DESIGN OF FIR FILTERS WITH CONSTRAINTS IN THE TIME OR FREQUENCY DOMAIN

- Up to now, we have concentrated on designing FIR filters to meet the given amplitude criteria in some sense. However, there exist applications where there are constraints in the time domain or in the frequency domain.
- For example, in some applications the transient part of the step response must be constrained to vary within the given limits
- Another example is the design of Nyquist filters or M-th band filters with every M-th impulse response value being zero except for the central value.
- Furthermore, in some cases, there are flatness constraints in the passband response of the filter.

How to solve these problems ?

- In some cases, the desired solution can be obtained by properly modifying the design methods proposed previously.
- In the remaining cases, new techniques are required.
- Perhaps the most flexible design method for finding the optimum solution to various constrained approximation problems is linear programming.
- The advantage of this technique is that the convergence to the optimum solution is guaranteed.
- With linear programming, it is also possible to find the optimum solution to the unconstrained minimax approximation problems considered previously.
- The disadvantage is, however, that the required computation to arrive at the desired solution is rather large.
- Therefore, it is preferred to use linear programming only in those cases which cannot be handled with other faster design techniques.

Linear Programming Approach for FIR Filter Design

- Linear programming is a very flexible approach for solving many constrained approximation problems in the minimax sense.
- Mathematically, the linear programming problem can be stated in the form of the following primal problem: Find the unkowns x_k, $k = 1, 2, \ldots, N$, subject to the constraints

$$x_k \geq 0, \quad k = 1, 2, \ldots, N$$

$$\sum_{k=1}^{N} \gamma_{lk} x_k = \beta_l, \quad l = 1, 2, \ldots, M \quad (M < N)$$

such that

$$\sigma = \sum_{k=1}^{N} \alpha_k x_k$$

is minimized.

- In this problem, γ_{lk}, α_k, and β_l are constants.

Dual Problem

- The above problem is mathematically equivalent to the following dual problem: Find the unknowns y_l, $l = 1, 2, \ldots, M$, subject to the constraints

$$\sum_{l=1}^{M} \gamma_{lk} y_l \leq \alpha_k, \quad k = 1, 2, \ldots, N$$

such that

$$\rho = \sum_{l=1}^{M} \beta_l y_l$$

is maximized.

- For digital filter design problems, the dual problem is the most natural form. There exist several well-defined procedures arriving at the desired solution within $M + N$ iterations. Lim has introduced an efficient special purpose algorithm for designing FIR filters. This is faster than general purpose algorithms.

When to apply linear programming

- Linear programming can be applied in a straightforward manner to those problems where the approximating function is linear, i.e. it can be expressed in the form

$$H(\omega) = \sum_{n=0}^{R} b[n]\Phi(\omega, n),$$

where the $b[n]$'s are unknowns.

- According to previous discussions the zero-phase frequency response of a linear-phase FIR filter can be expressed in all the four cases in the above form (see page 50 in the beginning of the pile of lecture notes for FIR filter synthesis).

- Also, in many other cases, the approximating function can be written in the above form. For instance, in the conventional frequency-sampling methods the approximating function is expressible in this form.

A General Constrained Frequency-Domain Approximation Problem

A general constrained frequency-domain approximation problem, which can be solved using linear programming, can be stated in the form: Find the unknowns $b[n]$ to minimize

$$\delta_1 = \max_{\omega \in X_1} |E(\omega)|,$$

where

$$E(\omega) = W(\omega)[H(\omega) - D(\omega)]$$

subject to

$$\max_{\omega \in X_2} |E(\omega)| \leq \delta_2.$$

Here, X_1 contains a part of the passband and stopband regions and X_2 contains the remaining part. For instance, by selecting X_1 and X_2 to be the stopband and passband regions of the filter, respectively, the stopband variation can be minimized for the given maximum allowable passband variation.

- Problems of this kind cannot be solved directly using the MPR algorithm.

- By sampling $W(\omega)$ and $D(\omega)$ along a dense grid of frequencies $\omega_1^{(1)}, \omega_2^{(1)}, \ldots, \omega_{K_1}^{(1)}$ on X_1 and along a grid of frequencies $\omega_1^{(2)}, \omega_2^{(2)}, \ldots, \omega_{K_2}^{(2)}$ on X_2, the problem can be stated in the form of the dual problem as follows: Find $b[0], b[1], \ldots, b[R]$, and δ_1 subject to the constraints

$$\sum_{n=0}^{R} b[n]\Phi(\omega_j^{(1)}, n) - \delta_1/W(\omega_j^{(1)}) \leq D(\omega_j^{(1)}), j = 1, \ldots, K_1 \tag{1a}$$

$$-\sum_{n=0}^{R} b[n]\Phi(\omega_j^{(1)}, n) - \delta_1/W(\omega_j^{(1)}) \leq -D(\omega_j^{(1)}), j = 1, \ldots, K_1 \tag{1b}$$

$$\sum_{n=0}^{R} b[n]\Phi(\omega_j^{(2)}, n) \leq D(\omega_j^{(2)}) + \delta_2/W(\omega_j^{(2)}), 1, \ldots, K_2 \tag{1c}$$

$$-\sum_{n=0}^{R} b[n]\Phi(\omega_j^{(2)}, n) \leq -D(\omega_j^{(2)}) + \delta_2/W(\omega_j^{(2)}), j = 1, \ldots, K_2 \tag{1d}$$

such that

$$\rho = -\delta_1 \tag{1e}$$

is maximized.

Comments

- Note that in the dual problem the constraints are formed in such a way that a linear combination of the unknowns is less than or equal to a constant.
- In the above problem, δ_2 is a constant and δ_1 is an unknown.
- This explains the difference between the equations.
- The above equations have been constructed such that, after finding the optimum solution with the minimum δ_1, $-\delta_2 \leq E(\omega_j^{(2)}) \leq \delta_2$ and $-\delta_1 \leq E(\omega_j^{(1)}) \leq \delta_1$ at the selected grid points.
- Note also that in the dual problem a linear combination of unknowns is maximized and maximizing $-\delta_1$ implies minimizing δ_1.

Other constraints

- It is easily include in the dual problem various constraints which are expressible in the desired form.
- For instance, it is straightforward to add constraints of the form

$$\frac{d^l H(\omega)}{d^l \omega}\Big|_{\omega=\omega_j} = \sum_{n=0}^{R} b[n] \frac{d^l \Phi(\omega, n)}{d^l \omega}\Big|_{\omega=\omega_j} \leq 0 \qquad (2a)$$

or

$$\frac{d^l H(\omega)}{d^l \omega}\Big|_{\omega=\omega_j} = \sum_{n=0}^{R} b[n] \frac{d^l \Phi(\omega, n)}{d^l \omega}\Big|_{\omega=\omega_j} \geq 0, \qquad (2b)$$

where l is an integer and ω_j is a grid point.

- The constraint expressed by Eq. (2a) is directly in the desired form. The constraint of Eq. (2b) can be written in the desired form by multiplying the left-hand side by -1 and replacing $\geq$ by $\leq$.

- By adding a constraint of the form of Eq. (2a) with $l = 1$ at all the grid points in the passband region, the passband response of the filter can be forced to be monotonically decreasing. Steiglitz has presented a FORTRAN code for designing filters of this kind.

- Furthermore, the first L derivatives of $H(\omega)$ can be forced to be zero at $\omega = \omega_j$ by simultaneously adding the constraints of Eqs. (1a) and (1b) for $l = 1, 2, \ldots, L$.

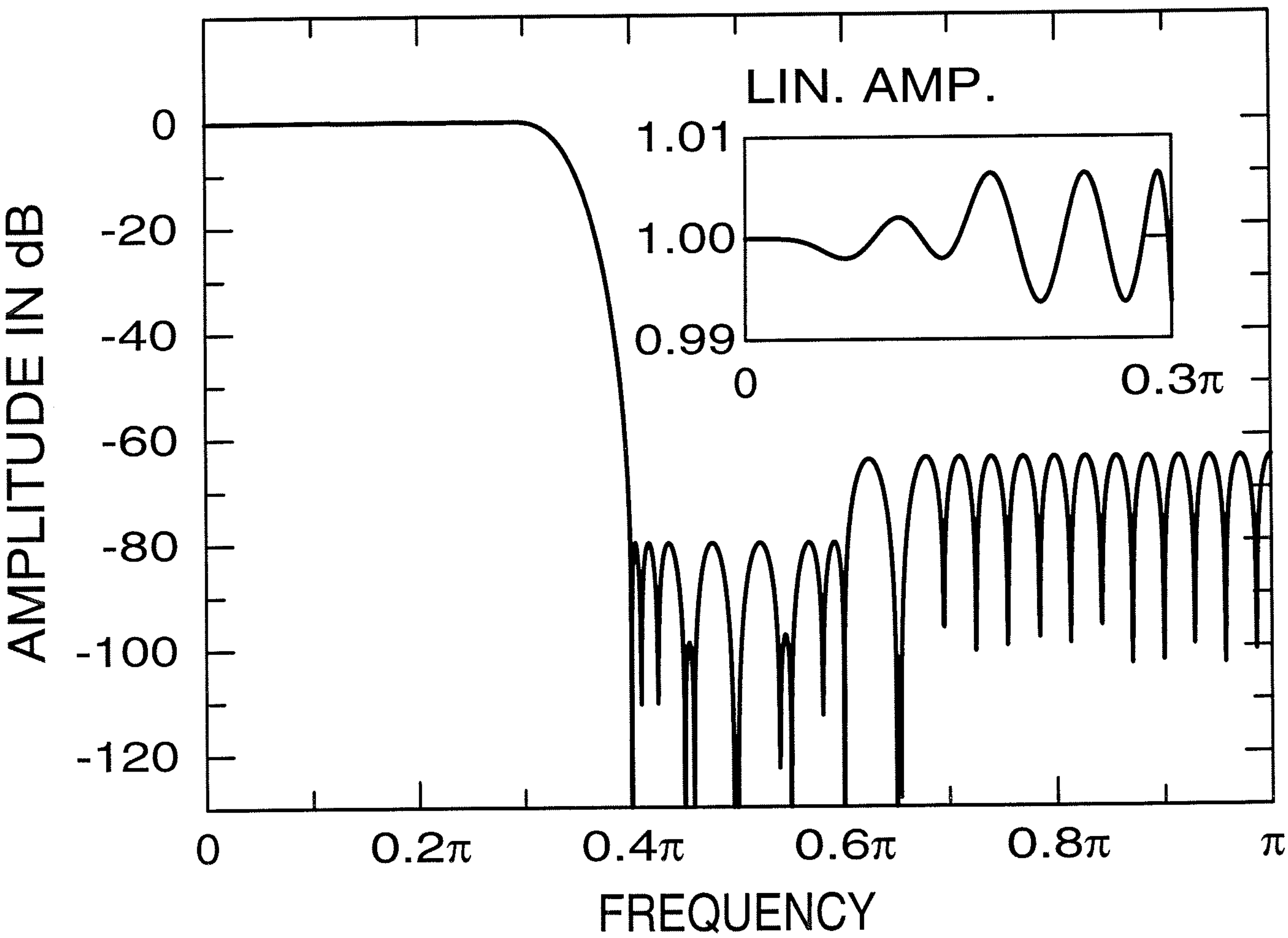
- In addition, if it is desired that $H(\omega)$ achieve exactly the value A at $\omega = \omega_j$, this condition can be included by using the following two constraints:

$$\sum_{n=0}^{R} b[n]\Phi(\omega_j, n) \le A, \quad -\sum_{n=0}^{R} b[n]\Phi(\omega_j, n) \le -A. \quad (3)$$

Illustrative Example 1

- Consider the design of a Type I linear-phase filter of order 70 having $[0, 0.3\pi]$ and $[0.4\pi, \pi]$ as the passband and stopband regions, respectively.
- The filter has fixed zero pairs at the angular frequencies $\pm 0.4\pi$, $\pm 0.45\pi$, $\pm 0.5\pi$, $\pm 0.55\pi$, $\pm 0.6\pi$, and $\pm 0.65\pi$, and $H(\omega)$ achieves the value of unity at $\omega = 0$ with its first four derivatives being zero at this point.
- The maximum deviation from unity on $[0, 0.15\pi]$ is 0.002 and the maximum deviation from zero on $[0.4\pi, 0.6\pi]$ is 0.0001, whereas the response is desired to be optimized in the remaining regions with weighting of unity on $[0.15\pi, 0.3\pi]$ and 10 on $[0.6\pi, \pi]$.
- The last part of this problem can be expressed in the form of Eq. (1) using $X_1 = [0.15\pi, 0.3\pi] \cup [0.6\pi, \pi]$ and $X_2 = [0, 0.15\pi] \cup [0.6\pi, \pi]$. $D(\omega)$ is 1 on $[0, 0.3\pi]$ and 0 on $[0.4\pi, \pi]$. $W(\omega)$ is 1 on $[0, 0.3\pi]$, 20 on $[0.4\pi, 0.6\pi]$, and 10 on $[0.6\pi, \pi]$, whereas $\delta_2 = 0.002$.

- To include the first part, Eq. (2) is used with $A = 0$ at the frequency points where the filter has fixed zeros and with $A = 1$ at the zero frequency.
- Equations (2a) and (2b) are used with $l = 1, 2, 3, 4$ at the zero frequency.
- The optimized filter response is shown below. The resulting ripple values on $[0.15\pi, 0.3\pi]$ and $[0.6\pi, \pi]$ are 0.00637 and 0.000637, respectively.



Other Contraints

- It is also easy to include time-domain constraints in the approximation problem. For instance, some of the unknowns $b[n]$, $n \in S$ can be fixed and the remaining ones can be optimized. In this case, the desired solution can be found by using the following approximating function

$$H(\omega) = \sum_{\substack{n=0 \\ n \notin S}}^{R} b[n]\Phi(\omega, n),$$

and by including the effect of the fixed terms in the desired function by changing it to be

$$\bar{D}(\omega) = D(\omega) - \sum_{n \in S}^{R} b[n]\Phi(\omega, n).$$

Example 2

- This example shows how linear programming can be used for designing filters with constraints on the step response, which is related to the impulse-response coefficients $h[n]$ through

$$g[n] = \sum_{m=0}^{n} h[m].$$

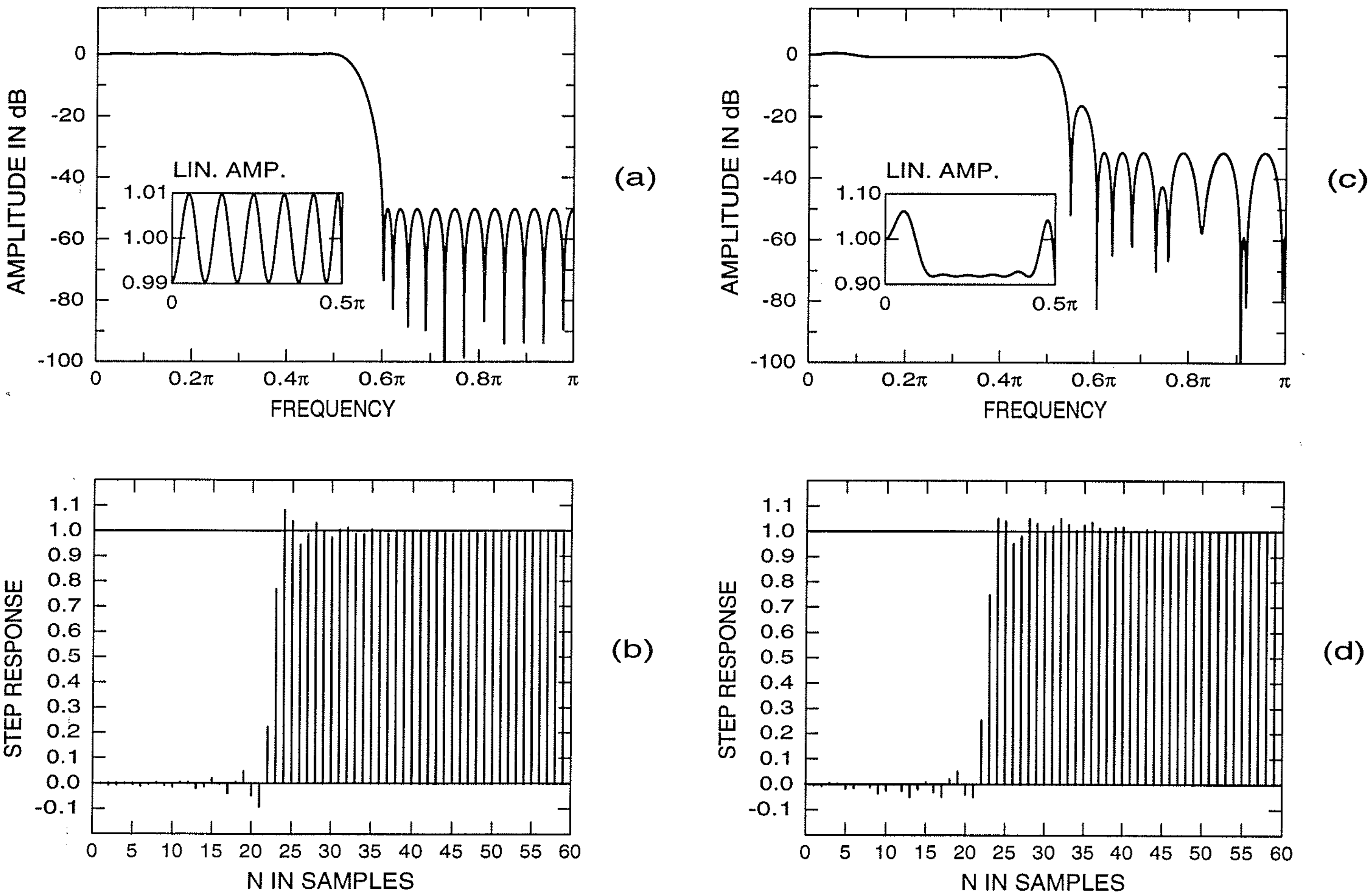
- In the following transparency, the figures on the left give the amplitude and step responses for a filter of order 46 optimized without any constraints in the time domain.
- For this filter, the passband and stopband ripples are related via $\delta_p = \sqrt{10}\delta_s$ and the maximum undershoot of the step response occuring at $n = 22$ is -0.0921 and $g[n] = 0.9903$ for $n \geq 46$.
- It is desired that $g[n] = 1$ for $n \geq 46$ and

$$-\delta_{\text{step}} \leq g[n] \leq \delta_{\text{step}} \qquad \text{for} \quad 0 \leq n \leq K,$$

where $K = 22$ and $\delta_{\text{step}} = 0.05$.

- The first condition can be satisfied by requiring that $H(0) = 1$.

Filter Responses



- The second constraint is linear in the $h[n]$'s and can thus be easily included in the dual problem.

- The second constraint is linear in the $h[n]$'s and can thus be easily included in the dual problem. Because of this condition, it is advantageous to express $H(\omega)$ directly in terms of the $h[n]$'s as ($M = N/2$ for Type I designs)

$$H(\omega) = h[M] + \sum_{n=1}^{M} h[M-n](2\cos n\omega),$$

so that $\Phi(\omega, M) = 1$ and $\Phi(\omega, M-n) = 2\cos n\omega$, $n > 0$.

- The amplitude and step responses for the filter optimized with the above constraints are shown in the figures on the right in the previous transparency.

Design of L-th Band (Nyquist) Filters

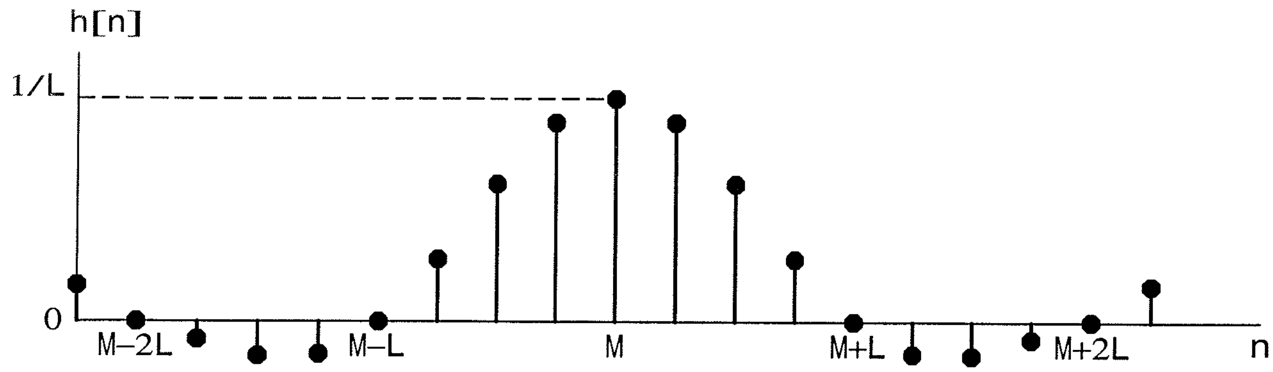
Consider again a Type I linear-phase FIR filter with transfer function

$$H(z) = \sum_{n=0}^{2M} h[n] z^{-n}$$

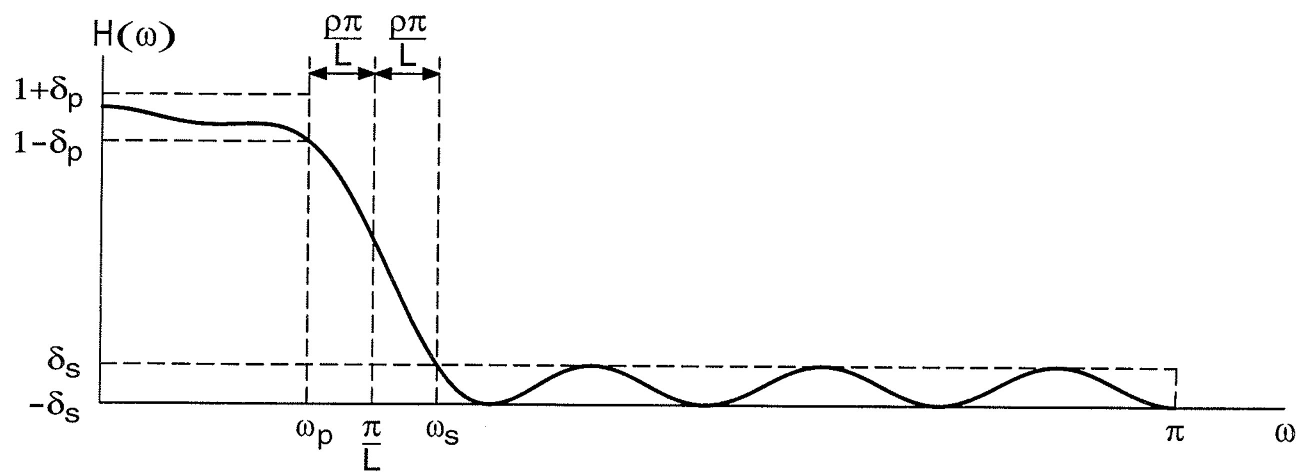
This filter is defined to be an L-th band filter if its coefficients satisfy

$$h(M) = 1/L \qquad (4a)$$

$$h(M + rL) = 0 \quad \text{for} \quad r = \pm 1, \pm 2, \ldots, \lfloor M/L \rfloor. \qquad (4b)$$



(a)



(b)

Importance of Lth-Band Filters

- These filters, also called Nyquist filters, play an important role in designing digital modem systems and filter-banks.
- They can also be used as efficient decimators and interpolators since every L-th impulse response coefficient is zero except for the central coefficient.
- They have also been used in designing alias-free QMF filter banks.
- An important subclass of these filters are half-band filters, which are considered in greater detail later on.

Properties

- It can be shown that the time-domain conditions state some limitations on the frequency response of the filter.
- First, the passband edge (in the lowpass case) is restricted to be less than π/L and the stopband edge to be larger than π/L.
- Usually, the edges are given in terms of an excess bandwidth factor ρ as follows

$$\omega_p = (1 - \rho)\pi/L, \quad \omega_s = (1 + \rho)\pi/L.$$

- Second, if the maximum deviation of $H(\omega)$ from zero in $[\omega_s, \pi]$ is δ_s, then the maximum deviation of $H(\omega)$ from unity in $[0, \omega_p]$ is in the worst case $\delta_p = (L - 1)\delta_s$.
- Usually, δ_p is much smaller than the above upper limit.

Approxination Criteria

- Since δ_p is guaranteed to be relatively small for a small value of δ_s, it can be concentrated on shaping the stopband response of the filter.
- The stopband response can be optimized either in the minimax sense or in the least-mean-square sense.
- In the case of the minimax criterion, the problem is to find the coefficients of $H(z)$ such that the time-domain conditions of Eq. (4) are satisfied and

$$\delta_s = \max_{\omega \in [\omega_s, \pi]} |W(\omega)H(\omega)|$$

is minimized, where $W(\omega)$ is a positive weighting function.

- In the case of the least-mean-square criterion, the quantity to be minimized is

$$E_2 = \int_{\omega_s}^{\pi} [W(\omega)H(\omega)]^2 d\omega.$$

- In some applications, it is desired to factorize $H(z)$ into the minimum-phase and maximum-phase terms.

- In this case, an additional constraint that $H(\omega)$ be nonnegative is required.

Design of L-th Band Filters in the Minimax Sense

- In order to find the filter $H(z)$ minimizing stop-band ripple δ_s and simultaneously meeting the time-domain conditions of Eq. (4), it is plit into two parts as follows:

$$H(z) = H_p(z)H_s(z) = \sum_{n=0}^{2K} h_p[n]z^{-n} \sum_{n=0}^{2(M-K)} h_s[n]z^{-n},$$

where

$$K = \lfloor M/L \rfloor.$$

- Both $H_p(z)$ and $H_s(z)$ are Type I linear-phase filters.
- $H_p(z)$ is determined such that the time-domain conditions of are satisfied, whereas $H_s(z)$ is used for providing the desired stopband response.
- For any $H_s(z)$, $H_p(z)$ can be determined such that the overall filter $H(z) = H_p(z)H_s(z)$ satisfies the time-domain conditions.

- This leads to a system of $2\lfloor M/L \rfloor + 1$ linear equations in the $2\lfloor M/L \rfloor + 1$ coefficients $h_p[n]$ of $H_p(z)$.
- Utilizing the fact that the coefficients of $H_p(z)$ as well as the time-domain conditions are symmetric, a system of $\lfloor M/L \rfloor + 1$ equations needs to be solved.
- The remaining problem is thus to find $H_s(z)$ to give the minimum value of δ_s.

Design Algorithm

The algorithm for iteratively determining the desired $H_s(z)$ consists of the following steps:

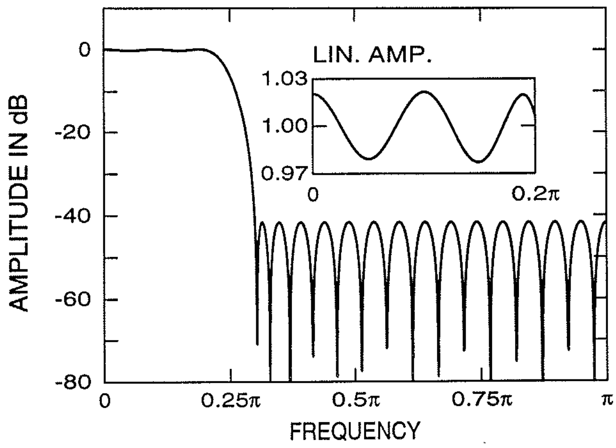
1. Set $H_p(\omega) \equiv 1$ and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{M-K+1}\} = \{0, 0, \ldots, 0\}$.
2. Find $H_s(\omega)$ such that $H_s(0) = 1$ and $W(\omega)H_p(\omega)H_s(\omega)$ alternatingly achieves at least at $M - K + 1$ consecutive points on $[\omega_s, \pi]$ the extremum values $\pm\delta_s$. Store the extremal points into $\bar{\Omega} = \{\bar{\omega}_1, \bar{\omega}_2, \ldots, \bar{\omega}_{M-K+1}\}$.
3. Find $H_p(z)$ such that the time-domain conditions of Eq. (4.131) are satisfied.
4. If $|\omega_k - \bar{\omega}_k| \leq \alpha$ for $k = 1, 2, \ldots, M - K + 1$ (α is a small number), then stop. Otherwise set $\Omega = \bar{\Omega}$ and go to Step 2.

About the Algorithm

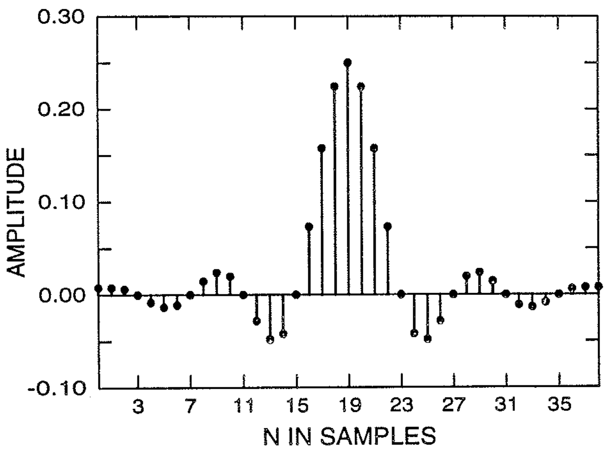
- The desired $H_s(\omega)$ at Step 2 can be found using the MPR algorithm.
- The desired function is zero in $[\omega_p, \pi]$ and weighting function is $W(\omega)/H_p(\omega)$.
- $H(\omega)$ can be forced to take the value unity at $\omega = 0$ by selecting a very narrow passband region $[0, \epsilon]$, setting $D(\omega) \equiv 1$, and by using a large weighting function in this region.
- When a very narrow passband region is used, then the MPR algorithm selects only one grid point in this region.
- Another approach for designing L-th band filters is to use linear programming. However, linear programming requires significantly more computation time than the above simple algorithm.

Example

- The specifications are: $L = 4$ and $\rho = 0.2$ ($\omega_p = 0.8\pi/4 = 0.2\pi$ and $\omega_s = 1.2\pi/4 = 0.3\pi$), and $\delta_s = 0.01$.
- The amplitude and the impulse responses for an optimized filter of order 38 are shown below.



(a)



(b)

Design of Factorizable Lth-Band Filters

- If it is desired that $H(z)$ be factorizable into the minimum- and maximum-phase terms, then the subfilter $H_s(z)$ is written in the form

$$H_s(z) = [\bar{H}_s(z)]^2, \quad \bar{H}_s(z) = \sum_{n=0}^{M-K} \bar{h}_s[n]z^{-n}.$$

- $\bar{H}_s(z)$ is either a Type I linear-phase filter ($M - K$ is even) or a Type II filter ($M - K$ is odd).
- The resulting overall zero-phase frequency response can be expressed as

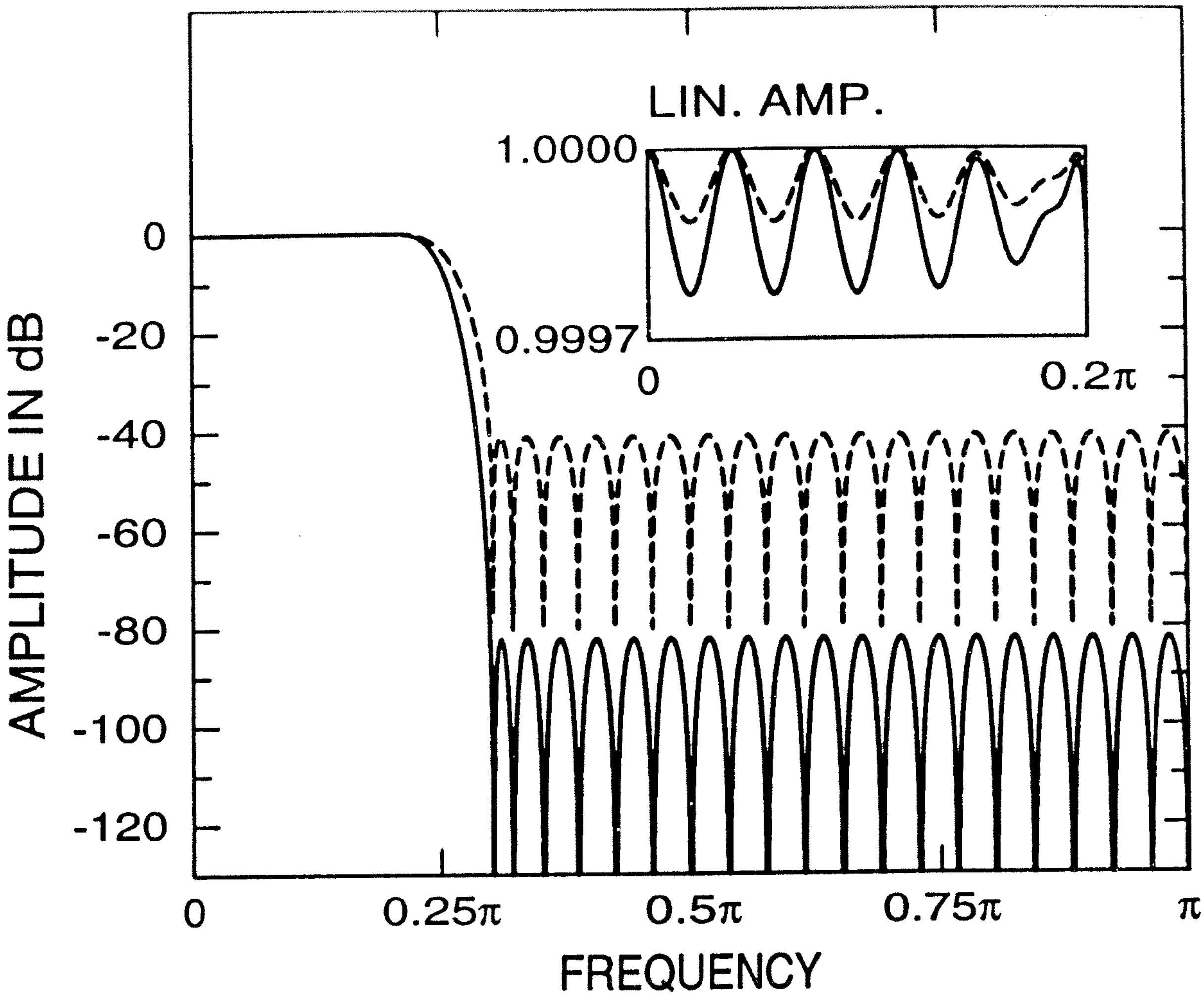
$$H(\omega) = H_p(\omega)[H_s(\omega)]^2.$$

- Since the zeros of $H_p(z)$ are off the unit circle, $H(\omega)$ is non-negative, as is desired.
- In this case, the minimization of the stopband ripple can be performed by slightly modifying the above algorithm.
- The basic difference is that now $\bar{H}_s(\omega)$ is determined at Step 2 such that $\bar{H}_s(0) = 1$ and $\sqrt{W(\omega)H_p(\omega)}\bar{H}_s(\omega)$ oscillates within the limits $\pm\bar{\delta}_s$ on $[\omega_s, \pi]$.

- Correspondingly, $W(\omega)H_p(\omega)[\bar{H}_s(\omega)]^2$ oscillates within the limits 0 and $\delta_s = (\bar{\delta}_s)^2$ on $[\omega_s, \pi]$.
- The advantage of this approach is that both the minimum- and maximum-phase terms of $H(z)$ contain $\bar{H}_s(z)$ and only $H_p(z)$ must be factored in order to get the overall maximum-phase and minimum-phase terms.

Example

- The specifications for the minimum-phase and maximum-phase terms are those in the previous example.
- The required stopband ripple for $H(\omega)$ is $(\bar{\delta}_s)^2 = 0.0001$.
- The amplitude responses for an optimized overall filter of order 106 (solid line) and and for the minimum-phase (or maximum-phase) term of order 53 are shown below.



Design of Half-Band Filters

- A very important subclass of L-th band filters in many applications are half-band filters ($L = 2$). For these filters,

$$h(M) = 1/2$$

$$h(M + 2r) = 0 \quad \text{for} \quad r = \pm 1, \pm 2, \ldots, \lfloor M/2 \rfloor.$$

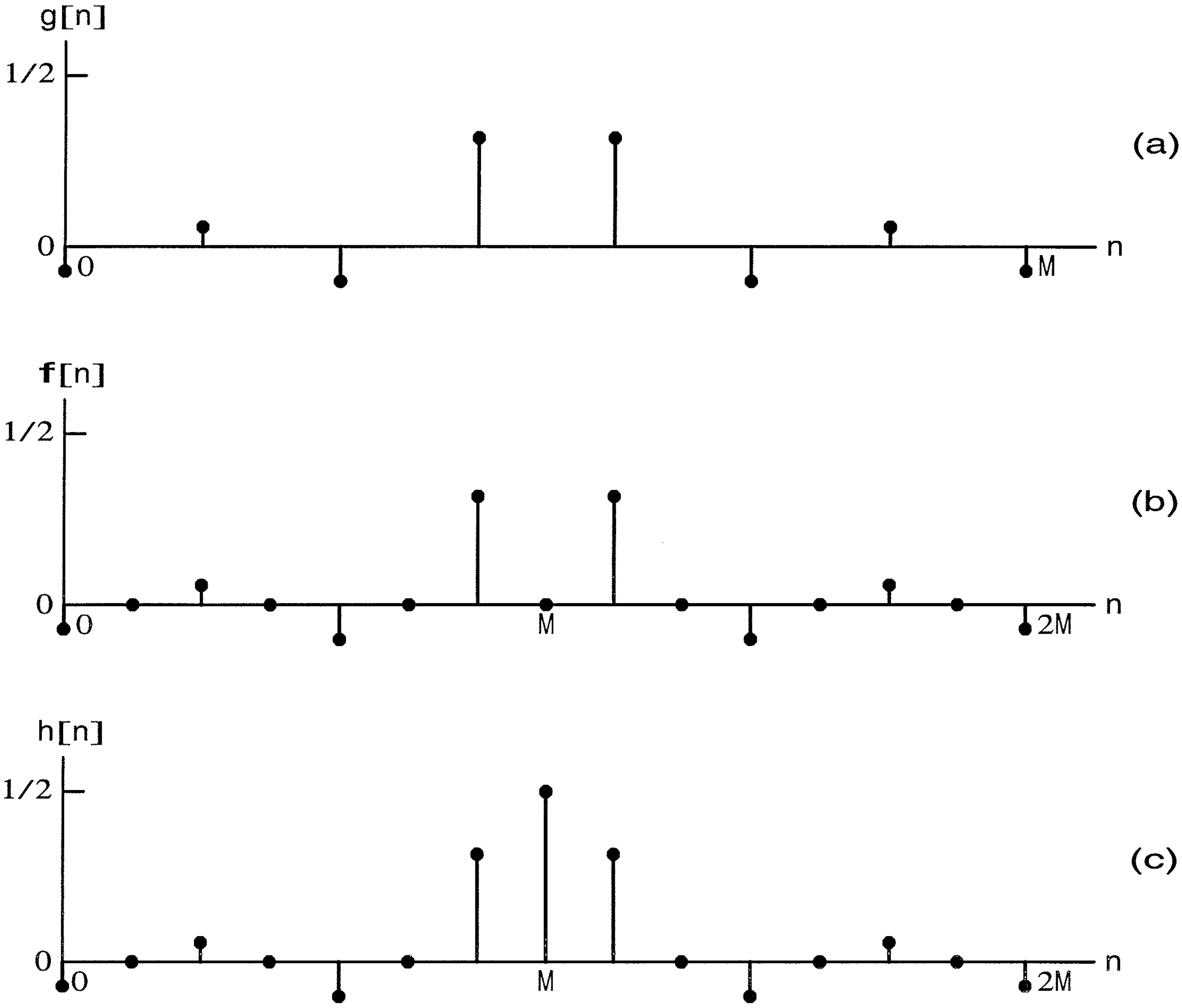
- A filter satisfying these conditions can be generated in two steps by starting with a Type II (M is odd) transfer function

$$G(z) = \sum_{n=0}^{M} g[n] z^{-n}, \quad g[n] = g[M - n].$$

- In the first step, zero-valued impulse-response values are inserted between the $g[n]$'s [see Figures (a) and (b) in the following transparency], giving the following Type I transfer function of order $2M$:

$$F(z) = \sum_{n=0}^{2M} f[n] z^{-n} = G(z^2) = \sum_{n=0}^{M} g[n] z^{-2n}.$$

Generation of the Impulse Response of a Half-Band Filter



- The second step is then to replace the zero-valued impulse-response value at $n = M$ by $1/2$ [see Figure (c) in the previous transprency], resulting in the desired transfer function

$$H(z) = \sum_{n=0}^{2M} h[n]z^{-n} = \frac{1}{2}z^{-M} + F(z) = \frac{1}{2}z^{-M} + \sum_{n=0}^{M} g[n]z^{-2n}.$$

- This gives $h[M] = 1/2$, $h[n] = g[n/2]$ for n even, and $h[n] = 0$ for n odd and $n \neq M$, as is desired.

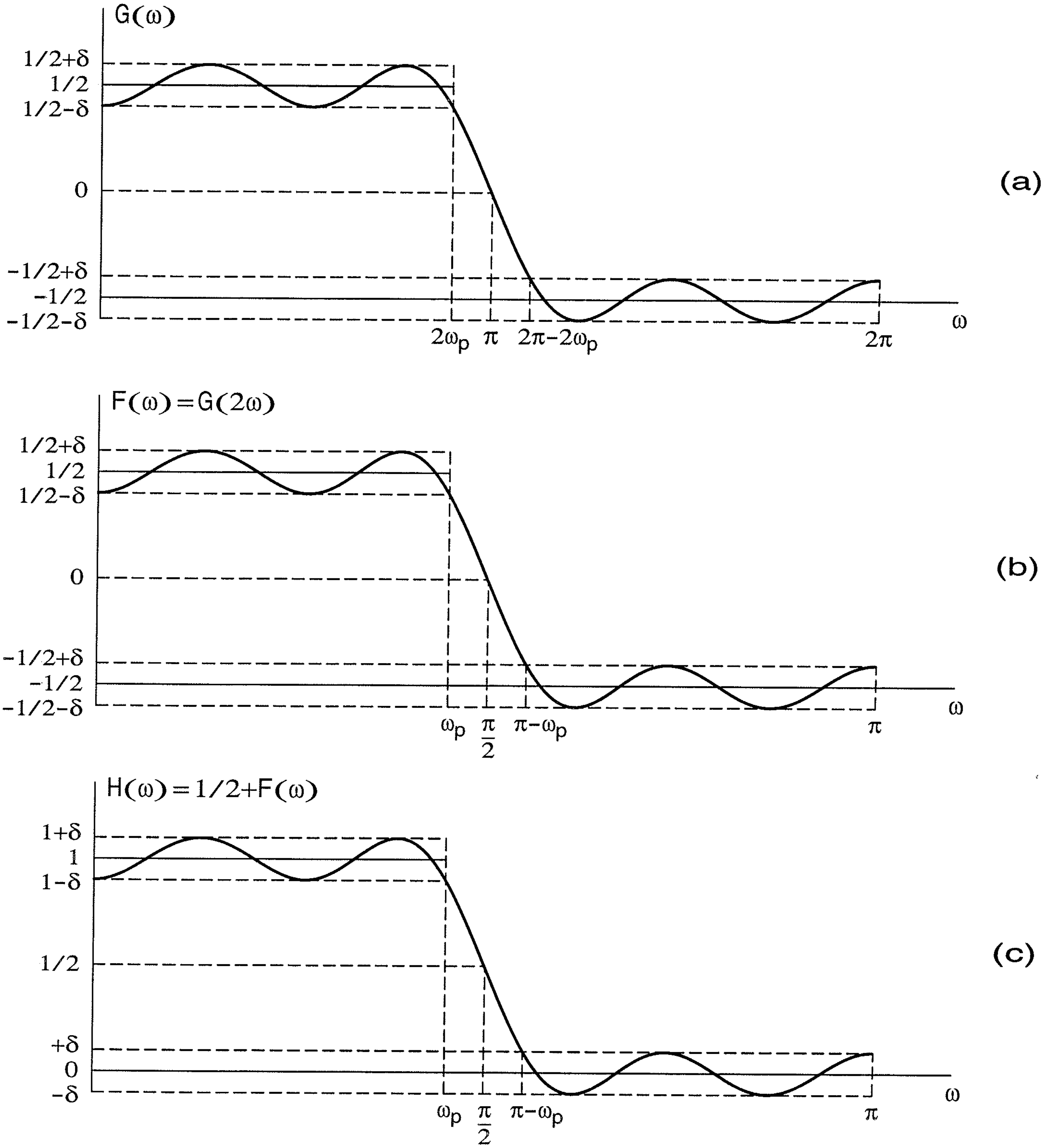
Filter Design

- The zero-phase frequency responses of $H(z)$, $F(z)$, and $G(z)$ are related through

$$H(\omega) = 1/2 + F(\omega) = 1/2 + G(2\omega).$$

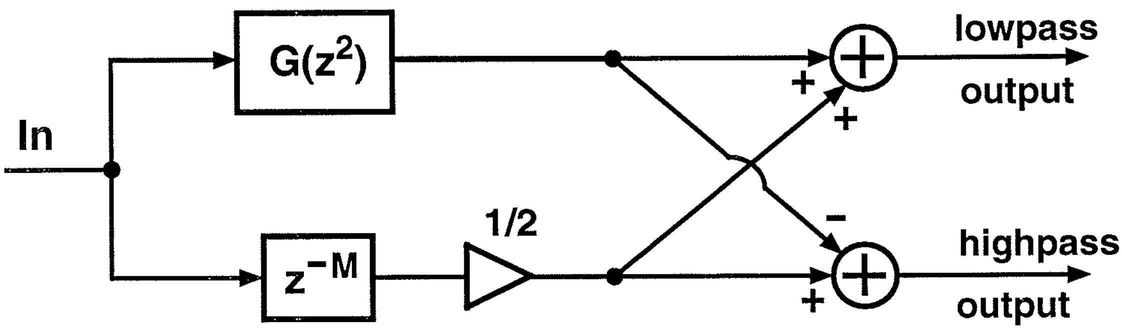
- Based on these relations, the design of a low-pass half-band filter with passband edge at ω_p and passband ripple of δ can be accomplished by determining $G(z)$ such that $G(\omega)$ oscillates within $1/2 \pm \delta$ on $[0, 2\omega_p]$ [see Figure (a) in the following transprency].
- Since $G(z)$ is a Type II transfer function, it has one fixed zero at $z = -1$ ($\omega = \pi$).
- $G(z)$ can be designed directly with the aid of the MPR algorithm using only one band $[0, 2\omega_p]$, $D(\omega) = 1/2$, and $W(\omega) = 1$.
- Since $G(z)$ has a single zero at $z = -1$, $G(\omega)$ is odd about $\omega = \pi$.
- Hence, $G(2\pi - \omega) = -G(\omega)$ and $G(\omega)$ oscillates within $-1/2 \pm \delta$ on $[2\pi - 2\omega_p, 2\pi]$.

Design of A Lowpass Half-Band Filter



Efficient Implementation of a Half-Band Filter

- An implementation for the half-band filter as a parallel connection of $G(z^2)$ and $(1/2)z^{-M}$ is shown below.
- This implementation is very attractive as in this case the complementary highpass output having the zero-phase frequency response $1 - H(\omega)$ is obtained directly by subtracting $G(z^2)$ from $(1/2)z^{-M}$.
- The term z^{-M} can be shared with $G(z^2)$.
- The number of non-zero coefficients in $G(z^2)$ is $M + 1$. By exploiting the symmetry in these coefficients, only $(M+1)/2$ multipliers (M is odd) are needed to implement a lowpass–highpass filter pair of order $2M$.



- The corresponding $F(\omega) = G(2\omega)$ stays within $1/2 \pm \delta$ on $[0, \omega_p]$ and within $-1/2 \pm \delta$ on $[\pi - \omega_p, \pi]$ [see Figure (b) in the previous transparency].
- Finally, $H(\omega)$ approximates unity on $[0, \omega_p]$ with tolerance δ and zero on $[\pi - \omega_p, \pi]$ with the same tolerance δ [see Figure (c) in the previous transparency].
- For the resulting $H(\omega)$, the passband and stopband ripples are thus the same and the passband and stopband edges are related through $\omega_s = \pi - \omega_p$.
- In general, $H(\omega)$ satisfies

$$H(\omega) + H(\pi - \omega) = 1.$$

- This makes $H(\omega)$ symmetric about the point $\omega = \pi/2$ such that the sum of the values $H(\omega)$ at $\omega = \bar{\omega} < \pi/2$ and at $\omega = \pi - \bar{\omega} > \pi/2$ is equal to unity [see Figure (c) in the previous transparency].

Responses for a complementary half-band filter pair of order 34 for $\omega_p = 0.4\pi$.

- The implementation of this filter pair requires only nine multipliers.