Homogeneous Observer Design for Finite-dimensional projections of Homogeneous PDEs

Sergiy Zhuk, Andrey Polyakov

Homogeneous Observer Design for Finite-dimensional projections of Homogeneous PDEs

Sergiy Zhuk and Andrey Polyakov

***Abstract*— Sufficient conditions for existence and uniqueness of solutions for a coupled system of homogeneous equations defining dynamics of the gain and observer for ODEs obtained as a Galerkin projection of homogeneous PDEs are proposed. The conditions rely upon fundamental concept of uniform complete observability which is also used to design an exponentially convergent observer. Convergence of the observer is confirmed by numerical experiments with ODEs obtained from a hyperbolic PDE in 1D (Burgers-Hopf equation).**

***Index Terms*— nonlinear filtering; uniform observability; Lyapunov exponents; bilinear systems; Riccati equations; fixed-time convergence**

I. Introduction

Stochastic filtering is fundamental in diverse fields including synchronization in complex networks, data assimilation and control engineering to name just a few. Theoretically, the optimal solution of stochastic filtering problem for Markov diffusions is given by the so-called Kushner-Stratonovich (KS) equation [1], a stochastic Partial Differential Equation (PDE) which describes evolution of the conditional density of the states of the underlying diffusion process. For linear systems, KS equation is equivalent to the Kalman-Bucy Filter (KF) equations.

In contrast, deterministic state estimators (observers), including the algorithm presented in this paper, assume that errors have bounded energy and belong to a given bounding set. The state estimate is then defined as the minimax center of the reachability set, and temporal dynamics of the minimax center is described by a minimax filter (or minimum energy estimate) [2]. There is a fundamental connection between stochastic filters and observers: namely, an observer can be obtained as asymptotic limit of KFs associated with the "noisy" version of the state equation and observations; when "noise" dissapears (e.g. noise variance approaches zero) the mean-squared error between stochastic and deterministic state estimates goes to zero, and, in fact, the conditional measure associated with the stochastic filter converges to a degenerate measure concentrated at the observer [2]. Having this in mind we consider the case of exact measurements and zero disturbances. In addition, we do not assume any form of diffusion (or damping): in fact, bilinear ODEs studied below can be obtained as Galerkin projection of nonlinear hyperbolic PDEs including Euler equation in two spatial dimensions [3], [4], or a finite-difference discretization of the Burgers equation in 1D (as discussed in the next section). Such systems have a number of conserved quantities, e.g. energy and volume, and so all the solutions evolve on a given bounded manifold (e.g. a sphere of a certain radius) and never approach zero. Hence, even the case of exact observations is very challenging for dynamical systems of this class.

Our key contribution is in using intrinsic symmetries provided by homogeneity of the state equation in combination with rather mild uniform observability assumptions along observer's trajectory to prove existence and uniqueness of solutions for the system of non-linear differential equations, which describe dynamics of the observer and its gain. The latter coincides with the classical Extended KF locally, on the basin of attraction. Our design relies upon Lyapunov methods like many other works on nonlinear observer design, however we do not require rather strict unifrom observability assumptions on the Jacobian which are imposed to prove existence and uniqueness of the observer and the gain (e.g. [2]), neither do we make an extra assumption of existense of the gain [5]. Yet another contribution is the description of the basin of attraction, which can be used in numerical computations as it is demonstrated below, and, more importantly, the description of the deformations of the basin in responce to rescaling of the set of "the true intial conditions". In addition, we quantify the effect of rescaling onto the estimation error convergence speed.

State estimation for bilinear systems is a challenging problem especially in the case of high dimension of the state vector. Indeed, in the latter case it is hard to implement the differential algebraic approach conventional for nonlinear systems [6], [7], [8]. Our design however is minimal in that one cannot make use of Lyapunov subspace reduction (e.g. [9], [10]) and restrict the observer's gain corrections just to the stable Lyapunov subspace due to the following fact: all the Lyapunov exponents of the considered dynamical system equal to zero. Nevertheless, unlike [9] we do not require to observe all "unstable directions" instead requiring the uniform complit observability along the observer's trajectory.

II. Mathematical preliminaries

In this section we introduce notations and collect all the required mathematical notions and results.

a) Notation: $\mathbb{R}^n$ – Euclidean space of n-dimensional real vectors with orthonormal basis $\{\psi_1 \dots \psi_n\}$, inner product $(x,y) = \sum_{i=1}^n (x,\psi_i)(y,\psi_i)$, and norm $\|x\|^2 = (x,x)$; $C(t_0,T,\mathbb{R}^n)$ – the space of continous $\mathbb{R}^n$-valued functions;

Sergiy Zhuk is with IBM Research, IBM Tech. campus, Damastown, Dublin, D15 HN66, Ireland sergiy.zhuk@ie.ibm.com

Andrey Polyakov is with Univ. Lille, Inria, CNRS, UMR 9189 CRIStAL, Centralle Lille, F-59000 Lille, France andrey.polyakov@inria.fr

$\dot{P}$ or $\dot{x}$ denotes the time derivative of the matrix- or vector-valued function of time; $\mathbb{S}^n$ – Hilbert space of symmetric non-negative definite $n \times n$-matrices. $\lambda_{\min}(P)$ and $\lambda_{\max}(P)$ denote minimal and maximal eigen-values of a matrix P; $\mathrm{Tr}(P)$ – trace of P; $I_n \in \mathbb{S}^n$ – the $n \times n$-identity matrix; $P = \{P_{ij}\}_{i,j=1}^n$ – stands for a matrix $P \in \mathbb{S}^n$ with components P_{ij}.

b) Trace Inequalities: If A is symmetric and B is skew-symmetric then $\mathrm{Tr}(AB + B^\top A) = 0$.

Lemma 1: If $0 < A \in \mathbb{S}^n$ then $\|A\|_\infty = \max_{ij} |a_{ij}| \leq \mathrm{Tr}\, A \leq \sqrt{n\, \mathrm{Tr}\, A^2}$.

A. Motivating examples: Finite dimensional projections

As noted in Section I the class of dynamical systems studied here includes finite dimensional projections of some important PDEs including Burgers(-Hopf) equation (in 1D) and Euler equations (in 2D).

Following [9] we consider the ODE obtained by discretizing the Burgers(-Hopf) equation $u_t = -\frac{\partial_\xi u^2}{2}$ on $(0,1)$ with periodic boundary conditions by using the finite difference scheme:

$$\dot{u}_i = -\frac{n}{6}\left(u_i(u_{i+1} - u_{i-1}) + (u_{i+1}^2 - u_{i-1}^2)\right) \tag{1}$$

taken on a periodic lattice ($i = 1 \dots n$, $u_{-1} = u_n$, $u_{n+1} = u_1$) which has the properties that

- the quadratic energy $\sum_i u_i^2$ is conserved, implying that every sphere in $\mathbb{R}^n$ is invariant under the motion of the system and $\|u\|$ is constant,
- the trace of the Jacobian of the r.h.s. of (1) is zero, implying that the flow conserves the volume of the phase element.

Denoting $x = (u_1 \dots u_n)^\top$ and setting $D = \{D_{ij}\}_{i,j=1}^n$ with $D_{ji} = -D_{ij}$ for $j \leq i$, and $D_{ii+1} = 1$, $D_{ij} = 0$ for $j > i+1$ except for $D_{1n} = -1$, we can rewrite (1) as follows: $\dot{x} = B(x)x$ with

$$B(x) = -\frac{n}{6}(\mathrm{diag}(x)D + D\,\mathrm{diag}(x)).$$

Clearly,

$$B^\top(x) = -\frac{n}{6}(D^\top \mathrm{diag}(x) + \mathrm{diag}(x)D^\top) = -B(x)$$

since $D^\top = -D$.

Euler[1] equations in 2D

$$\begin{aligned} &\partial_t \omega + \vec{u}\cdot\nabla\omega = 0\,, \quad -\Delta\psi = \omega\,, \\ &\vec{u} = \bar{u} + \nabla^\perp\psi\,, \omega(0) = \mathrm{curl}(\vec{u}_0)\,, \end{aligned} \tag{2}$$

with periodic boundary conditions at the boundary of a rectangular domain can be approximated by an ODE which has similar properties. Namely, assuming periodic boundary conditions, and applying Fourier-Galerkin (FG) approximation one can project Euler equation onto a $2N+1$-dimensional subspace generated by $\{e^{ikx}e^{isy}\}_{|k|,|s|\leq\frac{N}{2}}$ and obtain an ODE for the projection coefficients in the form $\dot{x} = B(x)x$ with skew-symmetric linear (in x) matrix B. In this case the components of x will represent projection coefficents of the solution. Analogously, for homogeneous Dirichlet conditions one can take sin-basis. We refer the reader to [3] for further details.

[1]More specifically, vorticity-streamfunction formulation of Euler equation

III. Main results

A. Problem statement

Let $x(t) \in \mathbb{R}^n$ and $y(t) \in \mathbb{R}^m$ denote the state vector and output of the following system:

$$\dot{x}(t) = B(x(t))\,x(t), \quad x(t_0) = x_0\,, \tag{3}$$

$$y(t) = C(t)x(t) \tag{4}$$

provided $B(x) = -B^\top(x)$ and $x \mapsto B(x)$ is a linear mapping, and $C(t) \in \mathbb{R}^{m\times n}$ is a given measurable matrix-valued function such that

$$\lambda_{\max}(C^\top(t)C(t)) \leq \bar{c} < +\infty\,.$$

Let $z(t_0) = z_0 \in \mathbb{R}^n$ and $P(t_0) = P_0 \in \mathbb{S}^n$ and consider the following system of equations:

$$\dot{z} = B(z)z + PC^\top R(y - Cz), \tag{5}$$

$$\dot{P} = \tilde{B}(z)P + P\tilde{B}^\top(z) - PC^\top RCP + Q, P(t_0) = P_0 \tag{6}$$

where $R \in \mathbb{S}^n$ and $Q \in \mathbb{S}^n$ are given continuous matrix-valued functions such that

$$0 < \underline{r} \leq \lambda_{\min}(R(t)) \leq \lambda_{\max}(R(t)) \leq \bar{r} < +\infty, \tag{7}$$

$$0 < \underline{q} \leq \lambda_{\min}(Q(t)) \leq \lambda_{\max}(Q(t)) \leq \bar{q} < +\infty, \tag{8}$$

$$\tilde{B}(z) = B(z) + B_1(z), \quad B_1(z) = \left[\begin{smallmatrix} B(\psi_1)z & \dots & B(\psi_n)z \end{smallmatrix}\right] \tag{9}$$

In the forthcoming sections we report the following results:

- existence and uniqueness of the unique bounded and continuous solution for the system (5)-(6) on any finite interval $[t_0, T]$, i.e. that neither z nor P solving (5)-(6) "blow-up" in finite time, see Section III-B, Theorem 1)
- z converges to x exponentially fast provided z_0 belongs to a certain vicinity of x_0, see Section III-D, Theorem 3
- construction of a homogeneous dilation which appropriately rescales (5)-(6) provided x_0 is rescaled as follows: $x_0 \mapsto \lambda x_0$, see Section III-E

B. Existence and uniqueness

Given $\rho > 0$ let ϕ_ρ be defined as follows

$$\phi_\rho(r) = \begin{cases} r & \text{if} \quad r \leq \rho, \\ \rho & \text{if} \quad r > \rho. \end{cases}$$

Theorem 1: Let $P_0 \in \mathbb{S}^n$ and $z_0 \in \mathbb{R}^n$ and $x_0 \in \mathbb{R}^n$. If $P_0, Q, R > 0$ and $x(t_0) = x_0$, $z(t_0) = z_0$ and $P(t_0) = P_0$ then for any fixed $\rho > 0$ the system of differential equations composed of (3) and

$$\dot{z} = B(z)z + PC^\top R(y - Cz), \tag{10}$$

$$\dot{P} = \tilde{B}_\rho(z)P + P\tilde{B}_\rho^\top(z) - PC^\top RCP + Q, \tag{11}$$

$$\tilde{B}_\rho(z) = B(z) + \phi_\rho(\|z\|)B_1\left(\frac{z}{\|z\|}\right) \tag{12}$$

has the unique solution (x, z, P) such that

(i) $x, y \in C(t_0, +\infty, \mathbb{R}^n)$ and $\|x(t)\| = \|x_0\|, \forall t \geq t_0$

(ii) $z \in C(t_0, T, \mathbb{R}^n)$ and $P \in C(t_0, T, \mathbb{S}^n)$ for any $T < +\infty$

(iii) $P(t) > p_1 I$ for some $p_1 > 0$.

Proof: Since $B(x) = -B^\top(x)$ it follows that $(x, \dot{x}) = (x, B(x)x) = 0$, i.e. $\|x(t)\| = \|x(t_0)\| = \|x_0\| < +\infty$ for all $t \geq t_0$. Hence $x \in C(t_0, +\infty, \mathbb{R}^n)$ and so is y as $\bar{c} < +\infty$. This demonstrates (i).

Denote by $F(z, P, t)$ a mapping from $\mathbb{R}^n \times \mathbb{S}^n \times \mathbb{R}^+$ to $\mathbb{R}^n \times \mathbb{S}^n$ which is defined by the right hand side of eqs. (10) and (11). Clearly, F is continuously differentiable w.r.t. z and P everywhere on $\mathbb{R}^n \times \mathbb{S}^n$, hence $F(\cdot, \cdot, t)$ is locally Lipschitz continuous w.r.t. z and P for all t ($y(t)$ is bounded!). By using a fixed-point iteration (the standard argument based on Picard theorem) we conclude that eqs. (10) and (11) have the unique continuous solution (z, P) defined on a maximal[2] segment $[t_0, T)$, $T > t_0$, and that one of the following cases holds:

1) $T = +\infty$
2) $T < +\infty$, $\sup_{t\in[t_0,T)} \|z(t)\| < +\infty$ and $\|P(t)\| \to +\infty$ as $t \to T$
3) $T < +\infty$, $\|z(t)\| \to +\infty$ as $t \to T$ and $\sup\limits_{t\in[t_0,T)} \|P(t)\| < +\infty$
4) $T < +\infty$, $\|z(t)\| \to +\infty$ and $\|P(t)\| \to +\infty$ as $t \to T$
5) $T < +\infty$, $\sup_{t\in[t_0,T)} \|z(t)\| < +\infty$, $\sup_{t\in[t_0,T)} \|P(t)\| < +\infty$ and $P(T)$ is not positive definite.

To conclude the proof let us show that the case 1) above is the only possible one, i.e. that for any $T < +\infty$ there exists the unique solution of eqs. (10) and (11) such that $z \in C(t_0, T, \mathbb{R}^n)$ and $P \in C(t_0, T, \mathbb{S}^n)$ (claim (ii)), and that $P(t) > 0$, $\|P^{-1}\| < +\infty$ for all $t \in [t_0, T]$ provided $P_0, Q, R > 0$ (claim (iii)).

Indeed, by contradiction, assume that either 2) or 4) holds true. Note that $\operatorname{Tr}\left(\tilde{B}_\rho P + P\tilde{B}_\rho^\top\right) = \phi_\rho \operatorname{Tr}\left(B_1(\frac{z}{\|z\|})P + PB_1^\top(\frac{z}{\|z\|})\right)$ as $\operatorname{Tr}(B(z)P) = 0$. Recalling that $\operatorname{Tr}(B^\top A)$ is the inner product of matrices B, A we apply Schwartz inequality:

$$\left|\operatorname{Tr}\left(PB_1^\top(\frac{z}{\|z\|})\right)\right| \leq \operatorname{Tr}^{\frac{1}{2}}(P^2)\operatorname{Tr}^{\frac{1}{2}}\left(B_1\left(\frac{z}{\|z\|}\right)B_1^\top\left(\frac{z}{\|z\|}\right)\right)$$

$$\leq \lambda_{\max}^{\frac{1}{2}}(\tilde{B}_1)\operatorname{Tr}(P), \quad \tilde{B}_1 = \{\operatorname{Tr}(B_1(\psi_i)B_1^\top(\psi_j)\}_{i,j=1}^n$$

Applying Tr to both sides of eq. (11) and recalling that $\phi_\rho(\|z\|) = \rho$ if $\|z\| > \rho$ we get:

$$\operatorname{Tr}(\dot{P}) = \operatorname{Tr}\left(\tilde{B}_\rho P + P\tilde{B}_\rho^\top\right) + \operatorname{Tr}(Q) - \operatorname{Tr}(PC^\top RCP)$$

$$\leq \operatorname{Tr}(Q) + 2\lambda_{\max}^{\frac{1}{2}}(\tilde{B}_1)\rho\operatorname{Tr}(P),$$

[2]Maximal means that the solution cannot be extended beyond $[t_0, T)$

By Bellman-Gronwall lemma it now follows that

$$\operatorname{Tr}(P(t)) \leq e^{2\rho(t-t_0)\lambda_{\max}^{\frac{1}{2}}(\tilde{B}_1)}\operatorname{Tr}(P(t_0)) + \operatorname{Tr}(Q)\frac{e^{2\rho(t-t_0)\lambda_{\max}^{\frac{1}{2}}(\tilde{B}_1)} - 1}{2\rho\lambda_{\max}^{\frac{1}{2}}(\tilde{B}_1)} \tag{13}$$

hence $\lambda_{\max}(P(t)) \leq \operatorname{Tr}(P(t)) < +\infty$ on $[t_0, T]$ which contradicts both 2) and 4).

Assume that 3) holds true. Since $P(t_0) > 0$ it follows that $P(t) > 0$ for all $t \in [t_0, T)$ (see [11]), and hence $P^{-1}(t)$ is well defined on $t \in [t_0, T)$, and it is bounded from below on $t \in [t_0, T)$ as $\|P(t)\|$ is uniformly bounded from above on $[t_0, T)$ by 3). Define $V_0 = z^\top P^{-1} z$, set $e = x - z$. If $V_0(z(T)) < +\infty$ then $\|z(T)\| < +\infty$. Recalling that $B(z)z = B_1(z)z$ consider

$$\begin{aligned}\dot{V}_0 &= -\|R^{\frac{1}{2}}Ce\|^2 + \|R^{\frac{1}{2}}Cx\|^2 - \|Q^{\frac{1}{2}}P^{-1}z\|^2 \\ &- 2\phi_\rho(P^{-1}z, B(\frac{z}{\|z\|})z) \leq \|R^{\frac{1}{2}}Cx\|^2 + \phi_\rho^2\|Q^{-\frac{1}{2}}B(\frac{z}{\|z\|})\| \\ &- \|Q^{\frac{1}{2}}P^{-1}z + \phi_\rho Q^{-\frac{1}{2}}B(\frac{z}{\|z\|})\|^2 \leq \|R^{\frac{1}{2}}Cx\|^2 + \rho^2\tilde{\lambda}V_0\end{aligned} \tag{14}$$

with $\tilde{\lambda} = \lambda_{\max}(P^{\frac{1}{2}}B(\frac{z}{\|z\|})Q^{-1}B(\frac{z}{\|z\|})P^{\frac{1}{2}})$. As P is uniformly bounded on $[t_0, T)$ by assumption it follows that $\tilde{\lambda} < +\infty$, $t \in [t_0, T]$. Thus (14) implies (e.g. using Bellman-Gronwall lemma) that $V_0(z(T)) < +\infty$, which contradicts 3).

Finally, assume that 5) holds true. By assumption, $P(t) > 0$ on $[t_0, T)$ and $P(T)x = 0$ for some $x \neq 0$. Hence, $W = P^{-1}(t)$ is defined for every $t \in [t_0, T)$, and $\lim_{t\uparrow T}\|W(t)\| = +\infty$. In this case, $\operatorname{Tr} W(t) \to \infty$ as $t \to T$ due to Lemma 1. For any $t \in [t_0, T)$ we have:

$$\dot{W} = -W\dot{P}W = -\tilde{B}_\rho^\top W - W\tilde{B}_\rho - WQW + C^\top RC\,.$$

Since $\operatorname{Tr}(A) = \sum_{j=1}^n \psi_i^\top A\psi_i$ and

$$\operatorname{Tr}(WQW) = \sum_{j=1}^n \psi_i^\top(WQW)\psi_i \geq \lambda_{\min}(Q)\sum_{j=1}^n \psi_i^\top W^2\psi_i$$

then

$$\begin{aligned}\operatorname{Tr}\dot{W} &= -\operatorname{Tr}(\tilde{B}_\rho^\top W + W\tilde{B}_\rho + \operatorname{Tr}(C^\top RC) - \operatorname{Tr}(WQW) \\ &\leq \operatorname{Tr}(C^\top RC) - \frac{q}{n}\operatorname{Tr}^2(W) + 2\lambda_{\max}^{\frac{1}{2}}(\tilde{B}_1)\rho\operatorname{Tr}(W).\end{aligned}$$

which shows that $\|W(T)\| < +\infty$. This completes the proof. ∎

Remark 1: Note that the theorem also holds for the case of $\rho = 0$ which was considered in [12].

C. Uniform bounds for DRE

Definition 1: Let z solve (10) and let $\dot{X} = \tilde{B}_\rho(z)X$, $X(0) = X_0$ and set $\Phi(t, s) = X(t)X^{-1}(s)$. A symmetric matrix $\mathcal{N}_\rho(t, t-\sigma; R, z)$ is said to be *a gramian along the trajectory z of* (10) if $\mathcal{N}_\rho(t, t-\sigma; R, z) = \int_{t-\sigma}^t \Phi^\top(s, t)C^\top RC\Phi(s, t)ds$.

Obviously, $\mathcal{N}_\infty$ is a gramian for z being a solution of (5).

It is well-known that for LTV systems invertibility of the gramian allows to prove that the observer $z(t)$ converges to the true state $x(t)$ for the case of exact observations [13], [11]. In what follows we generalize this result to systems of class (3). To that end note tha eqs. (3), (10) and (11) has the unique solution which is uniquely defined by the initial conditions x_0, z_0, P_0.

Definition 2: Given $\rho \geq 0$ we say that the system eqs. (3), (10) and (11) is *uniformly completely observable* from x_0, z_0, P_0 if there exist $\nu > 0$ and $\alpha, \beta > 0$ such that

$$\alpha I < \mathcal{N}_\rho(t, t-\sigma; R, z) < \beta I, \quad \forall t > t_0 + \sigma\,.$$

Theorem 2: Take $\rho > 0$. If the system eqs. (3), (10) and (11) is *uniformly completely observable* from x_0, z_0, P_0 then

$$\begin{aligned}\lambda_{\max}(P(t)) &\leq \frac{1}{\underline{r}\alpha} + \int_{t-\sigma}^{t} \lambda_{\max}(Q) e^{\rho\bar{\lambda}(t-s)} ds \\ &\leq \frac{1}{\underline{r}\alpha} + \frac{\bar{q}}{\rho\bar{\lambda}}(e^{\rho\bar{\lambda}\sigma} - 1)\,, \quad t \geq t_0 + \sigma,\end{aligned} \tag{15}$$

where $\bar{\lambda}^2 = \sum_{i=1}^n \lambda^2_{\max}(B_1(\psi_i) + B_1^\top(\psi_i))$.

Proof: Assume that $t < +\infty$ and consider the standard LQR design problem for the system $\dot{q} = -\tilde{B}_\rho^\top(z)q + C^\top u$, $q(t) = h$ with the cost

$$J(u) = \|P^{\frac{1}{2}}(t_0)q(t_0)\|^2 + \int_{t_0}^{t} \|R^{-\frac{1}{2}}u\|^2 + \|Q^{\frac{1}{2}}q\|^2 ds$$

It is known that the feedback $\hat{u} = RCPq$ minimizes J and $J(\hat{u}) = (P(t)h, h)$. Hence for any other control u we have that $J(\hat{u}) = (P(t)h, h) \leq J(u)$. Let us select u as follows: set $u(s) = RC(s)\Phi(s,t)\mathcal{N}_\rho^{-1}(t, t-\sigma; R, z)h$ for $s \in [t-\sigma, t]$ and set $u(s) = 0$ for $s \in [t_0, t-\sigma)$. Let us show that for this u one gets: $q(s) = 0$ for all $s \in [t_0, t-\sigma]$. To this end recall that

$$q(s) = \Phi^\top(t,s)h - \int_s^t \Phi^\top(\tau, s)C^\top(\tau)u(\tau)d\tau$$

and so $q(t-\sigma) = 0$ if[3]

$$\Phi^\top(t, t-\sigma)h = \Phi^\top(t, t-\sigma)\int_{t-\sigma}^{t} \Phi^\top(\tau, t)C^\top(\tau)u(\tau)d\tau$$

Clearly, for the above choice of u we get that $q(s) = 0$ for all $s \in [t_0, t-\sigma]$. Hence, for this u we get that

$$\begin{aligned}J(\hat{u}) = (P(t)h, h) &\leq \int_{t-\sigma}^{t} (R^{-1}u, u) + (Qq, q)ds \\ &= (\mathcal{N}_\rho^{-1}(t, t-\sigma; R, z)h, h) + \int_{t-\sigma}^{t}(Qq, q)ds \\ &\leq \frac{\|h\|^2}{\underline{r}\alpha} + \int_{t-\sigma}^{t}(Qq, q)ds\end{aligned} \tag{16}$$

since $\mathcal{N}_\rho(t, t-\sigma; R, z) \geq \underline{r}\mathcal{N}_\rho(t, t-\sigma; I, z) \geq \underline{r}\alpha$. To compute $\int_{t-\sigma}^{t} q^\top Q q ds$ we first note that

$$q(s) = \Phi^\top(t,s)(I - \mathcal{N}_\rho(t, s; R, z)\mathcal{N}_\rho^{-1}(t, t-\sigma; R, z))h$$

$$= \Phi^\top(t,s)\mathcal{N}_\rho(s, t-\sigma; R, z)\mathcal{N}_\rho^{-1}(t, t-\sigma; R, z)h$$

for $s \geq t - \sigma$ and so

$$\int_{t-\sigma}^{t}(Qq, q)ds = \|W^{\frac{1}{2}}(t, t-\sigma)\mathcal{N}^{-1}(t, t-\sigma; R, z)h\|^2$$

with

$$W(t,\tau) = \int_\tau^t \mathcal{N}(s, \tau; R, z)\Phi(t,s)Q\Phi^\top(t,s)\mathcal{N}(s, \tau; R, z)ds$$

Now, recall Vazhevskii estimate from [14, p. 110]:

$$\|\Phi(t,s)q(s)\| \leq \|q(s)\| e^{\int_s^t \frac{1}{2}\lambda_{\max}\left(\tilde{B}_\rho(z(\tau)) + \tilde{B}_\rho^\top(z(\tau))\right)d\tau}$$

so that $\|\Phi(t,s)\| \leq e^{\int_s^t \frac{1}{2}\lambda_{\max}(\tilde{B}_\rho(z(\tau)) + \tilde{B}_\rho^\top(z(\tau))d\tau}$, and the latter estimate is exact provided $\tilde{B} = B$. Clearly

$$\begin{aligned}\lambda_{\max}(\tilde{B}_\rho(z) + \tilde{B}_\rho^\top(z)) &\leq \frac{\rho}{\|z\|}\sum_{i=1}^{n} z_i \lambda_{\max}(B_1(\psi_i) + B_1^\top(\psi_i)) \\ &\leq \rho\bar{\lambda}\end{aligned}$$

Hence $\|\Phi(t,s)\| \leq e^{\frac{1}{2}\rho\bar{\lambda}(t-s)}$. Now

$$\begin{aligned}(w, \Phi(t,s)Q\Phi^\top(t,s)w) &\leq \lambda_{\max}(Q)\lambda_{\max}(\Phi(t,s)\Phi^\top(t,s))\|w\|^2 \\ &= \lambda_{\max}(Q)\lambda_{\max}(\Phi^\top(t,s)\Phi(t,s))\|w\|^2 \\ &= \lambda_{\max}(Q)\|\Phi(t,s)\|^2\|w\|^2\end{aligned}$$

so that:

$$W(t, t-\sigma) \leq \int_{t-\sigma}^{t} \lambda_{\max}(Q)e^{\rho\bar{\lambda}(t-s)}\mathcal{N}_\rho^2(s, t-\sigma; R, z)ds.$$

Finally, we note that $\mathcal{N}_\rho^2(s, t-\sigma; R, z) \leq \mathcal{N}_\rho^2(t, t-\sigma; R, z)$ hence $W(t, t-\sigma) \leq \mathcal{N}_\rho^2(t, t-\sigma; R, z)\int_{t-\sigma}^{t}\lambda_{\max}(Q)e^{\rho\bar{\lambda}(t-s)}ds$ and so $\int_{t-\sigma}^{t} z^\top Q z ds \leq \|h\|^2\int_{t-\sigma}^{t}\lambda_{\max}(Q)e^{\rho\bar{\lambda}(t-s)}ds$. This and (16) completes the proof. ■

Corollary 1: If $\bar{\lambda} = 0$ then

$$\begin{aligned}\lambda_{\max}(P(t)) &\leq \frac{1}{\underline{r}\alpha} + \int_{t-\sigma}^{t}\lambda_{\max}(Q)ds \\ &\leq \frac{1}{\underline{r}\alpha} + \bar{q}\sigma, \quad t \geq t_0 + \sigma\end{aligned} \tag{17}$$

Remark 2: Note that eq. (17) coincides with the upper bound derived in [12].

D. Convergence

Let us introduce the following definitions: $\beta^2 = \sum_i \|B(\psi_i)\|^2$ and

$$\begin{aligned}\bar{p} = \max\Big(&e^{2\rho\sigma\lambda_{\max}^{\frac{1}{2}}(\tilde{B}_1)}\operatorname{Tr}(P(t_0)) + \operatorname{Tr}(Q)\frac{e^{2\rho\sigma\lambda_{\max}^{\frac{1}{2}}(\tilde{B}_1)} - 1}{2\rho\lambda_{\max}^{\frac{1}{2}}(\tilde{B}_1)}, \\ &\frac{1}{\underline{r}\alpha} + \frac{\bar{q}}{\rho\bar{\lambda}}(e^{\rho\bar{\lambda}\sigma} - 1)\Big),\end{aligned} \tag{18}$$

Theorem 3: Set $\rho = \|x(t_0)\| + \frac{\bar{q}}{2\beta\operatorname{Tr}(P(t_0))}$, $V_0 = \frac{\bar{q}}{2\beta\bar{p}^{\frac{3}{2}}}$ and let z_ρ, P_ρ solve eqs. (10) and (11). Assume that system eqs. (3), (10) and (11) is *uniformly completely observable* from $x(t_0) = x_0$, $z(t_0) = z_0$ and $P(t_0) = P_0 > 0$. If

[3]We used the obvious equality $\Phi(\tau, t-\sigma) = \Phi(\tau, t)\Phi(t, t-\sigma)$

$V(e,t)=(P^{-1}e,e)$ and $V^{\frac{1}{2}}(e(t_0),t_0)\le\kappa V_0$ for $0<\kappa<1$ then

- $\|z_\rho(t)\|\le\rho$ for all $t\ge t_0$ and z_ρ,P_ρ coincide with the unique solution z,P of (5)-(6)
- $\mathcal{N}_\rho(t,t-\sigma;I,z_\rho)=\mathcal{N}(t,t-\sigma;I,z)$;
- $V(e(t),t)\le e^{-\frac{(1-\kappa)\bar q(t-s)}{\bar p}}V(e(s),s)$ for all $t\ge s$.

Proof: Note that since $\lambda_{\max}(P(t_0))\le\mathrm{Tr}(P(t_0))\le\bar p$ then $\bar p^{-\frac{1}{2}}\|x(t_0)-z(t_0)\|\le V^{\frac{1}{2}}(e(t_0),t_0)<V_0$ implies that

$$\begin{aligned}\|z(t_0)\|&\le\|e(t_0)\|+\|x(t_0)\|<V_0\bar p^{\frac{1}{2}}+\|x(t_0)\|\\&=\frac{\bar q}{2\beta\bar p}+\|x(t_0)\|\le\rho\end{aligned}\tag{19}$$

and so $\|z(t)\|<\rho$ on a small interval $[t_0,t_1]$. Then, on $[t_0,t_1]$ we have that $\phi_\rho(\|z\|)\tilde B(\frac{z}{\|z\|})=B(z)+B_1(z)$ and thus z_ρ,P_ρ also solve (5) and (6) on $[t_0,t_1]$. Thus, by direct calculation one finds that the estimation error equation takes the following form: $\dot e=B(z)e+B_1(x)e-PC^\top RCe$, $e(t_0)=e_0$. Hence, to complete the proof it is suffies to demonstrate that $\|z\|$ does not grow over time. To this end consider the following Lyapunov function $V(e,t)=(P^{-1}e,e)$. Compute $\dot V(e,t)=2(P^{-1}e,B(e)e)-(RCe,Ce)-\bar q\|P^{-1}e\|^2$. It is not hard to see that

$$\dot V(e,t)\le\|P^{-1}e\|\|e\|\left(2\|B(e)\|-\bar q\frac{\|P^{-1}e\|}{\|e\|}\right)-(RCe,Ce)$$

Note that $\lambda_{\min}(P^{-1})\le\frac{\|P^{-1}e\|}{\|e\|}$ and $\lambda_{\min}(P^{-1})=\lambda^{-1}_{\max}(P)\ge\bar p^{-1}$, and $\|B(e)\|\le\|e\|\beta$. Also $\lambda_{\min}(P^{-1})(e,e)\le(P^{-1}e,e)$ hence $\|e\|\le\lambda^{\frac{1}{2}}_{\max}(P)\|P^{-\frac{1}{2}}e\|\le\bar p^{\frac{1}{2}}V^{\frac{1}{2}}(e,t)$. Hence

$$2\|B(e)\|-\bar q\frac{\|P^{-1}e\|}{\|e\|}\le2\beta\|e\|-\frac{\bar q}{\lambda_{\max}(P)}\le2\beta\bar p^{\frac{1}{2}}V^{\frac{1}{2}}-\frac{\bar q}{\bar p}$$

and so $2\|B(e)\|-\bar q\frac{\|P^{-1}e\|}{\|e\|}<0$ provided $V^{\frac{1}{2}}(e(t),t)<\frac{\bar q}{2\beta\bar p^{\frac{3}{2}}}=V_0$. Assume that $V^{\frac{1}{2}}(e(t_0),t_0)<\kappa V_0$. The latter also holds for all $t\in[t_0,t_1-\varepsilon]$ for some $\varepsilon>0$, and on $[t_0,t_1-\varepsilon]$ we get:

$$\begin{aligned}\dot V(e,t)&\le-\|P^{-1}e\|\|e\|\big(\frac{\bar q}{\bar p}-2\beta\bar p^{\frac{1}{2}}V^{\frac{1}{2}}\big)\\&\le-V\big(\frac{\bar q}{\bar p}-2\beta\bar p^{\frac{1}{2}}\kappa V_0\big)\end{aligned}\tag{20}$$

as $0\le V(e(t),t)\le\|P^{-1}e\|\|e\|$ and so $-V\ge-\|P^{-1}e\|\|e\|$. Since $\frac{\bar q}{\bar p}-2\beta\bar p^{\frac{1}{2}}\kappa V_0=\frac{(1-\kappa)\bar q}{\bar p}>0$ it follows that $V(e(t),t)<V(e(t_0),t_0)<\kappa V_0$ on $[t_0,t_1-\varepsilon]$. This and (20) implies that $V(e(t),t)\le e^{-\frac{(1-\kappa)\bar q(t-s)}{\bar p}}V(e(s),s)\le\kappa V_0$ for all $t\ge t_0$. Hence, as above $\bar p^{-\frac{1}{2}}\|x(t)-z(t)\|\le V^{\frac{1}{2}}(e(t),t)<\kappa V_0$. The latter and (19) shows that $\|z(t)\|<\rho$ for $t\ge t_0$. ∎

E. Homogeneity: dilation symmetry

It is well known that the homogeneity (dilation symmetry) of a vector field is inherited by its flow. Indeed, if $f\in C^1(\mathbb{R}^m,\mathbb{R}^m)$

$$\dot\xi=f(\xi),\ f(\lambda\xi)=\lambda^2f(\xi),\forall\lambda>0,\forall\xi\in\mathbb{R}^m$$

then

$$\xi(t,\lambda\xi_0)=\lambda\xi(\lambda t,\xi_0),\tag{21}$$

where $\xi(\cdot,\bar\xi)$ is a solution of the ODE with the initial condition $\xi(0)=\bar\xi$. The obtained symmetry of solutions means that any solution with a scaled initial data $\lambda\xi_0$ can be obtained using a solution with $\xi(0)=\xi_0$ by means of the scaling of both time and state vector.

Let us rewrite the system (6) in the form

$$\dot P=\tilde BP+P\tilde B^\top-PC^\top RCP+\Xi^2,P(t_0)=P_0\tag{22}$$

$$\dot\Xi=0,\Xi(t_0)=\Xi_0,\tag{23}$$

where Ξ_0 is an arbitrary symmmtric positive definite matrix. Assuming that $t_0=0$, $C(t)\equiv C$, $R(t)\equiv R$ and $Q(t)\equiv Q$ we derive that the system (3),(5), (22), (23) can be rewritten as follows $\dot\xi=F(\xi)$, where $\xi(t)=(x(t),z(t),P(t),\Xi(t))$. It is easy to see that $F(\lambda\xi)=\lambda^2F(\xi)$ and the identity (21) holds for $\xi_0=(x_0,z_0,P_0,\Xi_0)$. Therefore, twe have proven the following claim.

Corollary 2: Let $t_0=0$, $C(t)=const$, $R(t)=const$ and $Q(t)=const$ and $z(t)\to x(t)$ as $t\to+\infty$ for some $x_0\in\mathbb{R}^n,z_0\in\mathbb{R}^n$ and $P_0\in\mathbb{S}^n$ then the observer (5), (6) will converge to the state of the system (3) with the scaled initial condition $x_0\to\lambda x_0,\lambda>0$ provided that initial conditions of the observer are scaled $z_0\to\lambda z_0$, $P_0\to\lambda P_0$ as well and the matrix Q is scaled quadratically with respect to λ, i.e. $Q\to\lambda^2Q$.

IV. Experiments

We perform a number of numerical experiments for the ODE obtained by discretizing the Burgers(-Hopf) equation $u_t=-\frac{\partial_\xi u^2}{2}$ on $(0,1)$ with periodic boundary conditions by using the finite difference scheme eq. (1). Recall that this ODE conserves energy and multidimensional volume: $\|x(t)\|=\|x_0\|$ and the Lebesgue measure of the set of initial conditions is preserved by ODE's semigroup.

We set $n=8$ and select C so that $Cx=(x_2,x_4x_6)^\top$, i.e. 3 out of 8 components of x were measured. We stress that the estimate of $\bar p$ given in eq. (18) is very conservative from the practical standpoint. To circumvent this we take the smallest possible $\bar p$ but setting $\rho=0$ so that

$$\bar p=\max\{\mathrm{Tr}(P(t_0))+\mathrm{Tr}(Q)\sigma,\frac{1}{\underline r\alpha}+\bar q\sigma\}$$

To estimate α and σ numerically we discretized eqs. (3), (5) and (6) together with the following Lyapunov equation

$$\dot{\mathcal{N}}(t,t_0)=-A^\top\mathcal{N}(t,t_0)-\mathcal{N}(t,t_0)A+C^\top(t)C(t),$$

with condition $\mathcal{N}(t_0,t_0)=0$ for the gramian $\mathcal{N}$, $A=B(z)+B_1(z)$ in time by using RK method of 4th order with $\Delta t=1.3\times10^{-4}$ on the time interval $[0,T]$ with $T=40$. $\mathcal{N}$ was computed on a set of intervals $(\Delta ti,\Delta ti+\sigma)$ with $\sigma=8$, and $i=0,1,2,\dots N$, N is such that $N\Delta t+\sigma\le T$. At the beginning of each interval $\mathcal{N}(\Delta t(i-1),\Delta t(i-1))$ was set to $0I_8$, and $\lambda_{\min}(\mathcal{N}(\Delta ti+\sigma,\Delta ti))$ was computed to get an estimate for α. In all experiments we observed that for $\sigma=8$ we always had $\lambda_{\min}(\mathcal{N}(\Delta ti+\sigma,\Delta ti))>1\times10^{-3}$.

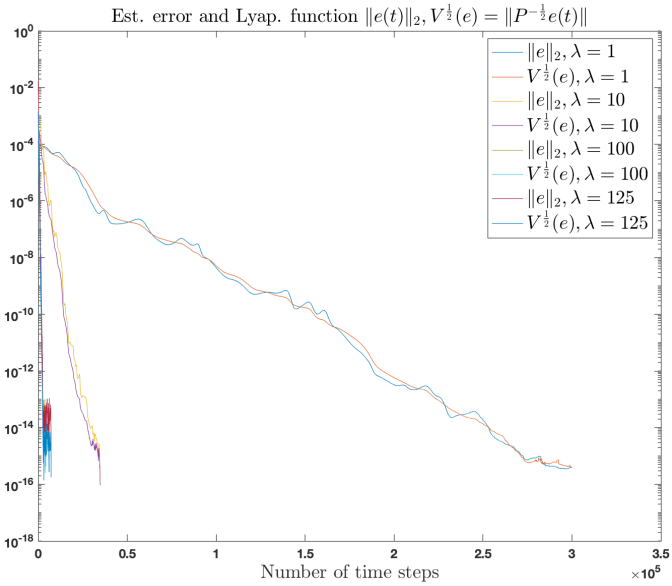

Fig. 1: Convergence of the estimation error and Lyapunov function for an ensemble of 4 members with rescaled initial conditions and parameters ($\lambda = 1, 10, 100, 125$, log-scale).

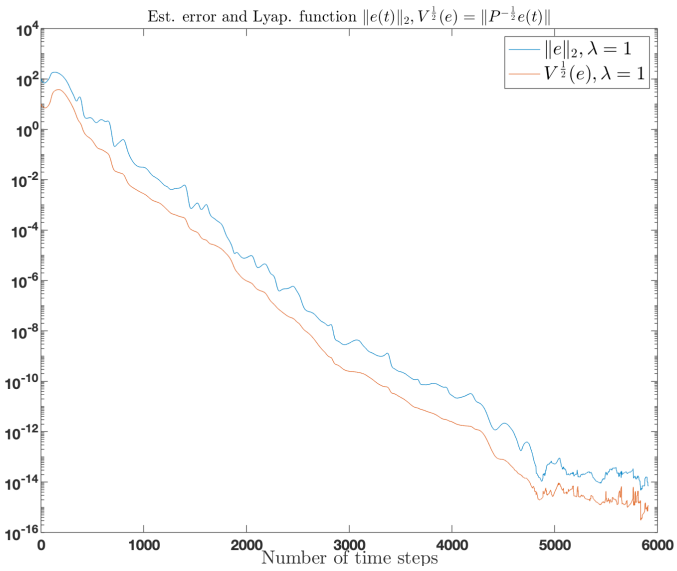

Fig. 2: Convergence of the estimation error and Lyapunov function for z_0 which is outside of the estimated basin of attraction.

Based on this we assumed that $\alpha > 1 \times 10^{-3}$ and $\sigma = 8$. We also computed that $\beta = 7.54$.

To compute the basin of attraction, $V^{\frac{1}{2}}(e(t_0), t_0) \leq \kappa V_0$ with $V_0 = \frac{\bar{q}}{2\beta\bar{p}^{\frac{3}{2}}}$ we took $P(0) = p_0 I_8$, $Q = \bar{q} I_8$, $R = \bar{r} I_3$ with $p_0 = 1$, $\bar{q} = 0.25$, $\bar{r} = 1 \times 10^2$. In this case we must select z_0 so that $\|x_0 - z_0\| \leq \theta_0 = \frac{\kappa p_0^{\frac{1}{2}} \bar{q}}{2\beta\bar{p}^{\frac{3}{2}}}$. Note that if we rescale $p_0(\lambda) = \lambda p_0$, $\bar{q}(\lambda) = \lambda^2 \bar{q}$ and $x_0(\lambda) = \lambda x_0$, $z_0(\lambda) = \lambda z_0$ then $\bar{p}(\lambda) = \lambda \bar{p}$ then the convergence radius rescales as follows: $\theta(\lambda) = \lambda \theta_0$.

Finally we set $\kappa = 0.9$ and generated x_0 randomly, projected it onto the unit ball and then performed 4 experiments for different values of λ:

- $\lambda = 1$: pick z_0 so that $\|x_0 - z_0\| \leq \theta(\lambda) = 1.27 \times 10^{-4}$
- $\lambda = 10$: pick z_0 so that $\|x_0 - z_0\| \leq \theta(\lambda) = 0.0013$
- $\lambda = 10$: pick z_0 so that $\|x_0 - z_0\| \leq \theta(\lambda) = 0.013$
- $\lambda = 125$: pick z_0 so that $\|x_0 - z_0\| \leq \theta(\lambda) = 0.015$

For each λ we rescaled parameters and the convergence radius as follows: $p_0(\lambda) = \lambda p_0$, $\bar{q}(\lambda) = \lambda^2 \bar{q}$ and $x_0(\lambda) = \lambda x_0$, $\theta(\lambda) = \lambda \theta_0$. The results are presented in Fig. fig. 1. Clearly, the decay of the Lyapunov function V is monotone (unlike the estimation error!), and rescaling the parameters speeds up the convergence. In Fig. fig. 2 we started z_0 outside of the estimated basin of attraction, yet after a transient period when V is increasing, the decay of V returns to monotone.

V. Conclusions

This contribution provides sufficient conditions for existense and uniquness of the system of non-linear differential equations eqs. (5) and (6), which describe dynamics of the observer and its gain. Basin of attraction for the "true solution" which evolves on a sphere and may have other invariants (e.g. conservation of volume) is described as a ball of certain radius. The radius is expressend in terms of upper bounds on the gain and parameters of the observer. It is also shown that the radius scales linearly if so are "true intial condition" together with observer's and gain's intial conditions, and if the constant term in the gain's equation scales quadratically.

References

[1] R. Stratonovich, "On the theory of optimal non-linear filtering of random functions," *Theory of Probability and Its Applications*, no. 4, p. 223–225.

[2] J. Baras, A. Bensoussan, and M. James, "Dynamic observers as asymptotic limits of recursive filters: Special cases," *SIAM Journal on Applied Mathematics*, vol. 48, no. 5, pp. 1147–1158, 1988.

[3] S. Zhuk, T. T. Tchrakian, and J. Frank, "Exponentially convergent data assimilation algorithm for navier-stokes equations," in *2017 American Control Conference (ACC)*, May 2017, pp. 3249–3256.

[4] S. Zhuk and T. Tchrakian, "Parameter estimation for euler equations with uncertain inputs," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 572–577.

[5] D. Hinrichsen and A. J. Pritchard, *Modelling, State Space Analysis, Stability and Robustness*. Springer, 2005.

[6] M. Fliess, J. Lévine, P. Martin, and P. Rouchon, "Flatness and defect of non-linear systems: introductory theory and examples," *International journal of control*, vol. 61, no. 6, pp. 1327–1361, 1995.

[7] A. Isidori, *Nonlinear Control Systems Design 1989: Selected Papers from the IFAC Symposium, Capri, Italy, 14-16 June 1989*. Elsevier, 2014.

[8] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Transactions on Automatic Control*, vol. 22, no. 5, pp. 728–740, October 1977.

[9] J. Frank and S. Zhuk, "A detectability criterion and data assimilation for nonlinear differential equations," *Nonlinearity*, vol. 31, no. 11, p. 5235, 2018.

[10] M. Tranninger, R. Seeber, S. Zhuk, M. Steinberger, and M. Horn, "Detectability analysis and observer design for linear time varying systems," *IEEE Control Systems Letters*, vol. 4, no. 2, pp. 331–336, 2019.

[11] B. Anderson, "Stability properties of kalman-bucy filters," *Journal of the Franklin Institute*, vol. 291, no. 2, pp. 137–144, 1971.

[12] S. Zhuk and A. Polyakov, "On practical fixed-time convergence for differential riccati equations," 2019.
[13] R. S. Bucy, "The riccati equation and its bounds," *Journal of computer and system sciences*, vol. 6, no. 4, pp. 343–353, 1972.
[14] L. Y. Adrianova, *Introduction to linear systems of differential equations*. American Mathematical Soc., 1995.