Proceedings of the
Estonian Academy of Sciences
2025, **74**, 4, 473–486

https://doi.org/10.3176/proc.2025.4.01

www.eap.ee/proceedings
Estonian Academy Publishers

CONTROL THEORY

RESEARCH ARTICLE

Received 29 October 2024
Accepted 27 December 2024
Available online 7 October 2025

Keywords:
nonlinear control system, discrete-time system, generalized observer form, parametrized state transformation, comparison, algebraic approach, differential geometric approach

Corresponding author:
Maris Tõnso
maris.tonso@taltech.ee

Citation:
Kotta, Ü., Mullari, T. and Tõnso, M. 2025. Transforming state equations into the generalized observer form: comparison of algebraic and differential geometric approaches. *Proceedings of the Estonian Academy of Sciences*, **74**(4), 473–486. https://doi.org/10.3176/proc.2025.4.01

Transforming state equations into the generalized observer form: comparison of algebraic and differential geometric approaches

Ülle Kotta[a], Tanel Mullari[b] and Maris Tõnso[a]

[a] Department of Software Science, Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia

[b] Department of Cybernetics, Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia

ABSTRACT

Two different approaches exist that allow transforming nonlinear discrete-time state equations into the generalized observer form. One of them is based on the algebraic approach of vector fields and the other on the standard differential geometric approach. This paper presents a comprehensive comparison of these two solutions, covering assumptions, two sets of solvability conditions, parametrized state transformation algorithms and domains of validity of the results.

1. Introduction

The main goal of this paper is to compare the results of the papers [7] and [9], which address practically the same nonlinear problem but with different mathematical tools. Such comparison papers are quite rare nowadays, though important for the comprehensive development of the field. The comparison highlights the advantages and disadvantages of both techniques and thus facilitates choosing the best approach for future extensions. The problem studied in these papers is the transformation of discrete-time nonlinear state equations into the generalized observer form. This form expands the possibilities of the classical observer form to construct the (joint) state (and disturbance) observers for nonlinear control systems since the solvability conditions are significantly less restrictive than in the case of the classical observer form and are, in the special case of the generalized form, always achievable. In spite of such generality, this form allows to construct the state observers with the same desirable properties as in the classical case. Both papers [7] and [9] give a complete solution to the problem: they provide the necessary and sufficient solvability conditions as well as the algorithms constructing the necessary parametrized state transformation. We compare the assumptions made, the two sets of solvability conditions, the state transformation algorithms, the domains of validity of the results and other minor aspects. To a certain extent, our comparison goes beyond merely specifying the differences and similarities of the two methods suggested in [7] and [9], respectively. In doing so, it also comments on the many aspects that are equally important in examining other control problems, frequently studied in parallel by these different mathematical tools. Moreover, we simplify one of the solvability conditions from the paper [7] and present one of the solvability conditions from the paper [9] in a non-confusing manner.

The mathematical approach in [7] is generic algebraic, whereas that in [9] is local differential geometric. The authors are not aware of papers that compare these two mathematical approaches, although some specific aspects have been explored in a few papers (see for instance [5]).

Note that although the results of [7] hold almost everywhere, there may exist the so-called singular (alternatively called non-regular) points where the dimensions and ranks drop or the coordinate transformation is not defined. In such points the results of [7] cannot be applied. The results of [9] are valid only locally around the equilibrium points (x^{eq}, u^{eq}) of the system. Therefore, it makes sense to make the formal comparison only around the regular equilibrium points. We additionally demonstrate by a simple illustrative example what happens at non-regular equilibrium points.

2. Preliminaries

Both papers [7] and [9] address the discrete-time nonlinear control system of the form

$$\begin{aligned} x^{\langle 1\rangle}(t) &= \overline{\Phi}(x(t), u(t)) \\ y(t) &= h(x(t)), \end{aligned} \tag{1}$$

where $x^{\langle 1\rangle}(t) := x(t+1), t \in \mathbf{Z}$. In [7], the state $x(t) \in \overline{X} \subset \mathbb{R}^n$, the control $u(t) \in U \subset \mathbb{R}$ and the output $y(t) \in Y \subseteq \mathbb{R}$, with $\overline{X}, U$ and Y being open subspaces. The paper [9] assumes that $x(t) \in M$, where M is n-dimensional C^∞ manifold, $u(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$. Throughout the paper, the notation $\zeta^{\langle \ell\rangle} := \zeta(t+\ell), \ell \in \mathbb{Z}$, is used.

Definition 1. *A point* (x^{eq}, u^{eq}) *is called the equilibrium point of the system* (1) *if it satisfies the relation* $x^{eq} = \overline{\Phi}(x^{eq}, u^{eq})$.

Both papers assume the system (1) to be submersive , which is a slightly more general assumption than state-reversibility.

Definition 2. *The system* (1) *is called submersive if*

$$\operatorname{rank} \frac{\partial \overline{\Phi}}{\partial (x,u)} = n \tag{2}$$

and reversible if

$$\operatorname{rank} \frac{\partial \overline{\Phi}}{\partial x} = n. \tag{3}$$

In [7], the rank is computed over the inversive difference field $\mathcal{K}$ (see below) and may drop at some points, while in [9], the rank is computed over $\mathbb{R}$ around the equilibrium point (x^{eq}, u^{eq}).

The paper [7] defines the spaces of differential one-forms $\mathcal{Y} := \operatorname{span}_{\mathcal{K}}\{dy^{\langle l\rangle}, l \geqslant 0\}$, $\mathcal{U} := \operatorname{span}_{\mathcal{K}}\{du^{\langle j\rangle}, j \geqslant 0\}$, $\mathcal{X} := \operatorname{span}_{\mathcal{K}}\{dx\}$. The subspace $\mathcal{O} := \mathcal{X} \cap (\mathcal{Y} + \mathcal{U})$ is called the observable space of the system (1).

Definition 3. *The system* (1) *is said to satisfy the generic observability condition if* $\dim_{\mathcal{K}} \mathcal{O} = n$.

Mathematical framework in [7]. The algebraic approach from [6], based on the vector fields, is used to derive the results. An inversive difference field $\mathcal{K}$ is associated with the system (1), and further computations are made over $\mathcal{K}$. Since $\mathcal{K}$ is supposed to be inversive, it is necessary to supplement the state equations (1) with the relation

$$z(t) = \chi(x(t), u(t)), \tag{4}$$

where the additional new independent variable at time instant t, i.e. $z(t) \in \mathbb{R}$, is chosen such that the map $\Phi := \left[\overline{\Phi}^{\mathrm{T}}, \chi\right]^{\mathrm{T}} : \overline{X} \times U \to \overline{X} \times \mathbb{R}$ has an analytic inverse that allows computing the backward shifts of the variables x and u [6]. In case the analytic inverse is not valid globally but just generically, the choice of the variable z may, in principle, be given by different functions in different regions. The inversive field $\mathcal{K}$ allows one to define the backward shift of the vector field, as shown in [6].

In this paper only the projection of the backward shift of the vector field is explicitly used, therefore we limit the theoretical exposition by stating that for the vector field

$$\Gamma = \sum_{i=1}^{n} a_i \frac{\partial}{\partial x_i} + \sum_{k\geqslant 0} b_k \frac{\partial}{\partial u^{\langle k\rangle}} + \sum_{l\geqslant 1} c_l \frac{\partial}{\partial z^{\langle -l\rangle}}$$

the projection of its backward shift is computed as

$$\Gamma^{\langle -1\rangle \pi} = \sum_{i=1}^{n} \left\langle \mathrm{d}\bar{\Phi}_i, \Gamma \right\rangle^{\langle -1\rangle} \frac{\partial}{\partial x_i}. \tag{5}$$

Mathematical framework in [9]. The paper [9] is based on the standard differential geometric approach for discrete-time nonlinear systems (see for instance [1–4,8]). In particular, the conditions are formulated in terms of the Ad operator, which has been defined only for a reversible map. Since the paper assumes the equations (1) to be submersive but not necessarily reversible with respect to the variable x, the standard differential geometric tools are not directly applicable. In order to employ the standard approach, an extended system, associated with the original system (1), is studied instead of (1). The extended system is given by the equations

$$\xi^{\langle 1\rangle} = F(\xi, v), \quad \tilde{y} = h_e(\xi) = h(\xi_1, \ldots, \xi_n), \tag{6}$$

where the extended state $\xi = (x, u, \ldots, u^{\langle r-1\rangle}) \in \bar{M}$, where $\bar{M}$ is an $(n+r)$-dimensional manifold, for some fixed integer r, and

$$F(\xi, v) = \begin{bmatrix} \bar{\Phi}(x,u) \\ u^{\langle 1\rangle} \\ \vdots \\ u^{\langle r-1\rangle} \\ h(x) + v \end{bmatrix}. \tag{7}$$

Remark 4. If the state equations (1) are reversible, then the last component of F is taken to be equal to $u + v$.

Observe that the extended system (6) is different from the classical input prolongation. The goal of such a non-standard choice (adding the output function $h(x)$ to v) is to guarantee reversibility of the equations (6) with respect to the extended state ξ, so that the Ad operator can be defined for the map $F_v(\xi) := F(\xi, v)$, for the constant v as follows[1].

Definition 5. [2,9] *Given a vector field X on $\bar{M}$ and the diffeomorphism $F_v(\xi)$, define the operator* $\mathrm{Ad}_{F_v} X := \left(\frac{\partial F_v}{\partial \xi} X \right)\Big|_{F_v^{-1}(\xi)}$ *and recursively* $\mathrm{Ad}^0_{F_v} X := X$, $\mathrm{Ad}^{k+1}_{F_v} X := \mathrm{Ad}_{F_v}\left(\mathrm{Ad}^k_{F_v} X\right)$, $k \geqslant 0$.

Comparison of the frameworks in [7] and [9]. In the special case when the vector field X is defined in the finite state space $\mathrm{span}_{\mathcal{K}}\{\partial/\partial x\}$, the backward shift of the vector field is the analogue of the Ad operator [6]. In the general case when the vector field is defined in the infinite space as in [6], the Ad operator is the analogue of the projection of the backward shift into the state space.

The extended system (6) corresponds to the choice $z = h(x)$ in [7]. Taking $z = h(x)$ is legitimate if the original system is non-reversible, and [9] focuses on such a case[2]. However, restricting the choice of z does not affect the generality of the results. The results from [7] are formulated in terms of the original system (1), though a new variable z is introduced to compute the backward shifts, rather than in terms of the extended system that also relies on the output feedback. This also means that [7] handles reversible and non-reversible cases simultaneously, not separately as [9]. Of course, the choice of the new variable z depends on whether the system is reversible or not, though this does not show up in theory but only in computations. Observe that the construction of the extended system (6) depends on *a priori* unknown integer r, while the advantage of [7] is that it does not require such a choice.

[1] Note that by Lemma 3 in [9], an additional assumption, namely the transformability of the equations (1) into the generalized observer form, is needed to guarantee reversibility. However, since in [7] it is proven that for an observable system the latter is always possible by taking $s = n - 1$ in (8) below, this assumption may be replaced by the observability assumption.

[2] The extended system in the reversible case corresponds to the choice $z = u$ in [7].

2.1. Results under comparison

The goal of both [7] and [9] is to transform the state equations (1), via a parametrized state transformation $X = \Psi(x, u, \ldots, u^{\langle r-1\rangle})$, into the so-called generalized observer form with the degrees (s, r):

$$\begin{aligned} X_i^{\langle 1\rangle} &= X_{i+1}, \quad i = 1, \ldots, s, \\ X_i^{\langle 1\rangle} &= X_{i+1} + \varphi_i(y, \ldots, y^{\langle s\rangle}, u, \ldots, u^{\langle r\rangle}), \\ &\qquad i = s+1, \ldots, n-1, \\ X_n^{\langle 1\rangle} &= \varphi_n(y, \ldots, y^{\langle s\rangle}, u, \ldots, u^{\langle r\rangle}), \\ y &= X_1. \end{aligned} \tag{8}$$

In [9], slightly different notations are used: the equations (8) are represented in the matrix form, the new state variables are denoted by z_i, instead of the non-negative integer r the letter l is used and instead of the functions $\varphi_{s+1}, \ldots, \varphi_n$, there are functions $\gamma_i(z_1, \ldots, z_{s+1}, u, \ldots, u^{\langle r\rangle}), i = s+1 \ldots, n$. The latter is immaterial since $z_i = y^{\langle i-1\rangle}, i = 1, \ldots, s+1$.

Both papers also formulate the results in terms of a single uniquely defined vector field, denoted by Ξ in [7] and by g in [9], both computed, in principle, from the state equations (1), though in [9], the extended system (6) is used. In order to find Ξ, [7] computes first the one-forms

$$\omega_k := \sum_{i=1}^{n} \frac{\partial y^{\langle k\rangle}}{\partial x_i} \mathrm{d}x_i, \quad k = 0, \ldots, n-1, \tag{9}$$

and then defines $\Xi \in \mathrm{span}_{\mathcal{K}}\{\partial/\partial x\}$ as the solution of the set of equations

$$\langle \omega_k, \Xi\rangle \equiv \delta_{k,n-1}, \quad k = 0, \ldots, n-1, \tag{10}$$

where $\delta_{k,n-1}$ is the Kronecker delta.

In order to construct g, the paper [9] starts from $F_0(\xi) := F(\xi, 0)$, defines recursively $F_0^0(\xi) := \xi$, $F_0^k(\xi) := F_0(F_0^{k-1}(\xi))$ and computes the one-forms

$$\mathrm{d}[h(F_0^k)], \quad k = 0, \ldots, n-1, \quad \mathrm{d}\xi_{n+1}, \ldots, \mathrm{d}\xi_{n+r}, \tag{11}$$

dropping the argument of $F_0(\xi)$ for brevity. The vector field g is found as the solution of the set of the equations

$$\langle \mathrm{d}\xi_{n+i},\ g\rangle = 0, \qquad i = 1, \ldots, r, \tag{12a}$$

$$\langle \mathrm{d}[h(F_0^{k-1})],\ g\rangle = \delta_{k,n}, \qquad k = 1, \ldots, n. \tag{12b}$$

Due to the structure of the extended system (6), the forward shifts of its output $\tilde{y}$ can never depend on the forward shifts of u of the order higher than $u^{\langle r-1\rangle}$.

Theorem 6. [7] *Under the generic observability assumption, the state equations* (1) *can be transformed, via a parametrized state transformation $X = \Psi(x, u, \ldots, u^{\langle r-1\rangle})$, into the generalized observer form* (8) *with the degrees (s, r) iff the following conditions are satisfied:*

(i)
$$\left[\frac{\partial}{\partial z^{\langle -q\rangle}}, \Xi^{\langle -l\rangle\pi}\right] \equiv 0, \quad q, l = 1, \ldots, n-s-1,$$

(ii)
$$\left[\Xi^{\langle -l\rangle\pi}, \Xi^{\langle -j\rangle\pi}\right] \equiv 0, \quad l, j = 0, \ldots, n-s-1,$$

(iii)
$$\frac{\partial y^{\langle s\rangle}}{\partial u^{\langle k\rangle}} \equiv 0, \quad k = r, \ldots, s-1,$$

(iv)
$$\left[\frac{\partial}{\partial u^{\langle k\rangle}}, \Xi\right] \equiv 0, \quad k = r, \ldots, n-2.$$

Theorem 7. [9] *Under the submersivity assumption, the state equations* (1) *can be locally, around the equilibrium point, transformed by a parametrized state transformation* $X = \overline{\Psi}(x, u, \ldots, u^{\langle r-1\rangle})$ *into the generalized observer form* (8) *iff the following conditions are satisfied:*

(A) *The one-forms* (11) *are linearly independent.*

(B) *The vector field g satisfies*

$$\left[\mathrm{Ad}^{p}_{F_0} g, \mathrm{Ad}^{q}_{F_0} g\right] = 0, \quad p, q = 0, \ldots, n - s - 1. \tag{13}$$

(C) *The output h satisfies*

$$\frac{\partial h^{\langle\kappa\rangle}}{\partial v} = 0, \quad \kappa = 0, \ldots, s. \tag{14}$$

(D) *The vector fields* $\eta_i := \mathrm{Ad}^{i-1}_{F_0} g, i = 1, \ldots, n - s - 1$ *satisfy*

$$\frac{\partial(\mathrm{Ad}_{F_v}\eta_i)}{\partial v} = 0. \tag{15}$$

Remark 8. The condition (i) in Theorem 6 may be replaced by a simpler condition (i*): $\left[\partial/\partial z^{\langle-1\rangle}, \Xi^{\langle-l\rangle\pi}\right] \equiv 0$, $l = 1, \ldots, n - s - 1$. The condition (i) for $q > 1$ follows from (i*).

The condition (D) may be replaced by a simpler and, since v is supposed to be a constant, a more natural condition (D*): $\left[\partial/\partial u^{\langle r-1\rangle}, \mathrm{Ad}^{i}_{F_0} g\right] \equiv 0$ (see the explanation in the proof of Proposition 11, part (a)).

Recall that the extended system (6) corresponds to the specific choice $z = h(x)$ in (4). Proposition 9 below holds under this choice. Of course, $z = h(x)$ is not the only possibility, but different functions $\chi(x, u)$ in (4) would correspond to different extended systems, not addressed in [9]. Therefore, such an assumption is legitimate.

Proposition 9. *Equality* $\Xi^{\langle-\ell\rangle\pi} = \mathrm{Ad}^{\ell}_{F_0} g$ *holds for* $\ell \geqslant 0$.

Proof. The proof is by induction. We first prove in two steps that the proposition holds for $\ell = 0$, i.e. that $\Xi = g$. In the first step we show that Ξ, the solution of (10), satisfies also the equations (12a). Given that $\xi = (x, u, \ldots, u^{\langle r-1\rangle})$, the equations (12a) can be rewritten as

$$\langle \mathrm{d}u^{\langle i\rangle},\ g\rangle = 0,\ i = 0, \ldots, r - 1. \tag{16}$$

In the proof of Theorem 4.1 in [7], it was shown that $u, \ldots, u^{\langle r-1\rangle}$ are the invariants of $\Xi^{\langle-l\rangle\pi}$ for $l = 0, \ldots, n-s-1$, meaning that $\langle \mathrm{d}u^{\langle i\rangle},\ \Xi\rangle = 0$ for $i = 0, \ldots, r-1$. This means that also Ξ satisfies (16), or alternatively (12a).

In the second step, we show that the equations (10) coincide with the equations (12b). Unlike in the system (1), in the extended system (6) (when $v = 0$) the variables $u^{\langle k\rangle}$, $k \geqslant r$, are no longer the independent variables of the difference field $\mathcal{K}$ because the equality $u^{\langle r\rangle} = h(x)$ holds. To find the relations between $\mathrm{d}[h(F_0^k)]$ and ω_k for $k = 0, ..., n - 1$, we use this fact, sometimes repeatedly, depending on the values of r and n. Direct but tedious computation results in

$$\mathrm{d}[h(F_0^k)] = \omega_k + \sum_{i=0}^{k-r} \alpha_{k,i}\omega_i + \sum_{j=0}^{r-1} \beta_{k,j}\mathrm{d}u^{\langle j\rangle}, \tag{17}$$

where the coefficients $\alpha_{k,i}, \beta_{k,j}$ are the functions of $x, u, \ldots, u^{\langle r-1\rangle}$ (some of them being possibly zero), and the coefficients of the one-forms ω_j, $j = 0, \ldots, k$, are computed at the point, given by the restriction $u^{\langle r\rangle} = h(x)$, defined by the extended system. Substituting $\mathrm{d}[h(F_0^k)]$ from (17) into (12b) and using the properties of the scalar product gives

$$\langle\omega_k, g\rangle + \alpha_{k,i}\left\langle\sum_{i=0}^{k-r}\omega_i, g\right\rangle + \beta_{k,j}\left\langle\sum_{j=0}^{r-1}\mathrm{d}u^{\langle j\rangle}, g\right\rangle = \delta_{k+1,n}, \quad k = 0, \ldots, n - 1. \tag{18}$$

The third term in the above sum is zero due to (16). By induction, one can easily show that the second term is also zero. For $k = 0$ (18) takes the form $\langle \omega_0,\ g\rangle = 0$ if $n > 1$. Assuming the second term is zero for the first k equations, i.e. $\langle \omega_0,\ g\rangle = \cdots = \langle \omega_k,\ g\rangle = 0$, concludes that it is also zero in the equation $k + 1$. Thus (10) coincides with (12b).

Assume now that the proposition is valid for ℓ and show that it also holds for $\ell + 1$. Using the definition of the Ad operator and the facts that $\mathrm{Ad}^{\ell}_{F_0} g \in \mathrm{span}\,\{\partial/\partial x_i, i = 1, \ldots, n\}$, $h(x)$ is the invariant of the vector fields $\Xi^{\langle -l\rangle\pi}$, allows one to write

$$\begin{aligned} \mathrm{Ad}^{\ell+1}_{F_0} g = \mathrm{Ad}_{F_0}\left(\mathrm{Ad}^{\ell}_{F_0} g\right) = \sum_{i=1}^{n} \left\langle \mathrm{d}\bar{\Phi}_i, \mathrm{Ad}^{\ell}_{F_0} g\right\rangle\Big|_{F_0^{-1}} \frac{\partial}{\partial x_i} \\ + \left\langle \mathrm{d}h, \mathrm{Ad}^{\ell}_{F_0} g\right\rangle\Big|_{F_0^{-1}} \frac{\partial}{\partial u^{\langle r-1\rangle}} = \sum_{i=1}^{n} \left\langle \mathrm{d}\bar{\Phi}_i, \Xi^{\langle -\ell\rangle\pi}\right\rangle\Big|_{F_0^{-1}} \frac{\partial}{\partial x_i}. \end{aligned} \tag{19}$$

Computing the values of the rightmost scalar products in (19) at F_0^{-1} by using the relation $u^{\langle r-1\rangle} = z^{\langle -1\rangle}$ means shifting the scalar products backwards. Taking this into account and comparing the result with the formula (5) completes the proof for $\ell + 1$. □

3. Comparison of the results

3.1. Assumptions

The paper [7] assumes the system (1) to be generically observable in the sense of Definition 3. In [9], observability is not explicitly assumed; however, the following holds.

Proposition 10. *The condition* (A) *in* Theorem 7 *is equivalent to the local observability assumption of the system* (1) *around an equilibrium point.*

Proof. Assume that the system (1) is generically observable, which, by [7], is equivalent to the fact that the one-forms $\omega_0, \ldots, \omega_{n-1}$ are linearly independent. Note that independence is necessary for the vector field Ξ to be uniquely defined. Relations (17) express the one-forms $\mathrm{d}[h(F_0^k)]$, $k = 0, \ldots, n-1$, in terms of $\omega_0, \ldots, \omega_{n-1}, \mathrm{d}u, \ldots, \mathrm{d}u^{\langle r-1\rangle}$. The transformation matrix with the elements $\alpha_{k,j}$ is a regular matrix since the elements above its main diagonal are all zeros, and the elements on the main diagonal are equal to one. Thus $\mathrm{d}[h(F_0^k)]$, $k = 0, \ldots, n-1$, are also linearly independent. The one-forms $\mathrm{d}\xi_{n+1} = \mathrm{d}u, \ldots, \mathrm{d}\xi_{n+r} = \mathrm{d}u^{\langle r-1\rangle}$ are linearly independent from $\mathrm{d}[h(F_0^k)]$, $k = 0, \ldots, n-1$, since $\omega_0, \ldots, \omega_{n-1}$ do not involve $\mathrm{d}u, \ldots, \mathrm{d}u^{\langle r-1\rangle}$. The proof also holds in reverse. Since in [7] the claim holds generically, it also holds locally around the regular equilibrium point. □

3.2. Comparison of solvability conditions

The main theorem in [7] has four necessary and sufficient solvability conditions, whereas that in [9] has only three (if we set aside the condition (A) on observability that is given as an assumption in [7]).

Proposition 11.

(a) *The conditions* (i) *and* (D) *are equivalent.*

(b) *The conditions* (ii) *and* (B) *are equivalent.*

(c) *The conditions* (iii) *and* (C) *are equivalent.*

Proof. (a) Observe first that in the condition (i) it is enough to require that $q = 1$; the validity of (i) for the remaining values of q follows from this. The reasoning is as below. The vector field $\Xi^{\langle -1\rangle\pi}$ can, by the construction of Ξ and the formula (5), depend only on $z^{\langle -1\rangle}$. If it depends on $z^{\langle -1\rangle}$, then (i) is already not satisfied for $l = 1$. Otherwise, $\Xi^{\langle -2\rangle\pi}$ cannot depend on $z^{\langle -q\rangle}$, $q > 1$. Continuing in the same manner, the observation follows. This fact means that by (i) one only has to check whether the coefficients of the vector fields $\Xi^{\langle -l\rangle\pi}$, $l = 1, \ldots, n-s-1$, are independent of $z^{\langle -1\rangle}$.

Note that by Proposition 9, the vector fields η_i in (D) are given as $\eta_i = \Xi^{\langle -i+1\rangle\pi}$, $i = 1, \ldots, n-s-1$. Compute next, as required in (D), $\mathrm{Ad}_{F_v}\eta_i$, $i = 1, \ldots, n-s-1$, taking into account the above equality and the fact that $\Xi^{\langle -i+1\rangle\pi}$, for $i = 1, \dots, n-s-1$, has components only in the directions $\partial/\partial x_j$:

$$\mathrm{Ad}_{F_v}\eta_i = \sum_{j=1}^{n} \left\langle \mathrm{d}\bar{\Phi}_j, \Xi^{\langle -i+1\rangle\pi} \right\rangle\Big|_{F_v^{-1}} \frac{\partial}{\partial x_j} + \left\langle \mathrm{d}h, \Xi^{\langle -i+1\rangle\pi} \right\rangle\Big|_{F_v^{-1}} \frac{\partial}{\partial u^{\langle r-1\rangle}}, \tag{20}$$

where $i = 1, \dots, n-s-1$. Since $y = h(x)$ is the invariant of $\Xi^{\langle -i+1\rangle\pi}, i = 1, \dots, n-s-1$, $\left\langle \mathrm{d}h, \Xi^{\langle -i+1\rangle\pi} \right\rangle \equiv 0$, and so the last term in (20) vanishes.

Shifting the last equation of the extended system (6) back results in $u^{\langle r-1\rangle} - v = [h(x)]^{\langle -1\rangle} = z^{\langle -1\rangle}$.

The condition (D) guarantees that the scalar products $\left\langle \mathrm{d}\bar{\Phi}_j, \Xi^{\langle -i+1\rangle\pi} \right\rangle, i = 1, \dots, n-s-1$, $j = 1, \dots, n$, computed at the point $F_v^{-1}(\xi)$, do not depend on v. This is so if and only if they do not depend on $\xi_{n+r} - v = u^{\langle r-1\rangle} - v$, or equivalently, $\left[\Xi^{\langle -i+1\rangle\pi}\right]^{\langle -1\rangle\pi} = \Xi^{\langle -i\rangle\pi}$ is independent of $z^{\langle -1\rangle}$. This completes the proof of part (a).

(b) By Proposition 9, it is obvious that the condition (ii) of Theorem 6 and the condition (B) of Theorem 7 coincide.

(c) The condition (iii) in Theorem 6 says that the first $s+1$ coordinates $y^{\langle j\rangle}$, $j = 0, \dots, s$, do not depend on the forward shifts of u of the order higher than $r-1$. Note that in [9], the output shifts of $\tilde{y}$ are defined by the extended system (7), where $u^{\langle r\rangle}$ is replaced by $h(x) + v$. The corresponding requirement is that the forward shifts of $\tilde{y}$ up to the order s may not depend on v, which is guaranteed if (C) holds. There is actually no need to check the validity of (C) for all $k = 0, \dots, s$ because when $y^{\langle k\rangle}$ for some $k < s$ already depends on v, then all higher order forward shifts of $\tilde{y}$ also do. Therefore it is enough to check the condition (C) only for $k = s$. □

Observe that Theorem 6 contains an additional condition (iv), which does not have an analogue in Theorem 7. However, the condition (iv) is not redundant. First, it is necessary for the vector fields $\Xi^{\langle -l\rangle\pi}$, $l = 0, ..., n-s-1$, to be well-defined in the $(n+r)$-dimensional subspace with the coordinates $(x, u, \dots, u^{\langle r-1\rangle})$ that guarantee the existence of the parametrized state transformation (see Lemma 2.3 from [7]). Note that well-definedness of the vector fields $\mathrm{Ad}_{F_0}^{q}$, $q = 0, \dots, n-s-1$, in [9] is guaranteed by the structure of the extended system.

The more detailed explanation of why in Theorem 6 we need the condition (iv) but not its analogue in Theorem 7 is the following. In [7], one operates in the infinite dimensional space with the coordinates x, $u^{\langle k\rangle}$, $k \geqslant 0$, $z^{\langle -q\rangle}$, $q \geqslant 1$, where the vector fields $\Xi^{\langle -l\rangle\pi}$ are computed. Note that the components of the new state X are computed as $n-s$ canonical parameters and s common invariants of $\Xi^{\langle -l\rangle\pi}, l = 0, \dots, n-s-1$. Therefore, if the coefficients of Ξ (and, consequently, also the coefficients of $\Xi^{\langle -l\rangle\pi}$, see Lemma 2.3 in [7]) do not depend on $u^{\langle k\rangle}$, $k \geqslant r$, then also the new state X cannot depend on these variables. Note also that in [9], the extended map F_0, used for defining the vector field g, defines the equation $u^{\langle r\rangle} = h(x)$. This ensures that the forward shift of u of the higher order than $r-1$ cannot appear in the equations. Therefore in [9], one operates in the $(n+r)$-dimensional space with the coordinates $x, u, \dots, u^{\langle r-1\rangle}$, and, consequently, the coefficients of g can also not depend on $u^{\langle k\rangle}$, $k \geqslant r$.

Second, the condition (iv) jointly with (iii) allows one to look for the minimal possible r value for which Theorem 6 is satisfied. The paper [9] does not explicitly address finding the lowest possible r value for which the problem is solvable. To use the results of [9] for such a purpose, one has to find it by trial and error, starting either from $r = 0$ or from $r = n-2$. Searching for the minimal r here is more complicated than finding it from the conditions (iii) and (iv).

3.3. Construction of the parametrized state transformation and computational complexity

The idea of constructing the parametrized state transformation is, in principle, similar in both papers. To find the coordinate transformation, the commuting vector fields, the number of which is less than n (in case the equations cannot be transformed into the classical observer form), need to be completed to constitute a commuting family. In the case of the classical observer form, the state coordinates can be computed from the system of differential equations $\langle \mathrm{d}X_i, \Xi^{\langle n-l\rangle\pi}\rangle \equiv \delta_{il}, i, l = 1, \dots, n$, where all vector fields commute [5]. If the n vector fields $\Xi^{\langle n-l\rangle\pi}$ do not all commute but are linearly independent, one can still define the one-forms θ_i such that $\langle \theta_i, \Xi^{\langle n-l\rangle\pi}\rangle \equiv \delta_{il}$, though not all of these one-forms are the total differentials of certain functions. The first few new coordinates can be redefined as $h(x)$ and its forward shifts. Completing this set into a commuting family is sometimes difficult, and there are many solutions (see the proof of Lemma 2.3 in [7]), depending on the choice of the integration constants.

Both in [7] and [9], the first s coordinates are defined as

$$X_i = \Psi_i(x, u, ..., u^{\langle r-1\rangle}) = y^{\langle i-1\rangle}, \quad i = 1, ..., s, \tag{21}$$

whereby these coordinates are the independent invariants of $\Xi^{\langle -l\rangle\pi}$ (or $\mathrm{Ad}^l_{F_0}g$), $l = 0, ..., n-s-1$. Note that in [7], the value of s is determined from the conditions (i) and (ii) of Theorem 6; in [9], it is determined from the conditions (D) and (B) of Theorem 7, respectively.

In [7], the remaining $(n-s)$ coordinates are found as the canonical parameters of the commuting vector fields $\Xi^{\langle -l\rangle\pi}$, $l = 0, ..., n-s-1$, using the method shown below. Construct the $[n \times (n-s)]$-matrix M, whose columns can be interpreted as the vector fields $\Xi^{\langle -l\rangle\pi}$. The rows of its left inverse M_L^{-1}, being the $[(n-s) \times n]$-matrix, can be interpreted as the one-forms θ_i, $i = s+1, ..., n$. Due to the construction, M and therefore also M_L^{-1} are of full rank; therefore the one-forms θ_i are linearly independent over the field $\mathcal{K}$. The independence allows one to find (non-uniquely) the new coordinates as functionally independent functions $X_i = \Psi(x, u, ..., u^{\langle r-1\rangle})$, whose total differentials are the linear combinations of θ_i and the total differentials of $u^{\langle k\rangle}$ and Ψ_q, $q = 1, ..., s$, which are the invariants of the vector fields $\Xi^{\langle -l\rangle\pi}$, $l = 0, ..., n-s-1$:

$$\mathrm{d}\Psi_i = \theta_i + \sum_{q=1}^{s}\beta_{iq}\mathrm{d}\Psi_q + \sum_{k=0}^{r-1}\bar{\beta}_{ik}\mathrm{d}u^{\langle k\rangle}, \quad i = s+1, ..., n. \tag{22}$$

Note that there is a restriction in defining the $(s+1)$th new coordinate. In order to satisfy the sth equation of (15), it has to be given by $\Psi_{s+1} = y^{\langle s\rangle}$. In (22), the value of r is found from the conditions (iii) and (iv) of Theorem 6. The coefficients β_{iq} and $\bar{\beta}_{ik}$, which are the functions of the variables $x, u, ..., u^{\langle r-1\rangle}$, are computed based on the Frobenius theorem. This yields a system of underdetermined (if $s > 0$) partial differential equations to find these coefficients.

In [9], the coordinates X_i, $i = s+1, ..., n$, are found in two steps. In the first step, one defines the intermediate coordinates $\varphi_i = y^{\langle i-1\rangle}$, $i = 1, ..., s$, whereas the remaining coordinates φ_i, $i = s+1, ..., n$, are defined as arbitrary independent functions of the variables $x, u, ..., u^{\langle r-1\rangle}$ to get a full set of coordinates. Then the vector fields $\mathrm{Ad}^l_{F_0}g$ are computed in the new coordinates. Because they do not have the components in the directions of $\partial/\partial\varphi_i$, $i = 1, ..., s$, one can construct a regular $[(n-s) \times (n-s)]$ matrix R, whose columns can be interpreted as the subvectors of $\mathrm{Ad}^l_{F_0}g$, $l = 0, ..., n-s-1$. Next, the inverse R^{-1} of this matrix is computed. The rows of R^{-1} are interpreted as the one-forms $\tilde{\theta}_i \in \mathrm{span}_{\mathcal{K}}\{\mathrm{d}\varphi_{s+1}, ..., \mathrm{d}\varphi_n\}$, where $u, \ldots, u^{\langle r-1\rangle}$ and $\varphi_1, \ldots, \varphi_s$ are treated as constants. In the second step, one defines the final new coordinates as $X_i = \varphi_i$, $i = 1, ..., s$, and $X_i = \int \tilde{\theta}_i + C_i(\varphi_1, ..., \varphi_s, u, \ldots, u^{\langle r-1\rangle})$, $i = s+1, ..., n$, where C_i are the integration constants[3].

As the example in subsection 4.2 demonstrates, an arbitrary integration constant C_{s+1} does not necessarily yield the generalized observer form given in the problem statement. Searching for the suitable integration constant requires additional effort. This drawback can be overcome by requiring the determination of C_{s+1} to guarantee that $X_{s+1} = y^{\langle s\rangle}$.

Note that although the method in [7] that finds the new coordinates in a single step is conceptually more transparent, the method from [9] is *computationally significantly simpler* than the one given in [7] for the following reasons. First, the computation of R^{-1} is easier than the computation of M_L^{-1}, the complexity of which increases as s increases, i.e. the fewer columns M has. Second, the method in [9], unlike the one in [7], does not require the application of the Frobenius theorem because the elements of R^{-1} are already the partial derivatives of the new coordinates with respect to the old ones, and therefore the new coordinates can be found by simple integration.

Depending on the state equations (1), it is not always possible to explicitly express the backward shifts $x^{\langle -1\rangle}$, $u^{\langle -1\rangle}$ in terms of elementary functions. Similarly, in such a case it is not possible to find the inverse operator F_0^{-1}, and thus the application of the Ad_{F_0} operator fails.

Finally, note that in [9], the new $(s+1)$th coordinate is not necessarily defined as $\tilde{y}^{\langle s\rangle}$. However, the choice in [7] for the $(s+1)$th coordinate can be recovered by suitably choosing the integration

[3] In [9], the variables $u, \ldots, u^{\langle r-1\rangle}$ were erroneously missing in C_i.

constant when integrating the one-form $\tilde{\theta}_{s+1}$. As demonstrated by the example in subsection 4.2, step 4), choosing $C_3 = 0$ leads to the system that is not in the generalized observer form (8), while $C_3 = u$ yields the proper form.

3.4. Domain of validity of the results

Note that the paper [7] operates in the $(n + r)$-dimensional space C with the coordinates $\bar{x} := (x, u, ..., u^{\langle r-1\rangle})$, whereas [9] operates in the $(n + r)$-dimensional C^∞ manifold $\bar{M}$.

The results of [9] are valid locally in a neighbourhood of an equilibrium point $(x^{eq}, u^{eq}, \ldots, u^{eq}) \in \bar{M}$ of the extended system, whereas the results of [7] are valid generically in the space C. The latter means that the transformation is valid almost everywhere in C. This is mathematically expressed by the words 'on an open and dense subspace of C'. Moreover, the assumptions made hold also on some open and dense subspace as well as the solvability conditions. Since one looks at the dimensions (or ranks) over the field of functions and not over $\mathbb{R}$, it is meaningless to speak about the constant dimensionality of the vector spaces of vector fields or one-forms. Moreover, the generic rank is the maximal rank (that is valid on an open and dense subspace), but it may drop at some points outside this subspace.

The higher generality of the generic approach has a price – one has to require that the state equations are analytic compared to the smoothness, i.e C^∞, assumption in the paper [9]. This drawback is not serious since many smooth control systems are described via analytic functions. The other drawback is related to possible singular points that can, in principle, be found easily from the results of a given system. However, the theory in [7] does not say anything about such points – whether the problem is solvable at such a point or not.

4. Example

Example 1. Consider the system

$$\begin{aligned} x_1^{\langle 1\rangle} &= x_3\\ x_2^{\langle 1\rangle} &= x_4\\ x_3^{\langle 1\rangle} &= u\,(1+x_2)\\ x_4^{\langle 1\rangle} &= x_2\\ y &= x_1, \end{aligned} \tag{23}$$

which is generically observable and submersive but not reversible.

4.1. Algebraic approach from [7]

Find first the backward shifts of the system variables under the simplest option $z = h(x) = x_1$:

$$u^{\langle -1\rangle} = \frac{x_3}{1+x_4}, \quad x_1^{\langle 1\rangle} = z^{\langle -1\rangle}, \quad x_2^{\langle -1\rangle} = x_4, \quad x_3^{\langle -1\rangle} = x_1, \quad x_4^{\langle -1\rangle} = x_2.$$

Forward shifts of the output are $y = x_1$, $y^{\langle 1\rangle} = x_3$, $y^{\langle 2\rangle} = u(1 + x_2)$, $y^{\langle 3\rangle} = u^{\langle 1\rangle}(1 + x_4)$ and the linearly independent one-forms

$$\omega_0 = \mathrm{d}x_1, \quad \omega_1 = \mathrm{d}x_3, \quad \omega_2 = u\mathrm{d}x_2, \quad \omega_3 = u^{\langle 1\rangle}\mathrm{d}x_4.$$

Next, find the vector field Ξ from the equations (10) and compute the projections of its backward shifts up to the order $n - 1 = 3$:

$$\begin{aligned} \Xi &= \frac{1}{u^{\langle 1\rangle}}\frac{\partial}{\partial x_4}, & \Xi^{\langle -2\rangle\pi} &= \frac{\partial}{\partial x_3} + \frac{1+x_4}{x_3}\frac{\partial}{\partial x_4},\\ \Xi^{\langle -1\rangle\pi} &= \frac{1}{u}\frac{\partial}{\partial x_2}, & \Xi^{\langle -3\rangle\pi} &= \frac{\partial}{\partial x_1} + \frac{1+x_2}{x_1}\frac{\partial}{\partial x_2}. \end{aligned} \tag{24}$$

The integer s is determined by the conditions (i) and (ii) of Theorem 6. The condition (i) states that the coefficients of $\Xi^{\langle -l\rangle\pi}, l = 1, \ldots, n - s - 1$, cannot involve the backward shifts of z. This condition is satisfied for any $l = 1, \ldots, n - s - 1$, meaning that s can be any of the integers $0, 1, 2, 3$, but one prefers it to be as small as possible. Checking the condition (ii) gives $[\Xi, \Xi^{\langle -1\rangle\pi}] = 0$, but $[\Xi, \Xi^{\langle -2\rangle\pi}] = \left[1/(u^{\langle 1\rangle}x_3)\right]\partial/\partial x_4 \neq 0$; thus $s = 2$.

The conditions (iii) and (iv) determine the integer r. The condition (iii) points to the possibility of $r = 1$ since $y^{\langle 3\rangle}$ depends on $u^{\langle 1\rangle}$, while $y^{\langle 2\rangle}$ does not. However, since Ξ depends on $u^{\langle 1\rangle}$, one concludes from (iv) that $r = 1$ is not large enough, and so $r = 2$.

By [7], construct the matrix M and compute its left inverse:

$$M = \begin{bmatrix} 0 & 0 \\ \frac{1}{u} & 0 \\ 0 & 0 \\ 0 & \frac{1}{u^{\langle 1\rangle}} \end{bmatrix}, \quad M_L^{-1} = \begin{bmatrix} 0 & u & 0 & 0 \\ 0 & 0 & 0 & u^{\langle 1\rangle} \end{bmatrix}.$$

The rows of M_L^{-1} specify the one-forms $\theta_3 = u\mathrm{d}x_2$, $\theta_4 = u^{\langle 1\rangle}\mathrm{d}x_4$. The new state variables are $Z_1 = y = x_1,\ Z_2 = y^{\langle 1\rangle} = x_3,\ Z_3 = y^{\langle 2\rangle} = u(1 + x_2)$ and

$$\mathrm{d}Z_4 = \theta_4 + \beta_{40}\mathrm{d}y + \beta_{41}\mathrm{d}y^{\langle 1\rangle} + \tilde{\beta}_{40}\mathrm{d}u + \tilde{\beta}_{41}\mathrm{d}u^{\langle 1\rangle},$$

where $\beta_{40}, \beta_{41}, \tilde{\beta}_{40}, \tilde{\beta}_{41}$ are the functions to be selected so that the right hand side becomes a total differential. It is admissible to add the linear combinations of $\mathrm{d}y, \mathrm{d}y^{\langle 1\rangle}, \mathrm{d}u, \mathrm{d}u^{\langle 1\rangle}$ to θ_4 because $y, y^{\langle 1\rangle}, u, u^{\langle 1\rangle}$ are common invariants of Ξ, $\Xi^{\langle -1\rangle\pi}, \ldots, \Xi^{\langle -n+s+1)\rangle\pi}$; thus the scalar products of $\mathrm{d}y, \mathrm{d}y^{\langle 1\rangle}$, $\mathrm{d}u$ and $\mathrm{d}u^{\langle 1\rangle}$ with Ξ, $\Xi^{\langle -1\rangle\pi}, \ldots, \Xi^{\langle -n+s+1)\rangle\pi}$ equal zero.

With the product's differentiation rule in mind, we notice that $\theta_4 = u^{\langle 1\rangle}\mathrm{d}x_4$ is lacking the term $x_4\mathrm{d}u^{\langle 1\rangle}$ to form a total differential of the function $u^{\langle 1\rangle}x_4$; thus we can choose $\beta_{40} = \beta_{41} = \tilde{\beta}_{40} = 0$ and $\tilde{\beta}_{41} = x_4$ yielding $\mathrm{d}Z_4 = u^{\langle 1\rangle}\mathrm{d}x_4 + x_4\mathrm{d}u^{\langle 1\rangle} = \mathrm{d}(u^{\langle 1\rangle}x_4)$. The generalized observer form is

$$Z_1^{\langle 1\rangle} = Z_2, \quad Z_2^{\langle 1\rangle} = Z_3, \quad Z_3^{\langle 1\rangle} = u^{\langle 1\rangle} + Z_4, \quad Z_4^{\langle 1\rangle} = u^{\langle 2\rangle}\left(\frac{y^{\langle 2\rangle}}{u} - 1\right), \quad y = Z_1. \tag{25}$$

4.2. Differential geometric solution from [9]

The conditions of Theorem 7 do not give direct instructions for finding the integer r but rather lead up to the trial and error method. Therefore we first consider the case $r = 1$ and examine if the conditions (A)–(D) are satisfied. The extended state equations are $\xi^{\langle 1\rangle} = F(\xi, v)$, which for $r = 1$ have the form

$$\xi_1^{\langle 1\rangle} = \xi_3, \quad \xi_2^{\langle 1\rangle} = \xi_4, \quad \xi_3^{\langle 1\rangle} = (\xi_2 + 1)\,\xi_5, \quad \xi_4^{\langle 1\rangle} = \xi_2, \quad \xi_5^{\langle 1\rangle} = \xi_1 + v, \quad y = \xi_1, \tag{26}$$

where

$$\xi_1 = x_1, \quad \xi_2 = x_2, \quad \xi_3 = x_3, \quad \xi_4 = x_4, \quad \xi_5 = u. \tag{27}$$

Setting $v = const$ allows one to compute the inverse of F_v:

$$F_v^{-1}(\xi) = \begin{bmatrix} \xi_5 - v & \xi_4 & \xi_1 & \xi_2 & \dfrac{\xi_3}{1 + \xi_4} \end{bmatrix}^{\mathrm{T}}, \tag{28}$$

leading to

$$F_0^{-1}(\xi) = \begin{bmatrix} \xi_5 & \xi_4 & \xi_1 & \xi_2 & \dfrac{\xi_3}{1 + \xi_4} \end{bmatrix}^{\mathrm{T}}.$$

We also need to find

$$\begin{aligned} \tilde{y} = h = \xi_1, \quad \tilde{y}^{\langle 1\rangle} = h(F_0) = \xi_3, \quad \tilde{y}^{\langle 2\rangle} = h(F_0^2) = (1 + \xi_2)\xi_5, \\ \tilde{y}^{\langle 3\rangle} = (\xi_1 + v)(1 + \xi_4) \not\equiv h(F_0^3) = \xi_1(1 + \xi_4). \end{aligned} \tag{29}$$

Since the one-forms $\mathrm{d}h, \mathrm{d}(F_0), \mathrm{d}h(F_0^2), \mathrm{d}h(F_0^3)$ and $\mathrm{d}\xi_5$ are linearly independent, the extended system (26) satisfies the condition (A). The vector field $g = (1/\xi_1)\partial/\partial\xi_4$. Taking into account that $\xi_4 = x_4$, $\xi_1 = \xi_5^{\langle 1\rangle} - v_{|v=0} = u^{\langle 1\rangle}$, observe that $\Xi = g$. Next, compute $\mathrm{Ad}_{F_0}g$ by Definition 5:

$$\mathrm{Ad}_{F_0}g = \left(\frac{\partial F_0}{\partial \xi}g\right)\Big|_{F_0^{-1}} = \begin{bmatrix} 0 \\ \frac{1}{\xi_1} \\ 0 \\ 0 \\ 0 \end{bmatrix}\Bigg|_{F_0^{-1}} = \begin{bmatrix} 0 \\ \frac{1}{\xi_5} \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Repeated application of the Ad operator yields

$$\mathrm{Ad}_{F_0} g = \frac{1}{\xi_5}\frac{\partial}{\partial \xi_2}, \quad \mathrm{Ad}^2_{F_0} g = \frac{\partial}{\partial \xi_3} + \frac{1+\xi_4}{\xi_3}\frac{\partial}{\partial \xi_4}, \quad \mathrm{Ad}^3_{F_0} g = \frac{\partial}{\partial \xi_1} + \frac{1+\xi_2}{\xi_1}\frac{\partial}{\partial \xi_2}.$$

Considering the relations (27), observe that $\mathrm{Ad}^{\ell}_{F_0} g = \Xi^{\langle -\ell\rangle\pi}$ for $\ell = 1, 2, 3$. The vector fields g and $\mathrm{Ad}_{F_0} g$ commute, while g and $\mathrm{Ad}^2_{F_0} g$ do not, thus the condition (B) indicates that $2 \leqslant s \leqslant n-1 = 3$. For $s = 2$, the condition (C) is satisfied since $\partial h^{\langle k\rangle}/\partial v = 0$ for $k = 0, 1, 2$. To check the condition (D), compute

$$\mathrm{Ad}_{F_v} g = \frac{1}{\xi_5 - v}\frac{\partial}{\partial \xi_2}.$$

Clearly, $\partial \mathrm{Ad}_{F_v} g/\partial v \neq 0$, thus (D) is not satisfied. Finally, $s = 3$ fails to satisfy the condition (C) since $h^{\langle 3\rangle}(\xi, v) = \tilde{y}^{\langle 3\rangle}$ depends on v; thus our guess $r = 1$ was wrong.

Consider the case $r = 2$. The equations $\xi^{\langle 1\rangle} = F(\xi, v)$ for $r = 2$ have the form

$$\xi_1^{\langle 1\rangle} = \xi_3,\ \xi_2^{\langle 1\rangle} = \xi_4,\ \xi_3^{\langle 1\rangle} = (\xi_2+1)\,\xi_5,\ \xi_4^{\langle 1\rangle} = \xi_2,\ \xi_5^{\langle 1\rangle} = \xi_6,\ \xi_6^{\langle 1\rangle} = \xi_1 + v,\ y = \xi_1, \tag{30}$$

where the states are given by

$$\xi_1 = x_1,\ \xi_2 = x_2,\ \xi_3 = x_3,\ \xi_4 = x_4,\ \xi_5 = u,\ \xi_6 = u^{\langle 1\rangle}. \tag{31}$$

The inverse is given by

$$F_v^{-1}(\xi) = \begin{bmatrix} \xi_6 - v & \xi_4 & \xi_1 & \xi_2 & \dfrac{\xi_3}{1+\xi_4} & \xi_5 \end{bmatrix}^{\mathrm{T}} \tag{32}$$

and

$$F_0^{-1}(\xi) = \begin{bmatrix} \xi_6 & \xi_4 & \xi_1 & \xi_2 & \dfrac{\xi_3}{1+\xi_4} & \xi_5 \end{bmatrix}^{\mathrm{T}}.$$

The forward shifts are $\tilde{y} = h = \xi_1$, $\tilde{y}^{\langle 1\rangle} = h(F_0) = \xi_3$, $\tilde{y}^{\langle 2\rangle} = h(F_0^2) = (1+\xi_2)\xi_5$, $\tilde{y}^{\langle 3\rangle} = h(F_0^3) = \xi_6(1+\xi_4)$. The condition (A) is satisfied. The vector fields are

$$g = \frac{1}{\xi_6}\frac{\partial}{\partial \xi_4},\ \mathrm{Ad}_{F_0} g = \frac{1}{\xi_5}\frac{\partial}{\partial \xi_2},\ \mathrm{Ad}^2_{F_0} g = \frac{\partial}{\partial \xi_3} + \frac{1+\xi_4}{\xi_3}\frac{\partial}{\partial \xi_4},\ \mathrm{Ad}^3_{F_0} g = \frac{\partial}{\partial \xi_1} + \frac{1+\xi_2}{\xi_1}\frac{\partial}{\partial \xi_2}. \tag{33}$$

Taking the relations (31) into account, we note that $g = \Xi$ and $\mathrm{Ad}^{\ell}_{F_0} g = \Xi^{\langle -\ell\rangle\pi}$ for $\ell = 1, 2, 3$. The vector fields g and $\mathrm{Ad}_{F_0} g$ commute, while g and $\mathrm{Ad}^2_{F_0} g$ do not; thus (B) is satisfied for $s = 2$. The condition (C) is satisfied. Since $\mathrm{Ad}_{F_v} g = (1/\xi_5)\partial/\partial \xi_2$ does not depend on v, the condition (D) is satisfied.

Unlike in [7], the state transformation has been found in two steps. In order to find the state transformation $Z = \Phi(x, u, u^{\langle 1\rangle})$, one has to complete the following steps:

1) Define $\varphi_1 := h = \xi_1, \varphi_2 := h(F_0) = \xi_3$.

2) To complete the functions φ_1, φ_2 into the set of new coordinates, one may choose $\varphi_3 := \xi_2, \varphi_4 := \xi_4$.

3) Compute η_i, $i = 1, \ldots, n-2$, in the φ-coordinates: $\eta_1 = \frac{1}{u^{\langle 1\rangle}}\frac{\partial}{\partial \varphi_4}$ and $\eta_2 = \frac{1}{u}\frac{\partial}{\partial \varphi_3}$. The matrix is defined by

$$\tilde{R}(\varphi, u, u^{\langle 1\rangle}) := \begin{bmatrix} \frac{1}{u} & 0 \\ 0 & \frac{1}{u^{\langle 1\rangle}} \end{bmatrix}.$$

4) The first s coordinates are $Z_1 := \varphi_1 = x_1$, $Z_2 := \varphi_2 = x_3$. The inverse matrix

$$\tilde{R}^{-1}(\varphi, u, u^{\langle 1\rangle}) = \begin{bmatrix} u & 0 \\ 0 & u^{\langle 1\rangle} \end{bmatrix}$$

yields the one-forms $\tilde{\theta}_3 = u\mathrm{d}\varphi_3$ and $\tilde{\theta}_4 = u^{\langle 1\rangle}\mathrm{d}\varphi_4$. Treating u and $u^{\langle 1\rangle}$ as integration constants allows one to integrate $\int \tilde{\theta}_3 = u\varphi_3 + C_3$ and $\int \tilde{\theta}_4 = u^{\langle 1\rangle}\varphi_4 + C_4$. The simplest option would be to take the integration constants $C_3 = C_4 = 0$; however, in this case the new state coordinates $Z_3 := ux_2$ and $Z_4 := u^{\langle 1\rangle}x_4$ would lead to the transformed system

$$Z_1^{\langle 1\rangle} = Z_2, \quad Z_2^{\langle 1\rangle} = Z_3 + u, \quad Z_3^{\langle 1\rangle} = Z_4, \quad Z_4^{\langle 1\rangle} = \frac{u^{\langle 2\rangle}(y^{\langle 2\rangle} - u)}{u}, \quad y = Z_1, \tag{34}$$

which is not in the generalized observer form (8). Since $u, u^{\langle 1\rangle}, \varphi_1, \varphi_2$ as the system invariants can be treated as the integration constants, one can also choose $C_3 = u$, resulting in $Z_3 = ux_2+u$. In such a case the transformed state equations are in the generalized observer form (25).

4.3. Study of singular points

The generic results of [7] are valid almost everywhere except in the so-called singular points. The goal of this subsection is to demonstrate, based on Example 1, whether the local approach from [9] can address such points. Since [9] studies the problem only around the equilibrium points, we have to first find the intersection of the sets of singular and equilibrium points.

Singular points of the generic approach appear at various steps of computations.

(a) Points where the system (23) is not submersive, i.e. points where the generic rank

$$\text{rank}_{\mathcal{K}} \frac{\partial f}{\partial(x,u)} = \text{rank}_{\mathcal{K}} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & u & 0 & 0 & x_2+1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} = n$$

drops. The rows of $\frac{\partial f}{\partial(x,u)}$ become dependent if $x_2 = -1$. However, this point is of no interest since submersivity is also the assumption in [9].

(b) Points where the equations (23) do not satisfy the generic observability assumption, i.e. where the forward shifts of $\mathrm{d}y$ up to the order $n-1$

$$\begin{bmatrix} \mathrm{d}y \\ \mathrm{d}y^{\langle 1\rangle} \\ \mathrm{d}y^{\langle 2\rangle} \\ \mathrm{d}y^{\langle 3\rangle} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & u & 0 & 0 \\ 0 & 0 & 0 & u^{\langle 1\rangle} \end{bmatrix}$$

are linearly dependent, resulting in the singular points $u = 0$ and $u^{\langle 1\rangle} = 0$.

(c) Points where Ξ or its backward shifts $\Xi^{\langle -i\rangle}, i = 1, \dots, n-1$, are not defined. From (24), one obtains the singular points $u = 0, u^{\langle 1\rangle} = 0, x_3 = 0, x_1 = 0$.

(d) Matrix M is not defined at the points $u = 0, u^{\langle 1\rangle} = 0$.

(e) Inspecting the state transformation and the new state equations reveals that the latter is not defined at $u = 0$.

Since six state equations of the extended system (30) depend on seven variables, there is, in general, an infinite number of equilibrium points, parametrized by one variable. To find the equilibrium points for the extended system (30) for $r = 2$, one has to solve the system of equations

$$\xi_1 = \xi_3, \ \xi_2 = \xi_4, \ \xi_3 = (\xi_2 + 1)\,\xi_5, \ \xi_4 = \xi_2, \ \xi_5 = \xi_6, \ \xi_6 = \xi_1 + v. \tag{35}$$

As the second and the fourth equations are identical, in this example one has, instead of one, two free parameters. Note that the solution of [9] is valid around the point $v = 0$, therefore fixing $v^{eq} := v^* = 0$ is necessary; for another free parameter we may pick, for instance, $\xi_4^{eq} := \xi_4^*$. From the first, fifth and sixth equations we obtain $\xi_1 = \xi_3 = \xi_5 = \xi_6$, the remaining equations yielding $\xi_2 = \xi_4 = \xi_4^*$. Our interest here is in the singular equilibrium points, hence we fix $\xi_1^{eq} = \xi_3^{eq} = \xi_5^{eq} = \xi_6^{eq} = 0$, which corresponds to the singular point (c) found above. The equilibrium points of the remaining coordinates are $\xi_2^{eq} = \xi_4^{eq} = \xi_4^*$, i.e. the equilibrium point under the study is $(\xi^{eq}, v^{eq}) = (0, \xi_4^*, 0, \xi_4^*, 0, 0, 0)$.

Finally, observe that the application of the differential geometric approach from [9] also fails in this point. Checking the condition (A) for the extended system (30) shows that at the point (ξ^{eq}, v^{eq}) the one-forms $\mathrm{d}h = \mathrm{d}\xi_1$, $\mathrm{d}(h \circ F_0) = \mathrm{d}\xi_3$, $\mathrm{d}(h \circ F_0^2) = (1 + \xi_4^*)\mathrm{d}\xi_5$, $\mathrm{d}(h \circ F_0^3) = (1 + \xi_4^*)\mathrm{d}\xi_6$, $\mathrm{d}\xi_5 = (1 + \xi_4^*)\mathrm{d}\xi_1$, $\mathrm{d}\xi_6 = (1 + \xi_4^*)\mathrm{d}\xi_3$ are not linearly independent; moreover, the vector field g, given by (33) for $r = 2$, is not defined whenever $\xi_6 = 0$.

5. Conclusions

Comparing the algebraic [7] and the differential geometric [9] approaches, both addressing the generalized observer form problem, revealed several noteworthy differences between them.

- The approach in [7] is generic, not local around an equilibrium point as in [9].
- The results of [7] are formulated in terms of the original system (though a new variable z is introduced to compute backward shifts), not in terms of the extended system relying also on the output feedback as in [9]. Furthermore, this means that [7], unlike [9], handles reversible and non-reversible cases simultaneously. Of course, the choice of the new variable z will depend on whether the system is reversible or not, but this does not show up in theory (only in computations).
- The paper [7], computes the new coordinates in a single step, whereas in [9], it is a two-step procedure. However, the computations for finding the parametrized state transformation are simpler in [9], though special care is needed in finding the integrating constant, associated with the $(s+1)$th new state.
- The paper [7] provides a geometric interpretation of the coordinate transformation as common invariants and canonical parameters of the set of vector fields Ξ and its backward shifts. We believe it is more transparent and allows easier comparison with the classical case. Of course, in principle, one gets the same parametrized state transformation in both papers.
- Unlike [7], the paper [9] does not address the problem of finding the minimal r.
- The approach in [7] allows generalizing the result to get the generalized observer form, depending on the backward but not on the forward shifts. It is unclear how to generalize the approach from [9].

Data availability statement

All data are available in the article.

Acknowledgements

The work of Maris Tõnso was supported by the Estonian Research Council grant PRG1463. The publication costs of this article were covered by the Estonian Academy of Sciences.

References

1. Califano, C., Monaco, S. and Normand-Cyrot, D. Canonical observer forms for multi-output systems up to coordinate and output transformations in discrete time. *Automatica*, 2009, **45**(11), 2483–2490. https://doi.org/10.1016/j.automatica.2009.07.003
2. Califano, C., Monaco, S. and Normand-Cyrot, D. On the observer design in discrete-time. *Syst. Control Lett.*, 2003, **49**(4), 255–265. https://doi.org/10.1016/s0167-6911(02)00344-4
3. Lee, H.-G. and Hong, J.-M. Algebraic conditions for state equivalence to a discrete-time nonlinear observer canonical form. *Syst. Control Lett.*, 2011, **60**(9), 756–762. https://doi.org/10.1016/j.sysconle.2011.06.001
4. Lee, W. and Nam, K. Observer design for autonomous discrete-time nonlinear systems. *Syst. Control Lett.*, 1991, **17**(1), 49–58. https://doi.org/10.1016/0167-6911(91)90098-y
5. Mullari, T. and Kotta, Ü. Transformation of nonlinear discrete-time state equations into the observer form: extension to non-reversible case. *Proc. Estonian Acad. Sci.*, 2021, **70**(3), 235–247. https://doi.org/10.3176/proc.2021.3.03
6. Mullari, T., Kotta, Ü., Bartosiewicz, Z., Pawłuszewicz, E. and Moog, C. H. Forward and backward shifts of vector fields: towards the dual algebraic framework. *IEEE Trans. Autom. Control*, 2017, **62**(6), 3029–3033. https://doi.org/10.1109/tac.2016.2608718
7. Mullari, T., Kotta, Ü., Kaldmäe, A., Kaparin, V. and Simha, A. Extended observer form with vector fields. *Int. J. Control*, 2024, **97**(10), 2399–2412. https://doi.org/10.1080/00207179.2023.2274060
8. Nam, K. Linearization of discrete-time nonlinear systems and a canonical structure. *IEEE Trans. Autom. Control*, 1989, **34**(1), 119–122. https://doi.org/10.1109/9.8665
9. Simha, A., Kaparin, V., Mullari, T. and Kotta, Ü. Extended observer forms for submersive discrete-time systems. *IEEE Trans. Autom. Control*, 2024, **69**(4), 2684–2688. https://doi.org/10.1109/tac.2023.3336253

Olekuvõrrandite teisendamine üldistatud vaatlejakujule: algebralise ja diferentsiaalgeomeetrilise lähenemise võrdlus

Ülle Kotta, Tanel Mullari ja Maris Tõnso

Artikkel võrdleb kahte erinevat metoodikat mittelineaarse diskreetse juhtimissüsteemi olekuvõrrandite teisendamiseks üldistatud vaatlejakujule. Üks lahendustest põhineb vektorväljade algebralisel aparatuuril, teises kasutatakse standardset diferentsiaalgeomeetrilist lähenemist. Artiklis võrreldakse meetodite rakendamiseks vajalikke eeldusi, lahenduse eksisteerimise tarvilikke ja piisavaid tingimusi, olekuteisenduse leidmise algoritme ja lahenduse kehtivuse piirkondi.