Předmět A2M31DSP/Př. 8

PS

Přednáška 8: Redukce šumů - Wienerova filtrace I

Obsah

Model zkreslení signálu

Model zkreslení signálu je obecně nelineární systém s pamětí (dynamický), do kterého vstupuje signál a šum (rušení).
Model určuje typ úlohy i typ metody, která umožňuje redukovat zkreslení signálu popsané daným modelem.
Pokud typ nelineárního zkreslení signálu není znám, je možné použít systém lineární časově invariantní (stacionární) nebo časově proměnný (nestacionární).

Podle typu modelu zkreslení lze definovat dvě základní úlohy:

- zkreslení signálu aditivním šumem
- zkreslení signálu konvolučním šumem

Pro tyto úlohy existují metody redukce šumů

Typy šumů

Základní typy šumů $\rightarrow$ **Modely zkreslení signálů**

- aditivní šum $x[n] = s[n] + u[n]$
- konvoluční šum - výstup LTI soustavy $x[n] = s[n] * h[n]$
- aditivní šum na vstupu LTI soustavy
  $y[n] = x[n] * h[n] = (s + v) * h = h[n] * s[n] + v[n] * h[n]$
- aditivní šum na výstupu LTI soustavy $y[n] = s[n] * h[n] + u[n]$
- šum na vstupu i výstupu LTI soustavy

Předpoklady

- $E[s[n]u[n+k]] = 0$ nebo $E[s[n]v[n+k]] = 0$
- $E[s[n]] = E[u[n]] = E[v[n] = 0$ - tento předpoklad může být v některých úlohách opuštěn

Korelační oblast - aditivní šum

Pro korelaci signálu s aditivním šumem platí[1]

$$R_x[k] = E[x[n]x[n+k]] = R_s[k] + R_u[k],\ k \geq 0$$

- bílý šum
  $R_x[0] = R_s[0] + R_u[0]$
  $R_x[k] = R_s[k]$, pro $k > 0$
- širokopásmový šum
  $R_x[k] = R_s[k] + R_u[k]$, pro $0 \leq k < k_0$
  $R_x[k] = R_s[k]$, pro $k \geq k_0$
- úzkopásmový šum - nízkofrekvenční nebo vysokofrekvenční
  $R_x[k] = E[x[n]x[n+k]] = R_s[k] + R_u[k],\ k \geq 0$

Pozor na nestranný odhad korelace - může „maskovat“složky korelace

[1] Uvažujeme pouze „pravou“ část sudé funkce. Ověření vztahu: do definice korelace dosadíme $x[n] = s[n] + u[n]$, roznásobíme a použijeme předpoklad nekorelovanosti signálu se šumem.

Spektrální oblast - aditivní šum

Pro PSD signálu s aditivním šumem platí[2]

$$S_x(e^{j\Theta}) = S_s(e^{j\Theta}) + S_u(e^{j\Theta}),\ \Theta \in \langle 0, \pi \rangle$$

- bílý šum
  $S_x(e^{j\Theta}) = S_s(e^{j\Theta}) + S_u(e^{j\Theta}),\ \Theta \in \langle 0, \pi \rangle$
- širokopásmový šum - signál zkreslen v pásmu $\Theta_0 < \Theta < \Theta_1$
  $S_x(e^{j\Theta}) = S_s(e^{j\Theta}) + S_u(e^{j\Theta}),\ 0 \leq \Theta_0 < \Theta < \Theta_1 \leq \pi$
- úzkopásmový šum
  limitní případ - sinusovka o kmitočtu Θ_0 a amplitudě A kontaminující signál
  $S_x(e^{j\Theta}) = S_s(e^{j\Theta}) + \frac{A^2}{2}\delta(\Theta - \Theta_0)$

[2] Uvažujeme pouze „pravou“ část sudé funkce. Vztah získáme použitím Wienerovy-Chinčinovy věty s využitím linearity Fourierovy transformace.

Korelační a spektrální oblast - aditivní šum

Předchozí vztahy umožňují podle typu šumu použít různé postupy redukce šumu

- pokud lze signál vzhledem k šumu považovat za úzkopásmový, pak lze použít filtraci pásmovou propustí
- pokud lze signál vzhledem k šumu považovat za širokopásmový, pak lze použít filtraci pásmovou zádrží
- pokud neplatí předchozí, pak lze signál zvýraznit pomocí „očištění korelací nebo spektra“: odhad korelace (spektra) signálu $s[n]$ lze získat např. tak, že odhadneme korelace nebo spektrum šumu a odečteme je od korelace (spektra) signálu se šumem, který máme k dispozici, a poté očištěné korelace (spektra) převedeme zpět na signál[3]

[3] V případě použití spektra se tato metoda nazývá spektrální odečítání

Korelační a spektrální oblast - konvoluční šum

Pro korelaci signálu zkresleného konvolucí platí[4]

$$R_x[k] = E[x[n+k]x[n]] = R_s[k] * R_h[k],\ k \geq 0$$

Pro PSD signálu zkresleného konvolucí platí[5]

$$S_x(e^{j\Theta}) = S_s(e^{j\Theta})|H(e^{j\Theta})|^2,\ \Theta \in \langle 0, \pi \rangle$$

[4] Odvozeno v přednášce 1. Zde uvažujeme pouze „pravou“ část sudé funkce. Při odvození použijeme linearitu konvoluce. .

[5] Uvažujeme pouze „pravou“ část sudé funkce. Vztah získáme opět Fourierovou transformací předchozí rovnice.

Model zkreslení

Metody zvýrazňování signálu jsou určeny MODELEM zkreslení

Modely zkreslení

- nelineární
- LTI - viz úvod přednášky
  - konvoluční šum
  - konvoluce+šum na výstupu
  - šum na vstupu + konvoluce
  - šum na vstupu + konvoluce+šum na výstupu

Příklad pro vybraný model: konvoluce a šum na výstupu
$y[n] = s[n] * h[n] + v[n]$
$S_{ys} = S_{ss}H + S_{vv}$ (viz Př.1)
bez šumu platí: $S_{ys} = S_{ss}H$
pro dekonvoluci je nutné určit $h[n]$ nebo $H(z)$ $\leftrightarrow$ měřením soustavy, popř. ze znalosti fyzikálního modelu

Vybrané metody redukce šumu - zvýrazňování signálu

Základní metody redukce aditivního nebo konvolučního šumu[6] (též zvýrazňování signálu)

- lineární
  - prostá dekonvoluce (inverzní filtrace), problémy: stabilita, zesílení šumu
  - pseudoinverze (nulování „slabých“ spektrálních čar), limitace, problém: subjektivní volba prahu
  - Wienerova filtrace – řeší výše uvedené problémy
  - systémy kombinující adaptivní filtraci a směrový příjem
- nelineární
  - nelinearita bez paměti -mediánový filtr pro redukci odlehlých hodnot pro 1-D i 2-D úlohy
  - nelinearita s pamětí - kombinace LTI filtru s pamětí a mediánového filtru

[6] CMS vynecháno

Metody zvýrazňování signálu - princip

Metody

A. Prostá dekonvoluce inverzním filtrem

$$H_I = \frac{1}{H}$$

Problémy:

- vliv šumu $\rightarrow$ šumová katastrofa
- neminimální fáze $\rightarrow$ problém stability

Metody zvýrazňování signálu - princip

Rozbor inverzní filtrace:

- **bez šumu** $x = h * s$:
  pro determinovaný signál:
  konvoluce: $X = HS \rightarrow$ dekonvoluce: $\hat{S} = \frac{X}{H} = XH_I = S$
  pro náhodný signál:
  $S_{xs} = S_{ss}H \rightarrow$ dekonvoluce: $\hat{S}_{ss} = \frac{S_{xs}}{H}$ - bez chyby (tedy $\hat{S}_{ss} = S_{ss}$), pokud má H minimální fázi[7]
- **se šumem** $x = h * s + u$:
  $S_{xs} = S_{ss}H + S_{uu} \rightarrow$ dekonvoluce: $\hat{S}_{ss} = \frac{S_{xs}}{H} = \frac{S_{ss}H + S_{uu}}{H} \rightarrow$
  $\hat{S}_{ss} = S_{ss} + \frac{S_{uu}}{H} \neq S_{ss}$ – $chyba \frac{S_{uu}}{H}$ závisí na úrovni a typu šumu (tvaru jeho spektra) a velikosti H a může vést k dělení velké hodnoty malou = šumová katastrofa (dojde k zesílení šumu místo jeho redukce)

[7] Minimální fáze H zajišťuje stabilitu H_I.

Metody zvýrazňování signálu - princip

Možnosti odstranění šumové katastrofy

I. Pseudoinverze

$H_I = 1/H$ pro $|1/H| < M$, $H_I = 0$ pro $|1/H| \geq M$

II. limitace

$H_I = 1/H$ pro $|1/H| < M$, $H_I = Me^{j\arg(1/H)}$ pro $|1/H| \geq M$

Nevýhoda: použitý postup je intuitivní a práh M je pevný (nezávisí na vlastnostech signálu)

Lze nalézt automatický postup pro adaptivní nastavování prahu M?

Metody zvýrazňování signálu - vlastnosti

Shrnutí vlastností metod

- prostá dekonvoluce za přítomnosti šumu nepoužitelná
- pseudoinverze odstraní frekvenční složky v okolí kmitočtů s malou hodnotou $|H|$
- limitace neodstraní složky spektra, fáze zachována $\rightarrow$ vhodné pro zpracování obrazů

Wienerova filtrace (WF)

Wienerova filtrace je často používaná metoda zvýraznění signálu (obrazu) v šumu

Typy úloh

- filtrace - kauzální (používá pouze minulé vzorky a aktuální vstup pro výpočet aktuálního výstupu)
- vyhlazování - nekauzální (používá minulé i budoucí vzorky pro výpočet aktuálního výstupu)
- predikce - kauzální (používá pouze minulé vzorky pro předpověď o jeden nebo více budoucích vzorků)
- dekonvoluce - (z minulých a aktuálního vzorku a impulsové odezvy získáme čistý signál)

Typy Wienerova filtru

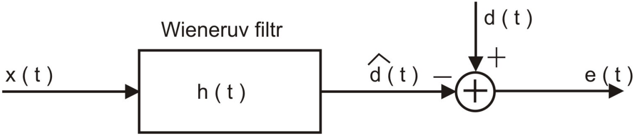
- FIR nebo IIR
- kauzální nebo nekauzální

Wienerova filtrace v časové oblasti

Wienerova filtrace v časové oblasti

Lineární časově invariantní Wienerův filtr[8] WF - pracující v časové oblasti je definován takto

- $x[n]$ je vstup do WF a je dán příslušným modelem zkreslení
- $d[n]$ je požadovaný signál, tedy nezkreslený signál vstupující do modelu zkreslení
- $\hat{d}[n] = h[n] * x[n]$ je výstupem WF a je to nejlepší estimace (odhad) požadovaného signálu $d[n]$
- signál $e[n] = d[n] - \hat{d}[n] = d[n] - h[n] * x[n]$ je chyba estimace



[8] V obrázku jsou signály spojité v čase, ale ve výkladu budeme používat signály diskrétní

Wienerova filtrace v časové oblasti

Cíl a kritérium

Cílem je nalézt takovou impulsovou odezvu filtru, aby hodnota účelové funkce[9] $J = E[e^2[n]]$ byla minimální[10]

Obecně se jedná o složitý problém, ale pro lineární systémy lze řešení nalézt poměrně snadno[11].

Kauzální lineární WF

Wienerův filtr realizujme jako kauzální FIR filtr řádu M, pak impulsová odezva $h[n]$ bude rovna koeficientům filtru $\mathbf{w} = [w_0, w_1, ..., w_M]^T$

Výstup WF pak bude

$\hat{d}[n] = \mathbf{w}^T\mathbf{x}$, kde $\mathbf{x} = [x[n], x[n-1], ..., x[n-M]^T]$

a chyba estimace

$e[n] = d[n] - \hat{d}[n] = d[n] - \mathbf{w}^T\mathbf{x}$

[9] Funkce J je střední kvadratickou hodnotou - MSE (Mean Square Error)

[10] Hledáme tedy MMSE - Minimum Mean Square Error: min $E[e^2[n]]$

[11] Potom ale používáme lineární MMSE, tedy LMMSE odhad (estimátor). LMMSE odhad je roven MMSE odhadu, pokud signály i odhadované parametry mají normální rozložení.

Wienerova filtrace v časové oblasti

Kauzální WF

Cílem je nalézt takovou impulsovou odezvu filtru, aby hodnota účelové funkce $J = E[e^2[n]]$ byla minimální

Výsledné rovnice se nazývají **Wienerovy-Hopfovy rovnice** a pro úlohu filtrace a kauzální FIR lze WF odvodit[12] pomocí kritéria MMSE ve tvaru[13]

$$\mathbf{R}_{xx}\mathbf{w} = \mathbf{r}_{dx},$$

kde $\mathbf{w}$ jsou koeficienty Wienerova filtru, $\mathbf{R}_{xx}$ je korelační matice, $\mathbf{r}_{dx}$ je vektor vzájemných korelací mezi $x[n]$ = vstupem do WF a $d[n]$ = požadovaným signálem

Koeficienty WF získáme inverzí korelační matice

$$\mathbf{w} = \mathbf{R}_{xx}^{-1}\mathbf{r}_{dx}$$

[12] Podobně jako Yuleovy-Walkerovy rovnice v přednášce o lineární predikci, kde chybu predikce nahradíme chybou estimace $e[n] = d[n] - \mathbf{w}^T\mathbf{x}$. Derivaci skalární funkce $E[e^2[n]]$ podle $\mathbf{w}$ položíme rovnu nule a tím získáme normální (Wienerovy-Hopfovy) rovnice.

[13] Toto je zcela obecný vztah pro koeficienty Wienerova FIR filtru.

Wienerova filtrace ve spektrální oblasti

Nekauzální WF

Použitím DTFT[14] na Wienerovy-Hopfovy rovnice[15] získáme

$$S_{dx}(e^{j\Theta}) = H_{WF}(e^{j\Theta})S_{xx}(e^{j\Theta}),$$

kde $S()$ jsou spektrální hustoty, $H_{WF}(e^{j\Theta})$ je frekvenční charakteristika Wienerova filtru daná Fourierovým obrazem koeficientů filtru **w**

$$H_{WF}(e^{j\Theta}) = \frac{S_{dx}(e^{j\Theta})}{S_{xx}(e^{j\Theta})}$$

Toto je **obecný vztah** pro frekvenční charakteristiku nekauzálního Wienerova filtru. Vzájemná PSD S_{dx} je určena modelem zkreslení.

[14] Proč používáme DTFT? Ahmed a Rao ukázali, že použití libovolné ortogonální transformace vede na redukci střední kvadratické chyby mezi požadovaným signálem a výstupem WF. Jako ortogonální transformaci lze použít DFT, DCT, KLT, apod.

[15] Wienerovy-Hopfovy rovnice představují konvoluci mezi korelační funkcí $R_{xx}[k]$ a koeficienty Wienerova filtru w_k

Wienerova filtrace pro aditivní šum

Zkoumejme, jak se jednotlivé modely zkreslení signálu projeví na tvaru vzájemné spektrální hustoty S_{dx} - volbou modelu zkreslení omezíme množinu úloh, pro které lze Wienerův filtr použít, ale získáme předpis pro realizaci WF

A. Použijeme-li **model pro aditivní šum**[16], získáme přenosovou funkci nekauzálního WF ve tvaru

$$H_{WF}(e^{j\Theta}) = \frac{S_{ss}(e^{j\Theta})}{S_{xx}(e^{j\Theta})},$$

Důkaz: položíme-li $x[n] = s[n] + u[n]$ a $d[n] = s[n]$, získáme

$$r_{dx} = E[d[n+k]x[n]] = E[d[n+k](d[n] + u[n])] = r_{dd} = r_{ss}$$

a tedy

$$S_{dx} = \mathcal{F}(r_{dx}) = \mathcal{F}(r_{ss}) = S_{ss}$$

[16] V tomto případě je požadovaným signálem $d[n]$ čistý signál $s[n]$, tedy $d[n] = s[n]$.

Wienerova filtrace pro aditivní šum

Teoretický vztah pro nekauzální WF pro redukci aditivního šumu pro stacionární signál i šum je

$$H_{WF}(e^{j\Theta}) = \frac{S_{ss}(e^{j\Theta})}{S_{xx}(e^{j\Theta})} = \frac{S_{xx}(e^{j\Theta}) - S_{uu}(e^{j\Theta})}{S_{xx}(e^{j\Theta})}$$

Poslední výraz v rovnici získáme použitím modelu pro aditivní šum ve spektrální oblasti[17].

Známe-li PSD čistého signálu $S_{ss}(e^{j\Theta})$ a PSD šumu $S_{uu}(e^{j\Theta})$, lze tento vztah přímo použít, neboť $S_{xx}(e^{j\Theta})$ umíme získat součtem obou výkonových spektrálních hustot. To přichází v úvahu při modelování signálů, kdy známe $S_{ss}(e^{j\Theta})$ a $S_{uu}(e^{j\Theta})$. K daným spektrálním hustotám samozřejmě umíme generovat příslušné realizace $s[n]$ a $u[n]$ a vytvořit směs $x[n]$ pomocí modelu pro aditivní šum směs $x[n] = s[n] + u[n]$.

[17] V tomto případě platí $S_{xx}(e^{j\Theta}) > S_{uu}(e^{j\Theta})$, a proto je čitatel posledního výrazu vždy kladný - to je zajištěno použitím modelu pro aditivní šum.

Implementace Wienerovy filtrace pro aditivní šum

Dosud jsme předpokládali, že spektrální hustoty známe. Nyní se zaměřme na situaci, kdy je potřeba spektrální hustoty odhadovat.

Zpracováváme-li signály (i šumy), ať už stacionární nebo nestacionární, je třeba signál segmentovat a pro každý segment odhadnout parametry Wienerova filtru.

Lze rozlišit tři případy

- stacionární signál i šum - oba signály máme k dispozici - segmentaci používáme pro odhad PSD pomocí Welchovy metody, ale parametry (čitatel i jmenovatel) WF jsou konstantní - v tomto případě lze použít vztah[18]

$$\hat{H}_{WF}(e^{j\Theta}) = \frac{\hat{S}_{ss}(e^{j\Theta})}{\hat{S}_{xx}(e^{j\Theta})} = \frac{\hat{S}_{xx}(e^{j\Theta}) - \hat{S}_{uu}(e^{j\Theta})}{\hat{S}_{xx}(e^{j\Theta})}$$

[18] Pokud by čitatel vlivem chyb odhadu PSD byl záporný, je třeba jej ošetřit - viz další výklad.

Implementace Wienerovy filtrace pro aditivní šum

- nestacionární signál a stacionární šum - jmenovatel odhadneme Welchem (je konstantní), čitatel WF odhadujeme pro každý nový segment
- nestacionární signál a nestacionární šum - čitatel i jmenovatel WF odhadujeme pro každý nový segment

Implementace Wienerovy filtrace pro aditivní šum

Pro oba uvedené případy lze použít jedno z možných řešení: spektrální odečítání[19]

$$\hat{S}_{ss}(e^{j\Theta}) = |\hat{X}(e^{j\Theta})|^2 - \hat{S}_{uu}(e^{j\Theta}),$$

kde odhad PSD šumu $\hat{S}_{uu}$ lze získat Welchovou metodou v pausách, kdy je signál $s[n]$ nepřítomen, zatímco $|\hat{X}(e^{j\Theta})|^2$ je v tomto případě periodogram[20] (výkonové spektrum) segmentu signálu $x[n]$. Obecně se samozřejmě, pro nestacionární signál a šum, mění jak čitatel tak i jmenovatel WF.

[19] Protože uvedený vztah nemůže nabývat záporné hodnoty - PSD je nezáporná funkce - je třeba rozdíl PSD ošetřit. Buď se vezme absolutní hodnota, tedy $\hat{S}_{ss} = |S_{xx} - \hat{S}_{uu}|$ - v tomto případě mluvíme o dvoucestném usměrnění, nebo se záporné hodnoty rozdílu nulují a použijí se pouze hodnoty kladné - v tomto případě mluvíme o jednocestném usměrnění. Spektrální odečítání v uvedeném tvaru je vhodné pro případ, kdy máme nestacionární signál a víceméně stacionární šum.

[20] Není tedy použito průměrování výkonových spekter (periodogramů).

Wienerova filtrace pro aditivní šum

Pro uvedené spektrální odečítání lze odhad PSD šumu získat:

- na začátku signálu $x[n]$, kdy není přítomen čistý signál - to ovšem vyžaduje stacionaritu šumu po celou dobu trvání signálu
- nalezení úseků, kde je přítomen čistý signál a kdy není - tedy v případě zpracování řeči jde o nalezení pauz v řeči[21] - tento způsob připouští nestacionaritu šumu, nicméně pro silně nestacionární šum může tato metoda generovat chyby projevující se různými „hvizdy“.

Alternativně lze pro spektrální odečítání použít i amplitudová spektra, nicméně pro WF potřebujeme odhadnout výkonová spektra - přesněji periodogram signálu a PSD (průměr peridogramů) šumu.

Pozn.: Existují i jiné metody odhadu PSD šumu, které nevyžadují detektor, a tudíž ani to, aby byl čistý signál intermitentní (přerušovaný), ale těmi se nebudeme zabývat.

[21] Algoritmus pro nalezení takových úseků v aplikacích zpracování řeči se nazývá detektor řečové aktivity - VAD (Voice Activity Detector)

Implementace Wienerovy filtrace pro aditivní šum

Odhadneme-li frekvenční charakteristiku nekauzálního WF pro redukci aditivního šumu

$$\hat{H}_{WF}(e^{j\Theta}) = \frac{\hat{S}_{ss}(e^{j\Theta})}{\hat{S}_{xx}(e^{j\Theta})}$$

některým z výše uvedených způsobů[22],
lze spektrum[23] $\hat{S}$ zvýrazněného signálu $\hat{s}[n]$ získat

$$\hat{S}(e^{j\Theta}) = \hat{X}(e^{j\Theta})\hat{H}_{WF}(e^{j\Theta})$$

Zvýrazněný signál[24] je pak

$$\hat{s}[n] = IDTFT\{\hat{S}(e^{j\Theta})\}$$

[22] Lze samozřejmě, kromě DFT, které bylo v předchozím výkladu použito, použít též DCT.

[23] Komplexní spektrum

[24] Pro segmentované signály se rekonstrukce signálu po použití inverzní Fourierovy transformace provádí typicky metodou OLA - sčítání segmentů s přesahem.