Probabilistic ODE Solvers for Integration Error-Aware Numerical Optimal Control

Amon Lahr AMLAHR@ETHZ.CH
ETH Zürich

Filip Tronarp FILIP.TRONARP@MATSTAT.LU.SE
Lund University

Nathanael Bosch NATHANAEL.BOSCH@UNI-TUEBINGEN.DE
University of Tübingen, Tübingen AI Center

Jonathan Schmidt JONATHAN.SCHMIDT@UNI-TUEBINGEN.DE
University of Tübingen, Tübingen AI Center

Philipp Hennig PHILIPP.HENNIG@UNI-TUEBINGEN.DE
University of Tübingen, Tübingen AI Center

Melanie N. Zeilinger MZEILINGER@ETHZ.CH
ETH Zürich

Editors: A. Abate, K. Margellos, A. Papachristodoulou

Abstract

Appropriate time discretization is crucial for real-time applications of numerical optimal control, such as nonlinear model predictive control. However, if the discretization error strongly depends on the applied control input, meeting accuracy and sampling time requirements simultaneously can be challenging using classical discretization methods. In particular, neither fixed-grid nor adaptive-grid discretizations may be suitable, when they suffer from large integration error or exceed the prescribed sampling time, respectively. In this work, we take a first step at closing this gap by utilizing probabilistic numerical integrators to approximate the solution of the initial value problem, as well as the computational uncertainty associated with it, inside the optimal control problem (OCP). By taking the viewpoint of probabilistic numerics and propagating the numerical uncertainty in the cost, the OCP is reformulated such that the optimal input reduces the computational uncertainty insofar as it is beneficial for the control objective. The proposed approach is illustrated using a numerical example, and potential benefits and limitations are discussed.

Keywords: numerical integration, nonlinear model predictive control, probabilistic numerics

1. Introduction

For direct optimal control, numerical integration methods are a crucial component for converting a continuous-time system model of ordinary differential equations (ODEs) into accurate discrete-time predictions based on which the input signal can be optimized. While negligible discretization error can generally be achieved by choosing a sufficiently fine time discretization, it can be computationally infeasible in the model predictive control (MPC) setting, where an optimal control problem (OCP) is to be solved online, i.e., in receding horizon fashion and at a fixed sampling interval. This is complicated by the fact that the accumulated discretization error at each time step

may depend on the chosen input signal, directly due to the control input affecting the vector field, or indirectly due to changing stiffness properties between different local solutions of the resulting state trajectory.

For such challenging scenarios, standard grid discretization approaches may fail to achieve negligible integration error given a limited computational budget. First, a uniform time discretization may lead to either unacceptable integration error for a coarse grid or excessive computational requirements for a fine grid. Second, a fixed, non-uniform time discretization (Qin and Badgwell, 1997; Tøndel and Johansen, 2002; Cagienard et al., 2007; Shekhar and Manzie, 2015; Yu and Biegler, 2016; Chen et al., 2020) may reduce discretization error without adding computational overhead (Quirynen et al., 2015). The offline design of fixed, non-uniform grids has been considered by Lazutkin et al. (2018) and Wolff et al. (2022). Lishkova et al. (2022) consider multi-rate time discretization for dynamical systems with varying time scales within the MPC. Still, the discretization accuracy of these approaches may vary as they do not consider the dependence of the integration error on the applied input. Third, adaptive mesh refinement methods can achieve a pre-specified integration accuracy by choosing smaller step sizes (Betts and Huffman, 1998; Binder et al., 2000; Schlegel et al., 2005), selecting higher-order local interpolants (Babuška and Suri, 1990; Darby et al., 2011), or relocating grid points (Tanartkit and Biegler, 1997; Zhao and Tsiotras, 2011). As refinement criteria, besides minimizing estimates of the integration error, the adjoint sensitivities can also be used to guide the refinement process towards lower open-loop costs (Tanartkit and Biegler, 1997; Paiva and Fontes, 2015). The former has been applied to MPC by, e.g., Gao et al. (2023), the latter, by Lee et al. (2018) and Paiva and Fontes (2018). Tailored to the receding-horizon strategy, Paiva and Fontes (2017) and Potena et al. (2018) propose to save computation time by enforcing a higher discretization accuracy only for the initial part of the time horizon. Nevertheless, due to the variable number of integration steps, the applicability of mesh refinement methods for online applications with a fixed sampling time can be complicated by their unpredictable solve times. Alternative to rendering the discretization error negligibly small by a suitable discretization strategy, Kögel and Findeisen (2015) propose to robustify the prediction against the worst-case integration error by employing a robust MPC strategy based on a Lipschitz bound of the dynamics subject to all possible input values; however, this might lead to unnecessarily conservative predictions.

Similarly to Kögel and Findeisen (2015), rather than accepting a possibly large integration error depending on the optimal input trajectory, or ensuring a specific error tolerance at the cost of unpredictable solve times, in this work we modify the input in response to non-negligible integration errors. Instead of a robust treatment, we propose to utilize probabilistic ODE solvers to predict the integration error as a function of the applied input. Encoding the trade-off between accurately solving the ODE and minimizing the control objective in a stochastic OCP formulation, uncertainty reduction can thus be actively targeted by the solver insofar as it is beneficial for the control objective. Viewing the proposed approach through the lens of probabilistic numerics thereby naturally exposes the implicit assumption of zero discretization error for classical direct transcription methods, as well as provides a dual-control interpretation of the proposed approach in terms of the informativeness of the numerical data processed by the probabilistic ODE solver.

To this end, we revisit the continuous-time OCP formulation and give an introduction to probabilistic ODE solvers in Section 2. In Section 3, we modify the OCP to allow for incorporating the integration error quantification provided by the probabilistic ODE solver, discussing the potential as well as limitations of the proposed approach. In Section 4, we illustrate the proposed method with a numerical example.

2. Problem formulation

We consider the continuous-time OCP

$$\min_{u(\cdot),x(\cdot)} \quad \int_0^T \phi(x(t),u(t))\mathrm{d}t \tag{1a}$$

$$\text{s.t.} \quad x(0) = x_0 \tag{1b}$$

$$\dot{x}(t) = f(t,x(t),u(t)), \forall t \in [0,T] \tag{1c}$$

where the aim is to find an open-loop input trajectory $u(t)$ minimizing the cost functional (1a), subject to nonlinear dynamics given by the initial value problem (IVP) (1b), (1c). The solution of the IVP is assumed to be unique and $(p+1)$-times continuously differentiable; this is achieved, e.g., when the vector field f is globally Lipschitz continuous and when f and u are p-times continuously differentiable, see, e.g., (Hennig et al., 2022, Thm. 36.6). The stage cost is considered quadratic and positive definite, i.e., $\phi(x(t),u(t)) = \frac{1}{2}\left(x(t)^\top W_x x(t) + u(t)^\top W_u u(t)\right)$, with positive definite matrices $W_x \in \mathbb{R}^{n_x \times n_x}$ and $W_u \in \mathbb{R}^{n_u \times n_u}$.

Remark 1 *A quadratic and positive definite terminal cost $\phi_f(x(T)) = x(T)^\top P x(T)$ has been omitted from the cost (1a) in the following exposition without loss of generality.*

Remark 2 *In principle, our proposed method also allows for path constraints $h(x(t),u(t)) \leq 0$ to be considered in the optimal control problem. Approaches for including constraints in the OCP formulation are discussed in Section 3.*

As the OCP (1) generally does not admit an analytical solution, we employ the direct transcription method to obtain a numerical approximation. For ease of exposition, we choose the single-shooting formulation of the OCP and eliminate the state variables. Therefore, the input function is parameterized by a parameter vector $\theta := (\theta_1, \ldots, \theta_{N-1})$ in terms of a time-dependent input policy $u(t) = \pi_\theta(t)$ defined on the subintervals $t \in [t_i, t_{i+1})$ of the grid $\{t_i\}_{i=0}^N \subset [0,T]$, with $0 = t_0 < \ldots < t_N = T$. By splitting the cost integral into the same subintervals as well as enforcing the path constraints only on the same grid, we obtain the direct single-shooting discretization of (1),

$$\min_\theta \quad \sum_{i=0}^{N-1} \Phi_i(x_\theta(\cdot), \pi_\theta(\cdot)) \tag{2}$$

where $x_\theta(t_i)$ denotes the *exact* solution to the continuous-time IVP (1b), (1c) given the control input function $u(t) = \pi_\theta(t)$, and the stage cost $\Phi : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}$ is given as

$$\Phi_i(x(\cdot),u(\cdot)) = \int_{t_i}^{t_{i+1}} \phi(x(t),u(t))\mathrm{d}t. \tag{3}$$

Lastly, the exact solution of the IVP (1b), (1c) as well as computation of the integral (3) are approximated in order to obtain a tractable OCP formulation. In this paper, the cost integral is approximated by the Riemann discretization of the quadratic stage cost, that is

$$\Phi_i(x(\cdot),u(\cdot)) \approx \hat{\Phi}_i(x(t_i),u(t_i)) = \frac{\Delta t_i}{2}\left(x(t_i)^\top W_x x(t_i) + u(t_i)^\top W_u u(t_i)\right), \tag{4}$$

where $\Delta t_i \doteq t_{i+1} - t_i$. For the ODE solution, in the following we investigate classical and probabilistic approximation methods through the lens of probabilistic numerics (Hennig et al., 2022).

Classical ODE solvers

The solution of the IVP (1b), (1c) as part of the optimization problem is typically performed using classical numerical integration methods, such as Runge–Kutta or collocation methods (Hairer et al., 1993). For the most part, these methods construct a point estimate

$$\hat{x}_\theta(t_i) \approx x_\theta(t_i) \tag{5}$$

by extrapolating a p-th order polynomial based on the initial condition at the last time step of the approximated solution, as well as $p \geq 1$ evaluations of the ODE vector field. Crucially, the extrapolation is performed under the assumption of a perfectly known initial condition, complicating the propagation of integration error across multiple integration steps (Schober et al., 2014).

In this paper, we employ probabilistic ODE solvers to obtain a numerical approximation of the ODE solution as part of the OCP (1). To this end, in the following section we present a brief introduction to probabilistic ODE solvers.

Probabilistic ODE solvers

Probabilistic numerics refers to a class of methods that attempt to quantify errors in numerical computations probabilistically (Hennig et al., 2015; Oates and Sullivan, 2019; Hennig et al., 2022). Thus, the output of a probabilistic solver to a numerical problem is a probability distribution over candidate solutions rather than a single point. Probabilistic solvers can roughly be put into two categories: methods that represent the probability distribution over candidate solutions implicitly via stochasticity (Teymur et al., 2016, 2018; Conrad et al., 2017; Abdulle and Garegnani, 2020), and methods that represent the distribution over solution candidates explicitly, via the Bayesian formulation (Cockayne et al., 2019; Schober et al., 2014; Kersting and Hennig, 2016a; Schober et al., 2019; Tronarp et al., 2019b, 2021; Kersting, 2021). This classification is not rigid: Indeed, there are methods that draw on ideas from both concepts (Chkrebtii et al., 2016).

In the Bayesian framework, a solver is constructed by first defining a prior distribution on the quantity of interest, $x(t)$, and then defining a likelihood reflecting the constraints on $x(t)$ implied by the numerical problem. In particular, for Bayesian ODE solvers, the likelihood is usually defined by subsampling the dynamics constraint (1c), though this can be extended to other information operators, such as sensor data, algebraic constraints and higher-order derivative information (Bosch et al., 2022). The aim is thus, given some prior on $x(t)$, to compute a posterior distribution of the form

$$p\left(x(t) \mid x(0) = x_0, \{\dot{x}(t_i) = f(t_i, x(t_i), u(t_i))\}_{i=0}^{N}\right), \tag{6}$$

where, for notational simplicity, we present the formulas using the same time discretization for the numerical ODE solver and the cost discretization. In this paper we consider the Gaussian state estimation approach to probabilistic ODE solvers, as introduced by Kersting and Hennig (2016b); Tronarp et al. (2019b); Schober et al. (2019); see also Hennig et al. (2015) for a thorough introduction. We briefly describe this approach in the following and formulate the numerical IVP solution as a Bayesian state estimation problem.

Bayesian ODE solvers jointly model our belief about the ODE solution $x(t)$ together with its first p derivatives with an extended state

$$X(t) = \left(x(t), \dot{x}(t), \ldots, \overset{(p)}{x}(t)\right). \tag{7}$$

Restricted to the discretization grid $\{t_i\}_{i=0}^N$, the prior is fully described by the density

$$p(X(t_{0:N}) \mid \theta, \kappa) = \prod_{i=0}^{N-1} p\left(X(t_{i+1}) \mid X(t_i), \theta, \kappa\right) p(X(0) \mid \theta, \kappa) \tag{8}$$

with the initial and transition densities

$$p(X(0) \mid \theta, \kappa) = \delta(\bar{X}_0), \tag{9}$$

$$p\left(X(t_{i+1}) \mid X(t_i), \theta, \kappa\right) = \mathcal{N}\left(X(t_{i+1}); A(\Delta t_i) X(t_i), \kappa Q(\Delta t_i)\right), \tag{10}$$

respectively, where $A(\Delta t_i)$ and $Q(\Delta t_i)$ are given by the exact discretization of the transition and diffusion matrix of the continuous-time integrated Wiener process prior (see, e.g., Kersting et al., 2020, for the formulas), and κ is a hyper-parameter scaling the prior uncertainty. Note that choosing the prior based on a given linear system can improve the numerical efficiency for semi-linear dynamics (Bosch et al., 2023). The initial condition of the extended state $\bar{X}_0$ is chosen such that it satisfies the given initial condition of the IVP (1b), (1c) for all derivative components exactly, and can be computed efficiently via Taylor-mode automatic differentiation (Krämer and Hennig, 2020).

The likelihood model is given by the differential equation. More precisely, we condition the prior process on noise-free observations $z_i := \dot{x}(t_i) - f(t_i, x(t_i), u(t_i)) = 0$, $i = 1, \ldots, N$, which can equivalently be described in terms of the extended state as

$$Z(t_i) = E_1 X(t_i) - f(t_i, E_0 X(t_i), \pi_\theta(t_i)) = 0, \tag{11}$$

where the matrices E_0, $E_1 \in \mathbb{R}^{n_x \times (p+1) n_x}$ are selection matrices extracting $x(t)$ and $\dot{x}(t)$, respectively, from $X(t)$.

A probabilistic estimate of the ODE solution and its derivatives is then obtained by computing the posterior density $p(X(t_{0:N}) \mid Z(t_{1:N}), \theta, \kappa)$ according to the following filtering problem:

$$p(X(t_0) \mid \theta, \kappa) = \delta(\bar{X}_0), \tag{12a}$$

$$p(X(t_i) \mid X(t_{i-1}), \theta, \kappa) = \mathcal{N}(X(t_i); A(\Delta t_{i-1}) X(t_{i-1}), \kappa Q(\Delta t_{i-1})), \tag{12b}$$

$$p(Z(t_i) \mid X(t_i), \theta, \kappa) = \mathcal{N}(Z(t_i); E_1 X(t_i) - f(E_0 X(t_i), \pi_\theta(t_i)), 0), \tag{12c}$$

for $i = 1, \ldots, N$. For linear (affine) vector fields, exact inference can be performed using a RTS smoother (Rauch et al., 1965), but in the general nonlinear case exact inference is intractable. We therefore resort to efficient linearization-based approximate inference methods such as the Extended Kalman Filter (EKF) or the (Iterated) Extended Kalman Smoother ((I)EKS). Finally, we obtain the Gaussian posterior time marginals

$$\gamma_{\text{prob}}(x_i) \doteq \mathcal{N}(x_i; E_0 \xi_\theta(t_i), \kappa E_0 \Lambda_\theta(t_i) E_0^\top) \approx p(E_0 X(t_i) \mid Z(t_{1:N}), \theta, \kappa), \tag{13}$$

where the formulas for the smoothing mean $\xi_\theta(t_i)$ and covariance $\Lambda_\theta(t_i)$ obtained by the (I)EKS can be found in appendix A.1; the calibration of κ, in appendix A.2.

Remark 3 (Convergence rates and smoothness of control input) *When the true solution to the IVP (1b), (1c) is of smoothness $p+1$, for a p-times integrated Wiener process prior probabilistic ODE solvers are expected to enjoy a global convergence rate of $\mathcal{O}(\delta^p)$ (Kersting et al., 2020; Tronarp et al., 2021), where $\delta \doteq \max\{\Delta t_i\}_{i=0}^{N-1}$. The smoothness of the solution is limited by the smoothness of the vector field (Arnold, 1992), which in turn is limited by the design of the control signal. Thus, it might be necessary to select a sufficiently smooth control function to guarantee the desired convergence order.*

3. Probabilistic ODE solvers for optimal control problems

In order to utilize the posterior uncertainty estimate of the approximate IVP solution returned by the probabilistic ODE solver, the OCP formulation needs to be adapted. We obtain a well-defined, deterministic cost expression by minimizing the expectation of the cost quadrature with respect to the probabilistic belief over the ODE flow, i.e., by solving $\min_\theta V(\theta)$, where

$$V(\theta) \doteq \mathbb{E}_{x_{0:N-1}}\left[\sum_{i=0}^{N-1} \hat{\Phi}_i(x_i, \pi_\theta(t_i))\right], \tag{14}$$

$$= \sum_{i=0}^{N-1} \mathbb{E}_{x_i}\left[\hat{\Phi}_i(x_i, \pi_\theta(t_i))\right], \tag{15}$$

$$= \sum_{i=0}^{N-1} \int \hat{\Phi}_i(x_i, \pi_\theta(t_i))\gamma(x_i)\mathrm{d}x_i. \tag{16}$$

Using the analytical expectation of the quadratic cost in (4), we propose to approximate the solution to (2) with respect to the numerical approximation of the IVP solution by solving the OCP

$$\min_\theta \quad \sum_{i=0}^{N-1} \frac{\Delta t_i}{2}\left(\|E_0\xi_\theta(t_i)\|^2_{W_x} + \|\pi_\theta(t_i)\|^2_{W_u} + \hat{\kappa}_\theta \cdot \mathrm{tr}\left\{W_x E_0 \Lambda_\theta(t_i) E_0^\top\right\}\right), \tag{17}$$

where $\xi_\theta(t_i)$ and $\Lambda_\theta(t_i)$ are the smoothing mean and covariance of the EKS, and $\hat{\kappa}_\theta$ is the prior diffusion scale calibrated according to Appendix A.2, as a function of the inputs θ.

Remark 4 *For a more accurate solution of the filtering problem* (12) *the IEKS estimate of the filtering problem* (12) *can be used by including the IEKS fixed-point equations into* (17)*, see appendix A.1.*

The reformulation of the deterministic cost in (2) as an expectation over the probabilistic belief of the ODE solution lends itself to the following interpretation from the viewpoint of probabilistic numerics. Clearly, by assigning a Dirac measure to the exact IVP solution at time t_i,

$$\gamma_{\text{exact}}(x_i) \doteq \delta(x_i - x_\theta(t_i)), \tag{18}$$

minimizing the expected cost in equation (16) with $\gamma(x_i)$ set to $\gamma_{\text{exact}}(x_i)$ still recovers the exact solution to (2). Yet, when computing an approximation to the IVP solution, we can use the same expression, but instead replace the exact state with the approximation computed by a classical method or a probabilistic ODE solver. For classical ODE solvers, it becomes evident that likewise assigning the Dirac density shifted by the computed point estimate $\hat{x}_\theta(t_i) \approx x_\theta(t_i)$, i.e.,

$$\gamma_{\text{clas}}(x_i) \doteq \delta(x_i - \hat{x}_\theta(t_i)), \tag{19}$$

leads to a classical direct single-shooting formulation of the continuous-time OCP, where the point estimate $\hat{x}_\theta(t_i)$ for the IVP is simply inserted into the cost (4). This formalizes the standard assumption of zero discretization error employed by classical direct transcription methods. By including the (approximate) posterior uncertainty description (13) provided by the probabilistic ODE solver,

i.e., propagating the computational uncertainty from the approximate IVP solution in the cost of the OCP (2), this assumption can thus be (partly) relaxed.

Since the uncertainty estimate of the probabilistic ODE solver depends on the input trajectory, an optimal solution to (17) captures the trade-off between obtaining an optimal mean prediction and reducing the computational uncertainty associated with it. Reduction of the computational uncertainty can be achieved by choosing a state-input trajectory that leads to a bigger uncertainty reduction in terms of the posterior covariance $\Lambda_\theta(t_i)$ in (17). In that sense, the computed optimal inputs can exhibit a dual-control effect (Bar-Shalom and Tse, 1974; Mesbah, 2018), in particular, for nonlinear ODEs. The connection of the proposed approach to dual control becomes clear when viewed through the lens of probabilistic numerics: Interpreting the ODE constraint (11) as (numerical) data observed within the prediction horizon, the proposed approach can be seen as a stochastic output-feedback MPC acting on the extended state $X(t)$ in (7) conditioned on the data. However, compared to the setting of most dual control approaches that incorporate the fact that (physical) data is going to be observed in the future, the (numerical) data in this setting is known *a priori*.

The potential of the proposed method in real-time applications with limited computational budget strongly depends on the computational overhead added by solving (17) with a probabilistic ODE solver instead of solving (2) with a classical ODE solver. Concerning the choice of ODE solver, while both classical and probabilistic ODE solvers share the same linear time complexity, the complexity in the state dimension for a vanilla probabilistic ODE solver implementation is cubic, compared to the linear complexity for most classical solvers. By assuming independence between different state dimensions (which classical solvers generally do since correlations are not modeled), probabilistic ODE solvers can enjoy linear space complexity as well (Krämer et al., 2022). Still, to which extent the independence assumption and the implied linearization strategy (Krämer et al., 2022, Sec. 2.2) affects the dynamics-dependent uncertainty reduction targeted in this paper requires further investigation. Concerning the choice of OCP formulation, optimizing over covariance matrices generally incurs a significant computational overhead, either due to increased nonlinearity or increased state dimension of the condensed or non-condensed OCP formulation, respectively. Going forward, computational efficiency of the proposed OCP formulation will thus have to be addressed in order for the uncertainty-aware approach to be competitive with the uncertainty reduction achieved by a (fixed) finer time discretization of a "certainty-equivalent" method given the same computational budget. This paper focuses on highlighting the potential of using probabilistic integrators for integration error-aware control, motivating the development of more efficient implementations in future work.

Remark 5 (Path and terminal constraints) *Due to the assumed Gaussian density distribution returned by linearization-based solvers such as the (I)EKS, discretized path-constraints and terminal constraints $h(x(t_i), u(t_i)) \leq 0$, $i = 0, \dots, N$ of the original open-loop problem have to be reformulated probabilistically, based on the probabilistic belief about the IVP solution. Computationally efficient approaches include linearization-based constraint tightenings on the posterior mean (Van Hessem and Bosgra, 2006; Kouvaritakis et al., 2010; Cannon et al., 2011), or penalties on the expected constraint violation (Messerer et al., 2023); see also Mesbah (2016) for an overview.*

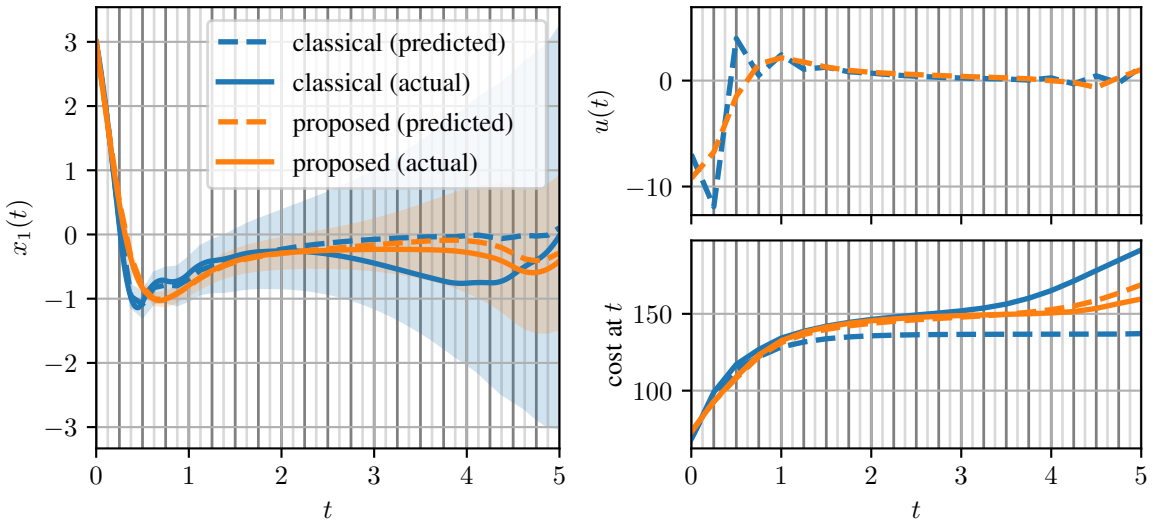

Figure 1: OCP solutions (dashed lines) compared to ground-truth simulations (solid lines) for the classical (blue) and proposed (orange) approach for $N_{\text{int}} = 40$ (left: state $x_1(t)$, top right: cost value, bottom right: applied input). The input computed by the proposed approach actively reduces computational uncertainty (shaded) for a lower expected cost in (17), which leads to a more accurate cost prediction and a lower ground-truth cost (4).

4. Numerical example

To illustrate the properties of the algorithm, we apply it to a numerical example[1]. For the probabilistic ODE solver, we use the JAX (Bradbury et al., 2018) implementation of a probabilistic ODE solver based on an EKS provided by the package `probdiffeq` (Krämer, 2023). The OCP (17) is implemented in JAX in condensed form using a low-rank variant of the BFGS (Nocedal and Wright, 2006) optimizer with bounds (L-BFGSB) optimizer, as implemented in the `jaxopt` package (Blondel et al., 2022). To test the incorporation of the uncertainty in the IVP solution separately from the integrator choice, the proposed approach is compared to a point estimate obtained as the mean prediction of the same probabilistic integrator, which is equivalent to solving (17) without considering the computational uncertainty, i.e., setting $\Lambda_\theta(t_i) = 0$ in (17). Hence, we denote as "proposed" and "classical" the proposed approach and proxy for the classical approach, respectively. The predicted state trajectories of both approaches are compared to a ground-truth simulation with the respective optimal control input, obtained by simulating the ODE using the IDAS (Cao et al., 2003; Hindmarsh et al., 2005) adaptive integration method with a relative error tolerance set to 10^{-10}.

We consider the logistic ODE with an added input, $\dot{x}_1(t) = (1 - x_2(t))^2 x_1(t) - x_2(t) + u(t)$, $\dot{x}_2(t) = x_1(t)$, with an initial condition $x(0) = (3, 1)$ and cost matrices $W_x = 50I_2$, $W_u = 1$. The problem is discretized into $N = 20$ control intervals for a time horizon of $T = 5$. Using a piece-wise affine parameterization of the input function, the probabilistic integrator is applied with integration order $p = 1$.

The simulation results for $N_{\text{int}} = 40$ integration steps is shown in Figure 1. It can be seen that the input computed by the proposed approach (top right image) reduces the uncertainty estimate,

1. The source code for the numerical experiments is available at doi:10.3929/ethz-b-000673668.

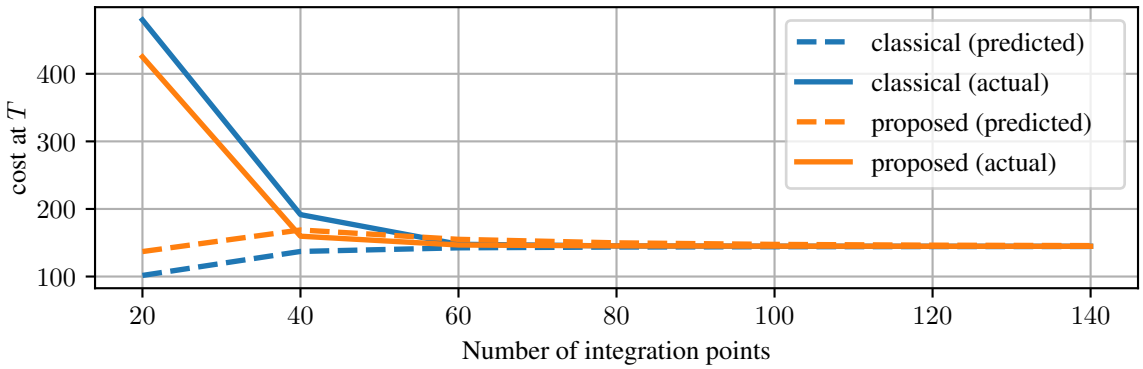

Figure 2: Predicted and ground-truth cost for an increasing number of integration points. For a low number of integration points, the proposed approach achieves a lower actual cost by adapting the input for a more accurate numerical discretization.

leading to a more precise cost estimate (left image), which, in this case, also leads to a lower total actual cost (bottom right image). Note that, to show the uncertainty reduction achieved by the proposed approach, the 2σ-confidence interval of the integrators' marginal uncertainty estimate is also plotted for the classical approach, even though it only considers the mean estimate.

Figure 2 shows the predicted and ground-truth cost values at time T for an increased number of integration points. Thereby, to isolate the effects of the ODE discretization from the cost discretization, the number of control intervals is kept constant while the number of total integration steps within each control interval is varied. In the range $N_{\text{int}} \in \{20, 40\}$, the proposed approach leads to a lower actual cost, which is also in better agreement with the predicted cost; for $N_{\text{int}} \geq 60$, the slightly underconfident uncertainty estimate leads to a marginally higher predicted cost for the proposed approach, albeit similar actual costs, while both approaches converge as the numerical accuracy increases. While the proposed approach achieves a better agreement of predicted and actual cost, as well as a lower actual cost for the coarser grids, for $N_{\text{int}} = 20$ it can be seen that there is still a considerable difference between the predicted and actual cost value. Closer investigations indicate that this discrepancy may be attributed to an overconfident uncertainty calibration; the development of improved uncertainty quantification strategies is reserved for future work.

5. Conclusion and Outlook

In this paper, we have investigated the idea of probabilistic numerics to propagate the numerical uncertainty, incurred by the approximate discretization of the IVP to predict the dynamics inside the OCP, in the cost prediction. Quantifying the input-dependent integration error using probabilistic ODE solvers, the computed input can achieve lower ground-truth cost in the presence of computational uncertainty as it automatically determines a trade-off between solving the IVP accurately and obtaining a lower nominal cost prediction with respect to the approximate IVP solution, which has been demonstrated in a numerical example. For future work, a computationally efficient implementation is to be addressed to achieve these performance benefits in real-time applications with limited computational budget.

Appendix A. Technical details on probabilistic ODE solvers

A.1. Extended Kalman Smoother equations

In the following, we present the equations used for approximating the solution to the smoothing problem (12). Thereby we mostly omit the subscript θ to avoid excessive notation. The EKF filtering pass starts from $\mu(t_0) = \bar{X}_0$, $\Sigma(t_0) = 0$ and iterates forward in time over the equations

$$\mu(t_i^-) = A(\Delta t_{i-1})\mu(t_{i-1}), \tag{20a}$$
$$\kappa\Sigma(t_i^-) = \kappa\big(A(\Delta t_{i-1})\Sigma(t_{i-1})A^\top(\Delta t_{i-1}) + Q(\Delta t_{i-1})\big), \tag{20b}$$
$$C(t_i) = E_1^\top - E_0^\top J_f(t_i, \tilde{x}(t_i), \pi_\theta(t_i)), \tag{20c}$$
$$b(t_i) = f(t_i, \tilde{x}(t_i), \pi_\theta(t_i)) - J_f(t_i, \tilde{x}(t_i), \pi_\theta(t_i))\tilde{x}(t_i) \tag{20d}$$
$$\kappa S(t_i) = \kappa C(t_i)\Sigma(t_i^-)C^\top(t_i), \tag{20e}$$
$$K(t_i) = \Sigma(t_i^-)C^\top(t_i)S^{-1}(t_i), \tag{20f}$$
$$\mu(t_i) = \mu(t_i^-) + K(t_i)\big(b(t_i) - C(t_i)\mu(t_i^-)\big), \tag{20g}$$
$$\kappa\Sigma(t_i) = \kappa\big(\Sigma(t_i^-) - K(t_i)S(t_i)K^\top(t_i)\big), \tag{20h}$$

where J_f denotes the Jacobian of f with respect to x. The RTS smoothing pass starts from the filtering mean and covariance, $\xi(t_N) = \mu(t_N)$, $\Lambda(t_N) = \Sigma(t_N)$, and iterates backwards as follows:

$$G(t_i) = \Sigma(t_i)A^\top(\Delta t_{i+1})\Sigma^{-1}(t_{i+1}^-), \tag{20i}$$
$$\kappa P(t_i) = \kappa(\Sigma(t_i) - G(t_i)\Sigma(t_{i+1}^-)G^\top(t_i)), \tag{20j}$$
$$\xi(t_i) = \mu(t_i) + G(t_i)(\xi(t_{i+1}) - \mu(t_{i+1}^-)), \tag{20k}$$
$$\kappa\Lambda(t_i) = \kappa(G(t_i)\Lambda(t_{i+1})G^\top(t_i) + P(t_i)). \tag{20l}$$

For the EKS, the linearization point $\tilde{x}(t_i) = E_0\mu(t_i^-)$ is the predictive mean of the EKF; for the IEKS, the smoothing mean $\tilde{x}(t_i) = E_0\xi(t_i)$ of the previous iteration. Note that covariances are invariant to scaling of κ (Tronarp et al., 2019a,b). Therefore, the filter can be run for $\kappa = 1$ and the resulting covariances can be post-multiplied once an appropriate value has been determined.

A.2. Calibration

To return meaningful uncertainty estimates, the scaling parameter κ needs to be estimated. We follow the approach described by Tronarp et al. (2019b) and Bosch et al. (2021) and choose $\hat{\kappa}$ as the quasi-maximum likelihood estimate (MLE). In view of (20), this estimate can be obtained by solving

$$\hat{\kappa}_\theta = \arg\max_\kappa \sum_{i=1}^N \log\mathcal{N}(b(t_i); C(t_i)\mu(t_i^-), \kappa S(t_i)) \tag{21}$$
$$= \frac{1}{Nn_x}\sum_{i=1}^N \big(b(t_i) - C(t_i)\mu(t_i^-)\big)^\top S^{-1}(t_i)\big(b(t_i) - C(t_i)\mu(t_i^-)\big). \tag{22}$$

Acknowledgments

The authors thank Nicholas Krämer for his support in using the `probdiffeq` software library, Andrea Carron, Johannes Köhler and Anna Scampicchio for providing helpful feedback, as well as Moritz Diehl and Armin Nurkanović for inspiring discussions on the initial motivation of the project. AL and MZ received funding by the European Union's Horizon 2020 research and innovation programme, Marie Skłodowska-Curie grant agreement No. 953348, ELO-X. FT was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. PH, NB and JS gratefully acknowledge financial support by the DFG Cluster of Excellence "Machine Learning - New Perspectives for Science", EXC 2064/1, project number 390727645; the German Federal Ministry of Education and Research (BMBF) through the Tübingen AI Center (FKZ: 01IS18039A), as well as funds from the Ministry of Science, Research and Arts of the State of Baden-Württemberg. The authors also thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting NB and JS.

References

A. Abdulle and G. Garegnani. Random time step probabilistic methods for uncertainty quantification in chaotic and geometric numerical integration. *Statistics and Computing*, 2020.

Vladimir I Arnold. *Ordinary Differential Equations*. Springer Science & Business Media, 1992.

Ivo Babuška and Manil Suri. The p- and h-p versions of the finite element method, an overview. *Computer Methods in Applied Mechanics and Engineering*, 80(1), 1990. doi:10.1016/0045-7825(90)90011-A.

Y. Bar-Shalom and E. Tse. Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Transactions on Automatic Control*, 19(5), 1974. doi:10.1109/TAC.1974.1100635.

John T. Betts and William P. Huffman. Mesh refinement in direct transcription methods for optimal control. *Optimal Control Applications and Methods*, 19(1), 1998. doi:10.1002/(SICI)1099-1514(199801/02)19:1<1::AID-OCA616>3.0.CO;2-Q.

T. Binder, A. Cruse, C.A. Cruz Villar, and W. Marquardt. Dynamic optimization using a wavelet based adaptive control vector parameterization strategy. *Computers & Chemical Engineering*, 24 (2-7), 2000. doi:10.1016/S0098-1354(00)00357-4.

Mathieu Blondel, Quentin Berthet, Marco Cuturi, Roy Frostig, Stephan Hoyer, Felipe Llinares-López, Fabian Pedregosa, and Jean-Philippe Vert. Efficient and Modular Implicit Differentiation, 2022.

N. Bosch, P. Hennig, and F. Tronarp. Calibrated adaptive probabilistic ODE solvers. In *24th International Conference on Artificial Intelligence and Statistics*, 2021.

Nathanael Bosch, Filip Tronarp, and Philipp Hennig. Pick-and-Mix Information Operators for Probabilistic ODE Solvers. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*. PMLR, 2022. URL https://proceedings.mlr.press/v151/bosch22a.html.

Nathanael Bosch, Philipp Hennig, and Filip Tronarp. Probabilistic Exponential Integrators. *Advances in Neural Information Processing Systems*, 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/7f64034009f4a5fa417a57e1a987c5cd-Abstract-Conference.html.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: Composable transformations of Python+NumPy programs, 2018. URL http://github.com/google/jax.

R. Cagienard, P. Grieder, E. C. Kerrigan, and M. Morari. Move blocking strategies in receding horizon control. *Journal of Process Control*, 17(6), 2007. doi:10.1016/j.jprocont.2007.01.001.

Mark Cannon, Basil Kouvaritakis, Saša V. Raković, and Qifeng Cheng. Stochastic Tubes in Model Predictive Control With Probabilistic Constraints. *IEEE Transactions on Automatic Control*, 56 (1), 2011. doi:10.1109/TAC.2010.2086553.

Yang Cao, Shengtai Li, Linda Petzold, and Radu Serban. Adjoint Sensitivity Analysis for Differential-Algebraic Equations: The Adjoint DAE System and Its Numerical Solution. *SIAM J. Sci. Comput.*, 24(3), 2003. doi:10.1137/S1064827501380630.

Yutao Chen, Nicoló Scarabottolo, Mattia Bruschetta, and Alessandro Beghi. Efficient move blocking strategy for multiple shooting-based non-linear model predictive control. *IET Control Theory & Applications*, 14(2), 2020. doi:10.1049/iet-cta.2019.0168.

O. A. Chkrebtii, D. A. Campbell, B. Calderhead, and M. A. Girolami. Bayesian solution uncertainty quantification for differential equations. *Bayesian Analysis*, 11(4), 2016.

Jon Cockayne, Chris J. Oates, T. J. Sullivan, and Mark Girolami. Bayesian Probabilistic Numerical Methods. *SIAM Rev.*, 61(4), 2019. doi:10.1137/17M1139357.

P. R. Conrad, M. Girolami, S. Särkkä, A. Stuart, and K. Zygalakis. Statistical analysis of differential equations: Introducing probability measures on numerical solutions. *Statistics and Computing*, 27(4), 2017.

Christopher L. Darby, William W. Hager, and Anil V. Rao. An hp-adaptive pseudospectral method for solving optimal control problems. *Optimal Control Applications and Methods*, 32(4), 2011. doi:10.1002/oca.957.

Feng Gao, Yu Han, Shengbo Eben Li, Shaobing Xu, and Dongfang Dang. Accurate Pseudospectral Optimization of Nonlinear Model Predictive Control for High-Performance Motion Planning. *IEEE Transactions on Intelligent Vehicles*, 8(2), 2023. doi:10.1109/TIV.2022.3153633.

Ernst Hairer, Gerhard Wanner, and Syvert P. Nørsett. *Solving Ordinary Differential Equations I*, volume 8 of *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 1993. ISBN 978-3-540-56670-0 978-3-540-78862-1. doi:10.1007/978-3-540-78862-1.

Philipp Hennig, Michael A. Osborne, and Mark Girolami. Probabilistic numerics and uncertainty in computations. *Proc. R. Soc. A.*, 471(2179), 2015. doi:10.1098/rspa.2015.0142.

Philipp Hennig, Michael A Osborne, and Hans P Kersting. *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press, 2022.

Alan C. Hindmarsh, Peter N. Brown, Keith E. Grant, Steven L. Lee, Radu Serban, Dan E. Shumaker, and Carol S. Woodward. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw.*, 31(3), 2005. doi:10.1145/1089014.1089020.

H. Kersting and P. Hennig. Active uncertainty calibration in Bayesian ODE solvers. In *Uncertainty in Artificial Intelligence (UAI) 2016*, 2016a.

Hans Kersting. *Uncertainty-Aware Numerical Solutions of ODEs by Bayesian Filtering*. PhD thesis, Universität Tübingen, 2021.

Hans Kersting and Philipp Hennig. Active Uncertainty Calibration in Bayesian ODE Solvers. In *UAI 2016 Proceedings*, NY, USA, 2016b. URL http://www.auai.org/uai2016/proceedings/papers/163.pdf.

Hans Kersting, T. J. Sullivan, and Philipp Hennig. Convergence rates of Gaussian ODE filters. *Stat Comput*, 30(6), 2020. doi:10.1007/s11222-020-09972-4.

Markus Kögel and Rolf Findeisen. Discrete-time robust model predictive control for continuous-time nonlinear systems. In *2015 American Control Conference (ACC)*, 2015. doi:10.1109/ACC.2015.7170852.

Basil Kouvaritakis, Mark Cannon, Saša V. Raković, and Qifeng Cheng. Explicit use of probabilistic distributions in linear predictive control. *Automatica*, 46(10), 2010. doi:10.1016/j.automatica.2010.06.034.

Nicholas Krämer. Probdiffeq, 2023. URL https://github.com/pnkraemer/probdiffeq.

Nicholas Krämer and Philipp Hennig. Stable Implementation of Probabilistic ODE Solvers, 2020.

Nicholas Krämer, Nathanael Bosch, Jonathan Schmidt, and Philipp Hennig. Probabilistic ODE Solutions in Millions of Dimensions. In *Proceedings of the 39th International Conference on Machine Learning*. PMLR, 2022. URL https://proceedings.mlr.press/v162/kramer22b.html.

Evgeny Lazutkin, Abebe Geletu, and Pu Li. An Approach to Determining the Number of Time Intervals for Solving Dynamic Optimization Problems. *Ind. Eng. Chem. Res.*, 57(12), 2018. doi:10.1021/acs.iecr.7b03361.

K. Lee, W. H. Moase, and C. Manzie. Mesh adaptation in direct collocated nonlinear model predictive control. *International Journal of Robust and Nonlinear Control*, 28(15), 2018. doi:10.1002/rnc.4235.

Yana Lishkova, Mark Cannon, and Sina Ober-Blöbaum. A Multirate Variational Approach to Nonlinear MPC. In *2022 European Control Conference (ECC)*, 2022. doi:10.23919/ECC55457.2022.9838240.

Ali Mesbah. Stochastic Model Predictive Control: An Overview and Perspectives for Future Research. *IEEE Control Systems Magazine*, 36(6), 2016. doi:10.1109/MCS.2016.2602087.

Ali Mesbah. Stochastic model predictive control with active uncertainty learning: A Survey on dual control. *Annual Reviews in Control*, 45, 2018. doi:10.1016/j.arcontrol.2017.11.001.

Florian Messerer, Katrin Baumgärtner, and Moritz Diehl. A Dual-Control Effect Preserving Formulation for Nonlinear Output-Feedback Stochastic Model Predictive Control With Constraints. *IEEE Control Systems Letters*, 7, 2023. doi:10.1109/LCSYS.2022.3230552.

Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operation Research and Financial Engineering. Springer, New York, second edition edition, 2006. ISBN 0-387-30303-0.

C. J. Oates and T. J. Sullivan. A modern retrospective on probabilistic numerics. *Statistics and Computing*, 29(6), 2019.

Luís Tiago Paiva and Fernando A. C. C. Fontes. Adaptive time–mesh refinement in optimal control problems with state constraints. *Discrete & Continuous Dynamical Systems - A*, 35(9), 2015. doi:10.3934/dcds.2015.35.4553.

Luís Tiago Paiva and Fernando A. C. C. Fontes. Sampled–Data Model Predictive Control Using Adaptive Time–Mesh Refinement Algorithms. In *CONTROLO 2016*, Lecture Notes in Electrical Engineering, Cham, 2017. Springer International Publishing. ISBN 978-3-319-43671-5. doi:10.1007/978-3-319-43671-5_13.

Luís Tiago Paiva and Fernando A. C. C. Fontes. Optimal Control Algorithms with Adaptive Time-Mesh Refinement for Kite Power Systems. *Energies*, 11(3), 2018. doi:10.3390/en11030475.

Ciro Potena, Bartolomeo Della Corte, Daniele Nardi, Giorgio Grisetti, and Alberto Pretto. Nonlinear model predictive control with adaptive time-mesh refinement. In *2018 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR)*, 2018. doi:10.1109/SIMPAR.2018.8376274.

S Joe Qin and Thomas A Badgwell. An overview of industrial model predictive control technology. In *AIche Symposium Series*, volume 93. New York, NY: American Institute of Chemical Engineers, 1971-c2002., 1997.

Rien Quirynen, Milan Vukov, and Moritz Diehl. Multiple Shooting in a Microsecond. In *Multiple Shooting and Time Domain Decomposition Methods*, Contributions in Mathematical and Computational Sciences, Cham, 2015. Springer International Publishing. ISBN 978-3-319-23321-5. doi:10.1007/978-3-319-23321-5_7.

H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8), 1965. doi:10.2514/3.3166.

Martin Schlegel, Klaus Stockmann, Thomas Binder, and Wolfgang Marquardt. Dynamic optimization using adaptive control vector parameterization. *Computers & Chemical Engineering*, 29(8), 2005. doi:10.1016/j.compchemeng.2005.02.036.

M. Schober, S. Särkkä, and P. Hennig. A probabilistic model for the numerical solution of initial value problems. *Statistics and Computing*, 29(1), 2019.

Michael Schober, David K Duvenaud, and Philipp Hennig. Probabilistic ODE Solvers with Runge-Kutta Means. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper/2014/hash/59b90e1005a220e2ebc542eb9d950b1e-Abstract.html.

Rohan C. Shekhar and Chris Manzie. Optimal move blocking strategies for model predictive control. *Automatica*, 61, 2015. doi:10.1016/j.automatica.2015.07.030.

P. Tanartkit and L. T. Biegler. A nested, simultaneous approach for dynamic optimization problems—II: The outer problem. *Computers & Chemical Engineering*, 21(12), 1997. doi:10.1016/S0098-1354(97)00014-8.

Onur Teymur, Kostas Zygalakis, and Ben Calderhead. Probabilistic linear multistep methods. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

Onur Teymur, Han Cheng Lie, Tim Sullivan, and Ben Calderhead. Implicit probabilistic integrators for ODEs. In *Advances in Neural Information Processing Systems (NIPS)*, 2018.

Petter Tøndel and Tor A. Johansen. Complexity reduction in explicit linear model predictive control. *IFAC Proceedings Volumes*, 35(1), 2002. doi:10.3182/20020721-6-ES-1901.00600.

Filip Tronarp, Toni Karvonen, and Simo Särkkä. Student's t -Filters for noise scale estimation. *IEEE Signal Processing Letters*, 26(2), 2019a.

Filip Tronarp, Hans Kersting, Simo Särkkä, and Philipp Hennig. Probabilistic solutions to ordinary differential equations as nonlinear Bayesian filtering: A new perspective. *Stat Comput*, 29(6), 2019b. doi:10.1007/s11222-019-09900-1.

Filip Tronarp, Simo Särkkä, and Philipp Hennig. Bayesian ODE solvers: The maximum a posteriori estimate. *Stat Comput*, 31(3), 2021. doi:10.1007/s11222-021-09993-7.

Dennis Van Hessem and Okko Bosgra. Stochastic closed-loop model predictive control of continuous nonlinear chemical processes. *Journal of Process Control*, 16(3), 2006. doi:10.1016/j.jprocont.2005.06.003.

Frank Wolff, Naoki Uchiyama, Mark Burkhardt, and Oliver Sawodny. Nonlinear Model Predictive Control with Non-Equidistant Discretization Time Grids for Rotary Cranes. In *2022 13th Asian Control Conference (ASCC)*, 2022. doi:10.23919/ASCC56756.2022.9828180.

Mingzhao Yu and Lorenz T. Biegler. A Stable and Robust NMPC Strategy with Reduced Models and Nonuniform Grids. *IFAC-PapersOnLine*, 49(7), 2016. doi:10.1016/j.ifacol.2016.07.212.

Yiming Zhao and Panagiotis Tsiotras. Density Functions for Mesh Refinement in Numerical Optimal Control. *Journal of Guidance, Control, and Dynamics*, 34(1), 2011. doi:10.2514/1.45852.