...

W. Hebisch.

11 maja 2022

# 1 Uzupełnienie o iteracji Czebyszewa

Powiedzieliśmy że błąd przy iteracji punktu stałego

$$x_{i+1} = x_i + \gamma_i(Ax_i - b)$$

spełnia

$$x_{i+1} - x_\infty = (x_i - x_\infty) + \gamma_i A(x_i - x_\infty).$$

gdzie $x_\infty$ jest dokładnym rozwiązaniem. Efekt iteracji można zapisać jako

$$x_n - x_\infty = P(A)(x_0 - x_\infty)$$

gdzie

$$P(z) = \prod_{j=1}^{n}(1 - \gamma_i z)$$

jest wielomianem takim że $P(0) = 1$. Zakładamy że $A$ jest macierzą symetryczną dodatnie określoną, taką że spektrum jest zawarte w przedziale $[m, M]$. Aby zminimalizować błąd szukamy wielomianu $P$ takiego że $P(0) = 1$ zaś

$$\sup_{z\in[m,M]} |P(z)|$$

jest minimalne. Powiedzieliśmy że rozwiązaniem jest przeskalowany wielomian Czebyszewa $T_n$, tzn. $P(z) = cT_n(az + b)$ gdzie $az + b$ przeształca przedział $[m, M]$ na $[-1, 1]$ zaś $c$ jest tak dobrane by $P(0) = 1$. Mamy

$$T_n(z) = 2^{-n}\cos(n\arccos(z)).$$

$n\arccos(z)$ przyjmuje wszystkie wartości w przedziale $[0, \pi n]$. Na tym przedziale cosinus zeruje się w punktach $k\pi/2$ gdzie $k = 1, 3, \ldots, 2n - 1$ jest liczbą nieparzystą. Daje to $n$ zer, a więc wszystkie zera wielomianu Czebyszewa spełniają

$$k\pi/2 = n\arccos(z)$$

czyli

$$z = \cos(\frac{k\pi}{2n})$$

z $k$ jak wyżej. Przekształcenie z $[m, M]$ na $[-1, 1]$ jest zadane wzorem

$$\frac{2(z-m)}{M-m} - 1$$

czyli $z$ jest zerem $P$ gdy

$$\frac{2(z-m)}{M-m} - 1 = \cos(\frac{k\pi}{n})$$

czyli

$$z = \frac{1}{2}\left((M-m)\cos(\frac{k\pi}{n}) + M + m\right).$$

Oznaczając przez $z_k$ wartość odpowiadającą $k$ mamy

$$P(z) = c\prod(z - z_k) = \prod(1 - \frac{z}{z_k}).$$

Oznaczmy przez $\theta_n(i)$ odwzorowanie 1-1 z liczb całkowitych od 1 do $n$ w liczby nieparzyste od 1 do $2n-1$. Wzór wyżej oznacza że $\gamma_i = \frac{1}{z_{\theta_n(i)}}$ dla pewnego $\theta_n$. Pozostaje dobrać $\theta_n$, tzn. dobrać kolejność zer. Różne kolejności nie są równoważne. By zminimalizować propagację błędów chcemy zminimalizować

$$\max_{j=1,\ldots,n} \sup_{z\in[m,M]} \prod_{i=j}^{n} |1 - \frac{z}{x_{\theta_n(i)}}|.$$

Jednakże, w praktyce trzeba więcej: warunek wyżej jedynie mówi o wpływie błędu z pośrednich etapów na końcowy wynik. Jeśli w pośrednich etapach błędy bedą zbyt silnie narastać to mogą stać się większe od poprawnego wyniku i przez to spowodwać znaczne błędy zaokrąglenia. Kryterium dobrego doboru kolejności nie jest jednoznaczne, poniżej pokażemy jedną z możliwych konstrukcji, która jest w pewnym sensie optymalna. Konstrukcja $\theta_n$ jest rekursywna. Dla parzystego stopnia przyjmujemy

$$\theta_{2n}(2i-1) = \theta_n(i),$$

$$\theta_{2n}(2i) = 4n - \theta_n(i)$$

dla $i = 1, \ldots, n$. Dla nieparzystego stopnia

$$\theta_{2n+1}(2i-1) = \theta_n(i),$$

$$\theta_{2n+1}(2i) = 4n + 2 - \theta_n(i)$$

dla $i = 1, \ldots, n$ oraz

$$\theta_{2n+1}(2n+1) = 2n+1.$$

Uzasadnienie wyboru wymaga trochę wysiłku, dlatego je pominiemy. Ale łatwo zobaczyć że przyjęcie $\theta_{2n+1}(2n+1) = 2n+1$ jest dobre. Mianowicie odpowiedni pierwiastek jest w środku przedziału, czyli odpowiedni czynnik ma najmniejsze

supremum na przedziale, toteż dodanie tego czynnika nie może powiększyć supremum, co oznacza że w minimalizujemy wpływ błędu na końcowy wynik.

Widać też że nasza reguła dodaje pierwiastki parami, tzn. po dodaniu pierwiastka $z_k$ wielomianu Czebyszewa dodamy $-z_k$ (to jest w terminach oryginalnych wielomianów Czebyszewa, faktycznie dodajemy przeskalowane pierwiastki). Czyli dodamy czynnik kwadratowy, symetryczny względem środka przedziału. Mamy wzór

$$T_{2n}(z) = 2^{-n} T_n(2T_2(z)).$$

Z tego wzoru wynika że nasz czynnik kwadratowy na taki sam wpływ na pośrednie produkty jak czynnik liniowy z $T_n$, tyle że przy innym podstawieniu (przedziale). Naturalne jest to dodawanie czynników kwadratowych do $T_{2n}$ w kolejności czynników liniowych dla $T_n$, pojedynczy czynnik nie powinien zbyt wiele zmienić, a jeśli kolejność dla $T_n$ była dobra to łącznie kolejność dla $T_{2n}$ też powinna być dobra. Nie jest to pełne uzasadnienie, ale pokazuje że reguła jest rozsądna.

Innymi słowy, w miarę łatwo można zrobić dość dobry wybór kolejności, to że pod pewnym względem wybór jest optymalny niewiele tu poprawia (choć jest to miła własność). Z drugiej strony, po prostu mnożąc czynniki po kolei dostaniemy wykładniczy wzrost maksimum, czyli mielibyśmy niestabilność numeryczną. Mnożąc czynniki w odwróconej kolejności w pierwszych etapach dostaniemy znaczny wzrost pośrednich wyników co też prowadzi do problemów.

## 2 Uzupełnienie o FFT

Dla niektórych operatorów znane są szybkie metody przejścia z bazy odpowiadającej wartościon w punktach do bazy złożonej z funkcji własnych i z powrotem. Typowym przykładem jest algorytm FFT i jego warianty. W zespolonym wariancie exponenty $e_j(k) = \exp(2\pi ijk/N)$, gdzie $i$ jest jednostką urojoną zaś $j = 0, 1, \ldots, N-1$, tworzą bazę dla operatora różnicy w przód $(Af)(k) = f(k+1) - f(k)$ z okresowymi warunkami brzegowymi $f(k+N) = f(n)$. Łatwo sprawdzić że $e_j$ tworzą układ ortogonalny względem standartowego produktu skalarnego

$$\langle f, g \rangle = \sum_{k=0}^{N-1} f(k)\overline{g(k)}$$

Z ortogonalności jeśli $f = \sum_j c_j e_j$ to

$$c_j = \frac{1}{N} \sum_{k=0}^{N-1} f(k) e_{-j}(k).$$

Podobny wzór daje $f$ w terminach $e_j$. Te wzory faktycznie oznaczają mnożenie przez odpowiednią macierz która jest gęsta (wszysktie elementy są niezerowe). Na pierwszy rzut nie widać żeby to się dało zrobić prościej.

Jednakże, oznaczając

$$(DFT(f))(j) = \sum_{k=0}^{N-1} f(k)a^{kj}$$

z $a = e_{-1}(1) = \exp(-2\pi i/N)$ widzimy że przejścia między bazami sprowadzają się do obliczania $DFT$. Ponadto $a$ jest pierwiastkiem pierwotnym stopnia $N$ z 1. Jeśli $N = ML$ to biorąc $j = Ls + t$, $k = Ml + n$, mamy

$$(DFT(f))(j) = \sum_{n=0}^{M-1} (a^L)^{ns} a^{nt} \sum_{l=0}^{L-1} f(Ml + n)(a^M)^{lt}$$

Mianowicie

$$jk = (Ml + n)(Ls + t) = Lns + nt + Mlt + Nst$$

Mamy $a^N = 1$, czyli

$$(DFT(f))(j) = \sum_{k=0}^{N-1} f(k)a^{kj} =$$

$$\sum_{n=0}^{M-1} \sum_{l=0}^{L-1} f(Ml + n)a^{(Ml+n)(Ls+t)} =$$

$$\sum_{n=0}^{M-1} \sum_{l=0}^{L-1} f(Ml + n)a^{Lns+nt+Mlt} =$$

$$\sum_{n=0}^{M-1} (a^L)^{ns} a^{nt} \sum_{l=0}^{L-1} f(Ml + n)(a^M)^{lt}$$

Powyższe wyrażenie można obliczać jako $M$ aplikacji $DFT$ długości $L$ do odpowiednich fragmentów $f$, mnożenie po składowych wektorów długości $N$ i $L$ aplikacji $DFT$ długości $M$. Dokładniej, niech $g_{n,t}$ będzie wewnętrzną sumę i ozanczmy

$$f_n(l) = f(Ml + n),$$

$$h_n(t) = a^{nt} g_{n,t}$$

Wtedy

$$g_{n,t} = DFT(f_n)(t),$$

i na mocy wzoru wyżej

$$DFT(f)(j) = DFT(h_n)(s)$$

Rekursywnie stosując ten wzór $DFT$ można obliczać przy pomocy $O(n \log(n))$ operacji. Dokładniej, jeśli $N = 2^r$, to biorąc $M = 2$ i $L = 2^{r-1}$ widzimy że

zewnętrzene $DFT$ zawierają sumę długości 2, czyli można je obliczyć kosztem $2N$ mnożeń i $N$ dodawań. Czynnik $a^{nt}$ przy naiwnym podejściu wymagałby $N$ mnożeń, ale te mnożenia i zewnętrzne $DFT$ można połączyć w jeden operator. Podobnie jak $DFT$ ten operator odpowiada mnożeniu przez macierz. Dzięki temu że zewnętrzne $DFT$ ma długość 2 odpowiednia macierz ma tylko dwa niezerowe elementy w każdym wierszu, w więc mnożenie przez tą macierz można wykonać przy pomocy $2N$ mnożeń i $N$ dodawań.

Wewnęrzne $DFT$ długości $2^{r-1}$ obliczamy rekursywnie w podobny sposób. Czyli łącznie mamy $r$ kroków, a w każdym kroku $2N$ mnożeń i $N$ dodawań co w sumie daje $2rN$ mnożeń i $rN$ dodawań. Jako że $r$ to $O(\log(N))$ to dostajemy obiecaną złożność $O(N\log(N))$.

Uwaga: przy naiwnych obliczaniu najbardziej kosztowne byłoby obliczanie potęg $a$. Ale przy ustalonym $N$ te potęgi można stablicować, dlatego pomijam koszt ich obliczania.

Powyższe oszacowanie można nieco polepszyć zauważając że w $DFT$ długości 2 współczynniki to 1 i $-1$. A więc zamiast mnożyć przez $-1$ możemy odejmować co oszczędza ponad połowę mnożeń. Nieco lepszy wynik można uzyskać stosując jako podstawowy blok $DFT$ długości 4. Wtedy współczynniki to 1, $i$, $-1$ i $-i$. Przy reprezentacji liczb zespoloych jako pary liczb rzeczywistych składające się z części rzeczywistej i urojonej mnożenie przez $i$ sprowadza się do zastąpienia części rzeczywistej przez urojoną zaś części urojonej przez minus część rzeczywistą. A więc $DFT$ długości 4 można obliczyć kosztem porównywalnym do 3 dodawań zespolonych. Tak jak poprzednio potrzeba $N$ mnożeń przez czynniki $a^{nt}$. Dla parzystego $r$ używając $DFT$ długości 4 potrzebujemy $r/2$ kroków, a więc łączny koszt jest porównywalny z $(3/2)rN$ dodawań zespolonych i $rN/2$ mnożeń zespolonych. Jako że dodawanie zespolone wymaga dwu dodawań rzeczywistych a mnożnie zespolone można zrealizować przy pomocy 4 mnożeń rzeczwistych i dwu dodawań to w terminach operacji rzeczywistych mamy $2rN$ mnożeń rzeczywistych i $4rN$ dodawań.

Dla porównania wersja ze $DFT$ długości 2 potrzebowała $4rN$ mnożeń rzeczywistych i $4rN$ dodawań.

Jeśli $N$ nie jest potęgą 2 to sytuacja się komplikuje. Możemy użyć metodę wyżej o ile $N$ ma niezbyt duży czynnik pierwszy $p$, koszt odpowiedniego kroku jest wtedy rzędu $O(Np)$, jednakże stałe są większe niż dla $p = 2$. W praktyce często możemy powiększyć $N$ do kolejnej potęgi 2. Większe $N$ powiększa koszt (potencjalnie nawet dwukrotnie, średnio rzędu $\sqrt{2}$) ale jest to proste rozwiązanie pozwalające stosować standartowe $FFT$. Przy dłuższych transformacjach może się opłacić użycie dodatkowych czynników, np. 3.

Dotychczas rozpatrywaliśmy $FFT$ jako operację jednowymiarową. Wielowymiarowe $FFT$ obliczamy transformując najpierw względem pierszej zmiennej, potem drugiej itd. Ideowo daje to podony efekt jak robicie $N = LM$ które użyliśmy w wymiarze 1. Ale wzory są prostsze.

Warto tu dodać że dla dużych $N$ na współczesnych komputerach istotny staje się czas dostępu do pamięci. Mianowicie, $FFT$ ma dość nieregularny dostęp do pamięci i to może zająć wiecej czasu niż obliczenie. Ale w algorytmie $FFT$ mamy nieco swobody w kolejności obliczeń co można użyć do zmniejszenia

kosztu dostępu do pamiąci przy zachowaniu liczby operacji arytmetycznych.

Zauważny że z macierzowego punktu widzenia $FFT$ długości $N = 2^r$ sprowadza się to inloczynu $r$ (czy $r/2$) macierzy unitarnych mających 2 lub 4 elementy w wierszu. A więc zgodnie z naszym wynikiem o błędzie zookrągleń błąd ten nie przekracza $Cr\varepsilon$ gdzie $\varepsilon$ to dokładność arytmetyki maszynowej zaś $C$ jest niezbyt dużą stałą. Oznacza to że z punktu widzenia błędu zaokrąglenia $FFT$ zachowuje się bardzo dobrze, znacznie lepiej niż bepośrednie obliczanie $DFT$ z definicji (gdzie zamiast $Cr$ mielibyśmy $N$).

Przy pomocy transformacji zespolonej można też obliczać transformacje rzeczywiste: sinusową i kosinusową. Jednakże można uzyskać nieco większą szybkość obliczając transformacje rzeczywiste bezpośrednio stosując do wyprowadznia wzorów wzory na sinus i kosinus sumy kątów.

# 3 Reprezentacja funkcji

Dotychczas reprezentowaliśmy funkcje przy pomocy wartości w punktach. Robiliśmy tak zarówno w przypadku równań zwyczajnych jak i w metodach na siatkach. Jednakże, taka reprezentacja oznacza że wartości w innych punktach nie są wyznaczone jednoznacznie, aby je otrzymać potrzebna jest np. interpolacja. Dla równań zwyczajnych nie jest to problemem. Jednakże dla równań cząstkowych sytuacja się komplikuje. W metodach wielosiatkowych chcemy przechodzić między siatkami. Aby efektywnie obsłużyć lokalne osobliwości chcielibyśmy używać metody adaptacyjne. Wymagałoby to niejednostajnych siatek, co dość znacznie komplikuje wzory różnicowe. Sytuacja się upraszcza jeśli reprezentujemy funkcje we wszystkich punktach. W dalszym ciągu wystarczy nam reprezentacja przez elementy skończenie wymiarowej przestrzeni wektorowej. Dokładniej, zakładamy że nasze funkcje należą do nieskończenie wymiarowej przestrzeni wektorowej $W$. W tej przestrzeni wyróżniamy skończenie wymiarową podprzestrzeń $V \subset W$ i w obliczeniach używamy tylko elementów $V$. Jakie przestrzenie pojawiają się w praktyce? Klasycznie $W$ może być przestrzenią funkcji $C^k$, tzn. funkcji mających $k$ ciągłych pochodnych. Dla równań cząstkowych często bardziej naturalne są przestrzenie Sobolewa. W najprostszym przypadku definijemy przestrzeń Sobolowa $H(k)$ jako przestrzeń funkcji mających $k$ słabych (dystrybucyjnych) pochodnych w $L^2$ z iloczynem skalarnym

$$\langle f, g\rangle_{H(k)} = \langle (1-\Delta)^k f, g\rangle.$$

Dla $k = 1$ daje to

$$\langle f, g\rangle_{H(1)} = \langle (1-\Delta) f, g\rangle = \langle f, g\rangle - \sum_{i=1}^{n} \langle \partial_i^2 f, g\rangle$$

$$= \langle f, g\rangle + \sum_{i=1}^{n} \langle \partial_i f, \partial_i g\rangle$$

Dla normy oznacza to że

$$\|f\|^2_{H(1)} = \|f\|^2_{L^2} + \sum_{i=1}^{n} \|\partial_i f\|^2_{L^2}.$$

Ogólniej oznaczając przez $\hat{f}$ transformację Fouriera $f$ mamy

$$\widehat{(1-\Delta)f}(\omega) = (1+|\omega|^2)\hat{f}(\omega).$$

Wzór ten pozwala zdefinować ułamkowe potęgi $(1-\Delta)^s$ jako operatory które mnożą transformację Fouriera przez $(1+|\omega|^2)^2$. To pozwala zdefiniować

$$\langle f, g\rangle_{H(s)} = \langle (1-\Delta)^s f, g\rangle$$

oraz

$$\|f\|_{H(s)} = \|(1-\Delta)^{s/2} f\|_{L^2}.$$

Dość łatwo pokazać że dla $k < s$ elementy przestrzeni $H(s)$ mają wszystkie pochodne dystrybucyjne $\partial^\alpha f$ rzędu do $k$ (tzn. $|\alpha| \leq k$) w $L^2$. Nieco więcej wysiłku pozwala pokazać że są to naprawdę tzw. pochodne mocne, tzn. pochodne można otrzymać jako granice w $L^2$ ilorazów różnicowych.

Klasyczny lemat Sobolewa mówi że dla $s > n/2$ przestrzeń $H(s)$ jest zawarta w przestrzeni funkcji ciągłych zaś wielokrotność normy Sobolewa majoryzuje normę supremum. A więc regularność w sensie przestrzeni Sobolewa tzn. przynależność do $H(s)$ z dużym $s$ implikuje również regularność w klasycznym sensie.

Dla funkcji z nośnikiem zwartym (czy ogólniej szybko malejących w nieskończoności) zwykła regularność implikuje regularność w przestrzeni Sobolewa. A więc zgrubnie zwykła regularność i regularność w sensie przestrzeni Sobolewa dają podobny efekt. Jednakże przestrzenie Sobolewa pozwalają znacznie bardziej precyzyjnie mierzyć regularność. Np. jeśli $g \in H(s)$, $f \in L^2$ i

$$\Delta f = g$$

to $f \in H(s+2)$. Czyli $f$ ma dwie pochodne więcej niż $g$. A więc $f$ ma dokładnie taką regularność jaka jest potrzebna by otrzymać $\Delta f \in H(s)$. Ogólniej mówimy o eliptycznej regularności: podobna własność jest prawdziwa dla operatorów eliptycznych.

Podobny wynik dla normy $L^\infty$ jest fałszywy, czyli nie ma podobne poprawy regularności: istnieje ciągłe $g$ takie że $f$ spełniające

$$\Delta f = g$$

nie ma dwu ciągłych pochodnych (pochodne są w $L^2$, ale nie są lokalnie ograniczone).

W zagadnieniach nieliniowych mogą się przydać przestrzenie Sobolowa $W(s,p)$ modelowane na $L^p$:

$$\|f\|_{W(s,p)} = \|(1-\Delta)^{s/2} f\|_{L^p}.$$

Dla $1 < p < \infty$ teoria operatorów różniczkowych w przestrzeniach $W(s,p)$ jest podobna do teorii w przestrzeni $L^2$, w szczególności jest poprawa regularności rozwiązań. Inną klasyczną rodziną przestrzeni są przestrzenia Hölderowskie. Dla $0 < s < 1$ definiujemy

$$\|f\|_{C(s)} = \|f\|_{L^\infty} + \sup_{x \neq y} \frac{|f(x) - f(y)}{|x - y|}.$$

Dla $s = k + s_0$ definujemy

$$\|f\|_{C(s)} = \|f\|_{C(k)} + \sum_{|\alpha|=k} \|\partial^\alpha f\|_{C(s_0)}.$$

Niekedy trzeba bardziej skomplikowane przestrzenie jak przestrzenie Trieblа-Lizorkina, przestrzenie Orlicza czy przestrzenie Morreya. Mogą się przydać normy mieszane gdzie funkcje traktujemy jako funkcję mniejszej ilości zmiennych o wartościach wektorowych w przestrzeni funkcji od pozostałych zmiennych. Przydatne też są przestrzenie wagowe, np. jeśli $w$ jest wagą to norma wagowej przestrzeni $L^p$ to

$$\|f\|_{L^p(w)} = (\int |f^p(x)w(x)dx)^{1/p} = \|w^{1/p}\|_{L^p}.$$

Dla wielu przestrzeni zachodzi eliptyczna regularność.

Wiele z występujących wyżej przestrzeni funkcyjnych jest zdefiniowane na całym $\mathbb{R}^n$. Większość definicji łatwo uogólnić na rozmaitości różniczkowe, ale nie będzie nam to potrzebne. Natomiast potrzebne są przestrzenie Sobolewa w obszarach. Tu pojawia się zależność of warunków brzegowych. Zerowy warunek brzegowy to po prostu funkcje które można przedłużyć przez 0 na całą przestrzeń, czyli te funkcje z $H(s)$ które są zerem poza $\Omega$. Dualna do niej jest przestrzeń tych funkcji które można przedłużyć do funkcji z $H(s)$ na całej przestrzeni. Warto też wspomnieć że dla $s > 1/2$ obcięcie funkcji z $H(s)$ do podprzestrzeni kowymiaru 1 należy do $H(s-1/2)$, co pozwala rozważać ogólniejsze warunki brzegowe.

Powyżej naszkicowaliśmy możliwe przestrzenie $W$. Teraz popatrzymy na podprzestrzenie skończenie wymiarowe przydatne w obliczeniach. Klasyczną taką przestenią jest przestrzeń wielomianów ustalonego stopnia. Jednakże sprawia ona kłopoty:

- wielomiany nie są elementami $L^p$ na całej przestrzeni, trzeba się ograniczać do zbiorów ograniczonych (czy ogólniej skończonej miary)
- klasyczna baza jednomianów jest źle uwarunkowana, trochę pomagają wielomiany ortogonalne
- wielomiany są z natury globalne i bardzo regularne, źle reprezentują osobliwości

Przy pracy z operatorami różniczkowymi naturalne jest użycie funkcji własnych wybranego operatora czy operatorów. W szczególności eksponenty z urojonym wykładnikiem są funkcjami własnymi wszystkich pochodnych. Prowadzi to naturalnie do reprezentacji szeregami Fouriera czy podobnymi szeregami funkcji własnych. Tu funkcje bazowe są zwykle ortogonalne i te reprezentacje są dość dobrze uwarunkowane. Jednakże, funkcje własne podobnie jak wielomiany łączą zachowanie lokalne z globalnym co często jest nieporządane. Dodatkowo, dla ciekawszych operatorów funkcje własne mogą być trudne do obliczenia.

## 3.1 Funkcje sklejane

Bardzo pożyteczną klasą funkcji są funkcje sklejane. Ogólne, dziedzina $\Omega$ jest podzielna na podzbiory $\Omega_j$ i na kazdym podzbiorze zadajemy funkcję $f_j$, tzn. dla $x \in \Omega_j$ mamy

$$f(x) = f_j(x).$$

Oczywiście, na przekrojach $\Omega_j$ definicje muszą się zgadzać. Jeśli chcemy lepszą regularność $f$ to pojawiają się dodatkowe warunki. Często jako $f_j$ bierze się wielomiany ustalonego stopnia. Np. interpolacja liniowa na prostej może być traktowana jako funkcja sklejana stopnia 1 (tzn. każde $f_j$ jest stopnia 1). Zauważmy, że jeśli na prostej sklejamy wielomiany stopnia $k$ tak by dostać funkcję klasy $C^k$ na zbiorze spójnym, to ta funkcja jest wielomianem. A więc, by dostać nietrywialy wynik musimy sklejać wielomiany stopnia wyższego niż $k$. Innymi słowy, sklejając funkcje liniowe dostaniemy funkcje ciągłe, lecz nie dostaniemy ciekawych funkcji $C^1$. Sklejając wielomiany kwadratowe można dostać funkcję klasy $C^1$. Np. zadając wartości w dyskretnym zbiorze punktów, zaś w jednym punkcie dodatkowo zadając pierwszą i drugą pochodną otrzymany jednoznacznie wyznaczoną funkcję sklejaną taką że nieciągłości drugiej pochodnej są tylko w zadanych punktach. Ogólniej, z wielomianów stopnia $k+1$ daje się zbudować funkcję $C^k$. Niekiedy warto osłabić warunek na gładkość, tak by dostać większą swobodę co np. może pozwolić na mniejszą normę $H(s)$. Można zauważyć że przestrzeń wielomianów stopnia $l$ ma wymiar $l+1$, zaś warunek $C^k$ w punkcie $x_i$ daje $k+1$ równań na współczynniki. Stąd można pokazać że przestrzeń takich funkcji sklejanych ma wymiar $m(l+1)-(m-1)(k+1) = (m-1)(l-k)+l+1$.

W przypadku wielowymiarowym sytuacja się komplikuje. Jak mówiliśmy mamy zbiory spójne bardziej skomplikowane niż przedziały. Zwykle zakłada się że dziedzina funkcji jest podzielona na w miarę proste figury geometryczne jak sympleksy czy kostki (na płaszczyźnie trójkąty i kwadraty). Jak poprzednio chcemy by funkcja na każdym podzbiorze podziału była wielomianem stopnia $l$ i $C^k$ w całości. Dla $l = 1$, $k = 0$ i podziału na sympleksy dalej sytuacja jest prosta: wielomian stopnia 1 jest jednoznacznie wyznaczony przez wartości na wierzchołkach sympleksu i dla zadanych wartości istnieje dokładnie jeden wielomian stopnia 1 z tymi wartościami w wierzchołkach. W efekcie, funkcja jest jednoznacznie zadana przez wartości w wierzchołkach sympleksów i tak otrzymana funkcja automatycznie jest ciągła (bo wartości na wspólnych brzegach się zgadzają). W podobny sprób można pokazać że wielomian stopnia 2 na trój-

kącie jest jednoznacznie wyznczony przez wartości w wierzchołkach trókąta i w środkach boków. A więc zadając wartości w wierzchołkach i środkach boków otrzymamy ciągłą funkcję sklejaną. Jednakże, gdy chcemy uzyskać funkcje klasy $C^1$ to przy podziale na sympleksy wielomiany stopnia 1 czy 2 nie wystarczą by uzyskać zadane wartości w wierzchołkach.

Na płaszczyźnie przy podziale na trójkąty prawie wystarczą wielomiany stopnia 3. Mianowicie można dowolnie zadać wartości i pierwsze pochodne w wierzchołkach. To spowoduje że wartości na brzegowych odcinkach będą się zgadzać. Zgodność wartości na brzegowym odcinku oznacza że pochodna w kierunku stycznym to odcinka będzie się zgadzać. By uzyskać $C^1$ potrzebowalibyśmy jeszcze zgodność pochodnych w kierynku normalnym. Jednakże mamy tylko jeden dodatkowy parametr do dyspozycji, co ogólnie nie wystarcza. Jednakże, wielomian stopnia 3 na trójkącie jest jednoznacznie wyznaczony przez wartości i pierwsze pochodne w wierzchołkach oraz wartość w srodku trójkąta. Daje to dogodną parametryzaję przestrzeni funkcji sklejanych z wielomianów stopnia 3.

Używając prostokąty można tworzyć funkcje w sposób produktowy. Biorąc po składowych wielomiany stopnia 3 dostaniemy w efekcie wielomian stopnia 6. Z takich kawałków można sklejać funkcję $C^1$ z zadanymi wartościami w wierzchołkach.

Powyżej mówiliśmy o kilku najprostszych przypadkach, w literaturze jest ich trochę więcej.

## 3.2 Bazy falkowe

W wielu problemach przydatne są dobrze zlokalizowane przestrzennie i regularne funkcje. Są ograniczenia dla istnienia takich funkcji. Np. nie istnieją funkcje analityczne o nośnikach zwartych. W $L^2$ zasada nieoznaczoności daje istotne ograniczenie: im lepiej funkcja jest zlokalizowana przestrzennie, tym mniej jest regularna. Z obliczeniowego punktu widzenia jest wygodne gdy nasze funkcje tworzą bazę ortogonalną. W najprostszym przypadku funkcja i jej przesuniecia dałyby bazę ortogonalną. Ale taki warunek jest zbyt kłopotliwy. Jeśli funkcja jest dobrze zlokalizowana i regularna to naturalne jest użycie dylatacji funkcji. Prowadzi to do pojęcia falki podstawowej i bazy falkowej. Dokładniej, powiemy że $\phi$ jest falką podstawową jeśli zestaw funkcji $\phi_{j,k}$ zadany wzorem

$$\phi_{j,k}(x) = 2^{-j/2}\phi(2^{-j}x - k)$$

jest bazą ortogonalną w $L^2$. Istnienie bazy falkowej nie jest sprawą oczywistą. Prostym przykładem jest baza Haara: $\phi(x) = -1$ dla $x \in (0, 1/2)$, $\phi(x) = 1$ dla $x \in (1/2, 1)$ i $\phi(x) = 0$ poza tym. Łatwo zobaczyć że przesunięcia $\phi$ są ortogonalne. Całka z dowolnej dylatacji $\phi$ to zero. Dla $j < 0$ nośnik $\phi_{j,k}$ jest odcinkiem o końcach będących całkowitymi wielokrotnościami $2^j$ i ma długość $2^j$, a więc $\phi$ jest stałe na nośniku $\phi_{j,k}$, czyli

$$(\phi, \phi_{j,k}) = c \int \phi_{j,k} = 0$$

a więc $\phi_{j,k}$ jest ortogonalne do $\phi$. To już pozwala pokazać że układ $\phi_{j,k}$ jest ortogonalny. Dość łatwo można pokazać że funkcja o nośniku w $(0,1)$, mająca całkę 0 i stała na odcinkach postaci $(k2^{-l},(k+1)2^{-l})$ jest kombinacją liniową $\phi_{j,k}$ z $l \leq j \leq 0$. Stąd wynika że układ $\phi_{j,k}$ jest zupełny. Trochę inny przykład można otrzymać następująco. Niech $I = (-2\pi,\pi) \cup (\pi,2\pi)$. Niech $\psi(\omega) = 1$ dla $\omega \in I$ i zero poza tym. $\phi$ definujemy wzorem

$$\hat{\phi} = \psi.$$

Ortogonalność i zupełność wystarczy pokazać po stronie transformacji Fouriera. Po stronie transformacji Fouriera przesunięcia całkowitoliczbowe przechodzą na mnożenie przez $\exp(ik\omega)$. Dylatacje pozostają dylatacjami. Jako że nośnik $\psi$ to $I$ to nośniki dylatacji są rozłączne i pokrywają z dokładnością do zbioru miary zero całą prostą. A więc wystarczy pokazać że

$$\exp(ik\omega)\psi$$

tworzą bazę ortogonalną w $L^2(I)$. Ale to jest standartowe twierdzenie o szeregach Fouriera.

Oba nasze przykłady są dość nieregularne. By dostać bardziej regularny przykład potrzeba więcej wysiłku. Dobra konstrukcja bazuje na filtrach zwierciadlanych (ang. quadrature mirror filter). Niech $h_k$ będzie ciągiem skończonym liczb rzeczywistych. Definujemy $g_k = (-1)^k h_{1-k}$ i

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum h_k \exp(-ik\omega)$$

$$m_1(\omega) = \frac{1}{\sqrt{2}} \sum g_k \exp(-ik\omega)$$

Powiemy że $h_k$ zadają filtr zwierciadlany jeśli $m_0(0) = 1$ i

$$|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1$$

Zauważmy że z określenia $m_1$ wynika że

$$m_1(\omega) = -\exp(-i\omega)\bar{m}_0(\omega+\pi)$$

czyli też

$$|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1.$$

Dalej będziemy zakładać że dla $|\omega| < \pi$ mamy $m_0(\omega) \neq 0$.

Falkę podstawiową $\phi$ i funkcję skalującą $\psi$ definiujemy wzorami

$$\hat{\phi}(\omega) = m_1(2^{-1}\omega) \prod_{j=2}^{\infty} m_0(2^{-j}\omega),$$

$$\hat{\psi}(\omega) = \prod_{j=1}^{\infty} m_0(2^{-j}\omega).$$

Z założenia $m_0(0) = 1$ i $m_0$ jest gładkie, więc

$$m_0(2^{-j}\omega) = 1 + O(|2^{-j}\omega|)$$

a więc nieskończone produkty wyżej są zbieżne. Jako że $|m_0(\omega)| \leq 1$ i $|m_1(\omega)| \leq 1$ to

$$|\hat{\phi}(\omega)| \leq 1$$

a więc mnożenie transformacji Fouriera przez $\hat{\phi}$ daje operator ograniczony na $L^2$. Oznaczmy

$$p_k(\omega) = \prod_{j=1}^{k} m_0(2^{-j}\omega).$$

$p_k$ ma okres $2^{k+1}\pi$ i dla $k > 1$ spełnia

$$p_k(\omega) = m_0(2^{-k}\omega)p_{k-1}(\omega).$$

A więc

$$\int_{-2^k\pi}^{2^k\pi} |p_k(\omega)|^2 d\omega = \int_0^{2^{k+1}\pi} |p_k(\omega)|^2 d\omega$$

$$= \int_0^{2^k\pi} |p_k(\omega)|^2 d\omega + \int_{2^k\pi}^{2^{k+1}\pi} |p_k(\omega)|^2 d\omega$$

$$= \int_0^{2^k\pi} |p_k(\omega)|^2 d\omega + \int_0^{2^k\pi} |p_k(\omega + 2^k\pi)|^2 d\omega$$

$$= \int_0^{2^k\pi} |m_0(2^{-k}\omega)|^2 |p_{k-1}(\omega)|^2 d\omega + \int_0^{2^k\pi} |m_0(2^{-k}\omega + \pi)|^2 |p_{k-1}(\omega + 2^k\pi)|^2 d\omega$$

$$= \int_0^{2^k\pi} (|m_0(2^{-k}\omega)|^2 + |m_0(2^{-k}\omega + \pi)|^2) |p_{k-1}(\omega)|^2 d\omega$$

$$= \int_0^{2^k\pi} |p_{k-1}(\omega)|^2 d\omega.$$

Indukcyjnie

$$\int_{-2^k\pi}^{2^k\pi} |p_k(\omega)|^2 d\omega = \int_{-2\pi}^{2\pi} |p_1(\omega)|^2 d\omega = \int_{-2\pi}^{2\pi} |m_0(2^{-1}\omega)|^2 d\omega$$

$$= 2\int_{-\pi}^{\pi} |m_0(\omega)|^2 d\omega = 2\int_0^{2\pi} |m_0(\omega)|^2 d\omega$$

$$= 2\int_0^{\pi} |m_0(\omega)|^2 + |m_0(\omega + \pi)|^2 d\omega = 2\pi$$

Wynika stąd że

$$\int_{-2^k\pi}^{2^k\pi} |\hat{\psi}(\omega)|^2 d\omega \leq 2\pi$$

czyli też

$$\int |\hat{\psi}(\omega)|^2 d\omega \leq 2\pi$$

co oznacza że $\psi$ i $\phi$ są w $L^2$. Nasze założenia o $m_0$ implikują że dla $|\omega| \leq \pi$ mamy $\hat{\psi}(\omega) \neq 0$, czyli na przedziale $[-\pi, \pi]$ moduł $\hat{\psi}(\omega)$ jest ograniczony z dołu. Definujemy $q_k(\omega) = p_k(\omega)$ dla $\omega \in [-2^k\pi, 2^k\pi]$ i zero poza tym. Mamy wzór

$$\hat{\psi}(\omega) = p_k(\omega)\hat{\psi}(2^k\omega)$$

co razem z oszacowaniem z dołu na $\hat{\psi}(\omega)$ oznacza że $|q_k(\omega)|$ jest ograniczone przez wielokrotność $|\hat{\psi}(\omega)|$. A więc na mocy twierdzenia Lebesque'a o zbieżności ograniczonej

$$\|\hat{\psi} - q_k\|_{L^2} \to 0$$

czyli

$$\|\hat{\psi}\|_{L^2} = 2\pi.$$

Jak wcześniej definujemy

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k).$$

Mamy

$$\hat{\psi}_{0,k}(\omega) = \exp(-ik\omega)\hat{\psi}.$$

Podobny rachunek jak wyżej pokazuje że

$$(\hat{\psi}, \exp(-ik\omega)\hat{\psi}) = \int_{-2\pi}^{2\pi} \exp(ik\omega)|m_0(2^{-1}\omega)|^2 d\omega$$

$$= \int_0^{4\pi} \exp(ik\omega)|m_0(2^{-1}\omega)|^2 d\omega$$

$$= \int_0^{2\pi} \exp(ik\omega)(|m_0(2^{-1}\omega)|^2 + |m_0(2^{-1}\omega + \pi)|^2) d\omega$$

$$= \int_0^{2\pi} \exp(ik\omega) d\omega = 0$$

dla $k \neq 0$. A więc $\psi_{0,k}$ tworzą układ ortogonalny.

Rozważając produkt dla $\hat{\phi}$ pokazujemy że

$$\|\hat{\phi}\|_{L^2}^2 = \int_{-2\pi}^{2\pi} |m_1(2^{-1}\omega)|^2 = 2\pi$$

i podobnie jak dla $\psi_{0,k}$ pokazujemy że $\phi_{0,k}$ tworzą układ ortogonalny.

Następnie

$$(\hat{\psi}, \exp(-ik\omega)\hat{\phi}) = \int_{-2\pi}^{2\pi} \exp(ik\omega) m_0(2^{-1}\omega)\bar{m}_1(2^{-1}\omega) d\omega$$

$$= 2\int_{-\pi}^{\pi} \exp(i2k\omega) m_0(\omega)\bar{m}_1(\omega)d\omega$$

$$= 2\int_{0}^{2\pi} \exp(i2k\omega) m_0(\omega)\bar{m}_1(\omega)d\omega$$

$$= -2\int_{0}^{2\pi} \exp(i2k\omega) m_0(\omega)\exp(i\omega) m_0(\omega+\pi)d\omega.$$

Jednakże

$$\int_{\pi}^{2\pi} \exp(i2k\omega) m_0(\omega)\exp(i\omega) m_0(\omega+\pi)d\omega$$

$$= \int_{0}^{\pi} \exp(i2k\omega) m_0(\omega+\pi)\exp(i(\omega+\pi)) m_0(\omega+2\pi)d\omega$$

$$= -\int_{0}^{\pi} \exp(i2k\omega) m_0(\omega)\exp(i\omega) m_0(\omega+\pi)d\omega$$

Toteż

$$-2\int_{0}^{2\pi} \exp(i2k\omega) m_0(\omega)\exp(i\omega) m_0(\omega+\pi)d\omega = 0$$

czyli $\phi_{0,k}$ są ortogonalne do $\psi$. Oznaczmy przez $V_j$ przestrzeń rozpiętą przez $\psi_{j,k}$ zaś przez $W_j$ przestrzeń rozpiętą przez $\phi_{j,k}$. Mamy

$$\hat{\psi}_{-1,k}(\omega) = 2^{-1/2}\exp(-ik\omega/2)\hat{\psi}(2^{-1}\omega) = 2^{-1/2}\exp(-ik\omega/2)\prod_{j=2}^{\infty} m_0(2^{-j}\omega).$$

Dla $f = \sum_k a_k\psi_{-1,k}$ mamy więc

$$\hat{f}(\omega) = 2^{-1/2}\psi(2^{-1}\omega)\sum_k a_k\exp(-ik\omega/2) = \hat{a}(2^{-1}\omega)\psi(2^{-1}\omega).$$

Jako że $m_0(2^{-1}\omega)$ i $m_1(2^{-1}\omega)$ są kombinacjami liniowymi $\exp(-ik\omega/2)$ to $\phi \in V_{-1}$ and $\psi \in V_{-1}$. Oznacza to że $V_0 \subset V_{-1}$, czyli ogólniej $V_j \subset V_l$ dla $j \geq l$. Mamy też $\phi_{j,k} \in V_{j-1}$, co oznacza że dla $j > 1$ $\phi_{j,k}$ jest ortogonalne do $\phi_{0,k}$. Stosując dylatacje widzimy teraz że $\phi_{j,k}$ tworzą układ ortogonalny. Chcielibyśmy jeszcze pokazać zupełność.

Dla $f_1, f_2 \in V_{-1}$ mamy też

$$(\hat{f}_1, \hat{f}_2) = \int \hat{a}_1(2^{-1}\omega)\overline{\hat{a}_2}(2^{-1}\omega)|\psi(2^{-1}\omega)|^2 d\omega$$

$$= \int_{-4\pi}^{4\pi} \hat{a}_1(2^{-1}\omega)\overline{\hat{a}_2}(2^{-1}\omega)|m_0(2^{-2}\omega)|^2 d\omega$$

$$= \int_{-2\pi}^{2\pi} \hat{a}_1(2^{-1}\omega)\overline{\hat{a}_2}(2^{-1}\omega)d\omega.$$

Wynika stąd że przekształcenie $f \mapsto \hat{a}$ utożsamia $V_{-1}$ z $L^2([-2\pi, 2\pi])$. Przy tym utożsamieniu $V_0$ przechodzi na funkcje postaci $m_0(2^{-1}\omega)\hat{a}(\omega)$ z $\hat{a}$ o okresie

$2\pi$. Podobnie $W_0$ przechodzi na funkcje postaci $m_1(2^{-1}\omega)\hat{a}(\omega)$ z $\hat{a}$ o okresie $2\pi$. Niech $P$ oznacza rzut ortogonalny z $V_1$ na $V_0$ zaś $Q$ rzut na $W_0$. W terminach $\hat{a}$ można wyrazić $P$ następująco: mnożymy $\hat{a}$ przez $\bar{m}_0(2^{-1}\omega)$, potem dodajemy wartości w $\omega$ i $\omega + 2\pi$ (co daje funkcję o okresie $2\pi$), potem mnożymy przez $m_0(2^{-1}\omega)$. Podobnie zapisuje się $Q$, tyle że używamy $m_1$. Obliczmy teraz $P+Q$. Jeśli $\hat{a}(\omega) = 1$, $\hat{a}(\omega + 2\pi) = 0$, to w $\omega$ dostaniemy

$$|m_0(2^{-1}\omega)|^2 + |m_1(2^{-1}\omega)|^2 = 1,$$

zaś w $\omega + 2\pi$ dostaniemy

$$m_0(2^{-1}\omega + \pi)\bar{m}_0(2^{-1}\omega) + m_1(2^{-1}\omega + \pi)\bar{m}_1(2^{-1}\omega)$$

$$= m_0(2^{-1}\omega+\pi)\bar{m}_0(2^{-1}\omega)+\exp(-i(\omega/2+\pi))\bar{m}_0(2^{-1}\omega+\pi+\pi)\exp(i\omega/2)m_0(2^{-1}\omega+\pi)$$

$$= m_0(2^{-1}\omega + \pi)\bar{m}_0(2^{-1}\omega) - \bar{m}_0(2^{-1}\omega)m_0(2^{-1}\omega + \pi) = 0.$$

A więc $P + Q = I$, czyli $V_{-1} = V_0 \oplus W_0$. Teraz widać że dla $m \leq l$ mamy

$$V_{m-1} = W_m \oplus W_{m+1} \oplus \cdots \oplus W_l \oplus V_l$$

czyli $\phi_{j,k}$ z $m \leq j \leq l$ tworzą bazę ortogonalną dopełnia ortogonalnego $V_l$ w $V_{m-1}$. Można też pokazać że $\bigcup_j V_j$ to całe $L^2$, zaś $\bigcap V_j = \{0\}$ co oznacza że nasz układ jest zupełny

Uzasadniliśmy że $\phi$ jest w $L^2$. Jednakże, można pokazać więcej. Mianowicie, zakładamy że nasz filtr jest rzędu $N$, tzn.

$$m_0(\omega) = (1 + \exp(i\omega))^N f(\omega)$$

gdzie $f$ jest gładkie i $|f(\omega)| \leq B$. Wtedy mamy

$$|(\hat{\phi})(\omega)| \leq C(1 + |\omega|)^{-N+\log(B)/\log(2)}$$

tzn. dla dostatecznie dużego $N - \log(B)/\log(2)$ nasze $\phi$ jest funkcją $C^k$.

Przykładem ciągu spełniającego warunek wyżej z $N = 2$ i $B = \sqrt{3}/4$ jest

$$h_0 = (1 + \sqrt{3})/(4\sqrt{2}),$$

$$h_1 = (3 + \sqrt{3})/(4\sqrt{2}),$$

$$h_2 = (3 - \sqrt{3})/(4\sqrt{2}),$$

$$h_3 = (1 - \sqrt{3})/(4\sqrt{2}).$$

Nasz wzór na $\phi$ i $\psi$ można też zapisać w terminach splotu. Mianowicie definujemy znakowaną miary dyskretne $\mu$ i $\nu$ wzorami

$$\mu = \frac{1}{\sqrt{2}} \sum h_k \delta_k$$

$$\nu = \frac{1}{\sqrt{2}} \sum g_k \delta_k$$

Wtedy $\hat{\mu} = m_0$, $\hat{\nu} = m_1$. Niech $D_t$ oznacza dylatację miary, tzn.

$$D_t\nu = \frac{1}{\sqrt{2}} \sum g_k \delta_{kt}$$

Wtedy

$$\psi = D_{2^{-1}}\mu * D_{2^{-2}}\mu * D_{2^{-3}}\mu * \dots$$

zaś

$$\phi = D_{2^{-1}}\nu * D_{2^{-2}}\mu * D_{2^{-3}}\mu * \dots$$

Stąd wynika że $\phi$ i $\psi$ mają nośniki zwarte. Mamy też

$$D_2\psi = \mu * \psi,$$

$$D_2\phi = \nu * \psi.$$

Dodatkowo, jeśli

$$f = \sum_k a_k \psi_{j,k}$$

to $f$ można rozbić na dwie części: jedną która jest podobną sumą z $j$ powiększonym o 1 i druga gdzie też $j$ jest powiększone o 1 a $\psi_{j+1,k}$ jest zastąpine przez $\phi_{j+1,k}$. Mianowicie, $f \in V_j$. Jak pokazaliśmy $V_j = W_{j+1} \oplus V_{j+1}$. Rzut na $V_{j+1}$ wyliczaliśmy następująco. Mamy

$$\hat{f} = \hat{a}(2^k\omega)\hat{\psi}(2^k\omega).$$

By dostać element $V_{j+1}$ mnożylśmy przez $\bar{m}_0(2^k\omega)$, dodajemy wartości w $\omega$ i $\omega + \pi$ po czym monożyliśmy przez $m_0(2^k\omega)$. Na poziomie współczynników ostatnie mnożenie można pominąć bo użycie $\psi_{j+1,k}$ automatycznie uwzględnia czynnik $m_0(2^k\omega)$. Dodanie wartości w $\omega$ i $\omega+\pi$ na poziomie współczynników to po prostu pominięcie nieparzystych współczynników Fouriera. Mnożenie przez $\bar{m}_0(2^k\omega)$ po stronie przestrzennej oznacza splot, czyli współczyniki $b_j$ rzutu na $V_{k+1}$ dostajemy wzorem

$$b_j = \sum_n h_{n-2j} a_n.$$

Komentarz: Bez sprężenia zespolonego byłoby $h_{2j-n}$ co jest normalnym wzorem na splot, ale wyżej uwzględniliśmy sprzężenie. Ze względu na skalowanie $\psi_{j,k}$ zniknął czynnik $1/\sqrt{2}$.

Podobnie współczynniki $c_j$ rzutu na $W_{k+1}$ są dane wzorem

$$c_j = \sum_n g_{n-2j} a_n.$$

Czyli to rozbicie wyliczamy w czasie liniowym, wykonując operację podobną do splotu. Człony z $\psi$ (czyli $b_j$) przetwarzamy rekursywnie. Jako że ilość członów w sumie do rekursywnego przetwarzania maleje w przybliżeniu do połowy, to cały algorytm jest asymptotycznie liniowy.

Wypadałby by jeszcze powidzieć jak znaleźć inne ciągi $h$. W tym celu zauważmy że wystarczy znaleźć $|m_0|^2$. Mianowicie, $|m_0|^2(\omega)$ jest postaci $\sum c_k \exp(-k\omega)$ z symetrycznymi rzeczywistymi $c_k$, tzn. $c_k = c_{-k}$. Zachodzi lemat

**Lemat 3.1** *Jeśli* $A(\omega) = \sum c_k \exp(-ik\omega)$, $c_k = c_{-k}$, $A(\omega) \geq 0$ *dla rzeczywistych* $\omega$, *i jedynym rzeczywistymi zerami* $A(\omega)$ *są* $\omega$ $z$ $\exp(-i\omega) = -1$, *to istnieją rzeczywiste* $h_k$ *takie że przy* $B(\omega) = \sum_k h_k(\exp(-ik\omega)$ *dla rzeczywistych* $\omega$ *mamy*

$$A(\omega) = |B(\omega)|^2$$

Szkic dowodu: Zamiast $\exp(-i\omega)$ wygodnie jest rozpatrywać zespolone $z$. Rzeczywiste $\omega$) odpowiada $|z| = 1$. Jako $B$ można wziąść wielomian od $z$ (ujemne potęgi elimunujemy mnożać przez potęgę $z$ co nie zmienia modułu). Piszemy

$$B(z) = c\prod_k (z - z_k)$$

Wtedy, uwzględniając że dla $|z| = 1$ mamy $z^{-1} = \bar{z}$

$$A(z) = c^2 \prod_k ((z - z_k)(z^{-1} - \bar{z}_k))$$

czyli aby znaleźć $B$ musimy wybrać niektóre z czynników $A$. Jako że $A$ ma symetryczne współczynniki, to jeśli $z$ jest pierwiatkiem $A$ to również $z^{-1}$ jest pierwiastkiem $A$. A więc by dostać $B$ spełniające $A(z) = |B(z)|^2$ musimy wybrać po jednym czynniku z każdej pary zer $z, z^{-1}$. Jeśli $z \neq z^{-1}$ to jest to możliwe. Potencjanie kłopotliwe jest $z \in \{1, -1\}$. Lecz z założenia jedyne zero $A$ o module 1 to $z = -1$. Ale $A$ ma parzystą iloć zer, więc zero w $-1$ ma parzystą krotoność, więc można dodać do $B$ czynnik $z + 1$ z krotnością podzieloną przez 2.

Ale trzeba też zagwarantować że $B$ ma współczynniki rzeczywiste, tzn. wraz każdym zerem dodać też jego sprzężenie zespolone. Jako że $A$ ma rzeczywste współczynniki to jeśli $z$ jest pierwiatkiem $A$ to również $\bar{z}$ jest pierwiastkiem $A$. Czyli jeśli $|z| \neq 1$ to zera $A$ są w czwórkach i wybieramy do $B$ parę zer sprzężonych z czwórki. Potencjalny kłopot to zera z modułem równym 1, lecz z założenia jedynym takim zerem jest $z = -1$, które jest rzeczywiste więc nie sprawia problemu. Założenie $A(z) \geq 0$ dla $|z| = 1$ oznacza że czynnik stały ma pierwiastek rzeczywisty. □

A więc pozostaje podać metodę konstrukcji wielomianów trygonometrycznych które są symetryczne, spełniają

$$A(\omega) + A(\omega + \pi) = 1$$

i są podzielne przez $(1 + \exp(i\omega))^{2N}$. Y. Meyer podał wzór

$$A(\omega) = 1 - c\int_0^{\omega} \sin(t)^{2N-1} dt$$

gdzie

$$c = \frac{(2N-1)!}{((N-1)!)^2 2^{2N-1}}.$$

Oczywiście $A$ jest wielomianem trygonometrycznym stopnia $2N$. Jako że sinus jest nieparzysty, to po scałkowaniu dostajemy funkcję parzystą. Zauważmy że $\sin(t+\pi) = -\sin(t)$, czyli

$$\int_0^{\omega} \sin(t)^{2N-1} dt = -\int_{\pi}^{\pi_{\omega}} \sin(t)^{2N-1} dt.$$

A więc

$$A(\omega) + A(\omega+\pi) = 2 - c\left(\int_0^{\omega} \sin(t)^{2N-1} dt + \int_0^{\omega+\pi} \sin(t)^{2N-1} dt\right)$$

$$= 2 - c\left(\int_0^{\omega} \sin(t)^{2N-1} dt + \int_{\pi}^{\pi+\omega} \sin(t)^{2N-1} dt + \int_0^{\pi} \sin(t)^{2N-1} dt\right)$$

$$= 2 - c\int_0^{\pi} \sin(t)^{2N-1} dt = 1.$$

$c$ wyżej jest dobrane właśnie tak by mieć ostatnią równość. Widać też że $A(\omega) \geq 0$ i dla $\omega = \pi$ mamy zero rzędu $2N$:

$$A(\omega+\pi) = -c\int_{\pi}^{\pi+\omega} \sin(t)^{2N-1} dt = c\int_0^{\omega} \sin(t)^{2N-1} dt$$

co ma zero rzędu $2N$ bo $\sin(t)^{2N-1}$ ma zero rzędu $2N-1$ a całkowanie powiększa rząd zera o 1.