Předmět B2M31DSP/Př. 12

PS

Přednáška 12: Kvadraturní banky filtrů a diskrétní vlnková transformace

Obsah

Úvod

Kurs CZS zavedl základy spojité vlnkové transformace - CWT a základní vlnky, též zmínil diskrétní vlnkovou transformaci a její souvislost s bankou filtrů

Tento kurs prohloubí tyto poznatky

A. CWT

- používá mateřskou vlnku, ze které se odvodí celá rodina vlnek pomocí změny měřítka (dilatací) vlnky a jejího posunu
- CWT[1] pak lze chápat jako prostředek pro výpočet vzájemné podobnosti (energie) mezi rodinou vlnek a signálem
- používá se např. pro časově frekvenční rozklad signálu - umožňuje větší flexibilitu než klasická STFT (krátkodobá Fourierova transformace), neboť v důsledku změny měřítka vlnek pracuje s různým časově-frekvenčním rozlišením, zatímco STFT má okno pevné délky[2]

[1] např. `https://cs.wikipedia.org/wiki/Vlnková_transformace`

[2] Existují ovšem též algoritmy STFT pracující s proměnnou délkou okna, kterou nastavují adaptivně podle vlastností signálu

Úvod

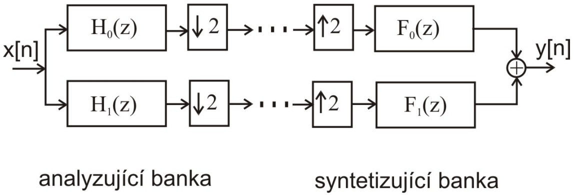
B. DWT

- DWT se klasicky definuje jako CWT, která má posuny a dilatace kvantovány v mocnině dvou
- používá, podobně jako DFT bázi, do které rozkládá signál
- podobně jako DFT lze realizovat bankou filtrů, ale na rozdíl od DFT je struktura banky jiná - je tvořena kaskádou dvoupásmových bank filtrů řazených za sebou a snížením vzorkovacího kmitočtu (decimací) - toto uspořádání nazýváme rozkladem (dekompozicí) signálu - postupně získáváme koeficienty na hladinách (scales)
- pro realizaci zpětné DWT (rekonstrukce signálu) je opět použita kaskáda dvoupásmových bank filtrů, před kterými dochází ke zvýšení vzorkovacího kmitočtu (interpolace)
- nejčastějšími aplikacemi DWT je ztrátová komprese signálů a redukce šumů - výhodou je, že existuje velké množství různých bází (tedy i vlnek), které mohou dobře vystihnout charakter signálu - tím se DWT přibližuje svou efektivitou metodě KLT, přičemž bázi není nutné opakovaně počítat jako u KLT

Dvoupásmová banky filtrů - opakování

Banky filtrů: skupina číslicových filtrů vyhovující jistým podmínkám = podmínky perfektní rekonstrukce

Dvoupásmová banka filtrů



Podmínky perfektní rekonstrukce

- **aproximace** = výstup dolní propusti $H_0(z)$
  detaily = výstup horní propusti $H_1(z)$
- **analyzující banka filtrů** = rozklad signálu do pásem a decimace
  syntetizující banka filtrů = interpolace a sloučení signálu
- **podmínky perfektní rekonstrukce**:
  - žádné překrývání: $H_0(-z)F_0(z) + H_1(-z)F_1(z) = 0$
  - žádné zkreslení signálu: $H_0(z)F_0(z) + H_1(z)F_1(z) = 2z^{-D}$

Potom $\rightarrow y[n] = x[n-D]$, D = zpoždění, x = vstup, $y = \hat{x}$ = výstup

Důsledek: **filtry jsou spolu svázány** a nemohou být libovolné

Platí $F_0(z) = H_1(-z)$ a $F_1(z) = -H_0(-z)$

Maticový zápis podmínek perfektní rekonstrukce

$$\frac{1}{2}\begin{bmatrix} F_0(z) & F_1(z) \end{bmatrix}\begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} = \begin{bmatrix} z^{-D} & 0 \end{bmatrix}$$

Matice **H** se nazývá modulační matice

Návrh dvoupásmové banky filtrů

Návrh filtrů vychází z prototypu, který musí mít jisté vlastnosti

1. návrh prototypu $H(z)$
2. výpočet filtrů banky vyhovující podmínkám perfektní rekonstrukce
   - $H_0(z) = H(z)$
   - $H_1(z) = H(-z) = H_0(-z)$
   - $F_0(z) = 2H(z) = 2H_0(z)$ - násobení 2 kvůli decimaci
   - $F_1(z) = -2H(-z) = -2H_0(-z)$

Z těchto podmínek lze pro prototyp $H(z)$ získat

$$H^2(z) - H^2(-z) = z^{-D}$$

z čehož plyne, že jeho frekvenční charakteristika musí být výkonově komplementární

$$|H(e^{j\Theta})|^2 + |H(e^{j(\Theta-\pi)})|^2 = 1,\ \Theta \in < 0, \pi >$$

První člen rovnice představuje dolní propust a druhý člen horní propust

Důsledky podmínek perfektní rekonstrukce

- **dvoupásmová banka filtrů** používá filtry, které dělí frekvenční pásmo vždy na dvě poloviny - vždy dolní propust a horní propust - tyto filtry se nazývají **zrcadlové filtry**
- podmínky perfektní rekonstrukce vyžadují, aby filtry (nejen prototyp) byly výkonově komplementární - takové filtry nazýváme **kvadraturními zrcadlovými filtry** - tuto podmínku ovšem nesplňují symetrické FIR filtry (tedy FIR s lineární fází) - nicméně, stále existuje možnost realizovat celou banku filtrů ze symetrických FIR filtrů i za cenu, že perfektní rekonstrukce nebude zcela dodržena - existují vhodné algoritmy návrhu
- budou-li filtry dekompoziční (analyzující) typu FIR, lze ukázat, že rekonstrukční (syntetizující) filtry budou typu IIR-v praxi nevýhodné
- je-li třeba celou banku filtrů (tedy dekompoziční i rekonstrukční část) realizovat pomocí FIR filtrů a dodržet perfektní rekonstrukci, je třeba opustit požadavek linearity fáze - podmínky perfektní rekonstrukce se potom lehce změní - takto se též realizuje ortogonální DWT

Kvadraturní zrcadlové filtry - vysvětlení

Podmínky perfektní rekonstrukce vedou na dříve uvedenou podmínku komplementarity kvadrátu modulu frekvenčních charakteristik (komplementarita výkonu)

$$|H(e^{j\Theta})|^2 + |H(e^{j(\Theta-\pi)})|^2 = 1,\ \Theta \in < 0, \pi >,$$

která znamená, že kvadrát modulu frekvenční charakteristiky dolní propusti

$$H_0(e^{j\Theta}) = H(e^{j\Theta})$$

a horní propusti

$$H_1(e^{j\Theta}) = H(e^{j(\Theta-\pi)})$$

ve dvoupásmové bance filtrů je zrcadlově symetrický podle poloviny vzorkovací frekvence

$\Rightarrow$

název: kvadraturní zrcadlové filtry (QMF - Quadrature Mirror Filters)

Návrh banky filtrů se symetrickými FIR filtry

Návrh **prototypu** $H(z)$ **v případě symetrickýh FIR filtrů** nemá analytické řešení a je nutné použít numerické optimalizační metody, které vedou na **přibližné splnění podmínek perfektní rekonstrukce**

- **účelová funkce (kritérium)** - hledáme její minimum

$$E_r + \alpha E_s,$$

  kde α je váhovací faktor a

- **podmínka komplementarity výkonu** má tvar

$$E_r = \int_{\Theta=0}^{\pi} (|H(e^{j\Theta})|^2 + |H(e^{j(\Theta-\pi)})|^2 - 1) d\Theta$$

- **podmínka maximálního potlačení v nepropustném pásmu** je

$$E_s = \int_{\Theta=\Theta_s}^{\pi} |H(e^{j\Theta})|^2 d\Theta,$$

kde $\Theta_s = (\frac{1}{4} + \Delta)2\pi$, $\Delta \to 0$

Návrh banky filtrů se symetrickými FIR filtry

Závěr: hledáme tedy takové koeficienty **symetrického FIR filtru typu dolní propust** = prototypu

- jehož frekveční charakteristika má co největší útlum v nepropustném pásmu
- a součet kvadrátů modulů dolní a horní propusti se co nejméně liší od jedničky - hledáme tedy co nejlepší aproximaci QMF filtru

Pozn.: byť se nejedná o perfektní rekonstrukci, stále pro praktické aplikace tento postup vyhovuje[3], neboť chyba se zmenšuje s rostoucím řádem filtru

[3] Tato úloha bude námětem cvičení

Přííklad návrhu prototypu

Příklad aproximace QMF prototypu $H(z)$ pro parametry:

- $\alpha = 0.1$
- normovaný konec propustného pásma $w = 2f/f_s = 0.3$
- délka filtru = počet koeficientů - musí být sudé číslo $N = 20$

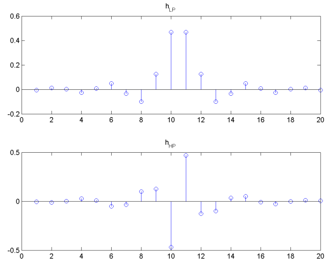
$h[n] =$

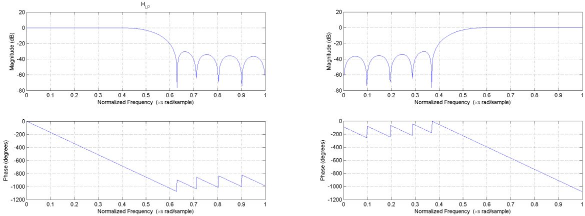
$-0.0059, 0.0129, 0.0013, -0.0274, 0.0086, 0.0510, -0.0338, -0.1001, 0.1243, 0.4688,$

$0.4688, 0.1243, -0.1001, -0.0338, 0.0510, 0.0086, -0.0274, 0.0013, 0.0129, -0.0059$

Pozn.: impulsová odezva je symetrická, proto má filtr nuly vně i uvnitř jednotkové kružnice - fáze je lineární

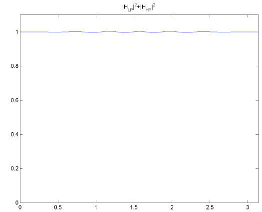
Aproximace kvadraturního zrcadlového filtru symtrickým FIR filtrem-příklad

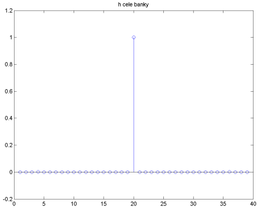


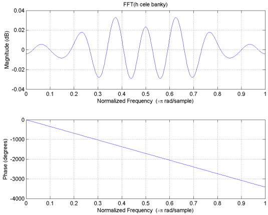


Nahoře: imp. odezvy QMF DP a HP; dole: frekvenční charakteristika QMF DP a HP

Aproximace kvadraturního zrcadlového filtru symtrickým FIR filtrem-příklad







Nahoře: součet kvadrátů modulů DP a HP; dole: impulsni odezva celé banky filtrů a její frekveční charakteristika

Je dobře patrné zvlnění v součtu kvadrátů modulů, tedy nepslnění komplementarity výkonu při ověření celé banky bylo zvlnění jejího přenosu cca 0.4 dB (viz obr vpravo) pro filtr řádu 10 a cca 0,2 dB pro filtr řádu 20

Perfektní rekonstrukce - ortogonální filtry

Má-li mít banka filtrů perfektní rekonstrukci při použití ortogonálních FIR filtrů, je třeba opustit předpoklad linearity fáze. Banku filtrů s perfektní rekonstrukcí s ortogonálními filtry lze např. získat následujícím způsobem:

1. návrh kauzálního prototypu $H(z)$ s nulami uvnitř jednotkové kružnice a tím i nelineární fází - lze jej získat ze symetrického QMF filtru, který má fázi lineární
2. výpočet filtrů banky vyhovující podmínkám perfektní rekonstrukce
   - $H_0(z) = H(z)$ - H_0=dolní propust s nelineární fází a s nulami uvnitř 1 kružnice
   - $H_1(z) = -z^{-N}H(-z^{-1})$ =horní propust s otočenou imp. odezvou oproti $h_0[n]$
   - $F_0(z) = 2z^{-N}H(z^{-1})$ =dolní propust s otočenou impuls. odezvou oproti $h_0[n]$
   - $F_1(z) = 2H(-z)$

Uvedené vztahy využívá též ortogonální DWT, tedy DWT, která využívá pro dekompozici a rekonstrukci tutéž vlnku. Uvedené typy filtrů se nazývají konjugované kvadraturní filtry.

Pozn.: člen z^{-N} zajišťuje kauzalitu filtru, argument z^{-1} představuje otočení impulsové odezvy filtru a člen $(-z)$ transformuje dolní propust na horní propust.

Diskrétní vlnková transformace a vlnky

Při použití DWT nás spíše než tvar vlnky zajímají parametry dekompozičních a rekonstrukčních filtrů příslušné banky filtrů.
Ty nemohou být libovolné, ale musí zajišťovat perfektní rekonstrukci a zároveň generovat funkci, která má vlastnosti vlnky[4]:

- splňuje **podmínku přípustnosti** (admissibility condition):
  $$C_\Psi = \int_0^\infty \frac{|\hat{\Psi}(\omega)|^2}{\omega} d\omega < \infty,$$
  kde $\hat{\Psi}(\omega)$ je Fourierova transformace vlnky. Tato vlastnost zaručuje invertibilitu transformace.
- a **konečné energie**:
  $$\int_{-\infty}^{\infty} |\Psi(t)|^2 dt < \infty$$
- z podmínky přípustnosti plyne **nulová střední hodnota**: $\int_{-\infty}^{\infty} \Psi(t) dt = 0$

[4] Slovně: konečná energie a nulová střední hodnota, hodnoty vlnky rychle ubývají k nule nebo jsou nulová vně konečného intervalu - prostě typický přechodový děj.

Diskrétní vlnková transformace - DWT

Jak bylo uvedeno, spíše než vlnka nás zajímají parametry fitrů v bance
Základní vztahy DWT:

- základem je **měřítková rovnice**[5]

$$\phi(t) = \sqrt{2} \sum_n h_0[n] \phi(2t - n),$$

  kde $\phi(t)$ je měřítková funkce, h_0 jsou **koeficienty rekonstrukční dolní propusti**[6]
- **vlnková rovnice** generuje vlnky pomocí konvoluce měřítkové funkce s koeficienty **rekonstrukční horní propusti**

$$\psi(t) = \sqrt{2} \sum_n h_1[n] \phi(2t - n),$$

[5] Ne všechny vlnky mají měřítkovou funkci, např. mexický klobouk, Morletova vlnka, Gaborova vlnka, apod. Většinou se jedná o vlnky, které lze vyjjádřit analyticky vztahem.

[6] Pro dané koeficienty lze měřítkovou rovnici použít iterativně a získat měřítkovou funkci $\phi(t)$. Poté lze tuto funkci dosadit do vlnkové rovnice a tím získat příslušnou vlnku $\psi(t)$. Nicméně tu, jak ukážeme, k výpočtu stejně nepoužíváme.

Aproximace a detaily

- syntéza signálu pomocí **diskrétní vlnkové transformace** je dána

$$f(t) = \sum_k c_{j0}(k) 2^{j0/2} \phi(2^{j0} t - k) + \sum_k \sum_j d_j(k) 2^{j/2} \psi(2^j t - k),$$

kde j je index hladiny rozkladu a $j0$ je konstanta (např. 0, nebo ∞)

- koeficienty rozkladu jsou pro ortogonální DWT dány skalárním součinem signály s měřítkovými funkcemi či vlnkami

aproximace:

$$c_j(k) =< f(t), \phi_{j,k}(t) >= \int f(t)\, 2^{j/2} \phi(2^j t - k) dt$$

detaily:

$$d_j(k) =< f(t), \psi_{j,k}(t) >= \int f(t)\, 2^{j/2} \psi(2^{j/2} t - k) dt$$

Vlnková transformace

Rekurentní výpočet koeficientů vlnkové tranformace

Použitím měřítkové rovnice, vlnkové rovnice, vztahů pro výpočet koeficientů $c_j(k)$ a $d_j(k)$ a vztahu pro syntézu signálu pomocí DWT lze odvodit vztahy pro koeficienty na různých hladinách rozkladu

A. Vztahy pro dekompozici (analýzu)

$$c_j(k) = \sum_m h_0[m-2k]c_{j+1}(m) = \sum_m h_0[-(m-2k)]c_{j+1}(m),$$

$$d_j(k) = \sum_m h_1[m-2k]c_{j+1}(m) = \sum_m h_1[-(m-2k)]c_{j+1}(m),$$

Pozn.1: výraz $h_1[m-2k]$ představuje decimaci 2 po provedené konvoluci – objem dat proto zůstane zachován a zmenší se zároveň šířka pásma

Pozn.2: Koeficienty na nejvyšší hladině $c_{j+1}(k)$, tedy hladině s největší hodnotu indexu j představují vzorky signálu

Vlnková transformace

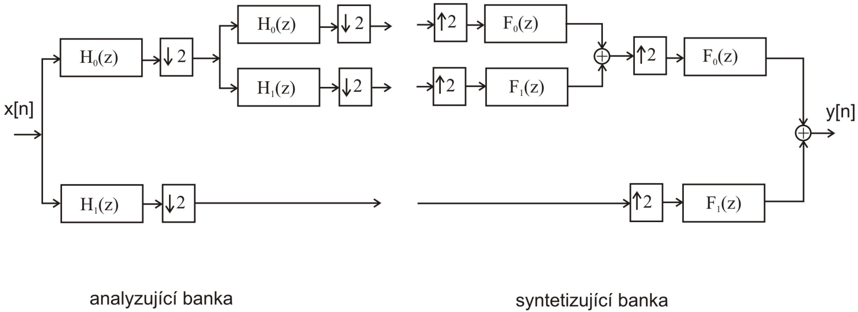
Rekurentní výpočet koeficientů vlnkové tranformace

B. Vztahy pro rekonstrukci (syntézu) jsou obdobné vztahům pro analýzu s těmito rozdíly:

- koeficienty rekonstručních filtrů mají opačné pořadí vzhledem k pořadí koeficientů dekompozičních filtrů
- místo decimace se provádí interpolace (vložení nul mezi vzorky a filtrace dolní nebo horní propustí)
- po interpolaci následuje součet koeficientů na příslušných hladinách

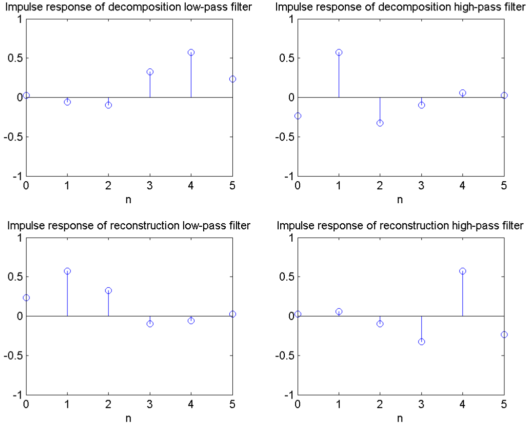
Realizace DWT bankou filtrů

Diskrétní vlnkovou transformaci lze tedy realizovat bankou filtrů s postupně se zvětšující šířkou pásma směrem k vyšším kmitočtům



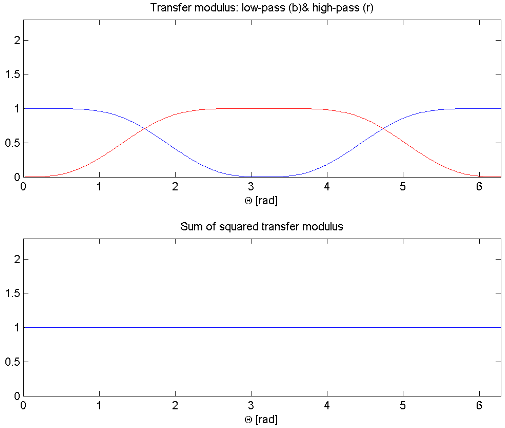
Výstupy analyzující banky filtrů (H_0 je dolní propust, H_1 je horní propust) označujeme jako hladiny (scales) - zde jsou 3 hladiny odpovídající 3 frekvenčním pásmům - odshora dolů (výstupy decimátorů): aproximace ($c_0 \leftrightarrow H_0 H_0$), detaily ($d_0 \leftrightarrow H_0 H_1$), detaily ($d_1 \leftrightarrow H_1$); dvě spodní frekvenční pásma pro c_0 a d_0 jsou stejně široká, horní frekvenční pásmo pro d_1 je dvakrát širší než frekvenční pásmo pro d_0. Vzhledem k mocnině dvou se jedná o oktávové dělení pásma.

Příklady perfektní rekonstrukce pro ortogonální DWT



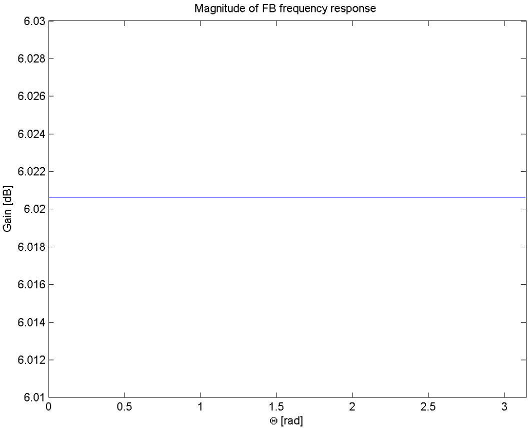
Impulsové odezvy filtrů dekompozičního (nahoře) a rekonstrukčního (dole) pro ortogonální vlnku Daubechies: db3

Příklady perfektní rekonstrukce pro ortogonální DWT



QMF filtr dekompoziční : dolní propust (modře) a horní propust (červeně)
Dole: součet kvadrátů modulů obou filtrů - jsou výkonově komplementární

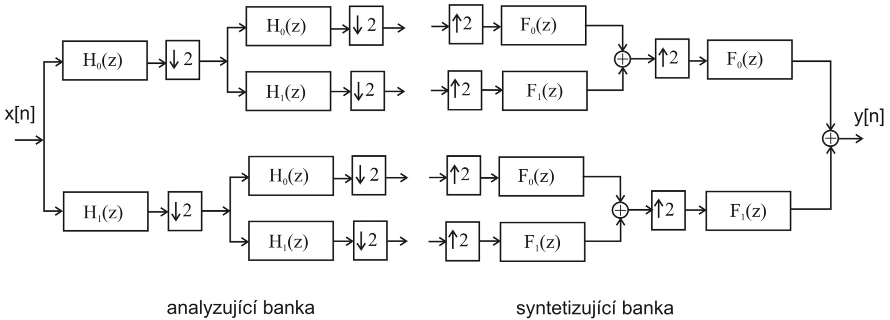
Příklady perfektní rekonstrukce pro ortogonální DWT



Celkový přenos v decibelech (detail) pro celou ortogonální banku filtrů s decimací a interpolací - žádné zvlnění :-)) - prostě perfektní rekonstrukce

Paketová DWT

Diskrétní vlnkovou transformaci lze též realizovat bankou filtrů s rovnoměrným dělením pásma, nicméně se jako DWT neoznačuje, ale používá se název paketová DWT (Wavelet Packet)



Pozn. V tomto případě se neprovádí kompletní rozklad, ale hledá se optimální cesta (graf, strom). Dělení na každém rozcestí se určuje pomocí vhodného kritéria založeného na entropii.

Biortogonální vlnková transformace

Biortgonální vlnková transformace

- tento typ transformace používá dvě různé vlnky: jednu pro dekompozici a druhou pro rekonstrukci
- ortogonální nejsou vektory dekompoziční báze navzájem[7], ale vektory báze dekompoziční a rekonstrukční navzájem
- výhodou je, že lze použít symetrické vlnky a tím zajistit linearitu fáze, kterou jsme kvůli podmínkám rekonstrukce u ortogonální DWT opustili - linearita fáze je nutná pro správnou rekonstrukci 1-D signálů a i obrazů
- dekompoziční a rekonstrukční filtry nejsou stejně dlouhé (u ortogonální DWT jsou stejně dlouhé), ale jsou symetrické
- zároveň lze podmínky perfektní rekonstrukce splnit s větší volností[8].
- neplatí Parcevalův teorém, který platí u ortogonální DWT

[7] Totéž platí pro rekonstrukční bázi

[8] I zde se ovšem občas používají aproximace splňující podmínky perfektní rekonstrukce přibližně

Redukce šumů - prahování koeficientů

Prahování koeficientů je používáno k redukci šumu pomocí DWT: po rozkladu jsou koeficienty na jednotlivých hladinách prahovány[9], a tím dochází k redukci šumu:

- tvrdé prahování - koeficienty menší než práh *th* se nahradí nulami - při tomto způsobu úpravy koefecientů c (c nahrazuje $c_j(k)$ nebo $d_j(k)$) vznikají nespojitosti

$$c' = \begin{cases} c & ; \quad |c| \geq th \\ 0 & ; \quad |c| < th \end{cases}$$

- měkké prahování - koeficienty menší než práh *th* se nahradí nulami a zbylé se zmenší o hodnotu prahu - nespojitosti se omezí, ale více se sníží energie signálu než v předchozím případě

$$c' = \begin{cases} sgn(c)(|c| - th) & ; \quad |c| \geq th \\ 0 & ; \quad |c| < th \end{cases}$$

[9] Malé koeficienty jsou nulovány, větší jsou ponechány nebo zmenšeny

Prahování koeficientů

Určení velikosti prahu[10]

Za předpokladu, že aditivní šum má normální rozdělení s nulovou střední hodnotu a rozptylem σ^2, je práh *th* dán vztahem

$$th = \sigma\sqrt{2ln(N)},$$

kde N je počet vzorků signálu

Rozptyl šumu σ^2 lze odhadnout pomocí vztahu

$$\sigma = \frac{1}{0.6745} med(|d_j(k)|),$$

kde $d_j(k)$ jsou detaily na nejjemnější úrovni rozkladu[11] a *med* označuje medián

[10] Jedna z mnoha možností

[11] Tedy v souladu se zavedeným značením je to nejvyšší hodnota indexu j pro hladiny detailů