Finite-Horizon Optimal State-Feedback Control of Nonlinear Stochastic Systems Based on a Minimum Principle

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

***Abstract*— In this paper, an approach to the finite-horizon optimal state-feedback control problem of nonlinear, stochastic, discrete-time systems is presented. Starting from the dynamic programming equation, the value function will be approximated by means of Taylor series expansion up to second-order derivatives. Moreover, the problem will be reformulated, such that a minimum principle can be applied to the stochastic problem. Employing this minimum principle, the optimal control problem can be rewritten as a two-point boundary-value problem to be solved at each time step of a shrinking horizon. To avoid numerical problems, the two-point boundary-value problem will be solved by means of a continuation method. Thus, the curse of dimensionality of dynamic programming is avoided, and good candidates for the optimal state-feedback controls are obtained. The proposed approach will be evaluated by means of a scalar example system.**

I. Introduction

Optimal control of nonlinear stochastic systems is still a challenging research field. One very useful approach to treat this problem in case of discrete problems is dynamic programming, which exploits Bellman's principle of optimality [1]. But even in the discrete case, dynamic programming suffers from the curse of dimensionality. Moreover, an analytical solution cannot be found in general [2] and numerical methods have to be employed.

Starting from the dynamic programming equation, Pontryagin's maximum principle offers necessary optimality conditions in case of deterministic systems. These conditions can be employed to reformulate the optimal control problem as a two-point boundary-value problem (TPBVP) that is numerically solvable. In case of stochastic, that is noise affected, systems, this approach is not directly applicable due to the fact that it is not possible to calculate a deterministic prediction of the system state. Thus, the determination of an optimal state-feedback law for nonlinear stochastic systems requires nonlinear optimization and stochastic state propagation.

Several approaches to obtain approximate solutions to the optimal control problem for nonlinear stochastic systems with continuous state spaces can be found in the literature. For an infinite horizon an approximation scheme of the value function by means of radial-basis functions is proposed in [3], which leads to a discretization of the problem. Based on the assumption of an underlying Ito-process, Pontryagin's maximum principle is extended to continuous-time stochastic systems in [4].

An expedient approach to solve the resulting TPBVP numerically is to employ a continuation method [5]. Thereby, a solution to an easily solvable initial problem can be calculated. While the initial problem is being continuously transformed into the original problem, the solution is being traced. In case of optimal control problems of nonlinear systems, a related linear system can be employed to initialize the continuation method. For a deterministic continuous-time system this idea has been successfully applied [6].

Nevertheless, an equivalent to the maximum principle or the TPBVP for stochastic systems in the technically important discrete-time case has not been found in the literature yet. In this work an approach is proposed that employs the idea of dynamic programming to find an approximate solution to the optimal nonlinear stochastic control problem. Starting from the dynamic programming equation, the stochastic problem will be reformulated, such that a minimum principle can be employed. Using this approach, a TPBVP will be derived and solved by means of a continuation method.

The remainder of this paper is structured as follows. In Section II the problem will be formulated. In Section III an approximation of the stochastic problem by employing Taylor series expansion of the value function up to second-order derivatives will be presented. Furthermore, the application of a minimum principle to the stochastic system is described. Section IV deals with the reformulation of the optimal control problem as a TPBVP, which is numerically solved by means of a continuation method. Moreover, the whole algorithm is described in more detail. In Section V the proposed approach will be evaluated by means of a scalar example system. Section VI summarizes the results of this paper and gives a survey of future work.

II. Problem Formulation

Exploiting the Markov property, dynamic programming is a common approach to solve nonlinear optimal closed-loop control problems by means of backward recursion in case of an additive cost function [2]. These assumptions will also be employed in the following.

M. P. Deisenroth is with the Department of Mechanical Engineering, Graduate School of Engineering, Osaka University, Japan, and with the Intelligent Sensor-Actuator-Systems Laboratory, Institute of Computer Science and Engineering, Universität Karlsruhe (TH), Germany. marc@newton.mech.eng.osaka-u.ac.jp

T. Ohtsuka is with the Department of Mechanical Engineering, Graduate School of Engineering, Osaka University, Japan. ohtsuka@mech.eng.osaka-u.ac.jp

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

Let the considered discrete-time system be given by

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

where $\underline{\boldsymbol{x}}_k \in \mathbb{R}^N$ denotes the system state, $\underline{u}_k \in \mathbb{R}^M$ the control law variable, and $\underline{f} : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}^N$ a nonlinear function. $\underline{\boldsymbol{w}}_k \in \mathbb{R}^N$ is an independent additive zero-mean noise term with covariance $\boldsymbol{\Sigma}_w$. $\hat{\underline{x}}_0$ is assumed to be known.

Throughout this work $\underline{\boldsymbol{x}}$ denotes a random variable, and $\hat{\underline{x}}$ is a concrete realization of the variable $\underline{x}$.

The objective is to establish an optimal control law $\underline{\mu}_k^*$ that maps the states $\underline{x}_k$ onto optimal controls $\underline{u}_k^* = \underline{\mu}_k^*(\underline{x}_k)$. Therefore, a value function has to be introduced. In case of stochastic systems, an apparent approach is to define the value function at time step k as the minimal expected cost-to-go from time step k to N, that is

$$J_k(\underline{x}_k) := \min_{\underline{u}_k} g_k(\underline{x}_k, \underline{u}_k) + \mathop{\mathrm{E}}_{w_k}\left[J_{k+1}(\underline{\boldsymbol{x}}_{k+1})\right] \ , \quad (2)$$

where $g_k(\underline{x}_k, \underline{u}_k)$ denotes the step cost from time k to $k+1$ depending on the current system state and the applied control action. $J_{k+1}(\underline{\boldsymbol{x}}_{k+1})$ summarizes the minimal expected cost to the terminal state $\underline{\boldsymbol{x}}_N$ starting from state $\underline{\boldsymbol{x}}_{k+1}$, which is obtained by (1) when $\underline{u}_k$ and $\underline{x}_k$ are given. The terminal cost

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

is independent of the state-feedback. Dynamic programming determines the optimal state-feedback control $\underline{u}_k^*$ for each time step instead of performing the minimization over all policies $\pi_k = (\underline{u}_k, \dots, \underline{u}_{N-1})$.

III. Application of the Minimum Principle

In the following, the stochastic minimization problem will be reformulated, such that a minimum principle can be applied. To simplify the derivation, the optimal values $\underline{u}_0^*, \dots, \underline{u}_{N-1}^*$ are assumed to be given at this point. The minimization problem itself will be treated in Section IV.

A. Approximation of the Value Function

Approximating J_{k+1} in (2) by means of Taylor series expansion up to second-order derivatives around the deterministic part of the state $\underline{\boldsymbol{x}}_{k+1}$ yields

$$\begin{aligned} J_{k+1}(\underline{\boldsymbol{x}}_{k+1}) \approx J_{k+1}(\underline{f}(\underline{x}_k, \underline{u}_k)) + \frac{\partial J_{k+1}(\underline{f}(\underline{x}_k, \underline{u}_k))}{\partial \underline{x}_{k+1}} \underline{\boldsymbol{w}}_k \\ + \frac{1}{2} \underline{\boldsymbol{w}}_k^T \boldsymbol{H}_{k+1}(\underline{f}(\underline{x}_k, \underline{u}_k)) \underline{\boldsymbol{w}}_k \ , \end{aligned} \quad (3)$$

where $\boldsymbol{H}_k$ denotes the Hesse matrix of the value function. Third- and higher-order derivatives are assumed to be negligible. Taking the expectation of (3), the approximation of $J_k(\underline{x}_k)$ in (2) is given by

$$\begin{aligned} J_k(\underline{x}_k) \approx g(\underline{x}_k, \underline{u}_k^*) + J_{k+1}(\underline{f}(\underline{x}_k, \underline{u}_k^*)) \\ + \frac{1}{2} \mathrm{tr}\Big(\boldsymbol{\Sigma}_w \boldsymbol{H}_{k+1}(\underline{f}(\underline{x}_k, \underline{u}_k^*))\Big) \ , \end{aligned} \quad (4)$$

where the property

$$\underline{\boldsymbol{w}}_k^T \boldsymbol{H}_{k+1} \underline{\boldsymbol{w}_k} = \mathrm{tr}\left(\underline{\boldsymbol{w}}_k \underline{\boldsymbol{w}}_k^T \boldsymbol{H}_{k+1}\right)$$

has been exploited. Moreover, the gradient in (3) vanishes due to the expectation value. A recursive calculation of the Hessian will be given in Theorem 3 after defining the Hamilton function.

Remark 1: It is important to mention that (4) is similar to the deterministic value function. The only difference is the last term, which is the contribution to the noise.

Considering (4), the value function J_k and its Hesse matrix $\boldsymbol{H}_k$ are evaluated at states, which would originate from a deterministic state propagation given by

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, \underline{u}_k) \ . \quad (5)$$

Because of that, the state propagation (5) is sufficient to determine the value of the value function, if (4) is employed at each time step. In this case, the expectation value needs not to be considered explicitly, since the additional term in the value function accounts for the noise that affects the system.

Employing Taylor series expansion to approximate the gradient of the value function up to second-order derivatives, the linearized gradient of the value function is given by

$$\begin{aligned} \frac{\partial J_{k+1}(\underline{\boldsymbol{x}}_{k+1})}{\partial \underline{x}_{k+1}} \approx \frac{\partial J_{k+1}(\underline{f}(\underline{x}_k, \underline{u}_k^*))}{\partial \underline{x}_{k+1}} \\ + \frac{\partial^2 J_{k+1}(\underline{f}(\underline{x}_k, \underline{u}_k^*))}{\partial \underline{x}_{k+1}^2} \underline{\boldsymbol{w}}_k \ . \end{aligned} \quad (6)$$

With the approximations (4), (5), and (6) of the stochastic system the minimum principle can be applied to the stochastic system.

B. Minimum Principle

In case of the stochastic system (1), a necessary minimum condition for the value function $J_k(\underline{x}_k)$ is given by

$$\begin{aligned} \frac{\partial J_k(\underline{x}_k)}{\partial \underline{u}_k} &= \frac{\partial g_k(\underline{x}_k, \underline{u}_k^*)}{\partial \underline{u}_k} + \mathop{\mathrm{E}}_{w_k}\left[\frac{\partial J_{k+1}(\underline{\boldsymbol{x}}_{k+1})}{\partial \underline{x}_{k+1}}\right] \frac{\partial \underline{f}(\underline{x}_k, \underline{u}_k^*)}{\partial \underline{u}_k} \\ &= \underline{0}^T \ , \end{aligned} \quad (7)$$

when the chain rule is employed.

In the following, for an optimal sequence of state-feedback controls $(\underline{u}_0^*, \dots, \underline{u}_{N-1}^*)$ minimizing (2) for $k = 0, \dots, N-1$, the corresponding state sequence is denoted by $(\underline{x}_0^*, \dots, \underline{x}_N^*)$ according to the state propagation (5).

Definition 1: The costate is defined as the gradient of the value function evaluated at $\hat{\underline{x}}_k$, that is

$$\underline{p}_k^T := \frac{\partial J_k(\hat{\underline{x}}_k)}{\partial \underline{x}_k} \ . \quad (8)$$

Theorem 1 (Costate Recursion): Employing the approximations (4), (5), and (6), a recursive calculation of the costate along the optimal sequence of state-feedback controls and the corresponding state sequence is given by

$$\underline{p}_N^T := \frac{\partial g_N(\hat{\underline{x}}_N^*)}{\partial \underline{x}_N} \ , \ k = N \ , \quad (9)$$

$$\underline{p}_k^T = \frac{\partial g_k(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{x}_k} + \underline{p}_{k+1}^T \frac{\partial \underline{f}(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{x}_k} \quad (10)$$

for $k = N-1, \dots, 0$.

Proof:
$k = N$:

$$\underline{p}_N^T := \frac{\partial J_N(\hat{\underline{x}}_N^*)}{\partial \underline{x}_N} = \frac{\partial g_N(\hat{\underline{x}}_N^*)}{\partial \underline{x}_N} \ .$$

$k \in \{N-1, \ldots, 0\}$: The dynamic programming equation (2) yields

$$\begin{aligned}\underline{p}_k^T := \frac{\partial J_k(\underline{x}_k^*)}{\partial \underline{x}_k} &= \frac{\partial g_k(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{x}_k} + \frac{\partial g_k(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{u}_k} \frac{\partial \underline{\mu}_k^*(\underline{x}_k^*)}{\partial \underline{x}_k} \\ &+ \mathop{\mathrm{E}}_{w_k}\left[\frac{\partial J_{k+1}(\underline{\boldsymbol{x}}_{k+1})}{\partial \underline{x}_{k+1}} \frac{\partial \underline{f}(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{x}_k}\right] \\ &+ \mathop{\mathrm{E}}_{w_k}\left[\frac{\partial J_{k+1}(\underline{\boldsymbol{x}}_{k+1})}{\partial \underline{x}_{k+1}} \frac{\partial \underline{f}(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{u}_k} \frac{\partial \underline{\mu}_k^*(\underline{x}_k^*)}{\partial \underline{x}_k}\right] \ , \quad (11)\end{aligned}$$

where $\underline{\boldsymbol{x}}_{k+1}$ denotes the one-step prediction by means of the system function (1) starting from the state $\underline{x}_k^*$. Employing the necessary minimum condition (7) for $J_k(\underline{x}_k^*)$, equation (11) can be rewritten as

$$\begin{aligned}\frac{\partial J_k(\underline{x}_k^*)}{\partial \underline{x}_k} &= \frac{\partial g_k(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{x}_k} \\ &+ \mathop{\mathrm{E}}_{w_k}\left[\frac{\partial J_{k+1}(\underline{\boldsymbol{x}}_{k+1})}{\partial \underline{x}_{k+1}}\right] \frac{\partial \underline{f}(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{x}_k} \quad (12)\end{aligned}$$

for $k = N-1, \ldots, 0$. Considering (8), (10), and (12), it remains to show that for a given state $\underline{x}_k^*$

$$\mathop{\mathrm{E}}_{w_k}\left[\frac{\partial J_{k+1}\left(\underline{f}(\underline{x}_k^*, \underline{u}_k^*)\right)}{\partial \underline{x}_{k+1}} + \underline{\boldsymbol{w}}_k)\right] = \frac{\partial J_{k+1}(\underline{x}_{k+1}^*)}{\partial \underline{x}_{k+1}} \quad (13)$$

is satisfied. Because of the assumptions, the states $\underline{x}_k \, , k = 1, \ldots, N$, are calculated by means of (5). Taking the expectation, (6) can be rewritten as (13) and the proof of (10) is concluded. ■

Remark 2: The consideration of higher-order derivatives in (6) would require the existence of an inverse mapping of $\frac{\partial J_{k+1}}{\partial \underline{x}_{k+1}}$ to satisfy (13).

Definition 2 (Stochastic Hamilton Function): To define a stochastic Hamilton function, the influence of noise has to be incorporated. In case of system (1), this leads to the definition

$$H_k(\underline{x}_k, \underline{p}_{k+1}, \underline{u}_k, \underline{\boldsymbol{w}}_k) := g_k(\underline{x}_k, \underline{u}_k) + \underline{p}_{k+1}^T\left(\underline{f}(\underline{x}_k, \underline{u}_k) + \underline{\boldsymbol{w}}_k\right)$$

for $k = N-1, \ldots, 0$.

Theorem 2: Along the optimal sequence of state-feedback controls and the corresponding state sequence, the following properties hold for $k = N-1, \ldots, 0$.

$$\frac{\partial}{\partial \underline{x}_k}\left(H_k(\underline{x}_k^*, \underline{p}_{k+1}, \underline{u}_k^*, \underline{\boldsymbol{w}}_k)\right) = \underline{p}_k^T \ , \quad (14)$$

$$\frac{\partial}{\partial \underline{x}_k}\left(H_k(\underline{x}_k^*, \underline{p}_{k+1}, \underline{u}_k^*, \underline{\boldsymbol{w}}_k)\right) = \frac{\partial J_k(\underline{x}_k^*)}{\partial \underline{x}_k} \ , \quad (15)$$

$$\frac{\partial}{\partial \underline{u}_k}\left(H_k(\underline{x}_k^*, \underline{p}_{k+1}, \underline{u}_k^*, \underline{\boldsymbol{w}}_k)\right) = \underline{0}^T \ . \quad (16)$$

Proof:
Equation (14): Let $k \in \{0, \ldots, N-1\}$. Then,

$$\frac{\partial H_k}{\partial \underline{x}_k} = \frac{\partial g_k(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{x}_k} + \underline{p}_{k+1}^T \frac{\partial \underline{f}(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{x}_k} = \underline{p}_k^T \ .$$

Equation (15): follows immediately from (8) and (14).
Equation (16): Because of the approximation of the gradient of the value function by means of Taylor series expansion given by (6) and the necessary minimum condition (7) for the value function $J_k(\underline{x}_k^*)$, (13) holds. With property (13),

$$\begin{aligned}\frac{\partial H_k}{\partial \underline{u}_k} &= \frac{\partial g_k(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{u}_k} + \underline{p}_{k+1}^T \frac{\partial \underline{f}(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{u}_k} \\ &= \frac{\partial g_k(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{u}_k} + \frac{\partial J_{k+1}(\underline{x}_{k+1}^*)}{\partial \underline{x}_{k+1}} \frac{\partial \underline{f}(\underline{x}_k^*, \underline{u}_k^*)}{\partial \underline{u}_k} \quad (17)\end{aligned}$$

is equivalent to (7), which concludes the proof. ■

Remark 3 (Stochastic Minimum Principle): With the assumptions of Theorem 1, a necessary minimum condition for the considered stochastic system along the optimal control sequence and the corresponding states is given by (16), which can be evaluated by means of the Hamiltonian.

Theorem 3 (Hessian Recursion): The Hesse matrix in (4) can be recursively calculated as follows, where the arguments of the functions are omitted to simplify the readability.

$$\begin{aligned}\boldsymbol{H}_N &:= \frac{\partial^2 g_N}{\partial \underline{x}_N^2} \\ \boldsymbol{H}_k &= \frac{\partial^2 H_k}{\partial \underline{x}_k^2} + \left(\frac{\partial \underline{f}}{\partial \underline{x}_k}\right)^T \boldsymbol{H}_{k+1} \frac{\partial \underline{f}}{\partial \underline{x}_k} \\ &- \left[\left(\frac{\partial \underline{f}}{\partial \underline{x}_k}\right)^T \boldsymbol{H}_{k+1} \left(\frac{\partial \underline{f}}{\partial \underline{u}_k}\right) + \frac{\partial^2 H_k}{\partial \underline{x}_k \partial \underline{u}_k}\right] \\ &\cdot \left[\left(\frac{\partial \underline{f}}{\partial \underline{u}_k}\right)^T \boldsymbol{H}_{k+1} \frac{\partial \underline{f}}{\partial \underline{u}_k} + \frac{\partial^2 H_k}{\partial \underline{u}_k^2}\right]^{-1} \\ &\cdot \left[\frac{\partial^2 H_k}{\partial \underline{u}_k \partial \underline{x}_k} + \left(\frac{\partial \underline{f}}{\partial \underline{u}_k}\right)^T \boldsymbol{H}_{k+1} \frac{\partial \underline{f}}{\partial \underline{x}_k}\right] \quad (18)\end{aligned}$$

for $k = N-1, \ldots, 0$, where $\boldsymbol{H}_k$ denotes the Hesse matrix of the value function and H_k refers to the Hamilton function.

Proof:
$k = N$:

$$\boldsymbol{H}_N = \frac{\partial^2 J_N}{\partial \underline{x}_N^2} = \frac{\partial^2 g_N}{\partial \underline{x}_N^2} \ .$$

$k \in \{N-1, \ldots, 0\}$: Because of (15), the Hessian $\boldsymbol{H}_k$ can be calculated by means of the Hamiltonian along the optimal sequence of controls and the corresponding state sequence. Considering the gradient of H_k as a function of $\underline{x}_k, \underline{p}_{k+1}$, and $\underline{u}_k$ and assuming that $\boldsymbol{H}_{k+1}$ has already been computed, $\boldsymbol{H}_k$ is given as the second partial derivative of the Hamiltonian with respect to $\underline{x}_k$, that is

$$\begin{aligned}\boldsymbol{H}_k &= \frac{\partial^2 H_k}{\partial \underline{x}_k^2} + \frac{\partial^2 H_k}{\partial \underline{x}_k \partial \underline{p}_{k+1}} \boldsymbol{H}_{k+1} \frac{\partial \underline{f}}{\partial \underline{x}_k} \\ &+ \left[\frac{\partial^2 H_k}{\partial \underline{x}_k \partial \underline{p}_{k+1}} \boldsymbol{H}_{k+1} \frac{\partial \underline{f}}{\partial \underline{u}_k} + \frac{\partial^2 H_k}{\partial \underline{x}_k \partial \underline{u}_k}\right] \frac{\partial \underline{\mu}_k^*}{\partial \underline{x}_k} \quad (19)\end{aligned}$$

with unknowns $\frac{\partial^2 H_k}{\partial \underline{x}_k \partial \underline{p}_{k+1}}$ and $\frac{\partial \underline{\mu}_k^*}{\partial \underline{x}_k}$.

$$\frac{\partial^2 H_k}{\partial \underline{x}_k \partial \underline{p}_{k+1}} = \left(\frac{\partial \underline{f}}{\partial \underline{x}_k}\right)^T \quad (20)$$

holds, if (14) and the costate recursion (10) are employed. Since (16) is satisfied for all $\underline{x}_k$,

$$\frac{\partial}{\partial \underline{x}_k}\left(\frac{\partial H_k}{\partial \underline{u}_k}(\underline{x}_k, \underline{p}_{k+1}, \underline{u}_k^*)\right) = \mathbf{0}$$

holds, that is

$$\frac{\partial^2 H_k}{\partial \underline{u}_k \partial \underline{x}_k} + \frac{\partial^2 H_k}{\partial \underline{u}_k \partial \underline{p}_{k+1}} \boldsymbol{H}_{k+1}\left(\frac{\partial \underline{f}}{\partial \underline{x}_k} + \frac{\partial \underline{f}}{\partial \underline{u}_k}\frac{\partial \underline{\mu}_k^*}{\partial \underline{x}_k}\right) + \frac{\partial^2 H_k}{\partial \underline{u}_k^2}\frac{\partial \underline{\mu}_k^*}{\partial \underline{x}_k} = \mathbf{0} \ ,$$

which leads to

$$\frac{\partial \underline{\mu}_k^*}{\partial \underline{x}_k} = -\left(\frac{\partial^2 H_k}{\partial \underline{u}_k \partial \underline{p}_{k+1}} \boldsymbol{H}_{k+1}\frac{\partial \underline{f}}{\underline{u}_k} + \frac{\partial^2 H_k}{\partial \underline{u}_k^2}\right)^{-1} \cdot \left(\frac{\partial^2 H_k}{\partial \underline{u}_k \partial \underline{x}_k} + \frac{\partial^2 H_k}{\partial \underline{u}_k \partial \underline{p}_{k+1}} \boldsymbol{H}_{k+1}\frac{\partial \underline{f}}{\partial \underline{x}_k}\right) \ , \quad (21)$$

where $\frac{\partial^2 H_k}{\partial \underline{u}_k \partial \underline{p}_{k+1}}$ is unknown. With

$$\frac{\partial H_k}{\partial \underline{u}_k} = \frac{\partial g_k}{\partial \underline{u}_k} + \underline{p}_{k+1}^T \frac{\partial \underline{f}}{\partial \underline{u}_k} \ ,$$

it follows that

$$\frac{\partial^2 H_k}{\partial \underline{u}_k \partial \underline{p}_{k+1}} = \left(\frac{\partial \underline{f}}{\partial \underline{u}_k}\right)^T \ . \quad (22)$$

Substitution of (22) in (21) and subsequent substitution of (20) and (21) in (19) yields proposition (18) and concludes the proof of Theorem 3. ■

IV. Two-Point Boundary-Value Problem

The boundary conditions (9) and the known state $\hat{\underline{x}}_0$, the state iteration (5), and the costate recursion (10) define a TPBVP for the considered system. For a given sequence of controls $(\underline{u}_0, \ldots, \underline{u}_{N-1})$, the corresponding states can be calculated by means of (5). After that, the corresponding costate sequence $(\underline{p}_N, \ldots, \underline{p}_0)$ is obtained by means of (9) and (10), starting from the final state $\hat{\underline{x}}_N$ of the system iteration. Thus, the knowledge of the $\underline{u}_k$-sequence is sufficient to obtain the remaining information. Introducing an augmented vector of the unknown optimal controls $\underline{u}_k^*$, $k = 0, \ldots, N-1$, as

$$\underline{U}^* := \left(\left(\underline{u}_0^*\right)^T \quad \cdots \quad \left(\underline{u}_{N-1}^*\right)^T\right)^T \ , \quad (23)$$

the optimal state-feedback control for the current state $\underline{x}_0^* := \hat{\underline{x}}_0$ is given by $\underline{u}_0^*$. Moreover, the necessary minimum condition (16) is rewritten by means of the nonlinear equation system

$$\underline{F}(\underline{U}^*) := \begin{pmatrix} \left(\frac{\partial H_0(\underline{x}_0^*, \underline{p}_1, \underline{u}_0^*, \underline{\boldsymbol{w}}_0)}{\partial \underline{u}_0}\right)^T \\ \left(\frac{\partial H_1(\underline{x}_1^*, \underline{p}_2, \underline{u}_1^*, \underline{\boldsymbol{w}}_1)}{\partial \underline{u}_1}\right)^T \\ \vdots \\ \left(\frac{\partial H_{N-1}(\underline{x}_{N-1}^*, \underline{p}_N, \underline{u}_{N-1}^*, \underline{\boldsymbol{w}}_{N-1})}{\partial \underline{u}_{N-1}}\right)^T \end{pmatrix} = \underline{0} \quad (24)$$

with N nonlinear equations for the N unknown optimal controls $\underline{u}_0^*, \ldots, \underline{u}_{N-1}^*$.

A. Solution with a Continuation Method

A continuation method provides an approach to solve the nonlinear equation system (24), which is a difficult task in general. The main idea of the continuation method is to embed (24) into a parameterized family of problems

$$\underline{F}(\underline{U}^*(\gamma)) := \begin{pmatrix} \left(\frac{\partial H_0(\underline{x}_0^*, \underline{p}_1(\gamma), \underline{u}_0^*(\gamma), \underline{\boldsymbol{w}}_0)}{\partial \underline{u}_0}\right)^T \\ \left(\frac{\partial H_1(\underline{x}_1^*(\gamma), \underline{p}_2(\gamma), \underline{u}_1^*(\gamma), \underline{\boldsymbol{w}}_1)}{\partial \underline{u}_1}\right)^T \\ \vdots \\ \left(\frac{\partial H_{N-1}(\underline{x}_{N-1}^*(\gamma), \underline{p}_N(\gamma), \underline{u}_{N-1}^*(\gamma), \underline{\boldsymbol{w}}_{N-1})}{\partial \underline{u}_{N-1}}\right)^T \end{pmatrix} = \underline{0} \ , \ \gamma \in [0,1] \ , \quad (25)$$

such that for the parameter $\gamma = 0$ the solution to an easy problem is obtained and for $\gamma = 1$ the original problem is described. With an increasing parameter $0 \leq \gamma \leq 1$, the easy problem is being continuously transformed into the original problem. During this process the solution to the problem is being traced. This means, that the solution for the previous value γ^- serves as an initial guess for the current continuation parameter. Then, the nonlinear equation system (25) can be solved, for example by means of a Newton method. The desired solution is obtained for $\gamma = 1$. Instead of applying a minimization method directly to (24), the continuation approach yields good initial guesses at each step, if the function $\underline{F}$ is sufficiently smooth.

In the considered case, the stochastic nonlinear system (1) can be parameterized, such that the easy problem is to find the optimal control for a linear system. For example, the system description (1) can be changed into

$$\underline{\boldsymbol{x}}_{k+1}(\gamma) = \gamma\, \underline{f}(\underline{\boldsymbol{x}}_k, \underline{u}_k) + (1-\gamma)\, \underline{l}(\underline{\boldsymbol{x}}_k, \underline{u}_k) + \underline{\boldsymbol{w}}_k \ , \quad (26)$$

such that the problem for $\gamma = 0$ consists in solving the LQ-problem. The original system (1) is obtained for $\gamma = 1$. In the linear case, the solution to the optimal control problem can be obtained by the discrete-time Riccati equation [2].

B. Implemented Algorithm

For a fixed terminal time, candidates for the desired optimal state-feedback controls $\underline{u}_k^*$, $k = 0, \ldots, N-1$, of the nonlinear system are determined as summarized in Algorithm 1. The known current state $\hat{\underline{x}}_k$ is accessible and is employed as a new initial value. The continuation method initially solves the LQ-problem and yields the solution $\underline{U}^*(0)$. This solution serves as an initial guess $\underline{U}_{\text{init}}(\gamma)$ for a Newton method that calculates $\underline{U}^*(\gamma)$ for increasing γ to satisfy condition (25). The desired state-feedback control $\underline{u}_k^*$ is given as the first entry of $\underline{U}^*(1)$.

Remark 4: The initial value to the numerical algorithm is a good choice, since the initial guess is the assumed correct solution of the previous step γ^- of the continuation. In case of sufficiently small continuation steps and a sufficiently smooth value function, the Newton iteration yields the correct solution, if the initial guess is close to the solution.

The extension of Algorithm 1 to the technically important model predictive control is straightforward.

Algorithm 1

init: $\delta > 0$; $T :=$ terminal time
for $k = 0$ to T **do**
  $\hat{\underline{x}}_0 := \hat{\underline{x}}_k$
  $N := T - k$
  $\underline{U}^*(0) = \mathtt{LQC}\ (\hat{\underline{x}}_0,\ N)$
  for $\gamma = \delta;\ \gamma \leq 1;\ \gamma = \gamma + \delta$ **do**
    $\gamma^- := \gamma - \delta$
    $\underline{U}_{\text{init}}(\gamma) = \underline{U}^*(\gamma^-)$
    $\underline{U}^*(\gamma) = \mathtt{newton}\ (\underline{U}_{\text{init}}(\gamma))$
  end for
  $\underline{u}_k^* := \underline{u}_0^*(1)$
  $\boldsymbol{\underline{x}}_{k+1} = \underline{f}(\hat{\underline{x}}_k, \underline{u}_k^*) + \underline{\boldsymbol{w}}_k$
end for

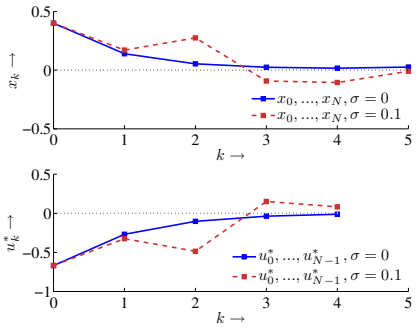


Fig. 1. Example state- and control sequences for $\sigma = 0$ and $\sigma = 0.1$. The state is shifted significantly due to the relatively strong noise influence on the system. The control has to be adapted due to this deviation.

V. Example

Let a scalar example system be given by

$$\boldsymbol{x}_{k+1} = \sin(q\,\boldsymbol{x}_k) + u_k + \boldsymbol{w}_k \ , \tag{27}$$

where $\boldsymbol{x}_k \in \mathbb{R}$, $q = \frac{3\pi}{4}$. The simulations are performed for $\hat{x}_0 \in \mathcal{X} := \{-1, -0.9, \ldots, 1\}$. $\boldsymbol{w}_k$ is a zero-mean independent Gaussian noise term with standard deviation $\sigma \in \mathcal{S} := \{0.05, 0.1, 0.2\}$. The initial horizon is set to five steps. After each time step the current system state is accessible and a candidate for the optimal state-feedback control of the shrunk horizon is determined. The parameterized system for the continuation is given by

$$\boldsymbol{x}_{k+1}(\gamma) = \gamma\,\sin\left(q\,\boldsymbol{x}_k\right) + (1-\gamma)\boldsymbol{x}_k + u_k + \boldsymbol{w}_k \ , \tag{28}$$

where $0 \leq \gamma \leq 1$. The solution of each continuation step is employed as the initial guess of the solution of the next continuation step. The system state is propagated by means of (5). Denoting the second derivative of J_{k+1} by h_{k+1}, the approximated value function according to (4) is given by

$$\begin{aligned} J_N(x_N) \ &= \frac{1}{2}(x_N - c)^2 \\ J_k(x_k, \gamma) &= \frac{1}{2}\left((x_k - c)^2 + a\,(u_k^*(\gamma))^2\right) + J_{k+1}\left(x_{k+1}(\gamma)\right) \\ &\quad + \frac{1}{2}\sigma^2 h_{k+1}(x_{k+1}(\gamma)) \end{aligned} \tag{29}$$

with a weighting factor $a = 2$ and the desired terminal state $c = 0$. The costate recursion and the Hamilton function yield the necessary minimum condition

$$\frac{\partial H_k(x_k^*(\gamma), p_{k+1}(\gamma), u_k^*(\gamma), \underline{\boldsymbol{w}}_k)}{\partial u_k} = a\,u_k^*(\gamma) + p_{k+1}(\gamma) = 0$$

and therefore an analytical solution

$$u_k^*(\gamma) = -a^{-1}\,p_{k+1}(\gamma) \tag{30}$$

to a candidate of the optimal state-feedback control $u_k^*(\gamma)$. Thus, (30) can be employed to verify the numerical solution of the algorithm.

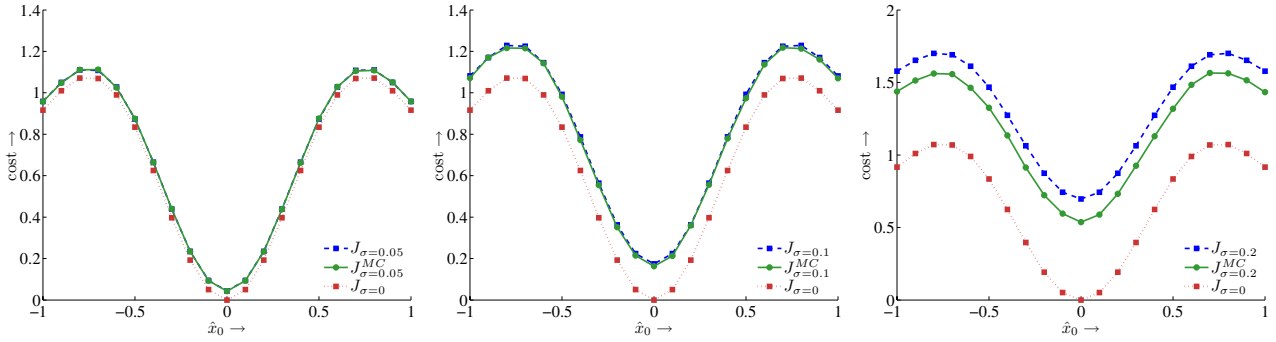
Remark 5: In contrast to an algorithm, which does not employ the continuation method, Algorithm 1 always converged and provided correct results. Therefore, the additional expenses arising from the continuation are justified.

Equation (27) reveals that the influence of noise is as strong as the influence of the control variable u_k. Moreover, the sine as the nonlinear part of the system function is bounded and attains values within the interval $[-1, 1]$. Thus, even the influence of noise with standard deviation $\sigma = 0.1$ can be regarded as a relatively strong influence on the considered system. This fact is stressed by Fig. 1 for one example simulation with $\hat{x}_0 = 0.4$, where the deviations of the state- and control trajectories can be seen easily.

In the following, $J_{\sigma=i}$ denotes the approximated value function for the initial horizon defined by (29) for $\gamma = 1$. The system is affected by noise with standard deviation σ. Therefore, $J_{\sigma=0}$ denotes the value function for the deterministic system, that is (29) without the additional term depending on the noise influence. For $\sigma > 0$ the arising cost $J_0(\hat{x}_0)$ of the simulation changes with each run. A Monte-Carlo simulation provides an approximate upper bound $J_{\sigma=i}^{MC}$ of the true value function depending on $\sigma = i$ by calculating the arithmetic mean of all arisen costs starting from $\hat{x}_0 \in \mathcal{X}$.

Remark 6: After 3000 runs, the result of the Monte-Carlo simulation is assumed to provide a sufficiently good estimate of the true value function (under deterministic control). This assumption is based on the uniqueness of the solution to the LQ-problem and the employment of the continuation method, which keeps the solution in the correct minimum.

In some practical applications, only the knowledge of the true value function is desired, instead of the optimal control leading to the value function. Thus, in Fig. 2 the Monte-Carlo estimate $J_{\sigma=i}^{MC}$, the approximated value function $J_{\sigma=i}$ given by (29), and the deterministic value function $J_{\sigma=0}$, which would result from the negligence of the noise in the system function (27), are compared. To calculate the Monte-Carlo estimate, a multitude of simulations is necessary in contrast to the value function approximations in Fig. 2, which can be calculated directly. Fig. 2(a) shows that $J_{\sigma=0.05}$ is very close to $J_{\sigma=0.05}^{MC}$. On the other hand, $J_{\sigma=0}$ would also be an acceptable approximation of the value function. For $\sigma = 0.1$, the approximation and the Monte-Carlo estimate almost coincide yet, in contrast to $J_{\sigma=0}$ as depicted in Fig. 2(b). Therefore, the proposed algorithm yields significantly better approximations of $J_{\sigma=0.1}^{MC}$. Since the influence of higher-



(a) $J^{MC}_{\sigma=0.05}$ and $J_{\sigma=0.05}$ almost coincide. A slight approximation improvement to $J_{\sigma=0}$ can be seen.

(b) $J^{MC}_{\sigma=0.1}$ and $J_{\sigma=0.1}$ almost coincide yet. The approximation by means of $J_{\sigma=0}$ is significantly worse.

(c) The structural approximation error becomes more significant for $\sigma = 0.2$.

Fig. 2. Estimated true value function and its approximations for a five-step horizon and different noise influences.

TABLE I

QUALITY OF THE APPROXIMATIONS FOR DIFFERENT NOISE INFLUENCES

	$d(J_{\sigma=0.05}, J^{MC}_{\sigma=0.05})$	$d(J_{\sigma=0}, J^{MC}_{\sigma=0.05})$	$d(J_{\sigma=0.1}, J^{MC}_{\sigma=0.1})$	$d(J_{\sigma=0}, J^{MC}_{\sigma=0.1})$	$d(J_{\sigma=0.2}, J^{MC}_{\sigma=0.2})$	$d(J_{\sigma=0}, J^{MC}_{\sigma=0.2})$
mean	0.0015	0.0404	0.0098	0.1533	0.1426	0.5096
max	0.0041	0.0430	0.0190	0.1671	0.1584	0.5454

order derivatives in the Taylor series expansion of the value function increases with increasing standard deviation, the error of the proposed approximation also increases in cases, where these derivatives do not vanish, which is stressed by Fig. 2(c). But even in this case, the employment of $J_{\sigma=0.2}$ is in fact preferable to the employment of $J_{\sigma=0}$ to approximate $J^{MC}_{\sigma=0.2}$. Table I summarizes the quality of the approximations $J_{\sigma=0}$ and $J_{\sigma=i}$ of $J^{MC}_{\sigma=i}$, $i \in \mathcal{S}$, where the distance measure d is defined pointwise, that is for all $\hat{x}_0 \in \mathcal{X}$

$$d(f_1, f_2) := \|f_1(\hat{x}_0) - f_2(\hat{x}_0)\|_2$$

for two functions f_1, f_2. Furthermore, the structural error for increasing σ due to (4) is revealed. Taking everything into account, the proposed approximation is preferable to the full disregard of the noise influence in the considered example.

VI. CONCLUSIONS AND FUTURE WORK

In this work, an approach to finite-horizon state-feedback control of nonlinear, stochastic, discrete-time systems has been proposed. Employing the idea of dynamic programming, the value function has been approximated by means of Taylor series expansion up to second-order derivatives. This approximation contains additional terms, which are a contribution to the influence of noise. Moreover, a minimum principle has been applied to the stochastic system. Similar to the deterministic case, a necessary condition has been derived to obtain the desired optimal state-feedback control. Employing these results, a two-point boundary-value problem has been formulated that was solved by means of a continuation method. This continuation method initially solves the LQ-problem and traces the solution while the linear system is being transformed into the desired nonlinear system. A nonlinear system has been simulated, which employs the proposed approach. The true value function has been estimated by means of a Monte-Carlo algorithm. In case of the considered example, the estimated value function and the approximated value function almost coincide, even in case of relatively strong noise. Moreover, the results reveal that the proposed approximation is superior to an approximation, which does not consider any influence of noise.

Future work will be aimed at extended incorporation of the stochastic behavior of the system into determination of the optimal control sequence.

REFERENCES

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

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

[3] 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, 2003, pp. 91–95. [Online]. Available: http://dit.unitn.it/~pistore/conferences/ICAPS03-wshop/Papers/Nikovski.pdf

[4] 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. [Online]. Available: http://www.sciencedirect.com/science/

[5] 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.

[6] J. D. Turner and J. L. Junkins, "Optimal Large-Angle Single-Axis Rotational Maneuvers of Flexible Spacecraft," in *2nd AIAA/VPI&SU Symposium on Dynamics and Control of Large Flexible Spacecraft*, Blacksburg, VA, June 21–23 1979, pp. 91–110.