Parametriset mallit

- Rakenteelliset mallit
  - parametreillä a priori tulkinta & merkitys
- Black box-mallit
  - parametrit vain laskennan/sovituksen apuvälineitä
- Tarkastellaan pääosin lineaarisia diskreettiaikaisia black-box-malleja
  - 3. harjoitustyössä malli epälineaarinen jatkuva-aikainen
- Tämän päivän teemoja
  - malliluokat
  - parametrien estimointi
  - mallin hyvyys vs. parametriestimaattien harhaisuus ja varianssi
  - systeemin / mallin identifioituvuus

Rakenteelliset mallit

- Fysikaalisin tai vastaavin perustein rakennettuja malleja
  - osa parametreista tunnettu: esim. massat, poikkipinta-alat jne...
  - osa estimoitava: kitkakertoimet,...
- Merkintä $$\frac{d}{dx}x(t) = f(x(t),u(t),\theta); \hat{y}(t\,|\,\theta) = h(x(t),u(t),\theta)$$

  jossa y^(t|θ) on mallin ennustettu ulostulo hetkellä t ja parametrivektorilla θ
- Edellä oletetaan valkoinen mittauskohina y(t)=h(x(t),u(t),θ)+e(t)
  - kohinaa ei voida ennustaa => asetetaan = 0
  - jos kohinalla on rakennetta, se kannattaa huomioida, esim. ARMA-malli
- Ennustevirhe ε(t)=y(t)-y^(t,θ)
- Estimointiongelma: etsi θ s.e. min $V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\theta)$

Black box -mallit

- Yleinen lineaarinen diskreettiaikainen malli on muotoa $y(t)=\eta(t)+w(t)$
  - $w(t)$ = häiriötermi
  - $\eta(t)$ = häiriötön ulostulo
- Termit muotoa $\eta(t)=G(q,\theta)u(t)$, $w(t)=H(q,\theta)e(t)$
  - G() ja H() lineaarisia suotimia
  - käytännössä rationaalisia (muotoa polynomi/polynomi) ja (asymptoottisesti) stabiileja
  - $G(q,\theta)=B(q,\theta)/F(q,\theta)$, $H(q,\theta)=C(q,\theta)/D(q,\theta)$
  - parametrivektori θ koostuu polynomien B,F,C ja D kertoimista b_i,f_i,h_i,d_i ja kohinan e(t) varianssista
  - "rakenneparametrit" n_b, n_c, n_d, n_f ja n_k (kuollut aika)

Erilaisia mallirakenteita

- Box-Jenkins (BJ): $G(q)=B(q)/F(q)$, $H(q)=C(q)/D(q)$
  - ”täydellinen” malli
  - systeemin ja kohinan mallit riippumattomia toisistaan
- Output error (OE): $H(q)=1$ eli $n_c=n_d=0$
  - kohina valkoista $w(t)=e(t)$, eli ei erityistä rakennetta
  - käytännössä $w(t)$ on poikkeama todellisen ja mitatun ulostulon välillä
- ARMAX: $A(q)y(t)=B(q)u(t)+C(q)e(t)$
  - kohina kokee saman dynamiikan kuin u
  - järkevää kun kohina tulee prosessiin sen alkupäässä
- ARX: $A(q)y(t)=B(q)u(t)+e(t)$ (myös *yhtälövirhemalli*)
  - ennustevirhe lineaarinen parametrien suhteen => parametrien estimointi ”helppoa”
  - kohina kokee saman dynamiikan kuin u; ei haittaa, jos signaali-kohina-suhde on hyvä
- Käyttö: määrää asteluvut, estimoi parametrit
  - käytännössä useita eri mallirakenteita ja astelukuja verrataan

Ennustaminen

- Usein mallin käyttötarkoitus ennustaminen
- Millainen on em. mallien ennuste?
  - OE: $\hat{y}(t,\theta)=G(q,\theta)u(t)$
  - ARX: $\hat{y}(t,\theta)=(1-A(q))y(t)+B(q)u(t)$
- Entä kun C poikkeaa 1:sta (H=C/D)?
  - jaetaan puolittain H:lla ja korvataan e(t) odotusarvollaan

$$\Rightarrow \hat{y}(t,\theta)=[1-H^{-1}(q,\theta)]y(t)+H^{-1}(q,\theta)G(q,\theta)u(t)$$

Mallien sovitus

- Ennustevirhemenetelmät: valitse θ s.e. ennuste on ”hyvä”
  - toimiva hyvyyskriteeri ennustevirheen (otos)varianssi
  - ennustevirhe $\varepsilon(t)=y(t)-\hat{y}(t,\theta)$ ja hyvyyskriteeri
    $$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t,\theta)$$
- $\hat{\theta}_N = \arg\min V_N(\theta)$
  - voidaan osoittaa, että $V_N(\theta)$ ei ole tuulesta temmattu
  - PNS on ennustevirhemenetelmien erikoistapaus
  - MIMO-mallit: $\varepsilon^2(t)$ matriisiarvoinen, tarvitaan sopiva reaaliarvoinen kuvaus: det, trace, ...
- $V_N(\hat{\theta})$ on kohinan e(t) varianssin estimaatti

Lineaarinen regressio

- Regressio: malli muotoa $y(t)=\theta^T\phi(t)+e(t)$, $\phi(t)$ vektori, jossa viivästettyjä u(t):n ja y(t):n arvoja
  - θ sisältää viiveoperaattoripolynomien kertoimet
- N mittausta => matriisi $X=[\phi(1),\phi(2),\ldots,\phi(N)]'$ sekä ulostulo $Y(t)=[y(1),\ldots,y(N)]'$ => $\hat{\theta}=(X^TX)^{-1}X^TY$
- Soveltuu ARX-mallien parametrien estimointiin
  - muut malliluokat epälineaarisia parametrien suhteen
- PNS-oletuksien syytä olla voimassa (ks esim. TAP luento nro 9):
  - viivästetyt y:n ja u:n arvot keskenään kollineaarisia (koesuunnitteluongelma) ⇔ parametrit tehottomia (suuri varianssi) ja harhaisia, mutta tarkentuvia
  - e:n homoskedastisuus (vakio varianssi) ⇔ parametrit tehottomia => painotettu PNS-estimointi
  - e:n korreloimattomuus ⇔ parametrit tehottomia, harhaisia eivätkä tarkentuvia! => kohinalle jonkinlainen rakenne

Iteratiivinen minimointi

- Mallit epälineaarisia parametrien suhteen
  - PNS ei onnistu
- Tarvitaan iteratiivinen menetelmä
  - kyseessä epälineaarinen rajoittamaton optimointitehtävä
- Gradienttimenetelmä: $\theta^{k+1}=\theta^k-\alpha^k V_N'(\theta^k)$
  - α^k viivahaun avulla haettava askelpituus
- Toisen kertaluvun menetelmät: sovelletaan Newton-iterointia optimiratkaisun välttämättömiin ehtoihin
  - välttämätön ehto $V_N'(\theta)=0$
  - iterointi $\theta^{k+1}=\theta^k-\alpha^k[V_N''(\theta^k)]^{-1}V_N'(\theta^k)$

Derivaattojen laskenta (kirja, Appendix 9.6)

- Kohdefunktio

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}(y(t_i) - \hat{y}(t_i,\theta))^2$$

- Gradientti

$$V_N'(\theta) = \frac{1}{N}\sum_{t=1}^{N} -\frac{d}{d\theta}\hat{y}(t_i,\theta)(y(t_i) - \hat{y}(t_i,\theta))$$

- Hessen matriisi

$$V_N''(\theta) = \frac{1}{N}\sum_{t=1}^{N}(\frac{d}{d\theta}\hat{y}(t_i,\theta))(\frac{d}{d\theta}\hat{y}(t_i,\theta))^T +$$

$$\frac{1}{N}\sum_{t=1}^{N}(\frac{d^2}{d\theta^2}\hat{y}(t_i,\theta))(y(t_i) - \hat{y}(t_i,\theta))$$

  - koko Hessen matriisi => Newton-Raphson –iterointi
  - hylätään viimeinen termi => Gauss-Newton
- Derivaatat lausekkeissa riippuvat mallirakenteesta

Mikä on hyvä malli?

- Mallin laatu on yhteydessä mallin käyttötarkoitukseen
  - hyvä säätösuunnittelumalli voi olla huono simulointimalli
- Mallin laatu liittyy sen kykyyn kuvata systeemin toiminta
  - systeemin ja mallin ulostulot riittävän samanlaiset
- Hyvä malli yleistää, i.e., mallin ominaisuudet eivät riipu estimointidatasta

$\Rightarrow$ Mallin tilastolliset ominaisuudet keino mitata tätä

  - parametrien varianssi:
    - kohinainen systeemi => samalla sisäänmenolla saadaan eri ulostulo => estimoitaessa saadaan eri malli
    - varianssia voidaan pienentää kasvattamalla havaintojen lukumäärää
    - kirjassa puhutaan mallin "variance error":sta
  - parametrien harhaisuus: estimaattien konvergenssi vääriin arvoihin
    - väärä mallirakenne / puutteellinen koesuunnittelu
    - erotettava <u>mallirakenteen</u> vaikutus ja <u>identifiointikokeen</u> vaikutus
    - kirjassa puhutaan mallin "bias error":sta

Parametriestimaattien harhasta

- Mitä tapahtuu, kun N->oo?
- Olkoon ennustevirheen varianssi $E\varepsilon^2(t,\theta)=V(\theta)$
- Jos ennustevirhe $\varepsilon(t,\theta)$ on valkoista kohinaa, niin

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\theta) \xrightarrow[N\to\infty]{} E\varepsilon^2(t,\theta) = V(\theta)\ \text{w.p.1}$$

  ja lisäksi $\theta \xrightarrow[N\to\infty]{} \theta^* = \arg\min V(\theta)$
- Jos ennustevirhe ei ole korreloimaton (esim. väärä malli), yo. pätee ”under very general conditions” edelleen, mutta estimaatti voi olla harhainen ja konvergoitua väärään arvoon vs. ”oikean” mallin estimaatin arvo
- Estimaatti minimoi edelleen ennustevirheen varianssin
  - paras malli väärästä malliluokasta
  - haittaako harha?

Esimerkki: ”väärä” mallirakenne

(”System Identification, Theory for the User,” Ljung 1999)

- Olkoon data peräisin prosessista $y(t)+a_0y(t-1)=b_0u(t-1)+e_0(t)+c_0e_0(t-1)$
- Sovitetaan ARX-malli $\hat{y}(t|\theta)+ay(t-1)=bu(t-1)$
- Ennustevirheen varianssi on $E(y(t)+ay(t-1)-bu(t-1))^2 = \ldots =$
  $r_0(1+a^2-2aa_0)+b^2-2bb_0+2ac_0$
  - ($r_0=Ey^2(t)$, ei riipu a:sta eikä b:stä)
- Ennustevirheen varianssin minimoivat $\hat{a}$, $\hat{b}$: $\hat{a}=a_0-c_0/r_0$;$\hat{b}=b_0$;
  - $E\hat{a}\neq a_0$ eli estimaatti on harhainen
- Ennustevirheen varianssi näillä $\hat{a}$, $\hat{b}$ on $1+c_0^2-c_0^2/r_0$
  - todellisilla parametriarvoilla a_0 ja b_0 varianssi on $1+c_0^2$ eli suurempi
- <u>Siis: vaikka parametriestimaatti on harhainen, se tuottaa pienemmän ennustevirheen varianssin</u>

Parametriestimaatin konvergenssi taajuustasossa

$$\theta^* = \lim_{N\to\infty} \hat{\theta}_N = \arg\min \int_{-\pi}^{\pi} | G_0(e^{i\omega}) - G(e^{i\omega}, \theta) |^2 \frac{\Phi_u(\omega)}{| H_*(e^{i\omega}) |^2} d\omega$$

- $V_N(\theta)$:n minimoiva estimaatti lähestyy arvoa, joka saa mallin taajuusvasteen mahdollisimman lähelle systeemin taajuusvastetta painotettuna
  - ohjauksen spektrillä
  - kohinamallin taajuusvasteen käänteisluvulla
- Koesuunnittelu: valitaan u:n taajuusominaisuudet sopivasti => hyvä sovitus mielenkiintoisilla taajuuksilla

Parametriestimaattien varianssi

- Oletetaan, että ennustevirhe $\varepsilon(t,\theta)$ on valkoista kohinaa
- Tällöin parametriestimaatin $\hat{\theta}_N$ kovarianssimatriisi P_N on

$$P_N = E(\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T \approx \frac{1}{N}\lambda \overline{R}^{-1}$$

jossa $R=\psi(t,\theta_0)\psi^T(t,\theta_0)$ ja $\psi(t,\theta)=d/d\theta\,\hat{y}(t,\theta)$

- P_N riippuu
  - kohinan varianssista
  - datapisteiden lukumäärästä
  - ennusteen gradientista (herkkyys!)
- Parametriestimaatit asymptoottisesti normaalijakautuneita
  - tilastollisen merkitsevyyden testaus t-testillä
- Taajuusvasteen varianssi annetulla parametriestimaatilla riippuu (kirja 9.61)
  - paramaterien lukumäärästä ja kohinan spektristä
  - datapisteiden lukumäärästä ja ohjauksen spektristä

Identifioituvuus

- Identifioituvuus: Voidaanko systeemin/mallin parametrit määrätä yksikäsitteisesti input-output –datasta?
- Olkoon θ_0 hyvyyskriteerin minimoiva parametriestimaatti
- Malli on identifioituva: $\hat{y}(t|\theta_0)=\hat{y}(t|\theta) \Rightarrow \theta=\theta_0$ (I)
- Milloin (I) ei päde?

  1) kaksi erilaista parametrivektoria tuottaa samanlaisen mallin input-output –käyttäytymiseen

  – rakenteellinen identifoituvuus, ymmärretään systeemin ominaisuutena

  2) kaksi erilaista parametrivektoria tuottaa erilaisen input-output-käyttäymiseen, mutta puutteellinen input aiheuttaa samanlaiset ennusteet

  – ”deterministinen identifioituvuus”

Esimerkki (kirja 9.1): tasavirtamoottori

- Valitaan tiloiksi kulma-asento y(t) ja –nopeus ω(t) ja ohjaukseksi jännite u(t):

$$\frac{d}{dt}x(t) = \begin{bmatrix} 0 & 1 \\ 0 & -1/\tau \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \beta/\tau \end{bmatrix} u(t),$$

- Tässä

$$\tau = \frac{JR_1}{fR_1 + k^2}, \quad \beta = \frac{k}{fR_1 + k^2}$$

- Systeemissä on 5 parametriä, mutta mallissa vain 2
  - vaikka mallin parametrit saataisiin estimoitua, niistä saadaan vain 2 yhteyttä systeemin parametrien välille
  - tällä parametroinnilla identifiointi ei onnistu – systeemi ei rakeenteellisesti identifioituva

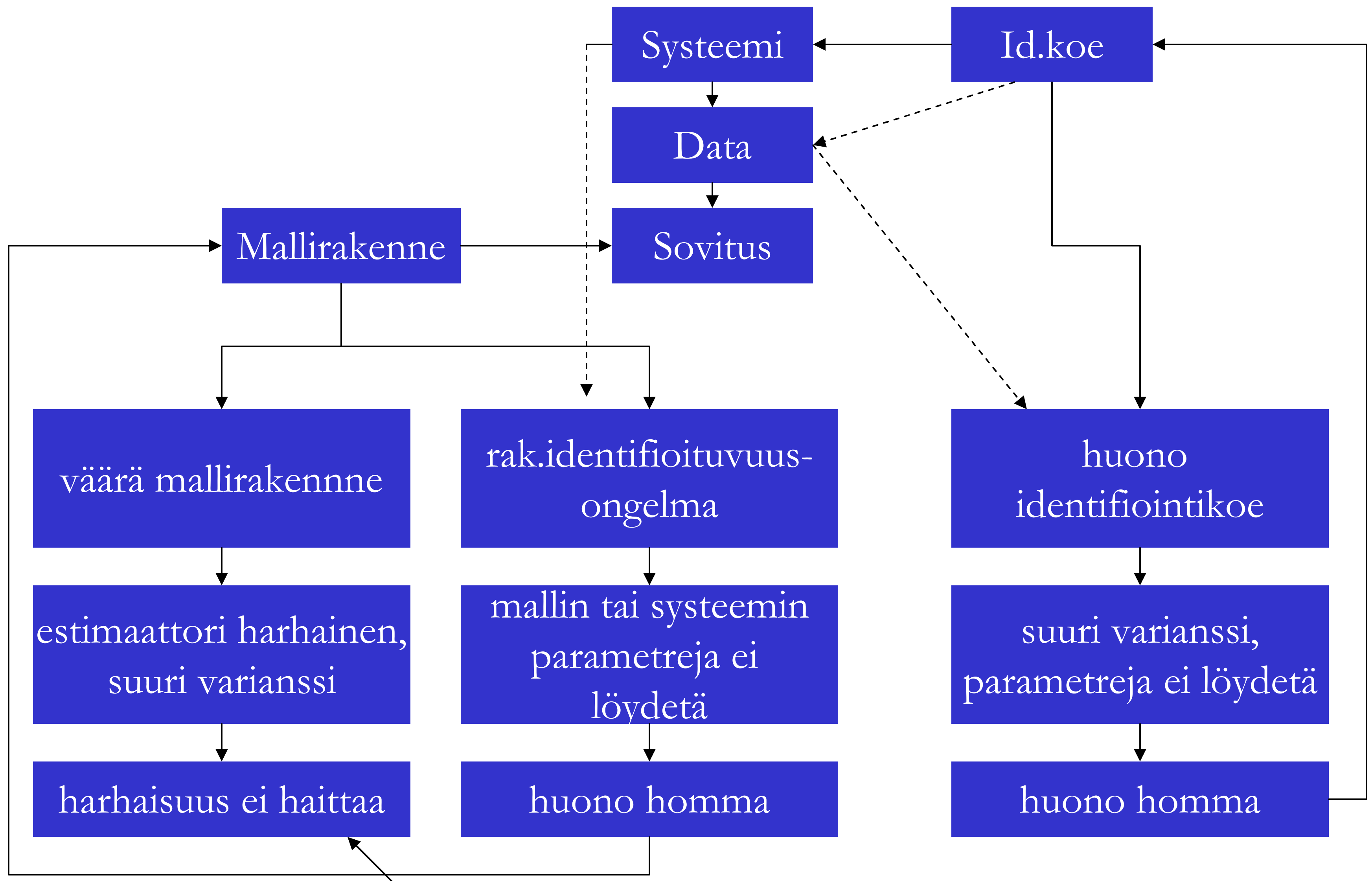
Vaatimukset herätteelle

- Millainen sisäänmeno tarvitaan, jotta parametriestimaatit ylipäätään konvergoisivat?
- Intuitiivisia tuloksia:
  - identifiointikokeen pitäisi herättää systeemin mielenkiintoiset moodit
  - sisäänmenon taajuussisältö oleellisessa asemassa
- Jatkuvasti herättävyyden käsite (seur. luento)
  - kvantitatiivisia tuloksia herätteen laadun ja parametriestimoinnin onnistumisen välille

Esimerkki (kirja 9.6)

- Ennustemalli $\hat{y}(t|\theta)=au(t-1)+bu(t-2)$, $\theta=(a\ b)'$
  - valitaan u vakio-ohjaus u_0:ksi
- Todellinen ennuste on tällöin $\hat{y}(t|\theta)=(a+b)u_0(t-1)$
  - kaikki parametrit a ja b, joiden summa on sama, antavat saman ennusteen
- a ja b eivät ole identifioituvia (tällä herätteellä)
  - systeemi on kuitenkin rakenteellisesti identifioituva

Estimoinnin ongelmalähteet



kunhan mallin käyttötarkoitus ei ole parametrien estimointi!

Yhteenveto

- Perusidea: haetaan parametrivektori joka minimoi ennustevirheen (otos)varianssin:
  - sitä kuvaa neliöllinen hyvyyskriteeri $V_N(\theta)$
  - sovitus: minimoi ennustevirheen varianssi
  - estimaatin varianssi riippuu kohinan varianssista, datan määrästä ja ennusteen herkkyydestä parametrin suhteen
- Lähestymistavan etuja:
  - yleinen käytettävyys, monipuolisuus
  - tuloksena simulointiin soveltuva malli
- Haittoja, muttei välttämättä ylitsepääsemättömiä:
  - periaatteessa tarvitaan näkemys systeemin rakenteesta
  - laaja laskentatuki tarpeen