A Simple and Fast Coordinate-Descent Augmented-Lagrangian Solver for Model Predictive Control

Liang Wu and Alberto Bemporad, *Fellow, IEEE*

***Abstract*—This article proposes a novel coordinate-descent augmented-Lagrangian (CDAL) solver for linear, possibly parameter-varying, model predictive control (MPC) problems. At each iteration, an augmented Lagrangian (AL) subproblem is solved by coordinate descent (CD), exploiting the structure of the MPC problem. The CDAL solver enjoys three main properties: 1) it is *construction-free*, in that it avoids explicitly constructing the quadratic programming problem associated with MPC; 2) is *matrix-free*, as it avoids multiplications and factorizations of matrices; and 3) is *library-free*, as it can be simply coded without any library dependency, 90-lines of C-code in our implementation. To favor the convergence speed, CDAL employs a reverse cyclic rule for the CD method, the accelerated Nesterov's scheme for updating the dual variables, a simple diagonal preconditioner, and an efficient coupling scheme between the CD and AL methods. We show that CDAL competes with other state-of-the-art methods, both in the case of unstable linear time-invariant and linear parameter-varying prediction models.**

***Index Terms*—Augmented Lagrangian (AL) method, coordinate descent (CD) method, model predictive control (MPC).**

I. Introduction

Model predictive control (MPC) has been widely used for decades to control multivariable systems subject to input and output constraints [1]. Apart from small-scale linear time-invariant (LTI) MPC problems whose explicit MPC control law can be obtained [2], deploying an MPC controller in an electronic control unit requires an embedded quadratic programming (QP) solver. In the past decades, the MPC community has made tremendous research efforts to develop embedded QP algorithms [3], based on interior-point methods [4], [5], active-set algorithms [6], [7], gradient projection methods [8], the alternating direction method of multipliers (ADMM) [9], [10], and other techniques [11], [12], [13], [14], [15].

A demanding requirement for industrial MPC applications is code simplicity, for easily being verified, validated, and maintained on embedded platforms. In this respect, the interior-point and active-set methods require more complicated arithmetic operations in their algorithm implementations when compared to first-order optimization methods like gradient projection and ADMM. The first-order optimization methods are quite appealing in embedded MPC since their embedded implementations could only involve additions and multiplications (no divisions, square roots, etc.). However, most of the proposed approaches require that the MPC-to-QP transformation is explicitly constructed for consumption by the solver, such as for preconditioning, estimating the Lipschitz constant of the cost gradient, and factorizing matrices. This may not be an issue for LTI MPC problems, in which the MPC-to-QP construction and other operations on the problem matrices can be done offline. But for some linear parameter-varying (LPV) or for linear time-varying MPC problems in which the linear dynamic model, cost function, and/or constraints change at run time, an explicit online MPC-to-QP construction increases the complexity of the embedded code and computation time. Avoiding an explicit MPC-to-QP construction can be called as construction-free property of an MPC solver. The barrier interior-point FastMPC solver [4] and the active-set based bounded-variable least-squares (BVLS) solver [15] are construction-free; they directly use the model and weight matrices to define the MPC problem without constructing a QP problem. Their complicated implementations are not matrix-free as involving Cholesky or QR factorizations arithmetic operations during iterations. The well-known simple and efficient first-order method OSQP [10] is not construction-free and matrix-free when applied to solve LPV-MPC problems, as it requires that matrix factorizations are computed and cached on each sampling time. The OSQP utilizes its own LDL^T solver to perform matrix factorizations, thus being library-free.

A. Contribution

By combining the coordinate descent (CD) and augmented-Lagrangian (AL) methods, in this article, we develop a construction-free, matrix-free, and library-free solver for LTI and LPV MPC problems that is particularly suitable for embedded industrial deployment.

Coordinate descent has received extensive attention in recent years due to its application to machine learning [16], [17], [18]. In this article, we will exploit the special structure arising from linear MPC formulations when applying CD. In [19], [20], and [21], the authors also use AL to solve linear MPC problems with input and state constraints using the fast gradient method [22] to solve the associated subproblems. The Lipschitz constant of the cost gradient and convexity parameters [19] are needed to achieve convergence, and computing them requires in turn the Hessian matrix of the subproblem, and hence constructing the QP problem. As the Hessian matrix of the AL subproblem is close to a block diagonal matrix, this suggests the use of the CD method to solve such a QP subproblem, due to the fact that CD does not require any problem-related parameter. Moreover, only small matrices are involved in running the CD method, namely the matrices of the linear prediction model and the weight matrices. As a result, the proposed CDAL algorithm does not require the QP construction phase and is extremely simple to implement. In addition, each update of the optimization vector has a computation cost per iteration that is quadratic with the state and input dimensions and linear with the prediction horizon.

To improve the convergence speed of CDAL, we propose four techniques: a reverse cyclic rule for CD, Nesterov's acceleration [22], preconditioning, and an efficient coupling between CD and AL. While the use of a reverse cyclic rule in CD still preserves convergence, when

Manuscript received 20 May 2022; revised 25 September 2022; accepted 26 January 2023. Date of publication 31 January 2023; date of current version 26 October 2023. Recommended by Associate Editor Eric C. Kerrigan. *(Corresponding author: Liang Wu.)*

The authors are with the IMT School for Advanced Studies Lucca, 55100 Lucca, Italy (e-mail: liang.wu@imtlucca.it; alberto.bemporad@imtlucca.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2023.3241238.

Digital Object Identifier 10.1109/TAC.2023.3241238

the MPC problem is solved by warm-starting it from the shifted previous optimal solution, the gap between the initial guess and the new optimal solution is mainly caused by the last block of variables, and computing the last block at the beginning tends to reduce the overall number of required iterations to converge, as we will verify in the numerical experiments reported in this article. We employ Nesterov's acceleration scheme for updating the dual vector to improve computation speed and a heuristic preconditioner that simply scales the state variables. In addition, an efficient coupling scheme between CD and AL methods is proposed to reduce the computation cost of each CD iteration. To analyze the role of each component of CDAL and its computational performance with respect to other solvers (FastMPC, μAO-MPC, OSQP, and MATLAB's `quadprog`), we conduct numerical experiments on an ill-conditioned problem of LTI-MPC control of an open-loop unstable AFTI-16 aircraft, and on LPV-MPC control of a continuously stirred tank reactor (CSTR).

B. Notation

$H \succ 0$ ($H \succeq 0$) denotes positive definiteness (semidefiniteness) of a square matrix H, H' (or z') denotes the transpose of matrix H (or vector z), H_i, j denotes the element of matrix H on the ith row and the jth column, and $H\_i, \cdot$ and $H\_\cdot, j$ denote the ith row vector and jth column vector of matrix H, respectively. For a vector z, $\|z\|\_2$ denotes the Euclidean norm of z and $z\_\neq i$ is the subvector obtained from z by eliminating its ith component z_i.

II. MPC

Consider the following MPC formulation for tracking problems:

$$\begin{aligned}
\min\ & \frac{1}{2}\sum_{t=0}^{T-1} \|W_y\,(y_{t+1}-r_{t+1})\|_2^2 + \frac{1}{2}\|W_u\,(u_t-u_t^r)\|_2^2 \\
& \qquad + \frac{1}{2}\|W_{\Delta u}\Delta u_t\|_2^2 \\
\text{s.t. } & x_{t+1} = Ax_t + Bu_t,\ t=0,\dots,T-1 \\
& y_{t+1} = Cx_{t+1},\ t=0,\dots,T-1 \\
& u_t = u_{t-1} + \Delta u_t,\ t=0,\dots,T-1 \\
& x_{\min} \le x_t \le x_{\max},\ t=1,\dots,T \\
& u_{\min} \le u_t \le u_{\max},\ t=0,\dots,T-1 \\
& \Delta u_{\min} \le \Delta u_t \le \Delta u_{\max}, t=0,\dots,T-1 \\
& x_0 = \bar{x}_0, u_{-1} = \bar{u}_{-1}
\end{aligned} \tag{1}$$

in which $x_t \in \mathbb{R}^{n_x}$ is the state vector, $u_t \in \mathbb{R}^{n_u}$ is the input vector, $\Delta u_t = u_t - u_{t-1}$ is the vector of input increments, $y_t \in \mathbb{R}^{n_y}$ is the output vector, r_t and u_t^r are the output and input set-points, and $\bar{x}_0$ and $\bar{u}_{-1}$ denote the current state and the previous input vectors, respectively. We assume that $W_y = W_y' \succeq 0$, $W_u = W_u' \succeq 0$, and $W_{\Delta u} = W_{\Delta u}' \succ 0$. The formulation (1) could be extended to include time-varying bounds on x and u along the prediction horizon, linear equality constraints or box constraints on the terminal state x_T for guaranteed closed-loop convergence, as well as affine prediction models. To simplify the notation, in the following, we consider the following reformulation of (1):

$$\begin{aligned}
\min\ & \frac{1}{2}\sum_{t=1}^{T} \hat{x}_t'(\hat{C}'\hat{W}\hat{C})\hat{x}_t - \hat{x}_t'(\hat{C}'\hat{W}\hat{r}_t) + \frac{1}{2}\hat{u}_{t-1}'W_{\Delta u}\hat{u}_{t-1} \\
\text{s.t. } & \hat{x}_{t+1} = \hat{A}\hat{x}_t + \hat{B}\hat{u}_t,\ t=0,\dots,T-1 \\
& \hat{x}_{\min} \le \hat{x}_t \le \hat{x}_{\max},\ t=1,\dots,T \\
& \hat{u}_{\min} \le \hat{u}_t \le \hat{u}_{\max},\ t=0,\dots,T-1 \\
& \hat{x}_0 = \begin{bmatrix} \bar{x}_0 \\ \bar{u}_{-1} \end{bmatrix}
\end{aligned} \tag{2}$$

where $\hat{x}_t = \begin{bmatrix} x_t \\ u_{t-1} \end{bmatrix} \in \mathcal{R}^{\hat{n}_x}$, $\hat{n}_x = n_x + n_u$, $\hat{u}_t = \Delta u_t \in \mathcal{R}^{n_u}$, $\hat{A} = \begin{bmatrix} A & B \\ 0 & I \end{bmatrix} \in \mathbb{R}^{\hat{n}_x \times \hat{n}_x}$, $\hat{B} = \begin{bmatrix} B \\ I \end{bmatrix} \in \mathbb{R}^{\hat{n}_x \times n_u}$, and $\hat{C} = \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix}$, $\hat{W} = \begin{bmatrix} W_y & 0 \\ 0 & W_u \end{bmatrix}$, $\hat{r}_t = \begin{bmatrix} r_t \\ u_{t-1}^r \end{bmatrix}$. The vector z of variables to optimize is

$$z = \begin{bmatrix} \hat{u}_0' & \hat{x}_1' & \hat{u}_1' & \dots & \hat{u}_{T-1}' & \hat{x}_T' \end{bmatrix}' \in \mathbb{R}^{T(\hat{n}_x+n_u)}.$$

The inequality constraints on state and input variables, whose number is $2T(\hat{n}_x + n_u)$, are

$$\underline{z} \le z \le \bar{z} \Leftrightarrow \begin{cases} \hat{x}_{\min} \le \hat{x}_t \le \hat{x}_{\max} & \forall t = 1,\dots,T \\ \hat{u}_{\min} \le \hat{u}_t \le \hat{u}_{\max} & \forall t = 0,\dots,T-1 \end{cases}$$

where $\hat{x}_{\min} = \begin{bmatrix} x_{\min} \\ u_{\min} \end{bmatrix}$, $\hat{x}_{\max} = \begin{bmatrix} x_{\max} \\ u_{\max} \end{bmatrix}$, $\hat{u}_{\min} = \Delta u_{\min}$, and $\hat{u}_{\max} = \Delta u_{\max}$. At each sample step, the MPC problem (1) can be recast as the following QP:

$$\begin{aligned}
& \min \frac{1}{2} z'Hz + h'z \\
& \text{s.t. } \underline{z} \le z \le \bar{z} \\
& \qquad Gz = g
\end{aligned} \tag{3}$$

where $H = H' \succeq 0$, $H \in \mathbb{R}^{n_z \times n_z}$, $n_z = T(\hat{n}_x + n_u)$, $h \in \mathbb{R}^{n_z}$, $G \in \mathbb{R}^{T\hat{n}_x \times n_z}$, and $g \in \mathbb{R}^{T\hat{n}_x}$ are defined as

$$H = \begin{bmatrix} R & 0 & \dots & 0 & 0 \\ 0 & Q & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & R & 0 \\ 0 & 0 & \dots & 0 & Q \end{bmatrix}, \quad \begin{aligned} R &= W_{\Delta u} \\ Q &= \hat{C}'\hat{W}\hat{C} \end{aligned}$$

$$G = \begin{bmatrix} \hat{B} & -I & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \hat{A} & \hat{B} & -I & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \hat{A} & \hat{B} & -I \end{bmatrix}$$

$$h = \begin{bmatrix} -\hat{C}'\hat{W}\hat{r}_1 \\ -\hat{C}'\hat{W}\hat{r}_2 \\ \vdots \\ -\hat{C}'\hat{W}\hat{r}_T \end{bmatrix}, \quad g = \begin{bmatrix} -\hat{A}\hat{x}_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Clearly, matrix G is full row-rank. Note that $A, B, C, W_y, W_u, W_{\Delta u}$, and the upper and lower bounds on x, u, and Δu in (1) may change at each controller execution.

III. Algorithm

A. Augmented Lagrangian Method

We solve the convex QP problem (3) by applying the augmented Lagrangian method. The bound-constrained Lagrangian function $\mathcal{L} : \mathcal{Z} \times \mathbb{R}^{T\times\hat{n}_x} \to \mathbb{R}$ is given by

$$\mathcal{L}(z,\Lambda) = \frac{1}{2}z'Hz + z'h + \Lambda'(Gz - g)$$

where $\mathcal{Z} = \{\underline{z} \le z \le \bar{z}\}$ and $\Lambda \in \mathbb{R}^{T\hat{n}_x}$ is the vector of Lagrange multipliers associated with the equality constraints in (3). The dual

problem of (3) is

$$\max_{\Lambda\in\mathbb{R}^{T\hat{n}_x}} \phi(\Lambda) \tag{4}$$

where $\phi(\Lambda)=\min_{z\in\mathcal{Z}}\mathcal{L}(z,\Lambda)$. Assuming that Slater's constraint qualification holds, the optimal value of the primal problem (3) and of its dual (4) coincide. However, $\phi(\Lambda)$ is not differentiable in general [23], so that any subgradient method for solving (4) would have a slow convergence rate. Under the AL framework, the augmented Lagrangian function

$$\mathcal{L}_\rho(z,\Lambda)=\frac{1}{2}z'Hz+z'h+\Lambda'(Gz-g)+\frac{\rho}{2}\|Gz-g\|^2 \tag{5}$$

is used instead, where the parameter $\rho>0$ is a penalty parameter. The corresponding augmented dual problem is defined as

$$\max_{\Lambda\in\mathbb{R}^{T\hat{n}_x}} \phi_\rho(\Lambda) \tag{6}$$

where $\phi_\rho(\Lambda)=\min_{z\in\mathcal{Z}}\mathcal{L}_\rho(z,\Lambda)$ is differentiable provided that $H+\rho G'G\succ 0$. The dual problem (4) and the augmented dual problem (6) share the same optimal solution [24, see Ch. 2 subsection 2.2], and most important $d_\rho(\Lambda)$ is concave and differentiable, with gradient [23], [25] $\nabla\phi_\rho(\Lambda)=Gz^*(\Lambda)-g$, where $z^*(\Lambda)$ denotes the optimal solution of the inner problem $\min_{z\in\mathcal{Z}}\mathcal{L}_\rho(z,\Lambda)$ for a given Λ. Moreover, the gradient mapping $\nabla\phi_\rho:\mathbb{R}^{T\times\hat{n}_x}\to\mathbb{R}^{T\times\hat{n}_x}$ is Lipschitz continuous, with a Lipschitz constant given by $L_\phi=\rho^{-1}$ [26].

Let $F_\rho(z;\Lambda^k)=\frac{1}{2}z'H_Az+(h_A^k)'z$, where $h_A^k=\frac{1}{\rho}h+G'\Lambda^k-G'g$, and $H_A=\frac{1}{\rho}H+G'G$ has the block-sparse structure

$$H_A=\begin{bmatrix}
\phi_1 & \phi_2 & 0 & 0 & 0 & \dots & 0 & 0 & 0\\
\phi_2' & \phi_3 & \phi_4 & \phi_5 & 0 & \dots & 0 & 0 & 0\\
0 & \phi_4' & \phi_1 & \phi_2 & 0 & \dots & 0 & 0 & 0\\
0 & \phi_5' & \phi_2' & \phi_3 & \phi_4 & \dots & 0 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\
0 & 0 & 0 & 0 & 0 & \dots & \phi_3 & \phi_4 & \phi_5\\
0 & 0 & 0 & 0 & 0 & \dots & \phi_4' & \phi_1 & \phi_2\\
0 & 0 & 0 & 0 & 0 & \dots & \phi_5' & \phi_2' & \phi_6
\end{bmatrix}$$

and $\phi_1=\frac{1}{\rho}R+\hat{B}'\hat{B}$, $\phi_2=-\hat{B}'$, $\phi_3=\frac{1}{\rho}Q+(I+\hat{A}'\hat{A})$, $\phi_4=\hat{A}'\hat{B}$, $\phi_5=-\hat{A}'$, and $\phi_6=\frac{1}{\rho}Q+I$. Since G is full rank, matrix $H_A\succ 0$. According to [24], the AL algorithm can be formulated in scaled form as follows:

$$z^{k+1}=\text{argmin}_{z\in\mathcal{Z}}F_\rho(z;\Lambda^k) \tag{7a}$$

$$\Lambda^{k+1}=\Lambda^k+(Gz^{k+1}-g) \tag{7b}$$

which involves the minimization step of the primal vector z and the update step of the dual vector Λ. As shown in [24], the convergence of AL can be assured for a large range of values of ρ. Typically, the larger the penalty parameter, the faster the AL algorithm is to converge, but the more difficult (7a) is to solve, due to a larger condition number of the Hessian matrix of subproblem (7a). The convergence rate of the AL algorithm (7) is $O(1/k)$ according to [27]. To improve the speed of the AL method, [28] proposed an accelerated AL algorithm, whose iteration-complexity is $O(1/k^2)$ for linearly constrained convex programs, by using Nesterov's acceleration technique. The accelerated AL algorithm is summarized in Algorithm 1.

For solving the strongly convex box-constrained QP (7a), the fast gradient projection method was used in [19] and [21]. Inspired by the fact that the Gauss–Seidel method in solving block tridiagonal linear systems is efficient [29], in this article, we propose the use of the cyclic CD method to make full use of block sparsity and avoid the explicit construction of matrix H_A. Note that in the gradient projection method or fast gradient projection method [21], the Lipschitz constant parameter deriving from matrix H_A needs to be calculated or estimated to ensure convergence. Therefore, for linear MPC problems that change at runtime such methods would be less preferable than cyclic CD. In this article, by making full use of the structure of the subproblem, we will implement a cyclic CD method that requires less computations, as we will detail in the next section.

Algorithm 1: Accelerated Augmented Lagrangian Method [28].

Input: Initial guess $z^0\in\mathcal{Z}$ and Λ^0; maximum number N_{out} of iterations; parameter $\rho>0$.
1. Set $\alpha_1\leftarrow 1$; $\hat{\Lambda}^0\leftarrow\Lambda^0$;
2. **for** $k=1,2,\dots,N_{\text{out}}$ **do**
2.1. $z^k\leftarrow\text{argmin}_{z\in\mathcal{Z}}F_\rho(z;\hat{\Lambda}^{k-1})$;
2.2. $\Lambda^k\leftarrow\hat{\Lambda}^{k-1}+(Gz^k-g)$;
2.3. **if** $\|\Lambda^k-\hat{\Lambda}^{k-1}\|_2^2\le\epsilon$, **stop**;
2.4. $\alpha_{k+1}\leftarrow\frac{1+\sqrt{1+4\alpha_k^2}}{2}$;
2.5. $\hat{\Lambda}^k\leftarrow\Lambda^k+\frac{\alpha_k-1}{\alpha_{k+1}}(\Lambda^k-\Lambda^{k-1})$;
3. **end**.

B. Coordinate Descent Method

The idea of the CD method is to minimize the objective function along only one coordinate direction at each iteration, while keeping the other coordinates fixed [30]. In [31], the authors showed that the CD method is convergent in convex differentiable minimization problems, and the rate of convergence is at least linear. We first give a brief introduction of the CD method to solve (7a). Under the assumption that the set of optimal solutions is nonempty and that the objective function F_ρ is convex, continuously differentiable, and strictly convex with respect to each coordinate, the CD method proceeds iteratively for $k=0,1,\dots$, as follows:

$$\text{choose } i_k\in\{1,2,\dots,n_z\} \tag{8a}$$

$$z_{i_k}^{k+1}=\text{argmin}_{z_{i_k}\in\mathcal{Z}}F_\rho(z_{i_k},z_{\neq i_k}^k;\hat{\Lambda}^k) \tag{8b}$$

where, with a slight abuse of notation, we denote by $F_\rho(z_{i_k},z_{\neq i_k}^k;\hat{\Lambda}^k)$ the value $F_\rho(z;\hat{\Lambda}^k)$ when $z_{\neq i_k}=z_{\neq i_k}^k$ is fixed. The convergence of the iterations in (8) for $k\to\infty$ depends on the rule used to choose the coordinate index i_k. In [31], the authors show that the *almost cyclic rule* and *Gauss–Southwell rule* guarantee convergence. Here, we use the almost cyclic rule that provides convergence according to the following lemma:

Lemma 1 ([31]): Let $\{z^k\}$ be the sequence of coordinate-descent iterates (8), where every coordinate index is iterated upon at least once on every N successive iterations, $N\ge n_z$. The sequence $\{z^k\}$ converges at least linearly to the optimal solution z^* of problem (7a).

In this article, we will use the *reverse cyclic rule*

$$i_k=n_z-(k\bmod n_z)$$

to exploit the fact that the shifted previous optimal solution is used as a warm start. The chosen rule clearly satisfies the assumptions of Lemma 1 for convergence. The implementation of one pass through all n_z coordinates using reverse cyclic CD is reported in Procedure 2. In Procedure 2, the Lagrangian variable $\hat{\Lambda}\in\mathbb{R}^{T\times\hat{n}_x}$ is divided into $\{\hat{\lambda}_0,\dots,\hat{\lambda}_{t-1},\dots,\hat{\lambda}_{T-1}\}$, where $\hat{\lambda}_{t-1}\in\mathbb{R}^{\hat{n}_x}$. For a given symmetric $M\in\mathbb{R}^{n_s\times n_s}\succeq 0$, $d\in\mathbb{R}^{n_s}$, the operator $\text{CCD}_{[\underline{s},\bar{s}]}\{M,d\}$ used in Procedure 2 represents one-pass iteration of the reverse cyclic CD method through all n_s coordinates $s_{n_s},\dots,s_1$ for the following box-constrained QP:

$$\min_{s\in[\underline{s},\bar{s}]}\frac{1}{2}s'Ms+s'd \tag{9}$$

Procedure 2 Full Pass of Reverse Cyclic Coordinate Descent on All Block Variables

Input: $\hat{\Lambda} = \{\hat{\lambda}_0, \ldots, \hat{\lambda}_{T-1}\}, U = \{\hat{u}_0, \ldots, \hat{u}_{T-1}\}$, $X = \{\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_T\}$; MPC settings $\hat{A}$, $\hat{B}$, Q, R, $\hat{u}_{\min}$, $\hat{u}_{\max}$, $\hat{x}_{\min}$, $\hat{x}_{\max}$; parameter $\rho > 0$.

1. $\sigma \leftarrow 0$;
2. $\{\hat{x}_T, \sigma\} \leftarrow \underset{\hat{x}_T \in [\hat{x}_{\min}, \hat{x}_{\max}]}{\text{CCD}} \{\frac{1}{\rho}Q + I, -\hat{\lambda}_{T-1} - \hat{A}\hat{x}_{T-1} - \hat{B}\hat{u}_{T-1} - \hat{C}'\hat{W}\hat{r}_T, \sigma\}$;
3. $\{\hat{u}_{T-1}, \sigma\} \leftarrow \underset{\hat{u}_{T-1} \in [\hat{u}_{\min}, \hat{u}_{\max}]}{\text{CCD}} \{\frac{1}{\rho}R + \hat{B}'\hat{B}, \hat{B}'(\hat{\lambda}_{T-1} + \hat{A}\hat{x}_{T-1} - \hat{x}_T), \sigma\}$;
4. **for** $t = T-2, T-3, \ldots, 0$ **do**
4.1. $\{\hat{x}_{t+1}, \sigma\} \leftarrow \underset{\hat{x}_{t+1} \in [\hat{x}_{\min}, \hat{x}_{\max}]}{\text{CCD}} \{\frac{1}{\rho}Q + I + \hat{A}'\hat{A}, -(\hat{\lambda}_t + \hat{A}\hat{x}_t + \hat{B}\hat{u}_t) + \hat{A}'(\hat{\lambda}_{t+1} + \hat{B}\hat{u}_{t+1} - \hat{x}_{t+2}) - \hat{C}'\hat{W}\hat{r}_t, \sigma\}$;
4.2. $\{\hat{u}_t, \sigma\} \leftarrow \underset{\hat{u}_t \in [\hat{u}_{\min}, \hat{u}_{\max}]}{\text{CCD}} \{\frac{1}{\rho}R + \hat{B}'\hat{B}, \hat{B}'(\hat{\lambda}_t + \hat{A}\hat{x}_t - \hat{x}_{t+1}), \sigma\}$;
5. **end**.

Output: $\hat{U}$, $\hat{X}$, σ.

Procedure 3 Evaluation of CCD in Step 4.2 of Procedure 2

Input: $\hat{\lambda}_t$, $\hat{u}_t$, $\hat{x}_t$, $\hat{x}_{t+1}$; MPC settings $\hat{A}$, $\hat{B}$, R, $\hat{u}_{\min}$, $\hat{u}_{\max}$; parameter $\rho > 0$; update amount $\sigma \geq 0$.

1. $V \leftarrow \hat{\lambda}_t + \hat{A}\hat{x}_t + \hat{B}\hat{u}_t - \hat{x}_{t+1}$;
2. **for** $i = n_u, \ldots, 1$ **do**
2.1. $s \leftarrow \frac{1}{\rho}R_{i,\cdot}\hat{u}_t + (\hat{B}_{\cdot,i})'V$;
2.2. $\theta \leftarrow [\hat{u}_{t,i} - \frac{s}{\frac{1}{\rho}R_{ii} + (\hat{B}'\hat{B})_{ii}}]_{\hat{u}_{min,i}}^{\hat{u}_{max,i}}$;
2.3. $\Delta \leftarrow \theta - \hat{u}_{t,i}$;
2.4. $\sigma \leftarrow \sigma + \Delta^2$;
2.5. $\hat{u}_{t,i} \leftarrow \theta$;
2.6. $V \leftarrow V + \Delta\hat{B}_{\cdot,i}$;
3. **end**.

Output: $\hat{u}_t, \sigma$.

that is to execute the following n_s iterations:

$$\begin{aligned} &\text{for } i = n_s, \ldots, 1 \\ &\quad s_i \leftarrow \left[s_i - \tfrac{1}{M_{i,i}}(M_{i,\cdot}s + d_i)\right]_{\underline{s}_i}^{\bar{s}_i} \\ &\text{end} \end{aligned} \tag{10}$$

where $[s_i]_{\underline{s}_i}^{\bar{s}_i}$ is the projection operator

$$[s_i]_{\underline{s}_i}^{\bar{s}_i} = \begin{cases} \bar{s}_i & \text{if} \quad s_i \geq \bar{s}_i \\ s_i & \text{if} \quad \underline{s}_i < s_i < \bar{s}_i \\ \underline{s}_i & \text{if} \quad s_i \leq \underline{s}_i. \end{cases} \tag{11}$$

Note that in Procedure 2, Steps 2, 3, 4.1, and 4.2 all involve the same operator CCD. In Procedure 3, we exemplify an efficient way to evaluate such an operator for Step 4.2 of Procedure 2, as the approach is similar for evaluating Steps 2, 3, and 4.1, where σ records the sum of squared coordinate variations.

C. Preconditioning

Preconditioning is a common heuristic for improving the computational performance of first-order methods. The optimal design of preconditioners has been studied for several decades, but such a computation is often more complex than the original problem and may become prohibitive if it must be executed at runtime. Diagonal scaling is a heuristic preconditioning that is very simple and often beneficial [32], [33]. In this article, we propose to make the change of state variables

Procedure 4 Modified Procedure 3 to Efficiently Couple CD and AL

Input: λ_t, $\hat{u}_t$; MPC settings $\hat{A}$, $\hat{B}$, R, $\hat{u}_{\min}$, $\hat{u}_{\max}$; parameter $\rho > 0$; update amount $\sigma \geq 0$.

1. **for** $i = n_u, \ldots, 1$ **do**
1.1. $s \leftarrow \frac{1}{\rho}R_{i,\cdot}\hat{u}_t + (\hat{B}_{\cdot,i})'\lambda_t$;
1.2. $\theta \leftarrow [\hat{u}_{t,i} - \frac{s}{\frac{1}{\rho}R_{ii} + (\hat{B}'\hat{B})_{ii}}]_{\hat{u}_{min,i}}^{\hat{u}_{max,i}}$;
1.3. $\Delta \leftarrow \theta - \hat{u}_{t,i}$;
1.4. $\sigma \leftarrow \sigma + \Delta^2$;
1.5. $\hat{u}_{t,i} \leftarrow \theta$;
1.6. $\lambda_t \leftarrow \lambda_t + \Delta \cdot \hat{B}_{\cdot,i}$;
2. **end**.

Output: $\hat{u}_t, \lambda_t, \sigma$.

$\bar{x} = E\hat{x}$, where $E \in \mathcal{R}^{\hat{n}_x \times \hat{n}_x}$ is a diagonal matrix whose ith entry is

$$E_{i,i} = \sqrt{Q_{i,i} + \hat{A}'_{\cdot,i}\hat{A}_{\cdot,i}} \tag{12}$$

and replace the prediction model $\hat{x}_{t+1} = \hat{A}\hat{x}_t + \hat{B}\hat{u}_t$ by

$$\bar{x}_{t+1} = \bar{A}\bar{x}_t + \bar{B}\hat{u}_t$$

where $\bar{A} = E\hat{A}E^{-1}$ and $\bar{B} = E\hat{B}$. The weight matrix Q and constraints $[\hat{x}_{\min}, \hat{x}_{\max}]$ are scaled accordingly by setting $\bar{Q} = E^{-1}QE^{-1}$ and $\bar{x}_{\min} = E^{-1}\hat{x}_{\min}$, $\bar{x}_{\max} = E^{-1}\hat{x}_{\max}$.

D. Efficient Coupling Scheme Between CD and AL Method

We are now ready to couple CD and AL to solve the posed MPC problem (1) efficiently. We first note that updating $\hat{u}_t$ and $\hat{x}_{t+1}$ for all t involves computing a similar temporary vector V in Procedure 3. As V is in fact the next update of the dual vector Λ in Algorithm 1, we modify Procedure 3 as shown in Procedure 4. The overall solution method described in the previous sections is summarized in Algorithm 5, that we call CDAL. Note that the main update of the Lagrangian variables in Algorithm 5 is placed early in Step 5, unlike in Algorithm 1, due to the use of the proposed efficient coupling scheme. The AL (outer) iterations are executed for maximum N_{out} iterations, the CD (inner) iterations for at most N_{in} iterations. The tolerances ϵ_{out} and ϵ_{in} are used to stop the outer and inner iterations, respectively. Algorithm 5 is matrix-free and library-free, and we could implement it in 90 lines of C code.

IV. Numerical Examples

We test the performance of the CDAL solver against other solvers in two numerical experiments. The first one is the ill-conditioned AFTI-16 control problem [34], [35] based on LTI-MPC, used in the Model Predictive Control Toolbox for MATLAB [36]. The main goals of this experiment include investigating whether our proposed simple heuristic preconditioner, reverse cyclic rule, and Nesterov's acceleration scheme are helpful, and provide a detailed comparison with other solvers. The second experiment demonstrates the benefits of the construction-free property in LPV-MPC of a CSTR [37], in which the prediction model is obtained by linearizing a nonlinear model of the process at each sample step. The reported simulation results were obtained on a MacBook Pro with 2.7 GHz 4-core Intel Core i7 and 16 GB RAM. Algorithm 5 is executed in MATLAB via a C-mex interface.

Algorithm 5: Accelerated Reverse Cyclic CDAL Algorithm for Linear (or Linearized) MPC.

Input: primal/dual warm-start $U = \{\hat{u}_0, \hat{u}_1, \ldots, \hat{u}_{T-1}\}$, $X = \{\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_T\}$, $\Lambda^{-1} = \Lambda^0 = \{\lambda_0, \lambda_1, \cdots, \lambda_{T-1}\}$; MPC settings $\{\hat{A}, \hat{B}, \hat{C}, W_y, W_u, W_{\Delta u}, \Delta u_{\min}, \Delta u_{\max}, u_{\min}, u_{\max}, x_{\min}, x_{\max}\}$; Algorithm settings $\{\rho, N_{\text{out}}, N_{\text{in}}\, \epsilon_{\text{out}}, \epsilon_{\text{in}}\}$

1. Obtain preconditioned $\bar{X} = \{\bar{x}_0, \ldots, \bar{x}_T\}$, $\bar{A}$, $\bar{B}$, $\bar{Q}$, $\bar{x}_{\min}, \bar{x}_{\max}$ according to Section III-C
2. $\alpha_1 \leftarrow 1$; $\hat{\Lambda}^0 \leftarrow \Lambda^0$;
3. **for** $k = 1, 2, \ldots, N_{\text{out}}$ **do**
3.1. **for** $t = 0, \ldots, T-1$ **do**
3.1.1. $\lambda_t^k = \hat{\lambda}_t^{k-1} + \bar{A}\bar{x}_t + \bar{B}\hat{u}_t - \bar{x}_{t+1}$;
3.2. **for** $k_{in} = 1, 2, \ldots, N_{\text{in}}$ **do**
3.2.1. $U, \bar{X}, \sigma \leftarrow$ Procedure 2 with use of Procedure 4;
3.2.2. **if** $\sigma \le \epsilon_{\text{in}}$ **break** the loop;
3.3. **if** $\|\Lambda^k - \hat{\Lambda}^{k-1}\|_2^2 \le \epsilon_{\text{out}}$ **stop**;
3.4. $\alpha_{k+1} \leftarrow \frac{1+\sqrt{1+4\alpha_k^2}}{2}$;
3.5. $\hat{\Lambda}^k \leftarrow \Lambda^k + \frac{\alpha_k - 1}{\alpha_{k+1}}(\Lambda^k - \Lambda^{k-1})$;
4. Recover X from $\bar{X}$
5. **end**.

Output: U, X, Λ

A. AFTI-16 Benchmark Example

The open-loop unstable linearized AFTI-16 aircraft model reported in [34] and [35] is

$$\begin{cases} \dot{x} = \begin{bmatrix} -0.0151 & -60.5651 & 0 & -32.174 \\ -0.0001 & -1.3411 & 0.9929 & 0 \\ 0.00018 & 43.2541 & -0.86939 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x \\ \qquad + \begin{bmatrix} -2.516 & -13.136 \\ -0.1689 & -0.2514 \\ -17.251 & -1.5766 \\ 0 & 0 \end{bmatrix} u \\ y = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x. \end{cases}$$

The model is sampled using zero-order hold every 0.05 s. The input constraints are $|u_i| \le 25^\circ, i = 1, 2$, the output constraints are $-0.5 \le y_1 \le 0.5$ and $-100 \le y_2 \le 100$. The control goal is to make the pitch angle y_2 track a reference signal r_2. In designing the MPC controller, we take $W_y = \text{diag}([10, 10])$, $W_u = 0$, $W_{\Delta u} = \text{diag}([0.1, 0.1])$, and the prediction horizon is $T = 5$.

To investigate the effects of the three techniques (reverse cyclic rule, acceleration, and preconditioning) that we have introduced to improve the efficiency of the CDAL algorithm, we performed closed-loop simulations on eight schemes with fixed $\rho = 1$. These are 0-CDAL, the basic scheme, without acceleration and reverse cyclic rule; R-CDAL, the scheme with the Reverse cyclic rule; A-CDAL, the accelerated scheme; AR-CDAL, the accelerated scheme with the reverse cyclic rule, and their respective schemes with preconditioners, namely P-0-CDAL, P-R-CDAL, P-A-CDAL; and finally CDAL, that includes all the proposed techniques. The stopping criteria are defined by $\epsilon_{\text{in}} = 10^{-6}$, $\epsilon_{\text{out}} = 10^{-4}$, and N_{out}, N_{in} are set to the large enough value 5000 in order to guarantee good-quality solutions.

The computational load associated with the above schemes is listed in Table I, in which the last column represents the closed-loop performance, which is the average value $\frac{1}{T}\sum_{t=0}^{T-1} \|W_y(y_{t+1} - r_{t+1})\|_2^2 + \|W_u(u_{t+1} - u_{t+1}^r)\|_2^2 + \|W_{\Delta u}\Delta u_t\|_2^2$ of the MPC cost over the duration T of the closed-loop simulation and is almost the same for all schemes. The associated closed-loop trajectories are reported in Fig. 1, which shows that the pitch angle correctly tracks the reference signal from 0° to 10° and then back to 0°, and that both the input and output constraints are satisfied.

TABLE I
COMPUTATIONAL PERFORMANCE OF DIFFERENT SCHEMES

method	sum of inner iters		outer iters		time (ms)		cost
	avg	max	avg	max	avg	max	
0-CDAL	8577	79615	339	2104	4.9	55.3	42.3
R-CDAL	7298	72693	340	2103	4.3	53.2	42.5
A-CDAL	7437	57026	45	297	4.0	41.1	42.5
AR-CDAL	6207	51884	44	205	3.8	39.5	42.5
P-0-CDAL	3467	13386	33	171	2.1	11.4	42.5
P-R-CDAL	1757	13430	33	171	1.0	10.9	42.5
P-A-CDAL	3299	12161	13	60	1.7	9.7	42.5
CDAL	1543	12508	13	60	0.85	9.5	42.5

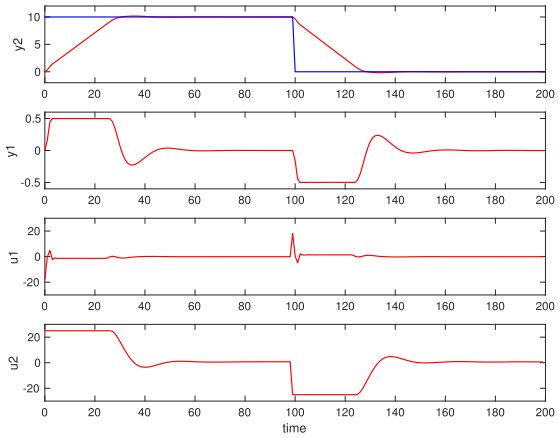

Fig. 1. Linear AFTI-16 closed-loop performance.

Since each MPC execution requires different numbers of inner and outer iterations, the average ("avg") and maximum ("max") number of iterations (or CPU time) are computed over the entire closed-loop execution. It can be observed that the maximum and average number of inner-loop iterations of R-CDAL are smaller than that of 0-CDAL (especially the maximum number), while their outer-loop iterations are almost the same, which shows that the reverse cyclic rule provides a significant improvement. Although A-CDAL has fewer outer-loop iterations, it has more inner-loop iterations than 0-CDAL on average. It therefore does not result in a significant reduction in total computation time. We can see that AR-CDAL achieves fewer iterations both in the inner and outer loops and has better average and worst-case computation performance. It can also be seen from Table I that preconditioning significantly reduces the number of outer-loop iterations.

Next, we investigate the effect on computation efficiency of parameter ρ that we expect to tend to trade off feasibility versus optimality. In particular, we expect larger values of ρ to favor feasibility, i.e., provide more inner-loop iterations and less outer-loop iterations, and vice versa. The computational performance results obtained by performing closed-loop simulations using the final *CDAL* algorithm for different values of ρ between 0.01 and 1 are listed in Table II. When the parameter value is between 0.01 and 0.1, the CDAL algorithm has very similar computational burden.

To further illustrate the efficiency of CDAL, Table II also lists the results obtained by using other solvers. Here, the FastMPC solver is also a construction-free solver that provides a free C-mex code. We also made comparison with the μAO-MPC solver v1.0.0-beta [38], which is based on an augmented Lagrangian method together with

TABLE II
COMPUTATIONAL LOAD OF CDAL WITH DIFFERENT VALUES OF ρ AND COMPARISON WITH OTHER SOLVERS

Solver	solver setting	time (ms) avg	time (ms) max	cost
CDAL	$\rho = 1$	0.85	9.5	42.561
	$\rho = 0.5$	0.72	7.1	42.590
	$\rho = 0.2$	0.53	4.2	42.612
	$\rho = 0.1$	0.47	3.8	42.619
	$\rho = 0.05$	0.42	3.3	42.618
	$\rho = 0.01$	0.41	3.2	42.618
FastMPC	$maxit = 5, k = 0.1$	0.54	4.2	42.627
μAO-MPC	$\mu = 0.05$ in_iter=100,ex_iter=100	7.0* 8**	68.1* 69**	42.627
OSQP	$N = 5000, \epsilon = 10^{-6}$	0.6* 1.5**	10.1* 13.8**	42.627
quadprog	default	10.3* 11**	20.6* 22**	42.622

*: pure solution time, without including matrix factorization
**: total time (MPC construction + solution)

Nesterov's gradient method. The μAO-MPC differs from CDAL in the way the subproblems are solved, and the outer loop not involving an acceleration scheme. The state-of-the-art first-order method for QP, OSQP solver v0.6.2 [10], and MATLAB's built-in QP solver (quadprog) are also used for comparison. For a fair comparison, each solver setting is chosen to at least ensure that each shares the same objective cost and constraint violation. When the parameter ρ of the CDAL is 0.01, the CDAL is faster than the other solvers. Regarding the μAO-MPC, OSQP, and quadprog solvers, we split between QP problem construction time (including the required matrix factorizations) and pure solution time. Note that in this case, the controller is LTI-MPC, and hence, the MPC problem construction and matrix factorizations required by these non-construction-free solvers can be performed offline. On the other hand, in the case of LPV-MPC problems, the total computation time would be spent online and the embedded code would also include routines for problem construction and matrix factorization functions. Instead, CDAL does not require any construction nor factorizations, thus making the solver very lean and fast also in a time-varying MPC setting, as investigated next.

B. Nonlinear CSTR Example

To illustrate the performance of CDAL when the linear MPC formulation (1) changes at runtime, we consider the control of the CSTR system [37], described by the continuous-time nonlinear model

$$\begin{aligned}\frac{dC_A}{dt} &= C_{A,i} - C_A - k_0 e^{\frac{-EaR}{T}} C_A \\ \frac{dT}{dt} &= T_i + 0.3T_c - 1.3\,T + 11.92 k_0 e^{\frac{-EaR}{T}} C_A \\ y &= C_A \end{aligned} \tag{13}$$

where C_A is the concentration of reagent A, T is the temperature of the reactor, $C_{A,i}$ is the inlet feed stream concentration, which is assumed to have the constant value 10.0 kgmol/m^3. The disturbance comes from the inlet feed stream temperature T_i, which has fluctuations represented by $T_i = 298.15 + 5\sin(0.05t)$ K. The manipulated variable is the coolant temperature T_c. The constants $k_0 = 34930800$ and $EaR = -5963.6$ (in MKS units). The reactor's initial state is at a low conversion rate, with $C_A = 8.57$ kgmol/m^3 and $T = 311$ K. The goal is to adjust the reactor state to a high reaction rate with $C_A = 2$ kgmol/m^3, which is a quite large condition. The controller manipulates the coolant temperature

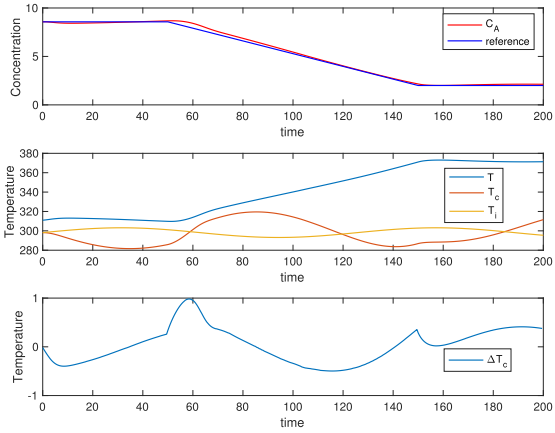

Fig. 2. Nonlinear CSTR closed-loop performance.

T_c to track a concentration reference as well as reject the measured disturbance T_i. Due to its nonlinearity, the model in (13) is linearized online at each sampling step:

$$\frac{dx}{dt} \approx f(x_t, u_{t-1}, p) + \frac{\partial f}{\partial x}\bigg|_{x_t, u_{t-1}, p} (x - x_t) + \frac{\partial f}{\partial u}\bigg|_{x_t, u_{t-1}, p} (u - u_{t-1})$$

where $f(x, u, p)$ is the mapping defined in (13) for $x = [C_A\ T]'$, $u = T_c$, $p = [C_{A,i}\ T_i]'$. By setting $A_c = \frac{\partial f}{\partial x}|_{x_t, u_{t-1}, p}$, $B_c = \frac{\partial f}{\partial u}|_{x_t, u_{t-1}, p}$, and $e_c = f(x_t, u_{t-1}, p) - A_t x_t - B_t u_{t-1}$, we get the following linearized continuous-time model:

$$\frac{d}{dt}x = A_c x + B_c u + e_c.$$

We use the forward Euler method with sampling time $T_s = 0.5$ min to obtain the following discrete-time model:

$$x_{t+1} = A_d x_t + B_d u_t + e_d$$

where $A_d = I + T_s A_c$, $B_d = T_s B_c$, and $e_d = T_s e_c$. Although held constant over the prediction horizon, clearly matrices A_d and B_d and the offset term e_d change at runtime, which makes the controller an LPV-MPC. Regarding the performance index, we choose weights $W_y = 1$, $W_u = 0$, and $W_{\Delta u} = 0.1$. The physical limitation of the coolant jacket is that its rate of change ΔT_c is subject to the constraint $[-1, 1]$ K when considering the sampling time $T_s = 0.5$ min. The prediction horizon is $T = 10$ steps.

We compare again CDAL with FastMPC, μAO-MPC, OSQP, and quadprog solvers in the LPV-MPC setting described above. CDAL is run with $\epsilon_{\text{in}} = 10^{-6}$, $\epsilon_{\text{out}} = 10^{-4}$, $\rho = 0.01$, and $N_{\text{out}} = N_{\text{in}} = 5000$. For a fair comparison, each solver setting is chosen to at least ensure each shares the same objective cost and constraint violation. The closed-loop simulation results of CDAL and other solvers almost coincide and are plotted in Fig. 2, from which it can be seen that C_A tracks the reference signal well, and the fluctuation of T_i is effectively suppressed. The computational load and closed-loop performance associated with CDAL and other solvers are reported in Table III. In this successive linearization-based MPC example, we found that the problem-construction time has a comparable computation time to the problem-solving time from the results of non-construction-free solvers. If we only compare the solution time, CDAL is faster than other solvers except for OSQP, but in fact the MPC construction time must be included for comparison, which leads to CDAL being faster than OSQP. Because of the construction-free, matrix-free, and library-free features, CDAL has an advantage in industrial embedded deployment when the optimization problem associated with MPC is constructed online and this operation has a cost that is comparable to the solution time.

TABLE III
COMPUTATIONAL PERFORMANCE OF CDAL AND OTHER SOLVERS

Solver	solver setting	time (ms) avg	time (ms) max	cost
CDAL	$\rho, \epsilon_{in}, \epsilon_{out} = 0.01, 10^{-6}, 10^{-4}$	0.3	0.6	0.02202
FastMPC	maxit=5,$k = 0.1$	0.5	7.2	0.030170
μAO-MPC	$\mu = 0.01$	1.4*	10.1*	0.02202
	in_iter=100,ex_iter=10	2.1**	15.2**	
OSQP	default	0.15*	0.37*	0.02219
		0.6**	5.5**	
quadprog	default	1.6*	9.7*	0.02219
		1.8**	13.3**	

* : solution time
** : MPC construction time + solution time

V. CONCLUSION

This article has proposed a construction-free, matrix-free, and library-free MPC solver, based on a cyclic coordinate-descent method in the augmented Lagrangian framework. We showed that the method is efficient and competes with other existing methods, thanks to the use of a reverse cyclic rule, Nesterov's acceleration, a simple heuristic preconditioner, and an efficient coupling scheme. Compared to many QP solution methods proposed in the literature, CDAL avoids constructing the QP problem, which makes it particularly appealing for some scenarios in which its online construction is required and has a comparable computation time to solving itself.

The proposed algorithm can be immediately extended to handle linear time-varying systems, in which the plant-model and/or cost-function matrices are allowed to vary over the prediction horizon. Future research will investigate the use of CDAL to solve nonlinear MPC problems and data-driven MPC formulations in which the model is adapted online by recursive system identification.

REFERENCES

[1] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Eng. Pract.*, vol. 11, no. 7, pp. 733–764, 2003.

[2] A. Bemporad, M. Morari, and E. N. P. V. Dua, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, no. 1, pp. 3–20, 2002.

[3] D. Kouzoupis, G. Frison, A. Zanelli, and M. Diehl, "Recent advances in quadratic programming algorithms for nonlinear model predictive control," *Vietnam J. Math.*, vol. 46, no. 4, pp. 863–882, 2018.

[4] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Trans. control Syst. Technol.*, vol. 18, no. 2, pp. 267–278, Mar. 2010.

[5] S. J. Wright, "Efficient convex optimization for linear MPC," in *Handbook of Model Predictive Control*. Berlin, Germany: Springer, 2019, pp. 287–303.

[6] H. J. Ferreau, H. G. Bock, and M. Diehl, "An online active set strategy to overcome the limitations of explicit MPC," *Int. J. Robust Nonlinear Control: IFAC-Affiliated J.*, vol. 18, no. 8, pp. 816–830, 2008.

[7] A. Bemporad, "A Quadratic programming algorithm based on nonnegative least squares with applications to embedded model predictive control," *IEEE Trans. Autom. Control*, vol. 61, no. 4, pp. 1111–1116, Apr. 2016.

[8] P. Patrinos and A. Bemporad, "An accelerated dual gradient-projection algorithm for embedded linear model predictive control," *IEEE Trans. Autom. Control*, vol. 59, no. 1, pp. 18–33, Jan. 2014.

[9] S. Boyd, N. Parikh, E. Chu, J. Eckstein, and B. Peleato, "Distributed optimization and statistical learning via the alternating direction method of multipliers" *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[10] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, "OSQP: An operator splitting solver for quadratic programs," *Math. Program. Comput.*, vol. 12, pp. 637–672, 2020.

[11] W. Li and J. Swetits, "A new algorithm for solving strictly convex quadratic programs," *SIAM J. Optim.*, vol. 7, no. 3, pp. 595–619, 1997.

[12] B. Hermans, A. Themelis, and P. Patrinos, "QPALM: A Newton-type proximal augmented Lagrangian method for quadratic programs," in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 4325–4330.

[13] A. Bemporad, "A Numerically stable solver for positive semi-definite quadratic programs based on nonnegative least squares," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 525–531, Feb. 2018.

[14] N. Saraf and A. Bemporad, "A bounded-variable least-squares solver based on stable QR updates," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 1242–1247, Mar. 2020.

[15] N. Saraf and A. Bemporad, "An efficient bounded-variable nonlinear least-squares algorithm for embedded MPC," *Automatica*, vol. 141, 2022, Art. no. 110293.

[16] C. J. Hsieh, K. W. Chang, C. J. Lin, S. S. Keerthi, and S. Sundararajan, "A dual coordinate descent method for large-scale linear SVM," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 408–415.

[17] K. W. Chang, C. J. Hsieh, and C. J. Lin, "Coordinate descent method for large-scale L2-loss linear support vector machines," *J. Mach. Learn. Res.*, vol. 9, no. 45, pp. 1369–1398, 2008.

[18] P. Richtárik and M. Takáč, "Distributed coordinate descent method for learning with Big Data," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2657–2681, 2016.

[19] S. Richter, C. N. Jones, and M. Morari, "Computational complexity certification for real-time MPC with input constraints based on the fast gradient method," *IEEE Trans. Autom. Control*, vol. 57, no. 6, pp. 1391–1403, Jun. 2012.

[20] V. Nedelcu, I. Necoara, and Q. Tran-Dinh, "Computational complexity of inexact gradient augmented Lagrangian methods: Application to constrained MPC," *SIAM J. Control Optim.*, vol. 52, no. 5, pp. 3109–3134, 2014.

[21] R. F. M. Kögel, "Fast predictive control of linear systems combining Nesterov's gradient method and the method of multipliers," in *Proc. IEEE 50th Conf. Decis. Control Eur. Control Conf.*, 2011, pp. 501–506.

[22] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Sov. Math. Dokl.*, vol. 27, no. 2, pp. 372–376, 1983.

[23] D. P. Bertsekas, "Nonlinear programming," *J. Oper. Res. Soc.*, vol. 48, no. 3, pp. 334–334, 1997.

[24] D. P. Bertsekas, *Constrained Optim. and Lagrange Multiplier Methods*. New York, NY, USA: Academic, 2014.

[25] Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program.*, vol. 103, no. 1, pp. 127–152, 2005.

[26] G. Lan and R. D. C. Monteiro, "Iteration-complexity of first-order augmented Lagrangian methods for convex programming," *Math. Program.*, vol. 155, no. 1, pp. 511–547, 2016.

[27] B. He and X. Yuan, "On the acceleration of augmented Lagrangian method for linearly constrained optimization," *Optim. Online*, vol. 3, 2010. [Online]. Available: https://optimization-online.org/2010/10/2760/

[28] M. Kang, M. Kang, and M. Jung, "Inexact accelerated augmented Lagrangian methods," *Comput. Optim. Appl.*, vol. 62, no. 2, pp. 373–404, 2015.

[29] P. Amodio and F. Mazzia, "A parallel Gauss–Seidel method for block tridiagonal linear systems," *SIAM J. Sci. Comput.*, vol. 16, no. 6, pp. 1451–1461, 1995.

[30] S. J. Wright, "Coordinate descent algorithms," *Math. Program.*, vol. 151, no. 1, pp. 3–34, 2015.

[31] Z. Q. Luo and P. Tseng, "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Optim. Theory Appl.*, vol. 72, no. 1, pp. 7–35, 1992.

[32] P. Giselsson and S. Boyd, "Diagonal scaling in Douglas-Rachford splitting and ADMM," in *Proc. IEEE 53rd Conf. Decis. Control*, 2014, pp. 5033–5039.

[33] R. Takapoui and H. Javadi, "Preconditioning via diagonal scaling," Oct. 2016, *arXiv:1610.03871*.

[34] P. Kapasouris, M. Athans, and G. Stein, "Design of feedback control systems for stable plants with saturating actuators," in *Proc. IEEE 27th Conf. Decis. Control*, 1988, vol. 1, pp. 469–479.

[35] A. Bemporad, A. Casavola, and E. Mosca, "Nonlinear control of constrained linear systems via predictive reference management," *IEEE Trans. Autom. Control*, vol. 42, no. 3, pp. 340–349, Mar. 1997.

[36] A. Bemporad, M. Morari, and N. L. Ricker, *Model Predictive Control Toolbox: User's Guide, Version 2*. Natick, MA, USA: MathWorks, 2004.

[37] D.E Seborg, T. F. Edgar, D. A. Mellichamp, and F. J. Doyle III, *Process Dynamics and Control*. Hoboken, NJ, USA: Wiley, 2016.

[38] P. Zometa, M. Kögel, and R. Findeisen, "μAO-MPC: A free code generation tool for embedded real-time linear model predictive control," in *Proc. Amer. Control Conf.*, 2013, pp. 5320–5325.