Dinamikus rendszerek paramétereinek becslése

Nemlineáris dinamikus rendszerek paramétereinek becslése

Hangos Katalin

PE Villamosmérnöki és Információs Rendszerek Tanszék

ISMÉTLÉS

A predikciós hiba minimalizálása

A paraméterbecslési módszer: $D^N \longrightarrow \hat{\theta}_N$

Az általános paraméterbecslési feladat:
Adottak

- mért értékek: $D[1,N] = D^N = \{(y(k), u(k)) \mid k = 1, \dots N\}$
- prediktív parametrizált modell: $\hat{y}(k|\theta) = g(k, D[1, k-1]; \theta)$
  predikciós hiba sorozat (diszkrét idejű jel): $\varepsilon(k,\theta) = y(k) - \hat{y}(k|\theta)\ \ ,\ \ k = 1, \dots, N$
- norma a predikciós hibán: $V_N(\theta, D^N) = \frac{1}{N}\sum_{k=1}^{N} \ell(\varepsilon(k,\theta))$ ahol $\ell(.)$ egy pozitív skalár értékű függvény, leggyakrabban: $\ell(\varepsilon) = \frac{1}{2}\varepsilon^2$

Az ismert D^N mért érték sorozatból és a θ paraméter vektor értékéből kiszámíthatjuk a $V_N(\theta, D^N)$ norma értéket. A $k = N$ időpillanatban válasszuk meg a becsült $\hat{\theta}_N$ paraméter vektort úgy, hogy

$$\hat{\theta}_N = \hat{\theta}_N(D^N) = \arg\min_{\theta} V_N(\theta, D^N)$$

FELADATKITÜZÉS

Nemlineáris rendszerek paraméterbecslési feladata

A nemlineáris és paraméterekben nemlineáris rendszerek paramétereinek becslése az általános paraméterbecslési optimalizálási feladat megoldását igényli.

A paraméterbecslési módszereket az egybemenetű-egykimenetű dinamikus rendszerek példáján mutatjuk be.

A paraméterbecslés, mint nemlineáris optimalizálási feladat 1

Adott:

1. Egy $D^N = \{(y(k), u(k)) | k = 1, \ldots, N\}$ **mért érték sorozat**;
2. **Prediktiv modell**: SISO eset

$$\hat{y}(k|\theta) = g(k, D^{k-1}; \theta)$$

   (Speciális eset: paraméterekben lineáris

$$\hat{y}(k|\theta) = \theta^T g^*(k, D^{k-1}; \theta)$$

   egy $g^*(.)$ adott mért adatokban nemlineáris vektor értékű függvénnyel);
3. Ebből és a mért kimenetekből számolható a **predikciós hiba sorozat**

$$\varepsilon(k, \theta) = y(k) - \hat{y}(k|\theta)$$

4. Egy **veszteségfüggvény** alkalmas $\ell(.)$ vektornormával

$$V_N(\theta, D^N) = \frac{1}{N} \sum_{k=1}^{N} \ell(\varepsilon(k, \theta))$$

A paraméterbecslés, mint nemlineáris optimalizálási feladat 2

Keressük:

azt a paraméterbecslési módszert, amely a mért értékekből előállít egy $\hat{\theta}_{modsz}(D^N)$ becslést úgy, hogy

$$\hat{\theta}_{modsz}(D^N) = \arg\min_{\theta} V_N(\theta, D^N)$$

Speciális eset: paraméterekben nemlineáris, de állandó paraméterű SISO rendszer paramétereit a **legkisebb négyzetes elvű** becsléssel becsüljük.

Veszteségfüggvény:

$$V_N(\theta, D^N) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{2}[y(k) - g(k, D^{k-1}; \theta)]^2$$

Optimalizálási feladat:

$$\hat{\theta}_{LKN}(D^N) = \arg\min_{\theta} \frac{1}{N}\sum_{k=1}^{N}\frac{1}{2}[y(k) - g(k, D^{k-1}; \theta)]^2$$

Az optimalizálási feladat lehetséges megoldási útjai 1

Az adott D^N minta melletti **optimalizálási feladat** célfüggvénye ugyan látszólag kvadratikus, a nemlineáris $g(k, D^{k-1}; \theta)$ függvény jelenléte miatt azonban **analitikusan általában nem megoldható**.

több lokális minimum is lehet

Analitikus megoldás hiányában az optimalizálási feladatot numerikus közelítő eljárásokkal oldhatjuk meg, az alábbi, elvileg különböző két módon.

1. **Nemlineáris egyenletrendszer megoldására vezető út**
   Ez esetben az optimum szükséges feltételét jelentő

$$\frac{1}{N}\sum_{k=1}^{N}\frac{1}{2}[y(k) - g(k, D^{k-1}; \theta)]\frac{\partial g}{\partial \theta_j} = 0 \quad , \quad j = 1, \ldots, N$$

   nemlineáris egyenletrendszert kell megoldani numerikus módszerrel.
   Probléma: **a megoldás általában nem egyértelmű,** és **a konvergencia** gyorsasága (és az, hogy melyik megoldáshoz konvergál) **erősen függ a kezdeti becsléstől**.

Az optimalizálási feladat lehetséges megoldási útjai 2

2. **Direkt megoldás a $\mathrm{V}_{\mathbf{N}}(\theta, \mathrm{D}^{\mathbf{N}})$ veszteségfüggvény θ szerinti minimalizálásával**
   Ez a feladat **numerikus függvényminimalizáló eljárásokkal**, például az úgynevezett gradiens módszerrel **oldható meg.**
   Itt is **problémát jelent, hogy** úgynevezett **globális optimalizálási eljárásokra van (lenne) szükség a több lehetséges lokális minimum miatt.**

A GRADIENS MÓDSZER

A gradiens módszer elve 1

A gradiens módszer egy általános függvény szélsőérték (minimum vagy maximum) kereső eljárás, amely **lokális szélsőérték-keresésre alkalmas**.
Adott egy

$$G \ : \ \mathbb{R}^m \rightarrow \mathbb{R}$$

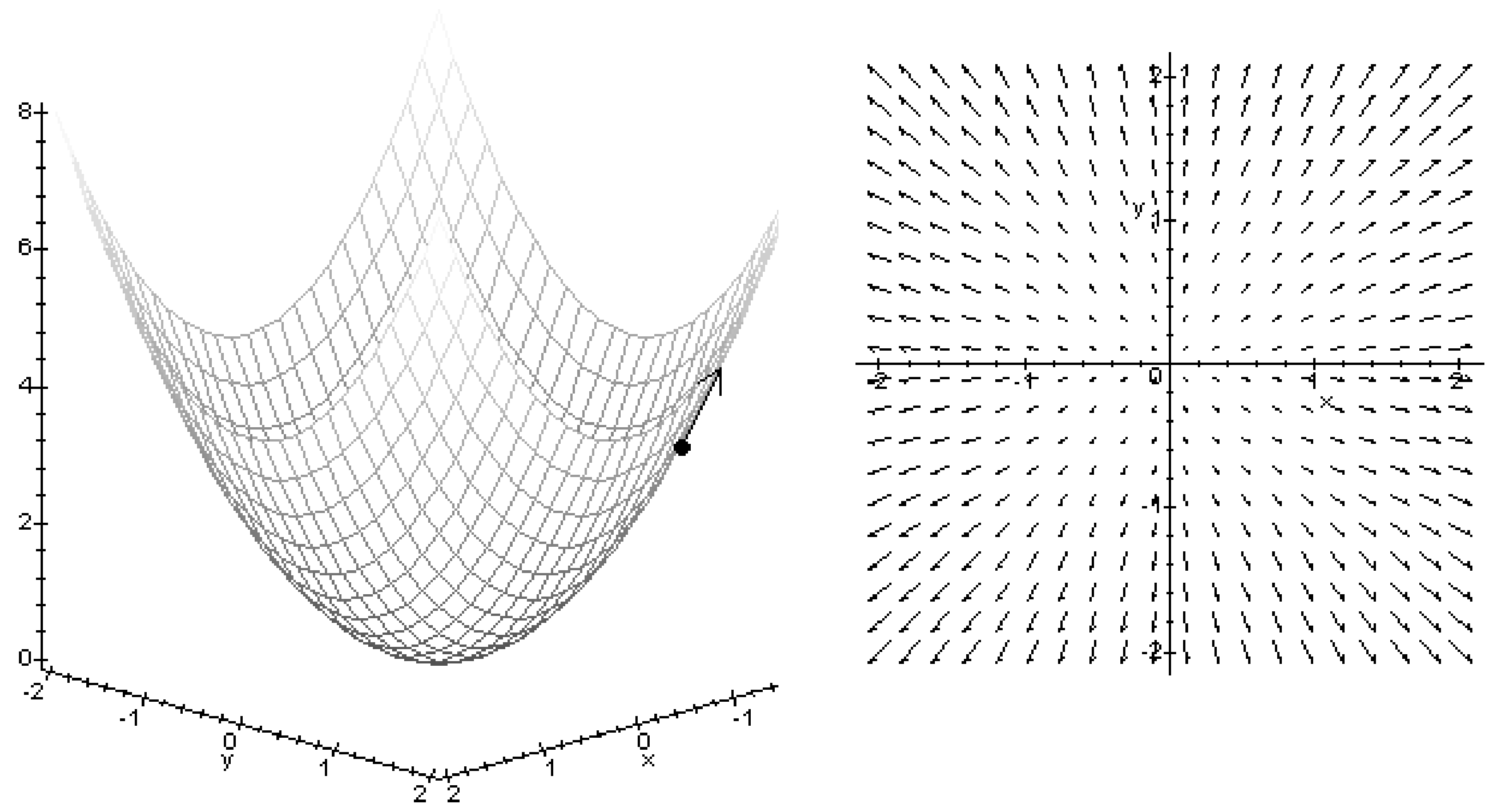
többváltozós skalárértékű nemlineáris függvény. Ennek **gradiense** (gradiens vektora) **valamely** $\mathbf{x}^* \in \mathbb{R}^{\mathbf{m}}$ **pontban** a

$$\mathbf{G_X}(\mathbf{x}^*) = \tilde{\mathbf{grad}}\ \mathbf{G}(\mathbf{x}^*) = [\ \frac{\partial \mathbf{G}}{\partial \mathbf{x_1}}(\mathbf{x}^*) \ \ \ldots \ \ \frac{\partial \mathbf{G}}{\partial \mathbf{x_m}}(\mathbf{x}^*)]^{\mathbf{T}} \qquad (1)$$

m-dimenziós vektor, ahol x_i az x vektor i-edik koordinátája. **A gradiens vektor a** G **függvény**, mint az $(m+1)$-dimenziós $\mathbb{R}^{m+1}$ térben értelmezett felület x^* **pontjában húzott érintők közül a legmeredekebb irányába mutat.**

A gradiens módszer elve 2

Az alábbi ábra egy álló forgási paraboloid felületet mutat. Meghúztuk a felület egy pontjában a gradiens vektort is, az ábra jobb oldali részében pedig a gradiens mező, azaz az egyes pontokban húzható gradiens vektorok vetületei láthatóak.



A gradiens módszer elve 3

A G függvény görbületét, azaz konvex vagy konkáv voltát valamely $x^* \in \mathbb{R}^m$ pontban a $H_G = G_{XX}$ mátrix, az úgynevezett *Hesse mátrix* mutatja, amelynek ij-edik eleme

$$[G_{XX}]_{ij} = \frac{\partial^2 G}{\partial x_i \partial x_j}$$

A $\mathbf{G}$ függvénynek minimuma (lehet, hogy csak lokális minimuma) **van az $\mathbf{x}^*$ pontban, ha**

$$\tilde{\mathbf{grad}}\ \mathbf{G}(\mathbf{x}^*) = \tilde{\mathbf{0}} \quad \textbf{és} \quad \mathbf{G_{XX}}(\mathbf{x}^*) > \mathbf{0}$$

azaz $G_{XX}(x^*)$ Hesse mátrix pozitiv definit.

A gradiens módszer

A gradiens módszer egy iteratív közelítő módszer egy többváltozós függvény szélsőértékének meghatározására. Használatához szükség van

- egy alkalmas $\mathbf{x_0}$ **kezdőérték**re,
- egy ε **pontossági korlát**ra,
- egy δ **lépésköz**re.

Az algoritmus fő lépései minimumkeresés esetén a következők:

1. Legyen $i := 0$, ahol i az iterációs lépések száma és $x_i = x_0$
2. Számítsuk ki a veszteségfüggvény x_i pontbeli $\vec{G}_X(x_i)$ gradiensvektorát.
3. Ha a gradiensvektor már "elég kicsi", azaz $||\vec{G}_X(x_i)|| < \varepsilon$, akkor elértük a minimumot, és $x_{min} = x_i$.
4. Egyébként lépünk egyet a negatív gradiens irányába, azaz

$$x_{i+1} = x_i - \vec{G}_X(x_i)\delta$$

megnöveljük a számlálót, azaz $i := i + 1$, majd a 2. lépéstől folytatjuk.

A veszteségfüggvény minimalizálása gradiens módszerrel

A $V_N(\theta, D^N)$ veszteségfüggvény θ szerinti minimalizálására is használhatjuk a gradiens módszer algoritmusát az alábbi megfeleltetésekkel:

$$\begin{array}{rcl} \mathbf{V_N}(\theta, \mathbf{D^N}) & \sim & \mathbf{G}(\mathbf{x}) \\ \theta & \sim & \mathbf{x} \end{array}$$

A szükséges **induló (a priori) adatok**:

- egy alkalmas θ_0 **kezdő paramétervektor érték**,
- egy ε **pontossági korlát**,
- egy δ **lépésköz a paraméterek terében**.

Fontos megjegyezni, hogy **a gradiens módszer**

- **lokális szélsőérték kereső módszer, azaz a becsült paraméterérték függhet a kezdeti érték jó**, azaz a valódi érték közeli **megválasztásától**,
- egy iterációs lépés **polinomiális időbonyolultságú,** és **a gyakorlati algoritmusok gondoskodnak a δ lépésköz** megfelelő változtatásáról, azaz a **minimum közelében megfelelő csökkentéséről**.

NEMLINEÁRIS MODELLEK MUNKAPONT KÖRÜLI LINEARIZÁLÁSA

Linearizálás: centrált változók

A nemlineáris modellek paramétereinek becslését gyakran munkapont körül linearizált változatuk segítségével végezzük, hiszen a lineáris, de legalábbis paraméterekben lineáris **modellek becslésére számos hatékony** és aszimptotikusan torzítatlan **módszer áll rendelkezésre**.

Irányítási feladatoknál, így dinamikus rendszerek paramétereinek becslésénél is általában $\widetilde{\mathbf{x}}$ **munkapont körüli,** azaz úgynevezett **perturbációs változókkal dolgozunk, amelyeket valamely $\mathbf{x_0}$ referencia pont vagy állandósult állapottól való eltérésükkel definiálunk:**

$$\widetilde{\mathbf{x}} = \mathbf{x} - \mathbf{x_0}$$

ahol x_0 az x változó állandósult állapotbeli értéke.

Munkapont körüli linearizálás 1

A linearizálás a linearizálandó nemlineáris függvény állandósult állapot körüli Taylor sorfejtésén alapszik. Egy egyváltozós dinamikus nemlineáris $y = f(x,t)$ függvénynél az x változó szerinti Taylor sorfejtés az alábbi alakban írható:

$$f(x,t) = f(x_0,t) + \left.\frac{df}{dx}\right|_{x_0} (x - x_0) + \left.\frac{d^2 f}{dx^2}\right|_{x_0} \frac{(x-x_0)^2}{2!} + \ldots \qquad (2)$$

ahol x_0 a referencia pont és $\frac{d^i f}{dx^i}$ az f függvény x szerinti i-edik parciális deriváltja, amelyet az x_0 referencia pontban veszünk.

A fenti sorfejtés első két tagja utáni magasabbrendű tagok elhanyagolásával kapjuk a lineáris közelítést:

$$f(x,t) \simeq f(x_0,t) + \left.\frac{df}{dx}\right|_{x_0} (x - x_0)$$

és így az eredeti nemlineáris egyenlet az

$$y = f(x_0,t) + \left.\frac{df}{dx}\right|_{x_0} (x - x_0)$$

formulával közelíthető.

Függvények linearizálása

A $y = f(x_0, t) + \left.\frac{df}{dx}\right|_{x_0} (x - x_0)$ közelítő alakból a $\widetilde{y} = y - y_0 = y - f(x_0, t)$-t helyettesítve

$$\widetilde{\mathbf{y}} = \left.\frac{\mathbf{df}}{\mathbf{dx}}\right|_{\mathbf{x_0}} \cdot \widetilde{\mathbf{x}}$$

amely a végső linearizált egyenlet az $\widetilde{x}$ perturbációs változóval kifejezve.

Többváltozós skalárértéku $f(x_1, \ldots, x_n, t)$ **esetre:**

$$\widetilde{\mathbf{y}} = \mathbf{J}|_{\mathbf{x_0}} \cdot \widetilde{\mathbf{x}}$$

ahol $\mathbf{J}$ a függvény úgynevezett **Jacobi mátrixa** (ez esetben sorvektor), amelynek i-edik eleme a megfelelő x_i változó szerinti parciális derivált, azaz

$$J_i = \frac{\partial f}{\partial x_i}$$

Nemlineáris prediktor linearizálása

A fenti elvet alkalmazzuk az alábbi nemlineáris prediktor linearizálására:

$$\hat{y}(k|\theta) = g(k, D^{k-1}; \theta)$$

Először definiálunk (keresünk) egy alkalmas x_* munkapontot:

$$x_* = (y_*, u_*, \theta_*)$$
$$D_*^N = \{y(k) = y_*, u(k) = u_* | k = 1, \ldots, N\}$$

Ezek után alkalmazzuk a **többváltozós lineáris Taylor közelítést**

$$\widetilde{\hat{\mathbf{y}}}(\mathbf{k}|\theta) = \sum_{\mathbf{i=1}}^{\mathbf{k}} \left[\left.\frac{\partial \mathbf{g}}{\partial \mathbf{u(i)}}\right|_{\mathbf{x}_*} \widetilde{\mathbf{u}}(\mathbf{i}) + \left.\frac{\partial \mathbf{g}}{\partial \mathbf{y(i)}}\right|_{\mathbf{x}_*} \widetilde{\mathbf{y}}(\mathbf{i}) \right]$$

felhasználva azt, hogy a θ paraméterek nem változnak, azaz $\widetilde{\theta} = 0$.