Automatica 48 (2012) 999–1002

Contents lists available at SciVerse ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Technical communique

A sparse and condensed QP formulation for predictive control of LTI systems☆

Juan L. Jerez [a,1], Eric C. Kerrigan [a,b], George A. Constantinides [a]

[a] Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ, United Kingdom

[b] Department of Aeronautics, Imperial College London, SW7 2AZ, United Kingdom

ARTICLE INFO

Article history:
Received 19 April 2011
Received in revised form
17 October 2011
Accepted 17 February 2012
Available online 27 March 2012

Keywords:
Predictive control
Optimization
Quadratic programming
Linear systems

ABSTRACT

The computational burden that model predictive control (MPC) imposes depends to a large extent on the way the optimal control problem is formulated as an optimization problem. We present a formulation where the input is expressed as an affine function of the state such that the closed-loop dynamics matrix becomes nilpotent. Using this approach and removing the equality constraints leads to a compact and sparse optimization problem to be solved at each sampling instant. The problem can be solved with a cost per interior-point iteration that is linear with respect to the horizon length, when this is bigger than the controllability index of the plant. The computational complexity of existing condensed approaches grow cubically with the horizon length, whereas existing non-condensed and sparse approaches also grow linearly, but with a greater proportionality constant than with the method presented here.

1. Introduction

In linear MPC, at every sampling instant, the optimal control input is determined through the solution of a convex optimization problem with a quadratic cost and linear constraints. MPC's natural ability for handling physical constraints has the potential to deliver significant performance benefits in many application areas. However, the very high computational demands mean that, if at all possible, expensive power-hungry hardware is often required to meet the application's sampling requirements. This has so far hindered the widespread use of the technology.

Given an estimate or measurement of the current state of the plant $\widehat{x}$, the constrained LQR problem that we will consider is

$$\min\ x_N^T \widetilde{Q} x_N + \sum_{k=0}^{N-1} \begin{bmatrix} x_k \\ u_k \end{bmatrix}^T \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} x_k \\ u_k \end{bmatrix} \tag{1}$$

subject to

$$x_0 = \widehat{x} \tag{2a}$$

$$x_{k+1} = Ax_k + Bu_k \quad \text{for } k = 0, 1, 2, \ldots, N-1 \tag{2b}$$

$$u_k = Kx_k + v_k \quad \text{for } k = 0, 1, 2, \ldots, N-1 \tag{2c}$$

$$Jx_k + Eu_k \le d \quad \text{for } k = 0, 1, 2, \ldots, N-1 \tag{2d}$$

where N is the horizon length, $x_k \in \mathbb{R}^n$ is the state vector at sample instant k, $u_k \in \mathbb{R}^m$ is the input vector, (A, B) is controllable, $(Q^{\frac{1}{2}}, A)$ is detectable, $\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \ge 0$, $R > 0$ to ensure uniqueness of the solution, $\widetilde{Q} \ge 0$, (2c) represents a possible affine transformation on the input, $J \in \mathbb{R}^{l\times n}$, $E \in \mathbb{R}^{l\times m}$, $d \in \mathbb{R}^l$ and l is the number of constraints. The techniques described in this note can easily be extended to problems with costs and constraints on the input rates, time-varying costs and constraints, as well as problems with linear terms in the cost function.

The method employed when formulating the constrained LQR problem as a QP has a big impact on the problem size and structure, the resulting computational and memory requirements, as well as on the numerical conditioning. The standard approach makes use of the plant dynamics to eliminate the states from the decision variables by expressing them as an explicit function of the current state and future control inputs (Maciejowski, 2001). This *condensed* formulation leads to compact and dense QPs. In this case, the complexity of solving the QP scales cubically in the horizon length when using an interior-point method. For MPC problems that require long horizon lengths, the *non-condensed* formulation, which keeps the states as decision variables and considers the system dynamics implicitly by enforcing equality constraints (Rao, Wright, & Rawlings, 1998; Wright, 1993, 1996), can result in significant speed-ups. With this approach the problem becomes larger but its structure can be exploited to find a solution in time linear in the horizon length.

☆ This work was supported by the EPSRC (grants EP/G031576/1 and EP/I012036/1) and the EC FP7 project EMBOCON. The material in this paper was presented at the 50th IEEE Conference of Decision and Control (CDC)/European Control Conference (ECC), December 12–15, 2011, Orlando, Florida, USA. This paper was recommended for publication in revised form by Associate Editor Mikael Johansson under the direction of Editor André L. Tits.

E-mail addresses: jlj05@imperial.ac.uk (J.L. Jerez), e.kerrigan@imperial.ac.uk (E.C. Kerrigan), gac1@imperial.ac.uk (G.A. Constantinides).

[1] Tel.: +44 2075946261; fax: +44 2075814419.

0005-1098/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2012.03.010

The non-condensed method is often also referred to as the sparse method due to the abundant structure in the resulting optimization problems. In this note, we show that this label does not provide the complete picture and that it is indeed possible to have a sparse condensed formulation that can be solved in time linear in the horizon length. In addition, we show that this method is at least as fast as the standard condensed formulation and it is faster than the non-condensed formulation for a wide variety of control problems. Our approach is based on the use of the linear feedback policy in (2c) as a mathematical trick to introduce structure into the problem. We choose K such that $A+BK$ is nilpotent and show that one can formulate a QP with banded matrices in cases where the horizon length is larger than the controllability index of the plant. The use of such feedback policies for pre-stabilizing predictions has been previously studied as a way of improving the conditioning of guaranteed stability MPC algorithms (Rossiter, Kouvaritakis, & Rice, 1998). However, we find it surprising that it has not yet been applied to introduce structure into the problem, as we will do here, considering the important practical implications.

Note that (2c) is effectively only a change of variables and it does not modify the optimal control problem, hence the computed optimal input is independent of the transformation used. Moreover, any procedure to guarantee stability and feasibility can still be used. For example, if the method in Scokaert and Rawlings (1998) is used, then $\widetilde{Q} \in \mathbb{R}^{n\times n}$ is the solution to the appropriate Riccati equation, which is independent of the choice of K in (2c).

We start by reviewing existing QP formulations and analyzing their computational complexity in the context of primal–dual interior-point methods. However, the results stated in this paper should have a similar impact on barrier-based interior-point methods and active-set methods. We then present our sparse condensed approach and compare it with existing formulations.

2. QP formulation problem

We consider the problem of formulating the optimal control problem (1)–(2) as a convex QP of the following form:

$$\min_{\theta} \frac{1}{2}\theta^T H\theta + h^T\theta \quad \text{subject to } F\theta = f,\ G\theta \le g. \tag{3}$$

Primal–dual interior-point methods can be used to solve for optimal θ. Algorithm 1 is a variant of an infeasible primal–dual method (Wright, 1996), where ν and λ are Lagrange multipliers for the equality and inequality constraints, respectively, s is a vector of slack variables, σ is a small constant between zero and one, $W_k := \Lambda_k S_k^{-1}$, Λ_k and S_k are diagonal matrices containing the elements of λ_k and s_k, respectively, and $\mu_k := (\lambda_k^T s_k)/(Nl)$ is a measure of sub-optimality that approaches zero at the optimum. In applications with fast dynamics, real-time requirements will impose a hard bound on the number of interior-point iterations, hence the number of interior-point iterations P is assumed fixed a priori.

At each interior-point iteration, computing the matrix triple product $G^T W_k G$ (line 1) and solving the system of linear equations $\mathcal{A}_k z_k = b_k$ (line 3) account for most of the computation, hence we will express the overall complexity considering the cost of these operations only.

3. Non-condensed approach

The future states can be kept as decision variables and the system dynamics can be incorporated into the problem by enforcing equality constraints (Rao et al., 1998; Wright, 1993, 1996). In this case, for any arbitrary K, if we let $\theta := [\mathbf{x}^T\ \mathbf{v}^T]^T$ where

$$\mathbf{x} := [x_0^T\ x_1^T\ \cdots\ x_N^T]^T, \qquad \mathbf{v} := [v_0^T\ v_1^T\ \cdots\ v_{N-1}^T]^T,$$

$h := 0$, then the remaining matrices have sparse structures that describe the control problem (1)–(2) exactly.

Algorithm 1 Primal–Dual Interior-Point Algorithm

Choose any initial point $(\theta_0, \nu_0, \lambda_0, s_0)$ with $[\lambda_0^T\ s_0^T]^T > 0$
for $k = 0$ to $P-1$ **do**

$$\mathcal{A}_k := \begin{bmatrix} H + G^T W_k G & F^T \\ F & 0 \end{bmatrix}$$

$$b_k := \begin{bmatrix} -h - F^T\nu - G^T(\lambda_k - W_k g + \sigma\mu_k s_k^{-1}) \\ -F\theta_k + f \end{bmatrix}$$

Solve $\mathcal{A}_k z_k = b_k$ for $z_k =: \left[(\theta_k + \Delta\theta_k)^T\ \Delta\nu_k^T\right]^T$

$\Delta\lambda_k := W_k(G(\theta_k + \Delta\theta_k) - g) + \sigma\mu_k s_k^{-1}$

$\Delta s_k := -s_k - (G(\theta_k + \Delta\theta_k) - g)$

$\alpha_k := \max\{\alpha \in (0,1] : (\lambda_k, s_k) + \alpha(\Delta\lambda_k, \Delta s_k) > 0\}$

$(\theta_{k+1}, \nu_{k+1}, \lambda_{k+1}, s_{k+1}) :=$
$\quad (\theta_k, \nu_k, \lambda_k, s_k) + \alpha_k(\Delta\theta_k, \Delta\nu_k, \Delta\lambda_k, \Delta s_k)$

end for

Assuming general constraints, the number of floating point operations (flops) for computing $G^T W_k G$ is approximately $Nl(n+m)^2$ operations. For solving $\mathcal{A}_k z_k = b_k$, the coefficient matrix $\mathcal{A}_k \in \mathbb{R}^{N(2n+m)\times N(2n+m)}$ is an indefinite symmetric matrix that can be made banded through appropriate row re-ordering (or interleaving of variables $\Delta\theta$ and $\Delta\nu$). The resulting banded matrix has a half-band of size $2n+m$. Such a linear system can be solved using a banded LDL^T factorization in $N(2n+m)^3 + 4N(2n+m)^2 + N(2n+m)$ flops (Boyd & Vandenberghe, 2004, Appendix C), or through a block factorization method based on a sequence of Cholesky factorizations in $\mathcal{O}(N(n+m)^3)$ operations (Rao et al., 1998). The memory requirements can be approximated by the cost of storing matrices H, G, F and $\mathcal{A}_k$, which are all sparse. For time-invariant problems, these matrices mostly consist of repeated blocks.

4. Condensed approach

The state variables can be eliminated from the optimization problem by expressing them as an explicit function of the current state and the controlled variables (Maciejowski, 2001):

$$\mathbf{x} = \mathbf{A}\widehat{x} + \mathbf{B}\mathbf{v}, \tag{4}$$

where $A_K := A + BK$ and

$$\mathbf{A} := \begin{bmatrix} I_n \\ A_K \\ A_K^2 \\ \vdots \\ A_K^{N-1} \\ A_K^N \end{bmatrix}, \qquad \mathbf{B} := \begin{bmatrix} 0 & & & & \\ B & 0 & & & \\ A_K B & B & \ddots & & \\ \vdots & & \ddots & & \\ A_K^{N-2}B & & & B & 0 \\ A_K^{N-1}B & A_K^{N-2}B & \cdots & A_K B & B \end{bmatrix}. \tag{5}$$

In this case, if we let $\theta := \mathbf{v}$, $F := 0$, $f := 0$, then we have an inequality constrained QP with

$$\begin{aligned} H &:= \mathbf{B}^T(\mathbf{Q} + \mathbf{K}^T\mathbf{R}\mathbf{K} + \mathbf{S}\mathbf{K} + \mathbf{K}^T\mathbf{S}^T)\mathbf{B} \\ &\quad + \mathbf{R} + \mathbf{B}^T(\mathbf{K}^T\mathbf{R} + \mathbf{S}) + (\mathbf{R}\mathbf{K} + \mathbf{S}^T)\mathbf{B}, \end{aligned}$$

$$h := \widehat{x}^T\mathbf{A}^T(\mathbf{Q}\mathbf{B} + \mathbf{S}(\mathbf{K}\mathbf{B} + I) + \mathbf{K}^T(\mathbf{R}(\mathbf{K}\mathbf{B} + I) + \mathbf{S}^T\mathbf{B})),$$

$$G := (\mathbf{J} + \mathbf{E}\mathbf{K})\mathbf{B} + \mathbf{E},$$

$$g := \mathbf{d} - (\mathbf{J} + \mathbf{E}\mathbf{K})\mathbf{A}\widehat{x},$$

where

$$\mathbf{Q} := \begin{bmatrix} I_N \otimes Q & 0 \\ 0 & \widetilde{Q} \end{bmatrix}, \qquad \mathbf{S} := \begin{bmatrix} I_N \otimes S \\ 0 \end{bmatrix},$$

$$\mathbf{R} := I_N \otimes R,$$

$$\mathbf{K} := \begin{bmatrix} I_N \otimes K & 0 \end{bmatrix}, \qquad \mathbf{J} := I_N \otimes J, \qquad \mathbf{E} := I_N \otimes E,$$

$\mathbf{d} := 1_N \otimes d$, $\otimes$ denotes a Kronecker product and 1_N denotes a vector of ones of length N.