



BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
GÉPÉSZMÉRNÖKI KAR
MECHATRONIKA, OPTIKA ÉS GÉPÉSZETI INFORMATIKA TANSZÉK



HUN-REN SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓINTÉZET

Appendix / Függelék

FIZIKAI MODELLEK HATÉKONY ADATALAPÚ KIEGÉSZÍTÉSE

Györök Bendegúz Máté
MSc IV. félév

Konzulensek:

Dr. Tóth Roland

Tudományos főmunkatárs / Full professor

HUN-REN SZTAKI

Eindhoven University of Technology

Témavezető:

Dr. Budai Csaba

Adjunktus

BME-GPK MOGI

Dr. Péni Tamás

Tudományos főmunkatárs

HUN-REN SZTAKI

BUDAPEST, 2024. november 9.

A Generalization capabilities of the multiplicative augmentation structure

The multiplicative model augmentation structure can be interpreted as a generalization of the additive case according to Theorem 1.

Theorem 1. *There exists a neural network parametrization $\chi \in \mathbb{R}^{n_\phi}$ for the multiplicative model augmentation structure (2.10) that realizes exactly the additive structure (2.7).*

Proof. Let us assume we have a neural network parametrization $\chi \in \mathbb{R}^{n_\phi}$ that realizes the following operation:

$$\mathcal{F}_\chi(x_b, x_a, u, \xi) = \begin{bmatrix} \xi \\ 0_{n_a} \end{bmatrix} + \tilde{\mathcal{F}}_\chi(x_b, x_a, u),$$

where $\tilde{\mathcal{F}}_\chi$ represents an arbitrary (possibly) nonlinear function. As $\xi \equiv f_\theta(x_b, u)$ by definition, this augmentation form is equivalent to the additive structure in (2.7a). $\square$

B Analyzing the LFR-based model augmentation

B.1 Generalization capabilities of the approach

The LFR-based approach can be interpreted as a hybrid additive structure, however, according to Theorem 2, the LFR-based structure is a direct generalization of the additive approach.

Theorem 2. *There exists a parametrization for the LFR-based model augmentation that realizes exactly the additive structure (2.7).*

Proof. Let us divide $\mathcal{F}_\phi$ into two parts, such that

$$\begin{bmatrix} x_a^+ \\ w \end{bmatrix} = \mathcal{F}_\phi(x_a, z) = \begin{bmatrix} \mathcal{F}_\phi^a(x_a, z) \\ \mathcal{F}_\phi^w(x_a, z) \end{bmatrix}.$$

Then by substitution, the LFR-based augmentation form can be expressed as

$$\begin{bmatrix} x_b^+ \\ x_a^+ \end{bmatrix} = \begin{bmatrix} f_\theta(x_b, u) + B_w \mathcal{F}_\phi^w(x_a, C_z x_b + D_{zu} u + D_{zw} w) \\ \mathcal{F}_\phi^a(x_a, C_z x_b + D_{zu} u + D_{zw} w) \end{bmatrix},$$

$$\hat{y} = h_\theta(x_b, u) + D_{yw} \mathcal{F}_\phi^w(x_a, C_z x_b + D_{zu} u + D_{zw} w).$$

Now let us set $B_w = I_{n_x}$, $D_{zw} = 0$, and $D_{yw} = 0$, furthermore select $\dim z = n_x + n_u$, $\dim w = n_x$. Furthermore if $C_z = \begin{bmatrix} I_{n_x} & 0_{n_u \times n_x} \end{bmatrix}^\top$ and $D_{zu} = \begin{bmatrix} 0_{n_u \times n_x} & I_{n_u} \end{bmatrix}^\top$, then,

$$z[k] = C_z x_b[k] + D_{zu} u[k] + D_{zw} w[k] = \begin{bmatrix} x_b[k] \\ u[k] \end{bmatrix} \quad \forall k \in \mathbb{N}.$$

Now, substituting all specifically set parameter values into the structure, we arrive to

$$\begin{bmatrix} x_b^+ \\ x_a^+ \end{bmatrix} = \begin{bmatrix} f_\theta(x_b, u) \\ 0_{n_a} \end{bmatrix} + \underbrace{\begin{bmatrix} \mathcal{F}_\phi^w(x_a, [x_b^\top u^\top]^\top) \\ \mathcal{F}_\phi^a(x_a, [x_b^\top u^\top]^\top) \end{bmatrix}}_{\mathcal{F}_\phi(x_a, [x_b^\top u^\top]^\top)},$$

$$\hat{y} = h_\theta(x_b, u),$$

which form is equivalent to the additive model augmentation framework in (2.7). $\square$

B.2 Extension of the LFR-based model augmentation

The LFR-based approach can be converted into a serially connected structure by introducing $\tau_x \in [0, 1]$ and $\tau_y \in [0, 1]$ tunable parameters into Equations (2.15) and (2.17), as

$$x_b^+ = \tau_x f_\theta(x_b, u) + B_w w, \quad (\text{B.1})$$

$$\hat{y} = \tau_y h_\theta(x_b, u) + D_{yw} w, \quad (\text{B.2})$$

and not modifying Equations (2.14) and (2.16). This is a promising direction for unifying all model augmentation structures, however, further investigation and theoretical analysis are necessary in the future.

C LTI state reconstruction with noise model

If a noise model is known for the system at hand, incorporating it into the LTI state-space model in (4.5a) and (4.5b), the following innovation form of the resulting dynamics can be given:

$$x[k+1] = Ax[k] + Bu[k] + Ke[k], \quad (\text{C.1})$$

$$y[k] = Cx[k] + Du[k] + e[k], \quad (\text{C.2})$$

where $e[k] \in \mathbb{R}^{n_y}$ is an i.i.d. zero-mean white noise process with finite variance Σ_e and $K \in \mathbb{R}^{n_x \times n_y}$ is an adjustable “parameter” which determines how much the process noise affects the state transition.

From (C.2), the noise realisation at time step k can be expressed as

$$e[k] = y[k] - Cx[k] - Du[k]. \quad (\text{C.3})$$

Then, the state transition can be expressed by substituting (C.3) into (C.1) as

$$x[k+1] = \underbrace{(A - KC)}_{\tilde{A}} x[k] + \underbrace{(B - KD)}_{\tilde{B}} u[k] + Ky[k]. \quad (\text{C.4})$$

Similarly to (4.8), $y[k+1]$, $y[k+2]$, $\dots$, $y[k+n]$ can be derived recursively, and the y_k^{k+n} can be expressed as:

$$y_k^{k+n} = \underbrace{\begin{bmatrix} C\tilde{A}^n \\ C\tilde{A}^{n-1} \\ \vdots \\ C\tilde{A} \\ C \end{bmatrix}}_{\tilde{O}_n} x[k] + \underbrace{\begin{bmatrix} D & C\tilde{B} & C\tilde{A}\tilde{B} & \dots & C\tilde{A}^{n-1}\tilde{B} \\ 0 & D & C\tilde{B} & \dots & C\tilde{A}^{n-2}\tilde{B} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & C\tilde{B} \\ 0 & 0 & 0 & 0 & D \end{bmatrix}}_{\tilde{\Gamma}_n} u_k^{k+n} + \underbrace{\begin{bmatrix} 0 & CK & C\tilde{A}K & \dots & C\tilde{A}^{n-1}K \\ 0 & 0 & CK & \dots & C\tilde{A}^{n-2}K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & CK \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\tilde{\Lambda}_n} y_k^{k+n} + e_k^{k+n}. \quad (\text{C.5})$$

In (C.5), it seems to be a contradiction that y_k^{k+n} appears both on the left and right side of the equation, but it can be seen, that for calculating y_{k+n} , output values are needed only up until y_{k+n-1} .

As previously, the time invariance of the system can be utilized as

$$y_{k-n}^k = \tilde{O}_n x[k-n] + \tilde{\Gamma}_n u_{k-n}^k + \tilde{\Lambda}_n y_{k-n}^k + e_{k-n}^k, \quad (\text{C.6})$$

and then from (C.6), the state vector can be expressed as

$$x[k-n] = \tilde{O}_n^{-1} \left[(I - \tilde{\Lambda}_n) y_{k-n}^k - \tilde{\Gamma}_n u_{k-n}^k - e_{k-n}^k \right], \quad (\text{C.7})$$

where, similarly to the noiseless scenario, $\tilde{O}_n$ always has a left inverse if the $(\tilde{A}, \tilde{B}, C, D)$ LTI state-space representation is observable and n is chosen sufficiently large.

Similarly to the noiseless scenario, $x[k+n]$ can be expressed in the same way, with the addition of the noise structure as

$$x[k+n] = \tilde{A}^n x[k] + \underbrace{\begin{bmatrix} 0 & \tilde{B} & \tilde{A}\tilde{B} & \dots & \tilde{A}^{n-1}\tilde{B} \end{bmatrix}}_{\tilde{\gamma}_n} u_k^{k+n} + \underbrace{\begin{bmatrix} 0 & K & \tilde{A}K & \dots & \tilde{A}^{n-1}K \end{bmatrix}}_{\tilde{\lambda}_n} y_k^{k+n}, \quad (\text{C.8})$$

then shifting back time:

$$x[k] = \tilde{A}^n x[k-n] + \tilde{\gamma}_n u_{k-n}^k + \tilde{\lambda}_n y_{k-n}^k. \quad (\text{C.9})$$

Now we can take the same substitution as in the noiseless case to get

$$x[k] = \tilde{A}^n \tilde{O}_n^{-1} \left[(I - \tilde{\Lambda}_n) y_{k-n}^k - \tilde{\Gamma}_n u_{k-n}^k - e_{k-n}^k \right] + \tilde{\gamma}_n u_{k-n}^k + \tilde{\lambda}_n y_{k-n}^k. \quad (\text{C.10})$$

However, compared to the noiseless case, we can not directly use (C.10) as the sample path of the noise $\{e[\tau]\}_{\tau=k-n}^k$ is unknown. However, we can take the conditional expectation of $x[k]$ w.r.t. the known samples of u_{k-n}^k and y_{k-n}^k , which, due to the assumption that $\mathbb{E}\{e[k]\} = 0$, gives

$$\mathbb{E}\{x[k] \mid y_{k-n}^k, u_{k-n}^k\} = \tilde{A}^n \tilde{O}_n^{-1} \left[(I - \tilde{\Lambda}_n) y_{k-n}^k - \tilde{\Gamma}_n u_{k-n}^k \right] + \tilde{\gamma}_n u_{k-n}^k + \tilde{\lambda}_n y_{k-n}^k, \quad (\text{C.11})$$

which is similar to the noiseless case (4.14) except the corrector terms $\tilde{\gamma}_n$ and $\tilde{\lambda}_n$ at the end of the equation. Furthermore, if the generating noise has a Gaussian distribution, i.e., $e[k] \sim \mathcal{N}(0, \Sigma_e)$, then (C.11) is a Maximum Likelihood estimator of $x[k]$.

Then finally, the parametrization of $P_{b,u}$, and $P_{b,y}$ can be given as

$$P_{b,u} = -\tilde{A}^n \tilde{O}_n^{-1} \tilde{\Gamma} + \tilde{\gamma}_n, \quad (\text{C.12})$$

$$P_{b,y} = \tilde{A}^n \tilde{O}_n^{-1} (I - \tilde{\Lambda}_n) + \tilde{\lambda}_n. \quad (\text{C.13})$$

D ACM Classification

concept_id	concept_desc	concept_significance
10010147.10010341.10010346.10010347	Computing methodologies~Systems theory	500
10010147.10010341.10010342.10010343	Computing methodologies~Modeling methodologies	500
10010147.10010178.10010213.10010214	Computing methodologies~Computational control theory	300