

# Accelerating soft-constrained MPC for linear systems through online constraint removal

S.A.N. Nouwens<sup>1</sup>, M.M. Paulides<sup>2,3</sup>, W.P.M.H. Heemels<sup>1</sup>

**Abstract**—Optimization-based controllers, such as Model Predictive Control (MPC), have attracted significant research interest due to their intuitive concept, constraint handling capabilities, and natural application to multi-input multi-output systems. However, the computational complexity of solving a receding horizon problem at each time step remains a challenge for the deployment of MPC. This is particularly the case for systems constrained by many inequalities. Recently, we introduced the concept of constraint-adaptive MPC (ca-MPC) to address this challenge for linear systems with hard constraints. In ca-MPC, at each time step, a subset of the constraints is removed from the optimization problem, thereby accelerating the optimization procedure, while resulting in identical closed-loop behavior. The present paper extends this framework to soft-constrained MPC by detecting and removing constraints based on sub-optimal predicted input sequences, which is rather easy for soft-constrained MPC due to the receding horizon principle and the inclusion of slack variables. We will translate these new ideas explicitly to an offset-free output tracking problem. The effectiveness of these ideas is demonstrated on a two-dimensional thermal transport model, showing a three order of magnitude improvement in online computational time of the MPC scheme.

## I. INTRODUCTION

Model predictive control (MPC) is a highly successful control technology in many industrial domains. Its strengths are, amongst others, the natural ability of constraint handling and straightforward application to multi-input multi-output systems. To implement this control technology, generally, MPC requires the solution of a finite horizon optimal control problem to be computed online. Despite MPC's success, for complex systems, the computational complexity of solving an optimization problem online can be a bottleneck for its real-time implementation. For this reason, there is a persistent pursuit to accelerate solving the online optimization problem. Approaches to accelerate MPC include, amongst others, advances in tailored solvers [1], [2], conditioning of the optimization problem [3], model reduction [4], (approximations of) explicit MPC [5]–[8], and constraint removal techniques [9]–[14].

In this paper, we are particularly interested in online constraint removal techniques. These techniques remove, at

each time step, a subset of the constraints to simplify the optimization problem (and thus speed up computations). Typically, online constraint removal techniques can remove more constraints from the optimization problem compared to offline methods [15], [16], as online methods can exploit specific “local” properties of the optimization problem, for instance, regions of activity [9], contraction properties of the cost function [10], [11], the refinement of coarse solutions [12], reachable sets [13], [14], and the availability of a feasible input sequence [14].

In our previous works [13], [14], we proposed a constraint-adaptive MPC (ca-MPC) framework that can be used to detect and remove constraints from the optimization problem, without changing the minimizer and thus with identical closed-loop behavior. Our framework exploits, amongst other information, feasible input sequences. Obtaining such sequences is natural in an MPC setting through, for example, extending the optimal solution from the previous time step based on the available terminal ingredients. Extending the previous optimal solution is often used to guarantee stability and recursive feasibility of the MPC feedback law. However, some MPC setups lack the formal terminal ingredients and use soft-constraints to guarantee feasibility of the optimization problem. For soft-constrained MPC setups the existing constraint removal strategies are not highly effective. Hereto, in this work, we utilize the ideas from [14] and propose a new computationally efficient constraint removal for soft-constrained MPC based on feasible input sequences. Crucially, for soft-constrained MPC, it will be trivial to generate feasible input sequences. We will demonstrate the method using an offset-free output tracking example, as this settings turns out to be particularly well-suited for our soft-constrained ca-MPC extension.

The remainder of this paper is structured as follows. We will start with the preliminaries in Section II. Second, in Section III, we will present the soft-constrained ca-MPC scheme, after which we apply it to an offset-free output tracking MPC setup. Fourth, in Section V, we demonstrate our method using a two-dimensional thermal regulation example containing 2030 inequality constraints. Finally, we state the conclusions in Section VI.

## II. PRELIMINARIES

In this section, we will introduce the system to be controlled, the MPC setup, and results for ellipsoidal sets.

This research is supported by KWF Kankerbestrijding and NWO Domain AES, as part of their joint strategic research programme: Technology for Oncology II, Learn-2-Act (17918). The collaboration project is co-funded by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

<sup>1</sup>Control Systems Technology section, Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

<sup>2</sup>Electromagnetics for Care & Cure, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

<sup>3</sup>Department of Radiotherapy, Erasmus University Medical Center Cancer Institute, Rotterdam, The Netherlands

### A. System description

In this paper, we consider systems described by the linear time-invariant (LTI) discrete-time model

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k, \quad (1a)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k. \quad (1b)$$

Here,  $\mathbf{x}_k \in \mathbb{R}^{n_x}$ ,  $\mathbf{u}_k \in \mathbb{R}^{n_u}$ , and  $\mathbf{y}_k \in \mathbb{R}^{n_y}$ , denote the state, input, and output, respectively at time  $k \in \mathbb{N}$ . Moreover,  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ ,  $\mathbf{D}$  denote matrices of appropriate dimensions.

### B. MPC setup

We will consider a MPC feedback law for an LTI model subject to linear inequality constraints and a quadratic cost function. By eliminating the dynamics, we obtain the following generic condensed MPC setup that is compatible with many typical MPC formulations with soft constraints [3], [17]:

$$\underset{\mathbf{v}, \boldsymbol{\varepsilon}}{\text{minimize}} \quad \frac{1}{2}\mathbf{v}^\top \mathbf{H}\mathbf{v} + \mathbf{v}^\top \mathbf{F}\mathbf{z}_k + \boldsymbol{\rho}^\top \boldsymbol{\varepsilon}, \quad (2a)$$

$$\text{subject to} \quad \mathbf{W}\mathbf{v} \leq \mathbf{c} + \mathbf{L}\mathbf{z}_k + \boldsymbol{\varepsilon}, \quad (2b)$$

$$\mathbf{0}_{n_c} \leq \boldsymbol{\varepsilon}. \quad (2c)$$

Here,  $\mathbf{v} \in \mathbb{R}^{n_v}$ ,  $\mathbf{z}_k \in \mathbb{R}^{n_z}$ , and  $\boldsymbol{\varepsilon} \in \mathbb{R}^{n_c}$  denote the sequence of predicted control inputs, the initial state and possibly reference information at time  $k$ , and slack variables, respectively. The matrices  $\mathbf{H} \in \mathbb{R}_{>0}^{n_v \times n_v}$  and  $\mathbf{F} \in \mathbb{R}^{n_v \times n_z}$  define the quadratic part of the cost function and  $\boldsymbol{\rho} \in \mathbb{R}_{>0}^{n_c}$  defines the penalty corresponding to a constraint violation. The matrices  $\mathbf{W} \in \mathbb{R}^{n_c \times n_v}$ ,  $\mathbf{L} \in \mathbb{R}^{n_c \times n_z}$  and vector  $\mathbf{c} \in \mathbb{R}^{n_c}$  parameterize the soft-constrained inequality constraints and  $\mathbf{0}_{n_c}$  is the zero vector of length  $n_c$  (when we omit the subscript, we assume that the dimensions are clear from context). Note that there always exists a  $\boldsymbol{\varepsilon} \geq \mathbf{0}$  sufficiently large such that a feasible solution to (2) exists.

*Remark 1:* In this paper, we consider an MPC setup where all constraints are soft constraints. Extensions to MPC setups where a subset of the constraints are soft constraints is straightforward. In such cases, it is needed to generate input sequences that are feasible with respect to the hard constraints. An easy scenario, where this is the case, is when all hard constraints are input constraints.

### C. Reduced MPC setup

As we aim to remove, at each time step, a subset of the constraints from the MPC problem (2), we introduce the reduced MPC setup given by

$$\underset{\mathbf{v}, \boldsymbol{\varepsilon}_\mathbb{A}}{\text{minimize}} \quad \frac{1}{2}\mathbf{v}^\top \mathbf{H}\mathbf{v} + \mathbf{v}^\top \mathbf{F}\mathbf{z}_k + \boldsymbol{\rho}_\mathbb{A}^\top \boldsymbol{\varepsilon}_\mathbb{A}, \quad (3a)$$

$$\text{subject to} \quad \mathbf{W}_\mathbb{A}\mathbf{v} \leq \mathbf{c}_\mathbb{A} + \mathbf{L}_\mathbb{A}\mathbf{z}_k + \boldsymbol{\varepsilon}_\mathbb{A}, \quad (3b)$$

$$\mathbf{0}_{n_r} \leq \boldsymbol{\varepsilon}_\mathbb{A}. \quad (3c)$$

Here, we use the notation  $\mathbf{W}_\mathbb{A} \in \mathbb{R}^{n_r \times n_v}$  to denote the matrix that consists of the rows in  $\mathbf{W}$  that have a row index in  $\mathbb{A} \subseteq \mathbb{N}_{[1, n_c]} = \{1, 2, \dots, n_c\}$ . This is similar for  $\mathbf{c}_\mathbb{A}$ ,  $\mathbf{W}_\mathbb{A}$ ,  $\mathbf{L}_\mathbb{A}$ ,  $\boldsymbol{\varepsilon}_\mathbb{A}$ . Note that  $\boldsymbol{\varepsilon}_\mathbb{A}$  implies that the number of decision variables is decreased as well.

As mentioned previously, we will adapt the constraints at each time step and, in particular, we will base this on the initial state  $\mathbf{z}_k$  and a feasible input sequence  $\tilde{\mathbf{v}}_k$ , which is assumed to be available at time  $k$ . More formally, given  $\tilde{\mathbf{v}}_k \in \mathbb{R}^{n_v}$ , we aim to compute  $\mathbb{A}(\mathbf{z}_k, \tilde{\mathbf{v}}_k) \subseteq \mathbb{N}_{[1, n_c]}$  such that the minimizers of (2) and (3) are identical.

### D. Ellipsoidal set

The ca-MPC method uses ellipsoidal bounds defined as

$$\mathcal{E}(\mathbf{P}, \mathbf{q}) := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{P}(\mathbf{x} - \mathbf{q})\|_2 \leq 1\} \subset \mathbb{R}^{n_v}. \quad (4)$$

Here,  $\mathbf{P} \in \mathbb{R}^{n_v \times n_v}$  is an invertible (possibly indefinite) matrix that controls the shape and size of the ellipsoid, and  $\mathbf{q} \in \mathbb{R}^{n_v}$  is a vector that represents the center of the ellipsoid.

### E. Ellipsoidal set and half-space intersection

In the proposed algorithm below, we will determine which constraint index is included in  $\mathbb{A}$  by evaluating if an ellipsoid  $\mathcal{E}(\mathbf{P}, \mathbf{q})$  is entirely contained in the half-space defined by an inequality constraint  $\{\mathbf{w} \mid \mathbf{w}\mathbf{v} \leq c\}$ . To this end, we introduce the ellipsoid half-space intersection check,

$$\|\mathbf{w}\mathbf{P}^{-1}\|_2 \leq |c - \mathbf{w}\mathbf{q}|. \quad (5)$$

If (5) holds, then  $\mathcal{E}(\mathbf{P}, \mathbf{q}) \cap \{\mathbf{w} \mid \mathbf{w}\mathbf{v} \leq c\} = \mathcal{E}(\mathbf{P}, \mathbf{q})$ , assuming  $\mathcal{E}(\mathbf{P}, \mathbf{q}) \cap \{\mathbf{w} \mid \mathbf{w}\mathbf{v} \leq c\} \neq \emptyset$ . This assumption is satisfied when  $\tilde{\mathbf{v}}_k$  does not violate any constraint [14].

## III. SOFT-CONSTRAINED CA-MPC

In this section, we will present the soft-constrained ca-MPC scheme. To this end, we will start by briefly summarizing one of the key insights from [14], where a feasible input sequence is used to bound the minimizer of a hard-constrained MPC problem. *A minimizer of a convex optimization problem satisfies first-order optimality conditions, i.e., the negative gradient of the cost function is an element of the normal cone of the constraint set [18]. Hence, outer approximations of the normal cone, based on a feasible input sequence, can be used to compute an ellipsoidal bound on the constrained minimizer.* We will, similar to this key idea in [14], use first-order optimality conditions to detect and remove constraints. In addition, we will exploit a key feature of soft constraints, i.e., it is trivial to generate feasible solutions for (2).

*Remark 2:* The presented approach can be extended to include the reachability analysis from [14] as well. We did not do this here, as we specifically focus on first-order optimality as this proved to be effective and requires minimal effort from the user to implement.

The first step in our ca-MPC algorithm is to compute a feasible solution  $(\tilde{\mathbf{v}}_k, \tilde{\boldsymbol{\varepsilon}}_k)$  to (2). To construct this solution, we start with  $\tilde{\mathbf{v}}_k \in \mathbb{R}^{n_v}$  as our predicted input sequence at time  $k$ . Then, we ensure feasibility by computing  $\tilde{\boldsymbol{\varepsilon}}_k$  as

$$\tilde{\boldsymbol{\varepsilon}}_k := \max(0, \mathbf{W}\tilde{\mathbf{v}}_k - \mathbf{c} - \mathbf{L}\mathbf{z}_k). \quad (6)$$

Trivially,  $\tilde{\boldsymbol{\varepsilon}}_k = \mathbf{0}_{n_c}$  when  $\tilde{\mathbf{v}}_k$  satisfies  $\mathbf{W}\tilde{\mathbf{v}}_k \leq \mathbf{c} + \mathbf{L}\mathbf{z}_k$ .

Based on the pair  $(\tilde{\mathbf{v}}_k, \tilde{\boldsymbol{\varepsilon}}_k)$ , we seek an ellipsoidal bound on the minimizer  $\mathbf{v}_k^* \in \mathcal{E}(\mathbf{P}, \mathbf{q})$ . This bound is then used to remove constraints from the optimization problem using the

result from Section II-E. Bounding only  $\mathbf{v}_k^*$  instead of both  $\mathbf{v}_k^*$  and  $\boldsymbol{\varepsilon}^*$  is the key insight that enables the soft-constrained ca-MPC scheme.

To obtain the ellipsoid bounding  $\mathbf{v}_k^*$ , we compute the first-order optimality conditions for the soft-constrained MPC problem (2), which yields

$$-\begin{bmatrix} \mathbf{H}\mathbf{v}_k^* + \mathbf{F}\mathbf{z}_k \\ \boldsymbol{\rho} \end{bmatrix} \in \mathcal{N}(\mathbf{v}_k^*, \boldsymbol{\varepsilon}_k^*), \quad (7a)$$

where the normal cone  $\mathcal{N}(\mathbf{v}_k^*, \boldsymbol{\varepsilon}_k^*) \subseteq \mathbb{R}^{n_v+n_c}$  is given by

$$\mathcal{N}(\mathbf{v}_k^*, \boldsymbol{\varepsilon}_k^*) = \{\mathbf{p} \in \mathbb{R}^{n_v+n_c} \mid \mathbf{p}^\top \left( \begin{bmatrix} \mathbf{v} \\ \boldsymbol{\varepsilon} \end{bmatrix} - \begin{bmatrix} \mathbf{v}_k^* \\ \boldsymbol{\varepsilon}_k^* \end{bmatrix} \right) \leq 0, \\ \text{for all } \mathbf{v}, \boldsymbol{\varepsilon} \text{ subject to (2b), (2c)}\}. \quad (7b)$$

Instead of evaluating (7b) at all pairs  $(\mathbf{v}, \boldsymbol{\varepsilon})$  subject to (2b), (2c), we outer approximate the normal cone by evaluating (7b) at  $(\tilde{\mathbf{v}}_k, \tilde{\boldsymbol{\varepsilon}}_k)$ . By evaluating (7b) for a single pair, we obtain the following necessary condition for the minimizer:

$$-\begin{bmatrix} \mathbf{H}\mathbf{v}_k^* + \mathbf{F}\mathbf{z}_k \\ \boldsymbol{\rho} \end{bmatrix}^\top \left( \begin{bmatrix} \tilde{\mathbf{v}}_k \\ \tilde{\boldsymbol{\varepsilon}}_k \end{bmatrix} - \begin{bmatrix} \mathbf{v}_k^* \\ \boldsymbol{\varepsilon}_k^* \end{bmatrix} \right) \leq 0. \quad (8)$$

After expanding and collecting the terms, we obtain

$$(\mathbf{H}\mathbf{v}_k^* + \mathbf{F}\mathbf{z}_k)^\top (\mathbf{v}_k^* - \tilde{\mathbf{v}}_k) \leq \boldsymbol{\rho}^\top (\tilde{\boldsymbol{\varepsilon}}_k - \boldsymbol{\varepsilon}_k^*). \quad (9)$$

The inequality (9) denotes an ellipse in  $\mathbf{v}_k^*$  that scales with  $\boldsymbol{\rho}^\top \boldsymbol{\varepsilon}_k^*$ . To obtain an ellipsoidal bound on  $\mathbf{v}_k^*$ , that does not depend on  $\boldsymbol{\varepsilon}_k^*$ , we use the observation that  $0 \leq \boldsymbol{\rho}^\top \boldsymbol{\varepsilon}_k^*$ , see (2c). This observation enables the upper bound  $\boldsymbol{\rho}^\top (\tilde{\boldsymbol{\varepsilon}}_k - \boldsymbol{\varepsilon}_k^*) \leq \boldsymbol{\rho}^\top \tilde{\boldsymbol{\varepsilon}}_k$ , which leads to the bound on  $\mathbf{v}_k^*$ ,

$$(\mathbf{H}\mathbf{v}_k^* + \mathbf{F}\mathbf{z}_k)^\top (\mathbf{v}_k^* - \tilde{\mathbf{v}}_k) \leq \boldsymbol{\rho}^\top \tilde{\boldsymbol{\varepsilon}}_k. \quad (10)$$

Which is equivalent to  $\mathbf{v}_k^* \in \mathcal{E}(\mathbf{P}, \mathbf{q})$ , with

$$\mathbf{P} = \frac{1}{\sqrt{\sigma}} \mathbf{G}, \quad \mathbf{q} = \frac{1}{2} (\tilde{\mathbf{v}}_k - \mathbf{H}^{-1} \mathbf{F}\mathbf{z}_k), \quad \mathbf{H} = \mathbf{G}^\top \mathbf{G}, \quad (11a)$$

$$\sigma = \boldsymbol{\rho}^\top \tilde{\boldsymbol{\varepsilon}}_k + \frac{1}{4} \|\mathbf{G}(\tilde{\mathbf{v}}_k + \mathbf{H}^{-1} \mathbf{F}\mathbf{z}_k)\|_2^2. \quad (11b)$$

Based on (11), we observe that  $\mathbf{v}_k^*$  is bounded by an ellipse centered at the mean between  $-\mathbf{H}^{-1} \mathbf{F}\mathbf{z}_k$  and  $\tilde{\mathbf{v}}_k$ , and inflated based on the degree of infeasibility and distance between  $\tilde{\mathbf{v}}_k$  and  $-\mathbf{H}^{-1} \mathbf{F}\mathbf{z}_k$ .

*Remark 3:* The expression for  $\sigma$  provides a guideline on how we should choose  $\tilde{\mathbf{v}}_k$ , as, typically, we want  $\sigma$  to be small. Hence, on one hand we want to be close to the unconstrained minimizer to reduce  $\|\mathbf{G}(\tilde{\mathbf{v}}_k + \mathbf{H}^{-1} \mathbf{F}\mathbf{z}_k)\|_2^2$ , while on the other hand we would like to reduce the infeasibility of our prediction to make  $\boldsymbol{\rho}^\top \tilde{\boldsymbol{\varepsilon}}_k$  small. Therefore, there is a balancing act between reducing infeasibility and remaining close to the unconstrained minimizer. Of course, when the unconstrained minimizer is feasible, we have trivially solved the optimization problem.

Based on the ellipsoidal bound, we use the inequality from Section II-E to define the set-valued mapping  $\mathbb{A}(\mathbf{z}_k, \tilde{\mathbf{v}}_k)$ ,

$$\mathbb{A}(\mathbf{z}_k, \tilde{\mathbf{v}}_k) = \{j \in \mathbb{N}_{[1, n_c]} \mid \sqrt{\sigma} \|\mathbf{W}_j \mathbf{G}^{-1}\|_2 > |\mathbf{c}_j + \mathbf{L}_j \mathbf{z}_k - \mathbf{W}_j \mathbf{q}|\}. \quad (12)$$

Here, we use  $\mathbf{W}_j$  to denote the  $j$ -th row from  $\mathbf{W}$  (the same holds for  $\mathbf{c}_j$  and  $\mathbf{L}_j$ ). An interesting observation is that (12) does not explicitly specify that the constraints violated by

$\tilde{\mathbf{v}}_k$  are included in  $\mathbb{A}$ . As it turns out, depending on the design of  $\boldsymbol{\rho}$ , e.g.,  $\boldsymbol{\rho}$  sufficiently large [17], these constraints are automatically included by (12). However, when  $\boldsymbol{\rho}$  is not sufficiently large, we may encounter the situation where all elements of  $\mathcal{E}(\mathbf{P}, \mathbf{q})$  violate a particular constraint. This violates the assumption that the intersection between the ellipsoid and half-space for a particular inequality constraint is non-empty, see Section II-E. To alleviate this problem and avoid complex design rules on  $\boldsymbol{\rho}$ , we modify (12) by including all constraints that are violated by  $\tilde{\mathbf{v}}_k$ ,

$$\mathbb{A}(\mathbf{z}_k, \tilde{\mathbf{v}}_k) = \{j \in \mathbb{N}_{[1, n_c]} \mid \sqrt{\sigma} \|\mathbf{W}_j \mathbf{G}^{-1}\|_2 > |\mathbf{c}_j + \mathbf{L}_j \mathbf{z}_k - \mathbf{W}_j \mathbf{q}| \text{ or } \tilde{\boldsymbol{\varepsilon}}_{j,k} > 0\}, \quad (13)$$

where  $\tilde{\boldsymbol{\varepsilon}}_{j,k}$  denotes the  $j$ -th element of  $\tilde{\boldsymbol{\varepsilon}}_k$ .

Observe that a significant fraction of (13) can be either pre-computed, e.g.,  $\|\mathbf{W}_j \mathbf{G}^{-1}\|_2$ , or is already required to set up the MPC problem, e.g.,  $\mathbf{F}\mathbf{z}_k$  and  $\mathbf{c}_j + \mathbf{L}_j \mathbf{z}_k$ , so, these do not lead to additional computational costs. The number of operations specific to the ca-MPC framework is  $2n_c(n_v + 1) + 3n_v^2 + n_v + 1$ , which, crucially, scales linear in  $n_c$ .

Last, we provide an overview of the method in Algorithm 1. Note that the only ca-MPC specific offline setup is pre-computing  $\|\mathbf{W}_j \mathbf{G}^{-1}\|_2$  in lines 2-5. In line 7, we measure  $\mathbf{z}_k$ , which is, for example, a stacked vector containing the current state, previous input, and information regarding a reference trajectory. In lines 8-11, we choose  $\tilde{\mathbf{v}}_k$  based on the shifted previous optimal minimizer  $\mathbf{v}_{[n_u+1:Nn_u], k-1}^*$ , where the subscript  $[n_u+1:Nn_u], k-1$  denotes that we select the  $(n_u+1)$ -th up to  $Nn_u$ -th elements at time  $k-1$ . When the minimizer at time  $k-1$  is not available, e.g., at time  $k=0$ , we use the unconstrained minimizer  $-\mathbf{H}^{-1} \mathbf{F}\mathbf{z}_k$ . Recall that the method allows all  $\tilde{\mathbf{v}}_k \in \mathbb{R}^{n_v}$ , but certain choices show better constraint removal properties, see also Remark 3 above. In line 20, we extract the rows from the constraint matrices and vectors based on  $\mathbb{A}$ ; this is required to build the reduced MPC problem. In lines 21-23, we solve the quadratic program, compute the optimal input based on the minimizer, and apply the input to the plant. Note that  $\mathbf{v}_k$  does not necessarily contain the input, e.g., when we optimize over delta's of the input relative to  $\mathbf{u}_{k-1}$ .

#### IV. OFFSET-FREE OUTPUT TRACKING MPC

The soft-constrained ca-MPC scheme described in the previous section applies to all MPC setups that can be captured in the formulation (2), which includes a large class of control problems for linear plant models. In this section, we will focus on an offset-free output tracking problem as it is both an industry-relevant MPC setup and particularly well-suited to our soft-constrained ca-MPC scheme.

The offset-free output tracking MPC setup is given by the

optimization problem [19]

$$\text{minimize} \quad \sum_{i=1}^N \|\mathbf{C}\mathbf{x}_{i|k} - \mathbf{y}_{k+i}^{\text{ref}}\|_Q^2 + \|\Delta\mathbf{u}_{i-1|k}\|_R^2, \quad (14a)$$

$$+ \boldsymbol{\rho}_x^\top \boldsymbol{\varepsilon}_{x,i} + \boldsymbol{\rho}_u^\top \boldsymbol{\varepsilon}_{u,i-1} + \boldsymbol{\rho}_\Delta^\top \boldsymbol{\varepsilon}_{\Delta,i-1},$$

$$\text{subject to} \quad \mathbf{x}_{i+1|k} = \mathbf{A}\mathbf{x}_{i|k} + \mathbf{B}\mathbf{u}_{i|k}, \quad i \in \mathbb{N}_{[0,N-1]}, \quad (14b)$$

$$\mathbf{u}_{i|k} = \mathbf{u}_{i-1|k} + \Delta\mathbf{u}_{i|k}, \quad i \in \mathbb{N}_{[0,N-1]}, \quad (14c)$$

$$\mathbf{x}_{0|k} = \mathbf{x}_k, \quad \mathbf{u}_{-1|k} = \mathbf{u}_{k-1}, \quad (14d)$$

$$\mathbf{M}_x \mathbf{x}_{i|k} \leq \mathbf{g}_x + \boldsymbol{\varepsilon}_{x,i}, \quad i \in \mathbb{N}_{[1,N]}, \quad (14e)$$

$$\mathbf{M}_u \mathbf{u}_{i|k} \leq \mathbf{g}_u + \boldsymbol{\varepsilon}_{u,i}, \quad i \in \mathbb{N}_{[0,N-1]}, \quad (14f)$$

$$\mathbf{M}_\Delta \Delta\mathbf{u}_{i|k} \leq \mathbf{g}_\Delta + \boldsymbol{\varepsilon}_{\Delta,i}, \quad i \in \mathbb{N}_{[0,N-1]}, \quad (14g)$$

$$\mathbf{0} \leq \boldsymbol{\varepsilon}_{x,i}, \quad i \in \mathbb{N}_{[1,N]}, \quad (14h)$$

$$\mathbf{0} \leq \boldsymbol{\varepsilon}_{u,i}, \quad \mathbf{0} \leq \boldsymbol{\varepsilon}_{\Delta,i}, \quad i \in \mathbb{N}_{[0,N-1]}. \quad (14i)$$

Here, we use  $\|\Delta\mathbf{u}\|_R^2$  to denote  $\Delta\mathbf{u}^\top \mathbf{R} \Delta\mathbf{u}$  and the  $i|k$  subscripts denotes the  $i$ -th prediction step made at time  $k$ . In (14), matrices  $\mathbf{Q} \in \mathbb{R}^{n_y \times n_y}$  and  $\mathbf{R} \in \mathbb{R}^{n_u \times n_u}$  denote a positive semi-definite and a positive definite matrix, respectively. The vector  $\mathbf{y}_k^{\text{ref}} \in \mathbb{R}^{n_y}$  denotes the reference and the matrix  $\mathbf{M}$  and vector  $\mathbf{g}$  define the inequality constraints for, respectively, the state, input, and delta input, depending on the subscript. We optimize (14) with respect to  $\mathbf{x}_{i|k}$ ,  $\mathbf{u}_{i-1|k}$ ,  $\Delta\mathbf{u}_{i-1|k}$ ,  $\boldsymbol{\varepsilon}_{x,i}$ ,  $\boldsymbol{\varepsilon}_{u,i-1}$ , and  $\boldsymbol{\varepsilon}_{\Delta,i-1}$  for  $i \in \mathbb{N}_{[1,N]}$ .

To transform (14) into (2), we define

$$\mathbf{z}_k = [\mathbf{x}_k^\top \quad \mathbf{u}_{k-1}^\top \quad \mathbf{y}_{k+1}^{\text{ref}\top} \quad \cdots \quad \mathbf{y}_{k+N}^{\text{ref}\top}]^\top, \quad (15a)$$

$$\mathbf{v}_k = [\Delta\mathbf{u}_{0|k}^\top \quad \cdots \quad \Delta\mathbf{u}_{N-1|k}^\top]^\top, \quad (15b)$$

$$\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_{x,1}^\top \quad \boldsymbol{\varepsilon}_{u,0}^\top \quad \boldsymbol{\varepsilon}_{\Delta,0}^\top \quad \cdots \quad \boldsymbol{\varepsilon}_{x,N}^\top \quad \boldsymbol{\varepsilon}_{u,N-1}^\top \quad \boldsymbol{\varepsilon}_{\Delta,N-1}^\top]^\top, \quad (15c)$$

after which transformations that condense the MPC problem can be applied to obtain  $\mathbf{H}$ ,  $\mathbf{G}$ ,  $\mathbf{F}$ ,  $\mathbf{W}$ ,  $\mathbf{c}$ ,  $\mathbf{L}$ , and  $\boldsymbol{\rho}$ .

Based on the MPC formulation (14), we can analyze why (14) is well-suited to the soft-constrained ca-MPC scheme. First of all, the delta-input enables a particularly good and easy-to-compute input sequence, namely,  $\tilde{\mathbf{v}}_k = [\Delta\mathbf{u}_{1|k-1}^{\star\top} \quad \cdots \quad \Delta\mathbf{u}_{N-1|k-1}^{\star\top} \quad \mathbf{0}_{n_u}^\top]^\top$ . Indeed, we can simply extend the previous optimal solution with a zero input, which is the steady-state solution for a constant reference. Moreover, recall that the size of the ellipsoid bounding the minimizers is determined, in part, by  $\boldsymbol{\rho}^\top \tilde{\boldsymbol{\varepsilon}}_k$ . The particular choice of  $\tilde{\mathbf{v}}_k$  is expected to result in small constraint violations, i.e., we expect  $\boldsymbol{\rho}^\top \tilde{\boldsymbol{\varepsilon}}_k$  to be small.

The second reason why (14) is particularly well suited to the soft-constrained ca-MPC scheme, is that the objective is to track the reference well. In other words, we can expect to be ‘‘close’’ to the reference in some sense. To see why this is valuable, recall that the ellipsoid  $\mathcal{E}(\mathbf{P}, \mathbf{q})$  also scales based on the distance between the unconstrained minimizer  $-\mathbf{H}^{-1}\mathbf{F}\mathbf{z}_k$ , and the predicted input sequence  $\tilde{\mathbf{v}}_k$ . This means that depending on the output reference, the unconstrained minimizer is expected to be close to the predicted input sequence. For example, when moving between two set points, the unconstrained minimizer will be closer to  $\tilde{\mathbf{v}}_k$ , when we ramp between them as opposed to instantaneously switching. Hence, we expect to remove more constraints when we have an appropriate reference that

connects the set points. We will demonstrate the effectiveness of the constraint removal method on an offset-free output tracking problem in the next section.

---

#### Algorithm 1: The soft-constrained ca-MPC scheme

---

```

1:  $k \leftarrow 0$ ,  $\mathbf{H} \in \mathbb{R}_{>0}^{n_v \times n_v}$ ,  $\mathbf{G} \in \mathbb{R}^{n_v \times n_v}$ ,  $\mathbf{F} \in \mathbb{R}^{n_v \times n_z}$ ,
    $\mathbf{W} \in \mathbb{R}^{n_c \times n_v}$ ,  $\mathbf{L} \in \mathbb{R}^{n_c \times n_z}$ ,  $\mathbf{c} \in \mathbb{R}^{n_c}$ ,  $\boldsymbol{\rho} \in \mathbb{R}_{>0}^{n_c}$ ,  $j \leftarrow 1$ 
2: for  $j \leq n_c$  do
3:    $\zeta_j \leftarrow \|\mathbf{W}_j \mathbf{G}^{-1}\|_2$ 
4:    $j \leftarrow j + 1$ 
5: while true do
6:   MEASURE  $\mathbf{z}_k$ 
7:   if  $k = 0$  then
8:      $\tilde{\mathbf{v}}_k \leftarrow -\mathbf{H}^{-1}\mathbf{F}\mathbf{z}_k$ 
9:   else
10:     $\tilde{\mathbf{v}}_k \leftarrow [\mathbf{v}_{[n_u+1:Nn_u],k-1}^{\star\top} \quad \mathbf{0}_{n_u}^\top]^\top$ 
11:     $\tilde{\boldsymbol{\varepsilon}}_k \leftarrow \max(0, \mathbf{W}\tilde{\mathbf{v}}_k - \mathbf{c} - \mathbf{L}\mathbf{z}_k)$ 
12:     $\mathbf{q} \leftarrow \frac{1}{2}(\tilde{\mathbf{v}}_k - \mathbf{H}^{-1}\mathbf{F}\mathbf{z}_k)$ 
13:     $\sigma \leftarrow \boldsymbol{\rho}^\top \tilde{\boldsymbol{\varepsilon}}_k + \frac{1}{4}\|\mathbf{G}(\tilde{\mathbf{v}}_k + \mathbf{H}^{-1}\mathbf{F}\mathbf{z}_k)\|_2^2$ 
14:     $j \leftarrow 1$ ,  $\mathbb{A} \leftarrow \emptyset$ 
15:    for  $j \leq n_c$  do
16:      if  $\sqrt{\sigma}\zeta_j > |\mathbf{c}_j + \mathbf{L}_j\mathbf{z}_k - \mathbf{W}_j\mathbf{q}|$  or  $\tilde{\boldsymbol{\varepsilon}}_{j,k} > 0$  then
17:         $\mathbb{A} \leftarrow \mathbb{A} \cup \{j\}$ 
18:       $j \leftarrow j + 1$ 
19:    EXTRACT  $\mathbf{W}_\mathbb{A}$ ,  $\mathbf{L}_\mathbb{A}$ ,  $\mathbf{c}_\mathbb{A}$ ,  $\boldsymbol{\rho}_\mathbb{A}$ 
20:     $(\mathbf{v}_k^\star, \boldsymbol{\varepsilon}_k^\star) \leftarrow \text{SOLVE (3)}$ 
21:    EXTRACT  $\mathbf{u}_k \leftarrow (\mathbf{v}_k^\star, \mathbf{z}_k)$ 
22:    APPLY  $\mathbf{u}_k$  TO PLANT
23:     $k = k + 1$ 

```

---

## V. NUMERICAL CASE STUDY

In this section, we will demonstrate the soft-constrained ca-MPC scheme using a two-dimensional thermal regulation example. Here, we aim to heat a particular region without exceeding temperature upper bounds on our entire two-dimensional domain, see Figure 1. This type of problem is relevant for hyperthermia-enhanced cancer treatments, see, e.g., [20]. We will start by defining our system and MPC setup. Hereafter, we will compare the computational complexity of our ca-MPC scheme and the original MPC.

### A. System definition and MPC setup

We consider a thermal system modeled by the following partial differential equation

$$\dot{T}(r, t) = \alpha \nabla^2 T(r, t) + \beta T(r, t) + \sum_{j=1}^3 \gamma_j u_j(t), \quad (16a)$$

for  $r \in \Omega$ , with boundary condition

$$\alpha \nabla T(r, t) \cdot \mathbf{n} = T(r, t), \quad r \in \partial\Omega. \quad (16b)$$

Here,  $\Omega = [0, 1] \times [0, 1]$ ,  $T : \Omega \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ ,  $\alpha = 2.5 \cdot 10^{-4}$ ,  $\beta = 2 \cdot 10^{-2}$ ,  $\gamma_j : \Omega \rightarrow \mathbb{R}_{\geq 0}$ ,  $u_j : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ , and  $\mathbf{n} \in \mathbb{R}^2$ , denote the domain, temperature, diffusivity, damping, distributed heat load, control input, and outward facing surface normal, respectively. As mentioned, we constrain the temperature below an upper bound, as given by  $T(r, t) \leq \bar{T}(r)$  for  $r \in \Omega$ , see Figure 1.

After spatially and temporally discretizing (16) on a square  $20 \times 20$  grid with a sample time of 1 second, we obtain a

discrete-time state-space model (1) with  $n_x = 400$  states and  $n_y = 25$  outputs, again, see Figure 1. To obtain an offset-free MPC scheme, we write the input in incremental form:  $\mathbf{u}_{k+1} = \mathbf{u}_k + \Delta \mathbf{u}_k$ . The heat loads  $\gamma_i$  on the discrete spatial grid are shown in Figure 2. Similarly, discretization of the temperature upper bound gives  $M_x = I_{n_x}$  and  $\mathbf{g}_x = \bar{T} \in \mathbb{R}^{n_x}$ , where  $I_{n_x}$  denotes the identity matrix of size  $n_x$  and  $\bar{T}$  is shown in Figure 1. Additionally, we constrain our input to be non-negative and limited in power by  $M_u = I_{n_u} \otimes \begin{bmatrix} 1 \\ -1 \end{bmatrix}$  and  $\mathbf{g}_u = \mathbf{1}_m \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ , where  $\otimes$  denotes the Kronecker product. The non-negative input in combination with the non-negative  $\gamma_i$  ensure we cannot actively apply cooling, as is typical in hyperthermia cancer treatments. We choose  $\mathbf{Q} = I_{n_y}$ ,  $\mathbf{R} = I_{n_u}$ , and  $\rho = \mathbf{1}_{n_c}$ . Using  $M_x$ ,  $\mathbf{g}_x$ ,  $M_u$ ,  $\mathbf{g}_u$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$  and choosing a horizon  $N = 5$ , we can transform the output tracking MPC (14) to (2) using the definitions from Section IV. Note that our MPC scheme has a total of  $n_c = 2030$  constraints and  $n_v = 15$  control inputs. Recall that we obtain many constraints as the upper bound on the continuous spatial domain is discretized. We define the output reference as a linear ramp to the target temperature,  $\mathbf{y}_k^{\text{ref}} = \mathbf{1}_{n_y} \min(10, 10 \frac{k}{30})$ . We solve the resulting quadratic programs using a primal-dual interior point solver from the Matlab MPC toolbox [21].

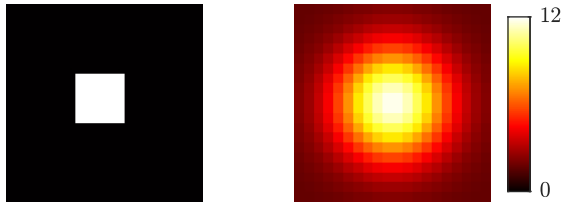


Fig. 1: Left: the union between the black and white pixels denote the domain  $\Omega$  and the white pixels denote the outputs. Right: the (Gaussian) temperature upper bound.

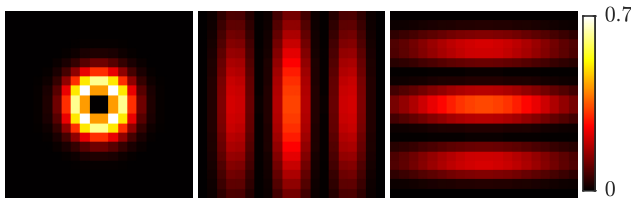


Fig. 2: The spatially distributed heat loads  $\gamma_1$ ,  $\gamma_2$ , and  $\gamma_3$  (from left to right).

### B. Results

In this section, we present the results comparing the soft-constrained ca-MPC scheme to the MPC setup (2). Recall that the ca-MPC setup has the exact same closed-loop behavior original soft-constrained MPC setup. Interestingly, when  $\rho$  is sufficiently large, the soft-constrained MPC scheme is equivalent to the corresponding hard-constrained MPC setup [17]. First, in Figure 3, we show the temperature evolution over time for the ca-MPC scheme in combination with a visualization of the constraints that are included in

the reduced MPC problem. As expected, the temperature in the center rises to follow the reference. However, crucially, only a small number of constraints at the discrete locations are considered in the reduced MPC problem. In Figure 4, we show how the number of constraints evolves over time. Note that the number of constraints does not exceed 62, which is 3.1% of the original number of constraints.

The significant reduction in constraints is also reflected in the computation time, see Figure 5. Here, our ca-MPC setup obtains a three orders of magnitude improvement in computation time compared to the original MPC setup. From Figure 5, we observe that the time spent to compute  $\mathbb{A}$  is negligible with respect to solving the resulting simplified quadratic program (QP), let alone compared to solving the original QP. This observation exemplifies the effectiveness and computational efficiency of the ca-MPC method. Indeed, for many MPC problems, determining which constraints can be excluded from the QP (computing  $\mathbb{A}$ ) can be performed in a negligible amount of time. Recall that by using soft constraints, removing a constraint from the optimization problem simultaneously removes a decision variable from the QP, further simplifying the resulting ca-MPC scheme.

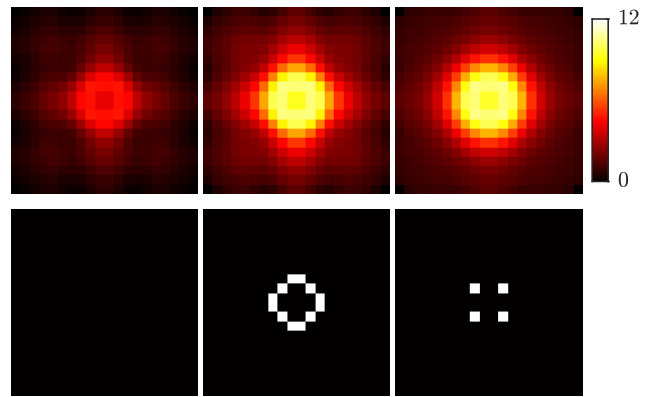


Fig. 3: Top: the temperature at time steps 15, 30, and 45, respectively. Bottom: white pixels indicate that the corresponding constraint is added to  $\mathbb{A}$  for at least one step in the horizon. Note that at time step 30, we approximately obtain the maximum number of constraints in the MPC problem.

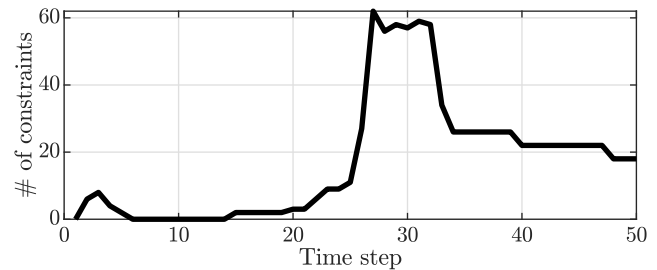


Fig. 4: The number of constraints in the reduced MPC problem over time. The maximum number of constraints (62) is obtained at time step 27. Recall that the total number of constraints is 2030.

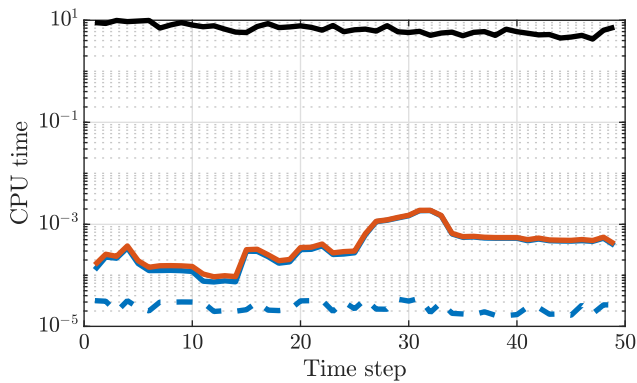


Fig. 5: Computation time comparison between the original MPC setup (2) (—), and the ca-MPC setup (3) (—). The time spent by soft-constrained ca-MPC is broken down in all time required to compute  $\Delta$  (- - -) and solving the resulting quadratic program (—).

## VI. CONCLUSION AND OUTLOOK

In this paper, we presented an extension to ca-MPC that allows for online constraint removal using easy-to-compute input sequences for soft-constrained MPC. Crucially, our constraint removal scheme does not change the minimizer of the original MPC problem and, thus, the ca-MPC scheme inherits all stability and performance properties from the original MPC setup. The presented method uses an a priori generated input sequence to compute ellipsoidal bounds on the constrained minimizer to detect and remove inequality constraints that cannot be reached within the ellipsoidal bound. Similar to existing ca-MPC schemes, our extension has a low computational overhead, making it applicable to a large class of linear systems and MPC setups. We demonstrated the ca-MPC scheme on a two-dimensional thermal regulation problem with a temperature upper bound. The regular MPC setup had 2030 constraints, while our ca-MPC scheme only required a maximum of 62, while guaranteeing the same closed-loop performance. This resulted in a computational speed-up of three orders of magnitude.

Exploiting the optimality properties of MPC proves to be effective at accelerating the computation of the online optimization problem. Future research interests include extensions to different cost functions and direct integration into specific optimization solvers, such as interior-point and active-set solvers. Other subjects of interest include different handling of the soft constraints, e.g., using one slack variable to bound all constraints and extensions to different slack variable penalties, such as mixed quadratic and linear penalties. Extensions to proximal point methods are also of interest. These methods solve a sequence of well-conditioned optimization problems, where each minimizer is close to the previous one. This property has the potential for highly effective constraint removal schemes.

## REFERENCES

[1] G. Frison and M. Diehl, "HPIPM: a high-performance quadratic programming framework for model predictive control," in *Proceedings of the IFAC World Congress*, 2020.

[2] D. Arnström, A. Bemporad, and D. Axehill, "A Dual Active-Set Solver for Embedded Quadratic Programming Using Recursive LDL<sup>T</sup> Updates," *IEEE Trans. Autom. Control*, vol. 67, no. 8, pp. 4362–4369, 2022.

[3] J. Jerez, E. Kerrigan, and G. Constantinides, "A condensed and sparse QP formulation for predictive control," in *IEEE Conf. Decis. Control*, 2011, pp. 5217–5222.

[4] S. Hovland, K. Willcox, and J. T. Gravdahl, "MPC for Large-Scale Systems via Model Reduction and Multiparametric Quadratic Programming," in *IEEE Conf. Decis. Control*, 2006, pp. 3418–3423.

[5] A. Bemporad, F. Borrelli, and M. Morari, "Model predictive control based on linear programming - the explicit solution," *IEEE Trans. on Automat. Contr.*, vol. 47, no. 12, pp. 1974–1985, 2002.

[6] A. Bemporad, A. Oliveri, T. Poggi, and M. Storaice, "Ultra-fast stabilizing model predictive control via canonical piecewise affine approximations," *IEEE Trans. Automat. Contr.*, vol. 56, no. 12, pp. 2883–2897, 2011.

[7] M. Kvasnica and M. Fikar, "Clipping-Based Complexity Reduction in Explicit MPC," *IEEE Trans. Autom. Control*, vol. 57, no. 7, pp. 1878–1883, 2011.

[8] B.A.G. Genuit, L. Lu, and W.P.M.H. Heemels, "Approximation of explicit model predictive control using regular piecewise affine functions: an input-to-state stability approach," *IET Control Theory & Applications*, vol. 6, no. 8, pp. 1015–1028, 2012.

[9] M. Jost and M. Mönnigmann, "Accelerating model predictive control by online constraint removal," in *IEEE Conf. Decis. Control*, 2013, pp. 5764–5769.

[10] M. Jost, G. Pannocchia, and M. Mönnigmann, "Online constraint removal: Accelerating MPC with a Lyapunov function," *Automatica*, vol. 57, pp. 164–169, 2015.

[11] R. Dyrská, M. Horváthová, P. Bakaráč, M. Mönnigmann, and J. Oravec, "Heat exchanger control using model predictive control with constraint removal," *Appl. Therm. Eng.*, vol. 227, p. 120366, 2023.

[12] Y. Nie and E. C. Kerrigan, "External Constraint Handling for Solving Optimal Control Problems With Simultaneous Approaches and Interior Point Methods," *IEEE Control Syst. Lett.*, vol. 4, no. 1, pp. 7–12, 2019.

[13] S. Nouwens, B. de Jager, M. Paulides, and W. Heemels, "Constraint Removal for MPC with Performance Preservation and a Hyperthermia Cancer Treatment Case Study," in *IEEE Conf. Decis. Control*, 2021, pp. 4103–4108.

[14] S. Nouwens, M. Paulides, and W. Heemels, "Constraint-adaptive MPC for linear systems: A system-theoretic framework for speeding up MPC through online constraint removal," *Automatica*, vol. 157, p. 111243, 2023.

[15] S. Paulraj and P. Sumathi, "A Comparative Study of Redundant Constraints Identification Methods in Linear Programming Problems," *Mathematical Problems in Engineering*, pp. 1–16, 2010.

[16] L. Roald and D. Molzahn, "Implied constraint satisfaction in power system optimization: The impacts of load variations," in *Allert. Conf. Commun. Control. Comput.*, 2019, pp. 308–315.

[17] E. C. Kerrigan and J. M. Maciejowski, "Soft constraints and exact penalty functions in model predictive control," 2000.

[18] J. Nocedal and S. Wright, *Numerical Optimization*, ser. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006, vol. 83.

[19] G. Pannocchia, "Offset-free tracking MPC: A tutorial review and comparison of different formulations," in *IEEE European Control Conference*, 2015, pp. 15–17.

[20] D. A. Deenen, B. Maljaars, L. C. Sebeke, B. de Jager, E. Heijman, H. Grill, and W. P. M. H. Heemels, "Offset-Free Model Predictive Temperature Control for Ultrasound-Based Hyperthermia Cancer Treatments," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 6, pp. 2351–2365, 2020.

[21] The MathWorks Inc., "MPC Toolbox version: R2021a," Natick, Massachusetts, United States. [Online]. Available: <https://www.mathworks.com>