

Neurális hálózatok házi feladat beszámoló

Pintér Bálint (I6QS0K), Szilágyi Gábor (NOMK01)

Koncentrált paraméterű RF szűrő optimalizációja
aktív tanulóval

2023. június 9.

Tartalomjegyzék

1. Bevezetés	1
1.1. Aktív tanulás	1
1.2. Az optimalizálandó probléma	2
1.3. A felhasznált könyvtár	3
2. A célfüggvény felírása	4
2.1. A hálózat kapcsolási rajza	4
2.2. Impedanciák és admittanciák	5
2.3. Láncparaméterek	7
2.4. Szórási paraméterek	7
3. A célfüggvény javítása	8
3.1. Az első verzió és a problémái	8
3.2. A paraméterek logaritmizálása	9
3.3. A függvényérték logaritmizálása	9
3.4. Büntetett margó	9
4. Optimalizált példák	11
4.1. Aluláteresztő szűrő	11
4.2. Összetett példa	12
A. Az általunk írt specifikáció osztály	16

1. Bevezetés

1.1. Aktív tanulás

A legtöbb neurális hálózatokat használó megoldás olyan problémára irányul, ahol sok rendelkezésre álló adat alapján kell a hálót betanítani egy feladat elvégzésére. Az a legtöbb esetben teljesül, hogy még több tanító adat felhasználásával jobb hálót lehetne tanítani, de ennek az extrém esetre tud megoldást nyújtani az aktív tanulás. A nem adathiányos problémáknál a rendelkezésre álló, címkézett adatpontok nagy részét felhasználva szokás tanítani a hálót, majd a fennmaradó adatpontokon ellenőrizni a háló teljesítőképességét olyan esetekre, amikkel nem találkozott a tanulás során. Az aktív tanulás folyamata ettől merőben eltér.

Aktív tanulás kiindulási helyzete, hogy nagyon sok címkézetlen adat áll rendelkezésre, de az egyes adatpontok címkézése rendkívül költséges. A címkézés költsége miatt vég-eredményben az a cél, hogy azt minél kevesebbszer kelljen elvégezni a tanulás során. A tanulási folyamat közben az eddig megkapott kevés címkézett adatpont alapján a háló jelöl ki következőnek címkézésre azt, amelyik várhatóan a leghasznosabb lesz számára. A hasznosság becslésére több megközelítés is létezik, erre a későbbiekben visszatérünk.

Az aktív tanulás egyik alosztálya a Bayes-optimalizáció. Itt nem egy osztályozót tanítunk minél kevesebb címkézett adat alapján, hanem egy „fekete doboz” függvény maximumát keressük a függvény minél kevesebb kiértékelése mellett. Ez a különbség már befolyásolni fogja a következőnek megcímkézendő adat választását, ami ebben az esetben a következő paraméterértékek megválasztását jelenti, ahol kiértékeljük a függvényt.

1.2. Az optimalizálendő probléma

Az optimalizálendő probléma egy koncentrált paraméterű rádiófrekvenciás (RF) szűrő elemeinek megadása, azaz méretezése. Egy koncentrált paraméterű elemekből felépülő szűrő alapesetben ellenállásból, tekercsekben és kondenzátorokból áll, de léteznek aktív elemek, azaz erősítőket is tartalmazó szűrők – de ez már kívül esik a vizsgálatunkon, csak passzív eszközökkel foglalkozunk. Feltesszük továbbá, hogy a vizsgálatunk során csillapítással nem foglalkozunk, így ellenállást sem használunk a szűrőnkhez.

Szűrőtervezéskor mindig egy előre megadott specifikációból indulunk ki, ebben elő van írva azon frekvenciatartományok csoportja, ahol a szűrni kell, illetve ahol csillapítás nélkül kell az RF teljesítményt áteresztetni. Ennek megfelelően beszélhetünk záró és áteresztő sávokról, ahol a szűrő impedanciakaraktisztikájának rendre nagy, illetve kicsinek kell lennie, ehhez a tekercsek és kondenzátorok soros és párhuzamos rezonanciáit használjuk ki. (S paraméter majd később).

A valóságban mind a tekercsek, mind a kondenzátorok 3D-s objektumok, induktivitásuk és kapacitásuk anyagparamétertől és geometriai méretektől függenek. Anyagparamétertől függés alatt a kondenzátorok dielektrikumában a relatív permittivitását, illetve a tekercsek belsejében a közeg relatív permeabilitását kell érteni – bár utóbbi nem jellemző RF szűrők esetében. Az anyagparaméter tehát egy fix érték, amit nem egyszerű változtatni, méretezés céljából ez nem járható út. Megoldást a geometriai méretek optimális megválasztása jelent, ami utat nyit az alapesetektől, az analitikus formulákkal egyszerűen kiszámolható geometriáktól való eltéréshez. Bár ezekben az alapesetekben, a síkkondenzátorban és az egyenes tekercsben, a szolenoidban is megjelennek geometriai jellemzők, a

$$C = \varepsilon_0 \varepsilon_r \frac{A}{d}$$
$$L = \mu_0 \mu_r \frac{N^2 A}{l}$$

formulák egyrészt csak közelítések, azaz nem írják le pontosan az L és a C értékeket, másrészt sokszor nem kivitelezhető az általuk leírt komponens.

A feladat innentől az, hogy olyan geometriai méreteket találjunk, amivel a komponensek pont az általunk kívánt L és C értékeket mutassák, ezzel a specifikációnak eleget tudunk tenni. Esetünkben a nem analitikusan kiszámolható értékekhez numerikus megoldást kell keresni, amire egy jó megoldás, ha 3D véges elemes szimulációt készítünk. Egy-egy szimuláció futási ideje a megkövetelt precizitás és a rendelkezésünkre álló erőforrás függvényében változhat, de mindenképp több mint csupán egyetlen egyenlet kiértékelése. Éppen ezért szükséges az optimalizálást úgy végezni, hogy szempont legyen a minél kevesebbszer történő kiértékelés.

A fent leírtak az RF szűrőtervezés témakörében egy releváns problémát takarnak, munkánkban azonban nem 3D véges elemes szimulációkat használunk, nem geometriai paraméterek optimalizálását végezzük el. Tesszük mindezt egyrészt azért, mert a szimulációs szoftverek programozott vezérlése meglehetősen nehéz, másrészt a neurális hálók szempontjából teljesen lényegtelen, hogy a kiértékelt függvényben mi történik. Azonban maga az optimalizálás amit csinálunk alkalmas egy ilyen feladatra, és érdemben tudja a tervezőt segíteni munkája során. Visszatérni tehát a koncentrált paraméterű, L és C elemeket tartalmazó hálózatra az optimalizáció szempontjából nem jelent kevesebb munkát, ebben az értelemben nem jár a feladat egyszerűsítésével.

A feladathoz már csak egy lépés hiányzik, amivel a problémát át tudjuk alakítani egy szélsőérték keresésre. Ennek az alapgondolata az, hogy az elvárt és a kapott kimenet különbségét vesszük figyelembe egy büntetőfüggvény segítségével. Magát az optimalizálást ezen keresztül tudjuk megtenni, ennek a függvénynek a kifejtésére a későbbiekben kerül sor.

1.3. A felhasznált könyvtár

Bayes-optimalizáció megvalósításához már léteznek elkészített könyvtárak, így azt külön nem készítettük el, hanem felhasználtuk. Az általunk használt ingyenesen elérhető és letölthető *Bayesian Optimization* könyvtár más, neurális hálózatokra készített, ugyancsak szabadon hozzáférhető könyvtárakon alapszik [3], itt most ennek a bemutatását adjuk [1].

A *Bayesian Optimization* használata rendkívül egyszerű, összesen két dolgot kell megtennünk. Elsőként példányosítjuk a *BayesianOptimization* osztályt, ami bemeneti paraméterként megkapja a maximalizálandó költségfüggvényt, illetve a határait annak a tartománynak, ahol a maximumot keresnie kell. Fontos kiemelni, hogy esetünkben a kapott és az elvárt kimenetek közötti minimalizálás a feladat, így a költségfüggvényben át kell térni maximumhely keresésre.

A szélsőértékhely keresést ezután a *maximize()* tagfüggvény hajtja végre. Ez a folyamat kezdetben random helyeken kiértékel pontokból indul el, majd iteratíván halad a modellterben az optimum felé. Ennek megfelelően a *maximize()* bemeneti paraméterei az inicializáló, véletlenszerű helyek száma és az iteráció száma. Ebben elsőként példányosul egy *UtilityFunction* nevű osztály, utána elindul az iteráció. Minden kör az előző lefutás utáni paraméterfrissítéssel, az *update_param()*-mal kezdődik, ami a *UtilityFunction* egy tagfüggvénye. A függvénynek több felderítési típusa lehet, ezek a UCB (Upper Confidence Bounds method), az EI (Expected Improvement method) és a POI (Probability Of Improvement), alapesetben az UCB-t használjuk. Az *update_param()*-nak három hiperparamétere van: χ , κ és κ_{decay} , az utóbbi a κ csökkentését végzi minden iterációban, de csak egy másik paraméter, κ_{delay} iterációtól kezdve. Frissítve a *UtilityFunction*-t, egy javaslattevő tagfüggvénynek, a *suggest()*-nek adjuk át inputként, ami javasol egy új pontot, az *x_probe*-ot a következő kiértékelésre. Ebben a *suggest()*-ben egy *argmax* vizsgálat történik egy akvizíciós függvényre, az *acq_max*-ra. Ez a vizsgálat két lépésből áll, elsőként egyenletes eloszlással felvesztünk mintavételi pontokat a *UtilityFunction*-ból a paramétertartomány határain belül, alapbeállításként 10^5 darabot. Második lépésként *L-BFGS-B* optimalizációs metódus futtatunk le minden mintavételezett pontra. Egy ciklusban kiválasztjuk a legnagyobbat, ezzel térünk vissza, ez lesz *x_probe*, a javasolt új kiértékelési hely. A *maximize()* iterációs ciklusának harmadik lépése maga a kiértékelés, erre a *probe()* tagfüggvény szolgál, ami az általunk megadott költségfüggvényt futtatja le a kiszámolt új *x_probe* helyen. Ezt a három iterációs lépést folytatja a kód a meghatározott iterációs számig.

Az *L-BFGS-B* optimalizáló solver nem ebben a könyvtárban van megírva, hanem a *scipy.optimizer* tartalmazza [3]. Ez egy minimalizálási feladatokra készített másodrendű optimalizációs algoritmus, az *L-BFGS* kiterjesztése korlátok kezelésére. Feloldva a rövidítést a Limited-memory Broyden-Fletcher-Goldfarb-Shanno algoritmus egy kvázi-Newton módszer, másodrendű deriváltak közelítésére hasznos, ahol az közvetlenül nehéz kiszámolni. Konkretizálva ez azt jelenti, hogy nem számolja ki a Hesse-mátrixot, csak annak inverzét közelíti. A módszer nevében az L előtag a limitált memóriára utal, azaz mindig csak az elmúlt m lépés koordináta- és gradiensvektorát tárolja el. Ez a memóriagazdál-

kodás fontos, különösen az *acq_max* függvényben, ahol mind a 10^5 mintavételi pontra elvégezzük az *L-BFGS-B*-t.

2. A célfüggvény felírása

A célfüggvényünket olyan formában definiáltuk, hogy az adott specifikációnak jobban megfelelő szűrőparaméterekhez nagyobb számot rendeljen. A specifikációban frekvenciasávok és a hozzájuk elvárt minimális vagy maximális átvitel van megadva. Az átvitel itt a teljes kétportú (vagyis négypólusú) rendszer szórási mátrixának (*S*-mátrixának) egyik elemét, név szerint az S_{21} -et jelenti. Mivel ez a reaktív elemek jelenléte miatt egy komplex szám, ennek az abszolút értékét vizsgáljuk, a specifikációbeli korlátok is az abszolút értékre vonatkoznak. Az S_{21} paraméter egy frekvenciafüggő mennyiség, ami azt az információt hordozza, hogy ha az 1-es porton (kapun) beküldünk a szűrőbe egy adott amplitúdójú és frekvenciájú szinuszos feszültség hullámot, akkor az a rendszer 2-es portján milyen amplitúdójú és fázisú kimenő feszültség hullámot eredményez. Az S_{21} komplex fázisa a két feszültség hullám relatív fázisát fejezi ki, ez számunkra nem érdekes, ezért vesszük az abszolút értéket. Ha $|S_{21}|(f) = 1$, akkor a szűrő az adott f frekvencián nem csillapít, de nem is erősít, tehát tökéletesen átereszt a rajta keresztülküldött jelet, legfeljebb más fázisban. Ez lenne a szűrő ideális viselkedése áteresztő sávban. Ha pedig $|S_{21}|(f) = 0$, akkor a beküldött jelből semmit nem enged át, azt általános esetben vagy elnyeli, vagy visszaveri, de mivel mi veszteségmentes, LC szűrőkkel foglalkoztunk, itt az elnyelés nem jöhet szóba. Ez utóbbi pedig az ideális viselkedés záró sávban. Természetesen véges frekvencián és véges elemértékekkel egyik szélső eset sem megvalósítható, még az ideális LC komponensekkel sem, amelyekkel mi számolunk. A szűrő tehát mindig valamilyen 0 és 1 közötti $|S_{21}|$ -et fog mutatni bármilyen frekvencián és bármilyen behelyettesített elemértékek mellett.

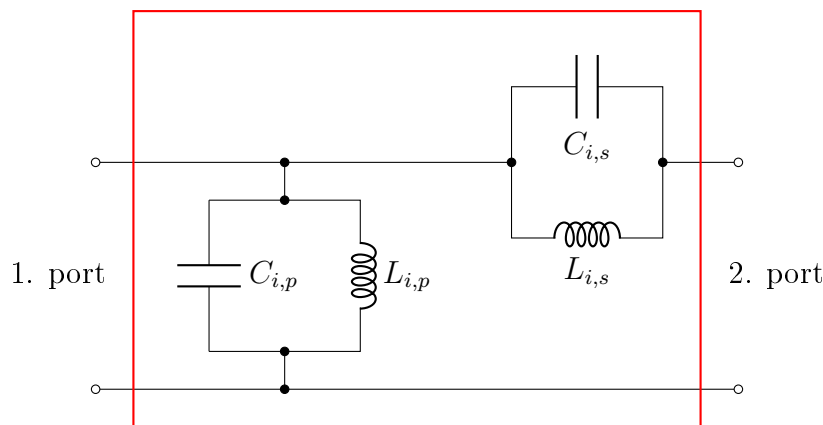
A célfüggvényünket, vagyis $c(\underline{x})$ -t, ahol \underline{x} a behelyettesített elemértékek vektora, egy összeggel definiáljuk. Az összegben diszkrét frekvenciákon felvett hibaértékeket, $h(f, \underline{x})$ -ket adunk össze.

$$c(\underline{x}) = \sum_{i=1}^{N_f} h(f_i, \underline{x}) \quad (1)$$

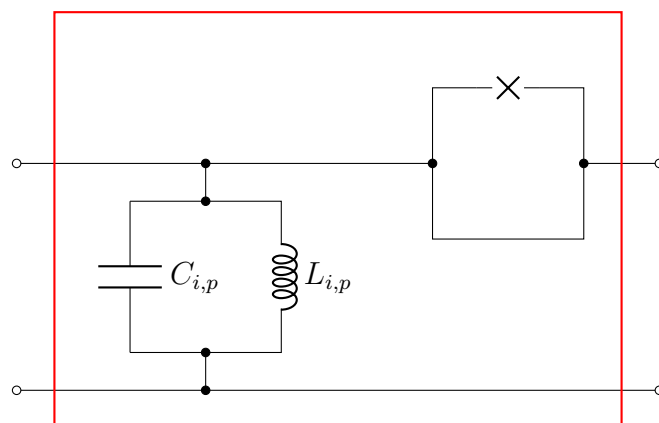
A vizsgált diszkrét f_i frekvenciákat logaritmikusan vettük fel a legkisebb és a legnagyobb specifikált frekvencia között, a kiértékelt frekvenciák száma, N_f megadható a specifikáció definiálásakor a programunkban. Ha egy adott frekvencián a kiszámolt $S(f) = |S_{21}|(f)$ kisebb, mint a megadott maximum, $S_{max}(f)$ vagy hasonlóan, ha nagyobb, mint a megadott minimum, $S_{min}(f)$, akkor a célfüggvény értékét nem változtatjuk az f_i frekvencián mutatott viselkedés alapján, itt nem büntetünk. Ez egyben azt is jelenti, hogy az adott diszkrét frekvencián megfelel a specifikációnak a szűrő. Ellenkező esetben minél jobban belóg az $S(f)$ a tiltott zónába, annál nagyobb abszolút értékű negatív szám lesz $h(f, \underline{x})$ értéke. A fent vázolt összegzős módszerrel egy olyan mennyiséget kapunk, ami az $S(f)$ átviteli függvény tiltott zónákba való belógásának integrálját közelíti. Hogy ez a belógás, vagyis a $h(f, \underline{x})$ pontosabban mit takar, később fejtjük ki a 3. fejezetben.

2.1. A hálózat kapcsolási rajza

A hálózatnak gyakorlatilag a kapcsolási rajza adott L és C elemértékekkel, innen kell valahogyan eljutni az $|S_{21}|$ paraméterig. Olyan struktúrájú szűrőt választottunk, amivel szűrőkarakterisztikák széles köre megvalósítható. Alapvetően egy periodikus szerkezetű



1. ábra. Az optimalizálandó szűrő i -edik egysége

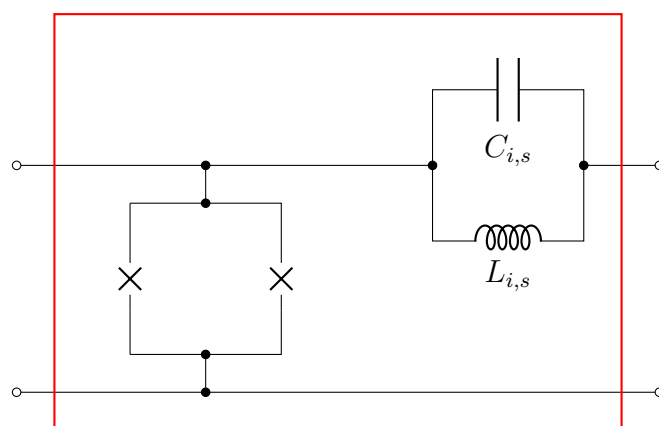


2. ábra. Az egységálózat párhuzamosan kötött párhuzamos LC rezgőkörnek használva. Itt érdekesség, hogy a soros ágba egyik alkatrészt sem használjuk, mégis összesen egy rövidzárral vannak helyettesítve, mivel ennek a helyettesítésnek van értelme.

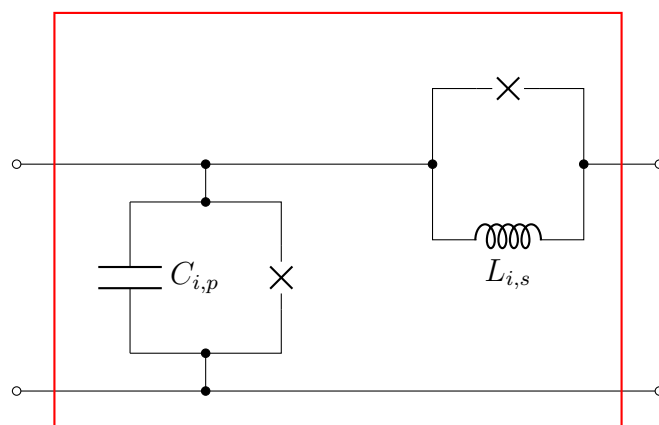
kapcsolásról van szó, de az egyes blokkokban, amik egymás után vannak kapcsolva, külön meg lehet adni az elemértékeket, eleve azt is, hogy az adott pozícióban van-e kondenzátor vagy induktivitás. A programkódunk végleges verziójában legfeljebb 3 db ilyen blokk használható, de a kód minimális módosításával bővíthető még több blokkal szükség szerint. Egy ilyen blokk kapcsolási rajza az 1. ábrán látható. Ahogy az ábrán is látszik, az egységálózat egy kétkapu, amiben egy párhuzamos és egy soros LC rezgőkör van. Ez csak abban az esetben néz ki pontosan így, ha mind a négy paraméterértéket beállítjuk. Ha egyes komponenseket nem használunk, akkor azokat szakadással, vagy rövidzárral helyettesíti a program, ahogy annak a teljes hálózat szempontjából értelme van. Erre láthatunk néhány példát a 2. 3. és 4. ábrán.

2.2. Impedanciák és admittanciák

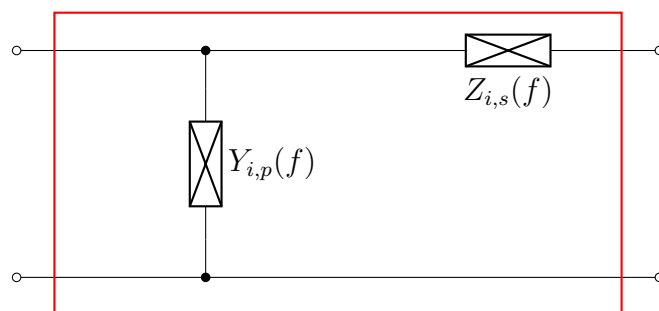
A célfüggvény számításának logikusan következő lépése, hogy az elemértékekből egy-egy diszkrét frekvencián helyettesítő impedanciát számolunk. A kódban egyébként az egyes párhuzamos rezgőkörökre direkte határozzuk meg annak az eredő impedanciát (vagy admittanciát) az *Impedance*() és *Admittance*() függvények segítségével. Ezek után a helyettesítőkép az 5. ábrán látható módon alakul. Itt a paraméterértékek az alábbi képletek alapján adódnak. Természetesen külön le vannak kezelve azok az esetek, amikor valame-



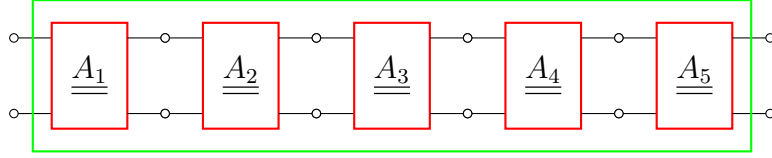
3. ábra. Az egységálózat sorosan kötött párhuzamos LC rezgőkörnek használva.



4. ábra. Az egységálózat elsőrendű LC aluláteresztő szűrőnek használva.



5. ábra. Helyettesítő impedancia és admittancia



6. ábra. A hálózat egységeinek kaszkád kapcsolása.

lyik impedancia vagy admittancia 0-ra vagy ∞ -re adódik és nem lehet direkt behelyettesíteni a képletekbe.

$$Z_{i,p,L} = i2\pi f \cdot L_{i,p} \quad (2)$$

$$Z_{i,p,C} = \frac{1}{i2\pi f \cdot C_{i,p}} \quad (3)$$

$$Z_{i,s,L} = i2\pi f \cdot L_{i,s} \quad (4)$$

$$Z_{i,s,C} = \frac{1}{i2\pi f \cdot C_{i,s}} \quad (5)$$

$$Z_{i,s} = \frac{Z_{i,s,C} \cdot Z_{i,s,L}}{Z_{i,s,C} + Z_{i,s,L}} \quad (6)$$

$$Y_{i,p} = \frac{Z_{i,p,C} + Z_{i,p,L}}{Z_{i,p,C} \cdot Z_{i,p,L}} \quad (7)$$

2.3. Láncparaméterek

A következő lépés egy köztes számítás, ami az egyes egységblokkok kaszkádkapcsolásának számítását könnyíti meg. Ehhez a láncparamétereket hívjuk segítségül, amelyek az angol szakirodalomban ABCD-paraméterekként szerepelnek. A láncparaméterek mátrixa, vagyis $\underline{\underline{A}} \in \mathbb{C}^{2 \times 2}$ lényegében a leírt rendszer két portján felvett feszültségek és áramok között teremt kapcsolatot, az impedanciaparaméterekhez (Z-paraméterek) nagyon hasonló módon. Egy adott frekvencián ezzel a 2×2 -es mátrixszal teljesen leírható a hálózat. A feszültségek és áramok referenciairányai úgy vannak megválasztva a láncparaméterek felírásakor, hogy a kaszkádba kapcsolt kétportú hálózatok eredő láncparamétereit az összekapcsolt alegységek $\underline{\underline{A}}$ mátrixainak szorzataként kaphatjuk meg, ahogy az a 6. ábrán látható.

$$\underline{\underline{A}} = \prod_i \underline{\underline{A}}_i = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (8)$$

2.4. Szórási paraméterek

Végül elérkeztünk a szórási paraméterek számításához. A 2.3. pontban tárgyalt láncparaméterek használatát az is indokolja, hogy létezik egyszerű képlet az $\underline{\underline{A}}$ mátrixból $\underline{\underline{S}}$ mátrixba való áttérésre. Az $\underline{\underline{S}}$ mátrix minden elemére külön analitikus képlet van [1],

de mivel nekünk csak az S_{21} -re van szükségünk, így csak az erre a konkrét paraméterre vonatkozót kellett használnunk. Az S-paraméterek használatához minden portra definiálni kell egy normalizáló impedanciát, Z_0 -t. Ezt mi a konvencionálisnak számító $50\ \Omega$ -ra választottuk.

$$S_{21} = \frac{2}{A + B/Z_0 + CZ_0 + D} \quad (9)$$

3. A célfüggvény javítása

A $c(\underline{x})$ célfüggvény a félév során átesett néhány változtatáson. Ezt az indokolta, hogy a naívan, első próbálkozásként felírt célfüggvény több szempontból sem viselkedett jól az optimalizálás szempontjából. Ebben a fejezetben ezeket a fejlesztési lépéseket írjuk le.

3.1. Az első verzió és a problémái

Eőször a következőképpen definiáltuk a célfüggvényt felépítő $h(f, \underline{x})$ hibafüggvényt.

$$h(f, \underline{x}) = \begin{cases} \frac{S}{S_{min}} - 1 & \text{ha } S < S_{min} \\ 1 - \frac{S}{S_{max}} & \text{ha } S > S_{max} \\ 0 & \text{egyébként} \end{cases} \quad (10)$$

Erre a felírára teljesül, amit fentebb célként vázoltunk, vagyis, hogy minél jobban átlóg valamelyik határon a függvény görbéje, annál jobban büntetjük és az egyáltalán nem belógó függvényt nem büntetjük. A hibafüggvényt is változtattuk később, de tapasztalatunk szerint már akár ilyen formában is használható lenne. Ebben a formában a $c(\underline{x})$ függvénnyel a fő probléma, hogy az \underline{x} vektorban egyszerűen a tekercsek induktivitásai és a kondenzátorok kapacitásai szerepeltek H-ben és F-ban. Ez magában még nem hangzik problémásnak, de ehhez azt is figyelembe kell vennünk, hogy az eredményes optimalizáláshoz *nagyságrendileg* elég széles tartományon belül kell megválasztania az optimalizáló algoritmusnak a következő kipróbálandó paraméterértéket. Az *Optimizer* objektum deklarálásakor meg kell adnunk a változtatható függvényparaméterek neveit és mindegyikhez azokat a határokat, amelyek közötti értékeket behelyettesíthet az optimalizálás során. Ezt a következőképpen lehet megadni a 4. ábrán vázolt elsőrendű aluláeresztő szűrő esetén, ha a kapacitás értékét 10^{-6} – 10^{-3} F között, az induktivitás értékét pedig 10^{-5} – 10^{-2} H között engedjük meg.

```

1 bounds = {'par1C': (1e-6, 1e-3), 'ser1L': (1e-5, 1e-2)}
2 optimizer = BayesianOptimization(
3     pbounds=bounds,
4     ...
5 )

```

A fent leírt alsó és felső határok egészen realiztikusak lehetnek, az optimalizációnak mégis nehezen tudja kezelni az ilyen eseteket. Ezt az okozza, hogy az intervallumok 0-hoz közeli végénél kis paraméterérték-változtatásra is nagyon jelentősen változik az átviteli karakterisztika, ez által a hibafüggvény és a célfüggvény értéke is. Más szóval a 0-közeli paraméterértékeknél nem tud az optimalizáló rátanulni a hirtelen nagyon megnövekvő parciális deriváltak miatt a célfüggvényre, pedig a feladat szempontjából ezek a kis abszolútértékű paraméterértékek nagyon is fontosak lennének.

Ez a probléma a kód futása során egyszerűen úgy jelentkezett, hogy ha mondjuk egy elemérték az 1×10^{-6} – 1×10^{-2} tartományban változhatott, akkor az optimalizáló egyáltalán nem is próbált ki olyan értékeket, amelyek körülbelül a 1×10^{-6} – 1×10^{-4} tartományba esnek.

3.2. A paraméterek logaritmizálása

Az előző pontban vázolt, sok nagyságrendet átötlő intervallumok problémájára találtunk megoldást. Ez gyakorlatilag annyi, hogy nem az elemértékeket adjuk meg a célfüggvény paramétereiként, hanem ezek (természetes alapú) logaritmusait. Ezek alapján majd a célfüggvény belül egyből visszaszámolja az eredeti elemértékeket és változatlanul számol tovább, csak az változik, hogy a célfüggvény összességében sokkal simább lesz és az optimalizáló jobban tudja kezelni. Ekkor a paraméter-határértékek megadása a következőképpen néz ki.

```
1 import math
2 bounds = { 'par1C': (math.log(1e-6), math.log(1e-3)),
3             'ser1L': (math.log(1e-5), math.log(1e-2)) }
4
```

Ezzel az említett mintavételezési probléma megoldódott.

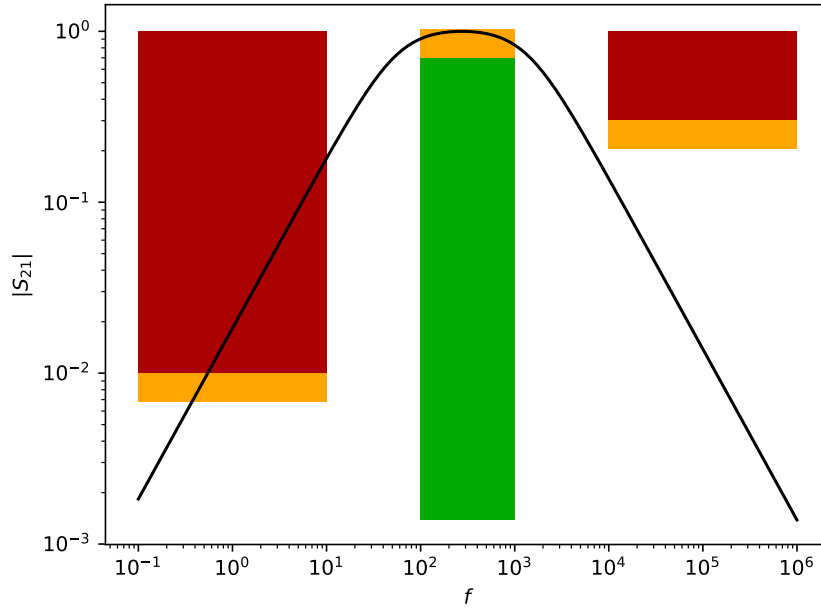
3.3. A függvényérték logaritmizálása

Egy másik, kevésbé jelentős problémára nyújtott megoldást a függvény kimenetének logaritmizálása. Ez egy konstans szorzó erejéig ekvivalens azzal, hogy dB-ben számoljuk ki az eltérést S_{max} és S_{min} , illetve az aktuális $S = |S_{21}|$ között. Ez a konstans szorzó a különböző logaritmus alapokból származik a dB skála és az általunk használt természetes alapú logaritmus között, de nincs jelentősége, mivel ilyen jellegű változtatásra érzéketlen az optimalizálónk. A függvényérték logaritmizálásával azt értük el, hogy loglog skálán ábrázolva az átviteli függvényt és a tiltott régiókat ténylegesen görbe alatti terület jellegű lesz a célfüggvény.

3.4. Büntetett margó

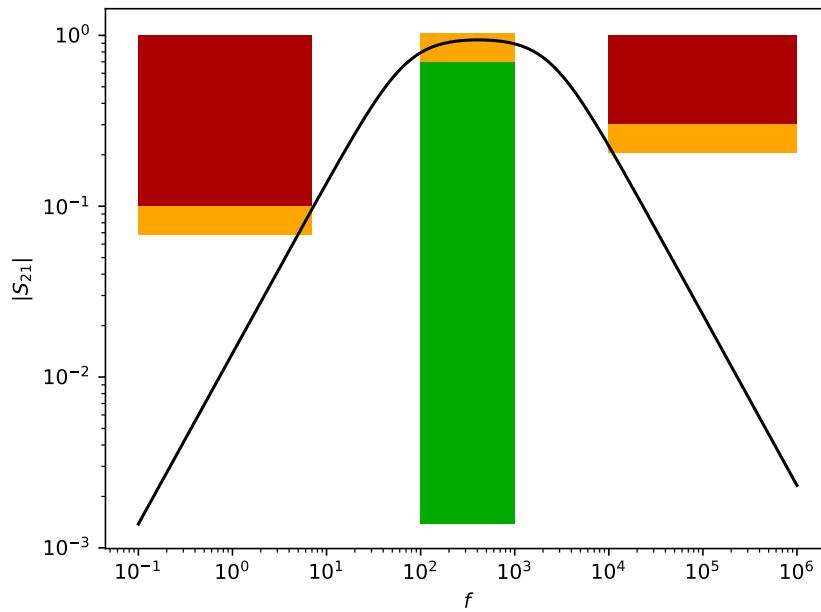
Az optimalizálás segítése érdekében bevezettünk egy opcionális `margin` nevű paramétert, amit a `Spec` objektum példányosításakor kell megadni. Ez egy további büntetett régiót ad az áteresztő sávok tiltott régiója fölé, illetve a záró sávok tiltott régiója alá. Ezzel effektíve azt érzük el, mintha ennyivel szigorúbbak lennének a határértékeink. A büntetett margót a 7. ábra szemlélteti. Az ábrán a bal oldali (kisebb frekvenciákhoz tartozó) záró sávban jól láthatóan az eredeti specifikáció által definiált tiltott régióba is belelóg az átviteli függvény. Emiatt természetesen ennek a sávnak nemnulla hozzájárulása van a célfüggvényhez. Az áteresztő sávban az eredeti specifikációnak ugyan megfelel az átviteli függvény, de a margóba még belóg, így itt is nemnulla büntetés adódik. A jobb oldali záró sávnak még a margóját is elkerüli az átviteli függvény, így ennek a sávnak 0 a célfüggvényhez a hozzájárulása.

A margó bevezetését az indokolta, hogy jobban fel lehessen oldani az olyan helyzeteket, amikor az átviteli függvény ugyan teljesíti az eredeti specifikációt, de egyes frekvenciákon csak éppen, mivel ilyenkor valós alkalmazásoknál érdemes valamennyi tartalékot rászámolni a tervezésnél. Ilyenkor a margó lehetővé teszi, hogy nagyjából egyforma mértékben közelítse meg a határértékeket az átviteli függvény, ne pedig az egyiknél nagyon nagy



7. ábra. Három specifikált frekvenciasáv, két záró (piros) és egy áteresztő (zöld) sáv, valamint a büntetett margó használata.

tartalékkal, a másikinál pedig kis tartalékkal haladjon el. A margó célszerű használatakor olyan megoldás születik, amikor a hálózat a specifikációnak, de a legtöbb sávhoz tartozó margóba kicsivel belelóg. Erre mutat példát a 8. ábra, még ha csak egy nagyon egyszerű esetre is, amit az optimalizáló algoritmus talált meg.



8. ábra. A margó ideális használata.

4. Példák

Az előző fejezetben egy sáváteresztő jellegű szűrőn keresztül mutattuk be a margókat, ebben a részben néhány további szűrőtípus optimalizálását mutatjuk be. Az egyik eset egy egyszerűbb aluláteresztő szűrő, a másik egy összetettebb, záró-áteresztő-záró-áteresztő struktúra.

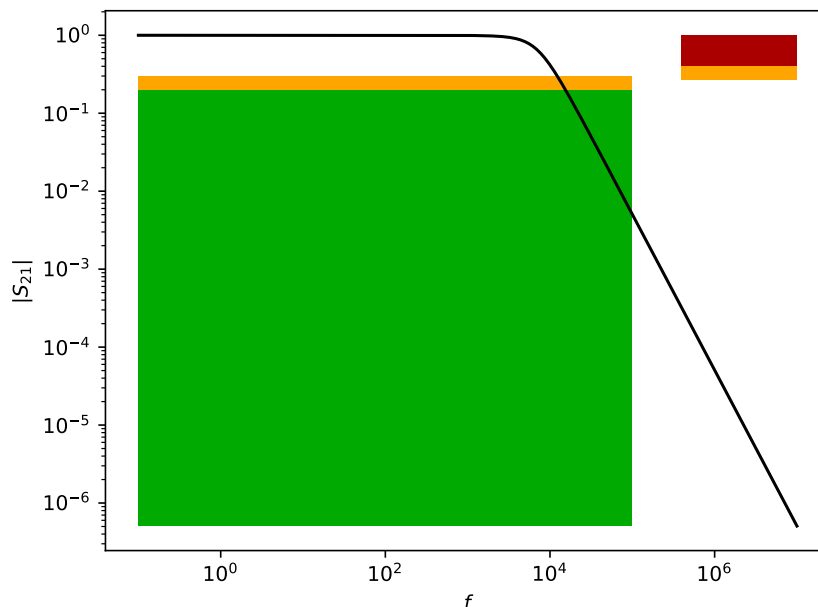
4.1. Aluláteresztő szűrő

Egy ilyen szűrőtípus esetén egy áteresztő és egy zárósávot kell definiálnunk, közte egy toleranciasávval – ennek hiányában nem adható helyes megoldás. Egy ilyen egyfokozatú szűrőnél legegyszerűbb esetben egyetlen elemet kell felhasználnunk, ez azonban nem biztosít elég meredek szűrőkarakterisztikát, továbbá túlzottan lekorlátozza a megoldható problémák halmazát.

Ha kételemű aluláteresztő szűrőt tervezünk, egy párhuzamos kondenzátor és egy soros tekercset kell méreteznünk. Ennek a szemléletes magyarázata, hogy kis frekvenciákon a kondenzátornak nagy, a tekercsnek kicsi impedanciája van, ezáltal kis csillapítást okoz, nagy frekvenciákra pedig fordítva, jelentős csillapítást okoz. A specifikált frekvenciatartományokat, a megválasztott limiteket és a frekvenciatartományok típusát az alábbi kódrészlet mutatja.

```
1 starts = [0.1, 400e3]
2 ends = [100e3, 10e6]
3 limits = [0.2, 0.4]
4 directions = ["pass", "stop"]
5 spec = Spec(starts, ends, limits, directions, margin=0.47, n=1000)
6
```

Felvesszünk $L = 1$ mH és $C = 1$ μ F paraméter értékeket kiindulásnak, de az eredmény 9. ábrán láthatóan nem felel meg a követelményeknek. Meghívjuk ezután az optimalizálót,



9. ábra. Aluláteresztő szűrő kezdeti paraméterekkel.

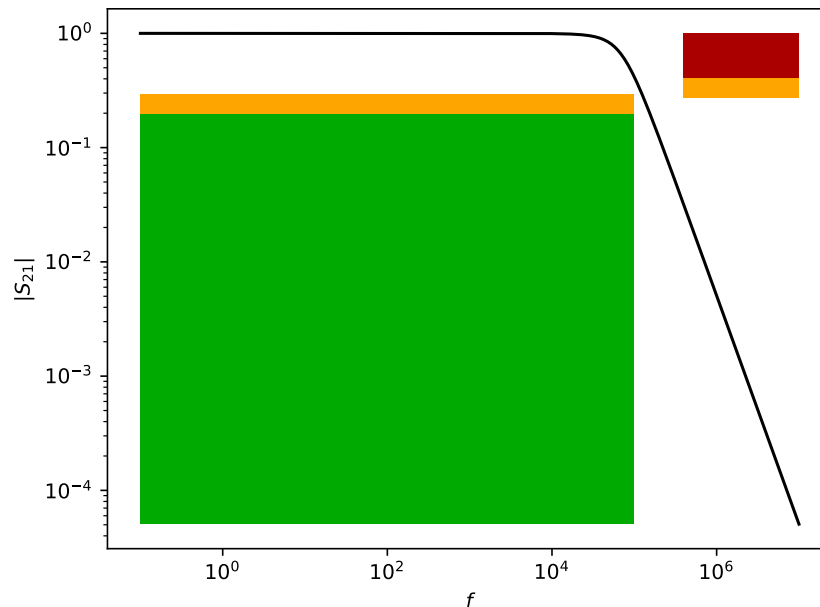
úgy, hogy a paramétereket az inicializált értékének ± 1 dekádos tartományában keresse, a határok megadását mutatja az alábbi kódrészlet.

```

1 bounds = {'par1C': (math.log(1e-7), math.log(1e-4)), 'ser1L': (math.log(1e-4), math.log(1
2           e-2))}

```

Az eredményül adódott szűrőkarakterisztikát a 10. ábra mutatja – megjegyezzük, hogy ennek a méretezését kézzel is meg lehet tenni. Az L és C paraméterek egyikének lefixálása után a Thomson-formulát kell használni: $f_{rez} = \frac{1}{2\pi\sqrt{LC}}$, ami megadja a rezonanciafrekvenciát, esetünkben a töréspontot.



10. ábra. Aluláteresztő szűrő optimalizált paraméterekkel, ami így megfelel a specifikációnak.

4.2. Összetett példa

Vizsgáljunk meg egy összetettebb példát, ahol záró-áteresztő-záró-áteresztő sávok vannak, meglehetősen keskeny toleranciasávval elválasztva. A frekvenciatartományok kezdeteit, végeit, a limiteket és a sávok típusait a következőképp választottuk meg:

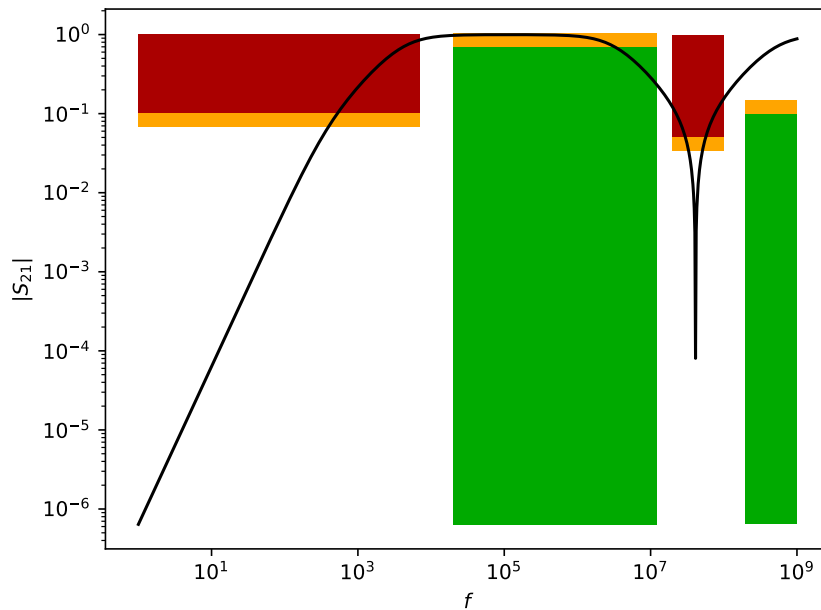
```

1 starts = [1, 2e4, 20e6, 200e6]
2 ends = [7e3, 12e6, 100e6, 1e9]
3 limits = [0.1, 0.7, 0.05, 0.1]
4 directions = ["stop", "pass", "stop", "pass"]
5 spec = Spec(starts, ends, limits, directions, margin=0.47, n=1000)
6

```

Az első zárósáv 7 kHz-es vége és következő áteresztő sáv 2 MHz-es kezdete között egy felüláteresztő jellegű, az előzőhöz hasonló kételemű blokkot használunk. A különbség itt a felüláteresztő viselkedés, azaz kis frekvenciákra kis átvitelt követelünk meg, ezt egy párhuzamos tekercsel és egy soros kondenzátorral tehetjük meg. A második fokozatban egy sávzáró szűrőt valósítunk meg, ugyanis a harmadik tartományunk előtt és után is áteresztőtartományok vannak. Ehhez első körben egy soros LC kört használunk, azaz a párhuzamos ág elemeire nincs szükség. Egy soros LC kör rezonanciafrekvencián nagy impedanciát mutat – ideális esetben végtelen nagyot – ami kis átvitelt, tehát az $|S_{21}|$ grafikonon egy leszívást mutat.

A fent látható inicializálóértékekkel a szűrőkarakterisztika elmetszi a záró és az át-eresztő sávokat, amint azt a 11. ábrán láthatjuk. Válasszuk meg a határokat, majd engedjük rá az optimalizálót.

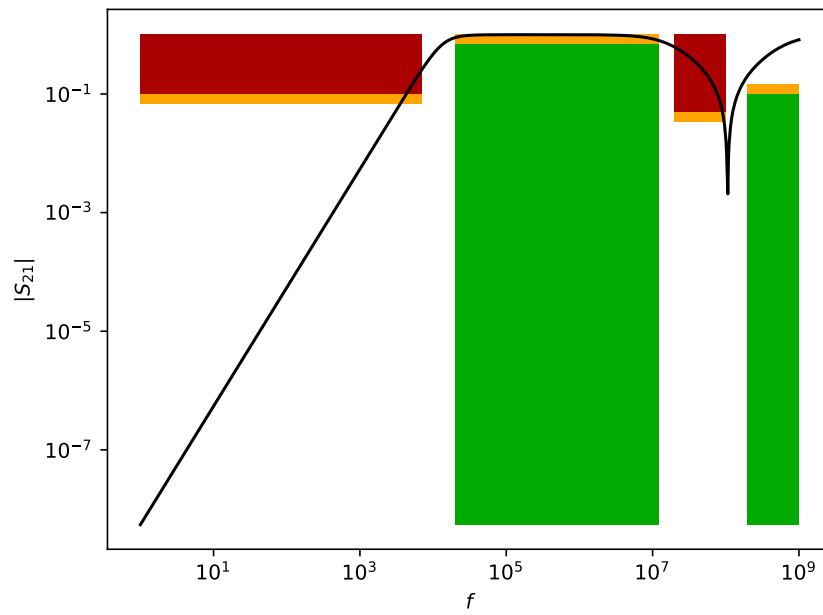


11. ábra. Összetett szűrő kezdeti paraméterekkel.

```
1 spec.cost(plot=True, par1L = math.log(9e-4), ser1C = math.log(9e-6), ser2L = math.log(5e
2 -6), ser2C = math.log(3e-12))
```

A 12. ábrán látható eredmény optimális, de nem helyes, a szűrőkarakterisztika mindkét zárótartományt elmetszi. A középső átteresztősávot viszont helyesen kikerüli, továbbá az első zárósávba is minimalizálta a belógást. Ezt akár meg is lehet szüntetni, ha meredekebb, 3 elemű felülatertesztő szűrő méretezünk rá, a sávzárónál ez már nem ilyen könnyű. Vélhetően specifikáció követel meg túl sokat, a további leszívások, azaz soros *LC* tagból álló fokozatok beiktatása javulást még okozhat, de a maradéktalan betartás nehéz lenne.

A második példa arra világít rá, hogy a neurális hálós optimalizáló nem csodaszer, csak a kezelője által meghatározott tartományon tud optimalizálni a megadott komplexitásig. A méretezés hosszadalmas folyamatát le tudja rövidíteni, de a tervező dolga figyelembe venni a szűrőkarakterisztika alapvető tulajdonságait.



12. ábra. Összetett szűrő optimalizálás után. Látható, hogy az általunk megadott lehetőségekkel nem lehet megfelelni a követelményeknek.

Hivatkozások

- [1] Láncparaméter – S-paraméter konverzió (elérve 2023.06.07.): <https://www.rfwireless-world.com/Terminology/abcd-matrix-vs-s-matrix.html>
- [2] BayesianOptimization python könyvtár GitHub oldala. (Elérve: 2023.06.07.) <https://github.com/bayesian-optimization/BayesianOptimization>
- [3] Scipy: L-BFGS-B minimalizáció dokumentációja <https://docs.scipy.org/doc/scipy/reference/optimize.minimize-lbfgsb.html>

A. Az általunk írt specifikáció osztály

```
1 import numpy as np
2 import math
3 import matplotlib.pyplot as plt
4 from matplotlib.patches import Rectangle
5 from matplotlib.backends.backend_pdf import PdfPages
6
7 pi = math.pi
8 inf = math.inf
9 nan = math.nan
10
11 #impedance of the series LC subcircuit
12 def Impedance(f, lnC, lnL):
13     match [math.isnan(lnC), math.isnan(lnL)]:
14         case [True, True]:
15             return 0.0
16         case [True, False]:
17             L = math.exp(lnL)
18             return f*1j*2*pi*L
19         case [False, True]:
20             C = math.exp(lnC)
21             return -1j*1/(f*2*pi*C)
22         case _:
23             C = math.exp(lnC)
24             L = math.exp(lnL)
25             ZC = -1j*1/(f*2*pi*C)
26             ZL = 0+f*1j*2*pi*L
27             return ZL*ZC/(ZL+ZC)
28
29 def Admittance(f, lnC, lnL):
30     match [math.isnan(lnC), math.isnan(lnL)]:
31         case [True, True]:
32             return 0.0
33         case [True, False]:
34             L = math.exp(lnL)
35             return 1/(f*1j*2*pi*L)
36         case [False, True]:
37             C = math.exp(lnC)
38             return f*2*pi*C*1j
39         case _:
40             C = math.exp(lnC)
41             L = math.exp(lnL)
42             ZC = -1j*1/(f*2*pi*C)
43             ZL = 0+f*1j*2*pi*L
44             return (ZL+ZC)/ZL*ZC
45
46 class Spec:
47     def __init__(self, starts, ends, limits, directions, margin, n):
48         # starts of pass- or stopbands in Hz
49         self.starts = starts
50         # ends of start- or stopbands in Hz
51         self.ends = ends
52         # pass- or stopband threshold values
53         self.limits = limits
54         # pass- or stopbands? possible values: "pass" or "stop" strings
55         self.directions = directions
56         # number of frequency points
57         self.n = n
58         # margin to still punish solution that only barely satisfies the specification
59         # on stopbands, with limit l and margin m, the margin range is from l*(1-m) to l
60         # on passbands, with limit l and margin m, the margin range is from l to l*(1+m)
61         self.margin = margin
62
63     #def cost(self, types, values):
64     def cost(self, plot=False, par1C=nan, par1L=nan, ser1C=nan, ser1L=nan, par2C=nan,
65             par2L=nan, ser2C=nan, ser2L=nan, par3C=nan, par3L=nan, ser3C=nan, ser3L=nan):
66         """Parameters define a ladder structure,
67         where each series or parallel element consists of
68         two discrete, ideal L or C components in parallel
69         with each other. Default values define a perfect
70         all-pass filter. In the function parameters, "parXY" means
```

```

70     the value of the Y-th sub-component of the X-th parallel
71     element. Similarly, "serXY" means the value of the Y-th
72     sub-component of the X-th series element. The ladder starts
73     with a parallel element.""
74     values = [par1C, par1L, ser1C, ser1L, par2C, par2L, ser2C, ser2L, par3C, par3L,
ser3C, ser3L]
75     nval = len(values)
76     nladder = int(nval/4)
77     minf = (min(self.starts))
78     maxf = (max(self.ends))
79     # frequency axis sample points (log spacing)
80     faxis = []
81     # no. of frequency sample points
82     #n = self.n
83     factor = (maxf/minf)**((1/self.n))
84     for i in range(self.n+1):
85         faxis.append(minf*factor**(i))
86     # reference impedance in Ohm, on both ports
87     Z0 = 50
88     Y0 = 1/Z0
89     # array of overall S21 values at the frequency sample points
90     S21 = []
91     #process 1 parallel and 1 series element:
92     for f in faxis:
93         # ABCD parameter matrix of the whole system
94         ABCD = np.matrix([[1,0],[0,1]])
95         for l in range(nladder):
96             # admittance of the two parallel components together from the current
step of the ladder
97             Y = Admittance(f, values[4*l+0], values[4*l+1])
98             mPar = np.matrix([[1, 0],[Y, 1]])
99             Z = Impedance(f, values[4*l+2], values[4*l+3])
100             mSer = np.matrix([[1, Z],[0, 1]])
101             ABCD = ABCD*mPar*mSer
102             # 2/(A+B/Z0+C*Z0+D)
103             A = ABCD.item(0,0)
104             B = ABCD.item(0,1)
105             C = ABCD.item(1,0)
106             D = ABCD.item(1,1)
107             S21.append(abs(2/(A+B/Z0+C*Z0+D)))
108         # natural log of margin+1
109         lnmargin=math.log(self.margin+1)
110         if(plot):
111             # green: passband; red: stopband; orange: ok, but close to not ok, still
punished
112             fig, ax = plt.subplots()
113             minS21 = min(S21)
114             for i in range(len(self.starts)):
115                 if(self.directions[i] == "pass"):
116                     # region forbidden by the original specification
117                     ax.add_patch(Rectangle((self.starts[i], minS21),
118                                             self.ends[i]-self.starts[i],
119                                             self.limits[i]-minS21,
120                                             facecolor='#00aa00'))
121                     # region close to original limit, but satisfying it, additional
penalty region
122                     ax.add_patch(Rectangle((self.starts[i], self.limits[i]),
123                                             (self.ends[i]-self.starts[i]),
124                                             self.limits[i]*self.margin,
125                                             facecolor='orange'))
126                 else: # "stop"
127                     # region forbidden by the original specification
128                     ax.add_patch(Rectangle((self.starts[i], self.limits[i]),
129                                             self.ends[i]-self.starts[i],
130                                             1.0-self.limits[i],
131                                             facecolor='#aa0000'))
132                     # region close to original limit, but satisfying it, additional
penalty region
133                     ax.add_patch(Rectangle((self.starts[i], self.limits[i]*(1/(1+self.
margin))),
134                                         (self.ends[i]-self.starts[i]),
135                                         self.limits[i]*(1-1/(1+self.margin)),
136                                         facecolor='orange'))

```

```

137         ax.loglog(faxis, S21, 'k-')
138         plt.ylabel(r'$|S_{21}|$')
139         plt.xlabel(r'$f$')
140         plt.savefig("plot.pdf", dpi=120, format='pdf', bbox_inches='tight')
141         plt.show()
142         cost = 0
143         # number of freq points in the regions where the S21 is specified
144         ncost = 0
145         for i in range(len(self.starts)):
146             for j in range(len(faxis)):
147                 if faxis[j]>self.starts[i] and faxis[j]<self.ends[i]:
148                     ncost = ncost + 1
149                     lnlimit = math.log(self.limits[i])
150                     lnS21 = math.log(S21[j])
151                     if self.directions[i] == "pass":
152                         cost += max(0, min(-lnS21, lnlimit+lnmargin-lnS21))
153                     else: # "stop"
154                         cost += max(0, lnS21-lnlimit+lnmargin)
155         # negative of average cost, for function maximizing
156         return -cost/ncost

```