Online-Computation Approach to Optimal Control of Noise-Affected Nonlinear Systems with Continuous State and Control Spaces

Marc P. Deisenroth, Florian Weissel, Toshiyuki Ohtsuka, and Uwe D. Hanebeck

***Abstract*— A novel online-computation approach to optimal control of nonlinear, noise-affected systems with continuous state and control spaces is presented. In the proposed algorithm, system noise is explicitly incorporated into the control decision. This leads to superior results compared to state-of-the-art nonlinear controllers that neglect this influence. The solution of an optimal nonlinear controller for a corresponding deterministic system is employed to find a meaningful state space restriction. This restriction is obtained by means of approximate state prediction using the noisy system equation. Within this constrained state space, an optimal closed-loop solution for a finite decision-making horizon (prediction horizon) is determined within an adaptively restricted optimization space. Interleaving stochastic dynamic programming and value function approximation yields a solution to the considered optimal control problem. The enhanced performance of the proposed discrete-time controller is illustrated by means of a scalar example system. Nonlinear model predictive control is applied to address approximate treatment of infinite-horizon problems by the finite-horizon controller.**

I. Introduction

System models and state information of dynamic systems are always to some degree uncertain. The consideration of noise representing such uncertainties offers the opportunity to increase the quality with which nonlinear systems can be controlled. Therefore, the consideration of noise is important when designing an optimal controller for noise-affected dynamic systems. The nonlinear optimal control problem can be reduced to the Hamilton-Jacobi-Bellman partial differential equation that is very difficult to solve [1]. Therefore, less ambitious methods are often employed to solve the optimal control problem. Nevertheless, this optimization is still a highly challenging task in case of nonlinear systems.

In almost all technical applications the state of a dynamic system is continuous valued. For instance, when a robot moves between two points, it attains all positions on the connecting path. The use of continuous-valued state spaces is a natural way to incorporate this property into the controller design. Moreover, continuous-valued control inputs are desirable features of a controller to keep the mentioned robot close to the optimal trajectory.

Nonlinear model predictive control (NMPC) is a common approach to sidestep the computational burden of infinite-horizon problems. In NMPC a finite horizon $[t, t+N]$ is chosen, within which a solution to the optimal control problem is determined. Application of the first control action of this solution, shift of the finite-horizon window, and re-solving the optimal control problem after each time step finally results in a locally optimal, but computationally tractable, closed-loop solution to an infinite-horizon problem [2].

The dynamic programming (DP) algorithm is the discrete-time equivalent of the Hamilton-Jacobi-Bellman equation. DP exploits Bellman's principle of optimality [3] and is a useful approach to optimal control of nonlinear systems with finite sets of discrete states and control inputs. Although DP suffers from the "curse of dimensionality", it allows efficient calculation of the optimal closed-loop control inputs for deterministic as well as stochastic systems with a small number of states and controls. Moreover, DP is the only general approach for sequential optimization in case of stochastic systems [1].

In case of deterministic optimal control, which leads to suboptimal results for noise-affected systems, Pontryagin's minimum principle offers an efficient way to determine the desired control inputs. Employing this theory, approaches to optimal control of nonlinear systems with continuous state and control spaces can be found in [4] and [5]. In [4] stabilizing continuation methods are proposed to derive a solution to the optimal control problem of nonlinear, continuous-time systems with general boundary constraints. In [5] the time domain of a continuous-time system with general boundary constraints is modified by means of a continuation method. The initialization of this method yields an optimal input to the nonlinear system for a one-point horizon. While the horizon length is being continuously transformed into the whole considered horizon, the solution is being traced, such that NMPC can be applied.

For continuous-time systems, an extension of Pontryagin's minimum principle to the stochastic case is given in [6]. Here, the assumption of an underlying Ito process is employed. For the considered discrete-time case, an equivalent of Pontryagin's minimum principle for noise-affected system has not been found in the literature yet. For continuous-time systems with linear control inputs, an approach to optimal control of stochastic nonlinear systems with continuous state spaces is given in [7] and [8]. With certain restrictions on

M. P. Deisenroth was with the Intelligent Sensor-Actuator-Systems Laboratory, Institute of Computer Science and Engineering, Universität Karlsruhe (TH), Germany. He is currently with the Max Planck Institute for Biological Cybernetics, Department of Empirical Inference for Machine Learning and Perception, Spemannstraße 38, 72076 Tübingen, Germany. deisenroth@tuebingen.mpg.de

F. Weissel and U. D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory, Institute of Computer Science and Engineering, Universität Karlsruhe (TH), Germany. {weissel|uwe.hanebeck}@ieee.org

T. Ohtsuka is with the Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan. ohtsuka@sys.es.osaka-u.ac.jp

the noise structure and the assumption of a cost function that is quadratic in the control input, the optimal control problem can be written as a path integral, for which an approximate solution can be found by Monte Carlo methods.

For discrete-time systems, an approach to infinite-horizon optimal control that considers the noise influence on a system with continuous state spaces, but only a finite set of control inputs, is presented in [9]. In this approach the DP value function is approximated by means of a radial basis function network with a finite number of Gaussian kernels. Evaluation of this network at the mean values of the kernels yields a finite Markov decision problem, which can be solved by approximate value iteration. In [10] the same problem class is considered for NMPC with finite prediction horizon. Here an approach is presented that provides a closed-form approximate solution. The method is based on Gaussian mixture representation of the cost function. In addition, transition densities are approximated by means of axis-aligned Gaussian mixtures. In [11] a value function approximation scheme for this framework employing DP is presented, which significantly lowers the computational demand.

Regarding the control problem as a reinforcement learning problem, an approach to derive a closed-form evaluation of the value function of a nonlinear, noise-affected system in discrete time with continuous state and control spaces is presented in [12]. Here, the value function is approximated by means of a Gaussian process (GP). GPs represent a distribution over functions and extend the properties of a set of support points to the entire continuous-valued space. Applying GPs to system identification and to value function approximation, policy iteration yields the desired optimal policy. Similarly, in [13] and [14], a nonlinear discrete-time system is identified by means of GPs. To solve the prediction of uncertain system states analytically, the distributions of the successor states are (pointwise) approximated by Gaussians. Within the NMPC framework, a controller is obtained that determines an optimal control input, which strongly depends on the quality of the learned system dynamics.

In this paper an approach is presented that considers the noise influence in the optimal control of a discrete-time nonlinear system with unconstrained continuous-valued state spaces and control inputs. Starting off by solving a corresponding noise-free optimal control problem, the state and the control spaces are restricted to an area around the corresponding trajectories provided by this initial solution. Here, an advanced solution incorporating the noise influence is derived by stochastic DP combined with value function approximation as well as nonlinear stochastic state prediction. This method can be treated as a solution to the closed-loop optimal control problem for noisy systems within NMPC.

This paper is structured as follows. The considered system and the corresponding optimal control problem are introduced in Section II. In Section III the proposed online-computation approach is described. Benefits gained by this algorithm are illustrated by means of a scalar example system in Section IV. Finally, in Section V the results of this paper are summarized, and a survey of future work is given.

II. Problem Formulation

We consider a discrete-time system, which is given by

$$\underline{\boldsymbol{x}}_{k+1} = \underline{f}(\underline{\boldsymbol{x}}_k, \underline{u}_k) + \underline{\boldsymbol{w}}_k\,,\ k = 0, \ldots, N-1\,, \tag{1}$$

where $\underline{x}_k \in \mathbb{R}^{n_x}$ is the system state, $\underline{u}_k \in \mathbb{R}^{n_u}$ the control input, and $\underline{f}$ a nonlinear function. $\underline{\boldsymbol{w}}_k \in \mathbb{R}^{n_x}$ is a zero-mean Gaussian white noise term with covariance matrix $\mathbf{C}_w$. The initial state $\underline{x}_0$ is assumed to be known, and the states $\underline{x}_k$ are directly accessible at each time step.

Notation: Throughout this paper, $\underline{x}$ is a vector-valued variable, x a scalar, and $\underline{\boldsymbol{x}}$ a vector-valued random variable. Matrices are denoted by capital boldface letters $\mathbf{X}$.

To determine an optimal solution to the finite-horizon control problem, a cost function is introduced. In the following, the important case of an additive cost function is considered. For a state $\underline{x}_k$ and a given policy $\pi_k \coloneqq (\underline{u}_k, \ldots, \underline{u}_{N-1})$, the *expected cost-to-go* from time step k to N within the N-step optimization horizon is defined as

$$V_k^{\pi_k}(\underline{x}_k) \coloneqq \mathop{\mathrm{E}}_{\underline{\boldsymbol{w}}_k, \ldots, \underline{\boldsymbol{w}}_{N-1}} \left[g_N(\underline{\boldsymbol{x}}_N) + \sum_{i=k}^{N-1} g_i(\underline{\boldsymbol{x}}_i, \underline{u}_i) \right]. \tag{2}$$

The function $g_N(\underline{x}_N)$ denotes the terminal cost, and $g_i(\underline{x}_i, \underline{u}_i)$ is the step cost from time i to $i+1$ depending on the system state and the applied control input at time i.

An optimal policy $\pi^* \coloneqq (\underline{u}_0^*, \ldots, \underline{u}_{N-1}^*)$ is desired, such that (2) is minimized for the initial state $\underline{x}_0$, that is,

$$\pi^* \coloneqq \arg\min_{\pi_0} V_0^{\pi_0}(\underline{x}_0)\,. \tag{3}$$

Without any additional assumptions, a naive approach would determine π^* in (3) by exhaustive minimization over all policies π_0. With the assumption that the system state satisfies the Markov property, the value function, that is the *minimal* expected cost-to-go, can be recursively calculated by the DP algorithm according to

$$J_N(\underline{x}_N) = g_N(\underline{x}_N)\,, \tag{4}$$

$$J_k(\underline{x}_k) = \min_{\underline{u}_k} \left(g_k(\underline{x}_k, \underline{u}_k) + \mathop{\mathrm{E}}_{\underline{\boldsymbol{w}}_k} \left[J_{k+1}(\underline{\boldsymbol{x}}_{k+1}) \right] \right), \quad k = N-1, \ldots, 0\,, \tag{5}$$

where $\underline{\boldsymbol{x}}_{k+1} = \underline{f}(\underline{x}_k, \underline{u}_k) + \underline{\boldsymbol{w}}_k$. The function $J_k(\underline{x}_k)$ summarizes the minimal expected cost to the terminal state $\underline{x}_N$ starting from state $\underline{x}_k$.

Remark 1: The step costs g_k, $k = 0, \ldots, N$, can be selected in a way, such that the system state attains a desired trajectory $(\underline{x}_1', \ldots, \underline{x}_N')$, when the optimal policy is applied. These functions are often quadratic in the state as well as in the control variable.

The DP algorithm allows for the determination of an optimal policy $(\underline{u}_0^*, \ldots, \underline{u}_{N-1}^*)$ by recursive application of Bellman's principle of optimality to calculate the optimal control inputs

$$\underline{u}_k^* \coloneqq \arg\min_{\underline{u}_k} \left(g_k(\underline{x}_k, \underline{u}_k) + \mathop{\mathrm{E}}_{\underline{\boldsymbol{w}}_k} \left[J_{k+1}(\underline{\boldsymbol{x}}_{k+1}) \right] \right) \tag{6}$$

for $k = N-1, \ldots, 0$. Clearly, the computational effort is lowered compared to the more general formulation (3), since

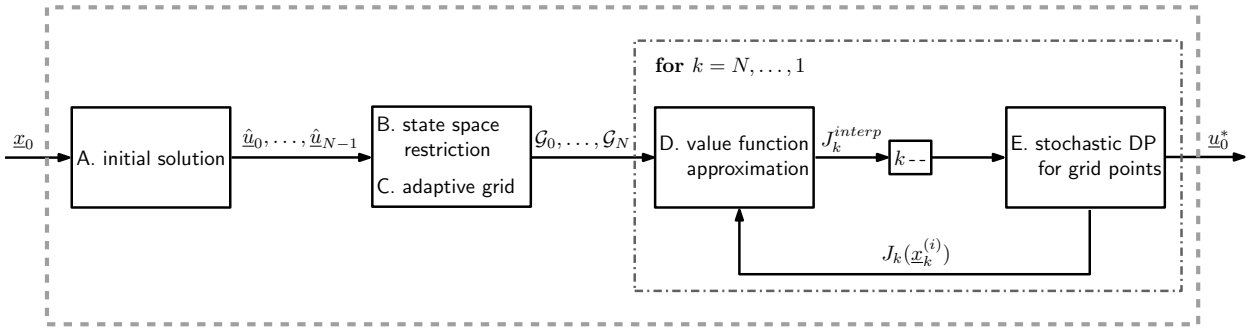


Fig. 1. Block diagram of the proposed controller. For a known state, the 5-step algorithm determines the desired state-feedback control $\underline{u}_0^*$ to be applied in the NMPC framework. An initial solution yields a control trajectory for the considered prediction horizon and serves as a basis for a restriction of the state space. After defining an adaptive grid within this restriction, value function approximation in stochastic DP is employed to yield the desired control input $\underline{u}_0^*$.

the minimization over policies in (3) is turned into recursive minimization over the individual control inputs.

In the considered case of nonlinear systems suffering from additive noise, recursion (5) is given by

$$J_k(\underline{x}_k) = g_k(\underline{x}_k, \underline{u}_k^*) + \int_{\mathbb{R}^{n_x}} \mathrm{p}_k^w\left(\underline{x}_{k+1} - \underline{f}(\underline{x}_k, \underline{u}_k^*)\right) J_{k+1}(\underline{x}_{k+1})\, \mathrm{d}\underline{x}_{k+1} \tag{7}$$

in case of continuous state spaces. p_k^w is the probability density function of the noise vector $\underline{w}_k$, and $\underline{u}_k^*$ is the assumed optimal control input given by (6). In general, the integral in (7) cannot be solved analytically. The complexity of (7) is caused by the nonlinearity of the system function and the noise influence, which requires the expected value.

Suitable approximations are necessary to obtain the desired results, since DP is not directly applicable in continuous-valued state and control spaces.

Example System

Throughout this paper, we use the nonlinear scalar system

$$\boldsymbol{x}_{k+1} = \sqrt{2}\,\sin\Big(\boldsymbol{x}_k + \frac{\pi}{4}\Big) + \frac{\boldsymbol{x}_k}{2} - 1 + u_k + \boldsymbol{w}_k \tag{8}$$

with a zero-mean Gaussian noise variable $\boldsymbol{w}_k$ with variance σ_w^2 for demonstration purposes. ■

III. Online-Computation Approach

Control without noise consideration leads to suboptimal results for the noise-affected system (1). We employ the following assumption to find a sound solution to (6) while avoiding the computational burden of the consideration of the entire state and control spaces.

***Assumption** 1: Given sufficiently smooth system and value functions, a noise-free solution is still in the vicinity of an optimal control incorporating the noise influence.*

This assumption leads to the proposed algorithm consisting of the following main steps to determine the optimal control input for a nonlinear, noisy system with continuous state and control spaces.

A. Determination of an initial (deterministic) solution.

B. Employment of the initial solution to restrict the state space.

C. Definition of an adaptive grid, which covers the considered part of the state space.

Recursion backward in time:

D. Approximation of the DP value function (5).

E. Stochastic dynamic programming within the restricted state space, where the approximated value function is employed and the control space is restricted.

A block diagram of a controller employing the proposed algorithm is depicted in Figure 1. For a known input $\underline{x}_0$, the controller determines an initial approximate solution $(\hat{\underline{u}}_0, \ldots, \hat{\underline{u}}_{N-1})$ as described in Section III-A. Using this control sequence, the state space is restricted depending on the mean values and the covariance matrices of the predicted successor states, which is described in Section III-B. Moreover, adaptive grids $\mathcal{G}_k$, $k = 0, \ldots, N$, are determined within this restriction, which is explained in Section III-C. Interleaving value function approximation along the grid points of $\mathcal{G}_k$, which yields the functions J_k^{interp}, and stochastic dynamic programming results in the expected optimal state-feedback control $\underline{u}_0^*$ for the state $\underline{x}_0$ as described in more detail in Sections III-D and III-E, respectively.

Remark 2: Only the control input $\underline{u}_0^*$ for the current system state $\underline{x}_0$ is required, since the employment of NMPC is proposed to approximately treat infinite-horizon problems. New state information is obtained after each time step, and the whole algorithm is repeated.

Several methods are conceivable to execute either of the steps A–E. In the following, a set of especially well-suited methods is described, based on which the proposed algorithm is evaluated in Section IV.

A. Initial Solution

In the proposed approach, a good candidate for the optimal policy for a corresponding deterministic system

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, \underline{u}_k) \tag{9}$$

is employed. Using this result, an *initial* solution to the original optimal control problem for a finite decision-making

horizon is found as described in [15]. There, the value function (5) of the stochastic dynamic programming algorithm is approximated by means of Taylor series expansion up to second order to simplify the problem. The approximation serves as a basis for the derivation of a stochastic minimum principle for the discrete-time case, where the properties of a stochastic Hamiltonian are employed. Using these theoretical results, the optimal control problem is reformulated as a two-point boundary-value problem. The arising nonlinear equations are solved numerically by means of a continuation method [16]. In [15] the continuation consists of transforming an initial linear system into the original nonlinear system, while the solution to the corresponding (non)linear equation system is being traced. This procedure yields a good candidate for the sequence of optimal state feedbacks of the simplified problem. After all, this control sequence is equivalent to the sequence solving the optimal control problem for (9), although initially a stochastic system was considered. However, according to Assumption 1, this solution can be employed as good prior knowledge in step B to restrict the state space.

In the following, the control inputs of this approximate initial solution are denoted by $\hat{\underline{u}}_0, \ldots, \hat{\underline{u}}_{N-1}$.

B. State Space Restriction

Discretization of state and control spaces is a common approach to apply dynamic programming to continuous-valued problems. If the state space can be restricted in a meaningful way, for instance, if there is knowledge about improbable or impossible system states, discretization can be concentrated there. Typically, this leads to a simplified problem with reduced computational demand.

In the proposed approach, the control sequence $(\hat{\underline{u}}_0, \ldots, \hat{\underline{u}}_{N-1})$ of the deterministic solution is employed to calculate the distributions of the states $\underline{\boldsymbol{x}}_0, \ldots, \underline{\boldsymbol{x}}_N$. After that, the corresponding mean values and covariance matrices are determined. Then, for $k = 0, \ldots, N$, the state space is restricted around the corresponding mean values by defining a symmetric region whose range proportionally depends on the covariance information.

Example System
In case of the scalar example system (8), based on empirical results, we choose the restricted state space to be

$$\left[\mu_k - \frac{3}{2}\sigma_k^x,\, \mu_k + \frac{3}{2}\sigma_k^x\right] \subset \mathbb{R}\,,$$

where μ_k denote the means and σ_k^x the corresponding standard deviations of the predicted states $\boldsymbol{x}_k$ for $k = 0, \ldots, N$. ■

In case of Gaussian noise affecting the system, the extended Kalman filter (EKF) provides a method to obtain the desired values by linearizing the nonlinear system function $\underline{f}$ around the mean value of the system state and subsequent application of the Kalman predictor for linear systems. An alternative approach is provided by the unscented transformation (UT), which is introduced in [17]. The UT determines estimates of the mean value and the covariance matrix of a nonlinearly transformed random variable $\underline{\boldsymbol{x}}$ given by

$$\underline{\boldsymbol{y}} = \underline{b}(\underline{\boldsymbol{x}})\,. \tag{10}$$

Instead of approximating the nonlinear function $\underline{b}$, which is for example done by the EKF, the *probability density function* of the random variable $\underline{\boldsymbol{x}}$ is approximated with a small fixed number of samples. These samples are individually transformed by the original function $\underline{b}$. The accuracy of the UT is superior to that of the EKF, while the computational efforts of the UT and the EKF are of the same order [18]. A more sophisticated approach to determine the desired estimates of the means and covariance matrices is given in [19], where a closed-form prediction for nonlinear, time-invariant systems is introduced, which provides more accurate predictions. This method suffers from higher computational cost, while the complexity of the predicted density stays constant over time.

In the present approach, the UT is employed to determine the desired mean values and covariance matrices of the successor states of $\underline{x}_0$ for the whole prediction horizon. This method results in a reasonable tradeoff between accuracy and computational effort. In the considered case, (10) is given by

$$\underline{\boldsymbol{x}}_{k+1} = \underline{f}(\underline{\boldsymbol{x}}_k, \hat{\underline{u}}_k) + \underline{\boldsymbol{w}}_k\,,\ k = 0, \ldots, N-1\,,$$

that is, the control inputs $\hat{u}_0, \ldots, \hat{u}_{N-1}$ of the initial solution from Section III-A are employed to predict the system state by means of (1). The incorporation of the noise term $\underline{\boldsymbol{w}}_k$ in the UT is treated as described in [20].

C. Definition of an Adaptive Grid

To approximate the DP value function (5) within the restricted part of the state space, grid (support) points have to be determined that cover the range of the considered part of the state space. Therefore, for $k = 0, \ldots, N$, a symmetric set $\mathcal{P}_k$ of $2p+1$ points around the mean value $\underline{\mu}_k$ can be heuristically determined, where p depends on the state dimension n_x. The set $\mathcal{P}_k$ depends on the covariance matrix $\mathbf{C}_k^x$ of the random variable $\underline{\boldsymbol{x}}_k$ through a function $\underline{s}$, that is,

$$\mathcal{P}_k := \left\{\underline{\mu}_k,\, \underline{\mu}_k \pm \underline{s}(\mathbf{C}_k^x, i),\, i = 1, \ldots, p\right\}\,. \tag{11}$$

Depending on the uncertainty of the random variable $\underline{\boldsymbol{x}}_k$, the sets $\mathcal{P}_k$ discretize the state space around the mean values $\underline{\mu}_k$ for $k = 0, \ldots, N$. The more uncertain the k-step prediction is, the wider the area becomes that is covered by $\mathcal{P}_k$. Employing Assumption 1 that the true, but unknown, trajectory is located within the considered region of the state space, the function $\underline{s}$ in (11) can be chosen such that the concentration of the grid points is higher around the mean values to improve the quality of the solution in the vicinity of the predicted trajectory.

Example System
In case of system (8), we define the sets $\mathcal{P}_k$, $k = 0, \ldots, N$, as

$$\mathcal{P}_k := \left\{\mu_k,\, \mu_k \pm \frac{3}{8}\sigma_k^x,\, \mu_k \pm \frac{3}{4}\sigma_k^x,\, \mu_k \pm \frac{3}{2}\sigma_k^x\right\}$$

with a higher concentration of the grid points in the vicinity of the predicted mean values μ_k. Thus, approximately the

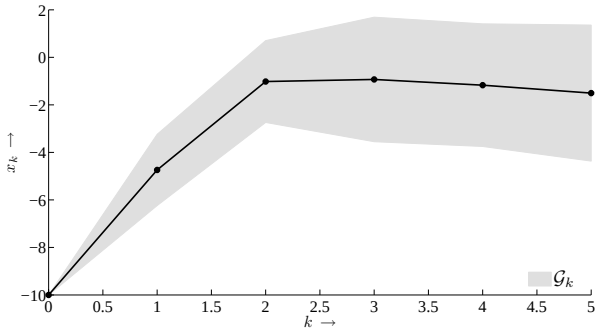


Fig. 2. Possible state space restriction. With the initial deterministic solution, the restrictions of the state space around the mean values of the successor states of x_0 are determined based on the corresponding variances.

same probability mass of a Gaussian is covered between neighboring grid points . ■

Instead of using the original set $\mathcal{P}_k$ as described in (11), a slightly modified set

$$\mathcal{G}_k := \left\{ \underline{x}_k^{(0)}, \ldots, \underline{x}_k^{(2p)} \right\} \tag{12}$$

can be determined to incorporate important properties of the system. A point $\underline{x}'_k$ of the desired state trajectory, which is implicitly defined through the step cost g_k in the DP algorithm, is substituted for one grid point of $\mathcal{P}_k$. The grid point to be substituted is the nearest neighbor of $\underline{x}'_k$. This modification assures exact consideration of the desired states. To maintain the symmetry of the grid point set, the symmetric equivalents of the substituted grid point are replaced, too.

Remark 3: It is important to note that the sets $\mathcal{G}_k$, $k = 0, \ldots, N$, cover the same parts of the state space as the sets $\mathcal{P}_k$, if no extremal point of $\mathcal{P}_k$ is replaced. Furthermore, the number of points does not change. In numerous simulations the employment of the modified grid $\mathcal{G}_k$ led to better results, since the desired points $\underline{x}'_k$ are explicitly included as knots of the subsequent interpolation scheme. The replacement is applied as soon as $\underline{x}'_k$ is in the scope of $\mathcal{P}_k$.

Example System

A possible state space restriction given by $\mathcal{G}_k$ for the scalar example system (8) is depicted in Figure 2. The black dots denote the sequence of the predicted mean values. The shaded area covers the range of the sets $\mathcal{G}_k$ for $k = 0, \ldots, N$ with $N = 5$. ■

D. Value Function Approximation

Dynamic programming on the grid $\mathcal{G}_k$ requires numerous grid points to achieve good accuracy. Because of the "curse of dimensionality", this procedure is not applicable in general. Therefore, interpolation using a small number of grid points is a promising approach to value function approximation. Then, good accuracy can be achieved in the DP algorithm, since DP is not restricted to a discrete set of points. Using value function interpolation, the "curse of dimensionality" cannot be removed, but noticeably reduced. The selection of appropriate interpolating functions possesses many degrees of freedom. In this paper we propose to piecewise interpolate the value function (5) by means of cubic splines within the range of the elements of the sets $\mathcal{G}_k$, $k = 0, \ldots, N$.

In order to be able to solve (7) in closed form, only the properties of interpolating polynomials are considered in the following. In case of higher-degree interpolating polynomials, oscillations tend to occur, and optimization becomes a serious problem. Because of that, piecewise defined lower-degree polynomials, that is, linear, quadratic [21], or cubic [22] functions, are often exploited to sidestep these problems. A common assumption in model-based control is the twice differentiability of the value function J_k in the Bellman equation (5) in the dynamic programming algorithm. For instance, higher-order Taylor series approximation of J_k requires at least second-order derivatives [6], [15]. To maintain this property, interpolating polynomials of at least third degree are required. In the scalar case, cubic polynomials allow for an analytical solution to the optimization problem [23] and represent a reasonable tradeoff between interpolation quality and function complexity.

Summarizing the discussed points, we conclude that with the employment of interpolating cubic splines, stochastic DP can be efficiently applied to solve the continuous-valued optimal control problem approximately.

E. Stochastic Dynamic Programming

Compared with the initial control sequence, an improving approximate solution to the considered finite-horizon optimal control problem with the incorporation of noise is obtained by stochastic DP within the restricted state space. There, the DP value function (5) is recursively approximated by interpolating the grid points of the sets $\mathcal{G}_k$, $k = N, \ldots, 0$.

Employing the assumption that J_{k+1} is already given by a continuous approximation, the aim is to obtain a similar description of J_k depending on J_{k+1}, such that the DP recursion can be applied. For $k+1 = N$, the value function is given by the terminal cost (4), which is independent of the control variable and, therefore, known.

At each time step, a restriction of the control space is determined by considering the sets $\mathcal{U}_k^{(i)}$ for all grid points $\underline{x}_k^{(i)}$, $i = 0, \ldots, 2p$, at time step k, where

$$\mathcal{U}_k^{(i)} := \left\{ \underline{u}_k^{(i,j)} : \mathop{\mathrm{E}}_{\underline{w}_k} \left[\underline{f}(\underline{x}_k^{(i)}, \underline{u}_k^{(i,j)}) + \underline{\boldsymbol{w}}_k \right] = \underline{x}_{k+1}^{(j)}, \; j = 0, \ldots, 2p \right\}. \tag{13}$$

The set $\mathcal{U}_k^{(i)}$ comprises the discrete set of control inputs that map $\underline{x}_k^{(i)}$ at time step k onto the grid points $\underline{x}_{k+1}^{(j)}$ at time step $k+1$. Employing a control action $\underline{u}_k^{(i,j)} \in \mathcal{U}_k^{(i)}$, the expected cost-to-go starting from a specific grid point $\underline{x}_k^{(i)}$ via the expected successor state $\underline{x}_{k+1}^{(j)}$ can be computed by

$$\begin{aligned} V_k(\underline{x}_k^{(i)}, \underline{u}_k^{(i,j)}) &:= g_k(\underline{x}_k^{(i)}, \underline{u}_k^{(i,j)}) + \mathop{\mathrm{E}}_{\underline{w}_k} \left[J_{k+1}(\underline{\boldsymbol{x}}_{k+1}^{(j)}) \right] \\ &= g_k(\underline{x}_k^{(i)}, \underline{u}_k^{(i,j)}) + \min_{\underline{u}_{k+1}, \ldots, \underline{u}_{N-1}} V_{k+1}^{\pi_{k+1}}(\underline{\boldsymbol{x}}_{k+1}^{(j)}), \end{aligned} \tag{14}$$

where the function J_{k+1} is known. For a fixed state $\underline{x}_k^{(i)}$, the computation of (14) for $\underline{u}_k^{(i,j)} \in \mathcal{U}_k^{(i)}$ yields the expected cost-to-go $V_k(\underline{x}_k^{(i)}, \underline{u}_k^{(i,j)})$. Subsequent interpolation of $V_k(\underline{x}_k^{(i)}, \underline{u}_k^{(i,j)})$ along the corresponding knots $\underline{u}_k^{(i,j)} \in \mathcal{U}_k^{(i)}$ yields a cost function $V_k^{interp}(\underline{x}_k^{(i)}, \underline{u}_k)$ that is continuous in the control variable. In case of piecewise interpolation by means of cubic splines, the optimization problem

$$J_k(\underline{x}_k^{(i)}) = \min_{\underline{u}_k} \left(V_k^{interp}(\underline{x}_k^{(i)}, \underline{u}_k) \right) \tag{15}$$

for one specific grid point $\underline{x}_k^{(i)} \in \mathcal{G}_k$ at time step k can be solved analytically. The same calculation is performed for all grid points in $\mathcal{G}_k$. Subsequent interpolation with respect to the grid points $\underline{x}_k^{(i)} \in \mathcal{G}_k$ yields an approximated value function $J_k^{interp}(\underline{x}_k)$ within the range of the state space restriction given by $\mathcal{G}_k$. Therefore, the desired continuous approximation of J_k depending on J_{k+1} is given by J_k^{interp} and can be employed in the dynamic programming algorithm.

In contrast to non-restricted spaces, the approximation by means of interpolating cubic splines is typically more accurate and, therefore, yields better results when using the sets $\mathcal{G}_k$, $k = 0, \ldots, N$.

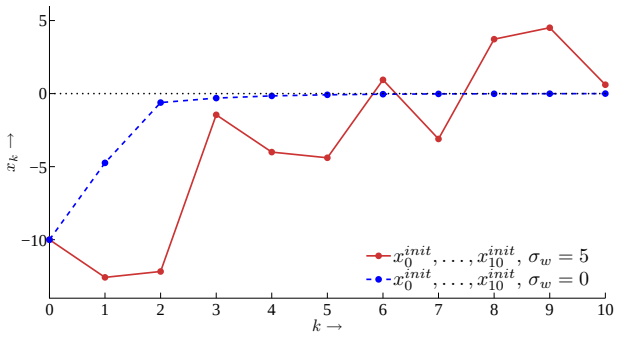
F. Properties and Limitations

After each time step, the proposed algorithm determines an initial control sequence as described in Section III-A and the resulting state space restriction with the corresponding adaptive grid as explained in Sections III-B and III-C. Because of this online computation, it is possible to incorporate current knowledge of the system state. Owing to the spline interpolation in Section III-D, the required number of grid points is dramatically reduced, and good accuracy of the DP algorithm is achieved. The first cubic spline interpolation of the discrete control inputs yields a continuous function that can be minimized analytically, which results in (15). The second interpolation of the value function along the states $\underline{x}_k^{(i)} \in \mathcal{G}_k$ yields a continuous expected minimal cost function within the considered restricted state space. However, the algorithm depends on the quality of the initial solution, a sufficiently good restriction of the state space, and the accuracy of the interpolation scheme employed. Moreover, the proposed algorithm only determines a local, suboptimal solution, if Assumption 1 does not hold.

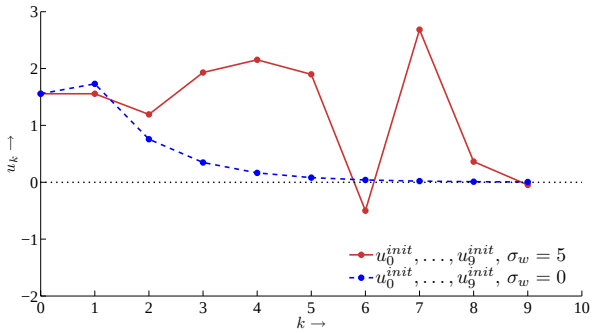
Several methods implementing the steps A–E of the algorithm can be employed to adapt the algorithm to specific problems, since the algorithm formulation in this paper is very general. With the chosen setting, a reasonable tradeoff between computational demand and accuracy of the solution is found as illustrated by simulations in the next section.

IV. Simulation Results

The proposed algorithm is evaluated by means of the scalar example system (8). NMPC is applied for 10 time steps with a prediction horizon of $N = 5$ time steps. Simulations are executed for initial values $x_0 \in \mathcal{X} := \{-10, -8, \ldots, 10\}$. The zero-mean noise variable $\boldsymbol{w}_k$ is Gaussian distributed



(a) State trajectories.



(b) Corresponding control trajectories.

Fig. 3. Example state and control trajectories for system (8). Comparing the trajectories for the deterministic system, that is, $\sigma_w = 0$, and the system affected by noise with standard deviation $\sigma_w = 5$, the differences are obvious.

with standard deviation $\sigma_w = 5$. The cost functions are given by

$$g_N(x_N) = \frac{1}{2} x_N^2 \,,$$
$$g_k(x_k, u_k) = \frac{1}{2}(x_k^2 + 2u_k^2) \,, \quad k = N-1, \ldots, 0 \,,$$

to attain the unstable equilibrium point 0 as rapidly as possible. Therefore, the implicitly encoded trajectory $(x'_1, \ldots, x'_N)$ introduced in Remark 1 is set to $(0, \ldots, 0)$.

As described in Section III-D, the value function is interpolated by means of piecewise defined cubic splines. Therefore, the solution to the integral in (7) reduces to an integral over the product of a Gaussian and a cubic polynomial. This solution can be analytically determined by using the first moments of the Gaussian and the error function, since the integral is restricted to the finite domains of the defining spline pieces.

In the following, the value functions J_k, the system states x_k, and the control inputs u_k for the initial solution of Section III-A are denoted by J_k^{init}, x_k^{init}, and u_k^{init}, respectively. The corresponding values for the whole proposed algorithm using the spline interpolation of the value function are denoted by J_k^{spline}, x_k^{spline}, and u_k^{spline}, respectively.

A. Noise Influence

To emphasize the noise influence on system (8), an example state trajectory and the corresponding control trajectory

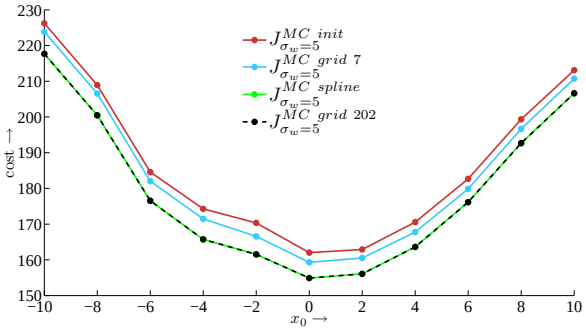


Fig. 4. MC estimates of the expected minimal cost for different controllers.

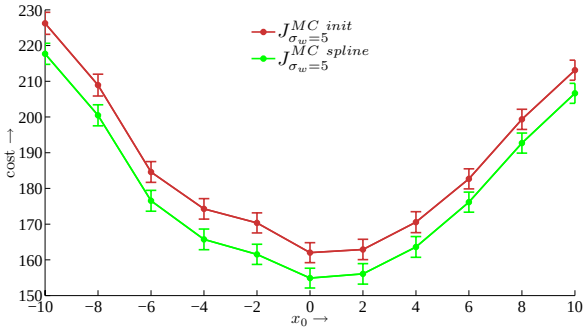


Fig. 5. 2-σ_s error bars for the sample mean values of the initial deterministic and the spline-based algorithms.

are depicted in Figure 3 for the initial solution in case of a deterministic system with $\sigma_w = 0$ and the considered noise-affected system with $\sigma_w = 5$. The deviation of the state trajectories between the deterministic and the stochastic systems in Figure 3(a) is caused by the noise influence. The corresponding control trajectories are given in Figure 3(b). Clearly, the considered system suffers from relatively strong noise disturbances, which motivates the consideration of noise in the controller design.

B. Quality of the Proposed Methods

In the following, the arising costs are compared to analyze the quality of the control sequence resulting from the application of the proposed methods.

For $\sigma_w > 0$ the costs change with each simulation. A Monte Carlo (MC) simulation provides an approximate upper bound and, thus, an estimate $J^{MC}_{\sigma_w=5}$ of the true value function depending on the noise standard deviation $\sigma_w = 5$ by calculating the arithmetic mean of all costs after 2816 simulations starting from each $x_0 \in \mathcal{X}$.

Four different controllers C^{init}, C^{spline}, $C^{grid\ 7}$, and $C^{grid\ 202}$ are considered. C^{init} and C^{spline} provide the controllers resulting from the work of [15] and the proposed improvement explained in Section III, respectively. $C^{grid\ 7}$ employs the same grid points as C^{spline}, but does *not* interpolate the value function by means of cubic splines. Therefore, the set of controls is also discretized and given by a 7-elements set. The employment of $C^{grid\ 202}$ with 202 possible states and controls within the restricted optimization space per time step is computationally very demanding, but is expected to be the best controller because of the relatively finely discretized state and control sets.

In Figure 4 the MC estimates of the expected minimal cost functions resulting from the applications of all controllers are depicted. In each simulation the controllers $C^{init}, \dots, C^{grid\ 202}$ suffer from the same noise vector, which explains the similar structure of the MC estimates. The first striking observation is that $J^{spline}_{\sigma_w=5}$ and $J^{MC\ grid\ 202}_{\sigma_w=5}$ are indistinguishable. Therefore, the proposed approach yields results of the same quality as $C^{grid\ 202}$, which employs finely discretized state and control spaces. As shown in Table I, however, the computational demand of C^{spline} is clearly lower than the one of $C^{grid\ 202}$. In Table I the relative computation times of the corresponding MATLAB codes are given. The computation time of the initial solution is set to 1, since it is employed in all other controllers to restrict the state space. Comparing the expected cost resulting from the initial solution, it is concluded that the first aim is achieved, that is, C^{spline} noticeably outperforms C^{init}, where the additional computation effort is still acceptable.

TABLE I
RELATIVE COMPUTATIONAL EFFORT OF THE CONTROLLERS.

C^{init}	C^{spline}	$C^{grid\ 7}$	$C^{grid\ 202}$
1	1.80	1.04	33.44

Remark 4: In the current setting the improvement is not only noticeable, but statistically significant. Considering the values $J^{MC\ init}_{\sigma_w=5}(x_0)$ and $J^{MC\ spline}_{\sigma_w=5}(x_0)$ for $x_0 \in \mathcal{X}$ as estimates of the mean values of the corresponding minimal cost functions (depending on the different controllers), a one-tailed t-test is performed to determine the significance level of the difference of these estimates. The p-values yield the result that the hypothesis of identical mean values has to be rejected with an error probability of less than 0.0010 along the sample points $x_0 \in \mathcal{X}$.

In the chosen setting the average improvement by C^{spline} is approximately 4%. The error bars along the sample mean values $J^{MC\ init}_{\sigma_w=5}(x_0)$ and $J^{MC\ spline}_{\sigma_w=5}(x_0)$ are given in Figure 5. The error bars desribe the 2-σ_s interval around the sample means, where σ_s is the standard error.

Simulations with smaller noises or other systems led to results similar to the ones described above.

V. CONCLUSIONS AND FUTURE WORK

In this paper a novel online-computation approach to optimal control of nonlinear, noise-affected systems with continuous state and control spaces is presented. At each time step, the algorithm is initialized with a candidate of the finite-horizon open-loop solution to the optimal control problem of a corresponding deterministic system. The noise is explicitly incorporated into the control in a post-processing algorithm comprising the following steps. Using the initial solution, estimates of the means and covariance matrices of the predicted successor states of the known initial state are obtained for the entire finite decision-making horizon.

In the vicinity of the means, the state space is restricted depending on the uncertainties of the successor states. Within this restriction, an improved solution is found by interleaving stochastic dynamic programming and value function approximation. Therefore, continuous state and control spaces are treated approximately. With the methods employed in the post-processing algorithm, a reasonable tradeoff between accuracy and computational effort is achieved.

The application of nonlinear model predictive control in the simulation of a scalar example system resulted in a noticeable and statistically significant improvement over the initial deterministic solution by using the proposed algorithm. Moreover, value function interpolation by means of cubic splines yielded better results than a purely grid based DP approach. Finally, the accuracy of the proposed approach is of the same quality as a computationally very demanding DP approach that discretizes the restricted part of the state space with a huge number of grid points.

Several points merit further investigation in future work. The proposed algorithm allows the employment of arbitrary methods to apply the main steps mentioned in Section III. Depending on a concrete application, other methods than the currently implemented ones might be preferable. As mentioned in Section III-D, solutions to the function approximation problem are desired, such that the expected value can be computed analytically. The true underlying function itself is, of course, not restricted to low-order polynomials, but is presumably covered by other function classes. With Gaussian approximations, the integral in the expected value (7) is analytically solvable. This property is exploited in the Gaussian process framework. Here, the Gaussian approximation holds pointwise for the function values. The employment of GPs to multi-step ahead prediction is for instance covered in [24], where the mean and the uncertainty information are predicted. The application of this method to state space restriction is of high interest. Therefore, the application of the GP framework in the proposed algorithm is worth being evaluated in future. The application of the proposed approach to a real experiment is a straightforward step to evaluate its practical use.

ACKNOWLEDGEMENTS

M. P. Deisenroth was supported in part by the German Research Foundation (DFG) through grant RA 1030/1.

The authors thank Jason Farquhar for proof-reading and helpful suggestions.

REFERENCES

[1] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed., ser. Optimization and Computation Series. Belmont, MA, U.S.A.: Athena Scientific, 2000, vol. 1.

[2] M. Diehl, *Real-Time Optimization for Large Scale Nonlinear Processes*, ser. Reihe 8: Mess-, Steuerungs- und Regelungstechnik. Düsseldorf: VDI Verlag GmbH, 2002, no. 920.

[3] R. E. Bellman, *Dynamic Programming*. Princeton, New Jersey, U.S.A.: Princeton University Press, 1957.

[4] T. Ohtsuka and H. Fujii, "Stabilized Continuation Method for Solving Optimal Control Problems," *Journal on Guidance, Control, and Dynamics*, vol. 17, pp. 950–957, November 1994.

[5] T. Ohtsuka, "A Continuation/GMRES Method for Fast Computation of Nonlinear Receding Horizon Control," *Automatica*, vol. 40, no. 4, pp. 563–574, 2004.

[6] V. Rico-Ramirez and U. M. Diwekar, "Stochastic Maximum Principle for Optimal Control under Uncertainty," *Computers & Chemical Engineering*, vol. 28, no. 12, pp. 2845–2849, November 2004.

[7] H. J. Kappen, "Linear Theory for Control of Nonlinear Stochastic Systems," *Physical Review Letters*, vol. 95, no. 20, p. 200201, November 2005.

[8] ——, "Path Integrals and Symmetry Breaking for Optimal Control Theory," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, p. P11011, November 2005.

[9] D. Nikovski and M. Brand, "Non-Linear Stochastic Control in Continuous State Spaces by Exact Integration in Bellman's Equations," in *13th International Conference on Automatic Planning & Scheduling (ICAPS '03)*, Trento, Italy, June 2003, pp. 91–95.

[10] F. Weissel, M. F. Huber, and U. D. Hanebeck, "A Closed-Form Model Predictive Control Framework for Nonlinear Noise Corrupted Systems," in *Proceedings of the 4th International Conference on Informatics in Control, Automation, and Robotics (ICINCO 2007)*, Angers, France, May 2007.

[11] ——, "Efficient Control of Nonlinear Noise Corrupted Systems Using a Novel Model Predictive Control Framework," in *Proceedings of the 2007 American Control Conference (ACC 2007)*, New York City, U.S.A., July 2007.

[12] C. E. Rasmussen and M. Kuss, "Gaussian Processes in Reinforcement Learning," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. K. Saul, and B. Schölkopf, Eds. Cambridge, MA, U.S.A.: The MIT Press, June 2004, pp. 751–759.

[13] J. Kocijan, R. Murray-Smith, C. E. Rasmussen, and A. Girard, "Gaussian Process Model Based Predictive Control," in *Proceedings of the 2004 American Control Conference (ACC 2004)*, Boston, MA, U.S.A., June–July 2004, pp. 2214–2219.

[14] J. Kocijan, R. Murray-Smith, C. E. Rasmussen, and B. Likar, "Predictive Control with Gaussian Process Models," in *Proceedings of IEEE Region 8 Eurocon 2003: Computer as a Tool*, B. Zajc and M. Tkalčič, Eds., Piscataway, NJ, U.S.A., September 2003, pp. 352–356.

[15] M. P. Deisenroth, T. Ohtsuka, F. Weissel, D. Brunn, and U. D. Hanebeck, "Finite-Horizon Optimal State Feedback Control of Nonlinear Stochastic Systems Based on a Minimum Principle," in *Proceedings of the 6th IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, Heidelberg, Germany, September 2006, pp. 371–376.

[16] S. L. Richter and R. A. DeCarlo, "Continuation Methods: Theory and Applications," in *IEEE Transactions on Automatic Control*, vol. AC-28, no. 6, June 1983, pp. 660–665.

[17] S. Julier and J. K. Uhlmann, "A General Method for Approximating Nonlinear Transformations of Probability Distributions," Robotics Research Group, Department of Engineering Science, University of Oxford, Oxford, OC1 3PJ United Kingdom, Tech. Rep., November 1996.

[18] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," in *Proceedings of the IEEE*, vol. 92, no. 3, March 2004.

[19] M. Huber, D. Brunn, and U. D. Hanebeck, "Closed-Form Prediction of Nonlinear Dynamic Systems by Means of Gaussian Mixture Approximation of the Transition Density," in *Proceedings of the 6th IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, Heidelberg, Germany, September 2006, pp. 98–103.

[20] K. Xiong, H.-Y. Zhang, and C. W. Chan, "Performance Evaluation of UKF-Based Nonlinear Filtering," *Automatica*, vol. 42, pp. 261–270, 2006.

[21] L. L. Schumaker, "On Shape-Preserving Quadratic Spline Interpolation," *SIAM Journal on Numerical Analysis*, vol. 20, no. 4, pp. 854–864, August 1983.

[22] S. A. Johnson, J. R. Stedinger, C. A. Shoemaker, Y. Li, and J. A. Tejada-Guibert, "Numerical Solution of Continuous-State Dynamic Programs Using Linear and Spline Interpolation," *Operations Research*, vol. 41, no. 3, pp. 484–500, May–June 1993.

[23] S. Bosch, *Algebra*, 6th ed. Springer-Verlag, 2006.

[24] M. Kuss, "Gaussian Process Models for Robust Regression, Classification, and Reinforcement Learning," Ph.D. dissertation, Technische Universität Darmstadt, Germany, February 2006.