j,:)] = chirpandnoise(0,10,0.1,0.5,500,10,-10);
end

xMrov = sum(XX)/M; % kumulace s rovnomernymi vahami a pevnym oknem
figure, hold on, plot(t,XX(10,:),'b:');
plot(t,xMrov,'k'); title('Rovnoměrná kumulace')
xlabel('t [s]'); ylabel('x(t), xM(t)');

```

```

q = (M-1)/(M+1); % vypocet koeficientu q tak, aby vlastnosti
                % exponenc. kumulace byly stejne jako vlastnosti
                % rovnomerne kumulace
xMexp = zeros(1,N); xMdif = zeros(1,N); gain = 0;

for j=M:-1:1 % od nejaktualnejsi po nejstarsi repetici
    xMexp = XX(j,:)+q*xMexp; % exponencialni kumulace
    gain = 1+q*gain; % celkove zesileni uzitecne slozky
    if mod(j,2)==0
        xMdif = xMdif+(1/M)*XX(j,:); % plus/minus prumerovani
    else
        xMdif = xMdif-(1/M)*XX(j,:);
    end
end

xMexp = xMexp/gain; % zachovani urovne uzitecne slozky

figure, hold on, plot(t,XX(20,:), 'b:');
plot(t,xMexp, 'k'); title('Exponenciální kumulace');
xlabel('t [s]'); ylabel('x(t), xM(t)');

Krov=ones(1,2*M)*sqrt(M); % dynamicke vlastnosti rovnomerne kumulace
Krov(1:M)=sqrt(1:M);

for j=1:2*M % dynamicke vlastnosti exponenc. kumulace
    Kexp(j)=sqrt((1-q^j)/(1+q^j))*sqrt((1+q)/(1-q));
end

figure, hold on, plot(Krov, 'k'), plot(Kexp, 'b:');
legend({'rovnoměrné', 'exponenciální'}); xlabel('j'); ylabel('Krms');

figure, hold on % vykreslení výsledku plus/minus prumerovani
                % a vypocet prumernych vykonu rusive slozky
                % pred a po rovnomerne kumulaci
plot(t,NN(20,:), 'b:');
plot(t,xMdif, 'k');
ms_n20=sum(NN(20,:).*NN(20,:),2)/N;
ms_xMdif=sum(xMdif.*xMdif,2)/N;
xlabel('t [s]'); ylabel('n20(t), xMdif(t)');
title('± průměrování');

```

**Obr. 3.13** Úloha 3.1 – řešení v Matlabu.

### 3.8 Shrnutí

Ve třetí kapitole byly představeny metody pro kumulační zvýrazňování užitečné složky repetičních časových řad ztracených v šumu. Byly vysvětleny různé typy repetičních časových řad a stručně probrána problematika určování referenčních časových souřadnic pro zajištění časové koherence užitečné složky v jednotlivých repetičích. Úspěch kumulačního zvýrazňování je možno vyjádřit průměrným amplitudovým nebo výkonovým zlepšením poměru signálu k šumu. V rovnicích bylo ukázáno, že toto zlepšení závisí pouze na počtu repetic a na nastavení váhových koeficientů. Z rovnic dále vyplynulo, že úspěch kumulačního zvýrazňování je podmíněn charakterem rušivé složky, která by měla být nezávislá na složce



užitečné, generována náhodným procesem s nulovou střední hodnotou a jejíž příspěvky v jednotlivých repeticích by měly být navzájem nezávislé. Dále byly ukázány tři kumulační metody, které se liší nastavením váhových koeficientů: kumulace s rovnoměrnými vahami, kumulace s exponenciálními vahami a tzv.  $\pm$  průměrování, které lze využít pro odhad zbytkové rušivé složky po rovnoměrné kumulaci. Vlivem různého nastavení vah se různí i dynamické vlastnosti kumulace, tj. závislosti průměrného zlepšení poměru signálu k šumu na počtu použitých repetic. Tyto vlastnosti se u rovnoměrné kumulace liší také podle implementačního algoritmu, přičemž diskutovány byly algoritmy s pevným a s klouzavým oknem. V řešené úloze byly všechny teoretické poznatky ověřeny na simulovaných datech.

## 4 Náhodné procesy a modely časových řad

Experimentální data v podobě časových řad získaných pozorováním reálných dynamických systémů mohou být využita pro konstrukci matematických modelů těchto systémů. Modely jsou z naměřených dat zpravidla vytvářeny za účelem hlubšího porozumění operacím a vazbám, které ovlivňují vznik hodnot časových řad, viz obr. 1.1. Dílčími změnami těchto operací nebo vazeb, prováděnými nad modelem, lze simulovat jejich vliv na změny hodnot časové řady. Matematický aparát modelování umožňuje nalézt v časových řadách ty komponenty, které nesou užitečnou informaci a které mají přímou interpretaci, popsat je ve formě deterministických funkcí času a separovat je od náhodné složky rušení. Analýzou časových řad ve formě modelování procesů jejich vzniku a s využitím simulací nad zkonstruovanými modely lze od prostého pozorování dospět až k možnostem ovlivnění či řízení systémů, které časové řady generují.

Při modelování vzniku časových řad se naměřené posloupnosti obvykle považují za realizace náhodných procesů. Pojem „náhodný“ je používán vzhledem k časté nepravidelnosti či komplikovanosti dějů reprezentovaných naměřenými průběhy, což však nutně neznamená, že by takové časové řady byly nepredikovatelné. Příkladem uváděným v [4] je elektrokardiogram, jehož hodnotu je sice možno vypočítat z elektrických potenciálů všech srdečních svalových buněk, avšak získávat takto složitá data pro rozhodování, zda pacient potřebuje kardiostimulátor, by bylo příliš náročné. Není tedy chybou považovat časovou řadu pro určitý účel za realizaci náhodného procesu a pro tutéž časovou řadu konstruovat kompletní deterministický model, vyžaduje-li to jiný účel. Postup modelování časových řad vyžaduje obvykle následující kroky.

- **Změření dat.** Konstrukce dobrého modelu závisí na tom, zda naměřená data reflektují chování modelovaného systému či náhodného procesu. Správné nastavení experimentu zahrnuje nejen sledování těch správných proměnných s dostatečnou přesností, ale jde také o to, aby záznam byl dostatečně dlouhý<sup>12</sup> a pořizovaný s vhodnou vzorkovací frekvencí tak, aby byly zachyceny veškeré důležité dynamické jevy.
- **Volba struktury modelu.** Strukturou se myslí matematický vztah mezi vstupními a výstupními proměnnými, přičemž tento vztah obsahuje neznámé parametry. Příkladem struktury lineárního modelu může být přenosová funkce  $H(z)$  s nastavitelnými póly a nulovými body. Za strukturu lineárního modelu lze považovat např. i tuto jednoduchou diferenční rovnici:  $y(n) + ay(n-1) = bx(n)$ , kde  $a$  a  $b$  jsou volitelné parametry modelu. Struktura modelu často souvisí s matematicko-fyzikálními principy, kterými se řídí pozorované jevy. Je-li struktura modelu dána známými zákony a závislostmi, pak je cílem modelování pouze odhad numerických hodnot parametrů modelu z dostupných dat a takové úlohy se označují jako **modelování šedé skřínky** (*gray-box modelling*). Pokud struktura modelu není předem známá, potom se hovoří o **modelování černé skřínky** (*black-box modelling*).
- **Odhad parametrů modelu podle zvolené struktury.** Výpočet numerických hodnot parametrů modelu vede na minimalizaci rozdílu mezi naměřenými daty  $x(n)$  a výstupem modelu  $\hat{x}(n)$ . Tento rozdíl:

---

<sup>12</sup> Pojem "dostatečně dlouhý" je nutno vnímat s ohledem na to, jak rychlé jsou změny v naměřených průbězích nebo s ohledem na periody opakujících se jevů.

$$v(n) = x(n) - \hat{x}(n) \quad (4.1)$$

se označuje jako **chyba predikce** a z vážené  $l_2$ -normy posloupnosti  $v(n)$  se odvozuje minimalizační kritérium označované také jako střední kvadratická chyba (*MSE – mean squared error*).

- **Hodnocení kvality modelu.** V závěrečném kroku je nutné posoudit kvalitu modelu a prověřit tak, zda je chování modelu adekvátní k zadané aplikaci. Hodnocení může být provedeno např. prostým srovnáním časových řad  $x(n)$  a  $\hat{x}(n)$  (např. graficky), nebo pomocí analýzy  $v(n)$ , která se často označuje jako analýza residuí. Residua reprezentují tu část dat, která není vysvětlena modelem.

## 4.1 Vybrané statistické momenty náhodných procesů

Jak bylo uvedeno v úvodu této kapitoly, vyčerpávající matematický popis sledovaných dynamických systémů ve formě diferenčních rovnic se známými parametry a počátečními podmínkami je k dispozici jen velmi zřídka. Časové průběhy sledovaných procesů jsou často tak složité, že se mohou jevit vnějšímu pozorovateli jako náhodné [19], a tak jen obtížně popsatelné matematickými funkcemi. Náhodné časové řady jsou charakterizovány pomocí statistických momentů, za kterými si lze představit jejich „průměrné vlastnosti“. Poněvadž takové charakteristiky (tj. statistické momenty) nemohou definovat časové průběhy sledovaných procesů kompletně, budou zřejmě platit pro širší soubor časových řad, které vykazují společné vlastnosti [4].

### 4.1.1 Časové průměry

Časový průměr posloupnosti nebo časové řady  $x(n)$  je definován jako:

$$\langle x(n) \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x(n). \quad (4.2)$$

Časovými průměry se má smysl zabývat pouze pro takové časové řady, pro které limita (4.2) existuje. Jedná se např. o periodické časové řady. U těchto je časový průměr roven nultému koeficientu  $a_0$  Fourierovy řady (1.6). Pro monotónně rostoucí časové řady (např. řady s lineárním nebo exponenciálním trendem) nebo časově ohraničené řady nelze tyto časové průměry definovat vůbec. Časové průměry je užitečné zavádět zejména pro časové řady, které se označují jako **stacionární**, tj. řady, jejichž statistické momenty se v čase nemění<sup>13</sup>. Toto omezení není tak restriktivní, jak se může na první pohled zdát, protože v mnoha případech lze nestacionární časové řady považovat v určitých intervalech za stacionární.

Dvě vlastnosti časových průměrů, které jsou důležité pro další odvozování, jsou:

1. **Linearita.** Časový průměr lineární kombinace časových řad  $x(n)$  a  $y(n)$  je roven lineární kombinaci časových průměrů pro libovolné konstanty  $a, b$ :

<sup>13</sup> Pojem **stacionarita** bude definován exaktně v kapitole 4.2.

$$\langle ax(n) + by(n) \rangle = a \langle x(n) \rangle + b \langle y(n) \rangle. \quad (4.3)$$

2. **Časová invariantnost.** Jakékoli zpoždění časové řady o  $n_0$  vzorků nemění její časový průměr a ten je vždy stejný jako její **střední hodnota**  $\mu_x$ :

$$\langle x(n - n_0) \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=n_0}^{n_0+N} x(n) = \mu_x. \quad (4.4)$$

Notaci časových průměrů lze zobecnit i na funkce časových řad:

$$\langle g(x(n)) \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N g(x(n)). \quad (4.5)$$

Příkladem může být **průměrný výkon časové řady**  $ms$  definovaný v (3.1). Souvislost průměrného výkonu časové řady s jejím **rozptylem**  $\sigma^2$  ukazují následující dvě rovnice:

$$\sigma_x^2 = \langle (x(n) - \mu_x)^2 \rangle, \quad (4.6)$$

$$ms_x = \sigma_x^2 + \mu_x^2, \quad (4.7)$$

kde  $\sigma_x$  je **směrodatná odchylka**.

#### 4.1.2 Pravděpodobnost a Čebyševova nerovnost

Časové průměry jsou velmi úzce provázané s pravděpodobnostmi, což lze ukázat na následujících rovnicích, ve kterých se pracuje s časovými průměry  $u(x(n))$ , kde  $u(x)$  je tzv. **jednotkový skok**:

$$\langle u(x(n)) \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N u(x(n)). \quad (4.8)$$

Funkce  $u(x(n))$  nabývá hodnoty 1, pokud  $x(n) \geq 0$  a jinak nabývá hodnoty 0. Suma v (4.8) vlastně vyjadřuje počet vzorků, pro které  $x(n) \geq 0$ , a časový průměr  $\langle u(x(n)) \rangle$  tedy vyjadřuje pravděpodobnost, že  $x(n)$  je nezáporné, tedy  $P(x(n) \geq 0)$ :

$$\langle u(x(n)) \rangle = \lim_{N \rightarrow \infty} \frac{\text{počet nezáporných vzorků } x(n) \text{ v intervalu } (1, N)}{\text{počet všech vzorků } x(n) \text{ v intervalu } (1, N)} = P(x(n) \geq 0). \quad (4.9)$$

Rovnice (4.8) a (4.9) lze zobecnit pro libovolnou prahovou hodnotu  $X_0$  a pomocí časových průměrů vyjádřit pravděpodobnost, že  $x(n)$  je větší nebo rovno  $X_0$ :

$$P(x(n) \geq X_0) = \langle u(x(n) - X_0) \rangle. \quad (4.10)$$

Z výše uvedeného vyplývá, že pravděpodobnosti lze považovat za speciální případy časových průměrů.

Rozptyly a pravděpodobnosti jsou mezi sebou vázány důležitou nerovností pojmenovanou podle ruského matematika Čebyševa. Tato nerovnost udává horní mez pro pravděpodobnost, že se hodnota časové řady  $x(n)$  liší od její střední hodnoty více než o libovolnou hodnotu  $\varepsilon$ :

$$P(|x(n) - \mu_x| \geq \varepsilon) \leq \frac{\sigma_x^2}{\varepsilon^2}. \quad (4.11)$$

Důkaz Čebyševovy nerovnosti lze najít např. v [4]. Je pozoruhodné, že tato nerovnost platí pro libovolné náhodné časové řady, přičemž pro konkrétní typy náhodných časových řad, vymezené např. konkrétní pravděpodobnostní funkcí, lze dospět ještě k těsnějším mezím. Čebyševova nerovnost je důležitá, neboť odůvodňuje běžnou matematickou praxi odhadů založených na minimalizaci kvadratické chyby. Např. minimalizace rozptylu chyby predikce  $\nu(n)$ , definované ve (4.1), dává podle Čebyševovy nerovnosti smysl, neboť je nepravděpodobné, že by hodnoty v chybové posloupnosti  $\nu(n)$  byly velké, pokud je její rozptyl malý.

#### 4.1.3 Korelační a kovarianční funkce

Pokud časová řada  $x(n)$  nabude pro konkrétní  $n$  hodnoty  $X$ , je pravděpodobné, že ještě alespoň krátkou dobu po  $n$  zůstanou i další hodnoty časové řady  $x(n)$  v okolí  $X$ . Vzorky časové řady s krátkým vzorkovacím intervalem obecně nejsou na sobě nezávislé. Je proto třeba znát časové průměry charakterizující tyto závislosti (nebo korelace) mezi vzorky v různých časech. Jedním z takových časových průměrů je **autokorelační funkce**.

$$R_{xx}(k) \cong \langle x(n)x(n-k) \rangle. \quad (4.12)$$

Díky stacionaritě časových průměrů nezávisí autokorelační funkce  $R_{xx}(k)$  na absolutním čase daným indexem vzorku  $n$ , ale jen na zpoždění mezi vzorky  $k$ . Z uvedeného vztahu lze intuitivně usoudit, že autokorelační funkce měří podobnost mezi po sobě jdoucími vzorky časové řady. Pokud jsou změny v časové řadě pomalé, budou mít vzorky v  $n$  a  $n-1$  často stejné znaménko, a tak bude průměrná hodnota jejich součinů  $R_{xx}$  velká. Naopak, pokud se časová řada s nulovou střední hodnotou mění velmi rychle a pro vzorky oddělené krátkými intervaly je stejně pravděpodobné, že mají shodná i různá znaménka, bude se průměrná hodnota jejich součinů  $R_{xx}$  nejspíš blížit nule [4].

Dalším časovým průměrem, který charakterizuje závislosti mezi vzorky v různých časech, se nazývá **autokovarianční funkce**.

$$C_{xx}(k) \cong \langle (x(n) - \mu_x)(x(n-k) - \mu_x) \rangle = R_{xx}(k) - \mu_x^2. \quad (4.13)$$

Jedná se o autokorelační funkci časové řady s odečtenou střední hodnotou. Protože jde o autokorelační funkci, budou mít  $C_{xx}$  a  $R_{xx}$  stejné matematické vlastnosti.

Autokorelační funkce jsou sudé funkce zpoždění:

$$R_{xx}(-k) = \langle x(n)x(n+k) \rangle = \langle x(n)x(n-k) \rangle = R_{xx}(k). \quad (4.14)$$

Autokorelační funkce časové řady  $x(n)$  vyčíslená v počátku odpovídá střednímu výkonu  $x(n)$  a autokovarianční funkce časové řady  $x(n)$  v počátku odpovídá rozptylu  $x(n)$ :

$$R_{xx}(0) \cong \langle x(n)^2 \rangle = ms_x. \quad (4.15)$$

$$C_{xx}(0) \cong \langle (x(n) - \mu_x)^2 \rangle = \sigma_x^2. \quad (4.16)$$

Autokorelační funkce nabývá maxima vždy v počátku:

$$|R_{xx}(k)| \leq R_{xx}(0), \quad -\infty < k < \infty. \quad (4.17)$$

Pro velká zpoždění  $k$  se autokovarianční funkce časové řady  $x(n)$ , je-li  $x(n)$  bez periodických komponent, blíží nule:

$$\lim_{|k| \rightarrow \infty} C_{xx}(k) = 0. \quad (4.18)$$

Předchozímu tvrzení odpovídá intuitivní představa o tom, že vzorky, které jsou v náhodné časové řadě oddělené velkým zpožďovacím intervalem  $k$ , jsou nekorelované. Toto odpovídá případům, kdy je časový průběh sledovaného náhodného procesu velmi nepravidelný a nepředvídatelný.

Pro velká zpoždění  $k$  se autokorelační funkce časové řady  $x(n)$  blíží kvadrátu střední hodnoty  $x(n)$ :

$$\lim_{|k| \rightarrow \infty} R_{xx}(k) = \mu_x^2. \quad (4.19)$$

**Autokorelační koeficient**  $\rho_{xx}(k)$  lze vyjádřit jako hodnotu autokovarianční funkce vztaženou k rozptylu:

$$\rho_{xx}(k) \cong \frac{C_{xx}(k)}{\sigma_x^2}. \quad (4.20)$$

O náhodné časové řadě  $v(n)$  se říká, že je **bílá**, když je její autokovarianční funkce  $C_{vv}$  váženým jednotkovým impulsem v počátku:

$$C_{vv}(k) = \sigma_v^2 \delta(k) = \begin{cases} \sigma_v^2 & \text{pro } k = 0 \\ 0 & \text{pro } k \neq 0 \end{cases}. \quad (4.21)$$

Pro autokorelační funkci bílé časové řady potom platí:

$$R_{vv}(k) = \sigma_v^2 \delta(k) + \mu_v^2. \quad (4.22)$$

Pojem „bílá“ pochází z fyziky světelného záření, o němž se říká, že je bílým světlem, pokud jeho zdroj vyzařuje energii rovnoměrně na všech vlnových délkách vnímaných lidským okem. Tato fyzikální představa má i svůj matematický základ, neboť se dá ukázat, že spektrum jednotkového impulsu je rovnoměrné, což znamená, že jeho harmonické složky jsou na všech frekvencích zastoupeny stejně. Při modelování časových řad hraje zásadní roli **bílý šum**, jak bude ukázáno později v této kapitole.

V mnoha úlohách se nevystačí pouze s analýzou jediné náhodné časové řady, ale je třeba se zabývat kombinací více časových řad. Například pro biosignál snímáný z lidské hrudi můžeme pozorovanou časovou řadu  $x(n)$  chápat jako aditivní směs užitečné složky, tj. elektrokardiogramu  $s(n)$  a rušivé složky  $v(n)$ , způsobené aktivitou jiných svalů a šumem snímací elektroniky:

$$x(n) = s(n) + v(n). \quad (4.23)$$

Autokorelační funkce časové řady  $x(n)$  je dána:

$$R_{xx}(k) = \langle x(n)x(n+k) \rangle = \langle (s(n) + v(n))(s(n+k) + v(n+k)) \rangle \quad (4.24A)$$

a s využitím (4.12) lze (4.24) přepsat na:

$$R_{xx}(k) = R_{ss}(k) + R_{vv}(k) + \langle s(n)v(n+k) \rangle + \langle v(n)s(n+k) \rangle. \quad (4.24B)$$

S využitím **křížové korelační funkce**<sup>14</sup> dvou náhodných časových řad  $x(n)$  a  $y(n)$ , která je definována jako [4]:

---

<sup>14</sup> Křížová korelační funkce je v české odborné literatuře uváděná také jako **vzájemná korelační funkce**.

$$R_{xy}(k) \equiv \langle x(n) + y(n+k) \rangle = \langle x(n-k)y(n) \rangle, \quad (4.25)$$

lze (4.23) a (4.24) přepsat na:

$$R_{xx}(k) = R_{ss}(k) + R_{vv}(k) + R_{sv}(k) + R_{vs}(k). \quad (4.26)$$

Podobně jako autokorelační funkce určuje míru závislosti mezi následujícími vzorky jediné časové řady, vyjadřuje křížová korelační funkce  $R_{xy}(k)$  podobnost mezi časovou řadou  $x(n)$  a zpožděnou časovou řadou  $y(n)$ . Křížová korelační funkce centrovaných časových řad  $x(n) - \mu_x$  a  $y(n) - \mu_y$  se označuje jako **křížová kovarianční funkce** [4]:

$$C_{xy}(k) \equiv \langle (x(n) - \mu_x)(y(n+k) - \mu_y) \rangle = R_{xy}(k) - \mu_x \mu_y. \quad (4.27)$$

Na rozdíl od autokorelačních funkcí nejsou křížové korelační funkce sudé:

$$R_{xy}(-k) = \langle x(n)y(n-k) \rangle = \langle x(n+k)y(n) \rangle = R_{yx}(k). \quad (4.28)$$

**Korelační koeficient**  $\rho_{xy}(k)$  lze vyjádřit jako křížovou kovarianční funkci normovanou součinem směrodatných odchylek:

$$\rho_{xy}(k) \equiv \frac{C_{xy}(k)}{\sigma_x \sigma_y}. \quad (4.29)$$

O časových řadách  $x(n)$  a  $y(n)$  se říká, že jsou nekorelované, je-li jejich křížová kovarianční funkce  $C_{xy}(k)$  nulová pro všechna zpoždění  $k$ . Taková situace často ukazuje na to, že časové řady  $x(n)$  a  $y(n)$  jsou realizace dvou na sobě nezávislých náhodných procesů. Z (4.27) vyplývá, že pokud je alespoň jeden z náhodných procesů charakterizován nulovou střední hodnotou, potom i křížová korelační funkce  $R_{xy}(k)$  je nulová pro všechna zpoždění  $k$ .

Křížové korelační a autokorelační funkce jsou užitečné pro detekci periodických komponent v časových řadách postižených aditivním rušením. Obr. 4.1 ukazuje tuto detekci ve směsi užitečné harmonické složky  $s(n) = A \cos(\omega n + \varphi)$  a aditivního rušení  $v(n)$  s nulovou střední hodnotou. Předpokládá se, že užitečná složka a rušivá složka jsou nekorelované. Křížová korelační funkce mezi aditivní směsí a jednotkovou kosinovou vlnou  $y(n) = \cos(\omega n)$  se stejnou úhlovou frekvencí jako  $s(n)$  je:

$$C_{yx}(k) = \langle y(n)(s(n+k) + v(n+k)) \rangle = C_{ys}(k) + C_{yv}(k). \quad (4.30)$$

Druhý člen v rovnici (4.30) je nulový, neboť  $y(n)$  a  $v(n)$  jsou nekorelované (viz výše uvedený předpoklad), a tak se křížová korelační funkce zjednoduší tak, že je až na amplitudu stejná jako užitečná složka:



$$C_{yx}(k) = C_{ys}(k) = \frac{A}{2} \cos(\omega n + \varphi). \quad (4.31)$$

V případech, kdy není perioda užitečné složky předem známá, lze pro určení periody využít autokorelační funkci směsi  $x(n)$ :

$$R_{xx}(k) = R_{ss}(k) + R_{vv}(k) + R_{sv}(k) + R_{vs}(k), \quad (4.32)$$

kde poslední dva členy budou nulové za předpokladu, že rušení  $v(n)$  a užitečná složka  $s(n)$  jsou nekorelované. Autokorelační funkce náhodné rušivé složky  $R_{vv}(k)$  bude pro velká zpoždění mezi vzorky zanedbatelná, a tak bude jistě snazší určovat periodu užitečné periodické složky z  $R_{xx}(k)$  než z průběhu  $x(n)$ , viz obr. 4.1D.

#### 4.1.4 Pravděpodobnostní funkce

V kapitole 4.1.2 bylo ukázáno, že pravděpodobnosti lze považovat za speciální případy časových průměrů. Ze vztahu (4.10) lze odvodit pravděpodobnost, že časová řada  $x(n)$  nabude hodnoty mezi  $X$  a  $X+\Delta X$ , a to pomocí časového průměru obdélníkové funkce s šířkou  $\Delta X$  [4]:

$$P(X \leq x(n) < X + \Delta X) = \langle \Pi_{\Delta X}(x(n) - X) \rangle, \quad (4.33)$$

kde  $\Pi_{\Delta X}(x) \equiv \begin{cases} 1 & \text{když } 0 \leq x < \Delta X \\ 0 & \text{jinak} \end{cases}$ .

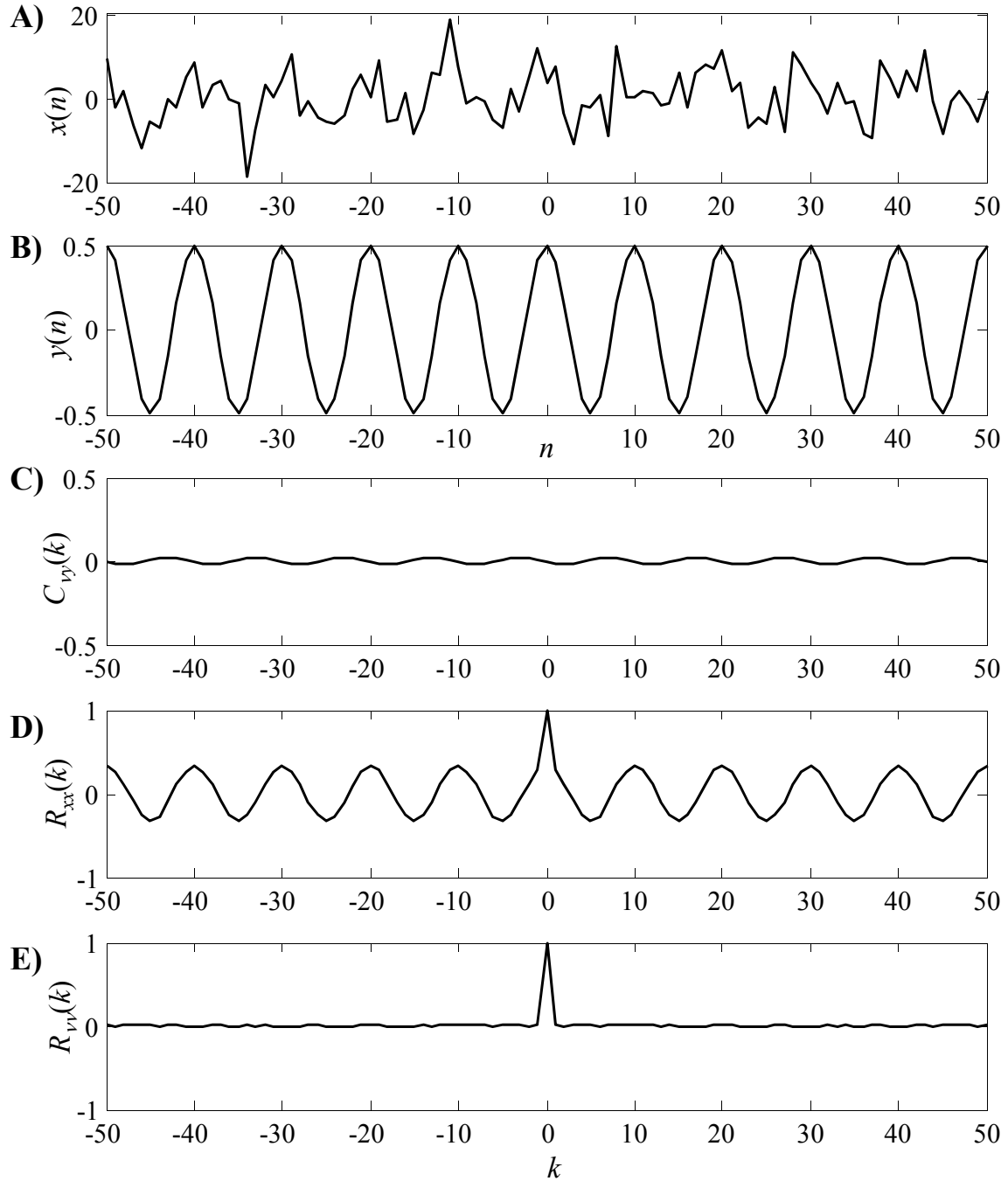
**Pravděpodobnostní funkce**  $f_x(X)$  je dána podílem této pravděpodobnosti a šířkou intervalu  $\Delta X$ , což při  $\Delta X$  blížící se v limitě nule dává:

$$f_x(X) \equiv \lim_{\Delta X \rightarrow 0} \frac{1}{\Delta X} P(X \leq x(n) < X + \Delta X) = \lim_{\Delta X \rightarrow 0} \left\langle \frac{1}{\Delta X} \Pi_{\Delta X}(x(n) - X) \right\rangle. \quad (4.34)$$

Plocha pod normalizovanou obdélníkovou funkcí v (4.34) bude jednotková pro jakoukoli šířku intervalu  $\Delta X$ . Pro  $\Delta X$  blížící se k nule bude tato funkce jednotkovým impulsem  $\delta(X)$ . To znamená, že pravděpodobnostní funkci lze formálně zapsat jako časový průměr funkce  $\delta$  [4]:

$$f_x(X) \equiv \langle \delta(x(n) - X) \rangle. \quad (4.35)$$

Pravděpodobnostní funkce  $f_x(X)$  vyjadřuje pravděpodobnost, že hodnota časové řady  $x(n)$  leží ve velmi úzkém intervalu se středem v  $x(n)=X$ , normovanou šířkou tohoto intervalu. Malé písmeno  $x$  v  $f_x(X)$  reprezentuje náhodnou časovou řadu  $x(n)$  a velké písmeno  $X$  reprezentuje konkrétní střed intervalu, pro který je pravděpodobnost počítána. Tato dvě písmena tedy spolu přímo nesouvisí. Např.  $f_y(X)\Delta X$  vyjadřuje pravděpodobnost, že hodnota časové řady  $y(n)$  je v intervalu o šířce  $\Delta X$  centrovaném v  $X$ .



**Obr. 4.1** Detekce sinusové vlny skryté v šumu. Amplitudy autokorelačních a křížových korelačních funkcí jsou normované tak, aby nabývaly jedné v počátku. Odhady jsou vypočítané z průběhů dlouhých 10000 vzorků. A) Časová řada  $x(n)$  je aditivní směsí užitečné složky  $s(n)$  – harmonická časová řada – a rušivé složky  $v(n)$  – bílý šum. B) Pomocná časová řada  $y(n)$  je harmonická časová řada bez rušení. C) Odhad křížové korelační funkce  $C_{vy}(k)$  ukazuje, že  $y(n)$  a  $v(n)$  jsou mezi sebou nekorelované. D) Z odhadu autokorelační funkce  $R_{xx}(k)$  časové řady  $x(n)$  lze vyčíst jasně periodu užitečné složky ve směsi  $x(n)$ . E) Odhad autokorelační funkce bílého šumu.

Pomocí pravděpodobnostní funkce lze vyjádřit střední hodnotu časové řady  $x(n)$ , alternativně k (4.2), jako:

$$\langle x(n) \rangle = \int_{-\infty}^{\infty} X f_x(X) dX, \quad (4.36)$$

přičemž se někdy uvádí, že pomocí (4.36) je definovaná tzv. **očekávaná hodnota** (*expected value*), což je v tomto kontextu jen jiný název pro střední hodnotu náhodné časové řady. Intuitivní důkaz k (4.36) lze ukázat, pokud se v rovnici (4.2) pro časový průměr provede sčítání nikoli přes indexy  $n$ , ale přes narůstající amplitudy časové řady  $x(n)$ . Rozdělením rozsahu všech hodnot  $x(n)$  na navazující malé intervaly s šířkou  $\Delta X$  centrované v  $X_i = i\Delta X$  lze sumu v rovnici (4.2) vyjádřit jako:

$$\sum_{n=1}^N x(n) = \sum_{i=-\infty}^{\infty} \left[ \sum_{X_i - \frac{\Delta X}{2} < x(n) \leq X_i + \frac{\Delta X}{2}} x(n) \right]. \quad (4.37)$$

Za předpokladu dostatečně úzkých intervalů  $\Delta X$  lze všechny hodnoty  $x(n)$  v jednotlivých intervalech aproximovat pomocí středu intervalu  $X_i$ , a sumu pak lze vyjádřit jako:

$$\sum_{n=1}^N x(n) \approx \sum_{i=-\infty}^{\infty} X_i \left( \text{počet vzorků, pro které } X_i - \frac{\Delta X}{2} < x(n) \leq X_i + \frac{\Delta X}{2} \right). \quad (4.38)$$

Vydělením  $N$  a zavedením limity pro velká  $N$  se již vyjádří časový průměr:

$$\langle x(n) \rangle \approx \sum_{i=-\infty}^{\infty} X_i P \left( X_i - \frac{\Delta X}{2} < x(n) \leq X_i + \frac{\Delta X}{2} \right), \quad (4.39)$$

který za předpokladu velmi úzkých intervalů  $\Delta X$  lze vyjádřit pomocí pravděpodobnostní funkce:

$$\langle x(n) \rangle \approx \sum_{i=-\infty}^{\infty} X_i f_x(X_i) \Delta X, \quad (4.40)$$

neboť pravděpodobnost, že  $x(n)$  leží v intervalu o šířce  $\Delta X$  centrovaném v  $X_i$  lze vyjádřit součinem  $f_x(X_i)\Delta X$ . V limitě se suma přes indexy  $i$  přibližuje integrálu uvedeném v (4.36).

#### 4.1.4 Souborové průměry, stacionarita, ergodicita

Modely založené na časových průměrech zanedbávají veškeré informace o časové variabilitě statistických momentů náhodných procesů, a proto jsou užitečné pouze

v případech, kdy lze zavést předpoklad stacionarity náhodných procesů a jimi generovaných časových řad.

**Stacionarita** je vlastnost náhodného procesu, která mu přisuzuje časovou invariantnost jeho statistických momentů. O stacionárním náhodném procesu lze říci, že jeho pravděpodobnostní funkce je nezávislá na čase. V důsledku toho je jeho autokorelační funkce závislá pouze na zpoždění mezi vzorky a nikoli na absolutním čase. V literatuře se lze setkat s mnoha různými definicemi stacionarity, včetně slabé a silné stacionarity, autokorelační stacionarity atd.

#### DRUHÁ POUČKA ODVOZENÁ SELSKÝM ROZUMEM:

**„Předpoklad stacionarity splňují ty časové řady či náhodné procesy, které jsou bez trendu, mají s měnícím se časem stejný rozptyl a stejný průběh autokorelační funkce.“**



U modelů živých systémů, které stárnou a adaptují se na měnící se podmínky prostředí, bude předpoklad stacionarity do jisté míry vždy omezující. V mnoha případech vykazují statistické momenty časových řad tak rychlé změny, že předpoklad stacionarity nelze zavést ani v omezených časových intervalech. Pro taková data je pak lepší volit abstraktnější model založený na **souborových průměrech**.

Vysvětlení rozdílu mezi časovými a souborovými průměry je ukázkově vysvětleno v [4] na příkladu plynových částic pohybujících se v uzavřeném prostoru. Každá částice vykonává neustále tzv. Brownův pohyb. Je-li zaznamenána trajektorie pohybu částic, může z ní být pro každou částici určena průměrná poloha, která se vypočítá způsobem odpovídajícím výše zavedenému časovému průměru. Alternativně lze však v různých časových okamžicích určit také průměrnou polohu všech částic, a to výpočtem, který se označuje jako **souborový průměr**. Pozici plynových částic lze tedy považovat za funkci dvou proměnných: času a souborové proměnné identifikující jednotlivé částice. Pro každý časový index  $n$  je možné vypočítat podíl částic, které leží mezi pozicemi  $X$  a  $X+\Delta X$  podél jedné z os. V limitním případě, když se  $\Delta X$  blíží k nule, definuje tento podíl časově závislou pravděpodobnostní funkci  $f_x(X, n)$ . Souborová průměrná pozice  $E(X(n))$ , neboli **očekávaná hodnota** pozice podél jedné z os se vypočítá:

$$E\{x(n)\} = \int_{-\infty}^{\infty} X f_x(X, n) dX, \quad (4.41)$$

což je rovnice, která je sice podobná na rovnici (4.36) pro výpočet časového průměru náhodné časové řady z pravděpodobnostní funkce, avšak obě rovnice mají zcela odlišnou interpretaci. V rovnici (4.34) byla pravděpodobnostní funkce definována jako časový průměr, takže (4.36) vyjadřuje vlastně vzájemný vztah mezi dvěma časovými průměry. Rovnice (4.41) definuje očekávanou hodnotu na základě souborové pravděpodobnostní funkce  $f_x(X, n)$ , která je a priori dána pro každé  $n$ . Tedy zatímco v (4.34) pravděpodobnostní funkce nese informaci o tom, jak se pozice jednotlivých částic mění v čase, tak v rovnici (4.41) je tato informace v pravděpodobnostní funkci obsažena.

Podobně jako na pohybující se plynové částice lze nahlížet také na mnohé časové řady. Sejme-li se např. vícekrát elektroencefalogram z povrchu hlavy stejného pacienta za identických experimentálních podmínek, není pochyb, že jednotlivé časové řady se budou

mezi sebou lišit, i když budou vykazovat velmi podobné statistické momenty. Na každý záznam elektroencefalogramu lze pohlížet jako na jeden prvek souboru časových řad, které by eventuálně mohly být naměřeny. V uvedeném příkladu není soubor přímo pozorovatelný – jedná se o abstrakci či model charakterizující nejistotu o časové řadě – elektroencefalogramu.

Teorie o pravděpodobnosti zavádí pojem *stacionární ergodický proces*, což je náhodný proces, který vykazuje stacionaritu i ergodicitu. *Ergodicita* je vlastnost náhodného procesu, která mu přisuzuje shodu mezi jeho souborovými a časovými průměry. V praxi to znamená, že stacionární ergodický náhodný proces jednak nemění své statistické momenty v čase a dále, že tyto jeho statistické momenty mohou být zjištěny z jediné dostatečně dlouhé realizace náhodného procesu.

## 4.2 Dekompozice časových řad a Boxovy-Jenkinsovy modely

Na problém modelování časových řad je možno pohlížet jako na analýzu směsí užitečných a rušivých komponent. Cílem modelování je oddělit užitečnou složku od náhodného rušení, přičemž užitečná (nebo také systematická) složka se v literatuře často ještě dělí na tzv. *trendovou*, *sezónní* a *cyklickou* složku. Je-li na směr užitečných a rušivých komponent pohlíženo jako na směr aditivní, pak lze časovou řadu  $x(n)$  zapsat ve tvaru:

$$x(n) = s(n) + v(n) = T(n) + C(n) + v(n), \quad (4.42)$$

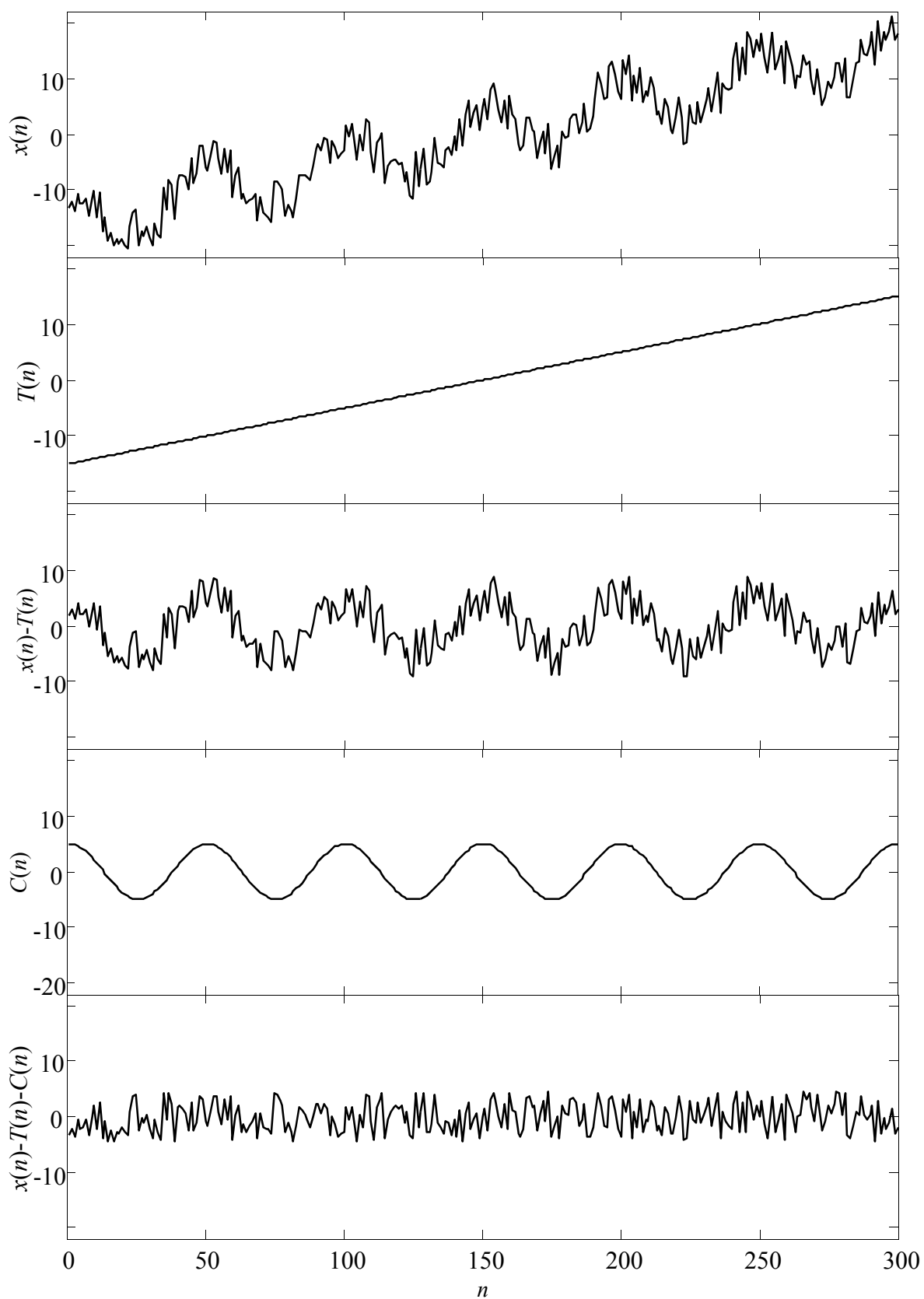
kde  $v(n)$  je náhodná chyba nebo rušení,  $T(n)$  je trendová složka vyjadřující dlouhodobý nárůst nebo pokles hodnot časové řady,  $C(n)$  zastupuje sezónní i cyklickou složku, neboť obě tyto komponenty lze považovat společně za periodické kolísání hodnot časové řady.

Detailnější rozklad  $C(n)$  na složku sezónní a cyklickou lze najít zejména v literatuře věnující se chování ekonomických systémů a autoři rozlišují periodické složky na sezónní a cyklické podle délky periody: je-li perioda dlouhá den, týden, měsíc, kvartál nebo rok, bývá taková složka označovaná jako sezónní; nevykazuje-li perioda kalendářní charakter nebo je delší než jeden rok, pak se hovoří o složce cyklické. S ohledem na znalosti získané v kapitole 1.3, o rozkladu periodických signálů nebo časových řad na harmonické komponenty pomocí Fourierovy řady, není třeba v tomto textu dále komplikovat názvosloví a složka  $C(n)$  zde bude označována jako *periodická*.

Kromě dekompozice aditivních směsí se někdy využívá také multiplikativní dekompozice. Zatímco v případě aditivního modelu jsou všechny složky vyjádřeny ve stejných jednotkách jako pozorovaná časová řada, v případě multiplikativního modelu je ve stejných jednotkách jako pozorovaná řada pouze trendová složka a ostatní (periodická složka a složka rušení) vyjadřují součinem relativní, okamžitý nárůst nebo pokles hodnot časové řady. Multiplikativní dekompozice je nad rámec tohoto učebního textu – dekompozice časových řad je zde probrána jen pro aditivní směsi.

Trendová složka je obvykle po grafické rozvaze vypočítána regresní analýzou, kdy je aplikací metody nejmenších čtverců proložen obecný polynom předem daného stupně nebo i jiná předem daná funkce (např. exponenciála, sigmoida aj.).

Detekce periodické složky může být provedena rozkladem na harmonické složky Fourierovou řadou (viz kapitola 1.3) nebo pomocí autokorelačních a křížových korelačních funkcí (viz kapitola 4.1.3).



**Obr. 4.2** Dekompozice časové řady  $x(n)$  na trendovou složku  $T(n)$ , periodickou složku  $C(n)$  a náhodné rušení  $v(n)$ .

Obr. 4.2 ukazuje aditivní dekompozici časové řady  $x(n)$  na trendovou, periodickou složku a složku rušení. V ilustračním jednoduchém příkladu byla použita časová řada obsahující lineární trend a periodická složka obsahovala jedinou harmonickou. V praxi může být aditivní dekompozice časové řady ztížena komplexnější trendovou funkcí (např. lineární trend s několika body zlomu, lokální polynommické trendy apod.) a/nebo více nezanedbatelnými harmonickými komponentami v periodické složce.

Pokud složka rušení  $v(n)$ , získaná očištěním časové řady od trendu a periodické složky, odpovídá realizaci bílého šumu, pak výsledek dekompozice dostatečně vysvětluje veškeré dynamické jevy zachycené v analyzované časové řadě. Pokud tomu tak není, což je v praxi velmi častý případ, znamená to, že složka rušení obsahuje ještě nezanedbatelné závislosti mezi vzorky a ty mohou být vysvětleny s využitím **Boxových-Jenkinsových modelů** stacionárních náhodných procesů. Tato metodika, pojmenovaná podle statistiků George Boxe a Gwilyma Jenkinse, mj. říká, že jakoukoli stacionární časovou řadu generuje náhodný proces, kterému lze přiřadit jeden z těchto modelů:

- čistě rekurzivní autoregresní model **AR**,
- nerekurzivní model s klouzavým průměrem **MA**,
- kombinovaný model **ARMA**,
- a bílý šum.

Bílý šum je náhodný proces nebo časová řada  $v(n)$ , která je bílá a její střední hodnota  $\mu_v$  je nulová. Potom podle rovnic (4.20) a (4.21) bude autokorelační funkce bílého šumu váženým jednotkovým impulsem<sup>15</sup>:

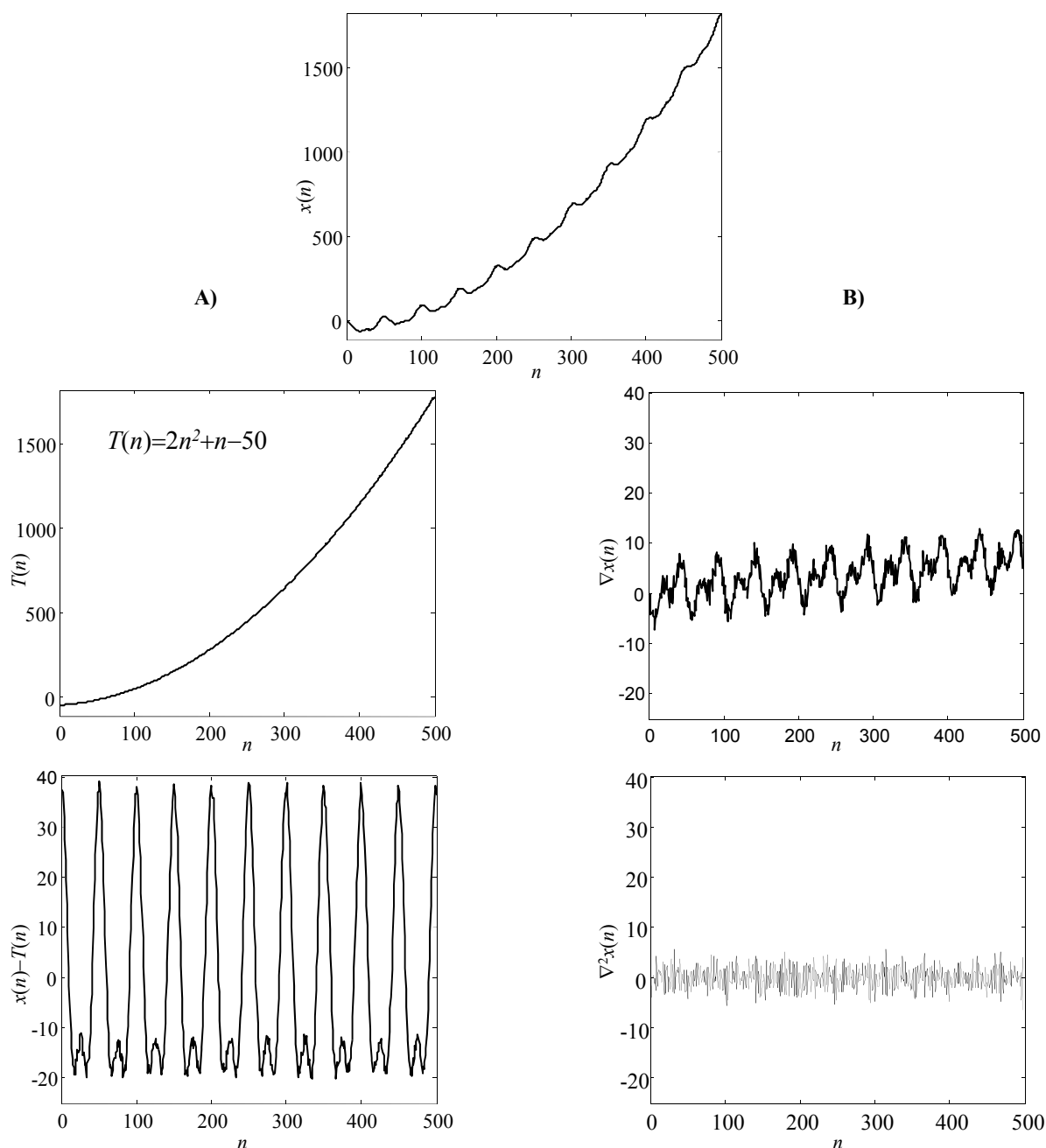
$$R_{vv}(k) = \sigma_v^2 \delta(k) = \begin{cases} \sigma_v^2 & \text{pro } k = 0 \\ 0 & \text{pro } k \neq 0 \end{cases} \quad (4.43)$$

Ukázka empirické autokorelační funkce bílého šumu je na obr. 4.1E. Označení **empirická** je zde použito proto, že se jedná o odhad teoretické autokorelační funkce náhodného procesu z jedné jeho realizace.

Box-Jenkinsova metodika zahrnuje postup pro konstrukci modelů, který obsahuje tři základní kroky a který odpovídá obecnému postupu uvedenému v úvodu čtvrté kapitoly. Tři základní kroky jsou pojmenovány jako [13]:

1. **Identifikace modelu.** Kromě volby struktury modelu zahrnuje tato fáze také zajištění stacionarity časové řady. Detekuje se trendová a periodická složka a případně se aplikuje diferencování časové řady, které je pak zpětně zahrnuto do struktury modelu.
2. **Odhad parametrů modelu.** Fáze, ve které se vypočítají numerické hodnoty parametrů modelu. Podle typu struktury se využívají lineární nebo nelineární metody nejmenších čtverců a metoda odhadu maximální věrohodnosti.
3. **Validace modelu.** Fáze, ve které se analýzou residuí (neboli analýzou posloupnosti chyb predikce) ověřuje, zda model dobře vystihuje pozorované dynamické děje popsané časovou řadou.

<sup>15</sup> Autokorelační funkce bývají často prezentovány v normované formě tak, aby měly pro nulové zpoždění jednotkovou hodnotu.



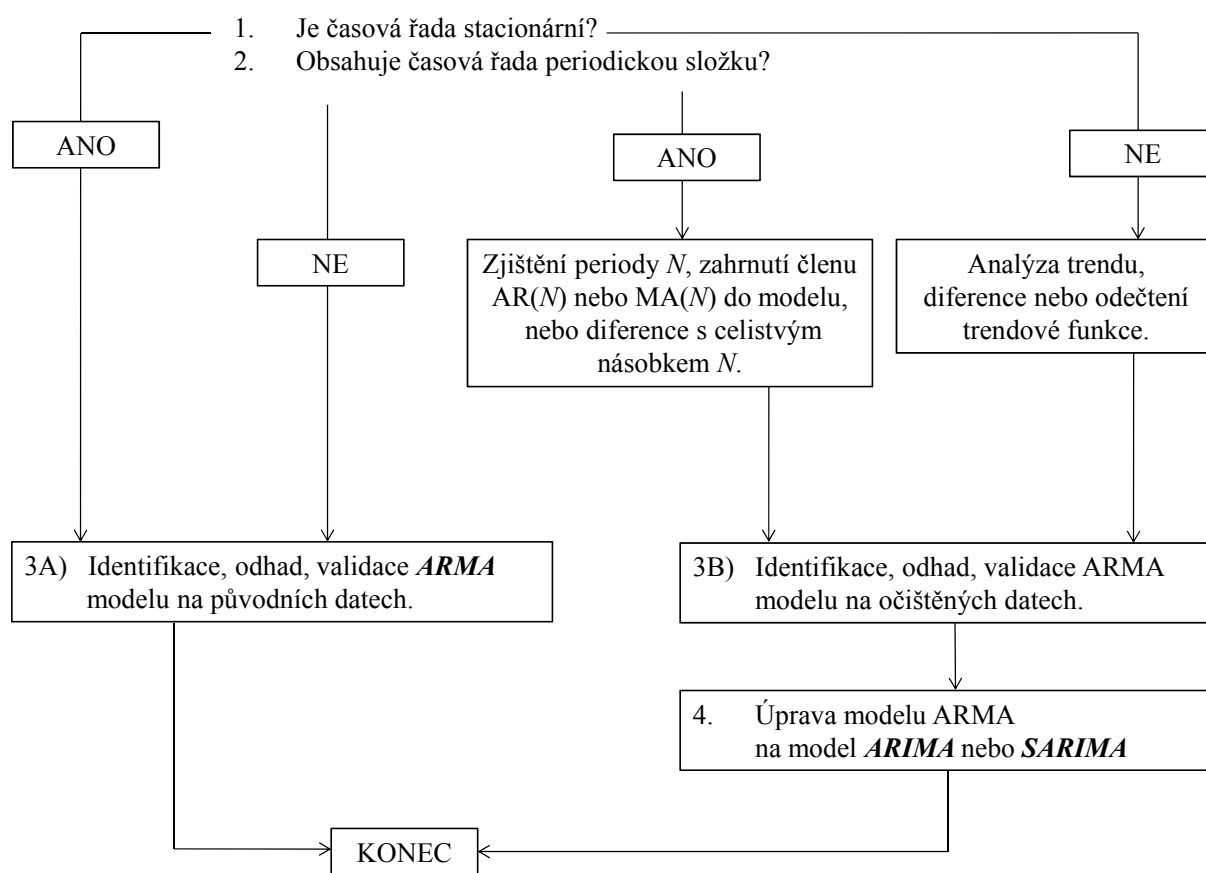
**Obr. 4.3** Stacionarizace časové řady  $x(n)$ : A) odečtením proloženého polynomu, B) jednoduchým a dvojitým diferencováním.

V případě nestacionárních časových řad začíná identifikace modelu předzpracováním časové řady na stacionární posloupnost. Tato fáze je velmi obdobná výše popsané dekompozici časových řad s následným očištěním časové řady od vlivů zjištěného trendu a případně i očištění od vlivů periodického kolísání. Odstranění trendu může být provedeno odečtením proložené trendové funkce, jak je to naznačeno např. na obr. 4.3A. Jinou možností pro stacionarizaci časové řady je její diferencování, jak je to naznačeno např. na obr. 4.3B. Jednoduché diferencování vede k potlačení lineárního trendu, dvojitě diferencování pak očistí časovou řadu od trendu kvadratického. Očištěná posloupnost  $\nabla x(n)$  vznikne diferencováním posloupnosti  $x(n)$  a po konstrukci stacionárního modelu na očištěných datech  $\nabla x(n)$  je pro výpočet  $\hat{x}(n)$  do modelu ještě zahrnuta postupná sumace (integrace) posloupnosti  $\nabla \hat{x}(n)$ .



Obecný stacionární model ARMA se tak změní na nestacionární model ARIMA (autoregressive integrated moving average). Jednoduché diferencování posloupnosti lze v Z-transformaci popsat přenosovou funkcí  $H(z)=1-z^{-1}$ . Sumaci (integraci) posloupnosti lze v Z-transformaci popsat přenosovou funkcí  $H(z)=1/(1-z^{-1})$ , což je přenosová funkce integrační soustavy na mezi stability [19], viz také kap. 1.3. Jednoduché diferencování s výše uvedenou přenosovou funkcí lze považovat za lineární filtraci s frekvenční charakteristikou odpovídající filtru typu horní propust, která byla vypočítána v řešené úloze 1.3, viz obr. 1.8.

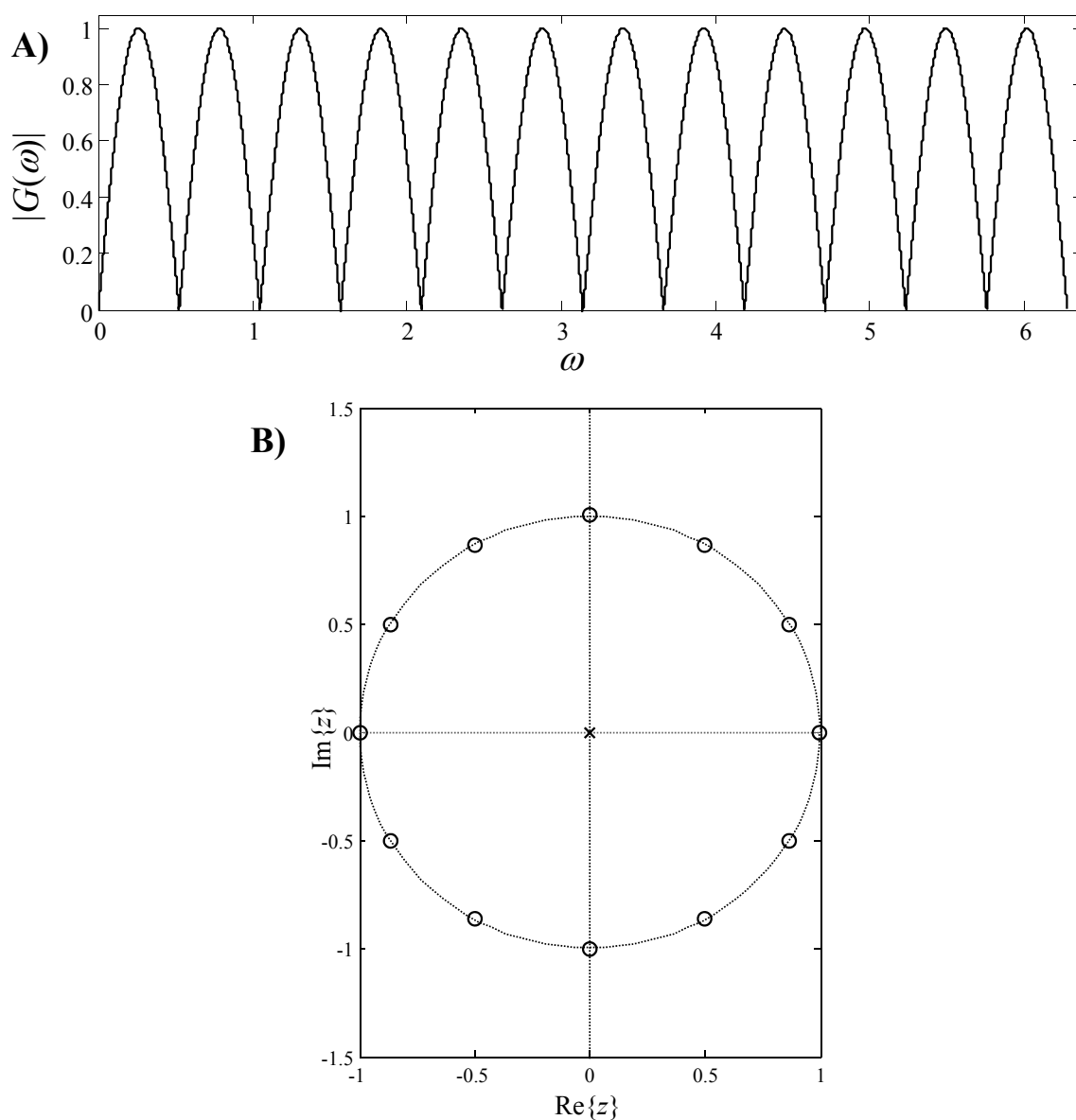
Na obr. 1.8 i na obr. 4.3B je vidět, že tato horní propust utlumuje nejen nízkofrekvenční složky odpovídající pozvolným změnám neboli trendu, ale že navíc částečně utlumuje i periodické složky na vyšších frekvencích, které v původní časové řadě  $x(n)$  mohou nést užitečnou informaci. Je-li stacionární model ARMA konstruován na takto nevhodně předzpracovaných datech, nebude ani výsledný model ARIMA modelovat dobře všechny dynamické jevy. Tento nedostatek však odborná literatura věnující se Boxovým-Jenkinsovým modelům nezmiňuje a diferencování uvádí jako doporučenou metodu.



**Obr. 4.4** Úprava modelu ARMA na model ARIMA nebo SARIMA v případech konstrukce modelu na diferencovaných datech. Diferencování časové řady je doporučeno podle Box-Jenkinsovy metodiky v případech, kdy je časová řada nestacionární nebo se v ní vyskytuje významná periodická složka [13].

Diferencování s celistvým násobkem periody, neboli tzv. **sezónní diferencování** (seasonal difference, seasonal adjustment) je rovněž součástí fáze identifikace modelu podle Boxovy a Jenkinsovy metodiky, a to v případě, kdy je z nějakého důvodu třeba očistit modelovanou časovou řadu od periodického kolísání jejích hodnot. Je-li sezónní diferencování aplikováno před konstrukcí modelu ARMA, pak je stejně jako v případě

jednoduchého nebo dvojitého diferencování pro potlačení trendu nutno tuto operaci zpětně zahrnout do výsledného modelu, který pak bývá označován jako model **SARIMA** (*seasonal ARIMA*).



**Obr. 4.5** Modulová frekvenční charakteristika (A) a rozložení nulových bodů a pólů (B) diskrétní soustavy realizující tzv. sezónní diferencování pro potlačení periodického kolísání – na uvedeném příkladu se jedná o diferenci s periodou 12 vzorků:  $\forall x(n) = x(n) - x(n-12)$ . Tato sezónní diference utlumuje děje nejen s periodou 12 vzorků, ale také vyšší harmonické, tj. děje s periodou 6, 4, 3, 2.5 a 2 vzorky.

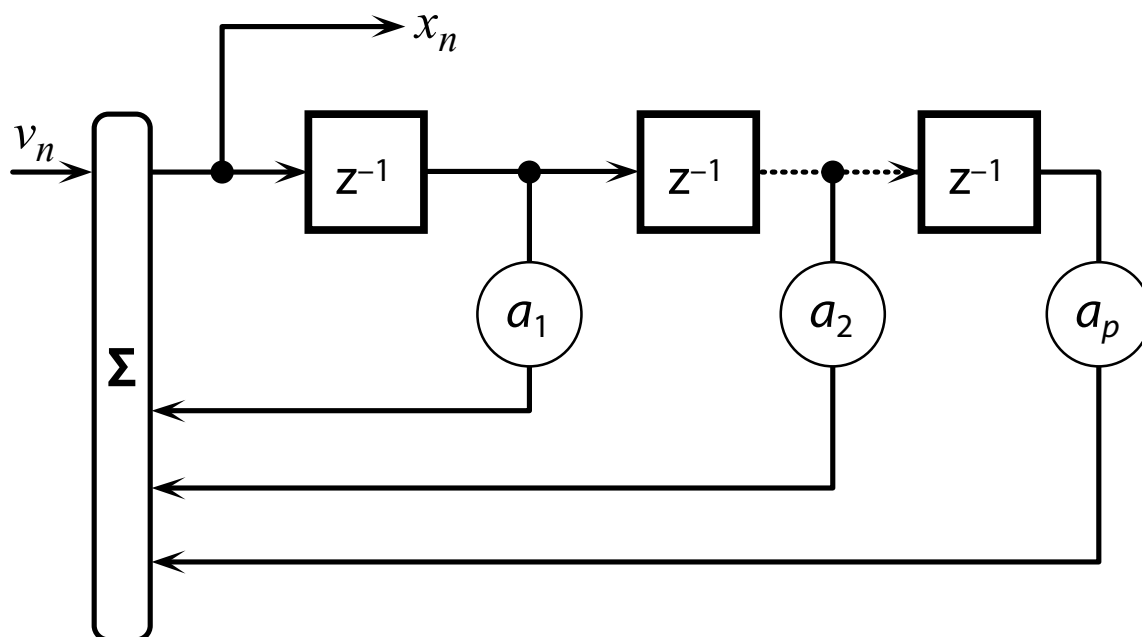
Tvar přenosové funkce diskrétní soustavy, která realizuje sezónní diferencování, je závislý na délce periody. Např. je-li časová řada  $x(n)$  vzorkována s intervalem jeden měsíc, pak jednomu roku odpovídá  $N=12$  vzorků. Je-li cílem analytika dat očistit  $x(n)$  od vlivu roční sezóny na sledované jevy a využije-li k tomu sezónní diferencování  $xd_{12}(n) = x(n) - x(n-12)$ , provádí vlastně lineární filtraci, a to s filtrem s přenosovou funkcí  $H(z) = 1 - z^{-12}$ , viz frekvenční charakteristiku a rozložení nul a pólů na obr. 4.5. Jedná se o tzv.

**hřebenový filtr**<sup>16</sup>, který si své pojmenování zasloužil podle typicky hřebenového tvaru modulové frekvenční charakteristiky – vlivem rovnoměrného rozložení nulových bodů na jednotkové kružnici. Obr. 4.5 ukazuje, že kromě utlumení základních dějů s periodou 12 měsíců jsou hřebenovým filtrem utlumeny i vyšší harmonické složky a částečný útlum nastává také pro frekvenční složky mezi body nulového a jednotkového přenosu. Analytik by si měl tedy klást před použitím sezónního diferencování otázku, zda zkreslení, které hřebenovým filtrem do očištěné verze časové řady zavede, skutečně nevádí při dalších krocích konstrukce modelu.

Alternativou k sezónnímu diferencování je zahrnutí členu  $AR(N)$  nebo  $MA(N)$  do modelu, což je jednak ukázáno ve schématu na obr. 4.4 a což také vyplývá z dalšího textu o určování řádů  $p$  a  $q$  u členů  $AR(p)$  a  $MA(q)$  v modelu  $ARMA(p,q)$ .

Struktura modelu  $AR(p)$  pro stacionární časovou řadu se střední hodnotou  $\mu$  je vyjádřena na obr. 4.6 a v následující rovnici:

$$x(n) = \mu + a_1x(n-1) + a_2x(n-2) + \dots + a_px(n-p) + v(n). \quad (4.44)$$



**Obr. 4.6** Model vzniku časové řady  $x(n)$  realizací náhodného procesu  $AR(p)$  si lze představit jako IIR filtr řádu  $p$  buzený bílým šumem  $v(n)$ .

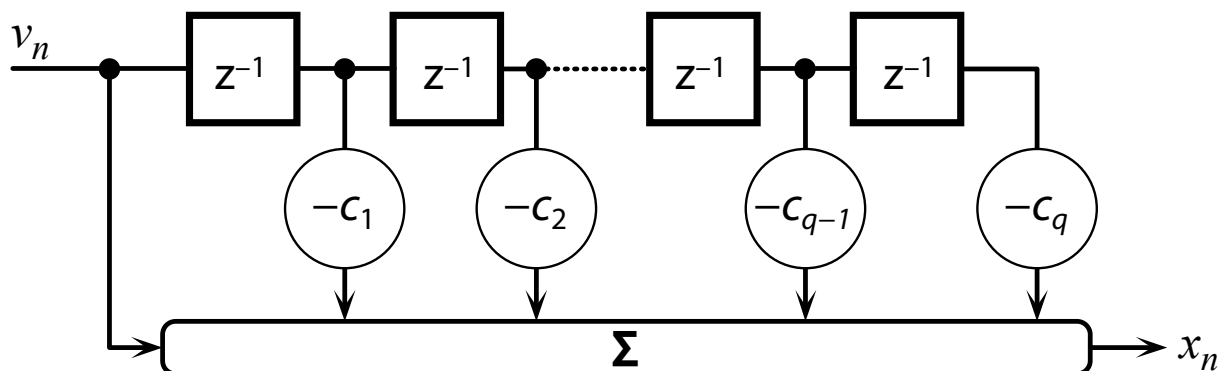
Diferenční rovnici (4.44) lze interpretovat jako lineární regresi aktuální hodnoty posloupnosti  $x(n)$  proti jedné a více jejím předchozím hodnotám. Jinými slovy lze rovnici popsat tak, že aktuální hodnota časové řady generované náhodným procesem  $AR(p)$  je dána lineární kombinací jejich  $p$  předchozích hodnot a náhodným přírůstkem  $v(n)$ , který v čase s indexem vzorku  $n$  tvoří v procesu jedinou novou informaci.

Struktura modelu  $MA(q)$  pro stacionární časovou řadu se střední hodnotou  $\mu$  je vyjádřena na obr. 4.7 a v následující rovnici (4.45). Jde opět o lineární regresi aktuálního

<sup>16</sup> Nejjednoduššímu hřebenovému filtru odpovídá diferencování s přenosovou funkcí  $H(z)=1-z^{-2}$ .

vzorku časové řady  $x(n)$ , ovšem tentokrát ne oproti jejím zpožděným hodnotám, ale oproti hodnotám bílého šumu  $v(n)$ :

$$x(n) = \mu + v(n) - c_1 v(n-1) - c_2 v(n-2) - \dots - c_q v(n-q). \quad (4.45)$$



**Obr. 4.7** Model vzniku časové řady  $x(n)$  realizací náhodného procesu  $MA(q)$  si lze představit jako FIR filtr řádu  $q$  buzený bílým šumem  $v(n)$ .

Kombinovaný model  $ARMA(p, q)$  pro stacionární časovou řadu se střední hodnotou  $\mu$  v sobě zahrnuje (4.44) i (4.45):

$$x(n) = \mu + a_1 x(n-1) + a_2 x(n-2) + \dots + a_p x(n-p) + v(n) - c_1 v(n-1) - c_2 v(n-2) - \dots - c_q v(n-q). \quad (4.46)$$

Pro určení struktury a složitosti modelu se využívá zkušeností, experimentování a srovnávání empirické autokorelační funkce pozorované časové řady s teoretickými tvary autokorelačních funkcí vybraných náhodných procesů. Autokorelační funkce stacionárního náhodného procesu je závislá na zpoždění mezi vzorky  $k$  a dále na parametrech procesu. Autokorelační funkce autoregresních náhodných procesů mají tvar klesající exponenciály nebo exponenciálně tlumené vlny. U náhodných procesů s klouzavým průměrem autokorelační funkce klesá náhle k nule od určitého konečného zpoždění. Kombinované náhodné procesy  $ARMA$  se vyznačují tím, že jejich autokorelace jsou pro úvodní zpoždění  $k=1,2,\dots,q$  komplikovaně závislé na parametrech členů  $AR$  i  $MA$ , avšak pro zpoždění  $k>q$  jsou již závislé pouze na parametrech autoregresního členu, a jejich průběh tedy opět vykazuje tvar klesající exponenciály nebo tlumené vlny. Jednoduchým příkladem je teoretická autokorelační funkce stacionárního náhodného procesu  $AR(1)$ :

$$R_{xx}(|k|) = E\{x(n)x(n-|k|)\} = a_1^{|k|}, \quad (4.47)$$

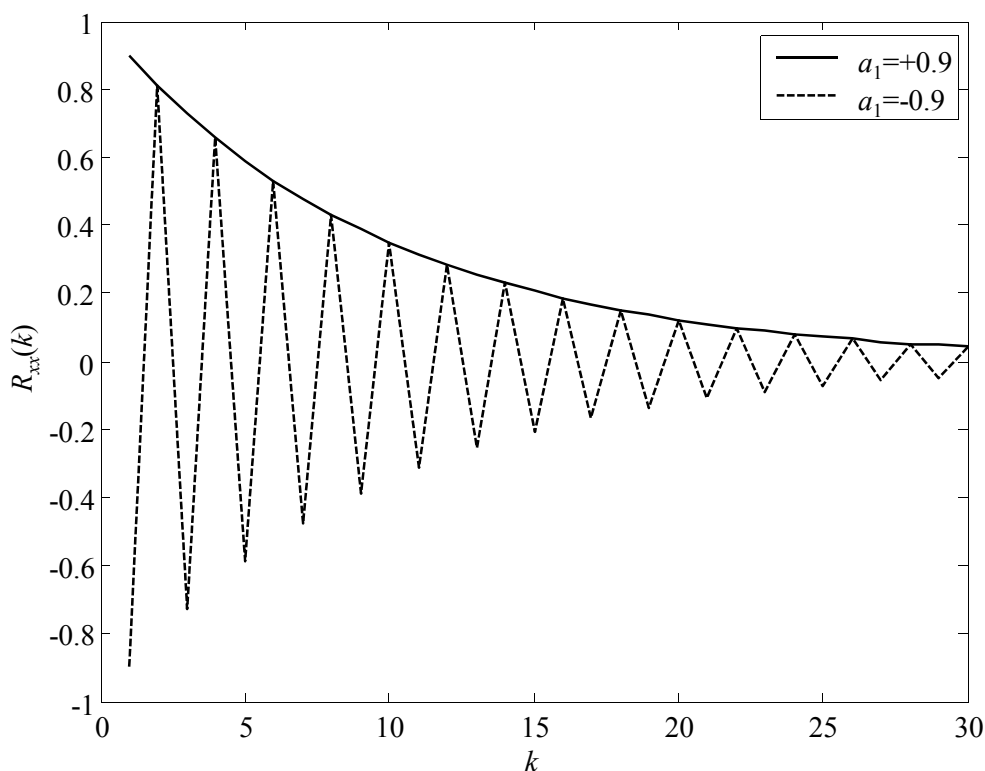
což je vlastně geometrická řada, která monotónně klesá, pokud je parametr  $a_1$  kladný a která má tvar tlumené vlny, pokud je parametr  $a_1$  záporný, viz obr. 4.8.

Tab. 4.1 sumarizuje heuristická pravidla pro určení struktury modelu na základě srovnávání empirických a teoretických autokorelačních funkcí. Pravidla zahrnují také

určování řádů  $p$  a  $q$ , a to na základě **parciální autokorelační funkce**. Hodnota parciální autokorelační funkce časové řady  $x(n)$  pro zpoždění  $k$  představuje závislost mezi dvěma vzorky časové řady  $x(n)$  a  $x(n-k)$ , přičemž do této závislosti se nezapočítává lineární vliv vzorků ležících mezi nimi. Jinými slovy se jedná o autokorelační koeficient pro zpoždění  $k$ , očištěný od vlivu hodnot vzorků  $x(n-1)$ ,  $x(n-2)$ , ...,  $x(n-k+1)$ . Podle [8] je parciální autokorelační funkce definována jako:

$$R_{xx}^*(k) = \text{korelace} [x(n) - \beta_1 x(n-1) - \beta_2 x(n-2) - \dots - \beta_{k-1} x(n-k+1), x(n-k)], \quad (4.48)$$

kde  $\beta = [\beta_1, \beta_2, \dots, \beta_{k-1}]$  je vektor regresních koeficientů. Průběh parciální autokorelační funkce je využíván pro určování řádu  $p$  autoregresního modelu: sleduje se zpoždění, pro které parciální autokorelační funkce nabývá nulové nebo zanedbatelné hodnoty, a hodnota zpoždění předchozího je pak dosazena za  $p$ , viz také tab. 4.1.



**Obr. 4.8** Teoretické průběhy normované autokorelační funkce pro různé parametry stacionárního náhodného procesu  $AR(1)$ .

Jednou z metod pro odhad numerických hodnot parametrů ve zvolené struktuře modelu je řešení **Yuleho-Walkerových rovnic**. V této metodě se rovněž uplatňují autokorelační a autokovarianční funkce náhodných procesů. Autoregresní model řádu  $p$  stacionární časové řady s nulovou střední hodnotou lze podle (4.44) zapsat také touto rovnicí:

$$x(n+1) = \sum_{j=1}^p a_j x(n-j+1) + v(n+1). \quad (4.49)$$

Pro zavedení autokorelací do rovnice (4.49) se každá strana této rovnice znásobí  $x(n)$  a aplikuje se operátor očekávané hodnoty (neboli souborového průměru):

$$E\{x(n)x(n+1)\} = \sum_{j=1}^p (a_j E\{x(n)x(n-j+1)\}) + E\{x(n)v(n+1)\}. \quad (4.50)$$

Parametry  $a_j$  jsou vytknuty před operátor  $E$ , neboť se jedná o deterministické a nikoli náhodné proměnné. Náhodné přírůstky  $v(n+1)$  nejsou nijak závislé na předchozích hodnotách  $x(n)$ , a tak bude druhý člen na pravé straně rovnice (4.50) nulový. Vzhledem k tomu, že autokorelační funkce jsou sudé funkce, lze (4.50) přepsat na:

$$R_{xx}(1) = \sum_{j=1}^p (a_j R_{xx}(j-1)). \quad (4.51)$$

Pokud se obě strany rovnice (4.49) vynásobí  $x(n-1)$ , lze stejnými úpravami jako v rovnicích (4.50)-(4.52) dospět k rovnici autokorelační funkce pro zpoždění  $k=2$ :

$$R_{xx}(2) = \sum_{j=1}^p (a_j R_{xx}(j-2)). \quad (4.52)$$

Obdobně lze odvodit rovnici autokorelační funkce pro maximální zpoždění mezi vzorky  $k=p$ :

$$R_{xx}(p) = \sum_{j=1}^p (a_j R_{xx}(j-p)). \quad (4.53)$$

Pro všechna možná zpoždění  $k=1, 2, \dots, p$  lze sestavit následující soustavu Yuleho-Walkerových rovnic:

$$\begin{aligned} R_{xx}(1) &= a_1 R_{xx}(0) + a_2 R_{xx}(1) + a_3 R_{xx}(2) + \dots + a_{p-1} R_{xx}(p-2) + a_p R_{xx}(p-1), \\ R_{xx}(2) &= a_1 R_{xx}(1) + a_2 R_{xx}(0) + a_3 R_{xx}(1) + \dots + a_{p-1} R_{xx}(p-3) + a_p R_{xx}(p-2), \\ &\dots \\ R_{xx}(p-1) &= a_1 R_{xx}(p-2) + a_2 R_{xx}(p-3) + a_3 R_{xx}(p-4) + \dots + a_{p-1} R_{xx}(0) + a_p R_{xx}(1), \\ R_{xx}(p) &= a_1 R_{xx}(p-1) + a_2 R_{xx}(p-2) + a_3 R_{xx}(p-3) + \dots + a_{p-1} R_{xx}(1) + a_p R_{xx}(0). \end{aligned} \quad (4.54)$$

Soustavu lze přepsat do maticové podoby:

$$\underbrace{\begin{pmatrix} R_{xx}(1) \\ R_{xx}(2) \\ \vdots \\ R_{xx}(p-1) \\ R_{xx}(p) \end{pmatrix}}_{\mathbf{r}} = \underbrace{\begin{pmatrix} 1 & R_{xx}(1) & R_{xx}(2) & \dots & R_{xx}(p-2) & R_{xx}(p-1) \\ R_{xx}(1) & 1 & R_{xx}(1) & \dots & R_{xx}(p-3) & R_{xx}(p-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{xx}(p-2) & R_{xx}(p-3) & R_{xx}(p-4) & & 1 & R_{xx}(1) \\ R_{xx}(p-1) & R_{xx}(p-2) & R_{xx}(p-3) & & R_{xx}(1) & 1 \end{pmatrix}}_{\mathbf{R}} \underbrace{\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{p-1} \\ a_p \end{pmatrix}}_{\mathbf{a}}. \quad (4.55)$$

Stručnější podoba Yuleho-Walkerových rovnic pro autoregresní stacionární náhodné procesy má tvar:

$$\mathbf{R}\mathbf{a} = \mathbf{r}. \quad (4.56)$$

Matice  $\mathbf{R}$  se označuje také jako *autokorelační matice*. Jedná se o symetrickou čtvercovou matici s tzv. *Toeplitzovou strukturou*, která má ve všech klesajících diagonálách shodné prvky. Výpočtem inverzní matice k autokorelační matici  $\mathbf{R}$  lze vypočítat nebo odhadnout<sup>17</sup> numerické hodnoty parametrů modelu ve vektoru  $\mathbf{a}$ :

$$\hat{\mathbf{a}} = \mathbf{R}^{-1}\mathbf{r}. \quad (4.58)$$

Yuleho-Walkerovy rovnice pro další typy stacionárních procesů  $\text{MA}(q)$  a  $\text{ARMA}(p,q)$  lze odvodit obdobně jako pro procesy  $\text{AR}(p)$ . Soustava Yuleho-Walkerových rovnic je soustavou lineárních rovnic, ovšem pouze v případě ryze autoregresních náhodných procesů. Odhad parametrů u  $\text{MA}(q)$  a  $\text{ARMA}(p,q)$  procesů vede na řešení soustavy nelineárních rovnic, což je přirozeně obtížnější úloha, než je tomu v případě lineárních rovnic. Výsledné rovnice pro všechny typy náhodných procesů lze najít např. v [8]. Řešení rovnic se v dnešní době ponechává v režii výpočetních programů, jako je např. Matlab, kde je tato metoda implementována v balíčku funkcí *System Identification Toolbox*.

Důležitou součástí návrhu modelu časové řady je na závěr jeho validace, a to formou kontroly reziduí mezi časovou řadou a její predikovanou verzí. Rezidua by měla v ideálním případě představovat realizaci bílého šumu. V časovém průběhu se sleduje, zda rozdělení reziduí vykazuje konstantně nulovou střední hodnotu a konstantní rozptyl. Pomocí autokorelační funkce posloupnosti reziduí se vyšetřuje jejich vzájemná nezávislost<sup>18</sup>. Při nesplnění těchto kritérií je třeba hledat jiný, vhodnější model. To pak znamená provést znovu krok identifikace modelu, přičemž pro nové určení řádů modelu  $p$  a  $q$  mohou poskytnout

<sup>17</sup> Jsou-li autokorelační koeficienty v matici  $\mathbf{R}$  odhadnuty z pozorovaných časových řad pomocí časových průměrů, nelze hovořit přímo o výpočtu numerických hodnot parametrů modelu, ale raději jen o odhadu parametrů modelu. Pokud jsou k danému náhodnému procesu k dispozici přesné hodnoty autokorelačních koeficientů, nebo je lze vypočítat pomocí souborových průměrů, pak lze hovořit přímo o výpočtu a ne o odhadu parametrů modelu.

<sup>18</sup> Někteří autoři [13] doporučují ověřovat u reziduí jejich normální rozdělení. Definice bílého šumu, viz (4.43), se však neopírá o určité rozdělení pravděpodobnosti. Požadavek na normální rozdělení je tedy dodatečný. Jestliže je rozdělení bílého šumu normální, pak se takový náhodný proces nazývá *gaussovský bílý šum* [19].

určité vodítko právě výsledky z analýzy reziduí. Až po té, co je nalezen vhodný model ve smyslu zpětného ověření předpokladů kladených na náhodná rezidua, je možné stanovit kvalitu předpovídání modelu pomocí chyby predikce ve stanoveném horizontu predikce.

**Tab. 4.1** Volba struktury modelu na základě srovnávání empirické autokorelační funkce vyšetřované časové řady s teoretickými tvary autokorelačních funkcí náhodných procesů. Heuristická pravidla pro tato srovnávání zahrnují také tvar parciální autokorelační funkce.

Tvar autokorelační funkce	Identifikace modelu
Exponenciála klesající k nule.	Model AR( $p$ ). Pro určení $p$ se vychází z parciální autokorelační funkce, která je pro AR( $p$ ) proces nulová od zpoždění $p+1$ .
Změny kladných a záporných hodnot, postupný pokles k nule.	Model MA( $q$ ). Řád $q$ odpovídá hodnotě zpoždění, od kterého je autokorelační funkce nulová nebo zanedbatelná.
Jeden nebo několik vrcholů, zbytek zanedbatelný, nulový.	Kombinovaný model ARMA.
Průběh klesající až po několika zpožděních.	
Všechny hodnoty kromě počátku zanedbatelné, nulové.	Data jsou náhodná, generuje je bílý šum.
Vysoké hodnoty opakující se ve stejných intervalech.	Zahrnout AR člen s řádem odpovídajícím periodě.
Neklesá k nule.	Nejedná se o stacionární časovou řadu.

Zvyšování řádů  $p$  a  $q$  nemusí vždy vést k modelům s lepšími výsledky (věrněji odpovídajícím modelovaným časovým řadám). Nadměrné zvyšování složitosti modelu vede nutně k větším nejistotám, neboť odhady parametrů modelu jsou navázány na znalost statistických momentů a ty jsou zpravidla k dispozici pouze v jejich empirické podobě (jedná se o výběrové statistické momenty získané z pozorovaných časových řad) a nikoli v podobě teoretické. Cílem konstrukce modelu by mělo být vždy nalezení nejjednoduššího možného modelu, který ještě dobře popisuje dynamiku systému.

### 4.3 Exponenciální vyhlazování a predikce

Exponenciální vyhlazování je postup pro kontinuální revizi predikce s využitím nejnovějších hodnot modelované časové řady. Starším vzorkům se přiřazují exponenciálně klesající váhy, a tak se na výpočtu predikované hodnoty pozorované časové řady podílejí její novější vzorky větší vahou než ty starší. Kromě modelování či predikce se využívá exponenciální vyhlazování také za účelem potlačení náhodného rušení ve zpracovávané časové řadě.

Výpočet krátkodobé predikce časové řady pomocí exponenciálního vyhlazování je dán rekurentní rovnicí [13]:

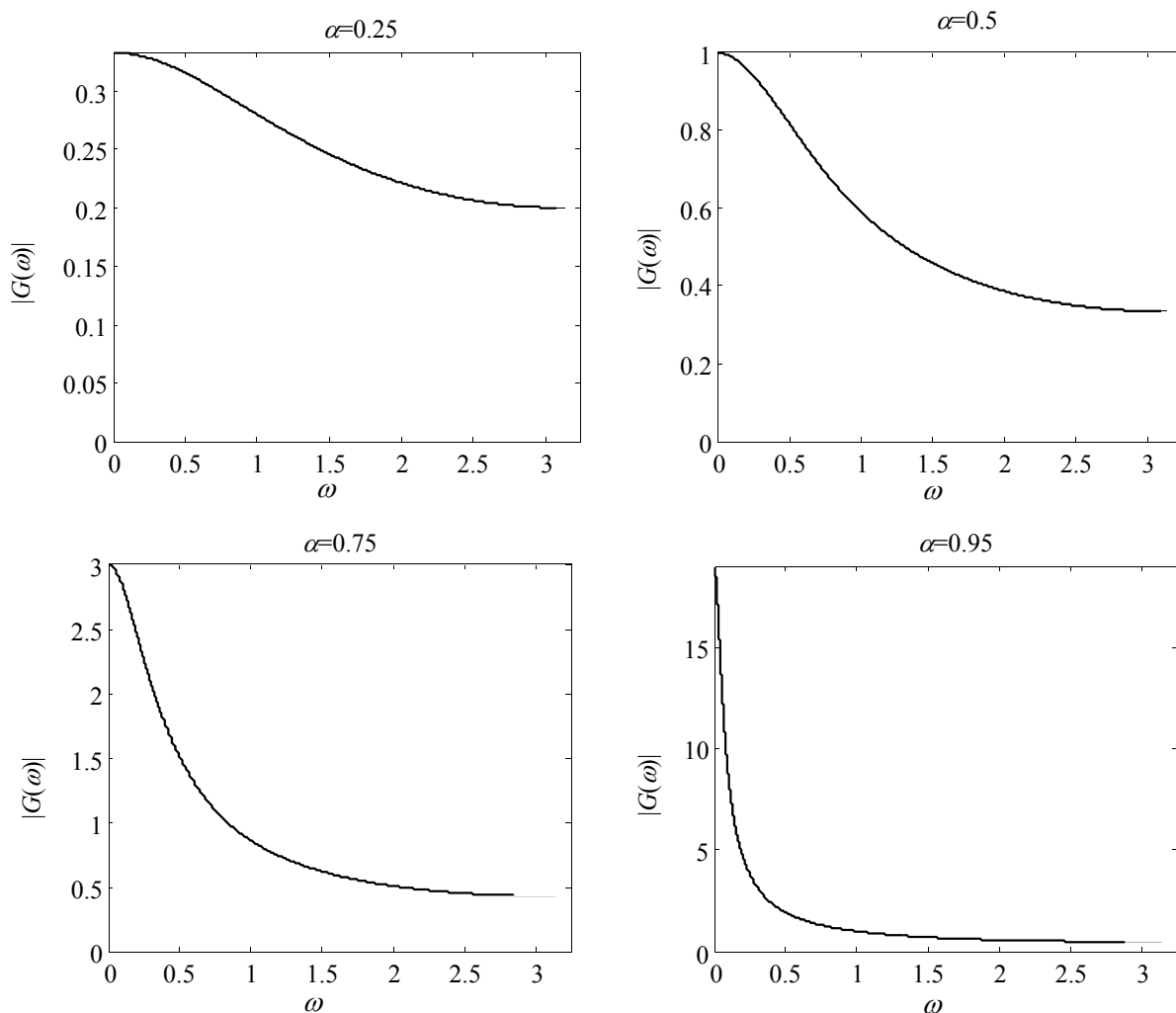
$$\hat{x}(n) = \alpha x(n) + (1 - \alpha) \hat{x}(n-1), \quad 0 < \alpha < 1, \quad (4.59)$$

ve které představuje  $\hat{x}(n)$  hodnotu vyhlazené časové řady a zároveň je i predikcí pro její příští vzorek  $x(n+1)$ . Parametr  $\alpha$ , který se označuje také jako **koeficient zapomínání**, má vliv na exponenciální pokles vah, pomocí kterých je vlastně realizován vážený průměr ze všech předchozích vzorků, neboť (4.59) lze přepsat také na [13]:



$$\begin{aligned}
\hat{x}_n &= \alpha x(n) + (1-\alpha)\hat{x}(n-1) = \alpha x(n) + (1-\alpha)[\alpha x(n-1) + (1-\alpha)\hat{x}(n-2)] = \\
&= \alpha x(n) + \alpha(1-\alpha)x(n-1) + (1-\alpha)^2[\alpha x(n-2) + (1-\alpha)\hat{x}(n-3)] = \\
&= \alpha \sum_{k=0}^{n-1} (1-\alpha)^k x(n-k) + (1-\alpha)^n \hat{x}(0).
\end{aligned} \tag{4.60}$$

Za iniciální hodnotu  $\hat{x}(0)$  se dosazuje první hodnota časové řady  $x(0)$  nebo někdy také průměr několika prvních pozorovaných vzorků časové řady.



**Obr. 4.9** Modulové frekvenční charakteristiky filtrů IIR pro realizaci exponenciálního vyhlazování s různými hodnotami koeficientu zapomínání  $\alpha$ .

Volba hodnoty koeficientu zapomínání  $\alpha$  může být ponechána na optimalizační proceduře, která najde hodnotu  $\alpha$  minimalizující střední kvadratickou chybu predikce pro konkrétní pozorovanou časovou řadu. Alternativně může být rozvaha nad hodnotou parametru  $\alpha$  provedena ve frekvenční doméně. Exponenciální vyhlazování lze realizovat pomocí diskrétní soustavy se strukturou odpovídající filtru IIR s přenosovou funkcí  $H(z) = \alpha / (1 - \alpha z^{-1})$ . Obr. 4.9 ukazuje modulové frekvenční charakteristiky exponenciálního vyhlazování pro různé hodnoty  $\alpha$ . Jde o dolní propusti lišící se mezi sebou šířkou propustného

pásma. Je-li pro konkrétní aplikaci důležité předpovídat pouze pozvolné změny v trendu časové řady, pak bude pro predikční model zvolen koeficient  $\alpha$  co nejbližší číslu 1. Pokud naopak bude v jiné aplikaci důležité předpovídat i děje s rychlou dynamikou, bude zřejmě zvolena nižší hodnota koeficientu  $\alpha$ . Má-li být rozvaha ve frekvenční doméně smysluplná, je nutno do ní zahrnout také frekvenční spektrum pozorované časové řady.

Exponenciální vyhlazování se označuje někdy také jako model EWMA (*exponentially weighted moving average*). Výše v textu však bylo ukázáno, že exponenciální vyhlazování ve své originální podobě podle rovnice (4.59) se realizuje IIR filtrem, a podobá se tak více autoregresním modelům než modelům s klouzavým průměrem. Pokud se rovnice (4.59) přepíše do tvaru pro předpovídání formou korekce chyby predikce:

$$\hat{x}(n) = \alpha x(n) + (1 - \alpha)\hat{x}(n-1) = \alpha(x(n) - \hat{x}(n-1)) + \hat{x}_{n-1} = \alpha v_n + \hat{x}(n-1), \quad (4.61)$$

je evidentní, že exponenciální vyhlazování svou strukturou odpovídá modelu ARIMA(1,1,0). Označovat exponenciální vyhlazování jako model EWMA je tedy nevhodné<sup>19</sup>.

## 4.4 Úlohy

**ÚLOHA 4.1** Sestavte model pro časovou řadu pozorovanou při měřeních koncentrace CO<sub>2</sub> v ovzduší. Soubor dat, který je dostupný na: <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4411.htm> obsahuje měsíční průměrné koncentrace CO<sub>2</sub> na observatoři Mauna Loa v letech 1974 – 1987. Koncentrace jsou udávány v bezrozměrných jednotkách, jako molární zlomek (*mole fraction*). Při konstrukci modelu využijte aditivní dekompozici společně s Boxovou-Jenkinsovou metodikou.

**Tab. 4.2** Ukázka datového souboru pro modelovanou časovou řadu v úloze 4.1. Soubor obsahuje celkem 161 řádků, kde každý řádek reprezentuje jedno měření průměrné měsíční koncentrace CO<sub>2</sub>.

CO <sub>2</sub>	Rok a měsíc	Rok	Měsíc
333.13	1974.38	1974	5
332.09	1974.46	1974	6
331.10	1974.54	1974	7
329.14	1974.63	1974	8
327.36	1974.71	1974	9
327.29	1974.79	1974	10
328.23	1974.88	1974	11
329.55	1974.96	1974	12
330.62	1975.04	1975	1
331.40	1975.13	1975	2
331.87	1975.21	1975	3
333.18	1975.29	1975	4
333.92	1975.38	1975	5

<sup>19</sup> Model EWMA může být podle definice v kapitole MA model s řádem, který odpovídá konečnému počtu exponenciálně klesajících vah podílejících se na výpočtu váženého průměru. U originálního exponenciálního vyhlazování je však počet vah vzhledem k rekurentní rovnici (4.59) teoreticky nekonečný, což odpovídá realizaci pomocí filtru IIR.

Data, která jsou na uvedené adrese k dispozici v textovém formátu, je třeba nejdříve převést běžnými softwarovými prostředky do souboru ve formátu CSV (*comma separated values*), kde jsou na každém řádku k dispozici všechny numerické hodnoty k jednomu měření oddělené čárkou, tabulátorem nebo středníkem. Každé měření je reprezentováno čtyřmi parametry, z toho tři reprezentují údaje na časové ose, jak ukazuje tab. 4.2.

Kód skriptu, který v Matlabu realizuje načtení dat, jejich vykreslení, analýzu trendu a periodických komponent je na obr. 4.10. Na obr. 4.11 je časová řada vykreslena spolu s vypočítanou lineární trendovou funkcí. Obr. 4.12 ukazuje grafickou analýzu čar v diskretním spektru vypočteném z časové řady očištěné od lineárního trendu. Výsledkem analýzy jsou dvě významné harmonické komponenty s periodou 12 a 6 měsíců. Sezónní diferencování s periodou 12 vzorků utlumí obě tyto komponenty (i další vyšší harmonické), jak bylo vysvětleno výše v textu a na obr. 4.5.

```
%%% ULOHA 4.1A MODELOVANI CASOVE RADY S KONC. CO2 V OVZDUSI
%%% PRVNI CAST: ANALYZA TRENDU A PERIODICKYCH SLOZEK

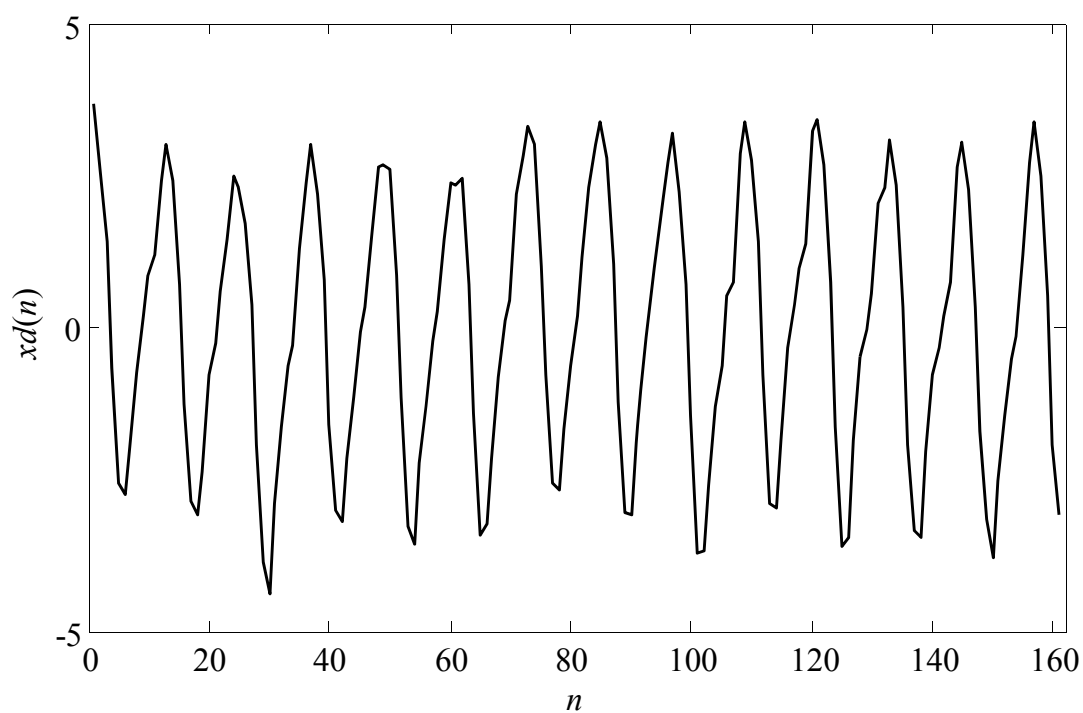
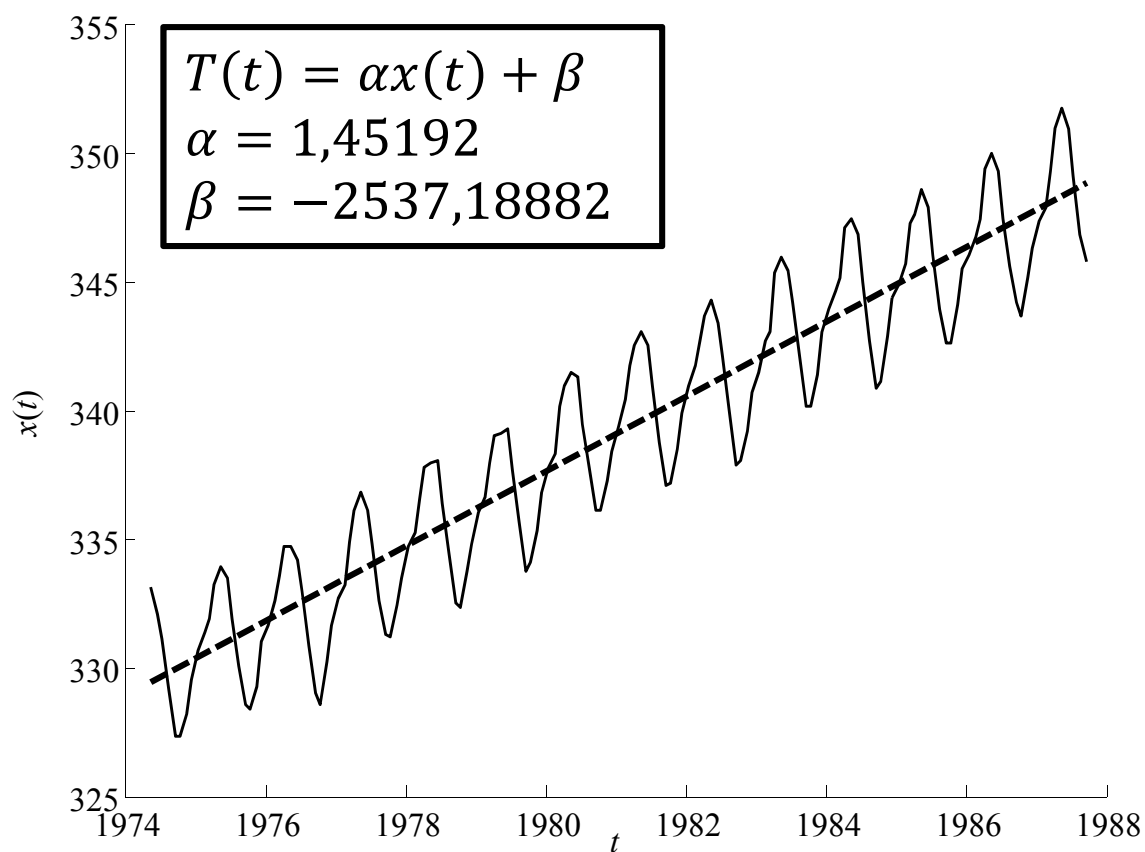
%%%%% nahrani dat a vykresleni casove rady
DATA = dlmread('co2.csv',';'); % csv soubor = hodnoty oddelene
strednikem
x = DATA(:,1); % koncentrace jsou v prvni sloupci
t = DATA(:,2); % casove udaje ve druhem sloupci
figure, hold on;
plot(t,x,'k'); % kreslime puvodni data
xlabel('čas'); ylabel('CO_2');

%%%%% ocisteni od trendu
c = polyfit(t,x,1); % prolozeni polynomem prvniho radu
trend = c(1)*t+c(2); % trendova posloupnost
h = line([min(t) max(t)],[trend(1) trend(end)]);
set(h,'LineWidth',2,'Color','red'); % trend vykreslime tlustou carou
xd = x - trend; % ocisteni od linearniho trendu

%%%%% zjisteni periody sezonnich/cyklickych jevu
spektrum = fft(xd); % vypocet diskretniho
spektra
faxis = linspace(0,1,length(spektrum)); % normovana frekvencni osa
figure, stem(faxis,abs(spektrum),'.'); % vykresleni modulu spektra
xlabel('normovana frekvence (fs...1)');
ylabel('|DFT(xd)|');

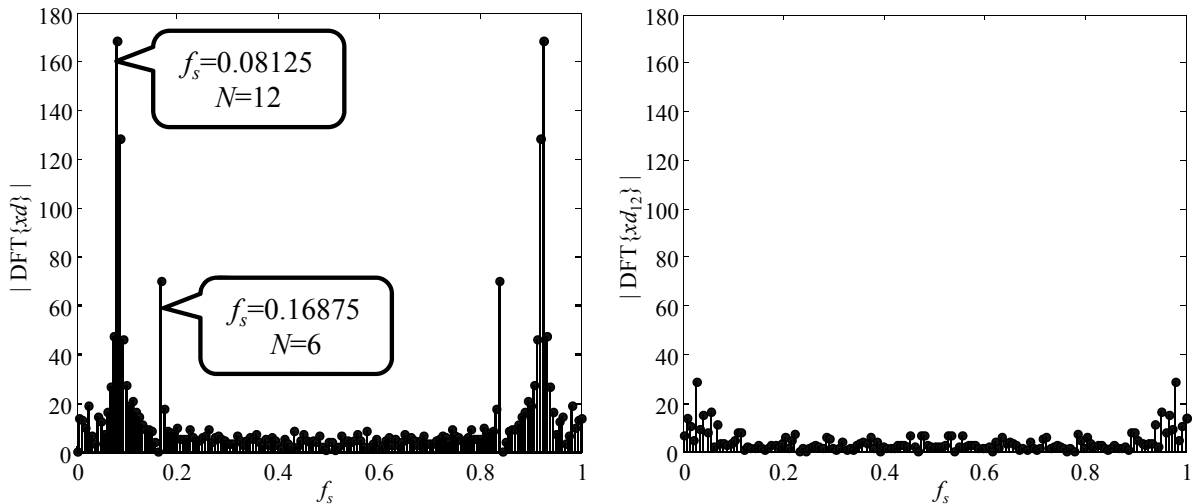
xd12 = xd(13:end)-ddata(1:end-12);
spektrum12 = fft(xd12); % vypocet diskretniho
spektra
faxis12 = linspace(0,1,length(spektrum12)); % normovana frekvencni
osa
figure, stem(faxis12,abs(spektrum12),'.');
xlabel('normovana frekvence (fs...1)');
ylabel('|DFT(xd12)|');
```

**Obr. 4.10** Úloha 4.1 – modelování časové řady s měsíčními koncentracemi CO<sub>2</sub>.  
První část: dekompozice.



**Obr. 4.11** Nahoře: časová řada  $x(t)$  s měsíčními koncentracemi  $\text{CO}_2$  v ovzduší a lineární trendová funkce  $T(t)$  vypočítaná z  $x(t)$  pomocí metody nejmenších čtverců. Dole: časová řada  $x(t)$  očištěná od lineárního trendu a s indexem vzorku  $n$  jako nezávisle proměnnou.

Pro následující konstrukci stacionárního modelu budou zvažovány obě varianty nesystematické složky časové řady  $x(n)$ : 1. posloupnost  $xd(n)$ , která se od původní časové řady liší pouze odečteným trendem a 2. posloupnost  $xd_{12}(n)$ , kterou od původní  $x(n)$  odlišuje navíc i sezónní diferencování o 12 vzorků.



**Obr. 4.12** Zjišťování periody významných cyklických/sezónních jevů. Amplitudová část diskrétního spektra: vlevo - časové řady očištěné od lineárního trendu  $dx(n)=x(n)-T(n)$ , vpravo - časové řady očištěné od lineárního trendu a po sezónním diferencování s periodou 12 vzorků. Při zjišťování periody ve frekvenční doméně je nutno mít na paměti, že spektrum je diskrétní a frekvenční osa je vzorkována podle počtu vzorků časové řady, viz (1.9) pro definici DFT. V časové řadě  $xd(n)$  je 161 vzorků a v časové řadě  $xd_{12}(n)$  je po sezónním diferencování 149 vzorků.

```
%%% ULOHA 4.1B MODELOVANI CASOVE RADY S KONC. CO2 V OVZDUSI
%%% DRUHA CAST: KONSTRUKCE MODELU

%%%%%%%% ZJISTOVANI DALSICH ZAVISLOSTI V NESYST. SLOZCE CASOVE RADY
figure, [Rxd,Cxd]=acfpacf(xd,50,50,1,0.05,0.05,'xd');
figure, [Rxd12,Cxd12]=acfpacf(xd12,50,50,1,0.05,0.05,'xd12');
y = iddata(xd12(:),[]); % datova struktura pro Sys. Ident. Toolbox
model = ar(y,2,'yw'); % reseni Yulovych-Walkerovych rovnic
% vysledny model je typu idpoly: A(q)y(t) = e(t)

%%%%%%%% VALIDACE MODELU
xpred = zeros(length(t),1); %do xpred ulozone predikovane hodnoty
a1=abs(model.A(2)); a2=abs(model.A(3));
for n = 15:length(t)
    xpred(n) = a1*x(n-1)+a2*x(n-2)+x(n-12)-a1*x(n-13)-a2*x(n-14);
end
figure, subplot(3,1,1), stem(e,'.'); ylabel('residua e(n)')
subplot(3,1,2), stem(acf_e(start:end),'.'); ylabel('acf_e(k)');

%%%%%%%% JAK MODEL PREDPOVIDA A JAKA JE SKUTECNOST
subplot(3,1,3), plot(t(15:end),x(15:end),'k'), hold on
subplot(3,1,3), plot(t(15:end),xpred(15:end),'b');
```

```

%%%%% ROZLOZENI NULOVYCH BODU A POLU PRENOSOVE FUNKCE
a = [1 a1 a2 0 0 0 0 0 0 0 0 0 1 -a1 -a2];
sys = firlt(1,a);           % definice citatele a jmen. prenosove funkce
figure, pzmap(sys);         % rozlozeni nulovych bodu a polu

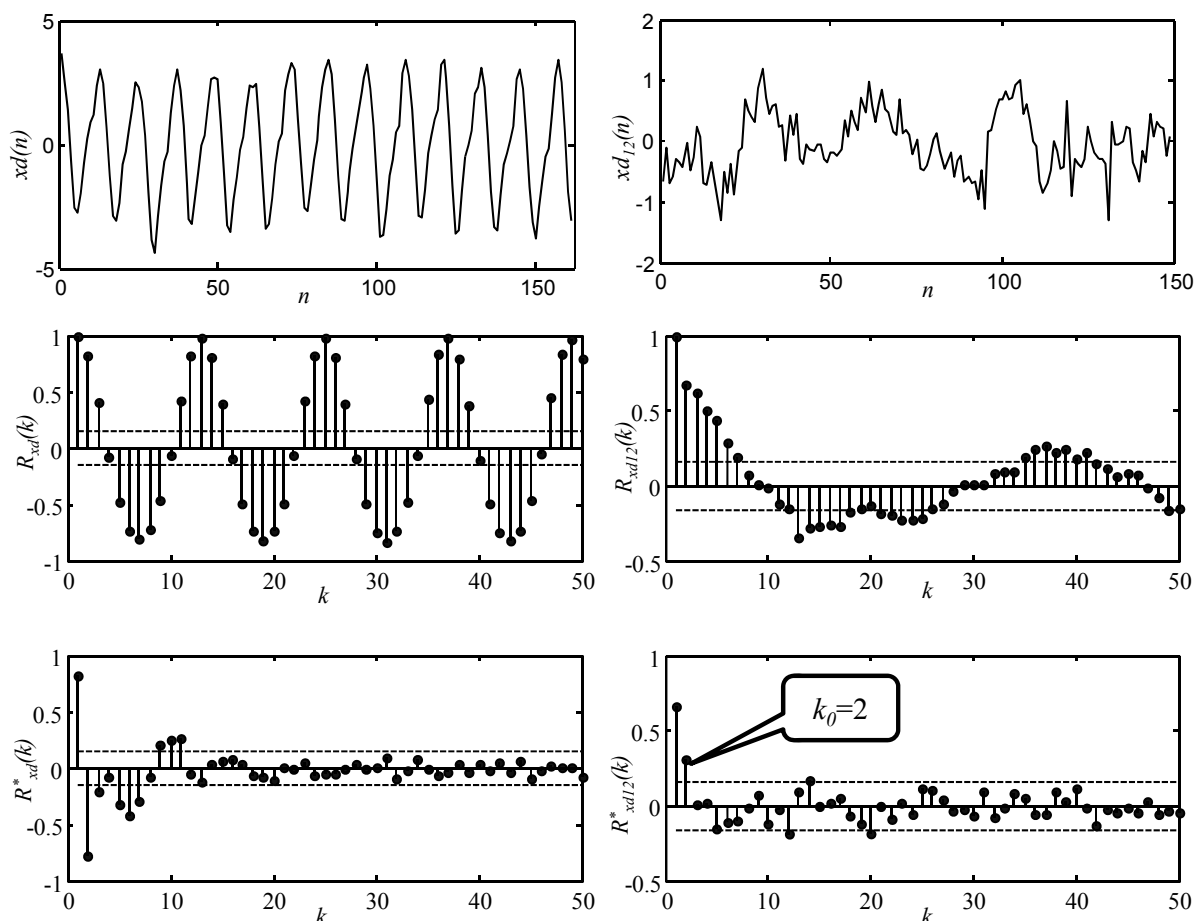
```

**Obr. 4.13** Úloha 4.1 – modelování časové řady s měsíčními koncentracemi  $\text{CO}_2$ .

Druhá část: konstrukce modelu a jeho validace s využitím balíčku funkcí

System Identification Toolbox pro Matlab.

Ve druhé části skriptu, jehož programový kód je na obr. 4.13, se zjišťují závislosti v nesystematické složce časové řady  $x(n)$ , a to pomocí autokorelační a parciální autokorelační funkce, viz obr. 4.14. Obě funkce zároveň slouží k určení řádů  $p, q$  členů  $\text{AR}(p)$  a  $\text{MA}(q)$  ve stacionárním modelu ARMA. Autokorelační funkce posloupnosti  $xd(n)$  potvrzuje výskyt periodické složky s periodou 12 vzorků – o posloupnosti  $xd(n)$  tedy nelze hovořit jako o náhodné, nesystematické časové řadě. Autokorelační funkce posloupnosti  $xd_{12}(n)$  vykazuje tvar tlumené vlny, a tak lze tuto posloupnost modelovat pomocí autoregresní soustavy. Řád této soustavy lze vyčíst z bodu useknutí v parciální autokorelační funkci, který nastává pro zpoždění  $k_0=2$ .



**Obr. 4.14** Zjišťování závislosti mezi vzorky v nesystematické složce pomocí autokorelační funkce  $R(k)$  a parciální autokorelační funkce  $R^*(k)$ .

Odhad parametrů modelu AR(2), provedený pomocí Yulových-Walkerových rovnic, vede na konstrukci stacionárního modelu:

$$xd_{12}(n) = v(n) + a_1 xd_{12}(n-1) + a_2 xd_{12}(n-2),$$

$$a_1 = 0,45543 \quad a_2 = 0,31384 .$$

Stacionární model je třeba upravit na nestacionární, a to zpětným promítnutím všech úprav, které byly provedeny při stacionarizaci původní časové řady  $x(n)$ , viz dále.

$$xd(n) = x(n) - (\alpha n + \beta),$$

$$xd_{12}(n) = xd(n) - xd(n-12) = x(n) - (\alpha n + \beta) - x(n-12) + (\alpha(n-12) + \beta) = x(n) - x(n-12),$$

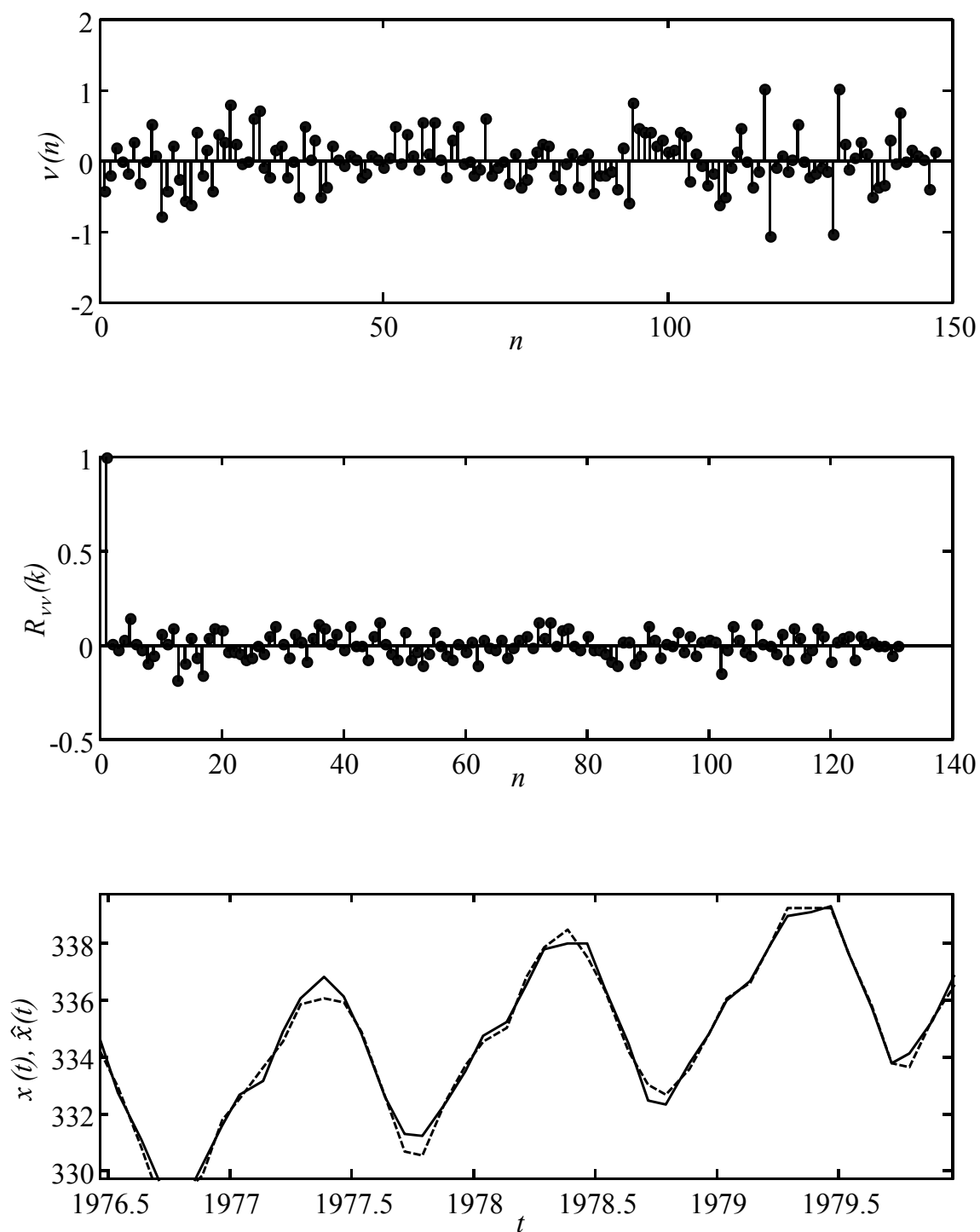
$$x(n) - x(n-12) = v(n) + a_1 [x(n-1) - x(n-13)] + a_2 [x(n-2) - x(n-14)]$$

Rovnice výsledného AR(14) modelu má tvar:

$$x(n) = v(n) + a_1 x(n-1) + a_2 x(n-2) + x(n-12) - a_1 x(n-13) - a_2 x(n-14),$$

$$a_1 = 0,45543 \quad a_2 = 0,31384 .$$

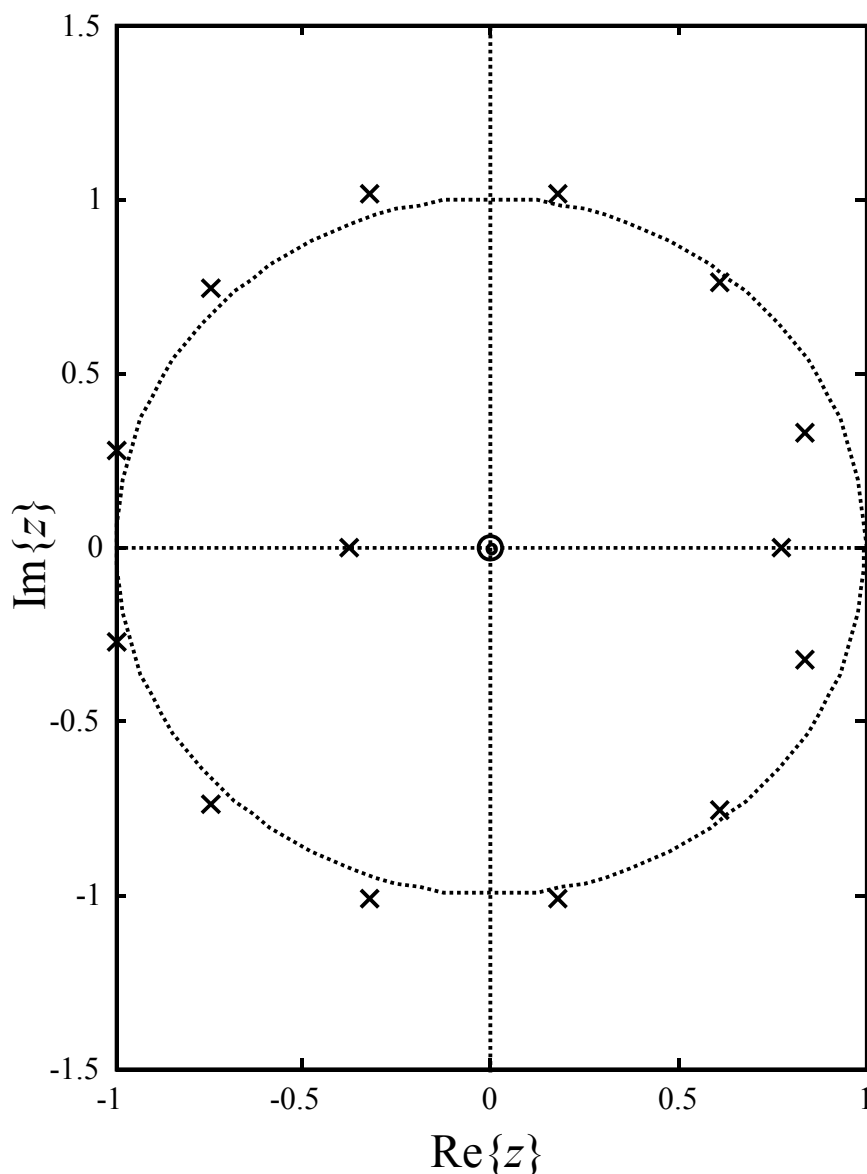
Na obr. 4.15 jsou znázorněny tři kroky validace modelu formou grafické analýzy reziduí neboli chyb predikce. Autokorelační funkce chybové posloupnosti  $R_{v_r}(k)$  se přibližuje autokorelační funkci bílého šumu. Část variability dat, kterou nebylo možné postihnout modelem, je tedy velmi náhodná.



**Obr. 4.15** Validace výsledného modelu zahrnuje a) grafickou analýzu residuí (neboli chyb predikce), b) včetně jejich autokorelační funkce a c) posouzení kvality předpovídání – vybrán detail, plná čára:  $x(n)$ , přerušovaná čára:  $\hat{x}(n)$ .

Obr. 4.16 ukazuje rozložení nulových bodů a pólů výsledného modelu AR(14) – několik pólů je vně jednotkové kružnice, a jedná se tedy o nestabilní soustavu. Tomu také odpovídá neustálý nárůst hodnot v časové řadě  $\hat{x}(n)$ , která je modelem generována.





**Obr. 4.16** Rozložení nulových bodů a pólů výsledného nestacionárního modelu  $AR(14)$ .

## 4.5 Shrnutí

Čtvrtá kapitola představuje jeden z možných pohledů na problematiku konstrukce modelů časových řad generovaných náhodnými procesy. V úvodu byly představeny vybrané statistické momenty, které se využívají pro popis vlastností náhodných časových řad. Byly vysvětleny základní rozdíly mezi časovými a souborovými průměry a dále byl čtenář seznámen s pojmy stacionarita a ergodicita, což jsou vlastnosti, které vymezují určité podmnožiny náhodných procesů. Jádro kapitoly tvoří metodika pro konstrukci stacionárních ARMA modelů náhodných časových řad podle Boxe a Jenkinse: probrány jsou zde všechny tři základní kroky od identifikace modelu přes odhad numerických hodnot parametrů modelu až po jeho validaci. Detailně byly rozebrány také možnosti a důsledky doporučených postupů pro předzpracování nestacionárních časových řad při konstrukcích modelů typu ARIMA nebo SARIMA. Byla ukázána důležitá role speciálního náhodného procesu, který se označuje jako

bílý šum. Jedna podkapitola se věnuje také jednoduchému exponenciálnímu vyhlazování pro predikci časových řad, přičemž je zde ukázáno, že se jedná vlastně o speciální model typu ARIMA. V závěru kapitoly je čtenář detailně proveden jedním z mnoha možných postupů při konstrukci modelu reálně naměřené nestacionární časové řady představující měsíční monitoring koncentrace  $\text{CO}_2$  v ovzduší.

## 5 Lineární predikce optimálními a adaptivními filtry

Konstrukce matematických modelů reálných systémů nebo procesů na základě pozorovaných časových řad pomocí postupů uvedených ve čtvrté kapitole vede ke stacionárním modelům. V případě nestacionárních časových řad je nutné v diskutovaných postupech takové posloupnosti nejdříve předzpracovat s cílem očistit je od trendu a/nebo od periodického kolísání, poté zkonstruovat model pro takto změněná data a nakonec do modelu zahrnout operace inverzní k předzpracování. Pro takovýto postup je ovšem nutné mít k dispozici apriorní znalosti o „průměrných vlastnostech“ modelovaných časových řad a dále musí charakter nestacionární složky umožňovat její oddělené modelování. V mnoha případech nelze takové znalosti a podmínky při modelování časových řad přepokládat, což platí zejména při skokových změnách podmínek vzniku časových řad a je-li cílovou aplikací modelu predikce budoucích hodnot časové řady.

Koncept *lineární predikce* byl naznačen již v první kapitole tohoto učebního textu, a to v diferenční rovnici LTI systému (1.16). Tento vztah lze interpretovat tak, že LTI systém uchovává v paměti starší vzorky vstupní i výstupní časové řady  $x(n)$ , respektive  $y(n)$  a že aktuální výstup diskrétního LTI systému lze predikovat výpočtem jejich lineární kombinace. Cílem lineární predikce je nalezení hodnot parametrů  $a_i$  a  $b_j$  v rovnici (1.16) stejně jako tomu bylo v případě modelování náhodných procesů ve čtvrté kapitole. Metodika, kterou se zabývá tato kapitola, zahrnuje *optimální filtraci* pro identifikaci neznámých, časově invariantních systémů, u kterých jsou hodnoty parametrů  $a_i$  a  $b_j$  v čase konstantní. Dále pak představuje také *adaptivní filtraci* pro identifikaci neznámých dynamických (neboli časově proměnných) systémů, u kterých jsou parametry  $a_i(n)$  a  $b_j(n)$  na čase závislé. Změnami hodnot těchto parametrů v čase je možné dosáhnout adaptivní přizpůsobení charakteristik filtru podle měnících se průměrných vlastností pozorovaných časových řad.

### 5.1 Optimální filtrace a identifikace

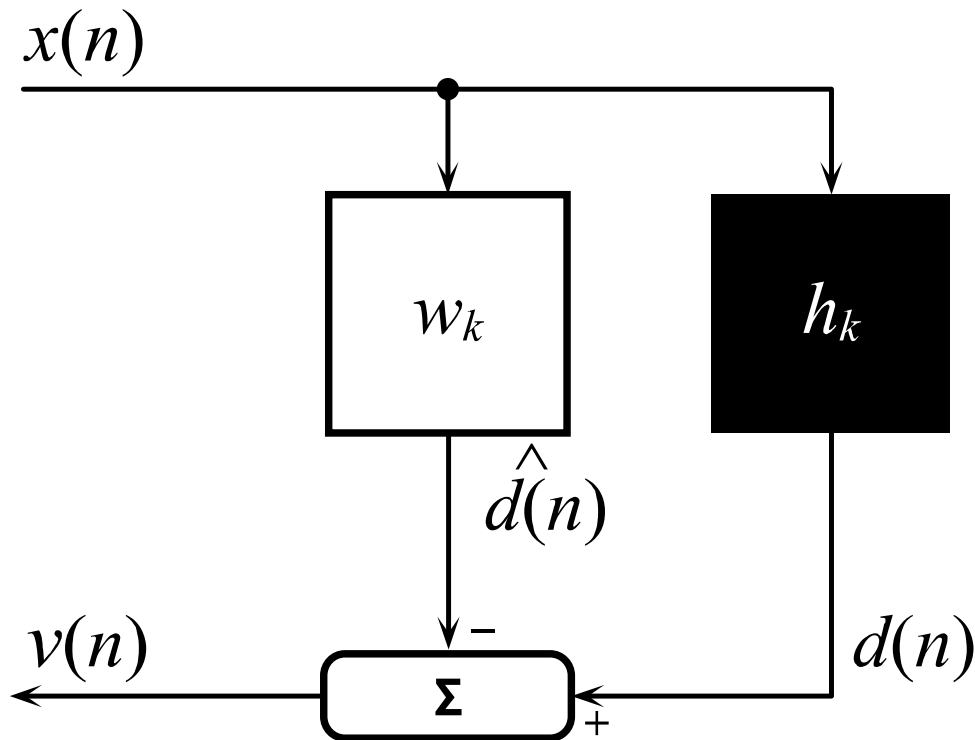
Jednou z aplikačních oblastí lineární predikce je identifikace systémů, jejíž obecný princip ukazuje obr. 5.1. Cílem identifikace je u predikčního filtru nastavit váhy  $w_k$  tak, aby vstupně–výstupní chování obou systémů bylo pokud možno shodné. Protože jsou identifikovaný systém i predikční model buzeny shodnou posloupností  $x(n)$ , jde tedy o minimalizaci rozdílu mezi výstupním signálem z identifikovaného systému  $d(n)$  a výstupním signálem z predikčního filtru  $\hat{d}(n)$ . Posloupnost residuí mezi těmito výstupy  $v(n)$  se označuje jako *chyba predikce*.

Minimalizační kritérium se odvozuje od kvadrátu chyby predikce, aby byly zohledněny stejnou měrou kladné a záporné odchylky mezi výstupními signály  $d(n)$  a  $\hat{d}(n)$ . Posloupnost chyb predikce  $v(n)$  lze považovat za realizaci náhodného procesu. Do kritéria minimalizace se proto zavádí ještě operátor očekávané hodnoty  $E$  (souborový průměr) a celé kritérium se pak označuje známým pojmem *střední kvadratická chyba* (*MSE – mean squared error*).

$$\varepsilon(n) = E\{v(n)^2\} = E\{[d(n) - \hat{d}(n)]^2\}. \quad (5.1)$$

Je-li střední kvadratická chyba uváděna s časovým indexem  $\varepsilon(n)$ , jedná se o časové sledování vývoje predikční chyby, např. v průběhu hledání vah optimálního filtru. Zápis

střední kvadratické odchylky  $\varepsilon$  bez časového indexu se využívá pro označení ustálené chyby predikce.



**Obr. 5.1** Základní schéma identifikace systémů pomocí optimální filtrace. Identifikovaný systém (černá skříňka) má neznámou impulsní charakteristiku  $h_k$ . Model identifikovaného systému v paralelní větvi je predikční filtr FIR s impulsní charakteristikou, které odpovídají váhy  $w_k$ .

Z obr. 5.1 je zřejmé, že pro různé vektory vah predikčního filtru  $\mathbf{w}$  budou výsledkem různé hodnoty střední kvadratické chyby predikce  $\varepsilon$ . Za předpokladu, že predikční filtr FIR je řádu  $M$ , uplatní se při výpočtu jeho aktuální výstupní hodnoty  $\hat{d}(n)$  pouze  $M$  posledních vzorků z časové řady  $x(n)$ . Operaci konvoluce lze zapsat pomocí vektorového součinu jako:

$$\begin{aligned}\hat{d}(n) &= \sum_{i=0}^{M-1} w_i(n)x(n-i) = \mathbf{w}_M^T(n)\mathbf{x}_M(n), \\ \mathbf{w}_M^T(n) &= [w_0(n), \dots, w_{M-1}(n)], \\ \mathbf{x}_M^T(n) &= [x(n), \dots, x(n-M+1)].\end{aligned}\tag{5.2}$$

Dosazením (5.2) do (5.1) lze získat funkční vztah mezi vahami  $\mathbf{w}$  a chybou  $\varepsilon$ :

$$\varepsilon(\mathbf{w}_M) = E\{v(n)^2\} = E\left\{\left[d(n) - \mathbf{w}_M^T \mathbf{x}_M(n)\right]^2\right\},\tag{5.3}$$

který je v následujících odstavcích předmětem podrobného zkoumání. Při identifikaci systémů pomocí optimální filtrace je cílem najít jediné, optimální nastavení predikčního filtru. Jedná se tedy o LTI soustavu s fixními parametry, a proto byl u vektoru vah  $\mathbf{w}_M$  v rovnici (5.3) záměrně vypuštěn časový index  $n$ . Zároveň je však nutno v takovém případě doplnit přirozený

požadavek na stacionaritu časových řad  $d(n)$  a  $x(n)$ . Jak bude ukázáno dále, v případě použití adaptivních predikčních filtrů tento omezující předpoklad není třeba zavádět.

Kvadrát rozdílu mezi časovými řadami  $d(n)$  a  $\hat{d}(n)$  lze rozepsat jako:

$$\left[d(n) - \mathbf{w}_M^T \mathbf{x}_M(n)\right]^2 = d^2(n) - 2\mathbf{w}_M^T d(n) \mathbf{x}_M(n) + \mathbf{w}_M^T \mathbf{x}_M(n) \mathbf{x}_M^T(n) \mathbf{w}_M. \quad (5.4)$$

Po aplikaci operátoru očekávané hodnoty na obě strany rovnice (5.4) je na levé straně získán opět výraz pro střední kvadratickou chybu:

$$\varepsilon(\mathbf{w}_M) = E\{d^2(n)\} - 2\mathbf{w}_M^T E\{d(n) \mathbf{x}_M(n)\} + \mathbf{w}_M^T E\{\mathbf{x}_M(n) \mathbf{x}_M^T(n)\} \mathbf{w}_M. \quad (5.5)$$

Na pravé straně rovnice (5.5), která je alternativou k rovnici (5.3) pro vyjádření funkčního vztahu mezi střední kvadratickou chybou a vektorem vah predikčního filtru, byly před operátory očekávané hodnoty vytknuty konstantní parametry časově invariantního predikčního filtru  $\mathbf{w}_M$ . Zbývající tři souborové průměry na pravé straně rovnice (5.5) odpovídají statistickým momentům, jejichž výčet následuje [1].

- **Střední výkon** výstupní časové řady  $\sigma_d^2 = E\{d^2(n)\}$ , který lze pro stacionární časovou řadu považovat za konstantní. Pro časové řady s nulovou střední hodnotou platí, že jejich střední výkon je roven jejich rozptylu.
- **Vektor křížových korelací** mezi vstupním a výstupním signálem identifikovaného systému  $\mathbf{p}_M = E\{d(n) \mathbf{x}_M(n)\}$ , který je pro stacionární časové řady nezávislý na čase.
- **Autokorelační matice** části vstupní časové řady  $\mathbf{R}_{MM} = E\{\mathbf{x}_M(n) \mathbf{x}_M^T(n)\}$ , která je pro stacionární časové řady nezávislá na čase.

Vzhledem k časové nezávislosti všech tří statistických momentů je minimalizační kritérium závislé pouze na nastavení vah predikčního filtru  $\mathbf{w}_M$ :

$$\varepsilon(\mathbf{w}_M) = \sigma_d^2 - 2\mathbf{w}_M^T \mathbf{p}_M + \mathbf{w}_M^T \mathbf{R}_{MM} \mathbf{w}_M. \quad (5.6)$$

Vektor křížových korelací je v následující rovnici (5.7) rozepsán na jednotlivé hodnoty korelací  $R_{xd}(k)$ , které jsou v případě stacionárních časových řad závislé pouze na časovém zpoždění  $k$  mezi vzorky a nikoli na absolutním času daným indexem vzorku  $n$ :

$$\mathbf{p}_M = \begin{pmatrix} E\{d(n)x(n)\} \\ E\{d(n)x(n-1)\} \\ \vdots \\ E\{d(n)x(n-M+1)\} \end{pmatrix} = \begin{pmatrix} R_{xd}(0) \\ R_{xd}(1) \\ \vdots \\ R_{xd}(M-1) \end{pmatrix}. \quad (5.7)$$

Skutečné hodnoty křížových korelací mezi vstupním signálem  $x(n)$  a výstupním signálem  $d(n)$  nejsou v případě neznámého systému k dispozici, a tak je nutné odhadnout je z dostupných dat. V případech, kdy je možné zavést další omezující předpoklad ergodicity

náhodných procesů, které generují vyšetřované časové řady, je možné odhady korelací  $\hat{R}_{xd}(k)$  vypočítat pomocí časových průměrů z jediné realizace časové řady:

$$\hat{R}_{xd}(k) = \frac{1}{N} \sum_{i=0}^{N-|k|-1} x(n-i)d(n-|k|-i), \quad (5.8)$$

kde  $N$  je počet vzorků v časovém okně, ze kterého jsou odhady korelací počítány. Je třeba mít na paměti, že při porušení podmínky ergodicity<sup>20</sup> je do výpočtu vah optimálního predikčního filtru vnesena chyba.

Obdobně se postupuje při vyjádření jednotlivých prvků autokorelační matice, která prostřednictvím hodnot autokorelační funkce popisuje závislosti mezi vzorky  $x(n)$ . Skládání prvků autokorelační matice již bylo probráno v kapitole 4.2 – autokorelační matice svou strukturou odpovídá struktuře Toeplitzovy matice<sup>21</sup>:

$$\begin{aligned} \mathbf{R}_{MM} &= E\{\mathbf{x}_M(n)\mathbf{x}_M^T(n)\} = \\ &= \begin{pmatrix} R_{xx}(0) & R_{xx}(1) & R_{xx}(2) & \dots & R_{xx}(M-2) & R_{xx}(M-1) \\ R_{xx}(1) & R_{xx}(0) & R_{xx}(1) & \dots & R_{xx}(M-3) & R_{xx}(M-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{xx}(M-2) & R_{xx}(M-3) & R_{xx}(M-4) & \dots & R_{xx}(0) & R_{xx}(1) \\ R_{xx}(M-1) & R_{xx}(M-2) & R_{xx}(M-3) & \dots & R_{xx}(1) & R_{xx}(0) \end{pmatrix}. \end{aligned} \quad (5.9)$$

Při předpokladu ergodicity náhodného procesu, který generuje časovou řadu  $x(n)$ , lze odhady hodnot autokorelační funkce  $\hat{R}_{xx}(k)$  získat z časových průměrů jediné realizace procesu:

$$\hat{R}_{xx}(k) = \frac{1}{N} \sum_{i=0}^{N-k-1} x(n-i)x(n-k-i). \quad (5.10)$$

Existují jistě i aplikace, při kterých jsou dostupné i teoretické statistické momenty náhodného procesu, který generuje časovou řadu  $x(n)$ . V takových případech je samozřejmě lepší nepracovat s výběrovými momenty, ale dosadit do korelací v matici (5.9) hodnoty teoretické autokorelační funkce.

Pro funkci  $\mathcal{A}(\mathbf{w}_M)$  popsanou rovnicí (5.6) byly výše vyjádřeny všechny její komponenty a na obr. 5.2 je průběh této funkce znázorněn konkrétně pro predikční filtr se dvěma vahami  $M=2$ . Délka kolmice vztyčené z bodu  $(w_1, w_2)$  k povrchu, který reprezentuje průběh kritériální funkce, odpovídá střednímu výkonu chybové posloupnosti  $E\{\nu(n)^2\}$ . Selským úsudkem, ale i podle rovnice (5.6) je tento výkon shodný se středním výkonem výstupní časové řady  $d(n)$ . Tento výkon bude dosahovat nejnižších hodnot při optimálním nastavení obou vah  $w_1=w_1^*$  a  $w_2=w_2^*$ .

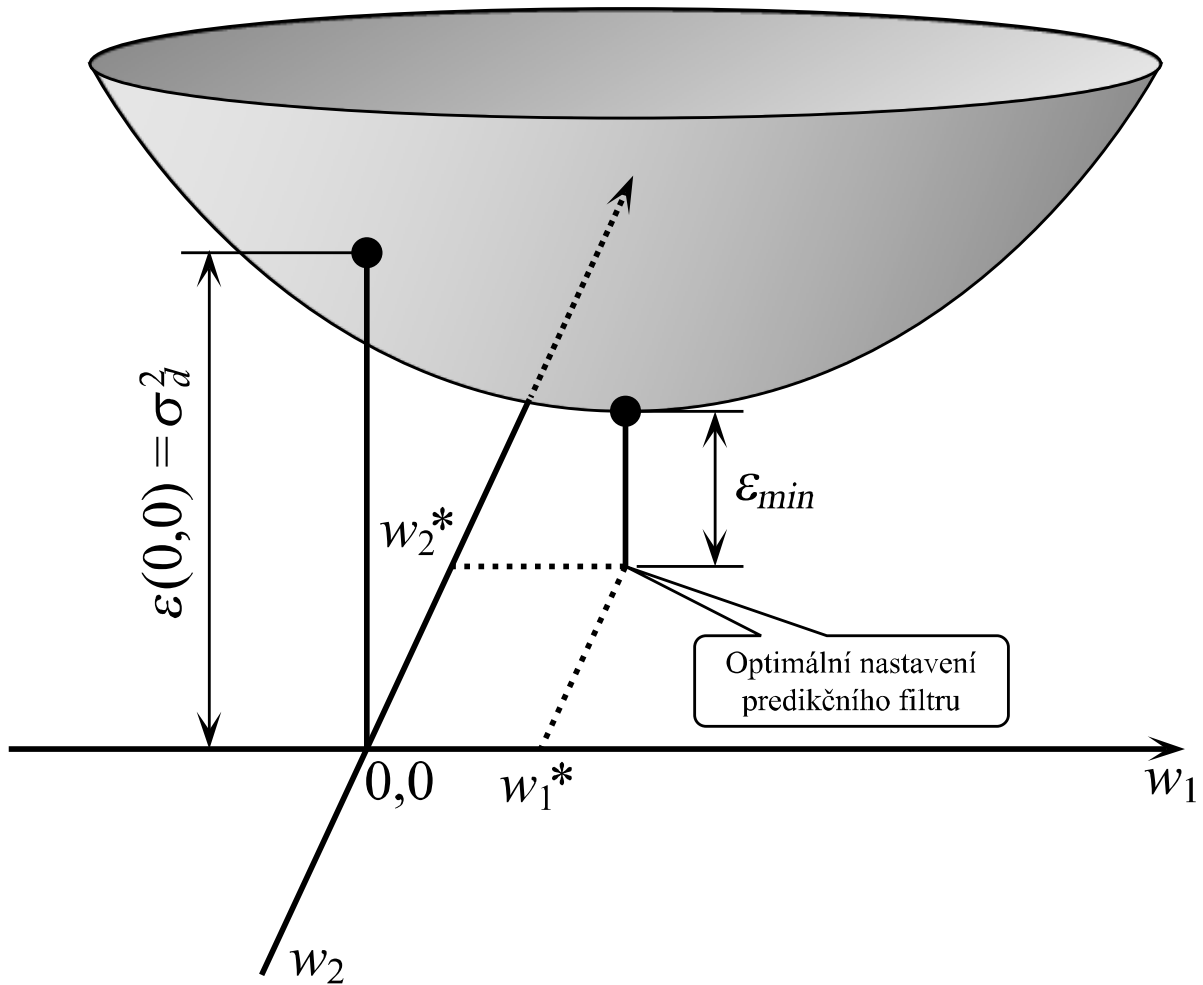
<sup>20</sup> Podmínka ergodicity je striktnější než podmínka stacionarity, viz kap. 4.1. Při identifikaci neznámého systému je možné splnění podmínky ergodicity ověřit jedinež pokusem.

<sup>21</sup> Toeplitzova matice má strukturu, ve které jsou zaručeny konstantní prvky na všech sestupných diagonálách.

Střední kvadratická chyba je funkce s jediným extrémem, pro jehož lokalizaci se vyjádří její první derivace a ta se položí rovna nule. Druhá derivace této funkce musí být v minimu kladná.

Pro  $M=2$  má střední kvadratická chyba predikce tvar:

$$\begin{aligned}\varepsilon(\mathbf{w}_M) &= \varepsilon(w_1, w_2) = \\ &= w_1^2 \phi_x(0) + 2w_1 w_2 \phi_x(1) + w_2^2 \phi_x(0) - 2w_1 \phi_{xd}(0) - 2w_2 \phi_{xd}(1) + \sigma_d^2\end{aligned}\quad (5.11)$$



**Obr. 5.2** Průběh kritériální funkce  $\varepsilon(w_1, w_2)$  pro optimalizaci vah predikčního filtru FIR 2. řádu při identifikaci neznámého systému. Tvar funkce  $\varepsilon(w_1, w_2)$  ovlivňují statistické momenty uvedené v rovnici (5.6), tj. rozptyl časové řady na výstupu systému  $\sigma_d^2$ , vektor vzájemných korelací mezi časovými řadami na vstupu a na výstupu systému a dále autokorelační matice náhodného procesu, který generuje časovou řadu přiváděnou na vstup identifikovaného systému. Body  $w_1^*$  a  $w_2^*$  označují optimální váhy odpovídající minimu kritériální funkce  $\varepsilon_{min}$ , tj. minimu střední kvadratické chyby predikce [1].

a lokalizace jejího extrému dopadne následovně:

$$\begin{aligned} \left[ \frac{\partial}{\partial w_1} \varepsilon(\mathbf{w}_M) \right]_{\mathbf{w}_M = \mathbf{w}_M^*} &= 0, & 2w_1^* \phi_x(0) + 2w_2^* \phi_x(1) - 2\phi_{xd}(0) &= 0, \\ \left[ \frac{\partial}{\partial w_2} \varepsilon(\mathbf{w}_M) \right]_{\mathbf{w}_M = \mathbf{w}_M^*} &= 0. & 2w_2^* \phi_x(0) + 2w_1^* \phi_x(1) - 2\phi_{xd}(1) &= 0. \end{aligned} \quad (5.12)$$

Výslednou soustavu rovnic je možno zobecnit pro libovolné  $M$ :

$$\mathbf{R}_{MM} \mathbf{w}_M^* = \mathbf{p}_M, \quad (5.13)$$

což jsou rovnice pro optimální filtraci založenou na minimalizaci střední kvadratické chyby – jsou známy také pod názvem **normální rovnice**<sup>22</sup>. Jak je vidět, pro jejich sestavení jsou vyžadovány momenty druhého řádu, jejichž znalost může být pro daný náhodný proces buď zaručena, nebo musí být odhadnuty z dostupných dat. Analytické řešení normálních rovnic představuje nalezení inverzní matice k autokorelační matici  $\mathbf{R}_{MM}$ :

$$\mathbf{w}_M^* = \mathbf{R}_{MM}^{-1} \mathbf{p}_M. \quad (5.14)$$

Pro úplnost zbývá ještě ověřit, že druhá derivace je v nalezeném extrému kladná. Tato podmínka pro minimum funkce evidentně platí, neboť druhým derivacím odpovídají diagonální prvky autokorelační matice, což jsou rozptyly nabývající vždy kladných hodnot:

$$\begin{aligned} \left[ \frac{\partial^2}{\partial w_1^2} \varepsilon(\mathbf{w}_N) \right]_{\mathbf{w}_N = \mathbf{w}_N^*} &= 2R_{xx}(0) = 2\sigma_x^2 > 0, \\ \left[ \frac{\partial^2}{\partial w_2^2} \varepsilon(\mathbf{w}_N) \right]_{\mathbf{w}_N = \mathbf{w}_N^*} &= 2R_{xx}(0) = 2\sigma_x^2 > 0. \end{aligned} \quad (5.15)$$

V rovnicích (5.5)–(5.15) byl detailně prozkoumán vztah mezi střední kvadratickou chybou predikce a váhami predikčního filtru. Výsledkem tohoto snažení je optimální řešení lineárního, časově invariantního FIR filtru, při jehož použití pro identifikaci neznámého systému je produkována chybová posloupnost  $\nu(n)$  s nejnižším středním výkonem.

Při dosazení optimálních vah predikčního filtru do rovnice (5.6):

$$\varepsilon_{\min} = \varepsilon(\mathbf{w}_N^*) = \sigma_d^2 - 2\mathbf{w}_N^{*T} \mathbf{p}_N + \mathbf{w}_N^{*T} \mathbf{R}_{NN} \mathbf{w}_N^* \quad (5.16)$$

---

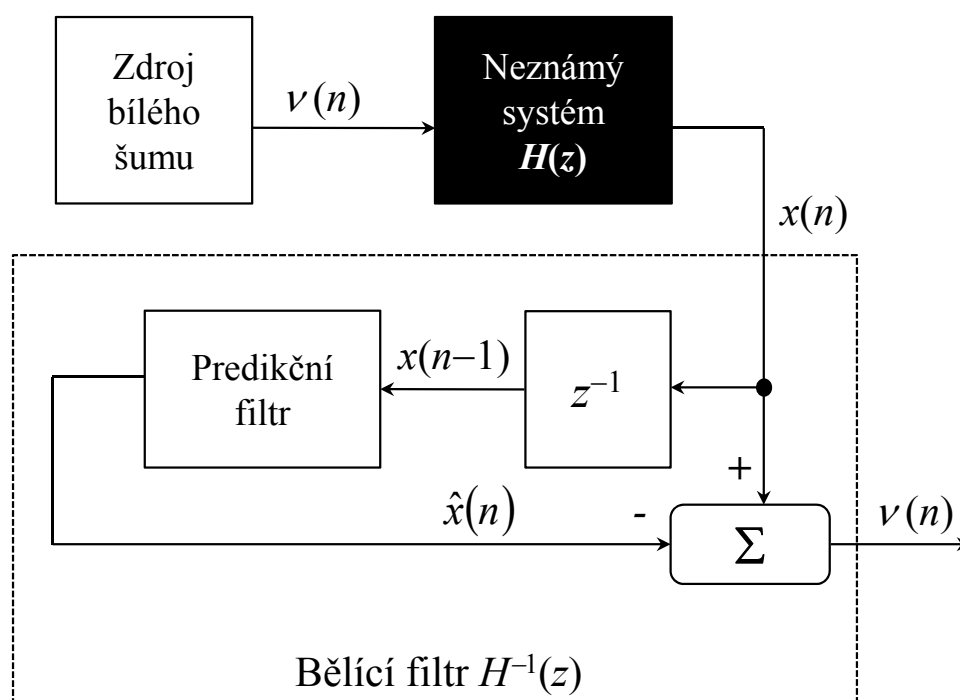
<sup>22</sup> Normální rovnice jsou známy také pod označením Yuleho-Walkerovy rovnice, viz kap. 4.



Lze dojít k zajímavé interpretaci: optimální predikční filtr zřejmě odstraňuje ty složky signálu  $d(n)$ , které korelují se signálem  $x(n)$ , čímž se snižuje výkon chybové posloupnosti. Tuto interpretaci je možno ověřit na dvou následujících extrémních teoretických případech [1].

1. Pokud bude identifikovaný systém pouze zpožďovat vzorky vstupní časové řady a zesilovat je násobením konstantou  $a$ :  $d(n)=ax(n-k_0)$ , budou posloupnosti  $x(n)$  a  $d(n)$  vykazovat pro zpoždění  $k_0$  maximální korelaci. Pokud bude mít vstupní časová řada vlastnosti bílého šumu, bude mít její autokorelační matice podobu diagonální matice, kde prvky na diagonále budou odpovídat rozptylům  $x(n)$ . Vzájemné korelace přejdou na autokorelace. Autokorelační funkce bude mít ovšem pro bílý šum tvar jednotkového impulsu – bude všude nulová kromě argumentu  $k_0$  odpovídajícímu zpoždění mezi vstupní a výstupní posloupností. Pravá strana rovnice (5.16) tak přejde na rozdíl dvou rozptylů, který bude nulový. Kriteriační funkce bude mít minimum v bodě  $\mathbf{w}_n^*=[0, 0, \dots, a, 0, 0, \dots]^T$ . Na obr. 5.2 si lze toto minimum představit v nulové výšce  $\varepsilon_{min}=0$ . Bude produkována nulová chybová posloupnost  $v(n)$ .
2. Opačným teoretickým extrémem je situace, kdy časové řady  $x(n)$  a  $d(n)$  nevykazují vůbec žádnou korelaci. V takovém případě bude mít autokorelační matice v rovnici (5.16) stejný tvar jako v předchozím případě, avšak vektor vzájemných korelací bude tentokrát všude nulový. Váhy predikčního filtru vyjdou všechny nulové, což odpovídá úvaze, že jakákoli predikce by v tomto případě jen zvýšila výkon chybové posloupnosti. Střední kvadratická chyba bude mít v tomto případě hodnotu odpovídající střednímu výkonu časové řady  $x(n)$ .

Při sériovém zapojení optimálního filtru a neznámého systému podle obr. 5.3 se v případě, že je na výstupu celé soustavy pozorován bílý šum, označuje optimální filtr jako **bělicí filtr**. Střední kvadratická chyba predikce bude v takovém případě shodná s rozptylem bílého šumu, kterým je buzen systém s neznámou přenosovou funkcí  $H(z)$ . Tento způsob modelování se označuje jako **inverzní identifikace** [12], neboť díky rovnoměrnému frekvenčnímu spektru bílého šumu musí být přenosová funkce bělicího filtru  $H^{-1}(z)$  inverzní k přenosové funkci systému nebo procesu, jenž generuje pozorovanou časovou řadu.



**Obr. 5.3** Inverzní identifikace neznámého systému bělicím filtrem.

## 5.2 Adaptivní filtrace a predikce

Identifikace neznámých systémů nebo predikce časových řad s využitím optimální filtrace postavené na fixním predikčním filtru (s konstantními parametry) je podmíněna stacionaritou souvisejících náhodných procesů. V případech, kdy nelze tuto podmínku splnit, je třeba najít časové intervaly, ve kterých mohou být související náhodné procesy považovány za stacionární a pro tyto intervaly pak odhadovat autokorelační funkce. Řešením nových soustav normálních rovnic pro každý další časový interval zavedený z důvodu změny podmínek, ve kterých je identifikace prováděna, je možné dospět k časově proměnnému filtru. Takový filtr pak bude v každém vymezeném časovém intervalu optimálním způsobem identifikovat neznámý systém. Toto řešení není ovšem použitelné pro aplikace, při kterých lze jen těžko určovat či dokonce předpovídat budoucí změny v souvisejících náhodných procesech [12].

Výše popsany návrh časově proměnného filtru pro identifikaci nestacionárních systémů a predikci jimi generovaných časových řad nelze považovat za adaptivní filtraci nebo adaptivní predikci, neboť tento návrh je založen na využití apriorních informací o náhodných procesech podílejících se na vzniku časových řad. V případě adaptivních filtrů nejsou žádné apriorní informace vyžadovány. **Adaptivní filtry** je možno definovat jako algoritmy, které přizpůsobují hodnoty svých parametrů pro každý vzorek časové řady a v relativně krátkém čase od změny ve vlastnostech časové řady jsou schopny předpovídat její budoucí hodnoty.

### 5.2.1 Algoritmus LMS

Smyslem adaptivní filtrace v oblasti predikce časových řad je opět nalézt vektor vah predikčního filtru, který se již ovšem nepovažuje za časově invariantní, jako tomu bylo v předchozí kapitole, ale předpokládá se u něj schopnost adaptace na měnící se charakteristiky modelovaných systémů, potažmo z nich generovaných a pozorovaných časových řad. Vektor vah predikčního filtru se proto zapisuje s časovým indexem  $n$ :

$$\mathbf{w}_M(n+1) = \mathbf{w}_M(n) + \alpha(n)\mathbf{S}(n), \quad (5.17)$$

kde  $\alpha(n)\mathbf{S}(n)$  je korekční krok, který je součinem korekčního směrového vektoru  $\mathbf{S}(n)$  a délkou kroku  $\alpha(n)$ . Při použití optimalizační metody nejstrmějšího sestupu je směrový vektor  $\mathbf{S}(n)$  dán záporným gradientem střední kvadratické chyby predikce  $\varepsilon(n)$ . Pro vyjádření derivací se zde tentokrát vychází z chybové posloupnosti  $v(n)$  a nikoli z rovnice (5.6), která je vázána na apriorní znalost momentů druhého řádu, čemuž je snaha se vyhnout:

$$\mathbf{S}(n) = -\nabla_{\mathbf{w}}[\varepsilon(n)] = -\frac{\partial}{\partial \mathbf{w}_M} E\{v^2(n)\} = -2E\left\{v(n)\frac{\partial}{\partial \mathbf{w}_M} v(n)\right\}. \quad (5.18)$$

Po vyjádření chybové posloupnosti jako rozdílu mezi časovými řadami na výstupu identifikovaného systému  $d(n)$  a predikčního filtru  $\hat{d}(n)$ :

$$v(n) = d(n) - \mathbf{w}_M^T(n)\mathbf{x}_M(n), \quad (5.19)$$

je pak možno následujícím způsobem určit parciální derivace  $\nu(n)$  podle jednotlivých vah:

$$\frac{\partial}{\partial \mathbf{w}_M} \nu(n) = -\mathbf{x}_M(n). \quad (5.20)$$

Dosazením (5.20) do (5.18) je nakonec vyjádřen směr korekčního členu  $\mathbf{S}(n)$  pomocí vektoru křížových korelací mezi chybovou posloupností  $\nu(n)$  a posloupností na vstupu  $x(n)$ :

$$\mathbf{S}(n) = -\nabla_{\mathbf{w}}[\varepsilon(n)] = -\frac{\partial}{\partial \mathbf{w}_M} E\{\nu^2(n)\} = 2E\{\nu(n)\mathbf{x}_M(n)\}. \quad (5.21)$$

Výpočet křížové korelační funkce ovšem vede na souborový průměr korelací všech možných realizací časových řad  $\nu(n)$  a  $x(n)$ . Apriorní znalost teoretické křížové korelační funkce mezi těmito dvěma časovými řadami se předpokládat nedá, a musí se tedy odhadnout pomocí časových průměrů, což s sebou opět přináší nutnost zavedení podmínky stacionarity a ergodicity souvisejících náhodných procesů. Řešení normálních rovnic metodou nejstrmějšího sestupu tedy rovněž nevede na návrh adaptivního predikčního filtru.

Místo výpočtu křížové korelace souborovým průměrem nebo místo odhadu časovým průměrem může být křížová korelace v rovnici (5.21) odhadnuta na základě jediného členu  $e(n)x(n)$ :

$$\hat{\mathbf{S}}(n) = -\hat{\nabla}_{\mathbf{w}}[\varepsilon(n)] = 2e(n)\mathbf{x}_M(n). \quad (5.22)$$

I když se jedná o velmi hrubý odhad, není nijak vychýlený a lze očekávat, že chyby budou v po sobě jdoucích krocích na sobě nezávislé, a budou se tedy kompenzovat. Kompenzací náhodných chyb je zajištěno, že se algoritmus bude po určitém počtu iterací velmi těsně přibližovat k optimálnímu nastavení vah predikčního filtru. Takto zjednodušený postup se označuje jako **LMS algoritmus** (*least mean squares*) a je zřejmé, že aproximuje metodu nejstrmějšího sestupu, což lze vidět na obr. 5.4. Jiným, méně obvyklým, avšak výstižným názvem pro tento algoritmus adaptivní filtrace je **filtr se stochasticky gradientní adaptací** [12].

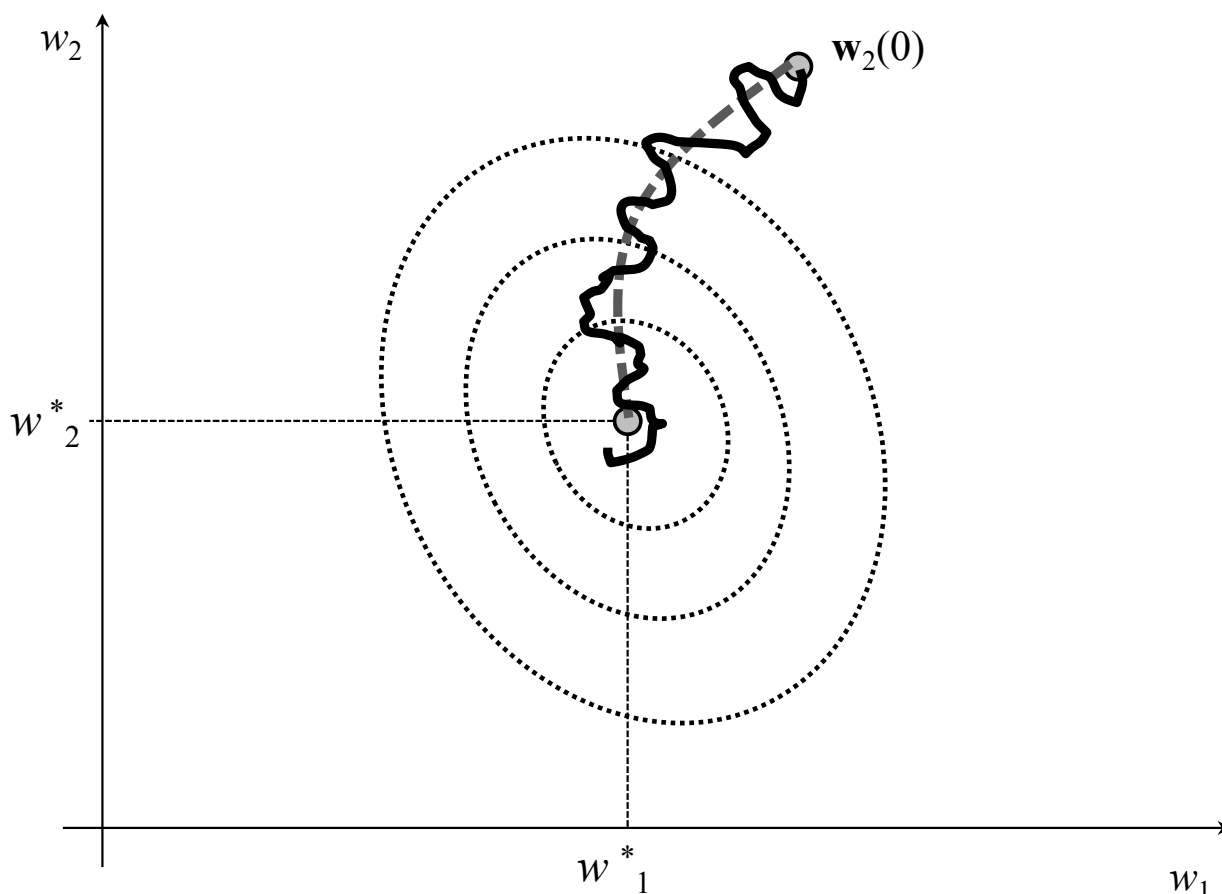
U LMS algoritmu se na rozdíl od jiných optimalizačních postupů nepracuje s proměnnou délkou korekčního kroku, takže lze po dosažení korekčního směru z (5.22) do (5.18) vynechat u  $\alpha(n)$  časový index  $n$ . Zahrne-li se do konstantní délky kroku  $\alpha$  také násobení dvěma, které zbylo v rovnici (5.22) po derivaci, je možné rovnici pro adaptivní filtraci pomocí LMS algoritmu zapsat jako:

$$\mathbf{w}_M(n+1) = \mathbf{w}_M(n) + \alpha \nu(n)\mathbf{x}_M(n) \quad (5.23)$$

Volba délky kroku alfa  $\alpha$  bývá zpravidla experimentální záležitostí – je třeba si ovšem uvědomit, že nastavení délky kroku výrazně ovlivní konvergenční vlastnosti LMS algoritmu. Velké hodnoty  $\alpha$  povedou k rychlé konvergenci ve smyslu menšího počtu nutných iterací, ovšem za cenu zvýraznění nepřesností způsobených odhadem gradientu. To se projeví zejména v blízkosti bodu optimálního nastavení vah  $\mathbf{w}^*$ . Nízké hodnoty  $\alpha$  mohou vést také

k nepřesným výsledkům predikce, a to v případech, kdy algoritmus není schopen rychle adaptovat váhy. V literatuře se doporučuje volit hodnotu  $\alpha$  nepřímo úměrně rozptylu časové řady  $x(n)$  [1].

Nevýhodou návrhu predikčního filtru pomocí LMS algoritmu je jistě větší počet iterací nutných k průchodu optimalizační trajektorií oproti metodě nejstrmějšího sestupu, viz obr. 5.4. Tento postup navíc vede zpravidla pouze k přiblížení se k optimálnímu řešení. Nevýhody jsou však vykoupeny příjemnou jednoduchostí a rychlostí výpočtu.



**Obr. 5.4** Průběh optimalizace vah predikčního filtru s využitím metody nestrmějšiho sestupu (čerchovaná čára – vypočítaný gradient) a pomocí LMS algoritmu (plná čára – odhad gradientu zatížený náhodnou chybou). Jedná se o ilustrační trajektorie dvourozměrné optimalizace predikčního filtru FIR druhého řádu [12].

### 5.2.2 Algoritmus RLS

Minimalizačním kritériem u dosud probíraných optimálních a adaptivních filtrů je střední kvadratická chyba MSE (*mean-square error*), jejíž výpočet je založen na souborových průměrech  $E\{x(n)x(n-k)\}$  a  $E\{d(n)x(n-k)\}$  nebo  $E\{v(n)x(n-k)\}$ . Vzhledem k tomu, že tyto statistické momenty nejsou obvykle k dispozici, je nutno odhadovat je z dostupných dat. Například u filtrů s algoritmem LMS jsou souborové průměry odhadovány pomocí okamžitých hodnot:  $E\{v(n)x(n-k)\} \cong v(n)x(n-k)$ , což do adaptačního postupu vnáší chybu,

která pak zpomaluje průběh konvergence. Alternativou k MSE je tzv. **celková kvadratická chyba** TSE (*total squared error*)<sup>23</sup>, která využívá pouze časových průměrů [16]:

$$\xi(n) = \sum_{i=0}^n v(i)^2 = \sum_{i=0}^n (d(i) - \hat{d}(i))^2 = \sum_{i=0}^n (d(i) - \mathbf{x}_n^T(i) \mathbf{w}_M(n))^2. \quad (5.24)$$

Adaptivní filtry s algoritmem RLS (*recursive least squares*) upravují své koeficienty  $\mathbf{w}_M(n)$  při zpracování každého vzorku časové řady tak, že minimalizují váženou kvadratickou chybu:

$$\begin{aligned} \xi'(n) &= \sum_{i=0}^n \lambda^{n-i} v(i)^2 = \sum_{i=0}^n \lambda^{n-i} (d(i) - \mathbf{x}_M^T(i) \mathbf{w}_M(n))^2, \\ 0 < \lambda &\leq 1, \\ \mathbf{x}(i) &= [x(i), x(i-1), \dots, x(i-M+1)]^T, \\ \mathbf{w}_M(n) &= [w_0(n), w_1(n), \dots, w_{M-1}(n)]^T. \end{aligned} \quad (5.25)$$

Parametr  $\lambda$  zajišťuje exponenciální váhování tak, že novější vzorky časových řad mají ve výpočtech větší vliv než ty starší. Pro  $\lambda=1$  mají všechny vzorky stejnou váhu, což je vhodné nastavení pouze pro predikci stacionárních časových řad. Minimalizace vážené kvadratické chyby vede na soustavu rovnic:

$$\mathbf{R}_{MM}(n) \mathbf{w}_M(n) = \mathbf{p}_M(n), \quad (5.26)$$

kde  $\mathbf{R}_{MM}(n)$  je exponenciálně vážená empirická (nebo výběrová) autokorelační matice pro časové řady  $x(n)$ . Vektor  $\mathbf{p}_M(n)$  obsahuje exponenciálně vážené empirické křížové korelace mezi časovými řadami  $d(n)$  a  $x(n)$ . Soustava rovnic (5.26) se označuje jako **deterministické normální rovnice**. Vzhledem k exponenciálnímu váhování nemá autokorelační matice v (5.26) strukturu Toeplitzovy matice jako tomu bylo v případě normálních rovnic (5.13). Opakovaný výpočet inverzní matice k matici  $\mathbf{R}_{MM}(n)$  je výpočetně náročný a navíc hrozí vzhledem k nezaručené Toeplitzově struktuře, že bude  $\mathbf{R}_{MM}(n)$  pro některé časy singulární. U adaptivních filtrů s algoritmem RLS se proto využívá rekursivní způsob výpočtu:

$$\mathbf{R}_{MM}(n) = \lambda \mathbf{R}_{MM}(n-1) + \mathbf{x}_M(n) \mathbf{x}_M^T(n), \quad (5.27)$$

a to postupným vážením předchozí verze autokorelační matice s přičítáním korelace  $\mathbf{x}(n)d(n)$ . Obdobně se postupuje i pro vektor křížových korelací:

$$\mathbf{p}_M(n) = \lambda \mathbf{p}_M(n-1) + d_M(n) \mathbf{x}_M^T(n), \quad (5.28)$$

<sup>23</sup> Někdy se toto kritérium označuje také jako nejmenší kvadratická chyba LSE (*least squares error*) [9].

Poměrně složitými úpravami, které jsou podrobně rozepsány např. v [9][15][16], pak lze dospět k výsledným rovnicím pro adaptaci koeficientů predikčního filtru:

$$\begin{aligned} \mathbf{w}_M(n) &= \mathbf{w}_M(n-1) + \mathbf{k}(n)(d(n) - \mathbf{x}_M^T(n-1)\mathbf{w}_M(n-1)), \\ \mathbf{k}(n) &= \frac{\mathbf{R}_{MM}^{-1}(n-1)\mathbf{x}_M(n)}{\lambda + \mathbf{x}_M^T(n)\mathbf{R}_{MM}^{-1}(n-1)\mathbf{x}_M(n)}, \end{aligned} \quad (5.29)$$

kde vektor  $\mathbf{k}(n)$  se označuje jako **Kalmanův zisk** nebo **Kalmanova zesílení** (*Kalman gains*)<sup>24</sup>.

## 5.3 Úlohy

**ÚLOHA 5.1** Identifikujte neznámý LTI systém metodou optimální filtrace. Jako model volte predikční filtr FIR s různými řády a porovnejte výsledné impulsní charakteristiky modelu. Pro ilustraci a kontrolu použijte pro identifikovaný systém např. soustavu MA, která provádí aritmetické průměrování z časových oken o čtyřech vzorcích.

Kód skriptu, který v Matlabu realizuje identifikaci LTI systému pomocí optimální filtrace, je na obr. 5.5. Výsledné impulsní charakteristiky modelu pro různé řády  $M$ :

```
M=1:    w=[0,250087]
M=2:    w=[0,251082 0,250141]
M=3:    w=[0,249691 0,249558 0,250642]
M=4:    w=[0,250000 0,250004 0,249951 0,249944]
M=5:    w=[0,250000 0,250000 0,250000 0,250000 0,000001]
M=6:    w=[0,250000 0,250000 0,250000 0,250000 0,000006 0,000007]
...
```

Pominou-li se zaokrouhlovací chyby a nepřesnosti v odhadech autokorelací pomocí časových průměrů, je vidět, že optimální filtr nevyužívá zbytečně složitější strukturu, než je potřebná pro optimální identifikaci neznámého systému.

```
%%% ULOHA 5.1 OPTIMALNI FILTRACE PRO IDENTIFIKACI LTI SYSTEMU
h = ones(1,4) ./4;          % impulsni char-ka neznámého systemu
x = randn(1,100000);        % gaussovsky bily sum - budici sekvence
d = filter(h,1,x);          % kyzeny (desired) vystupni signal
systemu

M = 4;                       % rad predikcniho filtru

% prevraceni poradi tak, aby 1. indexu odpovidal posledni vzorek
% casovych rad x(n) a d(n)
x = flipdim(x,2);
d = flipdim(d,2);

% vypocet autokorelacni matice R
[X,R] = corrmtx(x,M-1,'autocorrelation');
```

<sup>24</sup> Adaptivní filtr s algoritmem RLS je speciální typ tzv. **Kalmanova filtru**, pojmenovaném po maďarském matematikovi.

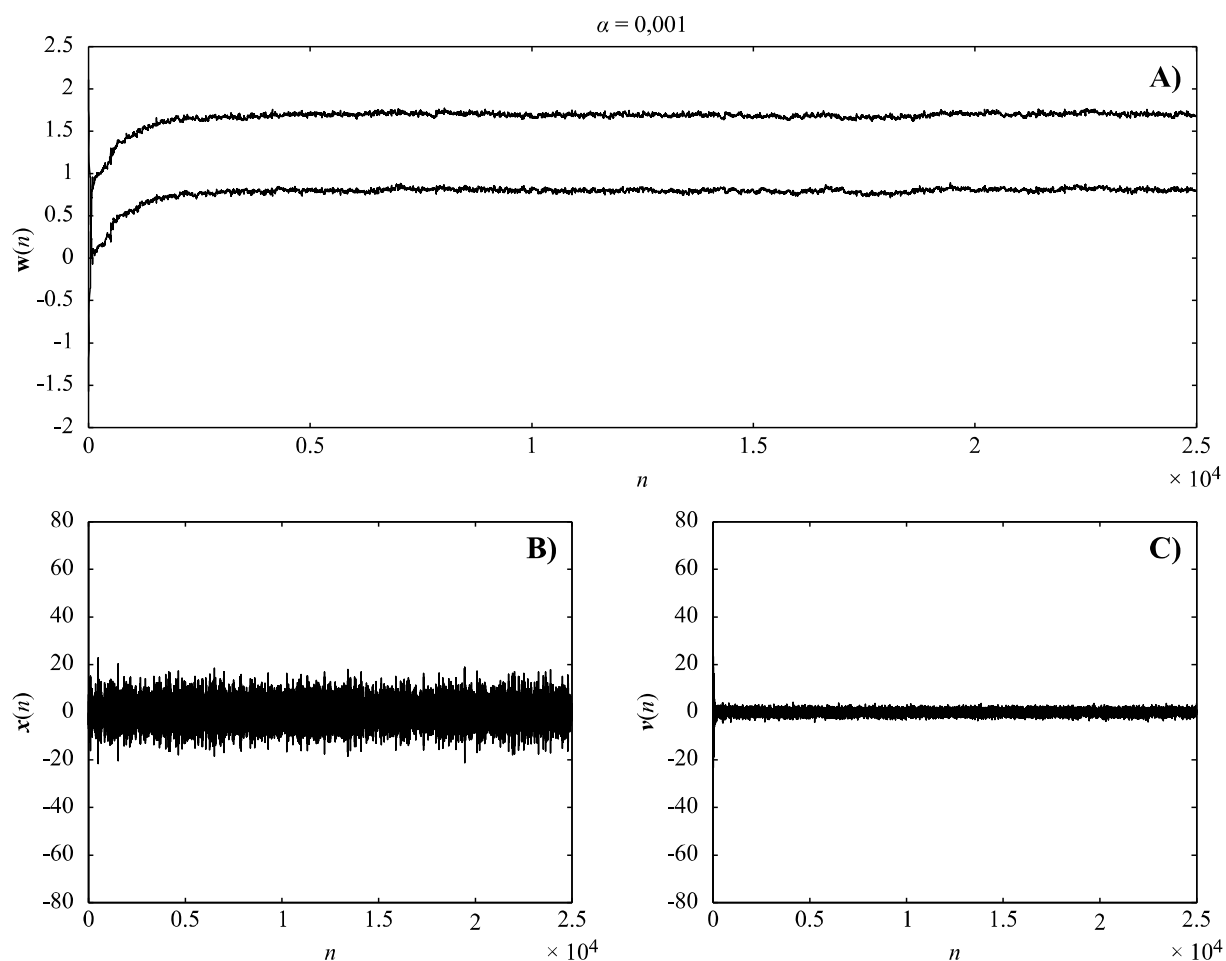
```

p = xcorr(x,d,M-1,'biased'); %vektor krizovych korelaci
delka = length(p);           %delka vektoru p: 2*(N-1)+1
p = p(ceil(delka/2):end)';   %orezani druhe poloviny vektoru.
                              %index ceil(delka/2) odpovida
                              %nulovemu zpozdeni.
                              %index end odpovida zpozdeni M-1

w=R\p;                        %vypocet vah optim. predikcniho filtru
                              %totez jako w=inv(R)*p;

```

**Obr. 5.5** Úloha 5.1 – identifikace „neznámého“ LTI systému metodou optimální filtrace.

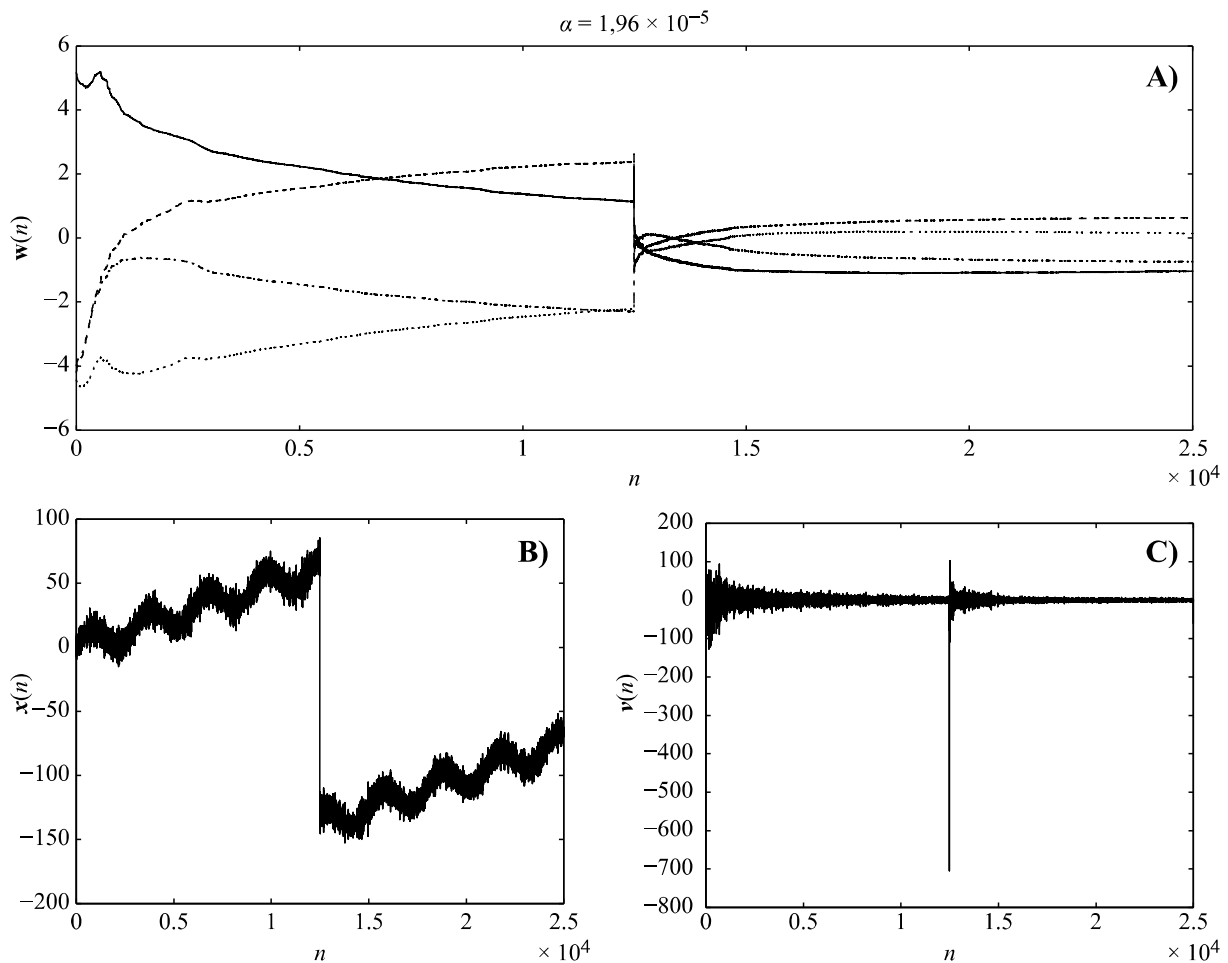


**Obr. 5.6** Úloha 5.2 – Adaptivní predikce časové řady generované stacionárním náhodným procesem AR(2) s koeficienty  $a_1=1,7$  a  $a_2=0,8$ .

A) Průběh konvergence vah predikčního filtru čtvrtého řádu, B) Pozorovaná časová řada  $x(n)$ ,  
C) Chybová sekvence  $v(n)$ .

**ÚLOHA 5.2** Sledujte konvergenci vah a chybu predikce adaptivního filtru s algoritmem LMS při modelování a) stacionárního procesu AR(2) s koeficienty  $a_1=1,7$  a  $a_2=0,8$ ; b) nestacionárního procesu s lineárním trendem a náhlou změnou střední hodnoty, harmonickým kolísáním a rušením generovaným výše popsáním AR(2) procesem.

Kód skriptu, který v Matlabu realizuje adaptivní predikci s algoritmem LMS, je na obr. 5.8. Průběhy konvergence vah při predikci stacionární časové řady lze vidět na obr. 5.6 a konvergence vah při predikci nestacionární časové řady je vidět na obr. 5.7.



**Obr. 5.7** Úloha 5.2 – Adaptivní predikce časové řady generované nestacionárním náhodným procesem s lineárním trendem, náhlou změnou střední hodnoty, harmonickým kolísáním a rušením.  
A) Průběh konvergence vah predikčního filtru čtvrtého řádu, B) Pozorovaná časová řada  $x(n)$ ,  
C) Chybová sekvence  $v(n)$ .

```
%% ULOHA 5.2 ADAPTIVNI PREDIKCE STAC. A NESTACIONARNI CASOVE RADY

N = 25000; % delka pozorovane casove rady
ni = randn(1,N); % excitacni sekvence AR procesu
a = [1 1.7 0.8]; b = 1; % koeficienty AR(2) procesu
x = filter(b,a,ni); % casova rada generovana AR(2)
x=x(:);

trend = 5e-3*(1:N); % linearni trend
trend(N/2:end)=trend(N/2:end)-200;% nahly zlom v trendu
trend = trend(:);
harm = 10*sin((2*pi/3000)*(1:N)); % harmonicke kolisani
harm=harm(:);
x2 = trend+harm+x; % nestacionarni casova rada
x = x2; % bud predikujeme x nebo x2
```



```

alfa = 0.001; % konvergenční konstanta
M = 4; % rad predikčního FIR filtru
w = zeros(M,N); % vahy ve sloupc. vektorech
w(1:M,1:M+1) = randn(M,M+1)*5; % počáteční podmínky náhodně
XP = zeros(N,1); % uchovávání predikovaných hodnot

for n = M+1:N-1
    XP(n) = - w(:,n)'*x(n-1:-1:n-M); % predikce
    e = x(n) - XP(n); % chyba predikce
    w(:,n+1) = w(:,n) - alfa*e*x(n-1:-1:n-M); % uprava vah
end

figure(1); xlabel('n'); ylabel('w'); hold on;
plot(w(1,M+1:end), 'b');
plot(w(2,M+1:end), 'g');
plot(w(3,M+1:end), 'c');
plot(w(4,M+1:end), 'k');
figure(2), plot(x(M+1:end));
figure(3), plot(x(M+1:end)-XP(M+1:end), 'k');

```

**Obr. 5.8** Úloha 5.2 – adaptivní predikce časové řady generované stacionárním a nestacionárním náhodným procesem.

## 5.4 Shrnutí

V poslední kapitole tohoto učebního textu byla na příkladech identifikace neznámých systémů představena oblast optimální a adaptivní filtrace nebo predikce. Čtenář byl seznámen s tím, jak lze postupnými úpravami minimalizačních kritérií odvozených z chyby predikce dospět k tzv. normálním rovnicím. Zatímco pro časové řady generované stacionárními procesy lze normální rovnice řešit běžnými postupy z lineární algebry (tj. výpočtem inverzní matice), pro nestacionární procesy byly ukázány dva rekursivní algoritmy, u nichž jsou koeficienty predikčního filtru adaptovány v každém kroku tak, že postupně konvergují k optimálnímu řešení pro predikci dalších hodnot časové řady. Tyto algoritmy, označované jako LMS (least-mean-squares) a RLS (recursive-least-squares), se liší minimalizačním kritériem. U adaptivních filtrů LMS se využívá střední kvadratická chyba počítaná pomocí souborových průměrů a nepřesnosti, které vznikají při odhadu statistických momentů, mohou zpomalit průběh konvergence koeficientů filtru. Filtry s algoritmem RLS vychází z vážené celkové kvadratické chyby, ve výpočtech používají pouze empirické (výběrové) statistické momenty a konvergují k optimálnímu řešení rychleji než je tomu u LMS.

## Literatura

- [1] Alexander, T.: Adaptive Signal Processing. Springer-Verlag, New York (1986)
- [2] Ambler, Z. et al.: Klinická neurologie. Triton (2008)
- [3] Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods. Springer, New York (2006)
- [4] Delgutte, B.: Biomedical Signal and Image Processing, Massachusetts Institute of Technology [Online] (2007) <http://ocw.mit.edu> (Cit.: 12/2011)
- [5] Devasahayam, S.R.: Signals and Systems in Biomedical Engineering. Kluwer Academic, New York (2000)
- [6] Drongelen, W.: Signal Processing for Neuroscientists: An Introduction to the Analysis of Physiological Signals. Elsevier, London (2008)
- [7] Eshel, G.: The Yule Walker Equations for the AR Coefficients, University of South Carolina, [Online] [http://www.stat.sc.edu/~vesselin/STAT520\\_YW.pdf](http://www.stat.sc.edu/~vesselin/STAT520_YW.pdf) (Cit.: 12/2011)
- [8] Greene W.H.: Econometric Analyses. Prentice Hall, New York (2003)
- [9] Hayes, M.H.: Statistical Digital Signal Processing and Modeling. John Wiley & Sons, New York (1996)
- [10] Holčík, J.: Signály, časové řady a lineární systémy. Akademické nakladatelství CERM, Brno (2012)
- [11] Iravanian, S., Tung, L.: A Novel Algorithm for Cardiac Biosignal Filtering Based on Filtered Residue Method. IEEE Transactions on Biomedical Engineering, 49, 1310–1317 (2002)
- [12] Jan, J.: Číslíková filtrace, analýza a restaurace signálů. VUTIUM, Brno (1997)
- [13] Croarkin, C., Tobias, P., Filliben, J., Hembree, B., Guthrie, W., Trutna, L., Prins, J.: e-Handbook of Statistical Methods. NIST/SEMATECH, [Online] (2010). <http://www.itl.nist.gov/div898/handbook> (Cit.: 10/2011)
- [14] Parks, T.W., Burrus, C. S.: Digital Filter Design. John Wiley & Sons, New York (1987)
- [15] Poularikas, A.D., Ramadan, Z.M.: Adaptive Filtering Primer with MATLAB. Taylor & Francis, New York (2006)
- [16] Rowell, D.: Introduction to Recursive-Least-Squares (RLS) Adaptive Filters. Massachusetts Institute of Technology [Online] (2007) <http://ocw.mit.edu> (Cit.: 12/2011)
- [17] Sayed, A.: Fundamentals of Adaptive Filtering. John Wiley & Sons, New York (2003)
- [18] Skalický, P.: Digitální filtrace a signálové procesory. ČVUT, Praha (1995)
- [19] Tůma, J.: Složité systémy řízení: regulace soustav s náhodnými poruchami. Vysoká škola báňská – Technická univerzita Ostrava, Ostrava (1997)

# Obsah

Předmluva .....	2
1 Systémy, signály a časové řady .....	3
1.1 Vlastnosti systémů .....	4
1.2 Lineární časově invariantní systémy, konvoluce .....	5
1.3 Lineární časově invariantní systémy a periodické signály, Fourierova řada .....	7
1.4 Z-transformace, přenosová funkce diskrétního systému .....	11
1.5 Úlohy .....	14
1.6 Shrnutí .....	16
2 Lineární filtrace .....	17
2.1 Klasifikace filtrů podle algoritmu (AR, MA, ARMA) .....	18
2.2 Klasifikace filtrů podle impulsní charakteristiky (FIR, IIR) .....	20
2.3 Filtry s konečnou impulsní charakteristikou .....	20
2.4 Filtry s nekonečnou impulsní charakteristikou .....	23
2.5 Úlohy .....	24
2.6 Shrnutí .....	31
3 Kumulační zvýrazňování užitečné složky časových řad v šumu .....	32
3.1 Poměr signálu k šumu SNR .....	33
3.2 Repetiční časové řady .....	34
3.3 Repetiční časové řady s náhodným rušením .....	35
3.3 Odhad zbytkové rušivé složky $\pm$ průměrováním .....	37
3.4 Repetiční časové řady s nenáhodným rušením .....	38
3.5 Kumulační zvýrazňování s rovnoměrnými vahami .....	39
3.6 Kumulační zvýrazňování s exponenciálními vahami .....	40
3.7 Úlohy .....	44
3.8 Shrnutí .....	48
4 Náhodné procesy a modely časových řad .....	50
4.1 Vybrané statistické momenty náhodných procesů .....	51
4.2 Dekompozice časových řad a Boxovy-Jenkinsovy modely .....	61
4.3 Exponenciální vyhlazování a predikce .....	72
4.4 Úlohy .....	74
4.5 Shrnutí .....	81
5 Lineární predikce optimálními a adaptivními filtry .....	83
5.1 Optimální filtrace a identifikace .....	83
5.2 Adaptivní filtrace a predikce .....	90
5.3 Úlohy .....	94
5.4 Shrnutí .....	97
Literatura .....	98
Obsah .....	99
Summary .....	100

## Summary

This textbook Linear and Adaptive Processing of Data is primarily intended for students and teachers of the study programme Computational Biology at the Faculty of Science, Masaryk University. This publication has two objectives. First, to explain the basics of theory of discrete linear systems to students or other readers interested in data analysis and processing. The systems are viewed here both as data generators and as means by which data can be analyzed or processed. And second, to demonstrate how selected problems in this area can be effectively solved using computational tools in MATLAB.

The first chapter shows that it is possible to describe a discrete system not only with difference equations, but also with its impulse response, frequency response and transfer function or with a configuration of zeros and poles. Links among various types of system description are also explained. Readers will acquire skills on how to build the system transfer function using the knowledge of difference equations, how to determine the configuration of its zeros and poles in the complex plane and how to determine whether the system is stable or not.

The second chapter introduces the concept of linear filtration with the use of LTI systems. The LTI systems are considered as filters because they selectively choose and pass required frequency components of time series and attenuate the others. Readers are shown how to determine the response of the filter excited with an arbitrary time series using discrete convolution in the time domain or using a discrete Fourier transform in the frequency domain. Furthermore, this chapter explains frequent terminology mistakes in filter classifications FIR, IIR and AR, MA, ARMA.

The third part of this textbook describes averaging methods for time series denoising. Various types of quasi-periodic time series are explained and the conditions for the correct use of the averaging methods to increase signal-to-noise ratio are stated. The influence of particular parameters on dynamic properties of the averaging methods is discussed. The averaging methods, which are very common tools for time series processing, are also shown in the light of their frequency characteristics.

The fourth chapter presents a way of looking at the issue of model construction for time series generated by random processes. The introduction will make the reader familiar with selected statistical moments which are used to describe the properties of random time series. Time averages as well as ensemble averages are discussed together with stationarity and ergodicity, which are properties of subsets of random processes. The core of the chapter consists in the methodology for stationary ARMA models by Box and Jenkins. Three basic steps from model identification through the estimation of numerical values of its parameters to the model validation are described.

The last chapter introduces the area of optimal and adaptive filtering or prediction, using the example of procedures for system identification. Readers will see how to derive the so-called normal equations from the minimization of prediction error. While common linear algebra can be used to solve the equations for stationary time series, two recursive algorithms are shown for non-stationary processes. Prediction filter coefficients are adapted at each time step in such a way that they gradually converge to the optimal solution in terms of minimizing the prediction error.



Lineární a adaptivní zpracování dat  
Ing. Daniel Schwarz, Ph.D.

Recenzenti: prof. Ing. Ivo Provazník, Ph.D., Ing. Zoltán Szabó, Ph.D.  
Jazyková korektura: Ing. Marie Juranová

Obálka: Radim Šustr, DiS.

Vydalo: AKADEMICKÉ NAKLADATELSTVÍ CERM, s.r.o. Brno,  
Purkyňova 95a, 612 00 Brno

[www.cerm.cz](http://www.cerm.cz)

Tisk: FINAL TISK s.r.o. Olomučany

Náklad: 200 ks

Vydání: první

Vyšlo v roce 2012

ISBN 978-80-7204-779-6