A Shepard interpoláció bemutatása

Sári Zoltán
Neumann János Informatikai Kar
Óbudai Egyetem
Budapest
Email: sari.zoltan.tamas@gmail.com

2015.03.23.

Kivonat

Az interpolációs technikák az alkalmazott matematika egyik leggyakrabban használt módszerei. Számos felhasználási területen[1] nélkülözhetetlen a többdimenziós alappontok interpolálása a folytonos felület előállítása érdekében. Az interpolálással kapott folytonos függvény segítségével támogatható az adatok elemzése, azonosíthatóak az outlier pontok, meghatározhatóak a gradiensek. Az interpolációs technika megválasztása függ az adatkészlet pontjainak eloszlásától, az alkalmazási területtől illetve az interpoláló függvény elvárt tulajdonságaitól. A többváltozós, szabálytalanul elszórt alappontú interpolációs problémák során hatékonyan alkalmazható Donald Shepard[1] nevével fémjelzett, a dolgozatban bemutatásra kerülő dimenziófüggetlen interpolációs módszer.

A dolgozatot három fő fejezet alkotja. Elsőként általánosan kerülnek bemutatásra a többváltozós interpolációs technikák. A következő fejezet a többdimenziós módszerek speciális eseteit, a távolság alapú interpolációkat tárgyalja. Majd a dolgozat fókuszát jelentő Shepard távolság alapú interpolációs technika kerül részletesen ismertetésre. Az alap Shepard módszer bemutatását követően a dolgozat kitér a technika értékelésére, előnyeinek és hátrányainak bemutatására, végül az egyik legfontosabb javítást, a lokalizált módszert részletezi.

1. Többváltozós interpoláció

A többváltozós interpoláció a multidimenziós (vektor-skalár) függvények közelítését jelenti a numerikus analízisben. Az alapfeladat egy n-dimenziós koordinátarendszer megadott pontjaihoz egy olyan felület keresése, amely mindegyik ismert ponton (alappont) áthalad.

1. Definíció. *A többváltozós interpoláció egy ismeretlen $f : \mathbb{R}^n \to \mathbb{R}$ függvény megtanulása $\mathcal{L} \subset \mathbb{R}^n$ tanítóhalmazon (adatkészlet) az $\mathbf{x} \in \mathcal{L}$ alappontokban ismert $f_i \in \mathbb{R}$ függvényértékek segítségével úgy, hogy $\forall i \in \{1 \ldots |\mathcal{L}|\}$ indexre $f(\mathbf{x}_i) = f_i$ függvényérték teljesüljön. Az eljárás eredményeként tetszőleges $\mathbf{x}' \in \mathbb{R}^n$ pontban megkapható $f(\mathbf{x}')$ függvényérték.*

Elvárás, hogy a függvénynek „jól kell viselkednie”, folytonosnak és differenciálhatónak kell lennie. Az interpoláció továbbá feltételezi a függvény valamilyen szintű simaságát, vagyis hogy a függvény második deriváltjának abszolút értéke sehol se legyen nagyobb egy elfogadható értéknél.

[1] geoinformációs rendszerek, térképészet, geofizika, meteorológia, adatbányászat, számítógépes grafika, képfeldolgozás, adatvizualizáció

A közelítő függvény tulajdonságait meghatározza az interpolálásba bevont adatpontok száma[4]

- **Globális** interpoláció esetén a teljes tanítóhalmaz felhasználásra kerül bármely interpolált érték esetén.
- **Lokális** esetben csak az adott pont szűkebb környezetében fellelhető pontok kerülnek figyelembe vételre. Ugyanakkor számolni kell azzal, hogy magasabb rendű deriváltakban nem lesz folytonos a függvény.
- **Spline** módszer alkalmazása a közelítő-függvény folytonosságának elvárása esetén célszerű. A függvény ebben az esetben egy meghatározott szakaszon értelmezett polinom (tehát lokális), de a változás a polinom együtthatóinál következik be, ezáltal a globális simaság is biztosított.

Az elvárásoknak eleget téve az egyváltozós esetben klasszikusan alkalmazott Lagrange-féle interpoláció során adott fokszámú polinom megkeresése a cél, míg az Hermite-féle interpolációk (melynek speciális esetének tekinthető a Lagrange-interpolációs és a Taylor polinom) során az alappontokban nemcsak függvényértékek, hanem deriváltértékek is megadhatók. A széleskörűen használt polinomiális interpolációs módszerek mellett elterjedtek a trigonometrikus és a racionális függvények bővebb osztályát használó eljárások. A többváltozós interpolációs eljárások speciális technikái a távolság-alapú interpolációs módszerek.

2. Távolság alapú interpoláció

A távolság alapú interpoláció nagy dimenziószám és az adatkészlet pontjai közötti távolság metrika ismertsége esetén bizonyul hasznosnak. Legegyszerűbb esetben a távolság alapú interpoláció egy csúszó ablakos módszer. Ekkor az eljárás lokális, a függvény értéke az interpolálandó pont meghatározott környezetében (szomszédság) található ismert pontok értékeinek számtani átlaga. A szomszédság két alapvető elv szerint határozható meg:

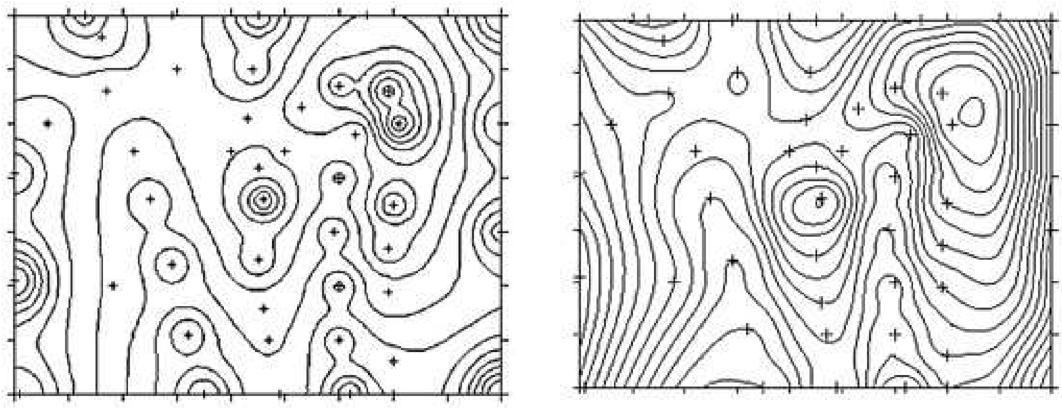
- az interpolálandó pont középpontú térbeli alakzaton belüli alappontok (pl. kör)
- meghatározott számú legközelebbi szomszédok

A módszer hátránya, hogy nem veszi figyelembe távolságok különbözőségét, minden szomszédos pont értéke ugyanakkora súllyal gyakorol befolyást a függvény értékre. Az inverz-távolság súlyozású (inverse distance weighting, IDW) eljárás explicit módon figyelembe veszi az ismert pontok interpolálandó ponttól való távolságát. Minden ismert pont függvény értékre gyakorolt súlya fordítottan arányos az interpolálandó ponttól való távolságával. A függvényértéket az ismert pontok megfigyelési értékeinek a távolság inverzével súlyozott átlaga adja. A módszer globális, az interpoláció során az összes ismert pont felhasználásra kerül.

előnyök	**hátrányok**
a távolságok eltérésének figyelembe vétele az eljárásba integrált: súlyfüggvénnyel szabályozható a távolságok befolyása az interpolálásra	nincs lehetőség az irányfüggő súlyozásra: a térbeli összefüggéseket figyelmen kívül hagyja (pl. magassági pont mellett a gerinc)
lehetővé teszi a gyors kiszámíthatóságot	,,bulls-eyes" effektus: az egyenlő értékeket reprezentáló körszerű alakzatok megjelenése az ismert pontok körül (1. ábra)

1. táblázat. Az alap inverz távolság súlyozású (IDW) eljárás főbb jellemzői [5]

A távolság alapú interpolációs technikák, a módszer egyszerűsége és általános jellege miatt a számos területen alkalmazhatók. Tipikus alkalmazási terület a geoinformációs felhasználás mellett a képfeldolgozás és a gépi tanulás.



(a) nem kívánt ,,Bulls eye” effektus (b) Shepard módosításával csökkentett hatás

1. ábra. A ,,Bulls eye” effektus és kezelése [6]

Néhány további alkalmazási terület[5]

- DNS szekvenciák összehasonlítása
- helyesírás-ellenőrzés során korrekciós javaslatok megkeresése
- adott szövegkörnyezetbe illő szavak javaslata szó-korrelációs analízisből
- színkészlet transzformációja, pl.: véges színkészlet szürkeárnyalatossá redukálása
- adatkészlet topológiai tulajdonságainak felderítése (pl.: klaszterek azonosítása)
- optimalizálási problémák megoldása
- neurális hálózatok kiegészítése csak releváns pontok hozzáadásával

Az eredeti inverz távolság súlyozású interpoláció Shepard nevéhez köthető. Az összes további IDW eljárás a Shepard módszer módosításának tekinthető.

3. Shepard interpoláció

3.1. Az alap Shepard módszer

A Shepard alap módszere egy szabályos rácson, 2 dimenzióban használható inverz távolság-alapú interpoláció. Az általános, dimenziószámtól független forma az $\mathbf{x} \in \mathbb{R}^n$ interpolálandó pontban határozza meg $f \in \mathbb{R}$ értéket az ismert $\mathbf{x_i} \in \mathcal{L}$ alappontok $f_i = f(\mathbf{x_i})$ értékei segítségével:

$$f(\mathbf{x}) = \begin{cases} \frac{\sum w_i(\mathbf{x}) f_i}{\sum w_i(\mathbf{x})} & \text{ha} \quad \forall \mathbf{x_i} \quad d(\mathbf{x}, \mathbf{x_i}) > 0 \\ f_i & \text{ha} \quad \exists \mathbf{x_i} \quad d(\mathbf{x}, \mathbf{x_i}) = 0 \end{cases} \tag{1}$$

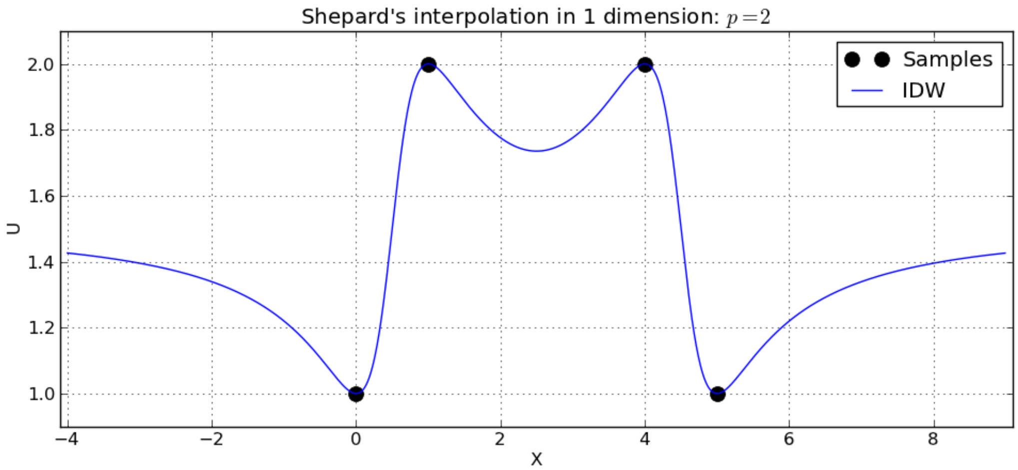
ahol a súlyfüggvény:

$$w_i(\mathbf{x}) = \frac{1}{d(\mathbf{x}, \mathbf{x_i})^p} \tag{2}$$

$d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ az x és x_i közötti távolságot meghatározó metrika és

$p \in \mathbb{R}$ az ún. *power parameter* mellett

Az interpolációs függvény eltolás, forgatás és skála invariáns. A függvény értékeit az összes alappontban vett érték lineáris kombinációjaként veszi fel, az együtthatóknál a megfelelő pont és $\mathbf{x}$ távolságának normalizált reciprokhatványát figyelembe véve. Kivételt képeznek azok az esetek, amikor az interpolációra kijelölt pont egybeesik valamelyik alapponttal. Ekkor az interpolációs tulajdonság érdekében a függvény érték megegyezik az alappontban felvett megfigyelési értékkel. A kivétel következménye a függvénynek az a nem kívánt jellemzője, hogy az alappontokban „kiegyenesedik”, vagyis parciális deriváltja zérus. A tulajdonság nagy alappontszám esetén rendkívül rontja az interpolációval előállítandó felület felismerhetőségét (analízisét).

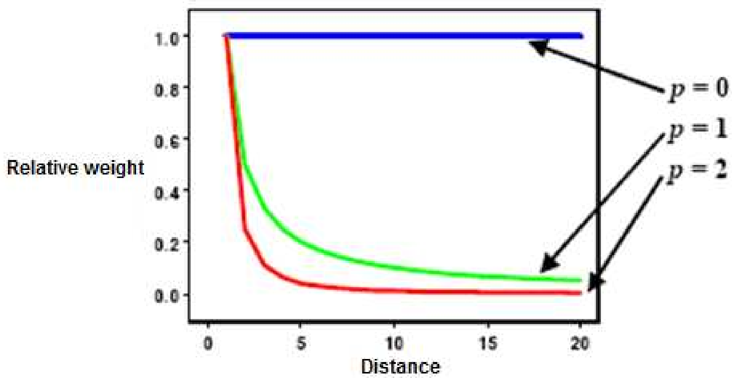


2. ábra. Példa a függvény kiegyenesedésére (zérus derivált) az alappontokban [7]

A módszer a súlyfüggvény két tulajdonságának módosításával finom-hangolható. A súlyfüggvény függ egyrészt a távolságfüggvény (norma) k paraméterének megválasztásától (általában $k = 2$, euklidészi norma):

$$d(\mathbf{x}, \mathbf{x_i}) = \left(\sum_j |x_{ij} - x_j|^k \right)^{\frac{1}{k}} \tag{3}$$

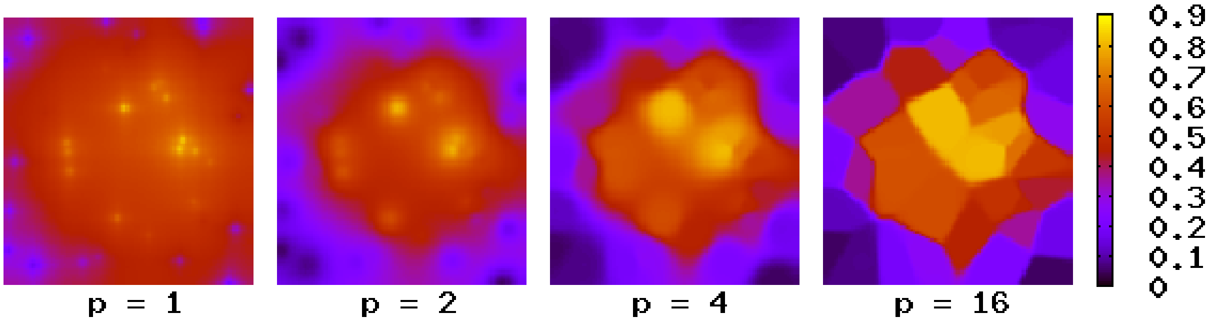
Másrészt a Shepard interpolácó súlyfüggvénye paraméterezhető az inverz távolság p (*power parameter*) kitevő módosításával.



3. ábra. A p power parameter értékének hatása a súlyfüggvényre [8]

A paraméter $p = 0$ értéke esetén az interpoláló függvény értéke konstans, az alappontokban felvett értékek egyszerű számtani átlaga képezi. Magas p érték esetén függvényt csak az interpolált ponthoz nagyon közeli pontok értéke befolyásolja. A p érték ideális megválasztását a kívánt simítási mérték, a tanítóadatok sűrűsége és eloszlása befolyásolja.

A gyakorlatban a paraméter tipikus értéke $p = 2$.



4. ábra. Shepard interpoláció különböző p értékek mellett [7]

Az optimális érték meghatározásához segítségül hívható az interpoláció pontosságát jellemző hibafüggvény minimalizálása (pl: az RMSPE, root mean square prediction error) [2]

A két különböző p paraméterű súlyfüggvény alkalmazására példa, a belső és külső környezet megkülönböztetése. Az alapértelmezett súlyfüggvény kitevője a belső környezetben $p = 2$, míg a külső környezetben $p = 4$. A külső súlyfüggvény szerepe a két környezet határán lévő folytonosság megőrzése.

3.2. Értékelés

A bemutatott, globális Shepard módszer legfőbb előnye az egyszerű megvalósíthatóság és az általános alkalmazhatóság. A formula megfelelő súlyfüggvény megadásával tetszőleges dimenzió számú térben értelmezhető. Az eljárás számításigénye alapesetben $O(n^2)$. A Shepard módszer előnyei mellett több olyan probléma is felmerül, amelyek a súlyfüggvény módosításával kezelhetőek.

probléma	**súlyfüggvény fejlesztése**
globális jelleg miatt magas mintaszám mellett a függvényérték kalkulációja proporcionálisan hosszabb	lokalizáció, csak a függvényérték kiszámítása szempontjából szignifikáns, közeli pontok figyelembe vétele távolságalapú súlyozási faktorral
súlyozásnál csak a távolság számít, az irány nem kerül figyelembe vételre, az interpoláció érzékeny az alappontok csoportosulására	irányfüggő súlyozási faktor bevonása (cosinus függvény), redundáns információkat hordozó, csoportosuló pontok átlagos súlyának mérséklése, a sűrű pontok alacsonyabb súlyt kapnak
az alappontokban a zérus iránymenti deriváltak egy nem-kívánt megszorítást okoznak a felületen a függvény felhasználhatóságát tekintve	meredekség befolyásolása: növekmények hozzáadása a függvényértékhez az alappontok közelében
numerikus instabilitás (kerekítési és csonkítási hiba) két majdnem megegyező alappont esetén	számítási hiba csökkentése: bizonyos minimum eltérésen belül a két érték átlagolása

2. táblázat. Az alap Shepard eljárás fejlesztési irányai [1]

A Shepard-módszer megjelenése óta számos publikáció született az interpoláció különböző javításainak vizsgálata tárgyában, amelynek köszönhetően sok módosított változat látott napvilágot. A következő fejezetben a Shepard interpoláció egyik fontos javítása, a lokalizált változat kerül bővebben bemutatásra.

[2]Ehhez az eredeti interpolációs függvényt approximációs függvénnyé kell módosítani, a $d(\mathbf{x}, \mathbf{x_i}) = 0$ feltétel törlésével

3.3. A lokalizált módszer

Shepard módszere alapesetben globális interpoláció, azonban a súlyfüggvény módosításával lokálissá tehető. A Shepard-interpoláció lokalizált változatának lényege, hogy egy interpolált érték számítása során nem az összes alappont, csak az interpolált ponthoz közeliek kerülnek figyelembe vételre. A módosítással az interpolációból kizárhatók a távoli, nem szignifikáns alappontok, ezáltal gyorsítható az eljárás. Lokalizálás esetén a függvény által előállított felszín nem csak a p power parameter értékétől, hanem a szomszédsági környezet méretezési stratégiájától is függ. A szomszédsági környezet méretét szabályozó kritérium az alábbiak szerint definiálható

- választott távolság alapján (sugár)
- választott alappontszám alapján (a legközelebbi pontok minimális és maximális száma)
- a kettő kombinációjával

A szomszédságra irányuló megszorítás során általánosan használt alakzat az R sugarú gömb. A módosított súlyfüggvény R sugarú gömbi környezetben megtalálható legközelebbi szomszédok figyelembe vétele esetén [2]:

$$w_i(\mathbf{x}) = \left(\frac{max(0, R - d(\mathbf{x}, \mathbf{x_i}))}{Rd(\mathbf{x}, \mathbf{x_i})} \right)^p \tag{4}$$

Shepard álláspontja alapján az interpolációhoz adott pont környezetében vett 4 alappont elegendő, 10-nél több pedig már nem okoz érdemi pontosság javulást. Két dimenzió esetén az átlagosan 7 pontot tartalmazó R sugarú kör meghatározásához irányadó képlet[1]:

$$R = \sqrt{\frac{7A}{N\pi}} \tag{5}$$

ahol

$N-$ az alappontok száma (adatkészlet mérete)

$A-$ az alappontok által kifeszített poligon területe

A szomszédsági környezetet leíró alakzat módosítását igényelheti, ha valamelyik irányban lévő alappontok értéke fontosabb. Ekkor alkalmazható például ellipszis illetve az adatpontok alakzaton belüli elhelyezkedése alapján tovább finomítható a súlyfüggvény (pl. alakzat szektorokra osztható).

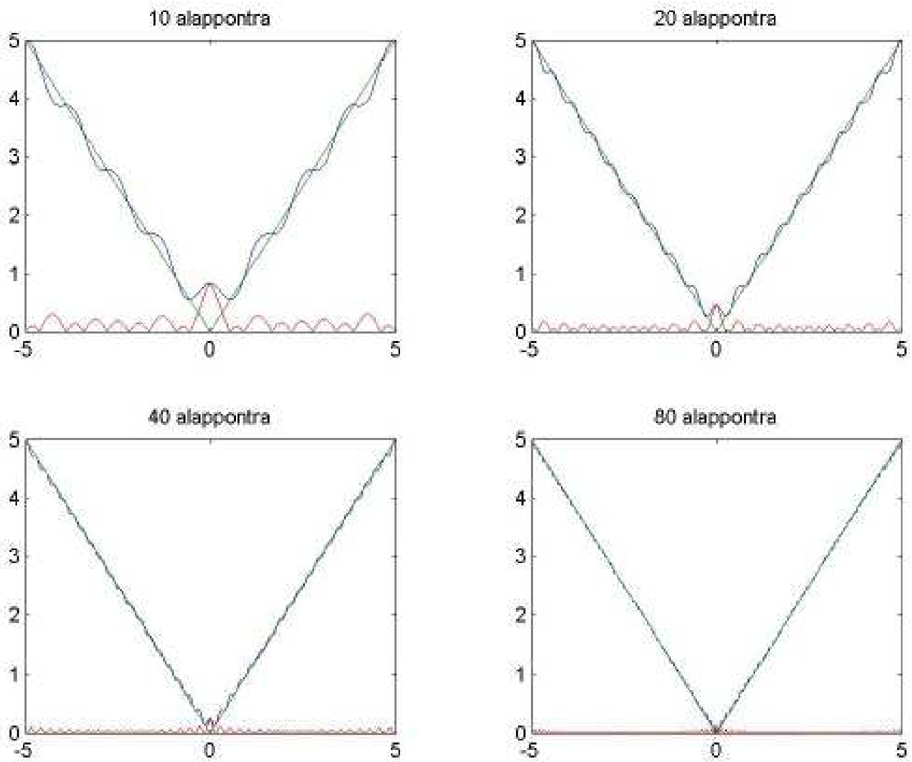
A távolság alapú korlátozás számítása egyszerű, de előfordulhat az a lehetőség, hogy a szabálytalan elhelyezkedés miatt egyes pontok környezetében sok alappont áll rendelkezésre, míg más esetében nem található alappont a megadott sugáron belül. Megoldást jelenthet többféle sugár megadása az alappontok adott környezetben való sűrűségétől függően. A módszer azonban nehézkes, mivel számos kivétel megadását és az alappontok elhelyezkedésének alapos ismeretét feltételezi.

Megoldást a választott alappontszám alapú kritérium alkalmazása kínál. Minden egyes interpolált pontban a távolság alapján rendezett első k darab alappont kerül figyelembe vételre. A módszer előnye, hogy minden pontban egyenletes, előre megadott darabszámú alappontot vesz figyelembe. Hátránya, hogy a rangsorolás rontja a módszer számításigényét.

A két kritérium kombinálása ötvözi az előnyöket. A sugár meghatározása esetén felülírható az alapfeltétel a minimum ill. maximum bevonandó pontok számával. A pontok számának megadása esetén pedig a szabály a minimális/maximális sugarat is tartalmazhatja. A felülbíráló feltételek életbe lépésekor kiterjesztődik vagy szűkül a környezet az alappontok sűrűségétől függően.

A lokalizált Shepard interpoláció fontos jellemző, hogy a lokalizációval a módszer elveszti azt a rossz tulajdonságát, hogy kisimuljon az alappontokban. A korrigált módszer az irreleváns elemek kizárásával a pontosság mellett a számítási bonyolultságot is javítja. Az eljárás gyors térbeli keresési struktúrákkal (pl. kd-tree) kombinálva hatékony, $O(NlogN)$ bonyolultságú interpolációs módszert ad. Ezzel a függvény vállalható költségek mellett illeszkedik a számítógépes alkalmazáshoz. Ugyanakkor az algoritmus implementációja az adatpárhuzamos jelleg kihasználása érdekében párhuzamos erőforrásokat, környezetet igényel.

A lokalizált változat a pontosság konvergenciája szempontjából is hasznosnak bizonyul. A Shepard interpoláció konvergenciáját Besenyői vizsgálta [3]. Az alap módszer egy dimenzióban, ekvidisztáns alappontrendszerben az alappontok számának növelésével egyértelműen konvergált. A hiba logaritmusa az alappontszám logaritmusának nagyjából lineáris függvényeként mutatkozott.



5. ábra. Az interpoláció pontosságának javulása az alappontszámok növelésével (zöld: abszolútérték függvény, kék: interpolációs függvény, piros: eltérés)[3]

A pontosság n-dimenziós esetben már nem konvergált. Az alap Shepard módszer esetében bármennyivel növelve is az alappontok számát nem biztosítható a javulás a pontosságot illetően. A lokalizált változatban azonban mind ekvidisztáns pontrendszerben, mind szabálytalanul elszórt alapontok esetében az alappontok sűrűsödésével a hibák csökkennek.

Hivatkozások

[1] D. Shepard "A two-dimensional function for irregularlly-spaced data" *Proceedings ACM National Conference*, pages 517–524, 1968

[2] R. Franke, G.M. Neilson "Smooth interpolation of large sets of scattered data" *International Journal of Numerical Methods in Engineering*, vol. 15., pages 1691–1704, 1980

[3] D. Besenyői "Lokalizált Shepard-módszerek alkalmazása kétváltozós, szabálytalan alappontú interpolációs problémákban" *Diplomadolgozat ELTE*, 2012

[4] http://hu.wikipedia.org/wiki/Interpol%C3%A1ci%C3%B3

[5] http://demonstrations.wolfram.com/ObtuseAngleShadowingNetworksAndDistanceBasedInterpolation/

[6] http://www.gitta.info/ContiSpatVar/en/html/Interpolatio_learningObject2.xhtml

[7] http://en.wikipedia.org/wiki/Inverse_distance_weighting

[8] http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//00310000002m000000

[9] http://www.ncgia.ucsb.edu/pubs/spherekit/inverse.html