



M Ű E G Y E T E M 1 7 8 2

Budapest University of Technology and Economics  
Faculty of Electrical Engineering and Informatics  
Department of Automation and Applied Informatics



Vrije Universiteit Brussel

Vrije Universiteit Brussel  
Faculty of Engineering  
Department of Fundamental Electricity and  
Instrumentation

# SYSTEM IDENTIFICATION IN HIGHLY NON-INFORMATIVE ENVIRONMENT

---

# RENDSZERIDENTIFIKÁCIÓ ERŐSEN INFORMÁCIÓHIÁNYOS KÖRNYEZETBEN

Ph.D. Tézisfüzet

Kolumbán Sándor

2016. március 11.

Témavezetők:

Prof. István Vajk

Prof. Johan Schoukens



Department of  
Automation and  
Applied Informatics



Vrije Universiteit Brussel  
Departement ELEC

# 1 Előzmények és célkitűzések

A rendszeridentifikáció, mint kutatási terület, továbbra is több irányban fejlődik. A figyelem eltolódott a ma már klasszikusnak számító SISO lineáris modellekről és megoszlik különböző modellstruktúrák és identifikációs helyzetek között. Ezeket a problémákat kísérlettervezési nézőpontból is megközelítik. A terület különböző kutatási irányai közül az egyik azokra a problémákra fókuszál, ahol a mérési idő viszonylag rövid. Ez alatt azt értjük, hogy a begyűjtött adatmennyiség éppen hogy lehetővé teszi értékelhető becslések készítését. A becslés bizonytalanságának kiértékelése ilyen esetekben más eszközöket igényel összehasonlítva azzal, amikor gyakorlatilag végtelen mennyiségű adat elérhető. Az alkalmazható eszközkészlet jelentős részét képezik az eloszlásfüggetlen eljárások, melyeknek fontosságát a 16. IFAC Rendszeridentifikációs szimpózium egyik plenáris előadása is hangsúlyozta [CCGW12]. Ezen módszerek filozófiája megragadta a figyelmemet és nagyban hozzájárult a témaválasztásomhoz.

Kutatásom célja a létező módszerek megértése volt és határaik kitolása, ahol ez lehetséges. A disszertáció e próbálkozásom eredményeit tartalmazza.

## 2 Új tudományos eredmények

Eredményeimet három tézisbe szerveztem. Ezek különböző irányokból közelítik meg az identifikációs problémát, melyek közös pontja valamilyen információ hiánya, amit az identifikációs fősodor adottnak tekint.

A méréseket terhelő zaj eloszlása általában ismertnek feltételezett, ez lehetővé teszi a paraméterbecslések bizonytalanságának értékelését. Ez az értékelés támaszkodhat aszimptotikusan érvényes eredményekre is. Az első tézis véletlen bizonytalanság kiértékelési eljárásokat mutat be, amelyek nagyban merítenek [CCW12a]-ból és kapcsolódó munkákból. Itt hipotézisvizsgálatok egy családját írom le, amiket adatperturbációs (DP) eljárásoknak nevezek. Leírom a feltételeket is, melyeknek teljesülniük kell a metodológia alkalmazhatóságához. Karakterizálom a zaj eloszlások egy családját, amelyek mellett összefüggő és korlátos konfidenciahalmazokat kapunk lineáris regressziós feladatokban bizonyos elégséges gerjesztési feltételek teljesülése esetén. Lineáris dinamikus rendszerek identifikációjára is kiterjesztettem a bevezetett adatperturbációs eljárásokat.

Az identifikációs feladat sok esetben megfogalmazható optimalizálási feladatként, aminek több lokális optimuma is van. A második tézis az ilyen optimalizálási feladatok globális optimumának megtalálására fókuszál. Az identifikációs feladat Lasserre hierarchiákra épülő konvex átfogalmazási lehetőségeit vizsgáltam meg. Explicit kifejezéseket adtam a szemidefinit programozási feladatok méretének növekedésére az identifikációs feladat különböző paraméterezései esetére. Azt a végkövetkeztetést vontam le, hogy számítási kapacitásigénye miatt ez az eljárás csak viszonylag rövid adathalmazok esetén alkalmazható.

A harmadik tézis részben véletlen gerjesztőjelű identifikációs problémákat tárgyal. A véletlen bemenőjelnek a becslés bizonytalanságára gyakorolt hatásait vizsgálom itt. Pontosabban, hogyan csökken ez a hatás a mintaszám növekedésével. Ennek alapján olyan szabályokat fogalmazok meg a mintaszámra vonatkozóan, amelyek teljesülése mellett, nagy valószínűséggel, a becslések bizonytalansága jól becsülhető.

A tézisfüzet hátralevő részében a fenti három tézist részletezem, idézve a hozzájuk kapcsolódó publikációimat sorolom fel.

# I. TÉZIS:

## Adatperturbációs eljárások

A tézishoz kapcsolódó publikációim: [1], [2], [5], [4].

Eloszlásmentes statisztikai módszereket főként a huszadik század első felében tanulmányoztak a statisztikusok. Ezek az eljárások [CW05, CCW12a] és a kapcsolódó munkák hatására szivárogtak be a rendszeridentifikációs közösségbe.

A legegyszerűbb paraméteres statisztikai példa, amikor ismert szórású normális eloszlású sokaság várhatóértékére végzünk hipotézisvizsgálatot. Az u-teszt konstrukciója során alapozunk arra, hogy a minták normális eloszlásúak (feltételezzük az eloszlást). Adott minták esetén az u-teszt determinisztikus választ ad arra nézve, hogy a null hipotézist elutasíthatjuk-e vagy sem.

Az eloszlásmentes eljárások célja, hogy a mérési adatok eloszlásának pontos ismerete nélküli hipotézisvizsgálatot is lehetővé tegyenek. Az SPS eljárás [CCW12a] egy példa erre, aminek két fő jellemzője van. Egy felől ez egy eloszlásmentes eljárás, másfelől az általa produkált eredmények véletlenek.

A [CCW12a]-ban bemutatott eredményekre építve egy sokkal általánosabb hipotézisvizsgálati eljáráscsaládot írok le, ami az SPS-hez hasonló jellemzőkkel bír és adatperturbációs eljáráscsaládnak neveztem el. Ezen eljárások tulajdonságait vizsgáltam lineáris regressziós valamint lineáris dinamikus rendszerek paraméterbecslési feladatai esetén.

### I.1 Altézis: Adatperturbációs eljárások

*Olyan hipotézisvizsgálati eljárásokat írok le, amelyek egzakt, tetszőlegesen megválasztott konfidenciaszintű döntéseket tesznek lehetővé anélkül, hogy a méréseket terhelő zaj eloszlásának teljes ismeretére támaszkodnának, részleges feltételekre azonban szükség van. Az ebbe a családba tartozó algoritmusok által hozott döntések véletlenek rögzített mérési adatok esetén is. Az SPS eljárás is ebbe a családba tartozik.*

Néhány feltételnek teljesülnie kell ahhoz, hogy adatperturbációs eljárásokat használhassunk egy adathalmazon.

**Feltétel 1 (Modellstruktúra feltétel)** *Feltesszük, hogy a méréseink egy ismert modellstruktúrához tartozó modelltől származnak, ami a következőképpen írható le*

$$Y = f(\theta_0, X, E),$$

*ahol  $X$  a mérhető zajjal nem terhelt adatokat tartalmazza,  $E$  tartalmazza a méréseket terhelő zaj forrását,  $Y$  a megfigyeléseket,  $f : \Theta \times \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{Y}$  egy ismert leképezés a modell paramétereiről, bemenetről és zajról a megfigyelésekre.*

Az 1. Feltétel interpretációja, hogy pontosan ismerjük, hogy a zaj hogyan szennyezi a méréseinket. Egy olyan eljáráscsaládot írok le, aminek célja a  $H_0 : \theta_0 = \theta$  null- és a  $H_1 : \theta_0 \neq \theta$  alternatív hipotézisek közötti hipotézisvizsgálat.

**Feltétel 2 (Invertálhatóság a zajra nézve)** *Egy  $f : \Theta \times \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{Y}$  modellstruktúra invertálható a zajra nézve, ha létezik egy  $f^* : \Theta \times \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{E}$  leképezés úgy, hogy*

$$\forall \theta \in \Theta, \forall X \in \mathcal{X}, \forall Y \in \mathcal{Y} : (\exists! E \in \mathcal{E} : Y = f(\theta, X, E)) \Rightarrow E = f^*(\theta, X, Y)$$

Az invertálhatósági feltétel a modellstruktúra feltétellel együtt azt jelenti, hogy minden lehetséges  $D = (Y, X)$  mérési adat esetén explicit és egyértelmű módon kiszámolható az  $E$  zajvektor minden lehetséges  $\theta$  modellez, ha az adat származhat a modellstruktúrához tartozó modelltől.

**Feltétel 3 (Transzformáció invariancia)** Legyen  $\mathbb{P} : \mathcal{E} \rightarrow [0, 1]$  egy vektorértékű valószínűségi változó eloszlásfüggvénye és  $G$  transzformációk egy halmaza, ami a természetes kompozíció operátorral egy  $(G, \cdot)$  kompakt csoportot alkot. A  $\mathbb{P}$  eloszlás invariáns a  $G$ -beli transzformációkra nézve, ha  $\forall E \subset \mathcal{E}, \forall g \in G F(E) = F(gE)$ .

A transzformáció invariancia egyenes következménye, hogy a  $G$  feletti Haar mérték létezik és véges, vagyis az egyenletes eloszlás definiálható  $G$ -n.

**Definíció 1 (Jól meghatározott sorrendezés egy permutációra nézve)** A  $Z_1, \dots, Z_m$  értékek jól meghatározott csökkenő sorrendezése az  $\{1, \dots, m\}$  értékek egy  $\pi$  permutációjára nézve az  $\{1, \dots, m\}$  értékek egy másik  $O$  permutációja, amit úgy definiálunk, hogy  $i$  megelőzi  $j$ -t  $O$ -ban, ha 1.  $Z_i > Z_j$  vagy ha 2.  $Z_i = Z_j$  és  $i$  megelőzi  $j$ -t  $\pi$ -ben.

Valós számok jól meghatározott sorrendje megegyezik a hagyományos sorrendezéssel, ha nincsenek egyező értékek. Azonos értékek sorrendjét az indexük és az adott  $\pi$  permutáció egyértelműen meghatározza.

**Definíció 2 (Teljesítménymérték)** Minden  $Z : \Theta \times \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  szerint értelmezett függvény felfogható teljesítménymértéknek.

Legyen  $D = (Y, X)$  a paraméterbecsléshez rendelkezésre álló adathalmaz és a becsült  $f$  modellstruktúra invertálható a zajra nézve. Ezenfelül legyen  $G$  transzformációk egy olyan halmaza, amire a zajeloszlás invariáns.  $\mathcal{A}$  az  $\{1, \dots, m\}$  értékek permutációinak egy tetszőleges részhalmaza és  $\alpha(\mathcal{A}) = \frac{|\mathcal{A}|}{m!}$ .  $\theta$  a hipotézisvizsgálatnak alávetett modell és  $Z$  egy tetszőleges teljesítménymérték.

**Tétel 1** Legyen  $D = (Y, X)$  egy adathalmaz, amit egy az 1. és 2. Feltételt teljesítő  $\theta_0$  modell generált. Ezen felül a méréseket terhelő zaj együttes eloszlása invariáns egy  $G$  transzformációs csoportra nézve.  $\mathcal{A}$  az  $\{1, \dots, m\}$  számok permutációinak egy tetszőleges részhalmaza.

A fenti feltételek mellett

$$\mathbb{P}(\text{TESTMODEL}(\theta_0, X, Y, \mathcal{A}) = \text{elfogad}) = \alpha(\mathcal{A}) = \frac{|\mathcal{A}|}{m!}$$

ahol a TESTMODEL eljárást az 1. Algoritmus definiálja.

Az SPS eljárás is egy adatperturbációs eljárás, ami esetén a zaj együttes eloszlása szimmetrikus és a transzformációk csoportja a vektorokra értelmezett elemenkénti előjelváltások csoportja. A teljesítménymérték a legkisebb négyzetek módszerében szereplő célfüggvény gradienseinek normáján keresztül kerül definiálásra.

---

**Algoritmus 1** Az adatperturbációs eljárás

---

- 1: **procedure** TESTMODEL( $\theta, X, Y, \mathcal{A}$ )
- 2:   **if**  $f^*(\theta, X, Y)$  nem definiált **then**
- 3:      **$\theta$  -t elutasítjuk .**
- 4:   **else**
- 5:     Számoljuk ki  $E(\theta) = f^*(\theta, X, Y)$ -t.
- 6:     Sorsoljuk a  $g_2, \dots, g_m$  véletlen elemeket egyenletesen  $G$ -ből.
- 7:     Készítünk  $m$  perturbált adathalmazt

$$D^{(1)} = (Y, X) \quad D^{(i)} = (f(\theta, X, g_i E(\theta)), X) \quad \forall i = 2, \dots, m$$

- 8:   Definiáljuk a  $Z_i$  értékeket mint  $\theta$  teljesítménye a különböző adathalmazokon

$$Z_i = Z(\theta, X, Y^{(i)}) \quad \forall i = 1, \dots, m \quad Y^{(i)} = f(\theta, X, g_i E(\theta))$$

- 9:   Egyenletesen választunk egy  $\pi$  permutációt az  $\{1, \dots, m\}$  számok felett.
  - 10:   Számoljuk a  $Z_1, \dots, Z_m$  számok  $\pi$ -re nézve vett  $O$  jól meghatározott sorrendjét.
  - 11:   **if**  $O \in \mathcal{A}$  **then**
  - 12:      **$\theta$ -t elfogadjuk.**
  - 13:   **else**
  - 14:      **$\theta$ -t elutasítjuk.**
  - 15:   **end if**
  - 16: **end if**
  - 17: **end procedure**
- 

**I.2 Altézis: Adatperturbációs eljárások lineáris regressziós problémákra**

*Behatóan vizsgáltam a lineáris regressziós modellstruktúrához tartozó adatperturbációs eljárásokat. Karakterizáltam zajeloszlások egy gazdag családját, amelyek lehetővé teszik adatperturbációs eljárások használatát és összefüggő konfidencialmazokat generálnak a megfelelő teljesítménymérték használatával. Megadtam a regresszorokra vonatkozó elégséges feltételeket, amelyek garantálják, hogy a konfidencialmazok korlátosak lesznek.*

A lineáris regressziós problémáknak megfelelő modellstruktúra a következőképpen írható le

$$Y = f(\theta_0, X, E) = X^T \theta_0 + E, \quad Y, E \in \mathbb{R}^N, \quad X \in \mathbb{R}^{n_\theta \times N}, \quad \theta_0 \in \mathbb{R}^{n_\theta}$$

Ez a modellstruktúra egyértelműen invertálható a zajra nézve

$$E = f^*(\theta, X, Y) = Y - X^T \theta$$

**Definíció 3 (Unitér csoport)** Az  $n$ -edrendű unitér csoportot, melyet  $U(n)$ -el jelöljük, az  $n \times n$  méretű unitér mátrixok alkotják a mátrixszorzás művelettel.

**Tétel 2** Ha az  $E$  zajvektor  $F$  együttes eloszlása invariáns  $U(n)$  egy tetszőleges  $G$  alcsoportjára, akkor készíthető olyan adatperturbációs algoritmus a  $\theta_0$  vektor hipotézisvizsgálatára, ami összefüggő konfidencialmazt eredményez.

*Az ezekhez az eljárásokhoz tartozó teljesítménymérték*

$$Z(\theta, X, Y) = ([X X^T]^{-1} X^T Y - \theta)^T [X X^T] ([X X^T]^{-1} X^T Y - \theta)$$

A sorrendek  $\mathcal{A}$  elfogadott halmazát úgy konstruáljuk, hogy a permutációkat az 1 bennük elfoglalt helye szerinti növekvő sorrendben válogatjuk be, mindaddig, amíg  $\alpha(\mathcal{A})$  el nem éri a kívánt konfidenciaszintet.

Az SPS eljárás két különböző teljesítménymértékkel is definiálva van [CCW12a]-ben. Ezek közül az egyik megfelel a 2. Tételben szereplőnek. Mindkét [CCW12a]-ben megadott teljesítménymérték esetén adtam elégséges feltételeket a regresszorokra vonatkozóan, amelyek korlátos konfidenciahalmazt eredményeznek. Megmutattam, hogy a [CCW12a]-ben szereplő másik teljesítménymérték nem általános abban az értelemben, hogy csak bizonyos bemenetek és zajeloszlások esetén eredményez korlátos konfidenciahalmazt.

### I.3 Altézis: Adatperturbációs eljárások lineáris dinamikus rendszerekhez

Adatperturbációs eljárásokat definiáltam lineáris dinamikus időinvariáns rendszerek paraméterbecsléséhez. Olyan teljesítménymértéket definiáltam, ami a lineáris regressziós példánál látott mérték természetes kiterjesztésének tekinthető. Megmutattam, hogy ezek az adatperturbációs eljárások nem összefüggő konfidenciahalmazokat generálnak a paramétereiktől való nemlineáris függés miatt. Megvizsgáltam a konfidenciahalmaz "központi" komponensének tulajdonságait és ennek használatát javasoltam, mint közelítő algoritmus.

A feladathoz tartozó modellstruktúra alakja a következő

$$\mathcal{A}(z)y[k] = \frac{\mathcal{B}(z)}{\mathcal{F}(z)}u[k] + \frac{\mathcal{C}(z)}{\mathcal{D}(z)}e[k]$$

ahol  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$ ,  $\mathcal{D}$  és  $\mathcal{F}$  a  $z^{-1}$  eltolás operátor véges  $n_{\mathcal{A}}$ ,  $n_{\mathcal{B}}$ ,  $n_{\mathcal{C}}$ ,  $n_{\mathcal{D}}$  és  $n_{\mathcal{F}}$  fokszámú polinomjai. A  $\mathcal{D}$  adathalmaz mérhető  $\mathcal{X}$  része a  $G(z) = \frac{\mathcal{B}(z)}{\mathcal{A}(z)\mathcal{F}(z)}$  folyamatmodell  $u[k]$  bemenő jelét tartalmazza, a megfigyelhető  $\mathcal{Y}$  megfelel a mért  $y[k]$  kimeneti értékeknek, a nem mérhető  $\mathcal{E}$  véletlen komponens a  $H(z) = \frac{\mathcal{C}(z)}{\mathcal{A}(z)\mathcal{D}(z)}$  zajmodell  $e[k]$  bemenő jele.

**Állítás 1 (Kezdeti érték feltételek és invertálhatóság)** Ahhoz, hogy a modellstruktúra megfeleljen a 2. Feltételnek, mind a  $G$  folyamatmodell, mind a  $H$  zajmodell kezdeti érték feltételei ismertek kell, legyenek.

Az 1. Állítás úgy értelmezhető, hogy a kezdeti érték feltételeknek vagy ismertnek kell lenniük, vagy becsült paraméterként a  $\theta$  vektor részét kell képezzék, annak érdekében, hogy adatperturbációs eljárást definiálhassunk a modellstruktúrára. Ez olyan  $\theta$  paramétervektort eredményez, ami a  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$ ,  $\mathcal{D}$  és  $\mathcal{F}$  polinomok együtthatóit tartalmazza, kiegészítve az ismeretlen kezdeti érték feltételekkel.

Amint a  $\theta$  paramétervektor és az egyéb feltételezések lehetővé teszik adatperturbációs módszer alkalmazását (a 2. és 3. Feltétel teljesül), az egyetlen hátralevő dolog a megfelelő teljesítménymérték kiválasztása. Az általam javasolt teljesítménymérték a lineáris regresszióknál alkalmazott mérték egy természetesebb kiterjesztése a [CCW12b]-ben megjelenhez képest.

Lineáris regressziók esetében a megfelelő teljesítménymérték a

$$Z(\theta, X, Y) = ([XX^T]^{-1}X^TY - \theta)^T [XX^T] ([XX^T]^{-1}X^TY - \theta)$$

volt, ami több különböző módon is interpretálható. Ezek közül az egyik, a célfüggvény gradienseinek súlyozott normájaként való értelmezés, ahogy [CCW12a]-ban szerepel. Egy

másik értelmezési lehetőségként  $\theta$  és az adathoz tartozó legkisebb négyzetes becslés közötti súlyozott távolságként is tekinthető  $Z$ . A harmadik lehetőség a tesztelt modell  $\theta$  és a legkisebb négyzetes becslés által produkált prediktált kimenetek differenciájának normája.

$$Z(\theta, X, Y) = (X^T [X X^T]^{-1} X^T Y - X^T \theta)^T (X^T [X X^T]^{-1} X^T Y - X^T \theta)$$

A harmadik interpretációt terjesztettem ki lineáris dinamikus rendszerek paraméterbecslési feladatára. Legyen  $D$  egy perturbált adathalmaz és  $\hat{\theta}$  egy paraméterbecslés ehhez az adathalmazhoz (tetszőleges becslési eljárás eredményeként),  $\hat{Y}_{\hat{\theta}}$  a  $\hat{\theta}$ -nak megfelelő folyamatmodell szimulált kimenete a  $D$ -ben lévő bemeneti jellel, hasonlóképpen  $\hat{Y}_{\theta}$  a  $\theta$  modellhez tartozó folyamatmodell kimenete ugyanarra a bemeneti jelre.

A lineáris regressziós esetben, ha a  $\hat{\theta} = [X X^T]^{-1} X^T Y$  legkisebb négyzetes becslést használjuk, akkor a teljesítménymérték a következő alakban írható

$$Z(\theta, X, Y) = \left\| \hat{Y}_{[X X^T]^{-1} X^T Y} - \hat{Y}_{\theta} \right\|^2$$

**Állítás 2 (Teljesítménymérték folyamatmodellre)** Az elérhető adathalmaz álljon az  $u[k]$  bemeneti értékekből és az  $y[k]$  megfigyelésekből. Legyen  $\hat{\theta}$  a paramétervektor egy becslése az elérhető adathalmazból. A

$$Z(\theta, U, Y) = \left\| \hat{Y}_{\hat{\theta}} - \hat{Y}_{\theta} \right\|^2$$

teljesítménymérték használata az adatperturbációs algoritmus hatodik lépésében olyan konfidenciahalmazt eredményez, ami tartalmazza  $\hat{\theta}$ -t. A konfidenciahalmaz  $\hat{\theta}$  körüli összefüggő részhalmaza jó közelítő konfidenciahalmaz a folyamatmodellre a kimeneti zajt feltételező modellstruktúrában.

A fent javasolt teljesítménymérték nem összefüggő konfidenciahalmazokat is eredményezhet. Ez igaz a legtöbb felmerülő teljesítménymértékre, arra is ami [CCW12b]-ben szerepel. Ennek fő oka a feladat nemlinearitása a paraméterek tekintetében.

A 2. Állítás a kimeneti zajt feltételező modellstruktúra esetére javasol teljesítménymértéket. Ez a teljesítménymérték csak a folyamatmodellről függ (hiszen a  $H(z) = 1$  zajmodell ismert). Általánosabb  $H(z)$  zajmodellek esetére is kiterjesztettem a fenti teljesítménymértéket.

A fentebb javasolt teljesítménymérték a legtermészetesebb választás, de más értelmes választásokat is bemutatok a dolgozatban. Ezek sok szempontból hasonlóan viselkednek. A kapcsolódó konfidenciahalmazok nem összefüggőek és hasonlóak ameddig a használt teljesítménymértékek értelmesek.

Ahogy az előző állításban említésre kerül, a  $\hat{\theta}$  körüli összefüggő komponens használatát javaslom, mint a konfidenciahalmaz egy alsó közelítése. Ezt a javaslatot különböző heurisztikus érvekkel támogattam meg, rámutatva, hogy a konfidenciahalmaz  $\hat{\theta}$ -t nem tartalmazó más komponensei elhanyagolható hozzáadott konfidenciaszintet jelentenek csak.

## II. TÉZIS:

### Polinomiális modellek globálisan optimális becslése

A tézishoz kapcsolódó publikációim: [6], [7], [8], [9], [10], [11], [12].

Sok esetben az identifikációs probléma megfelel egy több lokális optimummal rendelkező optimalizálási feladatnak. Ezekben az esetekben a globálisan optimális modell megtalálása a cél. Sok esetben kellően jó lokális optimumokat kaphatunk, de hasznos, ha elméleti garanciákat tudunk arra, hogy valójában a globális optimumot találtuk meg.

Amint azt az előző tézisben leírtam, a lineáris dinamikus paraméterbecsléséhez kapcsolódó adatperturbációs eljárások teljesítménymértékének számításához szükség van a perturbált adathalmazokhoz tartozó  $\hat{\theta}$  pontbecslések meghatározására. Ez tetszőleges becslési technikával számolható (úgy, mint predikciós hibát minimalizáló, alteres vagy segédváltozós eljárásokkal) és a  $\hat{\theta}$  körül szerkesztett közelítő konfidenciahalmaz tartalmazni fogja az eredeti adathalmazhoz tartozó becslést. Ez kiemeli a becslés megbízható számításának fontosságát. Példaként, ha a predikciós hibát minimalizáló becslő eljárás egy lokális optimumban áll meg, akkor az ekörül a pont körül szerkesztett konfidenciahalmaz konfidenciaszintje nem lesz megfelelő.

A fent említett okok miatt megvizsgáltam a szemidefinit programozási (SDP) feladatok sorozatából álló Lasserre hierarchia [Las01] alkalmazhatóságát polinomiális rendszerek identifikációs problémájának globális megoldására.

Ez az eljárás általános polinomiális korlátozások melletti polinomiális célfüggvényű optimalizálási feladatok megoldására használható. Ennek a megközelítésnek az identifikációs feladathoz kapcsolódó optimalizációs problémák szempontjából fontos tulajdonságait vizsgáltam. Ismert, hogy az általános megoldás számításigénye kezelhetetlen összetett, sokváltozós problémák esetén, ezért különböző megoldások jelentek meg, amelyek a probléma struktúrájára alapozva csökkentik a megoldandó SDP-k méretét. Megmutattam, hogy az identifikációs problémához kapcsolódó optimalizálási probléma korrelatív ritkasági értelemben ritka. Erre a ritkasági tulajdonságra alapozva az SDP-k numerikusan kezelhetőbbekké válnak [WKKM06].

Általában a feladat ritkaságára épített méretcsökkentés ára a globális optimumhoz való garantált konvergencia elvesztése. A korrelatív ritkaság esetében azonban ismertek a globális optimumhoz való konvergencia szükséges és elégséges feltételei [Las06]. Megmutattam, hogy az identifikációs feladat esetén ezek a feltételek teljesíthetőek.

A fentiek azt eredményezik, hogy SDP-k hierarchiája definiálható, amivel az identifikációs feladat globális optimumát közelíteni lehet. Ezek az SDP-k a korrelatív ritkasági struktúra miatt kezelhető méretűek maradnak úgy, hogy a globális optimumhoz való konvergencia megmarad. Adott relaxációs szinten belül az SDP-k mérete lineárisan nő az  $N$  mintaszámmal.

A megoldásra kerülő SDP-k pontos mérete változik a megoldásra használt szoftvercsomag függvényében, melyek speciális trükkökkel növelik a pontosságukat, de ezek nem változtatnak a méretek nagyságrendjén.

#### II.1 Altézis: Ritkaság nélküli optimalizálás

*Polinomiális modellek identifikációja esetén két alapvető paraméterezés létezik. Az egyikben csak a modellparaméterek az ismeretlenek. Ebben az esetben az identifikációs probléma egy magas fokszámú korlátozásmentes optimalizációs feladat. A másik esetben kiegészítő optimalizációs változók jelennek meg a zajmodell bemeneti jelére. Így négyzetes célfüggvényt*



kell optimalizálni egyenlőségkorlátozások mellett. Explicit formulákban megadtam a különböző paraméterezésekhez kapcsolódó SDP-k méretét a modell komplexitás, mintaszám és relaxációs szám függvényében.

Az alábbi modellstruktúrát tételezzük fel

$$\mathcal{A}(z)y[k] = \frac{\mathcal{B}(z)}{\mathcal{F}(z)}u[k] + \frac{\mathcal{X}_{\frac{\mathcal{B}}{\mathcal{F}}}(z)}{\mathcal{F}(z)}\delta[0] + \frac{\mathcal{C}(z)}{\mathcal{D}(z)}e[k] + \frac{\mathcal{X}_{\frac{\mathcal{C}}{\mathcal{D}}}(z)}{\mathcal{D}(z)}\delta[0] + \mathcal{X}_{\frac{1}{\mathcal{A}}}(z)\delta[0] \quad k \geq 1$$

ahol  $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$  és  $\mathcal{F}$  a  $z^{-1}$  eltolás operátor véges  $n_{\mathcal{A}}, n_{\mathcal{B}}, n_{\mathcal{C}}, n_{\mathcal{D}}$  és  $n_{\mathcal{F}}$  fokszámú polinomjai. A  $\mathcal{X}$ . polinomok a rendszerek kezdeti érték feltételeinek felelnek meg. A kezdeti érték polinomok és  $\mathcal{B}(z)$  kivételével minden polinom monikus.  $u[k]$  a  $G(z) = \frac{\mathcal{B}(z)}{\mathcal{A}(z)\mathcal{F}(z)}$  folyamatmodell bemenő jele,  $y[k]$  a mérhető kimeneti érték,  $e[k]$  a  $H(z) = \frac{\mathcal{C}(z)}{\mathcal{A}(z)\mathcal{D}(z)}$  zajmodell nem mérhető bemenő jele,  $\delta[k]$  az egység impulzus  $k$ -ban.

**Definíció 4 (Racionális törtfüggvények állapotteres leírása)** Az alábbi formában megadott racionális törtfüggvény

$$P(z) = \frac{\mathcal{B}(z)}{\mathcal{A}(z)} = \tilde{d} + \frac{\sum_{k=1}^n \tilde{b}_k z^{-k}}{1 + \sum_{k=1}^n a_k z^{-k}}$$

állapotteres leírása megadható

$$A = \begin{bmatrix} -a_1 & 1 & 0 & \cdots \\ \vdots & 0 & \ddots & 0 \\ \vdots & \vdots & 0 & 1 \\ -a_n & 0 & \cdots & 0 \end{bmatrix} \quad B = \begin{bmatrix} \tilde{b}_1 \\ \vdots \\ \tilde{b}_n \end{bmatrix} \quad C = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}^T \quad D = \tilde{d}$$

alakban. A rendszer  $U \in \mathbb{R}^N$  bemenetre adott  $Y \in \mathbb{R}^N$  válasza

$$Y = \mathbf{\Gamma}_{\frac{\mathcal{B}}{\mathcal{A}}} \mathbf{x}_{\frac{\mathcal{B}}{\mathcal{A}}}[0] + \mathbf{H}_{\frac{\mathcal{B}}{\mathcal{A}}} U$$

alakban számolható, ahol  $\mathbf{x}_{\frac{\mathcal{B}}{\mathcal{A}}}[0] \in \mathbb{R}^n$  a rendszer kezdeti értéke,  $\mathbf{\Gamma}_{\frac{\mathcal{B}}{\mathcal{A}}} \in \mathbb{R}^{N \times n}$  a rendszer kiterjesztett megfigyelhetőségi mátrixa

$$\mathbf{\Gamma}_{\frac{\mathcal{B}}{\mathcal{A}}} = [C^T (CA)^T (CA^2)^T \dots (CA^{N-1})^T]^T$$

és  $\mathbf{H}_{\frac{\mathcal{B}}{\mathcal{A}}}$  egy alsóháromszög Töplitz mátrix a

$$\mathbf{H}_{\frac{\mathcal{B}}{\mathcal{A}}}(:, 1) = [D \ C \ B \ C \ A \ B \ \dots \ C \ A^{N-2} \ B]^T$$

első oszlop eltolásával definiálva.

Az  $U, Y$  és  $E$  oszlopvektorokat definiálva a bemeneti, kimeneti és zaj értékekből, a modellstruktúra leírható az alábbi alakban

$$Y = \mathbf{\Gamma}_{\frac{1}{\mathcal{A}}} \mathbf{x}_{\frac{1}{\mathcal{A}}}[0] + \mathbf{H}_{\frac{1}{\mathcal{A}}} \left( \mathbf{\Gamma}_{\frac{\mathcal{B}}{\mathcal{F}}} \mathbf{x}_{\frac{\mathcal{B}}{\mathcal{F}}}[0] + \mathbf{H}_{\frac{\mathcal{B}}{\mathcal{F}}} U + \mathbf{\Gamma}_{\frac{\mathcal{C}}{\mathcal{D}}} \mathbf{x}_{\frac{\mathcal{C}}{\mathcal{D}}}[0] + \mathbf{H}_{\frac{\mathcal{C}}{\mathcal{D}}} E \right)$$

Egy adott  $\theta$  paramétervektorhoz tartozó predikciós hibák kifejezése

$$E(\theta) = \mathbf{H}_{\mathcal{D}} \left( \mathbf{H}_{\mathcal{A}} \left( Y - \Gamma_{\mathcal{A}} \mathbf{x}_{\mathcal{A}}[0] \right) - \Gamma_{\mathcal{F}} \mathbf{x}_{\mathcal{F}}[0] - \mathbf{H}_{\mathcal{F}} U - \Gamma_{\mathcal{D}} \mathbf{x}_{\mathcal{D}}[0] \right)$$

A legkisebb négyzetes predikciós hibát adó módszerhez kapcsolódó korlátozásmentes optimalizálási feladathoz tartozó célfüggvény

$$V_d(\theta) = \frac{1}{2N} E(\theta)^T E(\theta)$$

ahol a  $\theta$  paramétervektor a rendszer állapotteres leírásának paramétereit és az ismeretlen kezdeti érték feltételeket tartalmazza.

Az identifikációs probléma komplexitásának egy fontos mérőszáma a predikciós hibát kifejező polinomok fokszáma. Definiáljuk az  $\mathbb{1}$ . indikátorfüggvényt mint

$$\mathbb{1}_{\mathcal{P}} = \begin{cases} 1, & \text{a } \mathcal{P} \text{ polinom tartalmaz ismeretleneket} \\ 0, & \text{egyébként} \end{cases}$$

és jelölje  $\deg(\cdot)$  egy polinom elemű  $M$  mátrix maximális fokszámú elemének fokszámát.

$$\deg(M) = \max_{i,j} \{ \deg(M(i,j)) \}$$

A predikciós hiba kifejezésében megjelenő különböző tényezők fokszáma a következő módon számolható

$$\deg\left(\mathbf{H}_{\mathcal{A}}\right) = \mathbb{1}_{\mathcal{B}} + \mathbb{1}_{\mathcal{A}}(N-2) \quad \deg\left(\Gamma_{\mathcal{A}}\right) = \mathbb{1}_{\mathcal{A}}(N-1)$$

Az  $m$ -ed rendű SDP alapú relaxáció megoldása során legalább egy  $\binom{n_{\theta}+2m}{2m}$  méretű lineáris mátrix egyenlőtlenség (LMI) szerepel, ahol  $m$  legalább  $\frac{\deg V_d}{2} = \deg(\hat{E})$ .

**Állítás 3 (Korlátozásmentes általános SDP-k mérete)** A  $\deg(V_d)$  legkisebb négyzetes célfüggvény fokszáma a következő módon számolható

$$\frac{1}{2} \deg(V_d(\theta)) = \mathbb{1}_{\mathcal{C}} + (N-2)\mathbb{1}_{\mathcal{D}} + \max \left\{ \begin{aligned} & \mathbb{1}_{\mathcal{A}} + \mathbb{1}_{\mathcal{X}_{\mathcal{A}}} (1 + (N-1)\mathbb{1}_{\mathcal{A}}) \\ & \mathbb{1}_{\mathcal{X}_{\mathcal{F}}} (1 + (N-1)\mathbb{1}_{\mathcal{F}}), \\ & \mathbb{1}_{\mathcal{B}} + (N-2)\mathbb{1}_{\mathcal{F}}, \\ & \mathbb{1}_{\mathcal{X}_{\mathcal{D}}} (1 + (N-1)\mathbb{1}_{\mathcal{D}}) \end{aligned} \right\}$$

Modellstruktúrától függően az

$$\begin{aligned} \hat{\theta} &= \arg \min V_d(\theta) \\ \|\theta\|^2 &\leq R^2 \end{aligned}$$

optimalizációs feladat  $m$ -ed rendű SDP relaxációjának mérete az ismeretlenek számának és a célfüggvény fokszámának ismeretében számolható. Két teljes méretű LMI korlátozás szerepel az SDP reprezentációban. A modellstruktúra komplexitásától függően három különböző nehézségi osztályba sorolhatóak a feladatok.

1. *Modell komplexitásfüggő osztály: kezdeti értékek hiánya és az  $n_f = n_d = 0$  feltétel karakterizálja ezt az osztályt. A minimális relaxációs rend  $m_{\min} = 1$ . A modell bonyolultsága  $n_{\theta} = n_{\mathcal{A}} + n_{\mathcal{B}} + n_{\mathcal{C}}$ . Az SDP méretnövekedés kombinatorikusan függ  $n_{\theta}$ -től.*
2. *Modell komplexitás és mintaszám-függő nehézségi osztály: olyan problémák tartoznak ide, amelyekben a minimális relaxációs rend  $1N$ -től függ. Az ide tartozó modellek esetében  $\mathcal{D} = 1$  és a minimális relaxációs rend az  $N$  mintaszámmal nő.*
3. *Modell komplexitás és kétszeres mintaszám-függő nehézségi osztály: olyan problémák tartoznak ide, amelyekben a minimális relaxációs rend  $2N$ -től függ. Az előző két csoportból kimaradó minden modell ide tartozik. A minimális relaxációs rend a mintaszám duplájával nő.*

Az első csoportban, ahol a minimális relaxációs rend nem függ a mintaszámtól, az első relaxáció a globális optimumot meg is adja.

Minden más esetben magasabb rendű relaxációk szükségesek. Mivel a relaxációs rend a mintaszámtól lineárisan függ, ezért az SDP-k mérete exponenciálisan függ a mintaszámtól.

A másik megközelítésben az  $\varepsilon[k]$  zajmintákhoz kiegészítő változók tartoznak. Ebben az esetben a célfüggvény alakja

$$V_a(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \frac{1}{2N} \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$$

ahol a paramétervektor tartalmazza mind a rendszer  $\boldsymbol{\theta}$  paramétereit, valamint az  $\varepsilon[k]$  értékeket. A célfüggvény optimalizációja az alábbi korlátozások mellett történik

$$\begin{aligned} \mathcal{A}(z)\mathcal{F}(z)\mathcal{D}(z)y[k] &= F(z)\mathcal{D}(z)\mathcal{X}_{\mathcal{A}}(z)\delta[0] + \\ &+ \mathcal{D}(z)(\mathcal{X}_{\mathcal{F}}(z)\delta[0] + \mathcal{B}(z)u[k]) + \mathcal{F}(z)(\mathcal{X}_{\mathcal{D}}(z)\delta[0] + \mathcal{C}(z)\varepsilon[k]) \quad \forall k \in [N] \end{aligned}$$

amelyek minden mintára felírandóak.

Ez egy négyzetes célfüggvénnyel és polinomiális egyenlőségkorlátozásokkal történő optimalizációs probléma, amiben a változók száma lineárisan növekszik az  $N$  mintaszámmal, a korlátozások számához hasonlóan.

**Állítás 4 (Korlátozásos általános SDP-k mérete)** *Modellstruktúrától függően az*

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin} V_a(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \frac{1}{2N} \sum_{k=1}^N \varepsilon^2[k]$$

$$\begin{aligned} \mathcal{A}(z)\mathcal{F}(z)\mathcal{D}(z)y[k] &= \mathcal{F}(z)\mathcal{D}(z)\mathcal{X}_{\mathcal{A}}(z)\delta[0] + \\ &+ \mathcal{D}(z)(\mathcal{X}_{\mathcal{F}}(z)\delta[0] + \mathcal{B}(z)u[k]) + \mathcal{F}(z)(\mathcal{X}_{\mathcal{D}}(z)\delta[0] + \mathcal{C}(z)\varepsilon[k]) \quad \forall k \in [N] \\ &\|\boldsymbol{\theta}\|^2 + \|\boldsymbol{\varepsilon}\|^2 \leq R^2 \end{aligned} \tag{1}$$

*optimalizációs probléma  $m$ -ed rendű relaxációjához tartozó SDP-k mérete az egyenlőségkorlátozások fokszáma alapján számolható (mind a modell együtthatókat, kezdeti értékeket és hibaváltozókat figyelembe véve).  $2N + 2$  teljes méretű LMI korlátozás tartozik az SDP reprezentációkhoz.*

*Modell struktúra alapján a három nehézségi osztályba sorolhatóak a problémák.*

1. *Lineáris korlátozások: A legkisebb relaxációs rend  $m_{\min} = 1$ .*

2. Négyzetes korlátozások: A legkisebb relaxációs rend  $m_{\min} = 2$ .

3. Kőbös korlátozások: A legkisebb relaxációs rend  $m_{\min} = 3$ .

Az optimalizálási változók száma mindhárom esetben  $n_{\theta} + N$ . Ez mind az  $N$  mintaszámban, mind az  $m$  relaxációs rendben exponenciálisan növekvő méretű SDP-ket eredményez.

A legegyszerűbb problémáktól eltekintve (amelyek lineáris regressziós alakúak) mindkét megközelítés az  $N$  mintaszámban exponenciálisan növekvő méretű SDP-k megoldását igényli. Ez nagyon alacsony mintaszámokra korlátozza a módszer alkalmazhatóságát. Modellstruktúrafüggő, hogy melyik megközelítés indítható kisebb SDP mérettel.

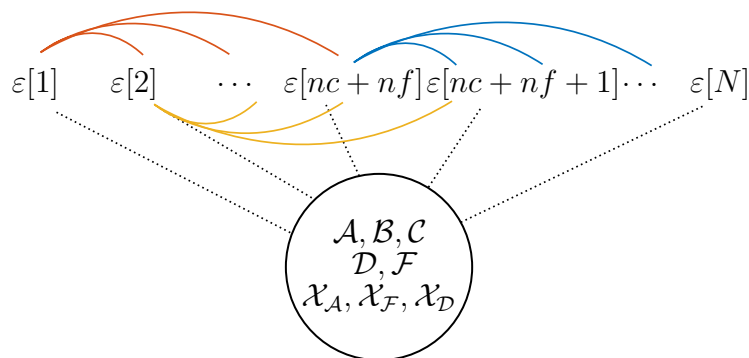
## II.2 Altézis: Optimalizálás ritkaság felhasználásával

Megvizsgáltam az egyenlőség korlátozásokat használó optimalizációs probléma ritkasági tulajdonságait. Megmutattam, hogy a problémához tartozó korrelatív ritkasági gráf (CSP) kiterjeszhető húrgráffá és meghatároztam a kiterjesztett gráf maximális klikkjeit. Ennek segítségével meghatároztam a feladathoz kapcsolódó ritka SDP relaxációk méretét. A fő következtetés az, hogy az optimalizálási feladat struktúrájának felhasználásával az SDP-k méretnövekedése a mintaszám függvényében lineárisra csökkenthető.

Egy optimalizálási probléma korrelatív ritkasági gráfja (CSP) a következőképpen definiált. Minden optimalizálási változónak megfelel egy csomópont a gráfban. Két csomópont között van él, ha a célfüggvény egy monomjában együtt jelennek meg, vagy ha együtt szerepelnek egy korlátozásban [WKKM06].

**Állítás 5 (A korlátozásmentes probléma ritkasága)** Ha a mintaszám magasabb, mint a becsülendő paraméterek száma, akkor a korlátozásmentes optimalizálási feladat CSP gráfja egy teljes gráf, vagyis a feladatnak nincs ritkasága.

Mivel a korlátozásmentes feladat megfogalmazása minimális számú optimalizálási változót használ, nem is várható, hogy egyszerűsítő ritkaság lenne még a feladatban. Ezzel ellentétben a predikciós hiba változóinak bevezetésével ez drasztikusan megváltozik. Az egyenlőségkorlátozásokkal megfogalmazott feladat CSP gráfja a 1. ábrán látható.



Ábra 1: Az egyenlőségkorlátozott identifikációs POP CSP gráfjának húrgráf kiterjesztése

**Állítás 6 (A CSP gráf húrgráf kiterjesztése)** *Az egyenlőségkorlátozott optimalizációs feladat CSP grádjának van kis húrgráf kiterjesztése. Ennek maximális klikkjei a modell paramétereit és kezdeti érték feltételeket tartalmazzák, valamint a rendszer fokszámának megfelelő számú zaj változót. A klikkek száma az  $N$  mintaszámmal, méretük pedig a modell komplexitással arányos.*

Jelentős számítási hatékonyságot nyerünk azzal, hogy a CSP gráf maximális klikkjeinek mérete sokkal kisebb, mint a változók teljes száma. A CSP gráf kiterjesztésének húrgráf tulajdonságából az is következik, hogy annak maximális klikkjei rendelkeznek az úgynevezett *folytonos metszetű* tulajdonsággal. A ritka SDP relaxációk konvergálnak az eredeti feladat globális optimumához, amennyiben a CSP gráf húrgráf kiterjesztésének maximális klikkjei rendelkeznek a folytonos metszetek tulajdonsággal és a globális optimum normájára rendelkezésre áll egy ismert felső korlát [Las06].

**Állítás 7 (Korlátozások ritka SDP-k mérete)** *A modell struktúra függvényében a*

$$\hat{\theta} = \operatorname{argmin} V_a(\theta, \varepsilon) = \frac{1}{2N} \sum_{k=1}^N \varepsilon^2[k]$$

$$\begin{aligned} \mathcal{A}(z)\mathcal{F}(z)\mathcal{D}(z)y[k] &= \mathcal{F}(z)\mathcal{D}(z)\mathcal{X}_A(z)\delta[0] + \\ &+ \mathcal{D}(z)(\mathcal{X}_F(z)\delta[0] + \mathcal{B}(z)u[k]) + \mathcal{F}(z)(\mathcal{X}_D(z)\delta[0] + \mathcal{C}(z)\varepsilon[k]) \quad \forall k \in [N] \\ \|\theta_C\|^2 + \|\varepsilon_{I_k}\|^2 &\leq R_k \quad k \in [p] \end{aligned}$$

*optimalizálási probléma  $m$ -ed rendű relaxációjának mérete az egyenlőségkorlátozások fokszáma alapján számolható (a rendszerparaméterek, kezdeti érték feltételek és hibaváltozók figyelembevételével).  $2N + 2p$  redukált méretű LMI korlátozás tartozik minden rendű relaxációhoz, ahol minden LMI mérete  $\binom{|I|+m}{m}$ , ahol  $p = N - n_C - n_F + 1$  és  $|I|$  a CSP gráf húrgráf kiterjesztésének maximális klikkmérete.*

*A modellstruktúra függvényében három nehézségi osztályba sorolhatóak az identifikációs problémák. Ez az osztályozás megegyezik a 4. Állításban meghatározottal.*

**Tétel 3 (Konvergens ritka relaxációk)** *Az 7. Állításban szereplő optimalizációs feladat megoldható a [Las06]-ben definiált ritka SDP relaxációkkal. A megoldandó SDP-k mérete kombinatorikusan nő a modell komplexitással és a relaxáció rendjével, de adott rendű relaxáción belül a növekmény a mintaszámnak csak lineáris függvénye.*

Mindig előnyösebb az egyenlőség korlátozásokkal megfogalmazott alak használata. Ez olyan LMI korlátozásokat eredményez, amelyek az SDP változóknak csak egy részhalmazára támaszkodnak, ami magasabb relaxációs rend elérését teszi lehetővé.

A polinomiális optimalizálási problémákra és a Lasserre hierarchiákra vonatkozó jelenlegi ismeretink alapján további redukálás nem tűnik valószínűnek. A jelenleg elérhető számítási kapacitás mellett a kezelhető problémák köre erősen korlátozott. Tíz-es nagyságrendű modellparaméter és 30-40-es mintaszám mellett megoldható a második relaxációs közelítés. Ennél nagyobb bonyolultságú feladatok esetén a megoldás numerikusan megbízhatatlanná válik, vagy egyenesen nem is számolható. A tapasztalat azt mutatja, hogy amikor a PEM eljárás egy lokális minimumban ragad, akkor az első vagy második relaxációból kinyerhető becslés használható kezdeti értéként a lokális PEM keresésben jobb becslés számolására.

### III. TÉZIS:

#### Véges mintaszám hatása bizonytalanság becslésére

A tézishoz kapcsolódó publikációim: [3], [13].

A paraméterbecslésekhez tartozó aszimptotikus konfidenciahalmazok használatának két fő oka van. Kellően megbízhatóak, ha elég adat áll rendelkezésre a megkonstruálásukhoz és nagyon egyszerű a struktúrájuk. Lineáris dinamikus rendszerek paraméterbecslése esetén ellipszoidok, amelyek könnyen karakterizálhatóak.

Az ilyen konfidenciahalmazok hátulütője, hogy csak a becslési probléma aszimptotikus tulajdonságaitól függenek, mint a bemenő jel és a zaj spektruma. Ha a feladat bizonyos regularitási feltételeknek eleget tesz, akkor a  $\theta_0$  nominális paraméter egy  $N$  minta alapján konstruált torzítatlan  $\hat{\theta}_N$  becslések eloszlása az alábbi összefüggéssel írható le

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_\infty) \quad \text{Var}(\hat{\theta}_N - \theta_0) = \frac{1}{N}\Sigma_\infty$$

ahol  $\Sigma_\infty$  egy megfelelő kovariancia mátrix. Ez a szabály csak az  $1/N$  szorzótényezőn keresztül veszi figyelembe a véges mintaszám hatását, ez viszont a variancia alulbecsléséhez vezet.

Ennek a tézisnek a célja: *i.* hogy rámutasson az  $1/N$  szabály hiányosságaira; *ii.* hogy javasoljon egy eljárást, ami képes a varianciabecslés hibájának karakterizálására; *iii.* hogy korrekciós módosításokat javasoljon a hagyományos aszimptotikus becslési eljáráshoz.

*Bevezettem a véges mintaszámú varianciaveszteség fogalmát, ami a becslés véges mintából várható valós varianciája és annak az  $1/N$  szabály által számított aszimptotikus becslése közötti különbséget jelenti. Lineáris dinamikus rendszerek torzítatlan becslése esetén algoritmust adtam a varianciaveszteség közelítő számítására és ennek alapján heurisztikus szabályokat javasoltam, amelyek segítségével eldönthető, hogy az aszimptotikus konfidenciahalmaz megbízható-e. Ennek alapján az aszimptotikus varianciabecslés korrigálható a varianciaveszteség valós vagy közelített értékével, hogy jobb közelítést kapjunk. Egy fontos következtetés, amit levonok, hogy a varianciaveszteség jelentős marad a vártnál sokkal magasabb mintaszámok esetén is.*

Ahogy a II. tézisben, úgy most is az alábbi modellstruktúrát tételezzük fel

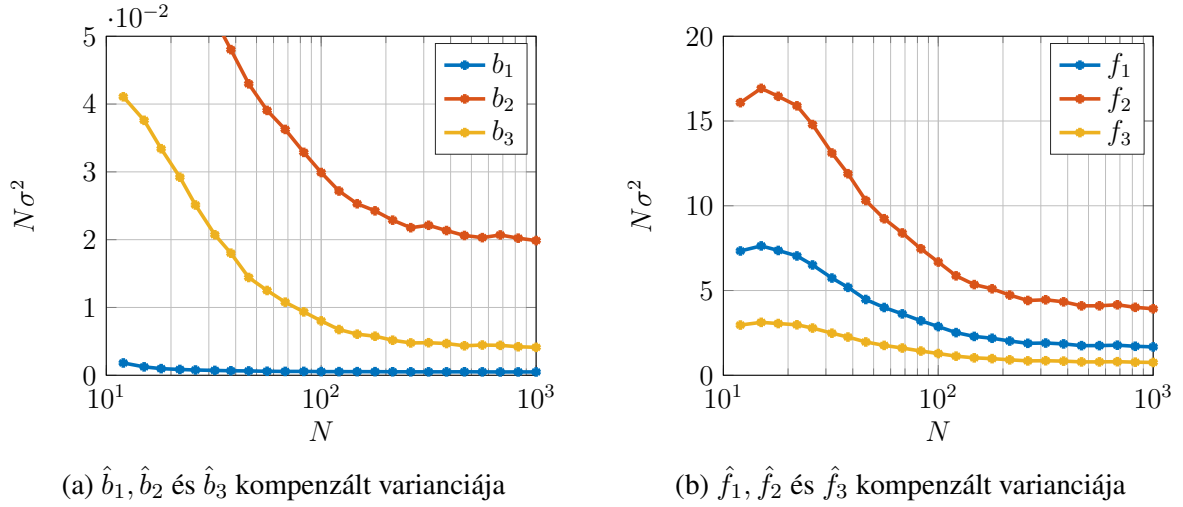
$$\mathcal{A}(z)y[k] = \frac{\mathcal{B}(z)}{\mathcal{F}(z)}u[k] + \frac{\mathcal{X}_{\frac{B}{F}}(z)}{\mathcal{F}(z)}\delta[0] + \frac{\mathcal{C}(z)}{\mathcal{D}(z)}e[k] + \frac{\mathcal{X}_{\frac{C}{D}}(z)}{\mathcal{D}(z)}\delta[0] + \mathcal{X}_{\frac{1}{A}}(z)\delta[0] \quad k \geq 1$$

ahol  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$ ,  $\mathcal{D}$  és  $\mathcal{F}$  a  $z^{-1}$  eltolás operátor véges  $n_A$ ,  $n_B$ ,  $n_C$ ,  $n_D$  és  $n_F$  fokszámú polinomjai.  $\mathcal{X}$ . polinomok a rendszerek kezdeti érték feltételeinek felelnek meg. A kezdeti érték polinomok és  $\mathcal{B}(z)$  kivételével minden polinom monikus.  $u[k]$  a  $G(z) = \frac{\mathcal{B}(z)}{\mathcal{A}(z)\mathcal{F}(z)}$  folyamatmodell bemenő jele,  $y[k]$  a mérhető kimeneti érték,  $e[k]$  a  $H(z) = \frac{\mathcal{C}(z)}{\mathcal{A}(z)\mathcal{D}(z)}$  zajmodell nem mérhető bemenő jele,  $\delta[k]$  az egység impulzus  $k$ -ban. Az  $u[\cdot]$  bemenetet és az  $e[\cdot]$  zajt függetlennek tételezzük fel.

Ahhoz, hogy össze tudjuk hasonlítani növekvő adatmennyiségből származó becslések tulajdonságait, az adatgyűjtési folyamatnak bizonyos értelemben stacionáriusnak kell lennie. A bemenőjel lehet determinisztikus periodikus jel, pseudo-véletlen bináris jel vagy egy stacionárius sztochasztikus folyamat egy realizációja. A dolgozatban a számítások fehér vagy színes Gauss zaj bemenetet feltételeznek, de az eljárás hozzáigazítható más

bemenőjel-típusokhoz is. A legkisebb négyzetes predikciós hibát minimalizáló  $\hat{\theta}_N$  becslésre koncentrálunk.

Egy egyszerű harmadrendű folyamatmodell és kimeneti additív zajmodell esetén a paramétervektor  $\theta = [b_1 \ b_2 \ b_3 \ f_1 \ f_2 \ f_3]^T$  alakú,  $\theta_0 = [0.0880 \ 0.0195 \ -0.0442 \ -2.0536 \ 1.4611 \ -0.3442]^T$  nominális értékekkel. A rendszer domináns időállandója 5 mintavételnyi. A 2. ábra mutatja a becslések várható szórásnégyzetét az  $1/N$  lecsengés kompenzálása után (különböző bemenőjel- és zaj realizációk átlagaként). A példán láthatóak a szórásnégyzet becslésének főbb jellemzői: *i.* az aszimptotikus varianciabecslés alulbecsli a valós varianciát; *ii.* a valós és az aszimptotikus értékek között jelentős eltérés lehet alacsony  $N$  mintaszám esetén; *iii.* a hiba különbözőképpen csökken számláló és nevező együtthatók esetében.



Ábra 2: A paraméterbecslések kompenzált varianciája a mintaszám függvényében.

**Definíció 5 (Véges mintaszámú varianciavesztés)** Egy  $\theta$  skalár paraméter esetén  $\sigma_{N|U}^2(\hat{\theta}_N)$  jelölje az  $N$  minta alapján számolt  $\hat{\theta}_N$  becslésnek az  $u$  bemenőjelre vett feltételes varianciáját. Az aszimptotikus varianciát a  $\sigma_{\infty|U}^2(\hat{\theta}) = \lim_{N \rightarrow \infty} N\sigma_{N|U}^2(\hat{\theta}_N)$  határértékként definiáljuk. Az  $N$  mintaszámhoz tartozó véges mintaszámú varianciavesztés

$$\alpha_{N|U}(\hat{\theta}_N) = \frac{N\sigma_{N|U}^2(\hat{\theta}_N)}{\sigma_{\infty|U}^2(\hat{\theta})}$$

A  $\theta$  paraméter várható véges mintaszámú varianciavesztése

$$\bar{\alpha}_N(\hat{\theta}_N) = \mathbb{E}_U \left( \alpha_{N|U}(\hat{\theta}_N) \right)$$

Olyan esetben, amikor az  $U$  bemenőjel különböző realizációi rendelkeznek ugyanazokkal az aszimptotikus tulajdonságokkal (p.l. multiszínusz jelek különböző kezdő fázissal, de azonos spektrummal, vagy egy stacionárius sztochasztikus folyamat realizációi), a  $\bar{\alpha}_N$  várható varianciavesztés a varianciavesztés azon részét ragadja meg, amelyek az aszimptotikus jellemzőktől függenek. Ellenben  $\alpha_{N|U}$  az aktuális  $U$  realizációtól függ.

A varianciavesztés pontos számolása ugyanolyan nehéz, mint a becslések varianciájának pontos megadása, ami nehéz és analitikusan nem kezelhető feladat a legtöbb esetben. Ez az

egyik oka annak, hogy miért használunk aszimptotikus eredményeket véges mintaszám esetén is. A célom egy olyan közelítő eljárás bemutatása, amely segítségével közelíthető a varianciaveszteség abban az esetben, ha az  $E$  zaj fehér Gauss zaj és a  $\hat{\theta}_N$  becslés torzítatlan. Ennek a közelítésnek jelentősen kisebb a számításiigénye és elfogadható eredményeket ad.

Bevezetve a  $\phi[k] \in \mathbb{R}^{1 \times n_\theta} = \left. \frac{\partial \hat{y}[k]}{\partial \theta} \right|_{\theta=\theta_0}$  sorvektorokat és a  $\Phi \in \mathbb{R}^{N \times n_\theta} = (\phi[k])_{k=1}^N$  mátrixot, valamint felhasználva a  $\hat{\theta}_N$  becslés torzítatlanságára tett feltételt, a becslés kovariancia mátrixa alulról becsülhető a Cramér-Rao alsó korláttal, vagyis

$$\mathbb{E} \left( (\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T \middle| U \right) \geq M_U^{-1}$$

ahol

$$M_U = \frac{1}{\sigma_y^2} \mathbb{E} (\Phi^T \Phi \mid U)$$

A 2. algoritmus egy momentumillesztés alapú közelítő módszer, ami a  $\bar{\alpha}_N((\hat{\theta}_N)_i)$  várható varianciaveszteség kiértékelését teszi lehetővé.

---

### Algoritmus 2 A várható varianciaveszteség közelítése

---

- 1: Számoljuk a  $\Psi = \mathbb{E}(\phi^T[N]\phi[N])$  kovariancia mátrixot.
- 2: Legyen  $X \in \mathbb{R}^{\kappa N \times n_\theta}$  egy  $\Psi$  kovariancia mátrixú normális eloszlású oszlopokból álló mátrix.
- 3: A  $\kappa N$  szabadsági fokot  $X^T X$  és  $\Phi^T \Phi$  közötti momentumillesztéssel határozzuk meg.
- 4: A várható varianciaveszteséget az alábbi formában közelítjük

$$\bar{\alpha}_N((\hat{\theta}_N)_i) \approx \frac{1}{\sigma_\infty^2((\hat{\theta}_N)_i)} \frac{N}{\kappa N - n_\theta - 1} (\Psi^{-1})_{i,i}$$


---

Az algoritmus harmadik lépésében szereplő momentumillesztést egy  $f(X^T X, \Phi^T \Phi)$  skalár egyenlet gyökének megkeresésével oldhatjuk meg. Egydimenziós problémák esetén a triviális választás a  $\mathbb{E}(X^T X - \Phi^T \Phi) = 0$  egyenlet. Magasabb rendű problémák esetén jó megoldások az  $i$ . diagonális elem illesztése, vagy a mátrixok determinánsának illesztése  $\mathbb{E}(|X^T X|) = \mathbb{E}(|\Phi^T \Phi|)$ .

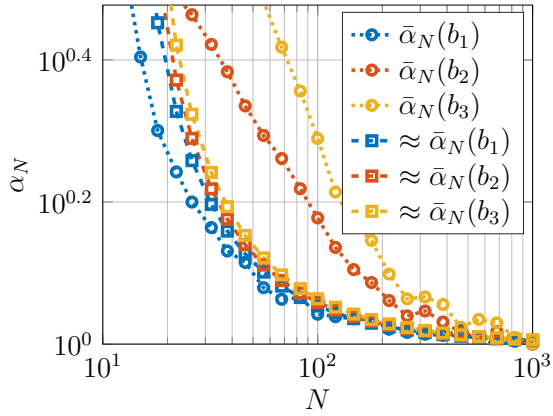
A példarendszer esetén a várható varianciaveszteség és annak közelítésének viselkedését a mintaszám függvényében a 3. ábra mutatja.

A várható varianciaveszteség viselkedéséből több következtetés is levonható. A varianciaveszteség eloszlása nem egy értékre koncentrálódik alacsony mintaszám esetén, tehát a várhatóértéke alatt és felett is lehet. Ez azt mutatja, hogy a hangsúly a gerjesztő jel aszimptotikus tulajdonságairól a konkrét realizációra kell tolnodjon, amikor rövid mérések gerjesztését tervezzük.

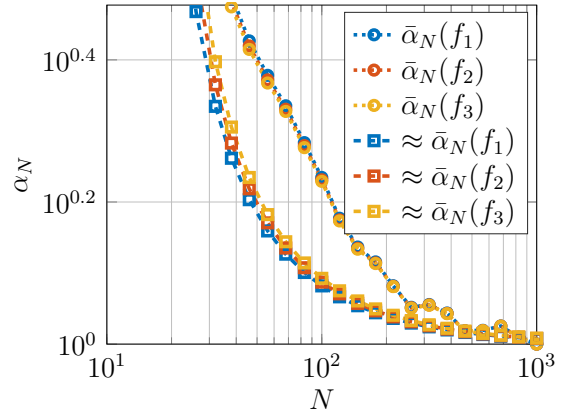
A varianciaveszteség jelentős marad elég magas mintaszám mellett is, 100 jellemző időállandó az példa esetében.

Mielőtt egy identifikációs feladatot megoldanánk, a mérési regisztrátum hosszát úgy választhatjuk meg, hogy a becslések közelítő varianciája, vagy azok varianciavesztesége egy előre meghatározott érték alá kerüljön. Ez hosszabb méréseket igényel mintha ugyanezeket a határokat az aszimptotikus eredményekből számolt varianciával szeretnénk teljesíteni.





(a)  $b$ . számláló együtthatók



(b)  $f$ . nevező együtthatók

Ábra 3: A paraméterek empirikus és közelítő várható varianciavesztése. Az empirikus görbéken kerek jelölők, míg a javasolt közelítés által szolgáltatott görbéken négyzetek jelzik a mért pontokat. Azonos színű vonalak jelölik az azonos paraméterre vonatkozó görbéket.

Miután a mérések rendelkezésre állnak és a becslés is megtörtént, az ehhez tartozó aszimptotikus varianciabecsléseket korigálni lehet, hogy a véges mintaszámú varianciavesztését is figyelembe vegyék. Minden  $\theta_i$  paraméter  $\alpha_{N|U}((\hat{\theta}_N)_i)$  varianciavesztése közelíthető (a javasolt algoritmussal, vagy szimulációk alapján). Feltéve, hogy a becslés aszimptotikus kovariancia mátrixa  $\Sigma_\infty$ , egy korigált véges mintaszámú kovariancia mátrixot kaphatunk  $\Sigma_\infty$  sajátvektorainak transzformációjával, a következő módon. Legyen  $\alpha \in \mathbb{R}^{n_\theta}$  az egyes paraméterek  $\left(\alpha_{N|U}((\hat{\theta}_N)_i)\right)_{i=1}^{n_\theta}$  varianciavesztését tartalmazó vektor és definiáljuk az  $r(v, \alpha)$  átskálázó operátort mint  $r(v, \alpha) : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_\theta} \rightarrow \mathbb{R}^{n_\theta}; (r(v, \alpha))_i = v_i \sqrt{\alpha_i}$ . A véges mintaszámú kovariancia mátrix a következőképpen számolható

$$\Sigma_N = N \sum_{i=1}^{n_\theta} \lambda_i r(v_i, \alpha) r^T(v_i, \alpha)$$

ahol  $\lambda_i$  és  $v_i$   $\Sigma_\infty$  sajátértékei és sajátvektorai. Ez az eljárás az aszimptotikus kovariancia mátrix egy többdimenziós átskálázását biztosítja, amivel megtartja a konfidenciahalmazok leírásának egyszerű strukturáját a véges mintaszám hatásainak figyelembevételével.

### 3 A dolgozathoz kapcsolódó publikációk listája

#### Folyóiratban publikált cikkek

- [1] S. Kolumbán, I. Vajk, and J. Schoukens, “Perturbed datasets methods for hypothesis testing and structure of corresponding confidence sets,” *Automatica*, vol. 51, no. 0, pp. 326 – 331, 2015.
- [2] S. Kolumbán, I. Vajk, and J. Schoukens, “Approximation of confidence sets for output error systems using interval analysis,” *Control Engineering and Applied Informatics*, vol. 14, no. 2, pp. 73–79, 2012.

#### Konferenciakiadványokban publikált cikkek

- [3] S. Kolumbán, “Effect of reduced sample count on estimated uncertainty,” in *Proceedings of the Automation and Applied Computer Science Workshop 2014*, 2014.
- [4] S. Kolumbán, “Sps toolbox for matlab,” in *Proceedings of the Automation and Applied Computer Science Workshop 2013*, 2013.
- [5] S. Kolumbán, “Reliable parameter estimation of pharmacokinetical double peak phenomenon models,” in *Proceedings of the Automation and Applied Computer Science Workshop 2012*, 2012.
- [6] S. Kolumbán, “On semidefinite representability of constraints,” in *Proceeding of Automation and Applied Computer Science Workshop 2011*, 2011.
- [7] S. Kolumbán, “Application of semidefinite programming in system identification,” in *Proceedings of the Automation and Applied Computer Science Workshop 2010*, 2010.
- [8] S. Kolumbán and I. Vajk, “Identification aspects of sdp based polynomial optimization relaxations,” in *16th IFAC Symposium on System Identification SYSID 2012*, 2012.
- [9] S. Kolumbán and I. Vajk, “Correlative sparsity structure of polynomial optimization problems in system identification,” in *Proceedings of the TAMOP PhD Workshop*, 2012.
- [10] S. Kolumbán and I. Vajk, “Rendszeridentifikációból származó polinom optimalizálási feladatok ritkasága,” in *SzámOkt 2011 XXI. Nemzetközi Számítástechnikai és Oktatási Konferencia kiadványa*, 2011.
- [11] S. Kolumbán and I. Vajk, “Using semidefinite programming for l1 identification of the arx model,” in *Proc. 11th IEEE International Symposium on Computational Intelligence and Informatics CINTI 2010*, 2010.
- [12] S. Kolumbán and I. Vajk, “Constrained identification of arx models using semidefinite programming,” in *Proceedings of 1st International Scientific Workshop on DCS*, 2010.
- [13] J. Schoukens and S. Kolumbán, “Study of the minimum experiment length to identify linear dynamic systems: a variance based approach,” in *Proc. 2015 IEEE International Instrumentation and Measurement Technology Conference*, 2015.

## 4 A dolgozathoz nem kapcsolódó publikációk listája

### Folyóiratban publikált cikkek

- [14] M. Balázs, G. Horváth, S. Kolumbán, P. Kovács, and M. Telek, “Fluid level dependent markov fluid models with continuous zero transition,” *Performance Evaluation*, vol. 68, no. 11, pp. 1149–1161, 2011.
- [15] A. Dudás, S. Juhász, and S. Kolumbán, “Performance analysis of multi-threaded locking in bucket hash tables,” *ANNALES Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae Sectio Computatorica*, vol. 36, pp. 63–74, 2012.

### Konferenciakiadványokban publikált cikkek

- [16] I. Albert, S. Kolumbán, H. Charaf, and L. Lengyel, “The solution area-based approach of the content-driven template-based layout system,” in *Proc. International Conference on System Science and Engineering*, 2013.
- [17] A. Dudás, S. Juhász, and S. Kolumbán, “Recalibrating fine-grained locking in parallel bucket hash tables,” in *Facing the Multicore-Challenge III 2012, LNCS 7686*, pp. 60–71, 2013.
- [18] A. Dudás and S. Kolumbán, “Performance evaluation of hash tables,” in *Proceedings of the Automation and Applied Computer Science Workshop 2008*, pp. 63–74, Budapest University of Technology and Economics, Department of Automation and Applied Informatics, Jun 2008.
- [19] A. Dudás, S. Kolumbán, and T. Schrádi, “Effect of cache lines in array-based hashing algorithms,” in *Proceedings of the 2011 International Conference on Foundations of Computer Science FCS2011*, pp. 10–15, July 2011.
- [20] S. Kolumbán, S. Juhász, and A. Dudás, “Modelling performance impact of caches on hash table performance,” in *Proc. of the IADIS European Conference on Informatics 2009, part of the IADIS Multiconference of Computer Science and Information systems 2009*, (Algarve, Portugal), pp. 34–42, June 2009.

## 5 Hivatkozások

- [CCGW12] Marco Campi, Balázs Csanád Csáji, Simone Garatti, and Erik Weyer. Certified system identification - towards distribution-free results. In *Proc. 16th IFAC Symposium on System Identification*, pages 245–255, 2012.
- [CCW12a] B.Cs. Csáji, M.C. Campi, and E. Weyer. Non-asymptotic confidence regions for the least-squares estimate. In *Proceedings of the 16th IFAC Symposium on System Identification (SYSID 2012)*, pages 227–232, 2012.
- [CCW12b] B.Cs. Csáji, M.C. Campi, and E. Weyer. Sign-perturbed sums (sps): A method for constructing exact finite-sample confidence regions for general linear systems. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 7321–7326, 2012.

- [CW05] M.C. Campi and E. Weyer. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41(10):1751 – 1764, 2005.
- [Las01] Jean Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.
- [Las06] Jean Bernard Lasserre. Convergent sdp-relaxations for polynomial optimization with sparsity. In Andrés Iglesias and Nobuki Takayama, editors, *Mathematical Software - ICMS 2006*, volume 4151 of *Lecture Notes in Computer Science*, pages 263–272. Springer Berlin / Heidelberg, 2006.
- [WKKM06] Hayato Waki, Sunyoung Kim, Masakazu Kojima, and Masakazu Muramatsu. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM J. on Optimization*, 17:218–242, January 2006.