Approximate Dual Control Maintaining the Value of Information with an Application to Building Control

Edgar D. Klenske[1], Philipp Hennig[1], Bernhard Schölkopf[1] and Melanie N. Zeilinger[2]

Abstract— **Dual control, the simultaneous identification and control of dynamic systems, is an idea that has been around for several decades without being widely used in applications, due to the fundamental intractability of the optimal solution. Available algorithms are either complex and computationally demanding, or reduce to a simple change of the cost function, which can lead to poor average performance. In addition, classic dual control schemes do not deal with constraints and economic cost structures present in many applications. In this paper, we aim at facilitating the use of dual control algorithms based on a series expansion of the cost function for such systems. In particular, this is realized by employing reference tracking of the optimal mean trajectory together with soft constraints. A key feature of the proposed formulation is that it maintains the value of information in the cost, making dual control tractable with all dual features. Evaluation on a simulated building control problem exhibits an advantage of using the proposed dual controller over simplified solutions.**

I. Introduction

DUAL CONTROL addresses the goal of solving the well-known exploration-exploitation trade-off optimally. The term was coined by [1], who noted that dual control can be seen as optimal control of not only the physical states of the system, but also of the parameters of the underlying dynamics model. See e.g., [2], [3] for an overview. It was realized early that this kind of problem is intractable [4], except for a few comparably simple systems, e.g., [5]. Approximate dual control methods have therefore been developed, as well as simpler alternatives based on re-formulations of state and control costs.

Many approximate methods are, however, computationally expensive or too simple to retain all features of dual control: *caution*, the downscaling of control signals when facing high uncertainty; *exploration*, the excitation of the system when cautious control does not learn fast enough; and the *value of information*, the selective exploration of system parameters that are important for future performance of the system [6].

An approximation derived by [7], [8] is conceptually close to optimal dual control and retains all three of the aforementioned features of dual control. In its classic form, however, this framework is not able to address aspects central to many modern control problems: nonlinear dynamics, constraints, and non-quadratic cost functions. The framework was recently extended to nonlinear systems [9]; the present work additionally addresses systems with constraints and non-quadratic cost functions. Dual control is particularly beneficial for systems with economic (linear) cost, since it can exploit time-varying cost structures to optimally identify the latent parameters of dynamical systems.

A problem of this type that has raised significant attention in recent years is building climate control, due to its potential impact on the world-wide energy consumption: A large amount of the globally consumed energy is used for buildings [10]. This has sparked recent interest in model predictive control (MPC) for buildings [11], [12], [13], [14], leveraging predictions of the model as well as external error sources, such as weather conditions and occupancy. These techniques require a sufficiently accurate system model. As the parameters of the model generally vary with the building and potentially with time, parameter identification has to be performed individually for each building during operation, which can be expensive.

Adaptive controllers offer the potential to obtain accurate models for a low energy footprint over the whole lifetime of the building. While passively learning adaptive control systems [15] can only learn by evaluating past measurements, dual controllers [16] can enhance the learning procedure by also reasoning about the effect of current actions on the control performance of the future. This way, dual control can make use of certain parts of the problem structure to identify the model more efficiently than purely passive control systems. For example, a dual controller can learn at times where the energy cost or demand is low (making use of real-time-pricing or day/night tariffs) to obtain more precise control at times of high control cost.

Dual control has regained attention over the past few years, not only in combination with MPC [17], but also in the use for building control [18]. In many dual control approaches exploration bonuses are used and added to the control cost. Other methods include constraining the minimal information gain [19], [20] and persistent excitation [21], [22]. In this work, we focus on maintaining the value of information explicitly, which cannot be expressed through exploration bonuses or excitation signals, but which is one of the key benefits of a dual controller.

After introducing the problem setting and approximate dual control based on series-expansion of the cost (Sec. II), we provide a procedure for using the method for economic cost and constrained systems using hierarchical tracking MPC and soft constraints (Sec. III). We apply the proposed technique to a simple building control problem and analyze the performance with respect to purely passively learning methods as well as simplistic dual control (Sec. IV).

[1] Edgar D. Klenske, Philipp Hennig and Bernhard Schölkopf are with the Max-Planck-Institute for Intelligent Systems, Spemannstraße 38, 72076 Tübingen, Germany {eklenske, bs, phennig} @tue.mpg.de

[2] Melanie N. Zeilinger is with the ETH Zürich, Sonneggstrasse 3, 8092 Zürich, Switzerland mzeilinger@ethz.ch

II. Preliminaries

A. Problem setting

We consider the continuous-time system

$$\dot{x}(t) = f(x(t), u(t), w(t)) + v(t), \tag{1}$$

with state $x \in \mathbb{R}^{n_x}$, input $u \in \mathbb{R}^{n_u}$, disturbance $w \in \mathbb{R}^{n_w}$ and white noise $v \in \mathbb{R}^{n_x}$. We assume that the dynamics f are not known, but can be described up to Gaussian uncertainty by a general linear model with linear and nonlinear features ϕ, and an unknown matrix M of appropriate size:

$$\dot{x}(t) = M\phi(x(t), u(t), w(t)) + v(t). \tag{2}$$

The linear part of the system can be discretized in numerous ways, but for maximal accuracy while retaining the possibility to directly calculate the Jacobian w.r.t. the parameters, we use element-wise zero-order-hold linearization: Each state dynamics is discretized as scalar differential equation, considering the other states as inputs. Using this method, we arrive at the discretized system

$$x_{k+1} = A_k x_k + B_k u_k + D_k w_k + M_k \phi^n(x_k, u_k, w_k) + \xi_k, \tag{3}$$

with time index k, matrices A_k, B_k, D_k of appropriate sizes and Gaussian disturbance $\xi_k \sim \mathcal{N}(0, \Xi)$. The matrix M_k is the result of a first-order Euler forward exponential integrator [23] of the nonlinear features ϕ^n. For simplicity of notation, we subsume the non-zero elements of matrices A_k, B_k, D_k and M_k into a parameter vector θ_k. The system is subject to possibly time-varying state and input constraints $x_k \in \mathbb{X}_k$ and $u_k \in \mathbb{U}_k$, where $\mathbb{X}_k \subset \mathbb{R}^{n_x}$ and $\mathbb{U}_k \subset \mathbb{R}^{n_u}$ are polytopes.

B. Approximate Dual Control

Dual control can be seen as a variant of adaptive control, where uncertain dynamics are identified and the belief about the dynamics is updated during runtime. For parametric settings, this means that the parameters of the dynamics are defined by a probability distribution.

In practice, a common procedure is to apply certainty equivalence (CE) control, using the current mean estimate of the parameters to compute the controller [24]. This approach can, however, lead to failure in cases of high uncertainty, which often occur in the beginning or after parameter changes. One approach to incorporate the uncertainty is stochastic optimal control [25], where the uncertain parameters are marginalized out while calculating the optimal controller. This leads to smaller control signals under higher uncertainty ("caution"), but it can also result in the so-called "turn-off phenomenon" [26] in the face of large uncertainties: The control is scaled down to zero, and, as a result, the system never acts or learns.

The—theoretically ideal, but intractable—way to deal with simultaneous identification and control is dual control: When the learning as response to current actions is taken into account, the turn-off characteristic vanishes in favor of explorative behavior[1] [1]. Because this is fundamentally intractable for general problems [4], tractable approximations have been developed. While many approximations only mimic some features of dual control (caution and exploration, e.g., [17], [18]), some methods build approximations of the cost structure of dual control (e.g., [8]), in order to take into account future belief updates.

The key observation in dual control is that both the states x and the parameters θ are subject to uncertainty and can therefore be subsumed in an *augmented state* $z_k^\top = \begin{pmatrix} x_k^\top & \theta_k^\top \end{pmatrix} \in \mathbb{R}^{n_x+n_p}$ [1], [27], [28]. The uncertainty of states and parameters can then be dealt with in the form of a joint probability density $p(z) = \mathcal{N}(z, \mu, \Sigma)$. In this framework, the dual control problem reduces to optimal stochastic control of the augmented system. It is important to note that we assume the parameters to be deterministic, but unknown; this is modeled by defining a prior distribution over the parameters $p(\theta_0) \sim \mathcal{N}(\mu_0^\theta, \Sigma_0^{\theta\theta})$ and deterministic parameter dynamics $p(\theta_{k+1}|\theta_k) = \delta(\theta_{k+1} - \theta_k)$, where δ is the Dirac delta distribution.

The optimal controller minimizing the *expected cost* over a finite horizon is defined by Bellman's equation

$$J_k(x_k) = \mathbb{E}_{z_k}\left[\ell_k(x_k, u_k) + \mathbb{E}_{x_{k+1}}\left[J_{k+1}(x_{k+1})\right]\right], \tag{4}$$

which can be solved using dynamic programming [29], [30]. In this formulation, all past knowledge is incorporated into the belief over z_k. The cost at each stage ℓ_k depends on both states x_k and inputs u_k except for the last stage of the horizon, N, where it only depends on the state x_N. The final element of the cost is defined as

$$J_N(x_N) = \mathbb{E}_{x_N}\left[\ell_N(x_N)\right]. \tag{5}$$

The optimal controller minimizing this cost will be denoted u_k^*, with associated cost

$$J_k^*(x_k) = \min_{u_k} \mathbb{E}_{z_k}\left[\ell_k(x_k, u_k) + \mathbb{E}_{x_{k+1}}\left[J_{k+1}^*(x_{k+1})\right]\right]. \tag{6}$$

This recursive formulation amounts to alternating minimization and expectation steps. As u_k influences both x_{k+1} and the belief θ_{k+1}, it enters the latter expectation nonlinearly, resulting in the loss of the closed-form solution. The alternative, i.e. optimizing the control inputs $u_k \cdots u_{N-1}$ jointly, is usually impractical due to the curse of dimensionality.

The basic idea of one class of approximate dual control [7], [8] is to use a second-order approximation of the nonlinear expected cost-to-go in combination with a nonlinear optimization scheme. The algorithm is outlined in the flowchart in Fig. 1. The optimization at each time step is initialized with the CE solution as starting point for the nonlinear optimization. In an inner loop, the gradient-free optimization algorithm evaluates the approximated cost function at different locations u_k. The optimization loop runs until convergence or until a number of function evaluations is reached.

The evaluation of the cost approximation is divided in three conceptual steps:

① With the given u_k, a one-step prediction is performed.
② From the predicted next state x_{k+1}, a CE trajectory is computed, including state and parameter covariances.

[1] This feature is sometimes also called "investigation" or "probing" in the dual control literature.

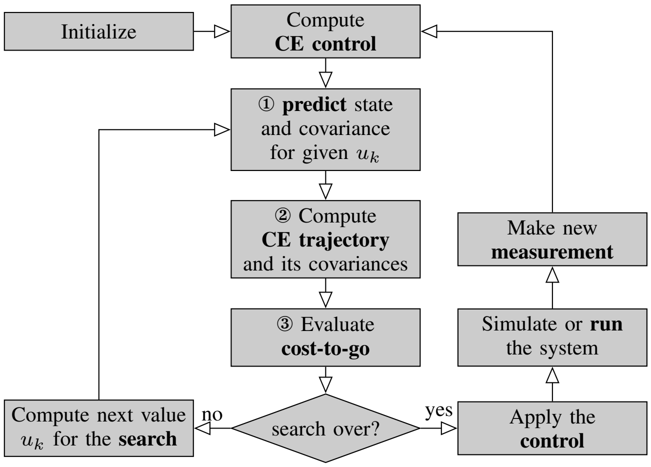

Fig. 1. Flow-chart of the approximate dual control algorithm to show the overall structure. Adapted from [8]. The left cycle is the inner loop, performing the nonlinear optimization.

③ The second-order approximation of the cost-to-go is evaluated, using the covariances along the trajectory.

For each evaluation of the cost function, ①-③ are computed to obtain the cost-to-go. More details on the cost approximation will be provided in the next section. Details of the overall method can be found in the original papers [7], [8] or in the recent nonlinear adaptation [9].

The main purpose of this algorithm is to reduce the highly nonlinear optimization algorithm in multiple dimensions (control inputs over the horizon) to nonlinear optimization of only the first control input, by approximating the cost-to go, in an iterative fashion. While retaining the possibility to explore, this approach alleviates the curse of dimensionality. This procedure also circumvents the difficulty of a receding horizon approach in the dual control setting as noted in [22]: If the excitation is planned for future time steps, in closed-loop the excitation may be delayed at every time instance, leading to non-explorative behavior. Forcing the excitation to occur in the first step only, this problem cannot arise.

C. The Value of Information

The value of information refers to the fact that not all parameters of an uncertain model are equally important. If a certain parameter is important for the future control performance, it can be beneficial to identify this parameter and it might pay off to invest some energy in its identification. If a parameter does not have an important impact on the future cost, its identification can be neglected.

The second-order approximation ③ defines a quadratic reference tracking problem based on the CE trajectory. The solution can be obtained with dynamic programming, projecting onto a quadratic form at every time step. This results in an approximation of the cost-to-go of the form [8]

$$J_k^*(x_{k+1}) = \frac{1}{2}[(x_{k+1} - x_{k+1}^{\text{ref}})^\top K_{k+1}^{xx}(x_{k+1} - x_{k+1}^{\text{ref}})] + \frac{1}{2}\,\text{tr}\sum_{j=k}^{N-1}\left\{Q_{j+1}\Sigma_{j+1}^{xx} + \left[\Sigma_{j+1|j} - \Sigma_{j+1}\right]K_{j+1}\right\}, \quad (7)$$

where $\cdot^{xx}$ defines the submatrix belonging to the state x from a matrix that is defined for the augmented state z. K_j is defined by the recursive Riccati equation

$$K_j = \bar{A}_j^\top K_{j+1}\bar{A}_j - K_{j+1}\bar{B}_j\left(\bar{B}_j^\top K_{j+1}\bar{B}_j + R_j\right)^{-1}\bar{B}_j^\top K_{j+1}\bar{A}_j + Q_j \quad (8a)$$

$$K_N = Q_N, \quad (8b)$$

where the Jacobian $\bar{A}_j$ and the input response $\bar{B}_j$ are calculated for the augmented system. The future beliefs Σ_j along the trajectory are generated with an extended Kalman filter (EKF) [31], where $\Sigma_{j+1|j}$ denotes the EKF prediction before the subsequent update.

While the first component of Eq. (7) is the usual deterministic cost, the trace term adds a cost that results from the uncertainty. The term $Q_{j+1}\Sigma_{j+1}^{xx}$ represents the cost of the state uncertainty in future time steps. Since the source of this uncertainty is mostly control actions with uncertain outcome (such as an unknown gain), this term results in cautious behavior of the control system. The final term $\left[\Sigma_{j+1|j} - \Sigma_{j+1}\right]K_{j+1}$ is the most interesting part, as it represents an approximate measure for the value of information: It introduces a cost that weighs the covariance update $\left[\Sigma_{j+1|j} - \Sigma_{j+1}\right]$ by the projected cost matrix K_{j+1}. This results in high cost for important parameters (indicated by large values in K_{j+1}) that are learned during the process (indicated by large values in the covariance update). If the parameters are either unimportant, precisely known or cannot be learned, this additional cost term vanishes.

III. Approximate Dual Control with Non-Quadratic Cost and Constraints

Classic approximate dual control algorithms were posed in the LQG setting, assuming linear dynamics, quadratic cost and Gaussian noise. In this setting, the optimal CE trajectory and the subsequent perturbation control can be obtained with dynamic programming, because there is a recursive solution for the optimal controller at each time step.

Many control problems where dual control may have an important impact involve economic costs. An example is the considered application to building control, where the cost is linear (energy prices) and the inputs and states are constrained (bounds on the temperature, heating/cooling limits). In this setting, dynamic programming is computationally expensive, since there is no simple recursive solution to obtain a second-order approximation to the cost.

In order to deal with more general cost structures and constraints, we therefore propose to use a common hierarchical tracking scheme: 1) An economic reference satisfying the constraints is computed using the CE system and standard MPC techniques; 2) The reference is tracked using a dual controller, where state constraints are considered in the form of soft constraints. The details of this scheme are outlined in the following sections.

A. Economic Reference

The economic reference for the controller is generated by solving a discrete-time MPC problem for the CE system.

$$(\mathbf{x}^{\text{ref}},\mathbf{u}^{\text{ref}}) \coloneqq \operatorname*{arg\,min}_{\mathbf{x},\mathbf{u}} \quad \sum_{n=0}^{N-1} \ell_n(x_n, u_n) \tag{9a}$$

$$\text{s.t.} \quad x_0 = x_k \tag{9b}$$

$$x_{n+1} = A_n x_n + B_n u_n + D_n w_n + M_d \phi(x_n, u_n, w_n) \tag{9c}$$

$$x_n \in \mathbb{X}_n \tag{9d}$$

$$u_n \in \mathbb{U}_n \tag{9e}$$

This nonlinear MPC problem is solved with standard algorithms, depending on the problem structure, e.g., sequential linear programming [32].

B. Soft Constraints and Uncertainty

Assuming Gaussian uncertainty, hard constraint satisfaction cannot be guaranteed when using the approximate dual control scheme in Sec. II-B. In order to capture the state constraints when tracking the reference, we introduce soft constraints. For constraints of the form $\mathbb{X}_k \coloneqq \{x_k \mid P_k x_k \leq p_k\}$, these take the form

$$\varepsilon_k(x_k) = \max(P_k x_k - p_k, \mathbf{0}), \tag{10a}$$

$$\ell_k^{\text{c}}(x_k) = \varepsilon_k(x_k)^\top W_k \varepsilon_k(x_k). \tag{10b}$$

With the $\max$ defined element-wise, ε_k captures the amount of constraint violation, whereas W_k penalizes the constraint violation in an extra cost term that is added to the stage cost considered by the dual controller.

In order to apply step ③ of the dual control scheme, one would now marginalize the Gaussian distributed state against the soft constraint penalty function. This calculation is of the form

$$\int_{-\infty}^{\infty} \varepsilon_k(x_k)^\top W_k \varepsilon_k(x_k) \cdot \mathcal{N}(x_k, \mu_k, \Sigma_k)\, \mathrm{d}x_k, \tag{11}$$

which has generally no closed-form solution because of the $\max$ operator in the definition of ε_k. Only when the mean μ_k coincides with the constraint boundary, there is a closed-form solution, amounting to

$$\int_{\mu_k}^{\infty} (Px_k - p)^\top W_k (Px_k - p) \cdot \mathcal{N}(x_k, \mu_k, \Sigma_k)\, \mathrm{d}x_k$$
$$= \frac{1}{2} \left((P\mu_k - p)^\top W_k (P\mu_k - p) + \operatorname{tr}\{W_k \Sigma_k\} \right). \tag{12}$$

For all states on the constraint boundary, this means that Gaussian marginalization of the soft constraints is equal to an additional quadratic tracking cost

$$\tilde{\ell}_k^{\text{c}}(x_k) = (x_k - x_k^{\text{ref}})^\top \tilde{W}_k (x_k - x_k^{\text{ref}}) \tag{13}$$

with $\tilde{W}_k = \frac{1}{2} W_k$. This can now be used to modify the second order approximation of the cost-to-go in (7), which is based on the CE reference trajectory: For states lying on the constraint boundary, the cost term in (13) is added. For states inside of the constraint boundaries, the additional cost can be reduced to zero, or to a small fraction of W_k to keep the cost positive definite if no other state cost is applied ($Q_k = 0$).

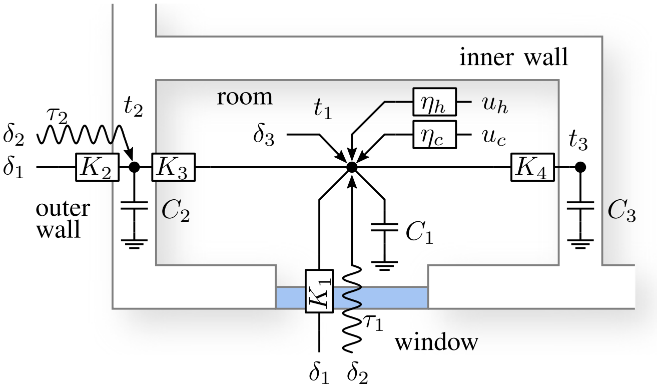

Fig. 2. Schematic overview of the building model. The room of which the temperature is to be controlled exchanges heat with the outside air through the window and the outer wall, and with the rest of the building through the inner wall. All symbols are explained in Table I.

With this approximation, we can capture some of the nonlinear effects of the state constraints and add them to the stage cost for the dual tracking controller

$$\ell_k^{\text{DC}} = (x_k - x_k^{\text{ref}})^\top Q_k (x_k - x_k^{\text{ref}})$$
$$+ (u_k - u_k^{\text{ref}})^\top R_k (u_k - u_k^{\text{ref}}) + \tilde{\ell}_k^{\text{c}}(x_k), \tag{14}$$

where Q_k and R_k are the cost matrices for states and inputs.

C. Maintaining the Value of Information

With the aforementioned modifications, the series-expansion based approximate dual control scheme presented in Sec. II-B can be applied to constrained linear programming problems. The basic idea is to obtain an approximation of the cost-to-go by tracking a CE trajectory satisfying constraints and minimizing an economic cost with a stochastic optimal controller. The rest of the procedure explained in Sec. II-B remains the same. The additional cost induced by parameter uncertainty and the inverse value of information can thereby be maintained also for more general cost functions and polytopic state constraints.

IV. Building Climate Control

A. The Building Model

In this section, we investigate the proposed dual controller for building climate control, with the goal of reducing overall energy consumption. We consider the simplified case of temperature control for a building equipped with a heat pump, a setup motivated by the increasing use of heat pumps in buildings. In this case, electrical energy can be assumed as the energy source for both heating and cooling. The simplified building model is shown in Fig. 2. The model is adapted from [11], [33], [34] and models the temperature in a single room inside a larger building.

Most parameters are relatively easy to identify and are therefore assumed to be known. The input efficiencies η_h and η_c in contrast are generally not known, but highly important

TABLE I

Overview of the model states, disturbances and parameters

symbol	meaning	unit
t_1	room air temperature	[°C]
t_2	exterior wall temperature	[°C]
t_3	interior wall temperature	[°C]
δ_1	outside air temperature	[°C]
δ_2	solar radiation	[kW]
δ_3	internal heat sources	[kW]
u	electrical input power	[kW]
u_h	electrical heating power ($u < 0$)	[kW]
u_c	electrical cooling power ($u > 0$)	[kW]
η_h	heating efficiency	-
η_c	cooling efficiency	-
τ_1	window radiation coefficient	-
τ_2	outer wall radiation coefficient	-
K_{1-4}	heat conductivities	[kW/°C]
C_{1-3}	heat capacities	[kJ]

and only identifiable while the respective inputs are active. For the simulation we consider 50 different buildings with parameters drawn from Gaussian distributions, $\eta_h \sim \mathcal{N}(4,2)$ and $\eta_c \sim \mathcal{N}(2,2)$, where we use rejection sampling to limit the range to $\eta_h \in [1,10]$ and $\eta_c \in [0.35,5]$.

The continuous-time dynamics of this building model are

$$\dot{t}_1 = \frac{1}{C_1}\left[K_3(t_2 - t_1) + K_1(\delta_1 - t_1) + K_4(t_3 - t_1) + \tau_1\delta_2 + \eta_h u_h + \eta_c u_c + \delta_3\right] \tag{15a}$$

$$\dot{t}_2 = \frac{1}{C_2}\left[K_2(\delta_1 - t_2) + K_3(t_1 - t_2) + \tau_2\delta_2\right] \tag{15b}$$

$$\dot{t}_3 = \frac{1}{C_3}\left[K_4(t_1 - t_3)\right] \tag{15c}$$

where all variables and parameters are defined in Table I. The model is simulated continuously, but state and control costs are defined for the discretized system. The model is simulated for a whole day with a discretization interval of $\Delta t = 600\,\mathrm{s}$.

The input constraints

$$-1000 \leq u_k \leq 1000 \tag{16}$$

are imposed at all times, representing the power limitations of the heat pump system. These constraints are chosen to retain feasibility also in the case of poor efficiency (low η_h and/or η_c). The input constraints are enforced through the CE MPC for generating the tracking trajectory. If the tracking dual controller violates the input constraints, the constraints are enforced by saturation. The state constraints are time-dependent to account for different temperature demands during and outside working hours:

$$\mathbb{X}_k : \begin{cases} 21 \leq t_1 \leq 26 & \text{from 08:00 to 18:00} \\ 19 \leq t_1 \leq 30 & \text{otherwise} \end{cases} \tag{17a}$$

The cost on constraint violation is also defined to be time-varying to account for the reduced importance of constraint satisfaction over night:

$$W_k = \begin{cases} 10^3 & \text{from 08:00 to 18:00} \\ 10^{-1} & \text{otherwise} \end{cases} \tag{18}$$

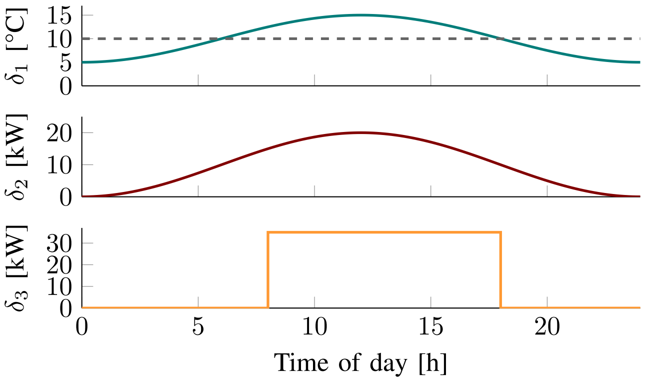

Fig. 3. The disturbances over 24 hours. **Top:** The outside air temperature (solid), around the mean temperature $\delta_0 = 10°\mathrm{C}$ (dashed). **Middle:** The solar radiation. **Bottom:** The internal heat gains.

The energy cost is linear

$$\ell_k^u(u_k) = d_k^\top u_k, \tag{19}$$

where the prices d_k are based on a day/night pricing, which is a common scheme for electricity used for heating:

$$d_k = \begin{cases} 0.025 & \text{from 06:00 to 22:00} \\ 0.010 & \text{otherwise} \end{cases} \tag{20}$$

The overall cost is the sum of energy and constraint cost

$$\ell_k(x_k, u_k) = \ell_k^u(u_k) + \ell_k^c(x_k). \tag{21}$$

For the purpose of controller comparison, we assume that an accurate prediction of outside air temperature and solar radiation is known. The disturbance trajectories are shown in Fig. 3. Nonetheless, not all simulated days are identical: The mean temperature is drawn from a Gaussian distribution $\delta_0 \sim \mathcal{N}(20,5)$ for each of the 50 building scenarios to provide a comparison of the controller types at days with different weather conditions.

B. Controller Types

In order to analyze the performance of the dual controller, we compare it to four other controllers. First of all, an optimal controller having access to the true parameter values is employed to serve as a lower bound (LB) to the cost for a specific instance of the problem.

The second approach is the CE controller, simply using the expectation of the uncertain parameters [24].

One of the more elaborate options when dealing with parameter uncertainties in MPC is the scenario approach (SA) [35]. Instead of relying on the mean value only, samples from the parameter distribution are used for marginalization. We use a simplified version of the scenario approach, where the MPC is solved for all sampled dynamics individually, averaging the optimal control afterwards. In order to obtain fast and reliable sampling, we use the latin hypercube sampling technique [36] for this process.

Since dual control is about the benefits of exploration, we also compare to a controller with modified cost function that favors exploration, also known as exploration bonus (EB) [37], [38]. This approach is often referred to as dual control,

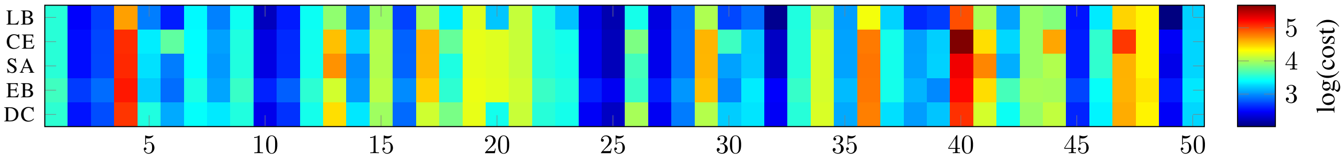

Fig. 4. Visual overview of the control performance for 50 different problem instances. The overall cost after one day is color-coded on a log scale. From top to bottom: Lower bound (LB), certainty equivalent (CE), scenario approach (SA), exploration bonus (EB), dual control (DC).

but it lacks the value of information. Exploration bonus based controllers cannot automatically decide which features of the dynamics are important. As a result, they aim at identifying as much as possible, defined by the trade-off between the superimposed (and purely virtual) uncertainty cost and the actual cost. We use an exploration bonus with an additional cost term of the form

$$\operatorname{diag}(\Sigma)^\top Q_{\text{EB}} \operatorname{diag}(\Sigma). \tag{22}$$

The matrix Q_{EB} has to be tuned for the exploration bonus to have some effect, but also not to dominate the certainty equivalent cost structure. In the experiments, it was chosen

$$Q_{\text{EB}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{23}$$

The last controller in the comparison is the dual controller (DC) as presented in this paper.

All controllers, except for the LB, use the element-wise zero-order hold discretization as described in Sec. II and all use a horizon length of one day ($N = 144$).

C. Experimental Results

In order to provide a fair comparison of the different controllers under uncertainty, we sampled 50 different buildings with 50 different weather conditions, as described in Sec. IV-A. For each of these setups, the performance of all five controllers was evaluated. Since already the optimal performance under full knowledge varies tremendously based on the temperature and the heat pump efficiencies, the performances of the tested controllers are also evaluated relative to the lower bound performance. The aggregated results are shown in Table II. Since the variability due to the different scenarios is high, it is difficult to draw strong general conclusions. Nonetheless it is noticeable that the dual controller shows the best average performance. Relative to the lower bound, the DC shows more than 50% improvement compared to the standard CE approach and about 28% compared to EB.

Fig. 4 shows the performance of the different controller types as color-coded entries of the result matrix, visualizing the performance differences. In most cases DC outperforms EB, but in some cases it is the other way round. This is due to the fact that, based on the weather, for certain days only the heating is necessary, for certain days only the cooling, and for some days both.

Note that for days where both cooling and heating are used, the EB and DC controllers perform almost equally well, since both input parameters have to be learned. Remaining differences are due to the used approximation and tuning. Fig. 5 illustrates such a case (problem instance 23), where the DC has not much benefit over the EB. Fig. 6 on the other hand shows a day (problem instance 20), where only heating, but no cooling is needed. This is an example of a situation where it is profitable to use DC instead of EB. Any controller with exploration bonus tries to identify *all* uncertain parameters, whereas the dual controller only identifies the parameters that are important, or *valuable* in this scenario.

TABLE II

SIMULATION RESULTS

	absolute			relative to LB		
	mean	std	SEM	mean	std	SEM
LB	34.45	27.24	3.85	0.00	0.00	0.00
CE	49.44	50.24	7.11	14.99	26.65	3.77
SA	45.15	40.76	5.76	10.70	18.09	2.56
EB	44.60	37.70	5.33	10.15	14.27	2.02
DC	41.75	33.45	4.73	7.30	10.61	1.50

Overall performance comparison. Aggregated costs over 50 different instances of the building control problem for the different controllers: Lower bound (LB), certainty equivalent (CE), scenario approach (SA), exploration bonus (EB), dual control (DC). Provided are the sample mean, sample standard deviation and the standard error of the mean (SEM).

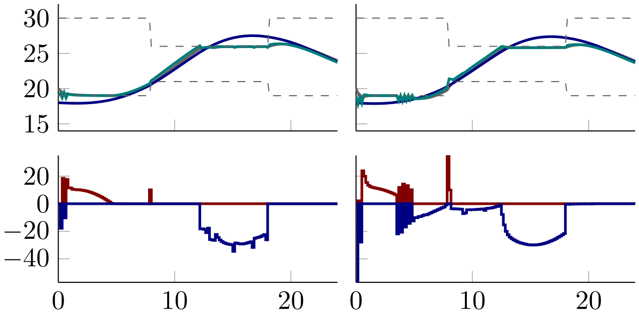

Fig. 5. Weather conditions where both heating and cooling are used (problem instance 23). **Top row:** Room temperature (green), outer wall temperature (blue) and inner wall temperature (grey). The constraints on the room temperature are dashed grey. **Bottom row:** Control inputs for heating (red) and cooling (blue). **Left:** Exploration bonus controller. **Right:** Dual controller.

V. Conclusions

The value of information is a feature of dual control often neglected. Using an approximation to the optimal dual control formulation in terms of a series expansion of the cost function, we constructed a controller that maintains an approximation of the value of information in systems with linear cost structure. This controller favors the identification of relevant features and ignores features that are not necessary for future control.

The method is based on the construction of a tracking reference found by solving the optimal control problem for the current mean estimate of the parameters. This reference is subsequently tracked by a quadratic low-level dual controller

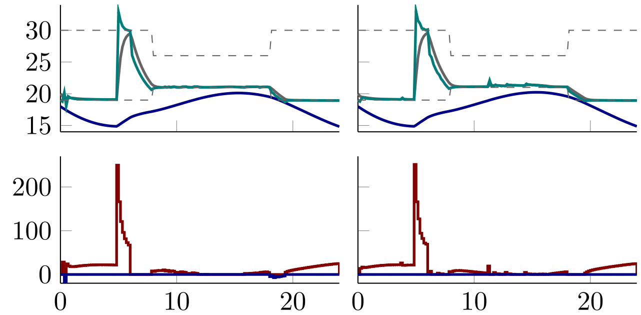

Fig. 6. Weather conditions where only heating is necessary (problem instance 20). Colors and controllers as in Fig. 5. The pre-heating around 5 am is due to the lower energy price at this time.

based on dynamic programming.

Since constraint satisfaction cannot be guaranteed by the low-level controller under Gaussian assumptions, we use soft constraints with relatively high cost to penalize constraint violation. Further, we propose a formulation that allows for marginalization of the augmented cost in closed form.

The proposed method combining reference tracking and soft constraint marginalization allows for the evaluation of the value of information. This can be used to increase the average control performance under high initial parameter uncertainty.

In simulation experiments with a simple building model, we illustrate that this method improves performance over simpler alternative approximations to dual control that are based on changes of the cost function, without an explicit model for future information gain.

References

[1] A. A. Feldbaum, "Dual Control Theory I-IV," *Avtomatika i Telemekhanika*, vol. 21(9), 21(11), 22(1), 22(2), 1960–1961.

[2] N. M. Filatov and H. Unbehauen, *Adaptive Dual Control*. Berlin: Springer Verlag, 2004.

[3] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*. NJ, USA: Prentice Hall, 1986.

[4] M. Aoki, *Optimization of Stochastic Systems*. New York - London: Academic Press, 1967.

[5] J. Sternby, "A simple dual control problem with an analytical solution," *IEEE Transactions on Automatic Control*, vol. 21, no. 6, pp. 840–844, 1976.

[6] Y. Bar-Shalom and E. Tse, "Caution, probing, and the value of information in the control of uncertain systems," *Annals of Economic and Social Measurement*, vol. 5, no. 3, pp. 323–337, 1976.

[7] E. Tse, Y. Bar-Shalom, and L. Meier III, "Wide-sense adaptive dual control for nonlinear stochastic systems," *IEEE Transactions on Automatic Control*, vol. 18, no. 2, pp. 98–108, 1973.

[8] E. Tse and Y. Bar-Shalom, "An actively adaptive control for linear systems with random parameters via the dual control approach," *IEEE Transactions on Automatic Control*, vol. 18, no. 2, pp. 109–117, 1973.

[9] E. D. Klenske and P. Hennig, "Dual control for approximate bayesian reinforcement learning," *arXiv:1510.03591*, 2015.

[10] "Energy efficiency requirements in building codes, energy efficiency policies for new buildings," 2008. [Online]. Available: https://www.iea.org/publications/freepublications/publication/Building_Codes.pdf

[11] M. Gwerder and J. Tödtli, "Predictive control for integrated room automation," in *REHVA World Congress for Building Technologies – CLIMA 2005*, 2005.

[12] F. Oldewurtel, A. Parisio, C. N. Jones, M. Morari, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and K. Wirth, "Energy efficient building climate control using stochastic model predictive control and weather predictions," in *American Control Conference (ACC)*, 2010, pp. 5100–5105.

[13] Y. Ma, J. Matusko, and F. Borrelli, "Stochastic model predictive control for building hvac systems: Complexity and conservatism," *Control Systems Technology, IEEE Transactions on*, vol. 23, no. 1, pp. 101–116, 2015.

[14] A. Aswani, N. Master, J. Taneja, D. Culler, and C. Tomlin, "Reducing transient and steady state electricity consumption in hvac using learning-based model-predictive control," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 240–253, 2012.

[15] K. J. Åström and B. Wittenmark, *Adaptive control*. Addison-Wesley, 1994.

[16] B. Wittenmark, "Adaptive dual control methods: An overview," in *In 5th IFAC symposium on Adaptive Systems in Control and Signal Processing*, 1995, pp. 67–72.

[17] Y. Cheng, S. Haghighat, and S. Di Cairano, "Robust dual control MPC with application to soft-landing control," in *American Control Conference (ACC)*, 2015, pp. 3862–3867.

[18] E. Zacekova, S. Privara, Z. Vana, J. Cigler, and L. Ferkl, "Dual control approach for zone model predictive control," in *European Control Conference (ECC)*, 2013, pp. 1398–1403.

[19] C. A. Larsson, "Application-oriented experiment design for industrial model predictive control," Ph.D. dissertation, KTH, Automatic Control, 2014.

[20] J. Rathouský and V. Havlena, "MPC-based approximation of dual control by information maximization," in *Proceedings of the 18th International Conference on Process Control*, Tatranská Lomnica, Slovakia, 2011, pp. 247–252.

[21] H. Genceli and M. Nikolaou, "New approach to constrained predictive control with simultaneous model identification," *AIChE Journal*, vol. 42, no. 10, pp. 2857–2868, 1996.

[22] G. Marafioti, R. R. Bitmead, and M. Hovd, "Persistently exciting model predictive control," *International Journal of Adaptive Control and Signal Processing*, vol. 28, no. 6, pp. 536–552, 2014.

[23] M. Hochbruck and A. Ostermann, "Exponential integrators," *Acta Numerica*, vol. 19, pp. 209–286, 2010.

[24] Y. Bar-Shalom and E. Tse, "Dual effect, certainty equivalence, and separation in stochastic control," *IEEE Transactions on Automatic Control*, vol. 19, no. 5, pp. 494–500, 1974.

[25] K. J. Åström, *Introduction to stochastic control theory*. New York, London: Academic press, 1970, vol. 70.

[26] D. J. Hughes and O. L. R. Jacobs, "Turn-off, escape and probing in nonlinear stochastic control," in *IFAC Symposium Adaptive Control*, Budapest, Hungary, 1974.

[27] M. O. G. Duff, "Optimal Learning: Computational procedures for Bayes-adaptive Markov decision processes," Ph.D. dissertation, University of Massachusetts, Amherst, 2002.

[28] P. Poupart, N. Vlassis, J. Hoey, and K. Regan, "An analytic solution to discrete Bayesian reinforcement learning," in *International Conference on Machine Learning (ICML)*, 2006.

[29] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.

[30] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, 2005.

[31] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013.

[32] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.

[33] R. Gondhalekar, F. Oldewurtel, and C. N. Jones, "Least-restrictive robust MPC of periodic affine systems with application to building climate control," in *IEEE Conference on Decision and Control (CDC)*, 2010, pp. 5257–5263.

[34] M. Maasoumy, M. Razmara, M. Shahbakhti, and A. S. Vincentelli, "Handling model uncertainty in model predictive control for energy efficient buildings," *Energy and Buildings*, 2014.

[35] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 742–753, 2006.

[36] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979.

[37] W. G. Macready and D. H. Wolpert, "Bandit problems and the exploration/exploitation tradeoff," *IEEE Transactions on Evolutionary Computation*, vol. 2, no. 1, pp. 2–22, 1998.

[38] J.-Y. Audibert, R. Munos, and C. Szepesvári, "Exploration-exploitation tradeoff using variance estimates in multi-armed bandits," *Theoretical Computer Science*, vol. 410, no. 19, pp. 1876–1902, 2009.