2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

A System-Theoretic Approach to Construct a Banded Null Basis to Efficiently Solve MPC-Based QP Problems *

Jiaheng Yang [1], T.J. Meijer [1,†], V.S. Dolk [2], Bram de Jager [1] and W.P.M.H. Heemels [1]

Abstract— **The null-space method is able to reduce the number of decision variables in the on-line optimization carried out in model predictive control. This method relies on the construction of a basis for the null space of the equality constraints. This paper proposes a systematic approach based on system-theoretic insights to construct such a basis with a banded structure. This banded structure carries over to the resulting lower-dimensional QP and can be exploited to compute a solution more efficiently. Specifically, solvers that exploit this structure result in a computational complexity that scales linearly with the prediction horizon. In contrast to similar approaches in the literature, the proposed method can be applied to uncontrollable, though stabilizable, systems with multiple inputs. This method is particularly interesting when dealing with systems with large state dimension and long prediction horizons. Finally, the method is applied to a numerical example in combination with both the alternating direction method of multipliers and the accelerated dual gradient projection method to demonstrate its benefits.**

I. Introduction

Model predictive control (MPC) encompasses a class of control algorithms which optimize the future response of a system based on a prediction model while accounting for state and input constraints [1]. In MPC, at every sampling instant, the next control action is computed by solving a constrained finite-horizon open-loop optimal control problem based on the current state. Originally, the high computational effort required for the on-line optimization restricted the application of MPC to systems with slow dynamics. However, recent advances in efficient optimization solvers have contributed significantly to a more widespread adoption of MPC. Among the available methods are interior-point methods [2], [3], active-set methods [4], accelerated dual gradient-projection method (GPAD) [5] and alternating direction method of multipliers (ADMM) [6].

In linear MPC with polyhedral constraints, the on-line optimization consists of solving a constrained finite-horizon linear-quadratic optimal control problem, which can be formulated as a quadratic programming (QP) problem. Traditionally, two types of MPC-QP formulations are encountered. Firstly, the so-called *sparse* formulation considers both the future states and inputs as decision variables while imposing equality constraints to incorporate the system dynamics implicitly. The sparse formulation features a large number of decision variables, but its Hessian and inequality constraints are sparse. Secondly, the *dense* formulation is obtained by considering the future states as an explicit function of the current state and future inputs and eliminating them from the decision variables [7]. This reduction in number of decision variables typically comes at the expense of a loss of sparsity.

The adopted QP formulation significantly affects the required computation power to compute a solution. As the sparse formulation incorporates states in the decision variables, the required computational complexity scales poorly for systems with large state dimension. However, the sparse structure can be exploited such that complexity scales linearly with the prediction horizon [3]. The dense formulation, on the other hand, results in more attractive scaling of the computational demands with state dimension while scaling poorly with the prediction horizon, e.g. cubically for interior-point methods [8]. Recently, two methods have been developed which construct a *sparse-condensed* formulation by eliminating the future states from the decision variables while preserving sparsity [8], [9]. Both methods achieve a computational complexity that scales linearly with the prediction horizon. However, the method in [8] requires the computation of a deadbeat feedback gain, which is not a trivial task and is only possible if the uncontrollable part of the system is nilpotent, i.e., for so-called null-controllable systems [10]. The second method, presented in [9], is an application of the so-called *null-space method* [11], [12] to the MPC problem. This method relies on the construction of a banded null basis for the equality constraints. The resulting computational complexity depends on the bandwidth of this null basis. General methods to compute such a banded null basis provide no theoretical guarantees regarding the bandwidth [13]. In [14], a method is proposed for *single-input* systems, which guarantees a sparse null basis when the prediction horizon is sufficiently large.

This paper proposes a novel approach to construct a banded null basis, which can be applied to uncontrollable, though stabilizable, multi-input systems. The proposed approach is based on the observation that any basis of the null space of the equality constraints can be interpreted as a set of linearly independent nonzero responses of the system that start and end in the origin. Using this system-theoretic insight, it is possible to provide guarantees regarding the sparsity of the sparse-condensed formulation based on the length of these responses. It is demonstrated how the banded structure of the condensed QP can be exploited by dedicated solvers based on ADMM and GPAD. Finally, the required computation power is connected to the bandwidth of the used

* This work is part of the research programme High Tech Systems and Materials (HTSM) with project number 13896, which is (partly) financed by the Dutch Research Council (NWO).

[1] Jiaheng Yang, T.J. Meijer, Bram de Jager and W.P.M.H. Heemels are with the Department of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. Email addresses: youngjh59@outlook.com; {t.j.meijer; a.g.de.jager; w.p.m.h.heemels}@tue.nl.

[2] Victor Dolk is with ASML, De Run 6665, 5504 DT Veldhoven, The Netherlands. Email address: victor.dolk@asml.com.

† Corresponding author: T.J. Meijer.

null basis and, hence, to the length of the responses used in its construction.

The content of this paper is organized as follows. Necessary background information on linear constrained MPC, the non-condensed MPC-QP formulation and the null-space method is presented in Section II. Section III defines the problem addressed in this paper. Section IV discusses the systematic construction of a null-space matrix using a system-theoretic approach. In Section V, the banded structure of the sparse-condensed QP formulation is discussed and exploited to efficiently compute a solution. Finally, a numerical example is covered in Section VI and conclusions are provided in Section VII.

A. Notation

The sets of n-by-n symmetric positive semidefinite and symmetric positive definite matrices are denoted by $\mathbb{S}^n_{\succeq 0}$ and $\mathbb{S}^n_{\succ 0}$, respectively. Moreover, for any two non-negative integers $0 \le i \le j$ the subset $\mathbb{N}_{[i,j]} := \{i, i+1, \ldots, j\} \subset \mathbb{N}$ is defined along with the shorthand $\mathbb{N}_{\le j}$ for the case $i = 0$. The notation $\mathbb{N}_{>0}$ refers to the set of positive integers. Finally, the Kronecker product of two matrices $A \in \mathbb{R}^{p\times q}$ and $B \in \mathbb{R}^{s\times t}$ is denoted by $A \otimes B \in \mathbb{R}^{ps\times qt}$.

II. Background

A. Linear constrained MPC

Consider the discrete-time LTI system

$$x_{k+1} = Ax_k + Bu_k, \tag{1}$$

where $x_k \in \mathbb{R}^n$ and $u_k \in \mathbb{R}^m$ denote the state and input at time $k \in \mathbb{N}$, respectively, and B is assumed to be full column rank. In linear constrained MPC, given the (estimated) current state $x_{0|k} = x_k$, the following constrained finite-horizon linear-quadratic optimal control problem is solved at every sampling instant k [7]:

$$\begin{aligned}
\min_{\boldsymbol{x}_k,\, \boldsymbol{u}_k} \quad & \frac{1}{2}x_{N|k}^T Q_N x_{N|k} + \frac{1}{2}\sum_{i=0}^{N-1} x_{i|k}^T Q x_{i|k} + u_{i|k}^T R u_{i|k}, \\
\text{s.t.} \quad & x_{i+1|k} = Ax_{i|k} + Bu_{i|k}, \quad i \in \mathbb{N}_{\le N-1}, \\
& Mx_{i|k} + Ju_{i|k} \le c, \qquad\quad i \in \mathbb{N}_{\le N-1}, \\
& M_N x_{N|k} \le c_N,
\end{aligned} \tag{2}$$

where $Q, Q_N \in \mathbb{S}^n_{\succeq 0}$, $R \in \mathbb{S}^m_{\succ 0}$ are the weighting matrices used in the MPC cost function and $N \in \mathbb{N}_{>0}$ denotes its prediction horizon. Moreover, $M \in \mathbb{R}^{p\times n}$, $J \in \mathbb{R}^{p\times m}$ and $c \in \mathbb{R}^p$ capture the polyhedral state and input constraints whereas $M_N \in \mathbb{R}^{q\times n}$ and $c_N \in \mathbb{R}^q$ characterize the terminal set constraints. The notation $x_{i|k}$ is used to denote the predicted state at future time instant $k+i$ in the optimal control problem solved at sampling instant k and similar notation is adopted for the future inputs $u_{i|k}$. Finally, the sequences of predicted states and future inputs are defined as $\boldsymbol{x}_k := \{x_{1|k}, x_{2|k}, \ldots, x_{N|k}\}$ and $\boldsymbol{u}_k := \{u_{0|k}, u_{1|k}, \ldots, u_{N-1|k}\}$. After the optimal input sequence $\boldsymbol{u}_k^* = \{u_{0|k}^*, u_{1|k}^*, \ldots, u_{N-1|k}^*\}$ is determined, the first control action is implemented, i.e. $u_k = u_{0|k}^*$, and the optimal control problem (2) is solved again at time $k+1$ for $x_{0|k+1} = x_{k+1}$ leading to the so-called receding horizon principle.

B. Non-condensed MPC-QP formulation

To solve the optimal control problem (2) it is formulated as a constrained QP. This formulation is not unique. In fact, each formulation has different consequences for the real-time implementation and the required computational effort. In this subsection, the non-condensed formulation is considered, which keeps both the predicted states and the future inputs as decision variables. Hence, the resulting QP contains $N(n+m)$ decision variables and the system dynamics are incorporated as equality constraints.

Define the vector of stacked decision variables $z_k \in \mathbb{R}^{N(n+m)}$ as

$$z_k := \begin{bmatrix} u_{0|k}^T & x_{1|k}^T & \ldots & u_{N-1|k}^T & x_{N|k}^T \end{bmatrix}^T. \tag{3}$$

The optimal control problem in (2) can then be rewritten to obtain the constrained QP

$$\begin{aligned}
\min_{z_k} \quad & \frac{1}{2} z_k^T H z_k, \\
\text{s.t.} \quad & F z_k = f, \\
& G z_k \le g,
\end{aligned} \tag{4}$$

where $H \in \mathbb{S}^{N(n+m)}_{\succ 0}$, $F \in \mathbb{R}^{Nn\times N(n+m)}$, $f \in \mathbb{R}^{Nn}$, $G \in \mathbb{R}^{(Np+q)\times N(n+m)}$ and $g \in \mathbb{R}^{Np+q}$ are given by

$$H = \begin{bmatrix} R & 0 & 0 \\ 0 & I_{N-1}\otimes\begin{bmatrix} Q & 0 \\ 0 & R\end{bmatrix} & 0 \\ 0 & 0 & Q_N \end{bmatrix}, \tag{5a}$$

$$F = \begin{bmatrix} -B & I & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -A & -B & I & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -A & -B & I \end{bmatrix}, \tag{5b}$$

$$f = \begin{bmatrix} \left(Ax_{0|k}\right)^T & 0 & \ldots & 0 \end{bmatrix}^T, \tag{5c}$$

$$G = \begin{bmatrix} J & 0 & 0 \\ 0 & I_{N-1}\otimes\begin{bmatrix} M & J \end{bmatrix} & 0 \\ 0 & 0 & M_N \end{bmatrix}, \tag{5d}$$

$$g = \begin{bmatrix} \left(c - Mx_{0|k}\right)^T & c^T & \ldots & c^T & c_N^T \end{bmatrix}^T. \tag{5e}$$

It should be noted that all matrices in the above formulation are sparse and feature a banded structure. Using general-purpose QP solvers, the computational cost for solving the non-condensed formulation (5) can be substantial. For instance, conventional interior point methods require $\mathcal{O}\left(N^3(n+m)^3\right)$ operations. However, solvers can exploit the sparsity and structure of the above QP to reduce the required computation power. For instance, a complexity of $\mathcal{O}\left(N(n+m)^3\right)$ operations is achieved in [3].

The dense formulation is obtained by eliminating the states from the decision variables through substitution of the equality constraints [7]. This yields a QP containing Nm decision variables for which the same general-purpose interior point methods require $\mathcal{O}\left(N^3m^3\right)$ operations. However, the condensed formulation no longer features sparsity that can be exploited to reduce the computational complexity. In the next subsection, a generalization of this condensing procedure is introduced, which can preserve the banded structure and the sparsity of the non-condensed QP.

C. Null-space method

The null-space method, alternatively referred to as the reduced Hessian approach, is well known in the optimization literature, see, e.g., [11], [12]. For its application to MPC-QP problems, see [9], [14]. The method works by decomposing the solution to the equality constraint in (5) into a homogeneous and a particular part. To this end, the notion of a null-space matrix is defined here.

Definition 1. *Given a matrix $A \in \mathbb{R}^{m\times n}$ with rank$(A) = r$ and $r < n$, the matrix $Z \in \mathbb{R}^{n\times n-r}$ is said to be a null-space matrix of A if it holds that $\mathrm{im}(Z) = \ker(A)$.*

The condition in Definition 1 is checked by verifying that $AZ = 0$ and rank$(Z) = n - r$, i.e. Z is full column rank.

Note that F in (5b) is full row rank, i.e., rank $(F) = Nn$, and, hence, $\dim\ker(F) = Nm$. Let $Z \in \mathbb{R}^{N(n+m)\times Nm}$ be a null-space matrix of F allowing the solution to the equality constraints to be decomposed as

$$z_k = Zy_k + \hat{z}_k, \tag{6}$$

where $y_k \in \mathbb{R}^{Nm}$ and $\hat{z}_k$ denotes a particular solution to the equality constraints, i.e. $F\hat{z}_k = f$. It is assumed that a $\hat{z}_k$ is given and it will be referred to as the particular solution.

Given a null-space matrix Z and the particular solution $\hat{z}_k$ for the equality constraints in (5), the parametrized solution (6) is substituted into (5) to obtain the following condensed QP:

$$\begin{aligned} \min_{y_k} \quad & \frac{1}{2}y_k^T\tilde{H}y_k + \tilde{h}^T y_k, \\ \text{s.t.} \quad & \tilde{G}y_k \le \tilde{g}, \end{aligned} \tag{7}$$

where $\tilde{H} = Z^T HZ$, $\tilde{h} = Z^T H\hat{z}_k$, $\tilde{G} = GZ$ and $\tilde{g} = g - G\hat{z}_k$. The resulting condensed QP contains $\dim\ker(F) = Nm$ decision variables, which is substantially lower (especially) when $m \ll n$ and it is the same as in the dense formulation. In fact, to justify calling the null-space method a generalization of the condensing procedure that leads to the dense formulation, it should be mentioned that the dense formulation is recovered from the null-space method by taking

$$Z = \begin{bmatrix} I & 0 & \dots & 0 \\ B & 0 & \dots & 0 \\ 0 & I & \dots & 0 \\ AB & B & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I \\ A^{N-1}B & A^{N-2}B & \dots & B \end{bmatrix} \text{ and } \hat{z}_k = \begin{bmatrix} 0 \\ Ax_{0|k} \\ 0 \\ A^2x_{0|k} \\ \vdots \\ 0 \\ A^N x_{0|k} \end{bmatrix}. \tag{8}$$

The columns of Z represent (delayed) impulse responses of the system (1) and $\hat{z}_k$ accounts for the initial condition $x_{0|k}$.

The null-space matrix Z, however, is not unique and this freedom can be exploited to introduce sparsity and desirable structure into the condensed problem. It has been shown that for banded matrices, such as F in (5b), a banded null-space matrix may exist [13]. Using a banded null-space matrix is beneficial since the matrices $\tilde{H}$ and $\tilde{G}$ in the condensed QP inherit this banded structure which, in turn, can be exploited by tailored QP solvers [9]. Hence, it is of interest to construct a null-space matrix of low bandwidth. The concept of bandwidth will be formally defined in Section III.

A method for constructing a banded null-space matrix for the equality constraints in MPC problems for single-input systems is presented in [14]. The method in [8] works for multi-input systems, however, it relies on the non-trivial task of computing a deadbeat feedback gain K and it computes a null-space matrix for an MPC formulation in terms of the system $x_{k+1} = (A+BK)x_k + Bu_k$ rather than for F in (5).

Finally, an important result related to the null-space method is presented below [12], [9].

Proposition 1. *If Z is a null-space matrix of F, then the condensed Hessian is positive definite, i.e. $\tilde{H} = Z^T HZ \in \mathbb{S}^{Nm}_{\succ 0}$.*

Proof. This is shown by proving that H is positive definite on the kernel of F or, equivalently, $\ker(H) \cap \ker(F) = \{0\}$. Let $Y = \begin{bmatrix} H & F^T \end{bmatrix}^T \in \mathbb{R}^{N(2n+m)\times N(n+m)}$, then $\ker(Y) = \ker(H) \cap \ker(F)$. By inspection, since $R \in \mathbb{S}^m_{\succ 0}$, Y is full column rank and, thus, $\ker(H) \cap \ker(F) = \{0\}$. □

III. Problem definition

As mentioned before, imposing a banded structure on the null-space matrix reduces the required computational complexity. For square matrices, the bandwidth provides a well-defined measure of bandedness. For matrices composed of N^2 blocks, e.g. Z and $\tilde{H}$, the concept *N-block bandwidth* is introduced to quantify the extend to which they are banded:

Definition 2. *The matrix*

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,N} \\ A_{2,1} & A_{2,2} & \dots & A_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N,1} & A_{N,2} & \dots & A_{N,N} \end{bmatrix} \in \mathbb{R}^{Ns\times Nt}, \tag{9}$$

where $A_{i,j} \in \mathbb{R}^{s\times t}$, is called an N-block banded matrix with N-block bandwidth $k \in \mathbb{N}_{\le N-1}$, denoted $\beta_N(A) = k$, if

$$A_{i,j} = 0, \quad \text{for } (i,j) \in \mathbb{N}^2_{[1,N]} \text{ with } |i-j| > k. \tag{10}$$

The problem addressed in the remainder of this paper is defined as follows: Given the finite-horizon linear-quadratic optimal control problem in (2) for multi-input state-space systems as well as the corresponding non-condensed QP formulation (5), can one systematically construct a banded null-space matrix Z with low $\beta_N(Z)$?

IV. System-theoretic construction of a banded null-space matrix

From a system-theoretic perspective, any vector in the null space of F in (5) can be seen as a sequence of inputs and corresponding states for the system in (1) with $x_0 = 0$ and delayed versions thereof. Therefore, a banded null-space matrix can be constructed using system responses that satisfy

$$x_k \ne 0, \qquad \forall k \in \mathbb{N}_{[1,\mu-1]}, \tag{11a}$$

$$x_k = 0, \qquad \forall k \in \{0\} \cup \mathbb{N}_{[\mu,N]}, \tag{11b}$$

where $\mu \le N$ and delayed versions of such responses. Such responses will be referred to as two-sided deadbeat responses of length $\mu - 1$ in the remainder of this paper. To be able to find such responses, the prediction horizon N is assumed to be sufficiently large in the remainder of this paper. It should be noted that the minimum achievable length of the two-sided deadbeat response is a property of the system and that no improvement in terms of N-block bandwidth can be obtained when this minimum length exceeds the prediction horizon.

Motivated by this system-theoretic insight, the construction of a banded null-space matrix is approached by computing m two-sided deadbeat responses. These responses along with N delayed versions of each response may then be used to construct the Nm columns of the null-space matrix Z. To ensure that the resulting Z is full column rank, as required by Definition 1, the first inputs of the original, i.e., without delay, responses are selected to be linearly independent. As a consequence, the resulting Z might not have minimal block bandwidth, but contains at most one additional nonzero block diagonal. Finally, two-sided deadbeat responses can also be constructed for uncontrollable systems since starting from the origin means that the uncontrollable dynamics are never excited. Hence, the null-space method can be applied to a larger class of systems than the method in [8], which requires null-controllability.

Observe that an input sequence that excites a two-sided deadbeat response of length $\mu - 1$ lies in the null-space of the μ-step reachability matrix. The k-step reachability matrix $R_k(A, B)$ is given by

$$R_k(A,B) = \begin{bmatrix} A^{k-1}B & \dots & AB & B \end{bmatrix}, \tag{12}$$

see also [15]. Using this insight, Algorithm 1 can be used to construct a banded null-space matrix for the equality constraints in (5).

Remark 1. *Algorithm 1 does not require U_0 to be the identity matrix as long as it is chosen to be full rank.*

Remark 2. *By the Rouché-Capelli theorem [17], Equation* (14) *admits a solution if $rank([\mathcal{A}|\mathcal{B}U_0]) = rank(\mathcal{A})$.*

Algorithm 1 computes a null-space matrix for F in (5):

Proposition 2. *The matrix Z as obtained by Algorithm 1 is a null-space matrix of F in* (5b).

Proof. To prove that Z is a null-space matrix of F, it suffices to verify that Z is full column rank and $FZ = 0$.

The first condition follows readily from the fact that the columns of Z as constructed in (16) are linearly independent as a result of U_0 being full rank.

To show that $FZ = 0$, consider the left-most block column given by $Z_1 = \begin{bmatrix} U_0^T & X_1^T & \dots & U_{\mu-1}^T & 0 \end{bmatrix}^T$. By inspection of F (5b), $FZ_1 = 0$ if $X_{i+1} - AX_i - BU_i = 0$ for $i \in \mathbb{N}_{\mu-1}$ with $X_0 = 0$ or, equivalently, $X_i = A^{i-1}BU_0 + \ldots + ABU_{i-2} + BU_{i-1} = R_i(A,B) \begin{bmatrix} U_0^T & \dots & U_i^T \end{bmatrix}^T$ which is precisely the construction used in (15). Hence, $FZ_1 = 0$ and, since after shifting Z_1 downwards by an integer multiple of $n{+}m$ rows and padding with zeros from above it still satisfies these conditions, it follows that $FZ = 0$. □

The N-block bandwidth of the computed null-space matrix, i.e. $\beta_N(Z)$, is related to the length $\mu{-}1$ of the two-sided deadbeat responses computed in Algorithm 1.

Proposition 3. *The null-space matrix Z computed by Algorithm 1 is an N-block banded matrix with $\beta_N(Z) = \mu - 1$.*

Algorithm 1: Banded null-space matrix

---

1. Transform (1) into its controllability staircase form using unitary similarity transformations [16], i.e.

$$T^T A T = \begin{bmatrix} A_c & A_{12} \\ 0 & A_{\bar{c}} \end{bmatrix} \text{ and } T^T B = \begin{bmatrix} B_c \\ 0 \end{bmatrix}, \tag{13}$$

where $A_c \in \mathbb{R}^{\nu \times \nu}$ and (A_c, B_c) is controllable;
2. Let $\mu = 2$, $U_0 = I_m$, $\mathcal{A} = B_c$ and $\mathcal{B} = A_c B_c$;
while *rank*$([\mathcal{A}|\mathcal{B}U_0]) \neq$ *rank*$(\mathcal{A})$ **do**
 3. $\mu = \mu + 1$;
 4. $\mathcal{A} = [\mathcal{B}\,|\mathcal{A}]$;
 5. $\mathcal{B} = A_c\mathcal{B}$;
end
6. Find a solution $\mathcal{U}$ to the system of equations

$$\mathcal{A}\mathcal{U} = -\mathcal{B}U_0, \tag{14}$$

which contains the remainder of the input sequence;
7. Partition according to $\mathcal{U} = \begin{bmatrix} U_1^T & \dots & U_{\mu-1}^T \end{bmatrix}^T$, where $U_i \in \mathbb{R}^{m \times m}$ for $i \in \mathbb{N}_{[1,\mu-1]}$;
8. Compute the corresponding state trajectories as

$$X_i = R_i(A,B) \begin{bmatrix} U_0^T & \dots & U_i^T \end{bmatrix}^T, \tag{15}$$

where $X_i \in \mathbb{R}^{n \times m}$ for all $i \in \mathbb{N}_{[1,\mu-1]}$;
9. Construct the null-space matrix Z of F in (5b) as

$$Z = \begin{bmatrix} U_0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ X_1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ U_1 & U_0 & \dots & 0 & 0 & \dots & 0 & 0 \\ X_2 & X_1 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ U_{\mu-1} & U_{\mu-2} & \dots & U_i & U_{i-1} & \dots & 0 & 0 \\ 0 & X_{\mu-1} & \dots & X_{i+1} & X_i & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & U_{\mu-1} & U_{\mu-2} & \dots & U_1 & U_0 \\ 0 & 0 & \dots & 0 & X_{\mu-1} & \dots & X_2 & X_1 \end{bmatrix} \tag{16}$$

Proof. The null-space matrix Z in (16) is composed of N^2 blocks $Z_{i,j} \in \mathbb{R}^{(n+m)\times m}$. The two-sided deadbeat responses satisfy $X_i = U_i = 0$ for $i > \mu$ and $U_{\mu-1} \neq 0$. By inspection of (16), it can be seen that $Z_{i,j} = 0$ for $|i - j| > \mu - 1$ and, hence, $\beta_N(Z) = \mu - 1$. □

Next, an upper bound on the length of the two-sided deadbeat responses and the N-block bandwidth of the null-space matrix computed by Algorithm 1 is provided. To this end, let ν denote the dimension of the controllable subspace, i.e. $A_c \in \mathbb{R}^{\nu \times \nu}$ in (13).

Proposition 4. *Given the system in* (1)*, Algorithm 1 leads to responses of length $\mu - 1 \leq \nu - m + 1$ and a null-space matrix with $\beta_N(Z) \leq \nu - m + 1$.*

Proof. Since B is full column rank in (1), B_c is also full column rank. Hence, by Corollary 6.1 in [18], the reachability matrix $R_{\nu-m+1}(A_c, B_c)$ is full rank as a result of the pair (A_c, B_c) being controllable. Observe that the while-loop in Algorithm 1 always breaks when $\mathcal{A}$ is full rank (it may break earlier). Moreover, after breaking at a given μ it holds that $\mathcal{A} = R_{\mu-1}(A_c, B_c)$ and, thus, the while-loop breaks when at most $\mu - 1 = \nu - m + 1$. Therefore, the algorithm computes responses of length $\mu - 1 \leq \nu - m + 1$ and, by Proposition 3, $\beta_N(Z) \leq \nu - m + 1$. □

V. Banded condensed MPC-QP formulation

In this section, it is assumed that a banded null-space matrix of bandwidth $\beta_N(Z) = \mu - 1$ is available and can be used to eliminate the equality constraints and arrive at a sparse-condensed formulation as in (7). As mentioned before, the sparse-condensed QP inherits the banded structure of the null-space matrix, i.e. the matrices $\tilde{H}$ and $\tilde{G}$ in the condensed QP are also banded.

Proposition 5. *Let Z be an N-block banded matrix with $\beta_N(Z) = \mu - 1$. Then, the condensed Hessian $\tilde{H}$ and inequality constraint matrix $\tilde{G}$ in (7) are also N-block banded matrices with $\beta_N(\tilde{H}) = \beta_N(\tilde{G}) = \mu - 1$.*

Proof. This can be seen by inspection of $\tilde{H} = Z^T H Z$ and $\tilde{G} = GZ$ combined with the block banded lower-triangular structure of Z. □

To illustrate how the structure of the sparse-condensed formulation can be exploited by QP solvers, the methods GPAD and ADMM are considered here, see [5] and [6] respectively. Let us remark here that gradient-projection and ADMM are considered to be amongst the most efficient and suitable methods for embedded MPC applications [19]. For both GPAD and ADMM the computationally most expensive step consists of solving a system of the form

$$\mathcal{H}y = h, \tag{17}$$

where $\mathcal{H} = \tilde{H}$ for GPAD and $\mathcal{H} = \tilde{H} + \rho \tilde{G}^T \tilde{G}$ for ADMM with $\rho > 0$ denoting the step size which remains fixed across iterations. Notice that in both cases $\mathcal{H}$ is constant, positive definite (by Proposition 1) and, since $\beta_N(\tilde{G}^T \tilde{G}) = \beta_N(\tilde{G})$, satisfies $\beta_N(\mathcal{H}) = \mu - 1$. One can solve (17) by pre-multiplying by $\mathcal{H}^{-1}$, however, the matrix inversion destroys the banded pattern and this results in a computational complexity of $\mathcal{O}(N^2 m^2)$ operations to solve (17). Instead, let L be a Cholesky factorization of $\mathcal{H}$, i.e. $\mathcal{H} = LL^T$, such that (17) can be solved using forward and backward substitutions in the following two steps:

$$Lv = b, \qquad L^T y = v. \tag{18}$$

Using this approach the banded structure of $\mathcal{H}$ is exploited since L is an N-block banded matrix with $\beta_N(L) = \mu - 1$ composed of m-by-m blocks. Specifically, L consists of $\sum_{i=0}^{\mu-1} N - i$ blocks, leading to a reduced computational complexity to solve (17) of $\mathcal{O}(N\mu m^2)$ operations for the forward and backwards substitutions [20]. Note that the computation of L is performed off-line. For the dense formulation, since $\mu = N$, no improvement over the approach using matrix inversion is obtained. The computation of the particular solution $\hat{z}_k$ as in (8) takes $\mathcal{O}(Nn^2)$ operations. Hence, the overall computational complexity scales linearly with the prediction horizon.

The ability to exploit the banded structure of the sparse-condensed formulation is not limited to GPAD and ADMM, in fact methods in which the optimality conditions involve solving a system of the form in (17) with constant $\mathcal{H}$ will be able to exploit structure in a similar fashion. However, for solvers such as interior-point methods which involve a time-varying Hessian applying the null-space method becomes less attractive as the condensed Hessian has to be recomputed on-line.

Both the dense formulation and sparse-condensed formulation are ill-conditioned for certain systems. Defining $u_k = Kx_k + v_k$ with pre-stabilizing K and constructing the null-space matrix for the transformed system $x_{k+1} = (A + BK)\,x_k + Bv_k$ can help to improve conditioning for some systems [21]. We have observed, however, that the choice of null-space matrix, e.g., dense formulation (8) versus sparse-condensed (16), also affects numerical conditioning. Hence, it may be possible to take the numerical conditioning of the condensed QP into account in Algorithm 1 when constructing the null-space matrix. This is the focus of future research.

VI. Numerical example

The proposed approach is implemented for the CD player benchmark example in [22]. The CD player model features 120 states, 2 inputs representing the swing arm and focus lens and 2 outputs corresponding to the tracking error and focus error [22]. Hence, $n = 120$, $m = 2$ and $m \ll n$. The continuous-time model is discretized with a sampling period of 0.1s. The resulting discrete-time system is asymptotically stable.

The control objective is to minimize the tracking errors while satisfying constraints on the tracking errors and the inputs. These constraints are captured by defining the matrices in (2) as

$$M = \begin{bmatrix} C \\ -C \\ 0 \\ 0 \end{bmatrix},\ J = \begin{bmatrix} 0 \\ 0 \\ I \\ -I \end{bmatrix} \text{ and } c = \begin{bmatrix} \bar{y} \\ -\underline{y} \\ \bar{u} \\ -\underline{u} \end{bmatrix}, \tag{19}$$

where $\bar{y} = -\underline{y} = \begin{bmatrix} 100 & 5 \end{bmatrix}^T$, $\bar{u} = -\underline{u} = \begin{bmatrix} 0.001 & 0.005 \end{bmatrix}^T$ and with $M_N = M$ and $c_N = c$. Moreover, the cost function in (2) is defined in terms of the matrices

$$Q = Q_N = C^T \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}^2 C \text{ and } R = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix} \cdot 10^{-3}. \tag{20}$$

In Algorithm 1, the system is first transformed to staircase controllability form using the method in [16]. Based on this formulation, the dimension of the controllable subspace is determined to be $\nu = 69$ and, thus, the system is (atleast numerically) uncontrollable. The algorithm computes several linearly independent two-sided deadbeat responses of length $\mu - 1 = 36$. The corresponding banded null-space matrix (16) is used to construct the condensed Hessian $\tilde{H}$ and condensed inequality constraint matrix $\tilde{G}$ which, by Proposition 5, satisfy $\beta_N(\tilde{H}) = \beta_N(\tilde{G}) = \mu - 1 = 36$. The banded structure of the condensed Hessian is illustrated for prediction horizons $N \in \{40, 100, 150\}$ in Fig. 1. It can be seen that the number of non-zero entries, denoted `nz` in Fig. 1, scales linearly with the prediction horizon.

Both MPC using the dense and the sparse-condensed formulation is simulated in closed-loop with the system for 200 time steps for a range of prediction horizons $N \in [40, 150]$ ($N > \mu$). Simulations are carried out using both GPAD and ADMM for each formulation and the mean computation time

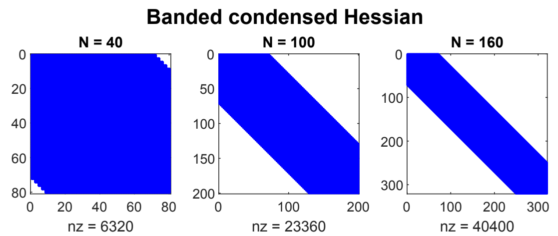

Fig. 1. Banded condensed Hessian, i.e. $\tilde{H}$ for GPAD and $\tilde{H} + \rho \tilde{G}^T \tilde{G}$ for ADMM, vs. prediction horizon N.

per iteration (TPI) over all time-steps and solver iterations is computed. The computed TPI excludes the time required to check the termination criteria. The simulations all have initial condition zero and Gaussian white noise is added to the inputs. Figure 2 depicts the resulting computation times. It can be seen that, as expected, the TPI scales linearly for both solvers applied to the sparse-condensed formulation. For all prediction horizons $N > \mu$, the sparse-condensed formulation is never (significantly) slower and, for large N, a substantial improvement in computational requirements can be achieved.

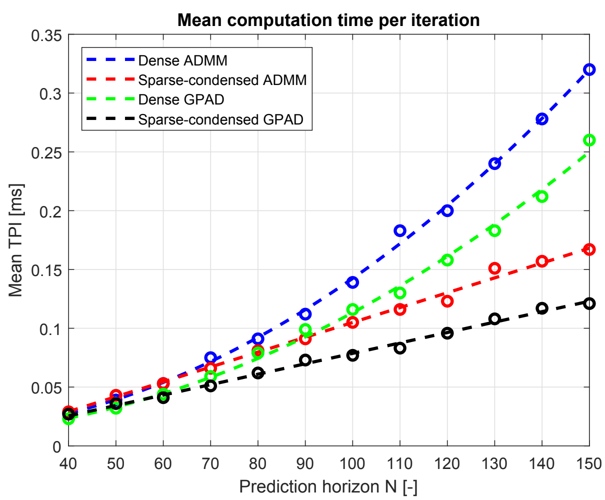

Fig. 2. Mean computation time per iteration (TPI) vs. prediction horizon for ADMM and GPAD applied to the dense and sparse-condensed formulations.

VII. Conclusions

This paper presented a systematic approach, inspired by system-theoretic insights, to construct a banded null-space matrix for the equality constraints encountered in MPC-QP problems. The constructed null-space matrix can be used to construct a sparse-condensed QP, which contains a reduced number of decision variables while preserving the banded structure of the sparse QP formulation. In contrast to existing approaches, this method is applicable to uncontrollable, thought stabilizable, systems with multiple inputs.

By exploiting the banded structure of the sparse-condensed formulation, the computational requirements to compute a solution scales linearly with the prediction horizon. This has been demonstrated for a numerical example using different solvers based on GPAD and ADMM. This reduction in computational power contributes to enabling MPC for large-scale systems and using longer prediction horizons. Understanding and exploiting how the construction of the null-space matrix affects the numerical conditioning of the sparse-condensed formulation remains a topic for future research.

References

[1] D. Q. Mayne, "Model predictive control: Recent developments and future promise," *Automatica*, vol. 50, no. 12, pp. 2967–2986, Dec 2014.

[2] R. Bartlett, A. Wachter, and L. Biegler, "Active set vs. interior point strategies for model predictive control," in *Proceedings of the 2000 American Control Conference*, vol. 6. IEEE, 2000, pp. 4229–4233.

[3] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 2, pp. 267–278, Mar 2010.

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

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

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

[7] J. M. Maciekowski, *Predictive Control with Constraints*. Harlow, UK: Prentice Hall, 2002.

[8] "A sparse and condensed QP formulation for predictive control of LTI systems," *Automatica*, vol. 48, no. 5, pp. 999 – 1002, 2012.

[9] T. V. Dang, K. V. Ling, and J. Maciejowski, "Banded null basis and ADMM for embedded MPC," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 13 170 – 13 175, 2017, 20th IFAC World Congress.

[10] P. Van Dooren, "Deadbeat control: A special inverse eigenvalue problem," *BIT Numerical Mathematics*, vol. 24, no. 4, pp. 681–699, Dec 1984.

[11] M. Benzi, G. H. Golub, and J. Liesen, "Numerical solution of saddle point problems," *Acta Numerica*, vol. 14, pp. 1–137, 2005.

[12] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY: Springer New York, 2006.

[13] M. W. Berry, M. T. Heath, I. Kaneko, M. Lawo, R. J. Plemmons, and R. C. Ward, "An algorithm to compute a sparse basis of the null space," *Numerische Mathematik*, vol. 47, no. 4, pp. 483–504, Dec 1985.

[14] T. V. Dang, "Efficient algorithms for embedded optimisation-based control," Ph.D. dissertation, Nanyang Technological University, 2018.

[15] L. Caccetta and V. Rumchev, "A survey of reachability and controllability for positive linear systems," *Annals of Operations Research*, vol. 98, no. 1, pp. 101–122, Dec 2000.

[16] A. Varga, "Numerically stable algorithm for standard controllability form determination," *Electronics Letters*, vol. 17, no. 2, pp. 74–75, Jan 1981.

[17] I. R. Shafarevich and A. O. Remizov, *Linear algebra and geometry*. Springer Science & Business Media, 2012.

[18] C.-T. Chen, *Linear System Theory and Design*, 3rd ed. New York: Oxford University Press, 1999.

[19] I. McInerney, G. A. Constantinides, and E. C. Kerrigan, "A Survey of the Implementation of Linear Model Predictive Control on FPGAs," *IFAC-PapersOnLine*, vol. 51, no. 20, pp. 381 – 387, 2018, 6th IFAC Conference on Nonlinear Model Predictive Control NMPC 2018.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[21] J. Rossiter, B. Kouvaritakis, and M. Rice, "A numerically robust state-space approach to stable-predictive control strategies," vol. 34, no. 1, 1998, pp. 65 – 73.

[22] P. M. R. Wortelboer, M. Steinbuch, and O. H. Bosgra, "Iterative model and controller reduction using closed-loop balancing, with application to a compact disc mechanism," *International Journal of Robust and Nonlinear Control*, vol. 9, no. 3, pp. 123–142, 1999.