

Úvod do vlnkové transformace

Radislav Šmíd

ČVUT FEL katedra měření, Technická 2, CZ-166 27 Praha 6

e-mail: smid@feld.cvut.cz, www: <http://measure.feld.cvut.cz/usr/staff/smid>

9. srpna 2001

Obsah

1	Spojité vlnková transformace	1
1.1	Vlastnosti CWT	2
1.1.1	Linearita	2
1.1.2	Invariance v čase	2
1.1.3	Dilatace	3
1.2	Požadavky na prototyp vlnky	3
2	Diskrétní vlnková transformace	3
3	Problém konečné délky signálu	5
4	Mateřské funkce	6
4.0.1	Vlnka Mexican hat	6
4.0.2	Morletova vlnka	6
4.0.3	Meyerova vlnka	7
4.0.4	Haarova vlnka	7
4.0.5	Vlnka Daubechies	8
4.0.6	Výběr vlnky	8

1 Spojitá vlnková transformace

Vznik vlnkové transformace (*wavelet transform*) je jedním z výsledků snahy získat časově-frekvenční popis signálu¹. Historicky starší Fourierova transformace poskytuje informaci o tom, které frekvence se v signálu nacházejí, nevypovídá však o jejich umístění (poloze) v čase, je tedy vhodná jen pro popis stacionárních signálů. Možným řešením uvedeného problému je použití okna, které v čase ohraničí krátký úsek signálu a umožní z něj určovat spektrum v daném časovém intervalu (tento postup se nazývá *Short-Time Fourier Transform*, varianta s Gausovským oknem Gaborova transformace, *Gabor Transform*). Z obdoby Heisenbergova principu neurčitosti vyplývá, že nelze současně určit přesně frekvenci a polohu jejího výskytu v čase. Proto má uvedené řešení pro časově konstantně široké okno pro všechny kmitočty velkou rozlišitelnost ve frekvenci a malou v čase a naopak pro časově úzké okno velkou rozlišitelnost v čase a malou ve frekvenci. Ideou vlnkové transformace je vhodnou změnou šířky "okna" v čase a jeho tvarem dosáhnout optimálního poměru rozlišitelnosti v čase a frekvenci. Pro nízké frekvence je "okno" širší, pro vysoké užší (obr. 1). Toto "okno" se nazývá mateřská vlnka ψ (*mother wavelet*). Pomocí parametru s , který se jmenuje měřítko, je možné měnit její šířku (dilatace), parametrem τ zvaným poloha, se mění umístění vlnky na časové ose (translace).

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad s, \tau \in \mathbf{R}, s \neq 0 \quad (1)$$

¹Nezávisle proměnnou samozřejmě nemusí být čas, ale v různých aplikacích jde např. o polohu, dráhu atd.

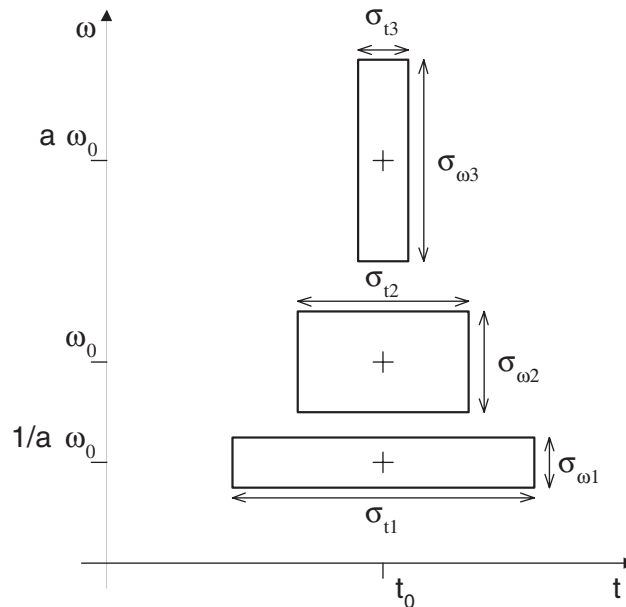
Člen $\frac{1}{\sqrt{s}}$ slouží k normalizaci energie vlnky při změnách měřítka, $\psi(t)$ je tzv. prototyp vlnky. Spojitá vlnková transformace (*Continuous Wavelet Transform, CWT*) je pak definována pro signály s konečnou energií², t.j. $f \in L^2(\mathbf{R})$, takto:

$$Wf(\tau, s) = \int_{-\infty}^{+\infty} f(t) \overline{\psi_{\tau,s}(t)} dt \quad (2)$$

kde $\overline{\psi}$ označuje číslo komplexně sdružené. Výsledkem pro jednorozměrný signál je dvojrozměrná funkce, která se nazývá vlnkové koeficienty $Wf(\tau, s)$. Po dosazení získáme tvar

$$Wf(\tau, s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \overline{\psi\left(\frac{t-\tau}{s}\right)} dt \quad (3)$$

Grafické vyjádření vlnkových koeficientů v rovině (τ, s) se často nazývá scalogram nebo vlnková mapa (*wavelet map*). Souvislost mezi měřítkem a frekvencí složek detekovaných CWT se musí stanovovat pro každou vlnku zvlášť. Doporučeným způsobem [4] je nalezení měřítka, při kterém nastává maximum vlnkových koeficientů při transformaci sinusovky referenční frekvence. U některých vlnek je souvislost přímá (Mexican hat, Morletova vlnka), u jiných (Daubechies) je vztah diskutabilní.



Obrázek 1: Časově-kmitočtové rozlišení vlnkové transformace, šířka v příslušné souřadnici je označena σ_t resp. σ_ω

1.1 Vlastnosti CWT

1.1.1 Linearita

$$(W(af_1 + bf_2))(\tau, s) = a(Wf_1)(\tau, s) + b(Wf_2)(\tau, s) \quad (4)$$

Linearita přímo vyplývá z vlastností skalárního součinu v (2).

1.1.2 Invariance v čase

$$W\hat{f}(\tau, s) = Wf(\tau - b, s) \quad \hat{f}(t + b) = f(t) \quad (5)$$

Invariance v čase popisuje skutečnost, že posun analyzované funkce po časové ose způsobí stejný posun vlnkových koeficientů po ose polohy. Tento fakt lze odvodit z představy, že CWT může být interpretována

² $\int_{-\infty}^{+\infty} |f(t)|^2 dt < \infty$

pomocí množiny lineárních časově-invariantních filtrů. Tato vlastnost se ztrácí při přechodu k diskrétní vlnkové transformaci a jak bude dále uvedeno, v některých aplikacích představuje velkou nevýhodu.

1.1.3 Dilatace

$$W\hat{f}(\tau, s) = (Wf)(a\tau, \frac{s}{a}) \quad \hat{f} = \sqrt{|a|}f(at) \quad a \neq 0 \quad (6)$$

Vztah popisuje závislost mezi CWT originální funkce a její roztaženou nebo zúženou podobou, ve vlnkových koeficientech dojde k adekvátnímu roztažení v ose polohy a k posunu v ose měřítka.

1.2 Požadavky na prototyp vlnky

Pro zajištění invertibility transformace nemůže být funkce pro vlnku ψ volena libovolně, ale musí splňovat určité podmínky [9]. Musí mít nulovou střední hodnotu (7) a vhodný frekvenční obsah (8).

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (7)$$

$$\int_0^{+\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty \quad (8)$$

kde $\Psi(\omega)$ Fourierův obraz $\psi(t) \in L^2(\mathbf{R})$.

Lze také říci, že každá netriviální absolutně integrovatelná funkce $g \in L^1(\mathbf{R})$ s nulovou střední hodnotou je přípustná vlnka.

2 Diskrétní vlnková transformace

Vhodnou dvojkovou závislostí parametrů s a τ můžeme vytvořit z vhodné vlnky ψ ortonormální bázi:

$$s = 2^p \quad \tau = 2^p k \quad p, k \in \mathbf{Z} \quad (9)$$

pak

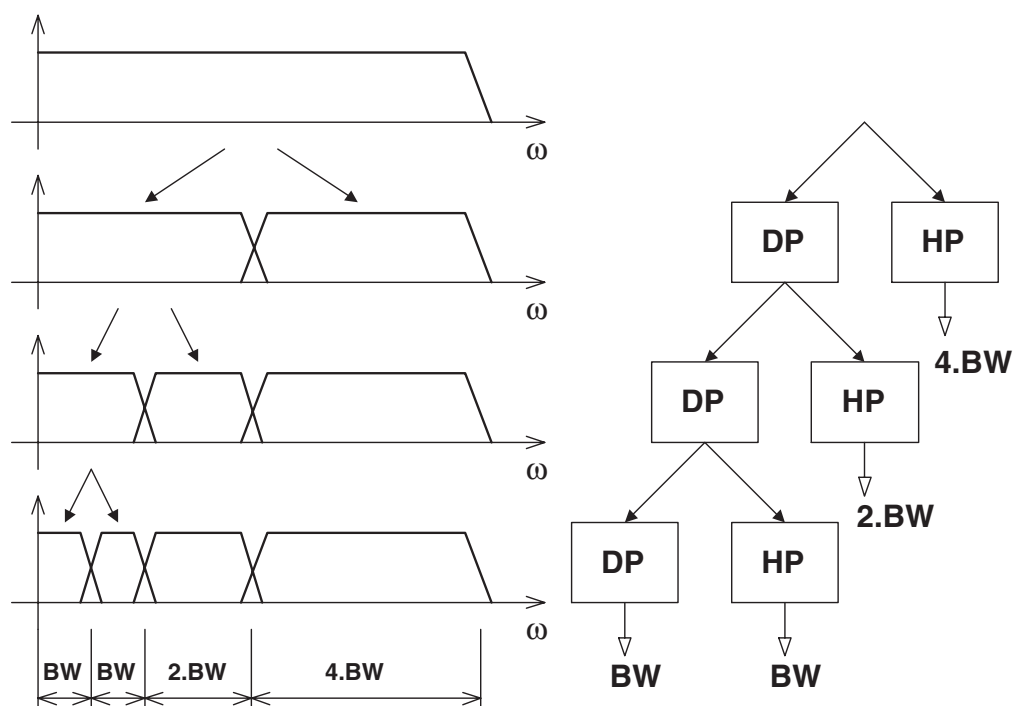
$$\psi_{k,p}(t) = \frac{1}{\sqrt{2^p}} \psi\left(\frac{t - 2^p k}{2^p}\right) \quad (10)$$

kde p odpovídá měřítku, k poloze. Díky ortonormalitě pak takto volená vlnka umožňuje neredundantní dekompozici signálu, tzv. analýzu s mnoha rozlišeními (*multiresolution analysis, decomposition*). Tento princip je základem diskrétní vlnkové transformace (*Discrete Wavelet Transform, DWT*). Vlnková funkce ψ se chová jako pásmová propust filtrující vstupní signál kolem centrálního kmitočtu, který je závislý na měřítku mocninou dvou, v následujícím měřítku je filtrována horní polovina pásma předchozí dolnofrekvenční části signálu (obr. 2). S rostoucím kmitočtem roste šířka pásma (BW) tohoto filtru, činitel jakosti Q je tak konstantní pro celou množinu měřítkem odvozených filtrů. Pro zvolené minimální měřítko však zůstává nepokryto pásmo od nižších kmitočtů do nuly. Proto je od vlnky ψ odvozena měřítková funkce ϕ (scaling function), která má charakter dolní propusti. Podobně jako u vlnky (7) musí ϕ splňovat podmínku přípustnosti:

$$\int_{-\infty}^{\infty} \phi(t) dt = 1 \quad (11)$$

DWT můžeme také chápat jako speciálně vzorkovanou CWT, která musí splňovat následující podmínky:

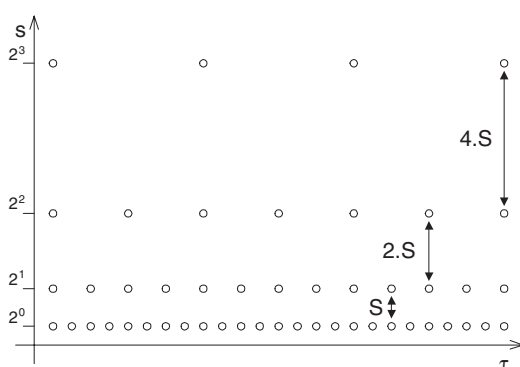
1. vzorkování časově-měřítkového prostoru musí probíhat na dvojkové mřížce, obr. 3
2. použitá vlnka musí vytvářet ortonormální bázi analyzovaného prostoru
3. analyzující vlnka musí mít kompaktní nosič



Obrázek 2: Frekvenční pohled na diskretní vlnkovou transformaci

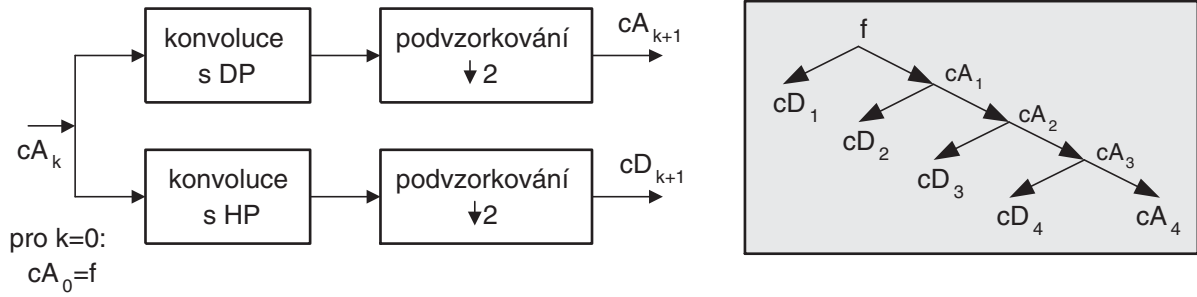
V tomto případě lze počítat DWT rychlým algoritmem, tvořeným filtrací FIR filtry a podvzorkováním (decimací), obr. 4. Volnější formulaci uvedených podmínek umožňuje tzv. *Overcomplete Wavelet Transform (OCWT)* [9], což je rozšíření WT o vzorkování CWT v rovině čas-měřítko na množině vzorků, která nemusí být dvojková jako u DWT. Pro její implementaci byly vyvinuty rychlé algoritmy podobně jako u DWT, nejčastěji používanější verzí je ekvidistantní vzorkování v čase a logaritmické v měřítku. Má následující výhody:

- možnost nastavení odolnosti vůči šumu změnou redundance reprezentace
- nepožaduje splnění podmínek ortonormality vlnky
- umožňuje nestejnoměrně vzorkovat signál s ohledem na jeho časový nebo frekvenční průběh



Obrázek 3: Dvojková mřížka v prostoru čas-měřítko

Oba filtry, dolní propust h (scaling filter) a horní propust g (wavelet filter), tvoří pár kvadraturních zrcadlových filtrů (QMF), které mají komplementární propustná pásma. Výstupy obou filtrů jsou podvzorkovány na polovinu vstupních vzorků. Horní propust poskytuje koeficienty tzv. detailů DWT (cD), dolní propust koeficienty tzv. aproximace (cA). Díky decimaci je celkový počet koeficientů po jednom kroku stejný jako počet vstupních vzorků. Koeficienty aproximace lze dále analyzovat shodným rozkladem filtry a obdržet tak další



Obrázek 4: Jeden krok DWT (vlevo), rozklad na aproximace a detaily (vpravo)

soubor koeficientů aproximace a detailů. Tak lze postupovat až do vyčerpání vstupní sekvence. Jednotlivé konvoluce s decimací lze formalizovat následujícími vzorci:

$$cA_{p+1}(k) = \sum_{m=1}^N h(m-2k) cA_p(m) \quad (12)$$

$$cD_{p+1}(k) = \sum_{m=1}^N g(m-2k) cA_p(m) \quad (13)$$

kde cA_p, cD_p jsou koeficienty aproximace resp. detailů měřítka $p \in \langle 0, P \rangle$ a N je délka vstupní sekvence $cA_p, cA_0 = f(n)$. Pár kvadraturních zrcadlových filtrů můžeme chápat jako ortonormální jádro transformace obdobné tzv. motýlku u FFT.

Podle výchozí formulace vlnky jsou určeny filtry h a g . Nejčastěji je z měřítkové funkce ϕ odvozen [5, 8] měřítkový filtr w , který má charakter dolní propusti. Po normalizaci w podělením normou poskytné koeficienty dolnoproputního filtru h . K němu se vypočítá horní propust g jako komplement (kvadraturní zrcadlový filtr). Rekonstrukční FIR filtry \bar{h} a \bar{g} jsou pak určeny jako časově obrácené sekvence koeficientů dekompozičních filtrů h a g . Jiné metody návrhu používají například iterační postupy (vycházejí z tzv. dilatační rovnice Daubechiesové) [7, 5].

Pokud má analyzovaný signál délku $N = 2^k$ a maximální měřítko dekompozice je $P \leq k$, pak obdržíme $N2^{-1} + N2^{-2} + \dots + N2^{-P+1} + N2^{-P}$ koeficientů detailů: $cD_1, cD_2, \dots, cD_{(P-1)}, cD_P$ a $N2^{-P}$ koeficientů aproximace cA_P . Celkem obdržíme N koeficientů, které můžeme označit $DW f$.

Protože je počet koeficientů shodný s počtem vzorků a nedochází ke ztrátě informace, popis signálu je neredundantní. Popis je také úplný, pomocí inverzního postupu k postupu na obr. 4 lze přesně rekonstruovat analyzovaný signál. Inverzní diskretní vlnková transformace se označuje IDWT. Operace podvzorkování je nahrazena převzorkováním, kdy za každým vzorkem původní sekvence následuje doplněný nulový vzorek. Místo původních filtrů jsou použity rekonstrukční filtry \bar{h} a \bar{g} . Výsledná aproximace cA_p je použita spolu se vstupními detaily cD_p jako vstup dalšího kroku IDWT (jde o pohyb nahoru ve schématu pyramidálního rozkladu v obr. 4 vpravo).

DWT není invariantní v čase, pokud $\hat{f}(n+b) = f(n)$ pak většinou neplatí $DW \hat{f} = DW f$.

3 Problém konečné délky signálu

Problém konečné délky signálu (*border distortion*) se projevuje na okrajích intervalu, na kterém je analyzovaný signál definován. Je důsledkem konečné délky obou signálů při konvoluci, u CWT v (3), u DWT při konvoluční filtraci. Pro zmírnění nebo odstranění lze použít různých metod podle charakteru signálu:

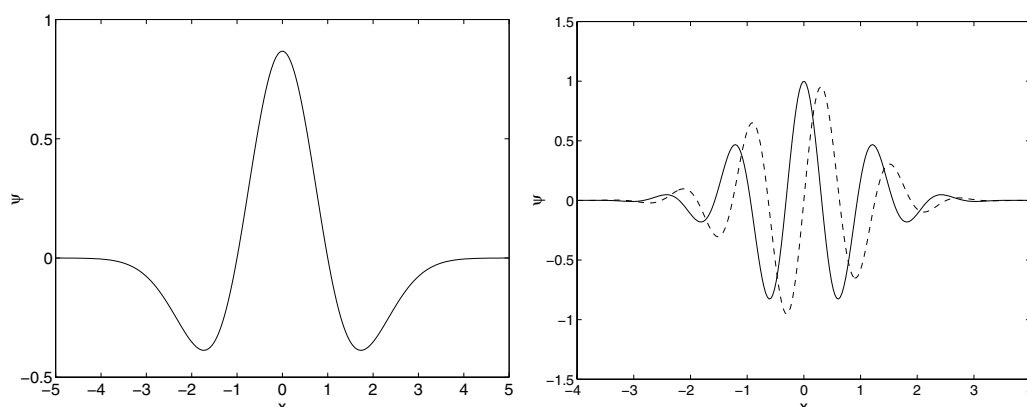
1. doplnění signálu:

- (a) nulami - chybějící část nutná pro výpočet konvoluce se doplní nulami. Jde o jednoduché řešení, které většinou zapříčiní vznik diskontinuit v analyzovaném signálu a odpovídajících artefaktů na krajích výsledku transformace, vhodné pouze pro signály s tvarem odpovídajícím modulaci vhodným oknem.

- (b) extrapolace konstantou - doplnění signálu konstantními hodnotami okrajových bodů, přináší stejné nevýhody jako předchozí metoda
 - (c) symetrizace - doplnění původním signálem symetrické kolem okrajového bodu. Vyvolá vznik diskontinuit v první derivaci a odpovídajících artefaktů na krajích výsledku transformace, vhodné zejména pro 2D transformaci obrazů.
 - (d) extrapolace s hladkou první derivací, vhodná pro signály reprezentované hladkou funkcí
2. výpočet ve frekvenční oblasti s využitím okénkové funkce
 3. periodizace - doplnění signálu periodickým opakováním původního, vhodná pro signály periodického charakteru s periodou kmitočtově nejnižší složky rovnou délce analyzovaného úseku, nevyvolává vznik artefaktů na okrajích výsledku transformace

4 Mateřské funkce

Následující přehled stručně uvádí některé rozšířené vlnky [3, 4, 1, 5, 9, 6]. Z vlastností je uvedena vhodnost pro CWT a DWT, typ nosiče a typ symetrie. V případě nekompaktního nosiče je numerický výpočet prováděn na efektivním nosiči.



Obrázek 5: Vlnka Mexican hat (vlevo) a Morletova vlnka (vpravo, imaginární část čárkovaně)

4.0.1 Vlnka Mexican hat

Vlnka Mexican hat (Mexický klobouk, Marrova vlnka, obr.5) má tvar druhé derivace průběhu hustoty pravděpodobnosti Gaussova rozdělení.

$$\psi(x) = \frac{2}{\sqrt{3}} \pi^{\frac{1}{4}} (1 - 2x^2) e^{-x^2} \quad (14)$$

Vlastnosti: symetrická, nemá kompaktní nosič, vhodná pro CWT, není ortogonální (nelze použít pro DWT). Vlnka je členem rodiny Gaussovských vlněk tvořené jednotlivými derivacemi průběhu hustoty pravděpodobnosti Gaussova rozdělení.

4.0.2 Morletova vlnka

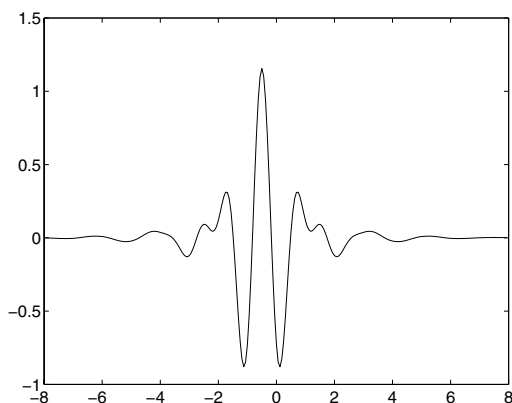
Má tvar komplexní sinusovky modulované Gaussovským oknem (obr.5). Je výsledkem kompromisu mezi polohovou lokalizací jednorázových dějů (lepší je např. vlnka Mexican hat) a frekvenčním rozlišením (Fourierova transformace).

$$\psi(x) = a \cdot e^{-\frac{1}{2}x^2} (\cos(5x) + j \sin(5x)) \quad (15)$$

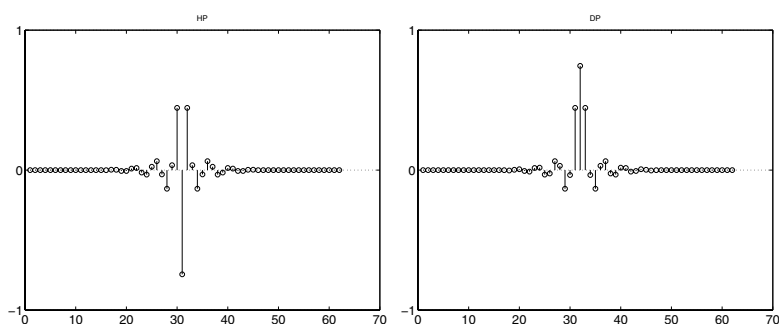
Vlastnosti: symetrická, komplexní, nemá kompaktní nosič, vhodná pro CWT, není ortogonální (nelze použít pro DWT).

4.0.3 Meyerova vlnka

Meyerova vlnka (obr. 6) je definována ve frekvenční doméně, nemá explicitní vzorec pro vyjádření v čase. V originálním tvaru nemůže být realizována FIR filtry a tudíž použita v rychlém algoritmu DWT, proto byla vytvořena její diskretní aproximace s filtry na obr. 7.



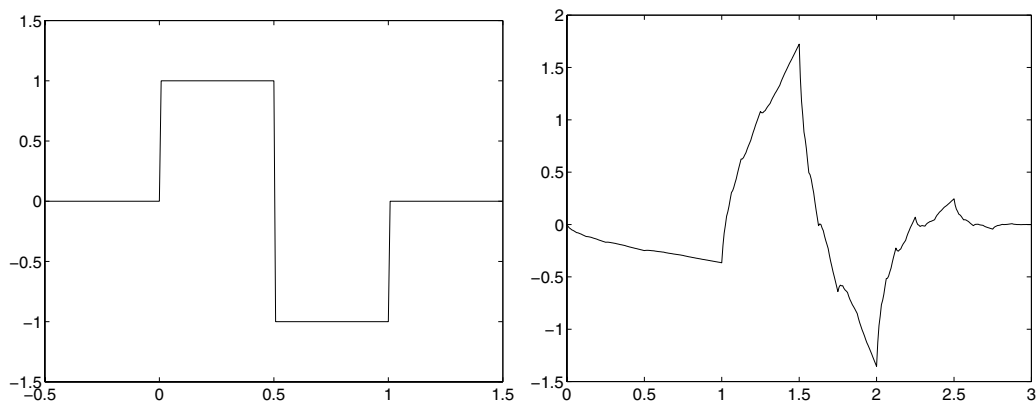
Obrázek 6: Meyerova vlnka



Obrázek 7: Konvoluční filtry pro Meyerovu vlnku

Vlastnosti: symetrická, nemá kompaktní nosič (aproximace má), vhodná pro CWT i DWT, ortogonální.

4.0.4 Haarova vlnka



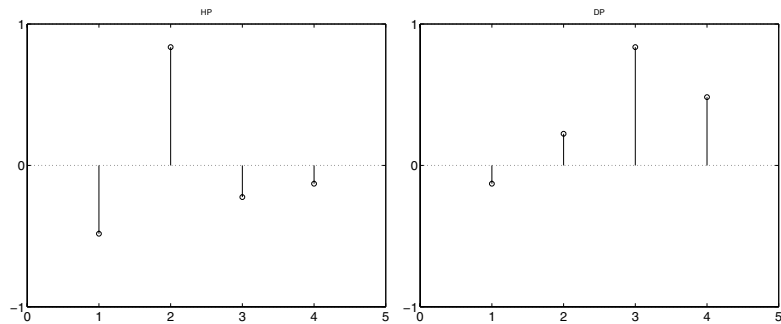
Obrázek 8: Haarova vlnka, vlnka Daubechies2

Haarova vlnka (obr. 8) představuje velmi jednoduchou vlnku, která ale neumožňuje hladkou rekonstrukci signálu. Bývá často nazývána Daubechies řádu 1.

$$\psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{jinde} \end{cases} \quad (16)$$

Vlastnosti: symetrická, má kompaktní nosič, vhodná pro CWT i DWT, je ortogonální, jednoduchá a efektivní implementace³.

Nespojitost Haarovy vlnky představuje přes všechny ostatní výhodné vlastnosti velkou nevýhodu v její aplikaci.



Obrázek 9: Konvoluční filtry pro vlnku Daubechies2

4.0.5 Vlnka Daubechies

Vlnky Daubechies (obr. 8, 9) představují skupinu vlněk různého řádu $N \geq 1$. Nemají (kromě Daubechies řádu 1) explicitní vyjádření $\psi(x)$.

Vlastnosti: asymetrická (kromě Daubechies řádu 1), má kompaktní nosič délky $2N - 1$, vhodná pro CWT i DWT, je ortogonální.

4.0.6 Výběr vlnky

Ve velké části dostupné literatury popisující aplikace vlnkové transformace se uvádí, že výběr použité vlnky byl proveden zkusmo nebo intuitivně. Bylo nalezeno několik souvislostí mezi řešenou úlohou (charakterem analyzovaného signálu) a vhodnou vlnkou. Tato pravidla lze shrnout do následujících doporučení:

- Komplexní vlnky jako Morletova detekují dobře oscilace, nejsou vhodné pro detekci osamocených singularit.
- Čistě reálné vlnky s málo oscilacemi dobře detekují špičky a singularity v signálu.
- Antisymetrické vlnky jsou vhodné k detekci změn gradientu.
- Symetrické vlnky nezpůsobují fázový posun mezi špičkou, singularitou, oscilací v signálu a příslušným projevem ve vlnkových koeficientech.
- Pro současnou detekci amplitudy a fáze je nutné použít komplexní vlnku (např. Morletovu)

Reference

- [1] M. O. Berger, editor. *12th International Conference on Analysis and Optimization of Systems. Images, Wavelets and PDEs*, Paris, June 1996.
- [2] T. Edwards. *Discrete wavelet transforms: Theory and implementation*. Technical report, Stanford University, September 1992.

³použití v hardwareově realizovaných vlnkových analyzátoch [2], DSP apod.

- [3] A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2(2), 1995.
- [4] J. Lewalle. Tutorial on continuous wavelet analysis of experimental data. Technical report, Syracuse University, April 1995.
- [5] A. K. Louis, P. Maaß, and A. Rieder. *Wavelets: Theory and Applications*. John Wiley and Sons Ltd., England, 1997.
- [6] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [7] A. Procházka, M. Kolínová, and J. Stříbrský. Signal segmentation using time-scale signal analysis. In *IX European Signal Processing Conference EUSIPCO-98*, Island of Rhodes, Greece, 1998.
- [8] G. Strang. Wavelets and dilation equations: A brief introduction. *SIAM Review*, 31:613–627, 1989.
- [9] A. Teolis. *Computational Signal Processing with Wavelets*. Birkhäuser, Boston, 1998.