

5. Modellillesztés (folyt.)

Iteratív modell-illesztés LMS-Newton módszerrel:

A (114) összefüggésben az $X(n)$ regressziós vektor normálása elvileg az R mátrixszal is lehetséges. Ha tehát abban a különös helyzetben lennénk, hogy ismerjük az R mátrixot, és a gradienst pedig pillanatnyi értékével becsüljük, akkor

$$W(n+1) = W(n) + 2\mu R^{-1} X(n)e(n) \quad (118)$$

$$V(n+1) = \left[\prod_{i=0}^n (I - 2\mu R^{-1} X(i)X^T(i)) \right] V(0) \quad (119)$$

Ennek a gondolatnak akkor van gyakorlati jelentősége, ha az R mátrixot a megfigyeléseinkből iteratív úton ugyancsak előállítjuk.

Iteratív modell-illesztés LMS-Newton módszerrel, R iteratív becslésével:

$$W(n+1) = W(n) + 2\mu R^{-1}(n+1) X(n)e(n) \quad (120)$$

ahol $R(n+1) = \lambda R(n) + \nu X(n)X^T(n)$, $0 < \lambda < 1$, $0 < \nu < 1$, $\lambda = 1 - \nu$, ($\lambda = 0.9 \dots 0.99$). (A rekurzív átalakítás származtatására lásd még a lineáris és az exponenciális átlagolásról szóló 3. megjegyzést.) $R(n+1)$ inverzét az iteratív számítást nagymértékben könnyítő, ún. mátrix inverziós lemma felhasználásával írjuk fel.

$$R^{-1}(n+1) = \frac{1}{\lambda} \left[R^{-1}(n) - \frac{R^{-1}(n)X(n)X^T(n)R^{-1}(n)}{\frac{\lambda}{\nu} + X^T(n)R^{-1}(n)X(n)} \right] \quad (121)$$

Megjegyzések:

1. A mátrix inverziós lemma: $[A + BC]^{-1} = A^{-1} - A^{-1}B[I + CA^{-1}B]^{-1}A^{-1}$. Figyeljük meg, hogy amennyiben, mint esetünkben, BC diád, akkor a jobboldali zárójeles inverz skalár érték lesz. Most $A = \lambda R(n)$, $BC = \nu X(n)X^T(n)$.
2. Az iterációt célszerűen $R(0) = \varepsilon I$ értékkel indítjuk, ahol $0 < \varepsilon \ll 1$.
3. A digitális jelfeldolgozás legalapvetőbb módszerei között kiemelt szerepet kapnak a rekurzív átlagolások.

- Egyszerű átlagolás: a megfigyelések lineáris átlagát képezzük, amivel a várható értéket becsüljük:

$$\hat{x}(n) = \frac{1}{n} \sum_{k=0}^{n-1} y(k) \Rightarrow \hat{x}(n+1) = \frac{1}{n+1} \sum_{k=0}^n y(k) = \frac{n}{n+1} \hat{x}(n) + \frac{1}{n+1} y(n) = \hat{x}(n) + \frac{1}{n+1} [y(n) - \hat{x}(n)]$$

Ugyanez a korrelációs mátrix közelítő számítására:

$$\begin{aligned} \hat{R}(n) &= \frac{1}{n} \sum_{k=0}^{n-1} X(k)X^T(k) \Rightarrow \hat{R}(n+1) = \frac{1}{n+1} \sum_{k=0}^n X(k)X^T(k) = \frac{n}{n+1} \hat{R}(n) + \frac{1}{n+1} X(n)X^T(n) = \\ &= \hat{R}(n) + \frac{1}{n+1} [X(n)X^T(n) - \hat{R}(n)]. \end{aligned}$$

- Exponenciális átlagolás:

$$\hat{x}(n+1) = a\hat{x}(n) + by(n),$$

ahol a és b konstansok. A frekvenciatartománybeli viselkedés leírására használható a z -transzformáció: $z\hat{X}(z) = a\hat{X}(z) + bY(z)$, amiből kifejezhető az exponenciális átlagoló átviteli függvénye:

$$H(z) = \frac{\hat{X}(z)}{Y(z)} = \frac{b}{z-a} = \frac{bz^{-1}}{1-az^{-1}}$$

Ez egy alul-áteresztő szűrő, amely, ha konstans jelet kap, akkor a tranziens lejátszódása kimenetén ugyanez a konstans jelenik meg. Ehhez pedig az kell, hogy megfelelően legyen normálva, vagyis $H(z) = 1$, ha $z=1$. Ezzel behelyettesítési értéke

$$\frac{b}{1-a} = 1, \text{ vagyis } a = 1-b. \text{ Ezzel}$$

$$\hat{x}(n+1) = \hat{x}(n) + b(y(n) - \hat{x}(n)).$$

Ennek megfelelően ilyenkor a megfigyelt új értéket egy konstanssal szorozzuk, ellentétben a lineáris átlagolással. A $R(n+1) = \lambda R(n) + \nu X(n)X^T(n)$ az exponenciális átlagolással azonos struktúrájú rekurzív eljárás, ahol $\lambda = 1-b$, ill. $\nu = b$ megfeleltetés alkalmazható.

További megjegyzések a modellillesztés témakör egészéhez:

1. A modellillesztés feladatát alapvetően kétféle módszerrel oldhatjuk meg:
 - „batch” vagy „off-line” jelleggel, amikor felvett regisztrátumot utólag dolgozunk fel.
 - iteratív, rekurzív, „on-line” jelleggel, amikor a felvétellel párhuzamos a feldolgozás.
2. A modellillesztés célját illetően is alapvetően két nagy csoportot különböztetünk meg:
 - identifikáció: a lehető legpontosabban szeretnénk megragadni a valóságot,
 - adaptáció: a lehető legjobban szeretnénk követni a valóság változásait.
3. Az adaptációt megvalósító ún. adaptív rendszerek esetében többnyire iteratív/rekurzív eljárásokat alkalmazunk, az identifikáció esetében a kétféle megközelítés lényegében egyenértékű.

Ha a kritériumfüggvény nem négyzetes → egy lehetőség Taylor sorba fejtése:

Ha valamilyen oknál fogva a kritérium függvény nem négyzetes, akkor egy adott $W(n)$ paraméter érték környezetében mindig megkísérelhetjük Taylor sorba fejteni:

$$\begin{aligned} C(y, \hat{y}) &= C(y, \hat{y}(W)) = C(W) \cong \\ &\cong C(W(n)) + [\nabla C(W(n))]^T [W - W(n)] + \frac{1}{2} [W - W(n)]^T H(W(n)) [W - W(n)] + \dots \end{aligned} \quad (122)$$

ahol $H(W(n)) = \left. \frac{\partial \nabla C(W)}{\partial W} \right|_{W=W(n)}$ a második derivált, ami megfeleltethető az R mátrixnak.

Iteratív modellillesztés a kritériumfüggvény Taylor sorfejtése alapján:

- a) A (122) összefüggés első két tagját vesszük figyelembe, és feltételezzük, hogy $C(W(n+1)) = 0$. Ekkor a $C(W(n)) + [\nabla C(W(n))]^T [W(n+1) - W(n)] = 0$ egyenletet kell megoldani.

$$W(n+1) = W(n) - \frac{C(W(n))}{[\nabla C(W(n))]^T \nabla C(W(n))} \nabla C(W(n)) \quad (123)$$

Ez a Newton-Raphson módszer. A tapasztalatok szerint az optimumtól távol kedvezően viselkedik, az optimumhoz „közeli” jósága attól függ, hogy $C(W^*) = 0$ teljesül-e.

- b) Ha a sorba fejtett $C(W)$ minimumát deriválással keressük, akkor a

$\nabla C(W(n+1)) = 0 = \nabla C(W(n)) + H(W(n))(W(n+1) - W(n))$ feltételt kapjuk, amiből a Newton módszer adódik:

$$W(n+1) = W(n) - H^{-1}(W(n))\nabla C(W(n)) \quad (124)$$

Megjegyzés: Az $\frac{1}{2}$ -es szorzó a korrekciós tagból most hiányzik, mert a Taylor sorban szerepel, ellentétben az eddigi négyzetes összefüggésekkel.

Adaptív IIR rendszerek

Hatékonyabb, de sok szempontból problematikusabb a modellillesztés/adaptáció, ha ún. végtelen impulzusválaszú (infinite impulse response: IIR) rendszereket alkalmazunk. Ezek sokféle (realizálási/kiszámítási) struktúra szerint megvalósíthatók, az alábbiakban csak az ún. direkt struktúra szerinti változatot tárgyaljuk. Formálisan továbbra is adaptív lineáris kombinátort alkalmazunk, de az aktuális becslő-érték kiszámításában korábbi becslő értékek is szerepet játszanak (n a diszkrét idő index):

$$\hat{y}(n) = \sum_{k=0}^{M-1} a_k(n)x(n-k) + \sum_{k=1}^{N-1} b_k(n)\hat{y}(n-k), \quad (125)$$

azaz

$$W^T(n) = [a_0(n), a_1(n), \dots, a_{M-1}(n); b_1(n), b_2(n), \dots, b_{N-1}(n)], \quad (126)$$

$$X^T(n) = [x(n), x(n-1), \dots, x(n-M+1); \hat{y}(n-1), \hat{y}(n-2), \dots, \hat{y}(n-N+1)]. \quad (127)$$

Ha ezekre alapozva használjuk az eddig megismert számítási eljárásokat, akkor ún. pszeudolineáris regressziót (PLR) hajtunk végre. Ennek során formálisan nem veszünk tudomást arról, hogy a regressziós vektor az illesztendő modell (azaz az adaptív szűrő) korábbi kimenőjel-mintáitól is függ (ún. implicit függés), aminek közvetlen következménye, hogy a hibafelület már nem paraboloid.

Visszavezetés véges impulzusválaszú (FIR) problémára (Equation-Error Formulation):

Az alábbiakban az adaptív szűrő átviteli függvényének ($H(z)$) számlálójával ($N(z)$) és nevezőjével ($D(z)$) mint „operátorokkal” hozakodunk elő, ezért keveredik látszólag a diszkrét idő, és a diszkrét frekvenciatartomány:

$$H(z) = \frac{N(z)}{D(z)} = \frac{\hat{Y}(z)}{X(z)}, \quad (128)$$

a közelítés hibája pedig, amit az eddigiekben négyzetes értelemben minimalizáltunk $e(n) = y(n) - \hat{y}(n)$. Ne ezt minimalizáljuk, hanem az $e_e(n) = D(z)e(n)$ hibát, mert (128) felhasználásával

$$e_e(n) = D(z)e(n) = D(z)y(n) - D(z)\hat{y}(n) = D(z)y(n) - N(z)x(n) \quad (129)$$

alakban írható, ami viszont nem függ $\hat{y}(n)$ -től. Bevezetve az

$N(z) = A(n, z) = \sum_{k=0}^{M-1} a_k(n)z^{-k}$, és a $D(z) = 1 - B(n, z)$, ahol $B(n, z) = \sum_{k=1}^{N-1} b_k(n)z^{-k}$ jelöléseket:

$$e_e(n) = y(n) - B(n, z)y(n) - A(n, z)x(n) = y(n) - y_e(n), \quad (130)$$

ahol (126) felhasználásával $y_e(n) = W^T(n)X_e(n)$. Itt (v.ö. a (127) összefüggéssel)

$$X_e^T(n) = [x(n), x(n-1), \dots, x(n-M+1); y(n-1), y(n-2), \dots, y(n-N+1)]. \quad (131)$$

Innentől bátran alkalmazható minden olyan módszer, ami az adaptív lineáris kombinátor megközelítésben megfelelőnek bizonyult. A módszer „blokkvázlata” a 27. ábrán látható.

Megjegyzés: zajos megfigyelés esetén torzítás lép(het) fel a paraméterbecslés során, azaz a becslő várható értéke nem fog megegyezni a paraméter „ideális” értékével.

A kimeneten értelmezett hibán alapuló modellillesztések (Output-Error Formulation):

Megjegyzés: az előadáson nem szerepelt, kérem, hogy az óravázlat alapján készüljenek fel!

1. Gradiens alapú eljárások: pillanatnyi gradiens (tehát LMS jellegű) esetben az $e^2(n) = e_o^2(n)$ „minimalizálására” törekszünk. Ilyenkor

$$\hat{\nabla}(n) = -2e(n) \frac{\partial \hat{y}(n)}{\partial W(n)} = -2e(n) \nabla \hat{y}(n), \quad (132)$$

ahol $\nabla^T \hat{y}(n) = \left[\frac{\partial \hat{y}(n)}{\partial a_0(n)}, \frac{\partial \hat{y}(n)}{\partial a_1(n)}, \dots, \frac{\partial \hat{y}(n)}{\partial a_{M-1}(n)}, \frac{\partial \hat{y}(n)}{\partial b_0(n)}, \frac{\partial \hat{y}(n)}{\partial b_1(n)}, \dots, \frac{\partial \hat{y}(n)}{\partial b_{N-1}(n)} \right]$, ill. ezen belül:

$$\frac{\partial \hat{y}(n)}{\partial a_k(n)} = x(n-k) + \sum_{i=1}^{N-1} b_i(n) \frac{\partial \hat{y}(n-i)}{\partial a_k(n)}, \quad \frac{\partial \hat{y}(n)}{\partial b_k(n)} = \hat{y}(n-k) + \sum_{i=1}^{N-1} b_i(n) \frac{\partial \hat{y}(n-i)}{\partial b_k(n)}. \quad (133)$$

A (133) összefüggések alapján jól látható az implicit függés mibenléte: Az IIR szűrők végtelen memóriája miatt az aktuális „paraméter-érzékenységre” valamennyi korábbi bemenőjel-minta kihat. Ha elegendően kis lépésközzel dolgozunk az iterációk során (azaz kicsi a μ bátorsági tényező), akkor alkalmazható a következő közelítés:

$$\frac{\partial \hat{y}(n-i)}{\partial a_k(n)} \cong \frac{\partial \hat{y}(n-i)}{\partial a_k(n-i)}, \quad \text{ill.} \quad \frac{\partial \hat{y}(n-i)}{\partial b_k(n)} \cong \frac{\partial \hat{y}(n-i)}{\partial b_k(n-i)}. \quad (134)$$

A (134) közelítés azért előnyös, mert a közelítő gradiens szűrőszerűen számítható: a szűrők bemenőjele $x(n-k)$, ill. $\hat{y}(n-k)$, kimenőjele pedig $\frac{\partial \hat{y}(n)}{\partial a_k(n)}$, ill. $\frac{\partial \hat{y}(n)}{\partial b_k(n)}$, és összesen annyi szűrő kell, ahány beállítandó paraméter. A szűrők blokkvázlata a 28. ábrán látható. Az így megvalósított eljárás Recursive Prediction Error (RPE) néven található az irodalomban. Az elnevezés arra utal, hogy az $n+1$ indexű, „jósolt” paraméter számításánál a korrekciós tagot rekurzív szűrővel számítjuk. A módszer nyilvánvaló hátránya, hogy annyi egyforma szűrő kell, ahány paramétert állítunk. Egy további közelítést alkalmazó egyszerűsítést a 29. ábrán láthatunk. Ez az egyszerűsített RPE eljárás.

2. Még inkább közelítő gradiens eljárások állíthatók elő úgy, hogy (133) „visszacsatolt” tagjait elhagyjuk, ezzel lényegében visszajutunk a pszeudolineáris regresszió (PLR) eljáráshoz. Formailag ebben az esetben is tipikusan az LMS, ill. az LMS/Newton eljárást alkalmazzuk.
3. A kimeneti hibára alapozott adaptív IIR eljárások esetén a paraméter-sík felett értelmezhető hibafelület nem paraboloid, és nem feltétlenül egyetlen minimumhellyel rendelkezik. Ilyenkor a gradiens módszerek könnyen juttatnak lokális minimumba. A hibafelületet egyetlen minimumhelyűvé transzformálhatjuk egy másfajta hibadefinícióval:

Szűrt-hiba eljárás (Filtered-error (FE) Algorithm):

$$e(n) \Rightarrow e_f(n) = [1 + C(n, z)]e(n), \quad \text{ahol } C(n, z) \text{-t úgy választjuk, hogy az } \frac{1 + C(n, z)}{1 + B_*(z)} \text{ átviteli}$$

függvény szigorúan pozitív valós részű (Strictly Positive Real: SPR), mert akkor a globális konvergencia biztosítható. $B_*(z)$ az ideális W^* -hoz tartozó $B(n, z)$.

Megjegyzés: Az SPR tulajdonság mibenlétét jól illusztrálja például egy soros RC tag impedanciájának függvénye: $Z(j\omega) = R + 1/j\omega C$. Ebből jól látható, hogy pozitív a valós rész, aminek következménye zárt áramkörben a kondenzátoron tárolt energia disszipációja, és ezzel az áramkör minimális energiaállapotának elérése, azaz a konvergencia.

Az adaptív IIR algoritmusok általános formája:

$$\boxed{W(n+1) = W(n) + 2\mu R^{-1}(n+1)X_f(n)e_f(n)}, \quad (135)$$

ahol $X_f(n)$ a “szűrt” információs (vagy regressziós) vektor, $e_f(n)$ pedig a “szűrt” hiba vektor. Elnevezések: Ha $X_f(n) = X_e(n)$, $e_f(n) = e_e(n)$, akkor Recursive Least Square (RLS) algoritmusról, ha ráadásul $R(n+1) = I$, akkor Least Mean Square (LMS) algoritmusról beszélünk.

Megjegyzés: Az alábbiak már szerepeltek az előadáson!

Stabilitáselméleti megközelítés: Az LMS eljárásán keresztül bemutatva.

$W(n+1) = W(n) + 2\mu e(n)X(n)$, $V(n+1) = [I - 2\mu X(n)X^T(n)]V(n)$. Ez utóbbi egy autonóm rendszer, amely globálisan aszimptotikusan stabil esetben: $V(n) \Rightarrow 0$, amivel $W(n) \Rightarrow W^*$.

Ljapunov módszerével keresünk egy alkalmas energifüggvényt: $G(n) = V^T(n)V(n)$ megfelelő. Azt keressük, hogy ez miként csökkenthető: $\Delta G(n+1) = G(n+1) - G(n) \leq 0$, minden n -re. Ha G_0 véges, akkor $\Delta G \rightarrow 0$.

$$\begin{aligned} \Delta G(n+1) &= V^T(n+1)V(n+1) - V^T(n)V(n) = \\ &= V^T(n)[I - 2\mu X(n)X^T(n)]^T [I - 2\mu X(n)X^T(n)]V(n) - V^T(n)V(n) = \\ &= -4\mu e^2(n)(1 - \mu X^T(n)X(n)). \end{aligned} \quad (136)$$

Mivel $\mu > 0$, ezért ha $0 < \mu < \frac{1}{X^T(n)X(n)} \quad \forall k$ -ra, akkor $\Delta G \Rightarrow 0$ magával vonja $\mu e^2(n) \rightarrow 0$, illetve $V^T(n)X(n) \rightarrow 0$.

Megjegyzések:

1. A $0 < \mu < \frac{1}{X^T(n)X(n)}$ implicit módon azt is tartalmazza, hogy az $X(n)$ korlátos, és $1/[1 - B(n, z)]$ stabil. (Az átviteli függvény pólusai az egységsugarú körön belül helyezkednek el.)
2. $e(n) = (W(n) - W^*)^T X(n)$ lehet azért nulla, mert a két vektor ortogonális. Ezt nyilván kerülni kell.

