Doktorantūros ataskaita
III studijų metų II pusmetis

Informatikos studijų programos doktorantas
Vytautas Dulskis

Vilniaus universiteto Duomenų mokslo ir skaitmeninių technologijų institutas

2022 m. rugsėjo mėn. 30 d.

- **Disertacijos tema:**
  Stochastinių dinaminių sistemų, stebimų su triukšmu, filtravimo, identifikavimo ir valdymo realiu laiku algoritmų sudarymas ir taikymas
- **Vadovas:**
  Prof. habil. dr. Leonidas Sakalauskas
- **Doktorantūros pradžios ir pabaigos metai:**
  2018 m. spalio mėn. 1 d. – 2023 m. rugsėjo mėn. 30 d. (akademinėse atostogose 2021 m. rugsėjo mėn. 1 d. – 2022 m. rugpjūčio mėn. 31 d.)

- **Tyrimo objektas:**
    - Stochastinės dinaminės sistemos, stebimos su triukšmu.
- **Tyrimo tikslas:**
    - Sudaryti ir pritaikyti rekursinius algoritmus stochastinių dinaminių sistemų filtravimui, identifikavimui ir valdymui realiu laiku, esant adityviajam sistemų stebėjimo triukšmui.

- **Tyrimo uždaviniai:**
  - Analitiškai apžvelgti su triukšmu stebimų stochastinių dinaminių sistemų filtravimo, identifikavimo ir valdymo realiu laiku uždavinių sprendimo metodus;
  - Sudaryti rekursinius algoritmus tiesinių ir netiesinių stochastinių dinaminių sistemų, stebimų su triukšmu, filtravimui, identifikavimui ir valdymui realiu laiku;
  - Sudarytus algoritmus ištirti statistinio modeliavimo būdu, įrodyti jų konvergavimą ir palyginti su esamais algoritmais;
  - Sudarytus algoritmus pritaikyti praktiniams uždaviniams spręsti.
- **Planuojami rezultatai:**
  - Sudaryti korektiški ir konkurencingi rekursiniai algoritmai, skirti stochastinių dinaminių sistemų, stebimų su triukšmu, filtravimui, identifikavimui ir valdymui realiu laiku;
  - Sudaryti algoritmai pritaikyti socialinių, verslo ir/ar technikos procesų/sistemų modeliavimui bei simuliavimui.

Visų studijų planas ir jo vykdymo suvestinė

1 lentelė: Visų studijų planas.

Studijų metai	**Egzaminai**		**Dalyvavimas konferencijose**		**Publikacijos**		
	Planas	**Įvykdyta**	**Planas**	**Įvykdyta**	**Planas**	**Įvykdyta**	**Būklė**
I (2018/2019)	1	1			1		
II (2019/2020)	1	1	1	1		1	
III (2020/2021)	**2**	**2**		**2**	**1**		**Įteikta**
IV (2022/2023)			1				

Ataskaitinių metų darbo planas ir jo įvykdymas

2 lentelė: Einamieji studijų metai (III: 2020/2021).

Egzaminai		Dalyvavimas konferencijose		Publikacijos	
Planas	**Įvykdyta**	**Planas**	**Įvykdyta**	**Planas**	**Įvykdyta**
Skaitinis intelektas investuojant į vertybinius popierius	Išlaikyta. Egzamino laikymo data: *2021 m. sausio 15 d.*; įvertinimas: *10 (puikiai).*		31st *European Conference on Operational Research*, 11-14 July 2021, Athens, Greece (https://euro2021.euro-online.org/). Pranešimas: *„Incremental Maximum Likelihood Estimation of Noisy Gaussian Random Walk"*.	Straipsnis *„Efficient Maximum Likelihood Batch Estimation With Pure Time Series Data of a One-Dimensional Cumulative Structural Equation Model"* į žurnalą *„R Journal"* (https://journal.r-project.org/)	Pakartotinai įteikta po recenzavimo
Statistinis modeliavimas ir stochastinis optimizavimas	Išlaikyta. Egzamino laikymo data: *2021 m. sausio 19 d.*; įvertinimas: *10 (puikiai).*		2nd International & European Conference *"Modelling and Simulation of Social-Behavioural Phenomena in Creative Societies"*, September 21-23, 2022, Vilnius, Lithuania (https://msbc.tech). Pranešimas: *„Efficient Maximum Likelihood Batch Estimation With Pure Time Series Data of a One-Dimensional Cumulative Structural Equation Model"*.		

Visų mokslinių tyrimų ir disertacijos rengimo etapai (1)

3 lentelė: Mokslinių tyrimų ir disertacijos rengimo etapai.

	Darbo pavadinimas	Atlikimo terminai	Pastabos
1	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje): 1. Atlikti stochastinių dinaminių sistemų analitinę apžvalgą; 2. Apžvelgti su stochastinėmis dinaminėmis sistemomis susijusių uždavinių sprendimo metodikas.	2018 m. spalio mėn. – 2019 m. rugsėjo mėn.	Apžvelgta dinaminių sistemų samprata, klasifikacija, aktualūs uždaviniai. Detaliau panagrinėtas šių uždavinių sprendimas tiesinės dinaminės sistemos modelio atveju. Pažvelgta į tiesinių dinaminių sistemų taikymą struktūrinių lygčių modeliavimo kontekste. Išsiaiškinta motyvacija šios srities tyrimams.
2	Mokslinio tyrimo vykdymas: 2.1. Tyrimo metodikos sudarymas: 1. Tyrimo metodikos iškeltiems uždaviniams spręsti parinkimas; 2. Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką. 2.2. Teorinis tyrimas: 1. Tiesinių stochastinių dinaminių sistemų, stebimų su triukšmu, rekursinių filtravimo, identifikavimo ir valdymo algoritmų pasirinkta metodika sudarymas; 2. Sudarytų algoritmų adaptavimas netiesinėms stochastinėms dinaminėms sistemoms. 2.3. Empirinis tyrimas: 1. Sudarytų algoritmų pritaikymas praktinių uždavinių sprendimui. 2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas: 1. Gautų rezultatų analizė; 2. Rezultatų apibendrinimas, esminių rezultatų išskyrimas; 3. Išvadų parengimas.	2019 m. spalio mėn. 2019 m. lapkričio mėn. – 2020 m. rugsėjo mėn. 2020 m. spalio mėn. – 2021 m. gegužės mėn. 2021 m. birželio mėn. – 2021 m. rugsėjo mėn.	Atsitiktinio Gauso klaidžiojimo, stebimo su triukšmu, modeliui sukonstruotas palaipsnis didžiausio tikėtinumo parametrų vertinimo algoritmas, kuriame įverčiai yra gaunami per išreikštinę taisyklę. Algoritmo veikimas yra išnagrinėtas eksperimentiškai, užsiimama jo teoriniu nagrinėjimu. Remiantis įgyta patirtimi, sukonstruotas tiesinio sudėtingumo didžiausio tikėtinumo parametrų vertinimo algoritmas, skirtas kumuliaciniam struktūrinių lygčių modeliui grynųjų laiko eilučių atveju. Taip pat sukonstruotas kultūros poveikio socialiniam kapitalui tikimybinis modelis bei kompiuterinės simuliacijos algoritmas. Tuo būdu kuriami algoritmai su stochastinėmis dinaminės sistemomis

Visų mokslinių tyrimų ir disertacijos rengimo etapai (2)

			susijusių uždavinių sprendimui bus pritaikyti socialinių sistemų modeliavimui ir simuliavimui.
3.	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas: 3.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas; 3.2. Analitinės disertacijos dalies parengimas; 3.3. Teorinės disertacijos dalies parengimas; 3.4. Eksperimentinės disertacijos dalies parengimas; 3.5. Bendrųjų išvadų formulavimas.	2022 m. spalio mėn. – 2023 m. balandžio mėn.	
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2023 m. gegužės mėn.	
5	Daktaro disertacijos gynimas	2023 m. rugsėjo mėn.	

Trumpas per pusmetį gautų mokslinių rezultatų pristatymas

- Dinaminių struktūrinių lygčių modeliai;
- Efektyvūs didžiausio tikėtinumo parametrų vertinimo algoritmai šiems modeliams;
- Rinktinės skaidrės iš konferencijos pranešimo šia tema medžiagos (anglų k.).

Dynamic Structural Equation Model (DSEM)

3.2 Dynamic panel structural equation model

In this section we consider a dynamic panel simultaneous equation model with latent variables and fixed effects (DPSEM(p,q)). A DPSEM(p,q) model for the individual $i = 1,\ldots,N$ at time $t = 1,\ldots,T$ can be written for the generic individual at any time period t using the " t-notation" as

$$\begin{aligned}
\boldsymbol{\eta}_{it} &= \sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{it-j} + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{it-j} + \boldsymbol{\zeta}_{it} && (3.1)\\
\boldsymbol{y}_{it} &= \boldsymbol{\Lambda}_y \boldsymbol{\eta}_{it} + \boldsymbol{\mu}_{yi} + \boldsymbol{\varepsilon}_{it} && (3.2)\\
\boldsymbol{x}_{it} &= \boldsymbol{\Lambda}_x \boldsymbol{\xi}_{it} + \boldsymbol{\mu}_{xi} + \boldsymbol{\delta}_{it} && (3.3)
\end{aligned}$$

where $\boldsymbol{\eta}_{it} = \left(\eta_{it}^{(1)}, \eta_{it}^{(2)}, \ldots, \eta_{it}^{(m)}\right)'$ and $\boldsymbol{\xi}_{it} = \left(\xi_{it}^{(1)}, \xi_{it}^{(2)}, \ldots, \xi_{it}^{(g)}\right)'$ are vectors of latent variables, $\boldsymbol{y}_{it} = \left((y_{it}^{(1)}, y_{it}^{(2)}, \ldots, y_{it}^{(n)}\right)'$ and $\boldsymbol{x}_{it} = \left(x_{it}^{(1)}, x_{it}^{(2)}, \ldots, x_{it}^{(k)}\right)'$ are vectors of observable variables, and $\boldsymbol{B}_j$ $(m \times m)$, $\boldsymbol{\Gamma}_j$ $(m \times g)$, $\boldsymbol{\Lambda}_x$ $(k \times g)$, and $\boldsymbol{\Lambda}_y$ $(n \times m)$ are coefficient matrices. The contemporaneous and simultaneous coefficients are in $\boldsymbol{B}_0$, and $\boldsymbol{\Gamma}_0$, while $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, $\ldots$, $\boldsymbol{B}_p$, and $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, $\ldots$, $\boldsymbol{\Gamma}_q$ contain coefficients of the lagged endogenous and exogenous latent variables. Finally, $\boldsymbol{\mu}_{yi}$ and $\boldsymbol{\mu}_{xi}$ are the $n \times 1$ and $k \times 1$ vectors of individual means, respectively. We treat $\boldsymbol{\mu}_{yi}$ and $\boldsymbol{\mu}_{xi}$ as vectors of coincidental (fixed) parameters, which makes the DPSEM model (3.1)-(3.3) a " fixed-effects" panel model. The statistical assumptions about the variables in (3.1)–(3.3) are as follows.

Cumulative Structural Equation Model (CSEM) Under Consideration (1)

Output measurement equation:

$$y_t = \eta_t + \epsilon_t, \tag{0.1}$$

where $\{y_t\}$, $t \in \mathbb{Z}^+$ is a sequence of scalar observed outcomes, $\{\eta_t\}$, $t \in \mathbb{Z}^+$ is a sequence of scalar latent spaces, and $\{\epsilon_t\}$, $t \in \mathbb{Z}^+$ is a sequence of independent and identically distributed $\mathcal{N}\left(0, \sigma_y^2\right)$ scalar observed process noises.

Transition equation:

$$\eta_{t+1} = \eta_t + \mu_\eta + L_\eta \xi_{t+1} + \zeta_{t+1}, \qquad \eta_0 = 0, \tag{0.2}$$

where μ_η is a scalar intercept term, L_η is an $1 \times k$ vector of latent input weights, $\{\xi_{t+1}\}$, $t \in \mathbb{Z}^+$ is a sequence of independent and identically distributed $k \times 1$ vectors of latent input to the latent process (common factors) where $\xi_{t+1} \sim \mathcal{N}\left(0_{k\times 1}, I_{k\times k}\right)$, and $\{\zeta_{t+1}\}$, $t \in \mathbb{Z}^+$ is a sequence of independent and identically distributed $\mathcal{N}\left(0, \sigma_\eta^2\right)$ scalar latent process errors.

Input measurement equation:

$$x_{t+1} = \mu_x + L_x \xi_{t+1} + \delta_{t+1}, \tag{0.3}$$

where $\{x_{t+1}\}$, $t \in \mathbb{Z}^+$ is a sequence of $m \times 1$ vectors of observed inputs, μ_x is an $m \times 1$ vector of intercept terms, L_x is an $m \times k$ matrix of factor loadings, and $\{\delta_{t+1}\}$, $t \in \mathbb{Z}^+$ is a sequence of independent and identically distributed $m \times 1$ vectors of specific factors where $\delta_{t+1} \sim \mathcal{N}\left(0_{m\times 1}, \operatorname{diag}\left(\sigma_{x_1}^2, .., \sigma_{x_m}^2\right)\right)$.

It is assumed that $\{\xi_{t+1}\}$, $\{\delta_{t+1}\}$, $\{\zeta_{t+1}\}$, and $\{\epsilon_t\}$, $t \in \mathbb{Z}^+$ are mutually independent and $t \leq T$ $(T \in \mathbb{N})$.

Algorithm Development (2)

We attach Δy_t to the end of x_t, $t = 1, ..., T$, and denote the resulting vector as z_t, additionally denoting $z = (z_1, ..., z_T)'$. Then

$$\mathsf{E} z_t = \mu$$

and

$$\mathsf{C}_{zz}(\tau) = \begin{cases} LL' + \Psi_0, & \tau = 0 \\ -\Psi_1, & \tau = 1 \\ 0_{(m+1)\times(m+1)}, & otherwise \end{cases},$$

where

$$\Psi_0 = \begin{bmatrix} \sigma^2_{x_1} & 0 & \dots & 0 & 0 \\ 0 & \sigma^2_{x_2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{x_m} & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_\eta + 2\sigma^2_y \end{bmatrix}, \quad \Psi_1 = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_y \end{bmatrix}.$$

Algorithm Development (5)

By taking the logarithm of (**??**), we obtain the logarithmic likelihood function:

$$\mathcal{L}(\theta) = \mathcal{L}(\theta; z_{vec}) = -\frac{1}{2}\left(\ln(|\mathsf{C}_{zz}|) + d'_{vec}\mathsf{C}_{zz}^{-1}d_{vec} + (m+1)T\ln(2\pi)\right),$$

where $\theta = (\mu_{vec}, \mathsf{C}_{zz})$ are parameters of the multivariate normal distribution.

Since it is true that

$$\arg\max_{\theta}\left[\mathcal{L}(\theta)\right] = \arg\max_{\theta}\left[\frac{\mathcal{L}(\theta)}{T}\right] = \arg\min_{\theta}\left[\frac{\ln(|\mathsf{C}_{zz}|) + d'_{vec}\mathsf{C}_{zz}^{-1}d_{vec}}{T}\right]$$

can be used for both normalisation and simplicity purposes, we further consider the minimisation of function $\mathcal{L}(\theta)$ defined as

$$\mathcal{L}(\theta) = \frac{\ln(|\mathsf{C}_{zz}|) + d'_{vec}\mathsf{C}_{zz}^{-1}d_{vec}}{T}. \tag{0.4}$$

Algorithm Development (10)

With the aforementioned choice of θ and given expressions (**??**) and (**??**), function (0.4) is ultimately reorganised as follows:

$$\begin{aligned}\mathcal{L}(s, \Sigma_x, \Sigma_s, \sigma_y^2, \mu_x, \mu_\eta) = \frac{1}{T}\Bigg(& \ln\left(\left(|\Sigma_x|\sigma_y^2\right)^T |Q|\right) \\ & + \operatorname{tr}\left(d_x \Sigma_x^{-1} d_x'\right) \\ & + \frac{1}{\sigma_y^2}\left(\left(d_y - d_x\Sigma_s\right)' Q^{-1} \left(d_y - d_x\Sigma_s\right)\right)\Bigg),\end{aligned} \tag{0.5}$$

where $d_y = z_y - 1_{T\times 1}\mu_\eta$, $Q = \left(\frac{s^2+1}{s}\right) I + J$, where $z_y = (\Delta y_1, ..., \Delta y_T)'$.

Algorithm Development (13)

By manipulating the derivatives and the estimators obtained from them (**??**)–(**??**), we can achieve a gradual way to obtain θ^*:

$$\mu_x^* = \hat{\mu}_x, \tag{0.6}$$

$$\Sigma_x^* = \hat{\Sigma}_x, \tag{0.7}$$

$$s^* = \arg\min_{s \in [0,1]} \tilde{\mathcal{L}}(s), \tag{0.8}$$

$$\Sigma_s^* = \hat{\Sigma}_s\left(s^*\right), \tag{0.9}$$

$$\mu_\eta^* = \hat{\mu}_\eta\left(s^*\right), \tag{0.10}$$

$$\sigma_y^{2*} = \hat{\sigma}_y^2\left(s^*\right), \tag{0.11}$$

Algorithm Development (16)

Let us denote $c_T = \frac{1}{T}\frac{\left([u]_{1:T,1}\right)'\left([Q]_{1:T,1:T}\right)^{-1}[v]_{1:T,1}}{q_T}$; here $q_T = \left[\left([Q]_{1:T,1:T}\right)^{-1}\right]_{T,T}$. Thus, instead of calculating terms of the form $u'Q^{-1}v$ in (**??**)–(**??**), we calculate terms of the form c_T since $\frac{1}{T}\frac{1}{q_T}$ simply cancels out in $\hat{\Sigma}_s$ and $\hat{\mu}_\eta$, while, for $\hat{\sigma}_y^2$ and $\tilde{\mathcal{L}}$, the use of c_T introduces robust expressions, as we now calculate them as follows:

$$\begin{aligned}\hat{\sigma}_y^2 &= q_T\, g \\ &= s\,\frac{f_T}{f_{T+1}}\,g\end{aligned}$$

and

$$\tilde{\mathcal{L}} = h + \ln(g),$$

where

$$\begin{aligned}h &= \frac{\ln(|Q|)}{T} + \ln q_T \\ &= \ln\left(\frac{f_T}{\left(f_{T+1}\right)^{\left(1-\frac{1}{T}\right)}}\right)\end{aligned}$$

and

$$g = \frac{\hat{\sigma}_y^2}{q_T},$$

where g is c_T with $u = v = \hat{d}_y - \hat{d}_x\hat{\Sigma}_s$ and $f_T = \sum_{i=0}^{T-1} s^{2i}$ is to be calculated recursively:

$$f_T = f_{T-1} + s^{2(T-1)}.$$

Algorithm Development (17)

Recursively, c_T is calculated as follows:

$$c_T = \left(1 - \frac{1}{T}\right) \frac{f_{T-1} f_{T+1}}{f_T^2} c_{T-1} + \frac{a_T b_T}{T},$$

where $a_T = \frac{\left([u]_{1:T,1}\right)' \left[\left([Q]_{1:T,1:T}\right)^{-1}\right]_{1:T,T}}{q_T}$ and $b_T = \frac{\left([v]_{1:T,1}\right)' \left[\left([Q]_{1:T,1:T}\right)^{-1}\right]_{1:T,T}}{q_T}$ are also to be calculated recursively:

$$a_T = s\frac{f_{T-1}}{f_T} a_{T-1} + [u]_{T,1},$$

$$b_T = s\frac{f_{T-1}}{f_T} b_{T-1} + [v]_{T,1}.$$

The starting values for all the recursions, namely a_1, b_1, c_1, and f_1, are

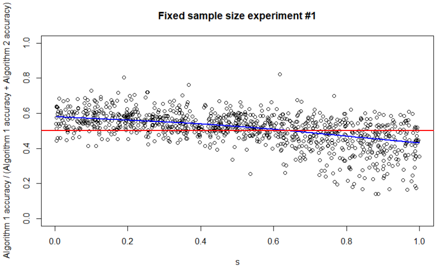
$$a_1 = [u]_{1,1} ,$$

$$b_1 = [v]_{1,1} ,$$

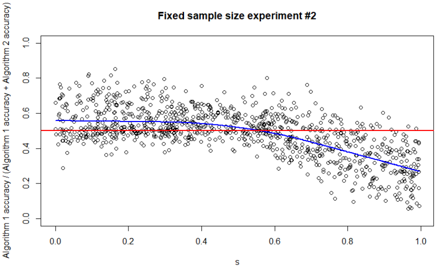
$$c_1 = [u]_{1,1} [v]_{1,1} ,$$

$$f_1 = 1.$$

Experimental Results (9)

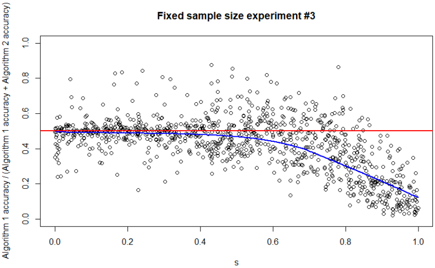


A fixed sample size experiment with $T = 10$.

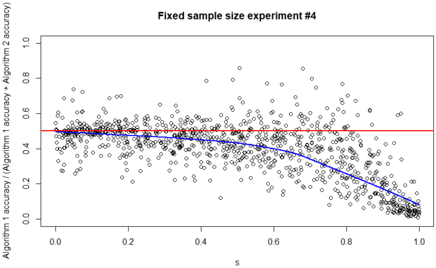


A fixed sample size experiment with $T = 20$.

Experimental Results (10)

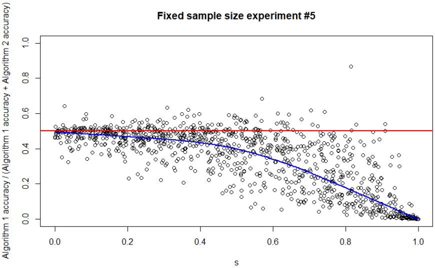


A fixed sample size experiment with $T = 50$.

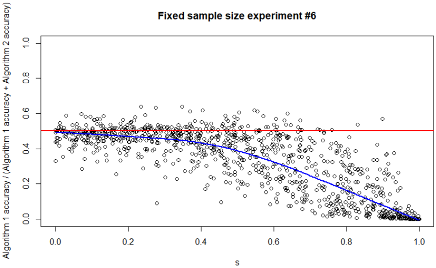


A fixed sample size experiment with $T = 100$.

Experimental Results (11)



A fixed sample size experiment with $T = 1000$.



A fixed sample size experiment with $T = 10000$.

Implementation in R

- EMLI – Efficient Maximum Likelihood Inference for Linear Dynamical Models (https://CRAN.R-project.org/package=EMLI);
- Three functions available that allow generating the considered CSEM data, running the developed algorithm, and evaluating the estimation accuracy;
- More models and features to be added.

Kito pusmečio darbo planas

2022/2023 m. m. darbo planas:

- Parengti ir publikuoti straipsnį, skirtą sukurtų algoritmų pritaikymui socialinių procesų modeliavimo ir simuliavimo kontekste;
- Sudalyvauti *The 23rd Conference of the International Federation of Operational Research Societies, July 10 to 14, 2023, Santiago, Chile*;
- Parengti disertaciją.