Andrei B. Bogatyrev*, Sergei A. Goreinov, and Sergei Yu. Lyamaev

Efficient synthesis of optimal multiband filter

DOI 10.1515/rnam-2017-0020
Received April 7, 2017; accepted ..., 2017

Abstract: A novel analytical approach to the synthesis of electrical (e.g., analogue, digital or microwave) filters is proposed. This approach allows to obtain the lowest possible degree filters with given involved specification including, e.g., many pass- and stopbands, narrow transition bands, high attenuation at the stopbands and low magnitude oscillations at the passbands. Comparison to other existing approaches is given.

Keywords: Electrical filter synthesis, Cauer filter, composite filter, Zolotarev fraction, Riemann surface, Calogero–Moser curve, Remez algorithm.

MSC 2010: 41A20, 93B50, 30E10, 14Hxx, 41A25

Every more or less complex piece of radio equipment contains electric filters — digital, analogue or microwave — which single out useful frequency ranges and suppress unwanted or spurious components as well as noise. The synthesis of filter with the best possible amplitude response is understood as search of rational function of the lowest degree satisfying the given specification and some additional constraints such as positivity or evenness. By specification we mean (see Fig. 1) a set of non-intersecting frequency ranges called passbands (E_+) and stopbands (E_-) together with required stopband attenuation (A_s) and passband ripple (A_p) levels. Sometimes the behaviour of amplitude response in transition band (frequency ranges joining passbands and stopbands) is also specified. This problem is rather complicated even for a numerical approach in the case of many frequency ranges, narrow transition bands, small passband ripple amplitude and low stopband attenuation levels. Its satisfactory solution is available only for the case of lowpass and highpass Cauer–Zolotarev filters (also known as elliptic filters), in which the amplitude response is expressed parametrically by means of elliptic functions.

1 Three approaches to optimal filter synthesis

1.1 Composite and other engineer approaches

An obvious approach to the synthesis of multiband filters is the composite one: a composite multiband filter is obtained by interconnecting several single-band filters. We have used the following rule for their construction: for each bandwidth, compute the frequency response function of the corresponding single-band filter; then take the sum of obtained functions, optimizing the parameters of elliptic filters manually to achieve the least possible order of the resulting multiband filter under the constraint that its amplitude response fits into the corridor given by the specification.

The mentioned single-band elliptic filters are non-optimal but close to optimal single-band filters based on Zolotarev fraction which (after linear scaling) becomes the solution of the third Zolotarev problem of the best rational approximation of the function sgn in uniform norm on the union of two intervals $[-1/k(\tau), -1] \cup [1, 1/k(\tau)]$, where $k(\tau)$ is the elliptic integral modulus. Zolotarev fraction of order n as a function of variable x has the following parametric representation,

***Corresponding author: Andrei B. Bogatyrev,** Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow 119333, Russia. E-mail: ab.bogatyrev@gmail.com
Sergei A. Goreinov, Sergei Yu. Lyamaev, Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow 119333, Russia

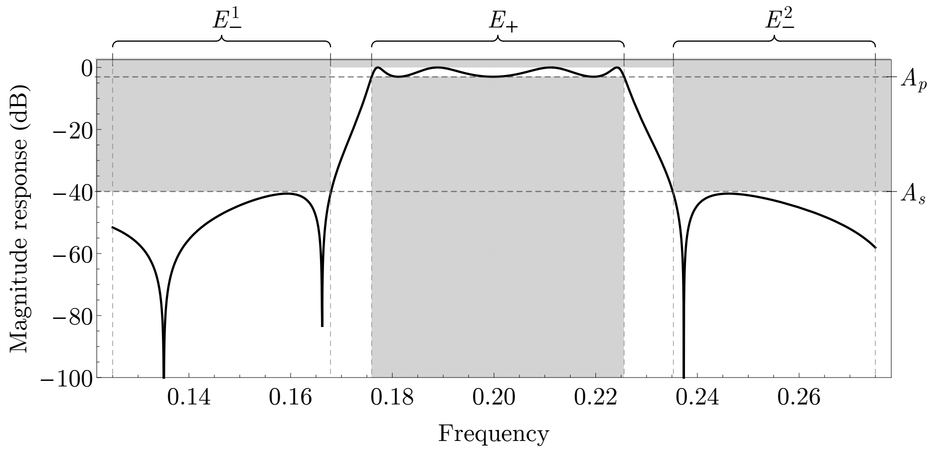

Figure 1: Amplitude response of a filter with two stopbands ($E_-^{1,2}$), one passband (E_+), passband ripple $A_p = -3$dB and stopband attenuation $A_s = -40$dB.

$$Z_n(x(u)) := \mathrm{sn}(K(\tau)u|\tau), \quad x(u) := \mathrm{sn}(K(n\tau)u|n\tau), \qquad u \in \mathbb{C}, \qquad n \in \mathbb{N}$$

where $\mathrm{sn}(\cdot|\tau)$ is the elliptic sine of modulus τ and $K(\tau)$ is the complete elliptic integral of the same modulus.

The disadvantage of composite approach is the significant non-optimality of the orders of generated multiband filters. This explains the interest in the development of other approaches based on more complex ideas but capable of generating filters having smaller orders with the same specification. Most of these approaches can be conventionally divided into the following categories [12]: (1) methods of microwave filter synthesis based on the use of multifrequency resonators; (2) methods that bring the single-band prototype filter to a multi-band configuration using certain substitutions of frequency variable; (3) optimization methods based on the Remez algorithm.

Much attention in the technical literature is given to the development of methods of the third type associated with the synthesis of optimal and close to optimal filters. Here by optimal we mean a filter that for a given multiband specification has the smallest order among all physically realizable filters that meet the requirements of this specification.

1.2 Optimal filters

The search for an optimal filter corresponding to a given specification can be reduced to series of generalized Zolotarev problem solutions. Consider two equivalent formulations of this problem. Let E be the union of m disjoint intervals of the real axis divided into two parts, the passbands E_+ and the stopbands E_-.

(1) Find a real rational function $R(x)$ of degree at most n for which the following quantity is minimal:

$$\frac{\max_{x \in E_+} |R(x)|}{\min_{x \in E_-} |R(x)|}. \tag{1.1}$$

(2) Another formulation is to find a real rational function $R_*(x)$ of degree at most n with n given for which the deviation from the ideal transfer function $F(x) := \pm 1$, $x \in E_\pm$, is minimal:

$$||R_* - F||_E := \max_{x \in E} |R_*(x) - F(x)|. \tag{1.2}$$

It is easy to show that the solutions of these two problems differ by a linear fractional substitution of the dependent variable and that the corresponding minimal deviations are related by a simple formula.

To construct a multiband filter with a given specification, one has to solve the least deviation problem with some high degree n and decrease it until the respective deviation satisfies the existing specification. The solution of the least deviation problem in the form (1.1) or (1.2) is unique in each of the 2^{m-2} classes which divide the set of rational functions. The solution in each class admits an equioscillation characterization (Chebyshev alternation) and can be found, e.g., by algorithms of Remez type [2, 15, 16]. Exactly the smallest of the deviations obtained for each class gives the global minimum value. The disadvantage of this approach is not only in the necessity of good initial approximation but also in the intrinsic instability present in methods of this type. With the use of double precision, Remez type methods do not allow obtaining solutions of a degree greater than 15-20.

1.3 Method of algebro-geometric ansatz

The method proposed in this paper is based on explicit analytical formulas obtained in [3] and generalizing Zolotarev fraction to the case of many intervals. Earlier, the ideology of this algebro-geometric approach was applied to the problems of polynomial optimization, in particular to the computation of Chebyshev polynomials for many intervals [9] which are used, e.g., in iterative solution of linear systems with large symmetric matrices whose spectrum is clustered on several intervals, and also in computation of optimal stability polynomials [7, 8] for multistage Runge–Kutta methods of the third order of accuracy.

The idea of this approach is as follows. The solutions of the least deviation problems satisfy the well-known equioscillation principle [1]: solution of degree n has $2n + 2$ equioscillation points on E in which the deviation value is attained with a successive change of sign. An equioscillation point in the interior of E will necessarily be a critical point of the solution function, which takes there one of four distinguished values (e.g., $\pm 1 \pm ||R_* - F||_E$ for the formulation 1.2). A rational function of degree n has only $2n - 2$ critical points counted with multiplicities, therefore the solutions of the least deviation problems satisfy the following property: *the great majority of critical values of the solution $R(x)$ are simple with their values in a four-element set* Q. The number of extremal critical points with respect to the set Q is given by the formula

$$g = 1 + \sum_{x:\, R(x)\notin \mathsf{Q}} \operatorname{ord}\, dR(x) + \sum_{x:\, R(x)\in \mathsf{Q}} \left[\frac{1}{2}\operatorname{ord}\, dR(x)\right] \tag{1.3}$$

where the sum is taken over all points in the Riemann sphere; ord $dR(x)$ is the branching order of the holomorphic map $R : \mathbb{C}P^1 \to \mathbb{C}P^1$ at the point x. For instance, this quantity is equal to zero at the simple poles of the function $R(x)$.

Functions with small g in (1.3) have a very special form and fill out the manifolds of low dimension in the space of all rational functions of a given degree. When solving the least deviation problems, it is reasonable to move from searching through the whole space of rational functions to searching over a small set of rational functions satisfying the above property. Such functions possess the following effective and stable low-parametric representation [3]:

$$R(x) = \operatorname{sn}\left(\int_{(e,0)}^{(x,w)} d\zeta + A(e) \,\Big|\, \tau \right), \qquad \mathsf{Q} = \{\pm 1, \pm 1/k(\tau)\} \tag{1.4}$$

where $d\zeta$ is a holomorphic Abelian differential on a Riemann surface X of genus g, whose periods lie in the lattice of periods of elliptic sine, while the phase shift $A(e)$ is also commensurable to this lattice. The surface X is determined by rational function $R(x)$ as the double covering of the Riemann sphere branched over the points where $R(x)$ takes the values from Q with odd multiplicity. Emerging Riemann surface is not arbitrary, it is the so-called Calogero–Moser curve which covers (with due branching) the torus defined by the set of distinguished values Q.

This parametrization of extremal functions allows us to control the behaviour of the solution in the filter transition bands (selection of the solution class) and directly solve the problem of the least degree filter for a given specification without considering series of the least deviation problems. The computational tools used for finding the extremal rational functions given by explicit analytical formula (1.4) include the previously

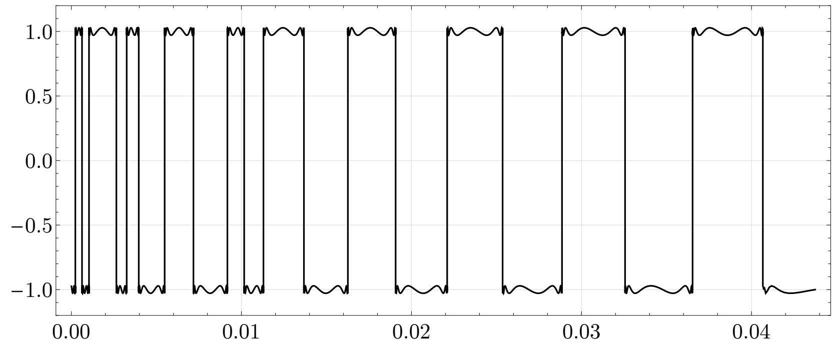

Figure 2: Graph of the solution of optimization problem (for a specification with $m = 21$).

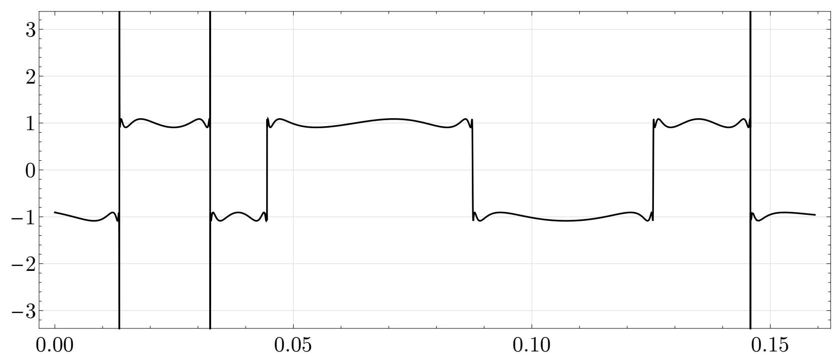

Figure 3: Graph of the solution of optimization problem from a class that allows for the presence of poles in the first, second and sixth transition bands (for a specification with $m = 7$).

developed [3–6] machinery for efficient computations on the Riemann surfaces and allow us to compute the solutions of degree n equal to at least several hundred in a stable way.

Using the ansatz (1.4) we computed the examples of solution of optimization problem (1.2). Figure 2 presents a graph of the solution of degree $n = 387$ for a specification with the number of bands $m = 21$ ($g = m = 21$). Figure 3 depicts a solution graph for a specification with $m = 7$ from a class that admits the presence of poles in the first, second and sixth transition bands.

1.4 Direct numerical optimization

The solutions of the optimization problem can be computed approximately, e.g., using the Remez type algorithms [2, 15, 16]. Starting point will be again the equioscillation principle in the formulation due to Akhiezer [1], bounding from below the number of equioscillation points on E of the exact solution. Assume that we have an approximate fraction $\varphi(w)/\psi(w)$ given by coefficients of numerator and denominator in some polynomial basis; and also the alternant A, i.e., the set of approximate equioscillation points. The algorithms in question consist in successive refinement of each of these two objects:

1. having the alternant set $A \subset E$, we can modify the fraction so that it will obey the equioscillation condition on A;
2. having the fraction, we can modify the positions of points in order to increase the uniform error norm over A (if this is impossible, then it is plain that the fraction is the exact solution of our problem).

Besides, we need an initial approximation, which is very important due to local convergence property of Remez type algorithms. We describe briefly our implementation of all the steps above.

Choice of initial approximation. In principle, good approximations can be devised for both involved objects, but we start with the alternant set A whose asymptotic behaviour is known. Namely, we use the $1/n$-

quantiles of the equilibrium measure for the set E which is computed using the numerical solution of the integral equation with the logarithmic kernel [13].

Step 2 *(refinement of A using the given fraction φ/ψ)*. First, we consider a wider set containing the endpoints of our m intervals and critical points of the error function $\varphi(w)/\psi(w) - F(w)$, which lie on E. Points corresponding to errors less than the level obtained previously at step 1 are discarded. The new set A must obey the rule of succession of signs and contain $2n + 2$ points. The presence of more or less of them is corrected with the help of previously introduced equilibrium measure, e.g., we add the points corresponding to the larger values of density of this measure. The search for critical points is carried out by the Brent method.

Step 1 *(refinement of the fraction φ/ψ using the given set A)*. Unlike polynomial approximations, the equioscillation condition for the set A cannot be represented, to our knowledge, in the form of a linear system; the best formulation we can obtain on this way is a generalized eigenvalue problem (see, e.g., [2]) with unsymmetric matrices. Even in the case of two intervals it may happen that standard mantissa computations are not sufficient for getting at least one decimal digit of coefficients of φ and ψ, let alone their roots. Besides, the search through all variants of equioscillation signs in adjacent intervals means an exponential in m complexity.

Some advantages with respect to the unsymmetric eigenvalue problems can be found in reduction of the problem (1.2) formulated for the given set A to a linear programming problem, although here again we see an exponential in m search associated with the choice of the sign of denominator of the fraction at each interval. Here is the problem: Minimize $t > 0$ under the linear constraints

$$-t\sigma(w)\psi(w) \leqslant \sigma(w)\varphi(w) - \sigma(w)\psi(w)F(w) \leqslant t\sigma(w)\psi(w), \qquad w \in A, \quad j = 1, \dots, m$$

where we assume $\operatorname{sign}\psi(w) =: \sigma(w)$ given *a priori* at each interval. Formally, we can remove this last assumption by some increase in the number of variables. Good results for this problem are shown by the primal-dual interior point method [10].

Linear parametrization of numerators and denominators of the fractions substantially limits the stability of Remez type algorithms [16]. The maximum degree of the solution R obtained by such methods working in double precision depends on configuration of the set E but does not exceed $n = 20$.

2 Comparison of the three approaches

This section provides a comparison of two approaches to the synthesis of optimal multiband filters: new algebro-geometric approach and the direct optimization method. Also, the comparison of optimal filters with (non-optimal) composite ones obtained by the corresponding approach is given. All considered examples pertain to the synthesis of digital filters. Optimal digital filters were obtained from the optimal analogue ones by means of a standard linear fractional frequency transform.

In Table 1, the orders of optimal and composite filters for six specifications are compared. The data from the first five columns of the table determine the filter specification, while the last two columns contain the orders of optimal and composite filters computed according to that specification. The composite filters obtained by the corresponding approach, as one can see from the table, have significantly higher orders, with respect to the optimal ones.

All the optimal filters whose orders are given in Table 1 are computed using the new algebro-geometric approach. Direct optimization did not yield any satisfactory results for all of the examples due to the complexity of specifications. Tables 2 and 3 show the results of using the direct optimization method for the specifications from the first and the second rows of Table 1 (single- and dual-band filters). In these computations, the bands were given by the specification, the passband ripple level was fixed at −2dB, and the stopband attenuation was determined by the amplitude response of the optimal filter scaled to fixed passband ripple. In both cases, the Remez type algorithm turned out to be unapplicable for finding the optimal filter of the order required to satisfy the original specification.

Table 1: Comparison of the orders of optimal and composite filters computed according to the same specification. We use the following notation: m — number of intervals in the set $E = E_+ \cup E_-$; the columns $E_\pm$ contain all intervals of respective sets; A_p — minimum possible passband ripple level, in dB; A_s — maximum possible stopband attenuation level, in dB; n_o — order of optimal filter; n_c — order of composite filter.

m	E_-	E_+	A_p (dB)	A_s (dB)	n_o	n_c
3	[0.00000, 0.14682] [0.23058, 0.50000]	[0.16249, 0.23056]	−2	−46.9	18	28
5	[0.00000, 0.11030] [0.21531, 0.32348] [0.36359, 0.50000]	[0.12255, 0.20279] [0.32664, 0.35590]	−2.6	−40	16	23
5	[0.25581, 0.25591] [0.25922, 0.25933]	[0.00000, 0.25574] [0.25598, 0.25915] [0.25941, 0.50000]	−0.25	−51.9	16	62
5	[0.00000, 0.05989] [0.25227, 0.25250] [0.36036, 0.50000]	[0.06723, 0.25222] [0.25267, 0.35920]	−2	−50	24	59
9	[0.00000, 0.03111] [0.05995, 0.11896] [0.18120, 0.19579] [0.27450, 0.28356] [0.33693, 0.50000]	[0.03133, 0.04665] [0.13359, 0.18111] [0.19660, 0.27429] [0.28360, 0.33539]	−2	−42.8	52	96
11	[0.00000, 0.04735] [0.07919, 0.10005] [0.26931, 0.28709] [0.30257, 0.33024] [0.36978, 0.37504] [0.38597, 0.50000]	[0.04762, 0.07870] [0.10026, 0.26436] [0.28795, 0.30241] [0.33042, 0.36959] [0.37517, 0.38444]	−2	−50	76	121

Table 2: Results of direct numerical optimization for single-band filter.

Order	Stopband attenuation A_s, dB	Number of iterations
3	−3.3812	102
4	−4.2272	154
5	−7.5894	211
6	−10.612	308
8	−19.355	457
9	−22.378	652
10	—	> 1000
18	**−46.9**	**n/a**

Table 3: Results of direct numerical optimization for dual-band filter.

Order	Stopband attenuation A_s, dB	Number of iterations
3	−5.3421	151
4	−9.1785	273
5	−12.649	351
6	−16.013	408
7	−19.215	594
8	–	> 1000
16	**−40**	**n/a**

3 Conclusion

The article contains a comparison of three approaches to the synthesis of multiband electrical filters: the new algebro-geometric approach, the direct numerical optimization based on Remez method and semi-analytical composite approach. Today, the direct numerical optimization has probably the widest algorithmic support: there exist mature packages for engineering computations. Unfortunately, the unavoidable instability of Remez type algorithms does not allow to solve too complicated problems: in double precision arithmetic (15 decimal digits) the filter's degree does not attain 20, which gives two or three bands provided that the stopband attenuation is not too deep, that the passband ripple is not too small and that the transitions between stop- and passbands are not too sharp. Composite approach consists in breaking up a complex problem into a series of simple ones and in successive solving of the latter by constructing the amplitude response of the bandpass filter using Zolotarev fraction. Its advantage is that in this way it is always possible to obtain an (ersatz) solution for a given specification. As a rule, it is far from optimal: the order of composite filter can be several times greater than the degree of optimal filter with the same mask. The situation gets worse with the increasing complexity of the filter specification (number of bands, selectivity, attenuation level, ripple level). From our point of view, the most promising, as well as the least studied from the algorithmic aspect, is the algebro-geometric approach based on a however complex mathematical apparatus. The authors intend to continue their research in this direction.

Funding: This work was supported by the Russian Science Foundation (grant 16-11-10349).

References

[1] N. I. Akhiezer, On a problem of Zolotarëv. *Izv. Akad. Nauk SSSR* (1929), No. 10, 919–931 (in Russian).
[2] R. A.-R. Amer and H. R. Schwarz, Contributions to the approximation problem of electrical filters. *Mitt. Inst. Angew. Math.* (1964), No. 9, Birkhäuser, Basel. 99pp.
[3] A. B. Bogatyrev, Chebyshev representation for rational functions. *Sb. Math.* **201** (2010), No. 11, 1579–1598. http://www.inm.ras.ru/library/Bogatyrev/BOGA11SE.pdf
[4] A. B. Bogatyrev, Computations in moduli spaces. *Comp. Methods and Function Theory* **7** (2007), No. 2, 309–324.
[5] A. B. Bogatyrev, *Extremal Polynomials and Riemann Surfaces*. Springer Monographs in Mathematics, 2012.
[6] A. B. Bogatyrev, Conformal mapping of rectangular heptagons. *Sb. Math.* **203** (2012), No. 12, 1715–1735.
[7] A. B. Bogatyrev, Effective solution of the problem of the optimal stability polynomial. *Sb. Math.* **196** (2005), No. 7, 959–981.
[8] A. B. Bogatyrev, Effective computation of optimal stability polynomials. *Calcolo* **41** (2004), No. 4, 247–256.
[9] A. B. Bogatyrev, Effective computation of Chebyshev polynomials for several intervals. *Sb. Math.* **190** (1999), No. 11, 1571–1605.
[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
[11] W. Cauer, *Theorie der linearen Wechselstromschaltungen*, Bd. 1. Becker und Erler, Leipzig, 1941; Bd. 2. Akademie, Berlin, 1960.
[12] V. Crnojevic-Bengin, *Advances in Multi-Band Microstrip Filters*. Cambridge University Press, 2015.
[13] W. Fuchs, On Chebyshev approximation on sets with several components. In: *Aspects of Contemporary Complex Analysis* (Eds. D. A. Brannan and J. G. Clunie). Academic Press, 1980, pp. 399–408.
[14] A. A. Gonchar, Zolotarev problems connected with rational functions. *Math. USSR-Sb.* **7** (1969), No. 4, 623–635.
[15] V. N. Malozemov, The synthesis of a multi-band electric filter. *USSR Comp. Math. Math. Phys.* **19** (1979), No. 3, 41–51.
[16] E. Ya. Remez, *Foundations of Numerical Methods of Chebyshev Approximation*. Kiev: Naukova Dumka, 1969 (in Russian).
[17] J. Shen and G. Strang, The asymptotics of optimal (equiripple) filters. *IEEE Trans. Sig. Proc.* **47** (1999), No. 4, 1087–1098.
[18] E. I. Zolotarëv, Applying elliptic functions to questions of functions of greatest and least deviation from zero. *Zap. S.-Petersburg Akad. Nauk.* **30** (1877), No. 5, 37–41, 54–71. (in Russian).