TECHNISCHE UNIVERSITÄT CHEMNITZ

Peter Benner Sabine Hein

Model Predictive Control Based on an LQG Design for Time-Varying Linearizations

CSC/09-07

Chemnitz Scientific Computing Preprints

Chemnitz Scientific Computing Preprints – ISSN 1864-0087

Impressum:

Chemnitz Scientific Computing Preprints — ISSN 1864-0087
(1995–2005: Preprintreihe des Chemnitzer SFB393)

Herausgeber:
Professuren für
Numerische und Angewandte Mathematik
an der Fakultät für Mathematik
der Technischen Universität Chemnitz

Postanschrift:
TU Chemnitz, Fakultät für Mathematik
09107 Chemnitz
Sitz:
Reichenhainer Str. 41, 09126 Chemnitz

http://www.tu-chemnitz.de/mathematik/csc/

[2] E. F. Camacho and C. Bordons, *Model Predictive Control*, Springer, 2004, 2.ed., London.

[3] R. Griesse, *Parametric Sensitivity Analysis for Control-Constrained Optimal Control Problems Governed by Systems of Parabolic Partial Differential Equations*, Dissertation, Mathematisches Institut, Universität Bayreuth, 2003.

[4] R. Griesse and S. Volkwein, *A Primal-Dual Active Set Strategy for Optimal Boundary Control of a Nonlinear Reaction-Diffusion System*, SIAM J. Cont. Optim., 44(2), 467–494, 2005.

[5] L. Grüne, *Private communication.*

[6] K. Ito and K. Kunisch, *On asymptotic properties of receding horizon optimal control*, SIAM J. Cont. Optim., 40(5), 1455–1472, 2001.

[7] K. Ito and K. Kunisch, *Receding horizon optimal control for infinite dimensional systems*, ESAIM, Control Optim. Calc. Var., 8, 741–760, 2002.

[8] K. Ito and K. Kunisch, *Receding Horizon Control with Incomplete Observations*, SIAM J. Cont. Optim., 45(1), 207–225, 2006.

[9] R.E. Kalman and R.S. Bucy, *New Results in Linear Filtering and Prediction Theory*, Trans. ASME, Series D, 83, 95–108,1961.

[10] H. Abou-Kandil and G. Freiling and V. Ionescu and G. Jank, *Matrix Riccati Equations in Control and Systems Theory*, Birkhauser, 2003, Basel.

[11] A. Locatelli, *Optimal Control: An Introduction*, Birkhauser, 2001, Basel, Boston, Berlin.

[12] H. Mena, *Numerical Solution of Differential Riccati Equations Arising in Optimal Control Problems for Parabolic Partial Differential Equations*, Dissertation, Escuela Politécnica Nacional, Quito (Ecuador), 2007.

[13] H. Schlitt, *Systemtheorie für stochastische Prozesse*, Springer, 1992, Berlin.

TECHNISCHE UNIVERSITÄT CHEMNITZ

Chemnitz Scientific Computing Preprints

Peter Benner Sabine Hein

Model Predictive Control Based on an LQG Design for Time-Varying Linearizations

CSC/09-07

Abstract

We consider the solution of nonlinear optimal control problems subject to stochastic perturbations with incomplete observations. In particular, we generalize results obtained by Ito and Kunisch in [8] where they consider a receding horizon control (RHC) technique based on linearizing the problem on small intervals. The linear-quadratic optimal control problem for the resulting time-invariant (LTI) problem is then solved using the linear quadratic Gaussian (LQG) design. Here, we allow linearization about an instationary reference trajectory and thus obtain a linear time-varying (LTV) problem on each time horizon. Additionally, we apply a model predictive control (MPC) scheme which can be seen as a generalization of RHC and we allow covariance matrices of the noise processes not equal to the identity. We illustrate the MPC/LQG approach for a three dimensional reaction-diffusion system. In particular, we discuss the benefits of time-varying linearizations over time-invariant ones.

Keywords. Receding horizon control, model predictive control, nonlinear systems, optimal control, linear quadratic Gaussian design, incomplete observations, noise, LTV systems.
AMS subject classifications (MSC2010). 93C10, 49M20, 49N35.

CSC/09-07 ISSN 1864-0087 **November 2009**

Contents

Author's addresses:

Peter Benner benner@mathematik.tu-chemnitz.de
Sabine Hein sabine.hein@mathematik.tu-chemnitz.de
TU Chemnitz
Fakultät für Mathematik
D-09107 Chemnitz

http://www.tu-chemnitz.de/mathematik/

when using efficient solvers for the DREs (sparse solvers) we can limit this time significantly. On the other hand, it does not seem to be worthwile to use DREs when LTI systems are available.

5 Conclusions

We have shown that the results of Ito and Kunisch in [8] can be generalized in several directions: first, we slightly generalize the receding-horizon approach to an MPC scheme. For the linearization on the time horizon, we study the use of instationary reference trajectories leading to LTV systems. Essentially, we can obtain analogous results as in [8] for the time-invariant case. We also allow stochastic disturbances in the input, the output/measurements, as well as in the initial conditions. This makes the suggested MPC/LQG/LTV scheme fairly robust against external disturbances and modeling errors. We also note that in summary, the proposed MPC/LQG/LTV scheme generalizes the RHC/LQG/LTI design. The presented example emphasizes the possible performance improvement obtainable using LTV systems embedded in an MPC/LQG approach. Future work will include improvement of the numerical methods used to compute the feedback control laws, in particular we will employ large-scale DRE solvers directly linked to the simulation software.

Acknowledgements

The work reported in this paper was supported by a grant of the DFG *(Deutsche Forschungsgemeinschaft)*, BE 3715/1-1: "Numerical Solution of Optimal Control Problems with Instationary Diffusion-Convection and Diffusion-Reaction Equations".

We would also like to thank Lars Grüne (Universität Bayreuth) for helpful discussions about the performance of the MPC strategy, in particular for pointing out the gap in the original proof in [8].

References

[1] P. Benner and H. Mena, *BDF methods for large-scale differential Riccati equations*, in B. De Moor, B. Motmans, J. Willems, P. Van Dooren, V. Blondel: Proceedings Sixteenth International Symposium on: Mathematical Theory of Network and Systems, MTNS, Leuven, Belgium, 2004.

We want to present some figures which emphasizes the results in the table.

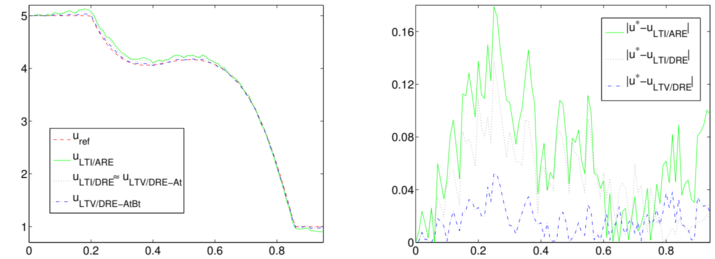

Figure 1: u (left) and $|u^* - u|$ (right) on $[0, 0.95]$

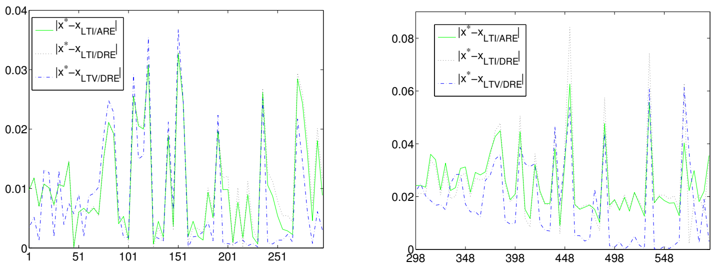

Figure 2: $|c_1^* - c_1|$ (left) and $|c_2^* - c_2|$ (right) on 59 discretization points at $t = 0.95$

The left figure in 1 shows the controls for the different cases choosing the length of the prediction and control horizon as $T_p = 0.1$ and $T_c = 0.05$, respectively. It confirms the results from the table. Since it is difficult to distinguish between the cases, we plotted the difference $|u^*(t) - u(t)|$ on the right side. There, the smaller deviation to the reference control for LTV systems becomes obvious.

In Figure 2 we plotted the difference $|c_i^*(0.95) - c_i(0.95)|$ for the first substance ($i = 1$, left side) and second substance ($i = 2$, right side) on 59 discretization points, respectively. Therefor we took every fifth point from the 297 discretization points of each concentration vector at time point $t = 0.95$. As seen in the table, we obtain similar deviations from the reference for the first concentrations and an improvement in the second concentration for the LTV case.

Finally, we can say that using a time-varying linearization improves the results for this example considerably. Of course, we have longer computation times, but

1 Introduction

In the following we consider the optimal control problem

$$\min \int_0^{T_f} y^T(t)Q(t)y(t) + u^T(t)R(t)u(t)\,dt + G(x(T_f)) \qquad (1)$$

subject to the semi-linear stochastic system with additive unmodeled disturbance

$$\dot{x}(t) = f(x(t)) + B(t)u(t) + F(t)v(t), \quad t > 0, \quad x(0) = x_0 + \eta, \qquad (2)$$

$u(t) \in \mathcal{U}$, $x(t) \in \mathcal{X}$, where $v(t)$ is an unknown Gaussian disturbance process with covariance V and η denotes the noise in the initial condition. Since in many applications the state is not completely available, we also consider the output function

$$y(t) = C(t)x(t) + w(t), \quad y \in \mathcal{Y}, \qquad (3)$$

where $w(t)$ is a measurement noise process which will also be assumed to be Gaussian with covariance W. The output equation (3) can be seen as a model for the measurements of (2) available in practice. If (2) is an ordinary differential equation (ODE), then we have a finite-dimensional problem with $\mathcal{X} = \mathbb{R}^n$, $\mathcal{Y} = \mathbb{R}^p$ and $\mathcal{U} = \mathbb{R}^m$. In the case of a partial differential equation (PDE) the problem is infinite-dimensional and $\mathcal{X}, \mathcal{Y}, \mathcal{U}$ are in general appropriate function spaces. Here we will only consider the finite-dimensional case, e.g., resulting from a spatial semi-discretization of a PDE to obtain an ODE. Extensions of the results in this paper to the infinite-dimensional case also treated in [7] will be reported elsewhere.

Based on the ideas of [8] we will derive an approximate solution of (1) by applying a model predictive control strategy and linear quadratic Gaussian design. In [8] Ito and Kunisch consider the system

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + Bu(t) + v(t), \quad t > 0, \quad x(0) = x_0 + \eta, \\ y(t) &= Cx(t) + w(t). \end{aligned}$$

There, the state equation is linearized about a given reference $(x^*(t), u^*(t))$ on small intervals $[t_i, t_i + T]$, $T < T_f$ (receding horizon control). Then, $x^*(t)$ is partially replaced by a stationary operating point $\bar{x}$ in order to obtain an LTI system. The resulting linear-quadratic optimal control problem is then solved by applying an LQG design, which requires in particular the solution of two algebraic Riccati equations (AREs) on each horizon. Furthermore, a quality criterion to estimate the performance of the strategy is developed.

We will generalize the results from [8] by using the resulting LTV system after linearization instead of the LTI system arising after partially replacing $x^*(t)$ by

$\bar{x}$. We expect a better approximation with respect to suboptimality since the linearization around a time-varying reference is more precise than replacing it by a constant operating point. Additionally, it is advantageous for practical problems, where the system matrices are time-varying, too. But it should also be noted that the computational effort is larger for LTV systems since we have to solve differential Riccati equations (DREs) instead of AREs if we want to apply an LQG controller. In studying the performance of the MPC strategy, we observed [5] what we believe is a gap in the proof of the stated bound for the combined estimation and tracking error in [8]. Besides generalizing the result to the more general situation considered here, we also provide a fix for this problem.

Furthermore we do not want to limit ourselves to receding horizon techniques (RHC), but also consider model predictive control (MPC). MPC means that we predict the system's behaviour on a small interval $[t_i, t_i + T_p]$ by using a model (prediction step), determine the control over an optimization horizon $[t_i, t_i + T_o]$, $T_0 \leq T_p$, in such a way, that a cost functional is optimized (optimization step) and apply the computed control on $[t_i, t_i + T_c]$, $T_c \leq T_o$ (implementation step) and recede the horizon to $[t_i + T_c, t_i + T_c + T_p]$ (receding horizon step). Thereby, we will set $T_o = T_p$ in the following, since in our case it makes no sense to linearize the problem on a larger horizon than optimization will be done. RHC is a special case of MPC for $T_c = T_p$. See [2] for further details on MPC.

The third additional (as compared to [8]) ingredient is the inclusion of noise covariance matrices different from the identity as it is suggested in text books dealing with LQG design. They will play a role in the DREs.

Another goal of this paper is to demonstrate that the MPC/LQG approach is applicable in the context of large-scale nonlinear dynamical systems, arising, e.g., from semidiscretized PDE control problems. For this purpose, we illustrate the performance of the proposed MPC/LQG strategy using a 3D reaction-diffusion system from [3, 4], resulting in a semilinear parabolic optimal control problem. Discretization by the finite element method then leads to a nonlinear finite-dimensional optimal control problem as considered here.

The outline of the paper is as follows. In the next section, we explain the MPC/LQG approach to solving the nonlinear optimal control (1). The analysis of the performance of this control strategy is then given in Section 3. The benefits of using time-varying linearizations are demonstrated using the numerical example mentioned above in Section 4. We close the paper by conclusions and an outlook in Section 5.

- **LTV/DRE:** We use LTV systems as introduced in Section 2 with time-varying A and B, where the latter means that the position of the nozzle changes in each time step.

We will compare the three cases for different prediction and control horizons.

T_p	T_c	Type	$\int_0^{\tilde{T}} z^T Q z + \tilde{u}^T R \tilde{u}\, dt$	$\int_0^{\tilde{T}} z_1^T z_1\, dt$	$\int_0^{\tilde{T}} z_2^T z_2\, dt$	$\int_0^{\tilde{T}} \tilde{u}^T \tilde{u}\, dt$
0.1	0.05	LTI/ARE	0.644872	0.067804	0.521638	0.005543
		LTI/DRE	0.623733	0.070703	0.524184	0.002885
		LTV/DRE	0.129287	0.068377	0.057168	0.000374
0.05	0.05	LTI/ARE	0.646785	0.067504	0.523364	0.005592
		LTI/DRE	0.612729	0.068944	0.529253	0.001453
		LTV/DRE	0.131773	0.068104	0.061985	0.000168
0.1	0.1	LTI/ARE	0.823303	0.061546	0.687169	0.007459
		LTI/DRE	0.812116	0.061004	0.724103	0.002701
		LTV/DRE	0.145055	0.067758	0.073827	0.000347

Table 1: Costs for LTI/ARE, LTI/DRE, LTV/DRE on $[0, 0.91]$

Table 1 shows the total costs (which should be minimized), the deviation from the reference substances and the deviation from the reference control. The first which can be noticed, is that total costs for the LTV case are much better than for the LTI cases. This is due to the fact, that we have a time-varying linearization and a time-varying nozzle. The partial costs for the concentration of the first substance are similar, but in the LTV case this is reached with much less effort, which is also reflected in the control costs and related to it in the costs for the second substance. If we compare the two LTI cases, one can see that there is only a small improvement of the total costs if we compute DREs with constant coefficients instead of AREs. But the control costs could be improved significantly. So for the two LTI cases we have to weigh the small improvement of the costs against the shorter computation times, since solving DREs is more extensive than solving AREs. Using the BDF method for the solution of an DRE means to compute an ARE in each time step instead of solving only the ARE on the prediction interval.

With respect to the choices of the prediction and control intervals we obtain expected results. So choosing $T_p = T_c$ (RHC) produces better results for smaller intervals. If we have a fixed control horizon, but different prediction horizons, the results are better for larger prediction horizons, since the LQG controller can work foresighted. So using an MPC approach is advantageous in contrast to the RHC approach.

The concentrations of the substances are denoted by c_i and the parameters D_i and k are the diffusion and reaction coefficients, respectively. We use Ω for a 3D annular cylinder where Ω_u is the upper surface. The function $\alpha(x,t)$ models a counter-clockwise revolving nozzle around the upper annular surface. This nozzle sprays one of the substances onto the reactor and the goal is to achieve a desired terminal concentration of the substances.

So we have to solve an optimal control problem subject to a nonlinear (semilinear) partial differential equation. Further we allow the presence of Gaussian white noise in measured outputs and process dynamics.

The discretization in space is performed by using piecewise linear and globally continuous finite elements on tetrahedra. On each time-horizon, a linearization around the reference trajectory is computed. This yields the following linear time-varying system for the difference $z(t)$ between reference and current estimate of the solution

$$\mathbf{M}\dot{z}(t) = \mathbf{A}(t)z(t) + \mathbf{B}(t)(\tilde{u}(t) + v(t)), \quad z(0) = \eta, \quad y(t) = Cx(t) + w(t),$$

see Section 2.

In order to be able to use standard software in MATLAB, we choose a coarse discretization with $n = 594$. The basic routines are coded in MATLAB, the FEM is done in FEMLAB and the DREs are solved with a BDF method as described in [1, 12]. Addressing larger problems requires efficient DRE solvers for sparse and low-rank data matrices; this is work in progress.

For this example, we assume full measurements and choose the following parameters:

$$\begin{array}{lllll} D_1 = 0.15, & D_2 = 0.2, & k = 1, & c_{10} = 1, c_{20} = 0, & T = 1, \\ dt = 0.01, & C = Q = I_{594}, & R = 10, & \sigma(v) = \sigma(w) = 0.5, & \eta = 0. \end{array}$$

So the aim is to steer c_1 to zero by spraying the second substance onto the reactor. Reference control and trajectory is computed by a primal-dual solution method whose code was provided by Roland Griesse, see [4].

We will distinguish the three cases

- **LTI/ARE:** We assume a time-invariant matrix A on each horizon, which is realized if we partially replace $x^*(t)$ by a constant operating point $\bar{x}$ and we fix the nozzle position in the middle of the control horizon. So we obtain an LTI systems and assume infinite prediction horizons to solve AREs.
- **LTI/DRE:** We take the same LTI system as a basis as described in the LTI/ARE case, but we use the real finite prediction intervals and obtain two DREs with constant coefficients on each interval.

2 MPC/LQG solution strategy

The MPC/LQG approach is based on a linearization of (2) on small intervals to obtain a linear time-varying problem. We solve this linear problem on this small interval by using an LQG design. So the strategy is the following:

(1) **Prediction and optimization step on $[t_i, t_i + T_p]$, $T_p < T_f$:**

linearize (2) around a given reference $(x^*(t), u^*(t))$ to obtain $A(t) = f'(x^*(t))$ and the linear state equation

$$\dot{z}(t) = A(t)z(t) + B(t)\tilde{u}(t) + F(t)v(t), \quad z(0) = \eta, \quad y(t) = C(t)x(t) + w(t),$$

with $z(t) = x(t) - x^*(t)$ and $\tilde{u}(t) = u(t) - u^*(t)$. Then solve the DRE

$$\begin{aligned} \dot{X}(t) &= -X(t)A(t) - A^T(t)X(t) + X(t)B(t)R^{-1}(t)B^T(t)X(t) \\ &\quad -C^T(t)Q(t)C(t), \\ X(t_i + T_p) &= G, \end{aligned} \tag{4}$$

in order to obtain $X_*(t)$ and $K(t) = -R^{-1}(t)B^T(t)X_*(t)$.

(2) **Implementation step on $[t_i, t_i + T_c]$, $T_c \leq T_p$:**

feed the original system with

$$u(t) = u^*(t) - K(t)(\hat{x}(t) - x^*(t)),$$

on $[t_i, t_i + T_c]$ while measuring $y(t)$ by solving the nonlinear ODE and estimating the next state $\hat{x}(t)$ through solving the linear ODE

$$\dot{\hat{z}}(t) = A(t)\hat{z}(t) + B(t)\tilde{u}(t) + L(t)\big(y(t) - C(t)\hat{x}(t)\big), \quad \hat{z}(t) = \hat{x}(t) - x^*(t),$$

with $L(t) = \Sigma_*(t)C^T(t)W^{-1}$, where $\Sigma_*(t)$ is the solution of the filter DRE (FDRE)

$$\begin{aligned} \dot{\Sigma}(t) &= A(t)\Sigma(t) + \Sigma(t)A^T(t) - \Sigma(t)C^T(t)W^{-1}C(t)\Sigma(t) + F(t)VF(t)^T, \\ \Sigma(t_i) &= \Sigma_0. \end{aligned} \tag{5}$$

V, W are the covariance matrices of the noise processes.

(3) **Receding Horizon Step:**

update $t_i = t_i + T_c$ and go to the first step.

Remark 2.1 *Matrix G in the terminal condition for the DRE (4) results from the terminal cost in the cost functional to penalize the state at the end of the horizon. If G is selected as an control Lyapunov function, Ito and Kunisch established the asymptotic stability and performance estimate for the receding horizon synthesis for finite-dimensional systems in [6]. The initial condition Σ_0 for the FDRE (5) can be chosen as the expected value of $\eta\eta^T$.*

One main item in this approach is to solve the DREs (4) and (5) and we refer the reader for example to [10] for detailed information about DREs. To solve large-scale DREs we will use the backward differentiation formulae (BDF) method, which requires the solution of an ARE in every time step. The BDF method is discussed in detail in [12].

The Kalman filter theory for the estimate of the next state can be found, for instance, in [9, 13].

3 Performance of the Compensator for LTV Systems

We introduce the *performance of the compensator* in a similar way as it was done by Ito and Kunisch in [8] for LTI systems.

Definition 3.1 *(Performance of the compensator)*

$$E(t) = \Big[\frac{1}{2}\big\langle x(t)-x^*(t), X(t)(x(t)-x^*(t))\big\rangle + \frac{1}{2}\big\langle x(t)-\hat{x}(t), \Sigma^{-1}(t)(x(t)-\hat{x}(t))\big\rangle\Big]^{\frac{1}{2}}, \quad (6)$$

where $X(t)$ and $\Sigma(t)$ are the solutions of the DRE (4) and the FDRE (5). So the performance contains the tracking error in the first part and the estimation error in the second part.

Note that in the definition of $E(t)$ in the LTI case in [8], the factors $\frac{1}{2}$ are missing and the brackets are not around the whole term. We include them as otherwise the definition would not be consistent with the achieved performance result.

Before presenting the main result we prove two propositions, which are the generalizations for LTV systems of Propostions 2.1 and 2.2 in [8].

Proposition 3.2 *Let $N(x) = \frac{1}{2}x^T(t)X(t)x(t)$. For $t \in [0, T_p]$ we have*

$$\begin{aligned}\frac{d}{dt}N(x-x^*) = & -\frac{1}{2}\Big[\big\langle X(t)B(t)R^{-1}(t)B^T(t)X(t)(x(t)-x^*(t)), x(t)-x^*(t)\big\rangle \\ & + \big\langle C^T(t)Q(t)C(t)(x(t)-x^*(t)), x(t)-x^*(t)\big\rangle\Big] \\ & + \langle F(t)v(t)+r(x,x^*), X(t)(x(t)-x^*(t))\rangle \\ & + \big\langle B(t)R^{-1}(t)B^T(t)X(t)(x(t)-\hat{x}(t)), X(t)(x(t)-x^*(t))\big\rangle,\end{aligned}$$

with $c = \max\{\sqrt{c_X}, \sqrt{c_\Sigma}\}$.

If we evaluate the value on the right hand side of the interval by exploiting (A11) we obtain with $i = 1$

$$\begin{aligned}E(t_2)|_{[t_1,t_2]} &\le e^{-\tilde{\omega}T_c}E(t_1)|_{[t_1,t_2]} + \int_0^{T_c} e^{-\tilde{\omega}(T_c-s)}\gamma(t_1+s)\,ds \\ &\le e^{-\tilde{\omega}T_c}c\frac{\delta}{2} + \frac{\delta}{2}\left(1-ce^{-\tilde{\omega}T_c}\right) = \frac{\delta}{2}.\end{aligned}$$

By induction we can show that from

$$E(t_i)|_{[t_{i-1},t_i]} \le \frac{\delta}{2},$$

it follows that

$$E(t_i)|_{[t_i,t_{i+1}]} \le c\frac{\delta}{2}$$

and (A11) implies

$$E(t_{i+1})|_{[t_i,t_{i+1}]} \le \frac{\delta}{2} \quad \forall i.$$

This implies

$$E(t) \le c\frac{\delta}{2}e^{-\tilde{\omega}t} + \int_0^t e^{-\tilde{\omega}(t-s)}\gamma(s)\,ds \quad \forall t > 0.$$

□

4 Numerical Example

The model for the following three dimensional example was provided by Roland Griesse, see [3, 4].

Our aim is to model a chemical or biological process where the species involved are subjected to diffusion and reaction among each other. This process can be modeled by a coupled system of reaction-diffusion equations ($i = 1, 2$)

$$(c_i)_t(x,t) = D_i\Delta c_i(x,t) - kc_1(x,t)c_2(x,t), \quad i = 1,2 \text{ on } \Omega\times(0,T)$$

with the boundary conditions

$$\begin{aligned}&\tfrac{\partial}{\partial n}c_1(x,t) = 0 \text{ on } \delta\Omega\times(0,T),\\ &\tfrac{\partial}{\partial n}c_2(x,t) = 0 \text{ on } (\delta\Omega\setminus\Omega_u)\times(0,T),\ \tfrac{\partial}{\partial n}c_2(x,t) = \alpha(x,t)u(t) \text{ on } \Omega_u\times(0,T)\end{aligned}$$

and the initial conditions $c_1(x,0) = c_{10}(x)$ and $c_2(x,0) = c_{20}(x)$.

$\frac{1}{\beta_2}I < \Sigma(t_i) < \frac{1}{\beta_1}I$. Moreover, the intervals need to be chosen small enough (or α_1, β_1 small and α_2, β_2 large enough) so that the Riccati solutions remain in the prescribed bounds.

It is quite difficult to show how to realize this but it is plausible that this can be fulfilled by choosing the intervals small enough.

Assumption (A10) (LTV-analogue to (3.2) in [8]) *The smallness condition (A10) is similar to Assumption (A6) in Theorem 3.5. Here, we require the smallness condition on every sub-interval $[t_i, t_{i+1}]$.*

Assumption (A11) (LTV-analogue to (3.3) in [8]) *(A11) is a more special assumption since we consider the integral on the whole sub-interval ($t = T_c$ in Assumption (A10)). Note, that there is an additional constant compared to that in [8], which is needed to overcome the difficulties with the discontinuity points discussed in Remark 3.7.*

Proof of Theorem 3.8:
Using the assumptions in the theorem we obtain analogously to (19)

$$E(t) \leq e^{-\tilde{\omega}(t-t_i)} E(t_i)|_{[t_i,t_{i+1}]} + \int_0^{t-t_i} e^{-\tilde{\omega}(t-t_i-s)} \gamma(t_i+s)\, ds \quad \text{for } t \in [t_i, t_{i+1}] \text{ and every } i,$$

if we shift the interval $[0, T_c]$ to $[t_i, t_i + T_c]$.

Now consider the first interval $[0, t_1]$. Then we obtain with $i = 0$, $t_0 = 0$ and (A7), (A11)

$$\begin{aligned} E(t_1)|_{[0,t_1]} &\leq e^{-\tilde{\omega}T_c} E(0)|_{[0,t_1]} + \int_0^{T_c} e^{-\tilde{\omega}(T_c-s)} \gamma(s)\, ds \\ &\leq e^{-\tilde{\omega}T_c} \frac{\delta}{2} + \frac{\delta}{2}\left(1 - e^{-\tilde{\omega}T_c}\right) = \frac{\delta}{2}. \end{aligned}$$

With (A9) we can move to the next interval and have

$$\begin{aligned} E(t_1)|_{[t_1,t_2]} &= \Big[\frac{1}{2}\big\langle x(t_1) - x^*(t_1), X_1(t_1)(x(t_1) - x^*(t_1))\big\rangle \\ &\qquad + \frac{1}{2}\big\langle x(t_1) - \hat{x}(t_1), \Sigma_1^{-1}(t_1)(x(t_1) - \hat{x}(t_1))\big\rangle\Big]^{\frac{1}{2}} \\ &\leq \Big[\frac{1}{2}c_X\big\langle x(t_1) - x^*(t_1), X_0(t_1)(x(t_1) - x^*(t_1))\big\rangle \\ &\qquad + \frac{1}{2}c_\Sigma\big\langle x(t_1) - \hat{x}(t_1), \Sigma_0^{-1}(t_1)(x(t_1) - \hat{x}(t_1))\big\rangle\Big]^{\frac{1}{2}} \\ &\leq cE(t_1)|_{[t_0,t_1]} \leq c\frac{\delta}{2}, \end{aligned}$$

where

$$r(x, x^*) = f(x) - f(x^*) - A(t)(x - x^*). \tag{7}$$

Proof: Consider

$$\frac{d}{dt}(x(t) - x^*(t)) = (f(x) + B(t)u(t) + F(t)v(t)) - (f(x^*) + B(t)u^*(t))$$

with

$$u(t) = u^*(t) - R^{-1}(t)B^T(t)X(t)(\hat{x}(t) - x^*(t)).$$

Then we obtain

$$\begin{aligned} &\frac{d}{dt}(x(t) - x^*(t)) \\ &\quad = f(x) - f(x^*) + B(t)u^*(t) - B(t)R^{-1}(t)B^T(t)X(t)(\hat{x}(t) - x^*(t)) \\ &\qquad - B(t)u^*(t) + F(t)v(t) \\ &\quad = A(t)(x(t) - x^*(t)) - A(t)(x(t) - x^*(t)) \\ &\qquad - B(t)R^{-1}(t)B^T(t)X(t)(x(t) - x^*(t)) \\ &\qquad + B(t)R^{-1}(t)B^T(t)X(t)(x(t) - \hat{x}(t)) + f(x) - f(x^*) + F(t)v(t). \end{aligned}$$

If we use the function $r(x, x^*)$ defined in (7) above, then it holds that

$$\begin{aligned} &\frac{d}{dt}(x(t) - x^*(t)) \\ &\quad = A(t)(x(t) - x^*(t)) - B(t)R^{-1}(t)B^T(t)X(t)(x(t) - x^*(t)) \\ &\qquad + B(t)R^{-1}(t)B^T(t)X(t)(x(t) - \hat{x}(t)) + F(t)v(t) + r(x, x^*). \end{aligned} \tag{8}$$

Since

$$\begin{aligned} \frac{d}{dt}N(x - x^*) &= \left[\frac{d}{dt}(x(t) - x^*(t))^T\right] X(t)(x(t) - x^*(t)) \\ &\quad + \frac{1}{2}(x(t) - x^*(t))^T \dot{X}(t)(x(t) - x^*(t)), \end{aligned}$$

we obtain with (8) and $z(t) = x(t) - x^*(t)$

$$\begin{aligned} \frac{d}{dt}N(z) &= \Big\langle A(t)z(t) - B(t)R^{-1}(t)B^T(t)X(t)z(t) \\ &\qquad + B(t)R^{-1}(t)B^T(t)X(t)(x(t) - \hat{x}(t)) \\ &\qquad + F(t)v(t) + r(x, x^*), X(t)z(t)\Big\rangle + \frac{1}{2}z^T(t)\dot{X}(t)z(t). \end{aligned} \tag{9}$$

Now we analyze the first two terms of the right-hand side in (9) and transform them by using the DRE (4).

$$
\begin{aligned}
&\langle (A(t) - B(t)R^{-1}(t)B^T(t)X(t))z(t), X(t)z(t)\rangle \\
&\quad = \frac{1}{2}z^T(t)(A^T(t)X(t) + X(t)A(t))z(t) - z^T(t)X(t)B(t)R^{-1}(t)B^T(t)X(t)z(t) \\
&\quad = \frac{1}{2}z^T(t)(A^T(t)X(t) + X(t)A(t) - X(t)B(t)R^{-1}(t)B^T(t)X(t))z(t) \\
&\qquad\qquad -\frac{1}{2}z^T(t)X(t)B(t)R^{-1}(t)B^T(t)X(t)z(t) \\
&\quad = \frac{1}{2}z^T(t)(-C^T(t)Q(t)C(t) - \dot{X}(t))z(t) \\
&\qquad\qquad -\frac{1}{2}z^T(t)X(t)B(t)R^{-1}(t)B^T(t)X(t)z(t) \\
&\quad = -\frac{1}{2}\Big[\left\langle C^T(t)Q(t)C(t)z(t), z(t)\right\rangle + \left\langle \dot{X}(t)z(t), z(t)\right\rangle \\
&\qquad\qquad + \left\langle X(t)B(t)R^{-1}(t)B^T(t)X(t)z(t), z(t)\right\rangle\Big]. \qquad (10)
\end{aligned}
$$

If we insert the result (10) into (9), the term $z^T(t)\dot{X}(t)z(t)$ is cancelled and we obtain

$$
\begin{aligned}
\frac{d}{dt}N(z) \;=\; -\frac{1}{2}\Big[&\left\langle X(t)B(t)R^{-1}(t)B^T(t)X(t)z(t), z(t)\right\rangle \\
&+ \left\langle C^T(t)Q(t)C(t)z(t), z(t)\right\rangle\Big] \\
&+ \langle F(t)v(t) + r(x, x^*), X(t)z(t)\rangle \\
&+ \left\langle B(t)R^{-1}(t)B^T(t)X(t)(x(t) - \hat{x}(t)), X(t)z(t)\right\rangle,
\end{aligned}
$$

which is the assertion for $z(t) = x(t) - x^*(t)$. □

Proposition 3.3 *Let $\tilde{N}(z) = \frac{1}{2}z^T(t)\Sigma^{-1}(t)z(t)$. For $t \in [0, T_p]$ we have*

$$
\begin{aligned}
\frac{d}{dt}\tilde{N}(x - \hat{x}) = -\frac{1}{2}\Big[&\left\langle W^{-1}C(t)(x(t) - \hat{x}(t)), C(t)(x(t) - \hat{x}(t))\right\rangle \\
&+ \left\langle VF^T(t)\Sigma^{-1}(t)(x(t) - \hat{x}(t)), F^T(t)\Sigma^{-1}(t)(x(t) - \hat{x}(t))\right\rangle\Big] \\
&+ \left\langle F(t)v(t) + r(x, x^*), \Sigma^{-1}(t)(x(t) - \hat{x}(t))\right\rangle \\
&- \left\langle w(t), W^{-1}C(t)(x(t) - \hat{x}(t))\right\rangle,
\end{aligned}
$$

where

$$r(x, x^*) = f(x) - f(x^*) - A(t)(x - x^*).$$

Theorem 3.8 *Assume*

- (A7) $E(0) \le \frac{\delta}{2}$;
- (A8) $(A1), (A2), (A4)$ *hold uniformly on all horizons* $[t_i, t_{i+1}]$;
- (A9) $\exists c_X, c_\Sigma$ *with* $1 < c_X \ll 2$ *and* $1 < c_\Sigma \ll 2$ *such that* $\|x(t_i)\|^2_{X_i(t_i)} \le c_X\|x(t_i)\|^2_{X_{i-1}(t_i)}$ *and* $\|x(t_i)\|^2_{\Sigma_i^{-1}(t_i)} \le c_\Sigma\|x(t_i)\|^2_{\Sigma_{i-1}^{-1}(t_i)}$ $\forall i > 0$ *with* $X_i(t_i), \Sigma_i(t_i)$ *being the solutions of the DREs on* $[t_i, t_{i+1}]$ *at* t_i;
- (A10) $\int_0^t e^{-\tilde{\omega}(t-s)}\gamma(s + t_i)\, ds < \frac{\delta}{2} \quad \forall t \in [0, T_c]$ *and* $\forall i = 0, 1, \ldots$;
- (A11) $\int_0^{T_c} e^{-\tilde{\omega}(T_c-s)}\gamma(s + t_i)\, ds \le \begin{cases} \frac{\delta}{2}\left(1 - e^{-\tilde{\omega}T_c}\right) & i = 0, \\ \frac{\delta}{2}\left(1 - ce^{-\tilde{\omega}T_c}\right) & i = 1, 2, \ldots \end{cases}$
  with $c = \max\{\sqrt{c_X}, \sqrt{c_\Sigma}\}$ *and* T_c *such that* $e^{-\tilde{\omega}T_c} < \frac{1}{c}$.

Then

$$E(t) \le e^{-\tilde{\omega}(t-t_i)}E(t_i) + \int_0^{t-t_i} e^{-\tilde{\omega}(t-t_i-s)}\gamma(t_i + s)\, ds \quad \forall t \in [t_i, t_i + T_p] \text{ and } \forall i = 0, 1, .. \qquad (21)$$

In particular this implies

$$E(t) \le c\frac{\delta}{2}e^{-\tilde{\omega}t} + \int_0^t e^{-\tilde{\omega}(t-s)}\gamma(s)\, ds \quad \forall t > 0. \qquad (22)$$

Remark 3.9 *Discussion of the assumptions in Theorem 3.8:*

Assumptions (A8) and (A9) *Assumption (A9) is needed to treat the discontinuity points discussed in Remark 3.7. Therefor we have to require that*

$$\|X_{i+1}(t_i) - X_i(t_i)\| \le \varepsilon_X \text{ and } \|\Sigma_{i+1}^{-1}(t_i) - \Sigma_i^{-1}(t_i)\| \le \varepsilon_\Sigma,$$

which are quite natural conditions, since the time-varying system matrices coincide at the time-points t_i.

Assumption (A1) from Theorem 3.5 has to hold uniformly on all horizons $[t_i, t_{i+1}]$. We discussed this in Remark 3.6 for the interval $[0, T_p]$. Here, it means that the solutions of all DREs and FDREs on every interval $[t_i, t_{i+1}]$ have to be uniformly bounded by $\alpha_i > 0$, $\beta_i > 0$, $i = 1, 2$. Again this means that the initial and terminal conditions satisfy $\alpha_1 I < X(t_i + T_p) < \alpha_2 I$ and

Assumption (A6) and continuity of $t \to \int_0^t e^{-\tilde{\omega}(t-s)}\gamma(s)\,ds$ implies

$$\exists \alpha \in (0,1) : \int_0^t e^{-\tilde{\omega}(t-s)}\gamma(s)\,ds \leq \frac{\alpha\delta}{2} \quad \forall t \in [0, T_p]$$

and with $0 < e^{-\tilde{\omega}t} < 1$, $\tilde{\omega}, t > 0$ it follows that

$$E(t) \leq \frac{\delta}{2} + \frac{\alpha\delta}{2} = \frac{\delta}{2}(1+\alpha) \quad \forall t \in [0, \tau]. \tag{20}$$

Step 4: So far we have proved the inequalities for the interval $[0, \tau]$. Now we need to extend this to the interval $[0, T_p]$.

If (20) holds on $[0, T_p]$ then (19) must hold on $[0, T_p]$ as we have seen in the proof. So we have to show that (20) holds on $[0, T_p]$.

Assume that (20) is not valid on $[0, T_p]$. Let $\tilde{\tau}$ denote the smallest value in $[0, T_p]$ such that $E(\tilde{\tau}) = \frac{\delta}{2}(1+\alpha)$. This implies that $\exists \varepsilon > 0$ so that (19) holds on $[0, \tilde{\tau}+\varepsilon]$. So it follows that (20) holds on $[0, \tilde{\tau}+\varepsilon]$ and with this $E(\tilde{\tau}+\varepsilon) < \frac{\delta}{2}(1+\alpha)$. But this contradicts $E(\tilde{\tau}) = \frac{\delta}{2} + \frac{\alpha\delta}{2}$. So (20) holds on $[0, T_p]$. $\square$.

It is obvious that if (20) holds on $[0, T_p]$, it also holds on $[0, T_c]$ with $T_c \leq T_p$.

Now we want to expand these results to the interval $[0, T_f]$.

Remark 3.7 *The following theorem is the time-varying generalization of Theorem 3.1 in [8]. Since we believe that there is a gap in their theorem we suggest a corrected version. The gap in their theorem occurs in the transition from one prediction interval to the next. There they assume that from*

$$E(t_i)|_{[t_{i-1}, t_i]} \leq \frac{\delta}{2},$$

it follows that

$$E(t_i)|_{[t_i, t_{i+1}]} \leq \frac{\delta}{2},$$

which does not have to be the case since we take different operating points for the LTI case and consequently different ARE solutions as a basis on different intervals [5]. As a consequence we have discontinuity points at the interval boundaries. The aim is to get a hold on these discontinuity points. The same has to be regarded for the LTV case. But the advantage is that we use the time-varying reference itself instead operating points.

Proof: Consider

$$\begin{aligned}
&\frac{d}{dt}(x(t) - \hat{x}(t)) \\
&\quad = \big[f(x) + B(t)u(t) + F(t)v(t)\big] - \big[A(t)(\hat{x}(t) - x^*(t)) + f(x^*) + Bu(t) \\
&\qquad\quad + \Sigma(t)C^T(t)W^{-1}(y(t) - C(t)\hat{x}(t))\big] \\
&\quad = -A(t)(\hat{x}(t) - x^*(t)) + A(t)(x(t) - x(t)) + f(x) - f(x^*) \\
&\qquad\quad + F(t)v(t) - \Sigma(t)C^T(t)W^{-1}\big(C(t)x(t) + w(t) - C(t)\hat{x}(t)\big) \\
&\quad = A(t)(x(t) - \hat{x}(t)) + F(t)v(t) - \Sigma(t)C^T(t)W^{-1}\big(C(t)x(t) + w(t) \\
&\qquad\quad - C(t)\hat{x}(t)\big) + r(x, x^*)
\end{aligned} \tag{11}$$

and

$$\begin{aligned}
\frac{d}{dt}\tilde{N}(x - \hat{x}) &= \left[\frac{d}{dt}(x(t) - \hat{x}(t))^T\right]\Sigma^{-1}(t)(x(t) - \hat{x}(t)) \\
&\quad + \frac{1}{2}(x(t) - \hat{x}(t))^T\left(\frac{d}{dt}\Sigma^{-1}(t)\right)(x(t) - \hat{x}(t)) \\
&= \left[\frac{d}{dt}(x(t) - \hat{x}(t))^T\right]\Sigma^{-1}(t)(x(t) - \hat{x}(t)) \\
&\quad - \frac{1}{2}(x(t) - \hat{x}(t))^T\Sigma^{-1}(t)\dot{\Sigma}(t)\Sigma^{-1}(t)(x(t) - \hat{x}(t)),
\end{aligned} \tag{12}$$

since from $\Sigma(t)\Sigma^{-1}(t) = I$ it follows $\dot{\Sigma}(t)\Sigma^{-1}(t) + \Sigma(t)\frac{d}{dt}\Sigma^{-1}(t) = 0$ which implies $\frac{d}{dt}\Sigma^{-1}(t) = -\Sigma^{-1}(t)\dot{\Sigma}(t)\Sigma^{-1}(t)$. Equation (12) together with (11) and $z(t) = x(t) - \hat{x}(t)$ implies

$$\begin{aligned}
\frac{d}{dt}\tilde{N}(z) &= \big\langle A(t)z(t) + F(t)v(t) - \Sigma(t)C^T(t)W^{-1}\big(C(t)x(t) + w(t) - C(t)\hat{x}(t)\big) \\
&\quad + r(x, x^*), \Sigma^{-1}(t)z(t)\big\rangle - \frac{1}{2}\big\langle z(t), \Sigma^{-1}(t)\dot{\Sigma}(t)\Sigma^{-1}(t)z(t)\big\rangle.
\end{aligned} \tag{13}$$

Now we analyze the term

$$\big\langle A(t) - \Sigma(t)C^T(t)W^{-1}C(t)z(t), \Sigma^{-1}(t)z(t)\big\rangle$$

by using the FDRE (5). So we obtain

$$\begin{aligned}
&\big\langle A(t)-\Sigma(t)C^T(t)W^{-1}C(t)z(t),\Sigma^{-1}(t)z(t)\big\rangle\\
&\quad= z^T(t)\big(A^T(t)-C^T(t)W^{-1}C(t)\Sigma(t)\big)\Sigma^{-1}(t)z(t)\\
&\quad= z^T(t)\Sigma^{-1}(t)\big(\Sigma(t)A^T(t)-\Sigma(t)C^T(t)W^{-1}C(t)\Sigma(t)\big)\Sigma^{-1}(t)z(t)\\
&\quad= \frac{1}{2}z^T(t)\Sigma^{-1}(t)\big(\Sigma(t)A^T(t)+A(t)\Sigma(t)\\
&\qquad -\Sigma(t)C^T(t)W^{-1}C(t)\Sigma(t)\big)\Sigma^{-1}(t)z(t)-\frac{1}{2}z^T(t)C^T(t)W^{-1}C(t)z(t)\\
&\quad= \frac{1}{2}z^T(t)\Sigma^{-1}(t)\big(-F(t)VF^T(t)+\dot\Sigma(t)\big)\Sigma^{-1}(t)z(t)\\
&\qquad -\frac{1}{2}z^T(t)C^T(t)W^{-1}C(t)z(t)\\
&\quad= -\frac{1}{2}\Big[\big\langle VF^T(t)\Sigma^{-1}(t)z(t),F^T(t)\Sigma^{-1}(t)z(t)\big\rangle\\
&\qquad -\big\langle\Sigma^{-1}(t)z(t),\dot\Sigma(t)\Sigma^{-1}(t)z(t)\big\rangle+\big\langle W^{-1}C(t)z(t),C(t)z(t)\big\rangle\Big]. \quad (14)
\end{aligned}$$

If we insert (14) into (13), the term $\langle\Sigma^{-1}(t)z(t),\dot\Sigma(t)\Sigma^{-1}(t)z(t)\rangle$ is cancelled and we obtain

$$\begin{aligned}
\frac{d}{dt}\tilde N(z) = &-\frac{1}{2}\Big[\big\langle VF^T(t)\Sigma^{-1}(t)z(t),F^T(t)\Sigma^{-1}(t)z(t)\big\rangle+\big\langle W^{-1}C(t)z(t),C(t)z(t)\big\rangle\Big]\\
&+\big\langle F(t)v(t)+r(x,x^*),\Sigma^{-1}(t)z(t)\big\rangle-\langle w(t),W^{-1}C(t)z(t)\rangle,
\end{aligned}$$

which is the assertion for $z(t)=x(t)-\hat x(t)$. □

Remark 3.4 *Using the notation from Proposition 3.2 and 3.3, (6) can be written as*

$$E(t)=\Big[N(x(t)-x^*(t))+\tilde N(x(t)-\hat x(t))\Big]^{\frac{1}{2}}.$$

Note that with the definition of $E(t)$ in [8], one would have

$$E(t)=\Big(2N(x(t)-x^*(t))\Big)^{\frac{1}{2}}+\Big(2\tilde N(x(t)-\hat x(t))\Big)^{\frac{1}{2}}.$$

Now we can formulate the main theorem for the interval $[0,T_p]$, which generalizes Theorem 2.1 in [8].

Theorem 3.5 *If the following assumptions are fulfilled:*

(A1) $\alpha_1\mathrm{I}\le X(t)\le\alpha_2\mathrm{I},\ \beta_1\mathrm{I}\le\Sigma^{-1}(t)\le\beta_2\mathrm{I},\ \alpha_1,\alpha_2,\beta_1,\beta_2>0,\forall t\in[0,T_p]$;

(A2) $\frac{1}{2}\Big[\big\langle R^{-1}(t)B^T(t)X(t)x(t),B^T(t)X(t)x(t)\big\rangle+\big\langle C^T(t)Q(t)C(t)x(t),x(t)\big\rangle$

$$\begin{aligned}
&= -\tilde\omega\left(N(x-x^*)+\tilde N(x-\hat x)\right)+\frac{1}{2}\big\langle F(t)v(t),X(t)(x(t)-x^*(t))\\
&\qquad +\Sigma^{-1}(t)(x(t)-\hat x(t))\big\rangle-\frac{1}{2}\big\langle w(t),W^{-1}C(t)(x(t)-\hat x(t))\big\rangle.
\end{aligned}$$

Consider now only the noise terms:

$$\begin{aligned}
&\langle F(t)v(t),X(t)(x(t)-x^*(t))\rangle+\big\langle F(t)v(t),\Sigma^{-1}(t)(x(t)-\hat x(t))\big\rangle\\
&\qquad -\big\langle w(t),W^{-1}C(t)(x(t)-\hat x(t))\big\rangle\\
&\le \sqrt{\alpha_2}\|F(t)\|\|v(t)\|Y(t)+\sqrt{\beta_2}\|F(t)\|\|v(t)\|Z(t)\\
&\qquad +\frac{1}{\sqrt{\beta_1}}\|C(t)\|\|W^{-1}\|\|w(t)\|Z(t)\\
&= \sqrt{\alpha_2}\|F(t)\|\|v(t)\|\Big(2N(x-x^*)\Big)^{\frac{1}{2}}\\
&\qquad +\left(\sqrt{\beta_2}\|F(t)\|\|v(t)\|+\frac{1}{\sqrt{\beta_1}}\|C(t)\|\|W^{-1}\|\|w(t)\|\right)\Big(2\tilde N(x-\hat x)\Big)^{\frac{1}{2}}\\
&\le \Big(\sqrt{\alpha_2+\beta_2}\|F(t)\|\|v(t)\|\\
&\qquad +\frac{1}{\sqrt{\beta_1}}\|C(t)\|\|W^{-1}\|\|w(t)\|\Big)\sqrt{2}\left(N(x-x^*)+\tilde N(x-\hat x)\right)^{\frac{1}{2}}\\
&= 2\gamma(t)\left(N(x-x^*)+\tilde N(x-\hat x)\right)^{\frac{1}{2}}.
\end{aligned}$$

So we obtain

$$E(t)\frac{d}{dt}E(t)\le-\tilde\omega\left(N(x-x^*)+\tilde N(x-\hat x)\right)+\gamma(t)\left(N(x-x^*)+\tilde N(x-\hat x)\right)^{\frac{1}{2}}.$$

This implies

$$\begin{aligned}
\frac{d}{dt}E(t) \;\le\; & E^{-1}(t)\Bigg[-\tilde\omega\underbrace{\left(N(x-x^*)+\tilde N(x-\hat x)\right)}_{=E(t)^2}\\
&\qquad +\gamma(t)\underbrace{\left(N(x-x^*)+\tilde N(x-\hat x)\right)^{\frac{1}{2}}}_{=E(t)}\Bigg]\\
=\; & -\tilde\omega E(t)+\gamma(t).
\end{aligned}$$

This differential inequality can be solved by variation of constants. The solution is

$$E(t)\le e^{-\tilde\omega t}E(0)+\int_0^t e^{-\tilde\omega(t-s)}\gamma(s)\,ds\quad\forall t\in[0,\tau]. \qquad (19)$$

we obtain

$$E(t)\frac{d}{dt}E(t) = \frac{1}{2}\Big\{h(x(t),x^*(t),\hat{x}(t)) + \big\langle F(t)v(t)+r(x(t),x^*(t)), X(t)(x(t)-x^*(t))\big\rangle + \big\langle F(t)v(t)+r(x(t),x^*(t)), \Sigma^{-1}(t)(x(t)-\hat{x}(t))\big\rangle - \big\langle w(t), W^{-1}C(t)(x(t)-\hat{x}(t))\big\rangle \Big\}.$$

By using assumption (A2) and multiplying $h(x(t),x^*(t),\hat{x}(t))$ with -1 this can be bounded by

$$h(x(t),x^*(t),\hat{x}(t)) \le -\omega\left(N(x(t)-x^*(t))+\tilde{N}(x(t)-\hat{x}(t))\right).$$

It follows that

$$E(t)\frac{d}{dt}E(t) \le \frac{1}{2}\Big[-\omega\left(N(x(t)-x^*(t))+\tilde{N}(x(t)-\hat{x}(t))\right) + \big\langle F(t)v(t)+r(x(t),x^*(t)), X(t)(x(t)-x^*(t))+\Sigma^{-1}(t)(x(t)-\hat{x}(t))\big\rangle - \big\langle w(t), W^{-1}C(t)(x(t)-\hat{x}(t))\big\rangle\Big].$$

In (18) we estimate the term

$$\big\langle r(x(t),x^*(t)), X(t)(x(t)-x^*(t))+\Sigma^{-1}(t)(x(t)-\hat{x}(t))\big\rangle.$$

By using assumption (A5) we obtain

$$\big\langle r(x(t),x^*(t)), X(t)(x(t)-x^*(t))+\Sigma^{-1}(t)(x(t)-\hat{x}(t))\big\rangle \le \left(\frac{\omega}{2}-\tilde{\omega}\right)(Y^2(t)+Z^2(t)).$$

This implies

$$\begin{aligned} E(t)\frac{d}{dt}E(t) &\le -\frac{\omega}{2}\left(N(x-x^*)+\tilde{N}(x-\hat{x})\right)+\frac{1}{2}\left(\frac{\omega}{2}-\tilde{\omega}\right)(Y^2(t)+Z^2(t)) \\ &\quad +\frac{1}{2}\left\langle F(t)v(t), X(t)(x(t)-x^*(t))+\Sigma^{-1}(t)(x(t)-\hat{x}(t))\right\rangle \\ &\quad -\frac{1}{2}\left\langle w(t), W^{-1}C(t)(x(t)-\hat{x}(t))\right\rangle \\ &= -\frac{\omega}{2}\left(N(x-x^*)+\tilde{N}(x-\hat{x})\right)+\left(\frac{\omega}{2}-\tilde{\omega}\right)\left(N(x-x^*)+\tilde{N}(x-\hat{x})\right) \\ &\quad +\frac{1}{2}\left\langle F(t)v(t), X(t)(x(t)-x^*(t))+\Sigma^{-1}(t)(x(t)-\hat{x}(t))\right\rangle \\ &\quad -\frac{1}{2}\left\langle w(t), W^{-1}C(t)(x(t)-\hat{x}(t))\right\rangle \end{aligned}$$

$$+\big\langle W^{-1}C(t)z(t), C(t)z(t)\big\rangle + \big\langle VF^T(t)\Sigma^{-1}(t)z(t), F^T(t)\Sigma^{-1}(t)z(t)\big\rangle\Big] - \big\langle R^{-1}(t)B^T(t)X(t)\Sigma(t)\Sigma^{-1}(t)z(t), B^T(t)X(t)x(t)\big\rangle \ge \omega\left(N(x)+\tilde{N}(z)\right) \quad \forall x,z\in\mathbb{R}^n;$$

(A3) $E(0) \le \frac{\delta}{2}$ *on* $[0,T_p]$ *for some* $\delta > 0$;

(A4) $\exists\ L = constant: \|r(x,x^*)\| \le L(x-x^*)^2$
$\Big[$*note:* $r(x,x^*) = f(x)-f(x^*)-A(t)(x-x^*)\Big]$;

(A5) $\tilde{\omega} = \frac{\omega}{2} - L\delta\frac{\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}}{2\alpha_1} > 0$;

(A6) $\int_0^t e^{-\tilde{\omega}(t-s)}\gamma(s)\,ds < \frac{\delta}{2} \qquad \forall t\in[0,T_p]$, *where*
$\gamma(t) = \frac{\sqrt{2}}{2}\sqrt{\alpha_2+\beta_2}\|F(t)\|\,\|v(t)\| + \frac{\sqrt{2}}{2\beta_1}\|C(t)\|\,\|w(t)\|\,\|W^{-1}\|$.

Then

$$E(t) \le e^{-\tilde{\omega}t}E(0) + \int_0^t e^{-\tilde{\omega}(t-s)}\gamma(s)\,ds \tag{15}$$

holds $\forall t\in[0,T_p]$.

Remark 3.6 *Discussion of the assumptions in Theorem 3.5 in comparison to the LTI case in [8]:*

Assumption (A1) ((2.10) in [8]). *Now we have to ensure that the solutions of the DREs are uniformly bounded by* $\alpha_i > 0$, $\beta_i > 0$, $i = 1,2$ *on* $[0,T_p]$. *Especially that means that the initial and terminal conditions satisfy* $X(T_p) = G$, *with* $\alpha_1 I < G < \alpha_2 I$, *and* $\Sigma(0) = \Sigma_0$, *with* $\frac{1}{\beta_2}I < \Sigma_0 < \frac{1}{\beta_1}I$. *Moreover, the intervals need to be chosen small enough (or* α_1,β_1 *small and* α_2,β_2 *large enough) so that the Riccati solutions remain in the prescribed bounds.*

Assumption (A2) ((2.11) in [8]). *This inequality can be assumed as in the LTI case but with time dependent matrices and inserted covariance matrices* V *and* W *and noise input matrix* $F(t)$.

Assumptions (A3) and (A4) ((2.12) and (2.13) in [8]). *These assumptions can be simplified since we do not need to consider a stationary operating point.*

Assumption (A5) ((2.14) in [8]). *We have to adapt* $\tilde{\omega}$ *to our (corrected) version of the performance and for the LTV case this parameter has to be modified.*

Assumption (A6) ((2.15) in [8]). *We insert the covariance matrices* V *and* W *and the noise input matrix* $F(t)$ *and we need the factor* $\frac{\sqrt{2}}{2}$ *for the corrected version of the performance.*

Proof of Theorem 3.5:
Step 1: Assumption (A3) implies the existence of $\tau > 0$ with $E(\tau) \leq \delta$. Also,

$$\sqrt{\alpha_1}\|x(t) - x^*(t)\| \leq E(t), \text{ implies } \|x(t) - x^*(t)\| \leq \frac{\delta}{\sqrt{\alpha_1}} \text{ on } [0,\tau]. \tag{16}$$

Define

$$\begin{aligned} Y(t) &= \sqrt{\langle x(t) - x^*(t), X(t)(x(t) - x^*(t))\rangle}, \\ Z(t) &= \sqrt{\langle x(t) - \hat{x}(t), \Sigma^{-1}(t)(x(t) - \hat{x}(t))\rangle}. \end{aligned}$$

Step 2: If we consider (A3), (A4) we obtain

$$\begin{aligned} &\langle r(x(t), x^*(t)), X(t)(x(t) - x^*(t))\rangle + \big\langle r(x(t), x^*(t)), \Sigma^{-1}(t)(x(t) - \hat{x}(t))\big\rangle \\ &\qquad \leq L\|x(t) - x^*(t)\| \left(Y^2(t) + \frac{\sqrt{\beta_2}}{\sqrt{\alpha_1}} Y(t)Z(t)\right), \end{aligned}$$

since $\langle x(t) - x^*(t), X(t)(x(t) - x^*(t))\rangle = Y^2(t)$ and

$$\begin{aligned} &\big\langle x(t) - x^*(t), \Sigma^{-1}(t)(x(t) - \hat{x}(t))\big\rangle \\ &\qquad = \left\langle \Sigma^{-\frac{1}{2}}(t)X^{-\frac{1}{2}}(t)X^{\frac{1}{2}}(t)(x(t) - x^*(t)), \Sigma^{-\frac{1}{2}}(t)(x(t) - \hat{x}(t))\right\rangle \\ &\qquad \leq \sqrt{\beta_2}\frac{1}{\sqrt{\alpha_1}} Y(t)Z(t). \end{aligned}$$

Factoring out $\frac{1}{\sqrt{\alpha_1}}$, employing $\sqrt{\alpha_1} < \sqrt{\alpha_2}$ and (16) yields

$$\begin{aligned} &\langle r(x(t), x^*(t)), X(t)(x(t) - x^*(t))\rangle + \big\langle r(x(t), x^*(t)), \Sigma^{-1}(t)(x(t) - \hat{x}(t))\big\rangle \\ &\qquad \leq L\frac{\delta}{\alpha_1}\left(\sqrt{\alpha_2}Y^2(t) + \sqrt{\beta_2}Y(t)Z(t)\right). \end{aligned} \tag{17}$$

Now we use the Cauchy inequality with epsilon

$$ab = \left(\sqrt{2\varepsilon}a\right)\left(\frac{b}{\sqrt{2\varepsilon}}\right) \leq \frac{(\sqrt{2\varepsilon}a)^2 + \left(\frac{b}{\sqrt{2\varepsilon}}\right)^2}{2} = \varepsilon a^2 + \frac{b^2}{4\varepsilon},$$

with $\varepsilon = \frac{\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}}{2}$ for the second term $\sqrt{\beta_2}Y(t)Z(t)$ and we obtain

$$\begin{aligned} &\langle r(x(t), x^*(t)), X(t)(x(t) - x^*(t))\rangle + \big\langle r(x(t), x^*(t)), \Sigma^{-1}(t)(x(t) - \hat{x}(t))\big\rangle \\ &\leq L\frac{\delta}{\alpha_1}\left(\sqrt{\alpha_2}Y^2(t) + \sqrt{\beta_2}Y(t)Z(t)\right) \\ &\leq L\frac{\delta}{\alpha_1}\left(\sqrt{\alpha_2}Y^2(t) + \sqrt{\beta_2}Y(t)Z(t)\right) \\ &\leq L\frac{\delta}{\alpha_1}\left(\sqrt{\alpha_2}Y^2(t) + \frac{\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}}{2}Z^2(t) + \frac{\beta_2}{2\left(\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}\right)}Y^2(t)\right) \\ &= L\frac{\delta}{\alpha_1}\left(\sqrt{\alpha_2}Y^2(t) + \frac{\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}}{2}Z^2(t) + \frac{\beta_2+\alpha_2-\alpha_2}{2\left(\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}\right)}Y^2(t)\right) \\ &= L\frac{\delta}{\alpha_1}\left(\sqrt{\alpha_2}Y^2(t) + \frac{\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}}{2}Z^2(t)\right. \\ &\qquad\qquad \left. + \frac{\left(\sqrt{\beta_2+\alpha_2}-\sqrt{\alpha_2}\right)\left(\sqrt{\beta_2+\alpha_2}+\sqrt{\alpha_2}\right)}{2\left(\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}\right)}Y^2(t)\right) \\ &= L\frac{\delta}{\alpha_1}\left(\sqrt{\alpha_2}Y^2(t) + \frac{\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}}{2}Z^2(t) + \frac{\sqrt{\beta_2+\alpha_2}-\sqrt{\alpha_2}}{2}Y^2(t)\right) \\ &= L\frac{\delta}{\alpha_1}\left(\frac{\sqrt{\alpha_2}+\sqrt{\alpha_2+\beta_2}}{2}(Y^2(t)+Z^2(t))\right) \qquad \text{for } t \in [0,\tau]. \end{aligned} \tag{18}$$

Step 3: Consider

$$E(t)\frac{d}{dt}E(t) = \frac{1}{2}\frac{d}{dt}\left(E(t)^2\right) = \frac{1}{2}\frac{d}{dt}\left(N(x(t) - x^*(t)) + \tilde{N}(x(t) - \hat{x}(t))\right).$$

(Note again that in [8] the factor $\frac{1}{2}$ is missing which is once more in conflict with their definition of the performance index.)

If we apply Propositions 3.2 and 3.3, rearrange terms and substitute some of the terms by

$$\begin{aligned} &h(x(t), x^*(t), \hat{x}(t)) \\ &\quad = -\frac{1}{2}\Big[\big\langle X(t)B(t)R^{-1}(t)B^T(t)X(t)(x(t) - x^*(t)), x(t) - x^*(t)\big\rangle \\ &\qquad + \big\langle C^T(t)Q(t)C(t)(x(t) - x^*(t)), x(t) - x^*(t)\big\rangle\Big] \\ &\qquad + \big\langle B(t)R^{-1}(t)B^T(t)X(t)(x(t) - \hat{x}(t)), X(t)(x(t) - x^*(t))\big\rangle \\ &\qquad - \frac{1}{2}\Big[\big\langle W^{-1}C(t)(x(t) - \hat{x}(t)), C(t)(x(t) - \hat{x}(t))\big\rangle \\ &\qquad + \big\langle VF^T(t)\Sigma^{-1}(t)(x(t) - \hat{x}(t)), F^T(t)\Sigma^{-1}(t)(x(t) - \hat{x}(t))\big\rangle\Big], \end{aligned}$$