

ADMM for MPC with state and input constraints, and input nonlinearity

Sebastian East¹, Mark Cannon¹

Abstract—In this paper we propose an Alternating Direction Method of Multipliers (ADMM) algorithm for solving a Model Predictive Control (MPC) optimization problem, in which the system has state and input constraints and a nonlinear input map. The resulting optimization is nonconvex, and we provide a proof of convergence to a point satisfying necessary conditions for optimality. This general method is proposed as a solution for blended mode control of hybrid electric vehicles, to allow optimization in real time. To demonstrate the properties of the algorithm we conduct numerical experiments on randomly generated problems, and show that the algorithm is effective for achieving an approximate solution, but has limitations when an exact solution is required.

I. INTRODUCTION

In Model Predictive Control (MPC), control inputs are optimized by solving an open-loop optimal control problem over a finite prediction horizon, whilst explicitly accounting for constraints on system states and inputs. This approach is widely used in process control applications, in which slow plant dynamics allow the optimization to be solved in real time, and the control of mechanical and electrical systems is becoming more feasible with improvements in embedded controller hardware [1]. To further increase MPC applicability, recent research has considered exploiting properties of optimization problems to improve computational efficiency.

Active set [2] and interior point methods [3] are commonly employed in MPC, but for reduced computational complexity first order approaches such as fast gradient methods [4] and alternating direction method of multipliers (ADMM) [5], [6] have received recent attention. ADMM has been shown to be particularly effective in distributed applications (D-ADMM) [7], [8] where the separability of the cost function can be leveraged. This paper likewise exploits separability of the cost and constraints, although distributed optimization and control are not considered here. We refer the reader to [9] for a full description of ADMM and a survey of applications.

This paper is motivated by the problem of minimizing fuel consumption in a hybrid electric vehicle, achieved by distributing the demanded load between two available power sources (an internal combustion engine and an electric motor) in a blended mode [10]. The problem is constrained by limits on available power and bounds on the battery state of charge, and MPC is a suitable framework since its feedback mechanism provides a degree of robustness to discrepancies between the predicted and actual power demanded. Although dynamic programming is typically used for the solution of the corresponding optimization [11], [12],

the computation required for a sufficiently accurate solution prohibits a real-time implementation. A projected Newton method for solving this problem subject to a terminal state of charge constraint was proposed in [13], but the method becomes intractable for the general case of constraints on the state charge at each instant on a prediction horizon. Here, we use ADMM to leverage a separable cost function to enforce constraints at all future time-steps.

The paper is organised as follows: Section 2 gives the problem definition; Section 3 states the proposed ADMM algorithm; Section 4 describes the energy management problem more fully and discusses its solution for randomly generated examples; Section 5 provides conclusions. Convergence and optimality analyses are included in the appendix.

II. PROBLEM DEFINITION

Consider a system represented by the discrete time model

$$x_{k+1} = A_k x_k + b_k(u_k)$$

where $x_k \in \mathbb{R}^{n_x}$ is the system state, $u_k \in \mathbb{R}^{n_u}$ is the control input, $A_k \in \mathbb{R}^{n_x \times n_x}$, and $b_k(\cdot)$ is convex. The states and control inputs are subject to elementwise bounds and the associated optimal control problem is defined as

$$\begin{aligned} & \underset{x, u}{\text{minimize}} && f(x) + g(u) \\ & \text{subject to} && \left. \begin{aligned} x_{k+1} &= A_k x_k + b_k(u_k) \\ \underline{x}_{k+1} &\leq x_{k+1} \leq \bar{x}_{k+1} \\ \underline{u}_k &\leq u_k \leq \bar{u}_k \end{aligned} \right\} k = 0, \dots, N-1 \end{aligned} \quad (1)$$

where $x = (x_1, \dots, x_N)$ is the vector of future states over a prediction horizon of N steps, $u = (u_0, \dots, u_{N-1})$ is the vector of control inputs over the prediction horizon, and f, g are state and control cost functions. It is assumed that f and g are convex and separable. We say that a function $a(c)$, where $c = (c_1, \dots, c_{N_c})$, is separable if

$$a(c) = \sum_{i=1}^{N_c} a_i(c_i).$$

Note that problem (1) is in general nonconvex due to the presence of nonlinear dynamics and state constraints.

Equality constraints: x and x_k can be written in terms of $b(u) = [b_0^\top(u_0) \cdots b_{N-1}^\top(u_{N-1})]^\top$ as

$$\begin{aligned} x &= \Phi x_0 + \Psi b(u) \\ x_k &= \Phi_k x_0 + \Psi_k b(u) \end{aligned} \quad (2)$$

¹Department of Engineering Science, University of Oxford, OX1 3PJ, UK

where Φ_k and Ψ_k are the k th block rows of Φ and Ψ :

$$\Phi = \begin{bmatrix} A_0 & & & \\ \vdots & & & \\ \prod_{k=0}^{N-1} A_k & & & \end{bmatrix}, \quad \Psi = \begin{bmatrix} I & \dots & 0 \\ \vdots & \ddots & \vdots \\ \prod_{k=1}^{N-1} A_k & \dots & I \end{bmatrix}.$$

Here $\Psi \in \mathbb{R}^{Nn_x \times Nn_x}$ is lower-triangular with ij th block given by $\prod_{k=j}^{i-1} A_k$ for $j < i$ and by the identity matrix for $i = j$, and the i th block-row of $\Phi \in \mathbb{R}^{Nn_x \times n_x}$ is $\prod_{k=0}^{i-1} A_k$.

Inequality constraints: Define elementwise indicator functions h_k^x , h_k^u and h^x , h^u as

$$h_k^x(x_k) = \begin{cases} 0 & x_k \in [\underline{x}_k, \bar{x}_k] \\ \infty & \text{otherwise} \end{cases} \quad h^x(x) = \sum_{k=1}^N h_k^x(x_k),$$

$$h_k^u(u_k) = \begin{cases} 0 & u_k \in [\underline{u}_k, \bar{u}_k] \\ \infty & \text{otherwise} \end{cases} \quad h^u(u) = \sum_{k=0}^{N-1} h_k^u(u_k).$$

Using these definitions we rewrite (1) in a more convenient form without explicit inequality constraints as

$$\begin{aligned} & \underset{x,u}{\text{minimize}} && f(x) + g(u) + h^x(x) + h^u(u) \\ & \text{subject to} && x = \Phi x_0 + \Psi b(u). \end{aligned} \quad (3)$$

III. ADMM ALGORITHM

Introducing v as a substitute for $b(u)$, we rewrite (3) as

$$\begin{aligned} & \underset{u,x,v}{\text{minimize}} && f(x) + g(u) + h^x(x) + h^u(u) \\ & \text{subject to} && \Phi x_0 + \Psi v - x = 0 \\ & && b(u) - v = 0. \end{aligned} \quad (4)$$

The associated augmented Lagrangian function is

$$L(u, v, x, y, z) = f(x) + g(u) + h^x(x) + h^u(u) + \frac{\rho_1}{2} \|b(u) - v + y\|^2 + \frac{\rho_2}{2} \|\Phi x_0 + \Psi v - x + z\|^2, \quad (5)$$

and the ADMM iteration is obtained (see e.g. [9]) as

$$\begin{aligned} u_k^{j+1} &= \arg \min_{u_k} \left(h_k^u(u_k) + g_k(u_k) + \frac{\rho_1}{2} (b_k(u_k) - v^j + y_k^j)^2 \right) \\ v^{j+1} &= \arg \min_v \left(\frac{\rho_1}{2} \|b(u^{j+1}) - v + y^j\|^2 \right. \\ & \quad \left. + \frac{\rho_2}{2} \|\Phi x_0 + \Psi v - x^j + z^j\|^2 \right) \\ x_{k+1}^{j+1} &= \arg \min_{x_{k+1}} \left(h_k^x(x_{k+1}) + f_{k+1}(x_{k+1}) \right. \\ & \quad \left. + \frac{\rho_2}{2} (\Phi_{k+1} x_0 + \Psi_{k+1} v^{j+1} - x_{k+1} + z_k^j)^2 \right) \\ y^{j+1} &= y^j + b(u^{j+1}) - v^{j+1} \\ z^{j+1} &= z^j + \Phi x_0 + \Psi v^{j+1} - x^{j+1} \end{aligned}$$

for $k = 0, \dots, N-1$, where j is the iteration counter. We assume that a suitable solver is available for updating u^{j+1} and x^{j+1} . For the application considered in Section IV analytical solutions exist for the minimizers u_k^{j+1} and x_{k+1}^{j+1} , which are computed by finding the roots of cubic equations and projecting these onto the inequality constraints in (1).

The update for v can be written explicitly as

$$v^{j+1} = (\rho_1 I + \rho_2 \Psi^\top \Psi)^{-1} \left[\rho_1 (b(u^{j+1}) + y^j) + \rho_2 \Psi^\top (-\Phi x_0 + x^j - z^j) \right].$$

For the application considered in Section IV, where $A_k = 1$, the matrix Ψ is a lower triangular matrix of 1s, so $(\rho_1 I + \rho_2 \Psi^\top \Psi)^{-1}$ becomes a linear time-invariant filter. Approximating this matrix by setting the elements that do not exceed a given threshold to zero then yields a banded matrix, making the computation of v^{j+1} straightforward. For the general case in which the matrices A_k defining Ψ are arbitrary, the results in [4] and [14] provide conditions under which $(\rho_1 I + \rho_2 \Psi^\top \Psi)^{-1}$ is approximately banded and bound the rates of decay of elements with distance from the diagonal.

The iteration is terminated when primal and dual residual variables have fallen below pre-defined thresholds, ϵ^{primal} and ϵ^{dual} , chosen based on the required accuracy of the optimization and the typical magnitudes of decision variables:

$$\|r^{j+1}\|_2 \leq \epsilon^{\text{primal}}, \quad \|s^{j+1}\|_2 \leq \epsilon^{\text{dual}}.$$

In the appendix we analyse the optimality and convergence properties of the algorithm, and provide definitions of the residuals r^j and s^j . We also give conditions for the algorithm to converge to the global minimum, despite the problem being nonconvex, however this condition cannot generally be determined *a priori*, and the algorithm may converge to a local minimum, or even a maximum, if it is not met.

IV. APPLICATION TO PHEV SUPERVISORY CONTROL

We are interested in the problem of minimizing the fuel consumption of a plug-in hybrid electric vehicle (PHEV) by controlling the power balance between the electric motor and internal combustion engine in a blended mode. The power, d , demanded by the driver is split between the motor m and engine u so that $u + m = d$. The system can be modelled as

$$x_{k+1} = x_k + b_k(u_k)$$

where x_k is the state of charge of the battery, b_k accounts for the charge loss-dynamics, and the engine power u_k is the control input (i.e. $n_x = n_u = 1$ and $A_k = 1$). The fuel minimization problem over a horizon of N time steps is

$$\begin{aligned} & \underset{u}{\text{minimize}} && \sum_{k=0}^{N-1} g_k(u_k) \\ & \text{subject to} && \left. \begin{aligned} x_{k+1} &= x_k + b_k(u_k) \\ \underline{x}_{k+1} &\leq x_{k+1} \leq \bar{x}_{k+1} \\ \underline{u}_k &\leq u_k \leq \bar{u}_k \end{aligned} \right\} k = 0, \dots, N-1 \end{aligned} \quad (6)$$

where g_k is the fuel consumption determined from a quasi-static model at time-step k ; \bar{u}_k and \underline{u}_k represent upper and lower limits on the engine power output; and \underline{x}_{k+1} and \bar{x}_{k+1} represent the bounds on state of charge. The cost and input loss functions are represented by time-varying quadratic functions as

$$\begin{aligned} g_k(u_k) &= \alpha_{2,k} u_k^2 + \alpha_{1,k} u_k + \alpha_{0,k} \\ b_k(u_k) &= -\beta_{2,k} (d_k - u_k)^2 - \beta_{1,k} (d_k - u_k) - \beta_{0,k} \end{aligned}$$

See [13] for details on how these functions are obtained from the fuel map and electrical loss maps for a particular PHEV.

To demonstrate the performance of the ADMM algorithm without reference to a particular PHEV powertrain, we generate random sets of nominal systems of the same form as the motivating problem in (6). We generate uniformly distributed random disturbance values $d_k \in [-1, 1]$; uniformly distributed α and β coefficients $\alpha_{2,k}, \beta_{2,k} \in [0, 0.1]$, $\alpha_{1,k}, \beta_{1,k} \in [0, 1]$, and $\alpha_{0,k}, \beta_{0,k} = 0$, and a uniformly distributed initial state $x_0 \in [-0.5, 0.5]$. We use $\underline{u}_k = -0.5$ and $\bar{u}_k = 0.5$ as control input limits, and state limits of $\underline{x}_k = -2$ and $\bar{x}_k = 2$. The ADMM iteration is initialised assuming that the inequality constraints in (6) are inactive:

$$u_k^0 = \pi_k^u \left[\arg \min_{u_k} g_k(u_k) \right], \quad v^0 = b(u^0),$$

$$x_{k+1}^0 = \pi_{k+1}^x \left[\Phi_{k+1} x_0 + \Psi_{k+1} v^0 \right], \quad y^0 = 0, \quad z^0 = 0,$$

for $k = 0, \dots, N-1$, where π_k^u, π_k^x are projection operators:

$$\begin{aligned} \pi_k^u(u_k) &= \min \{ \bar{u}_k, \max \{ \underline{u}_k, u_k \} \} \\ \pi_k^x(x_k) &= \min \{ \bar{x}_k, \max \{ \underline{x}_k, x_k \} \}. \end{aligned}$$

If $\Phi x_0 + \Psi v^0 = x^0$, then these initial values are optimal and the state constraints are inactive. The stopping conditions used are equal i.e $\epsilon^{\text{primal}} = \epsilon^{\text{dual}} = \epsilon$. The simulations are run in Matlab, using an Intel 2.60 GHz i7-6700HQ CPU.

A. Variation of iteration count with ρ

The rate of convergence of the ADMM algorithm is affected by the values of ρ_1 and ρ_2 . For the algorithm to be practically useful, it is desirable for the optimal values for these parameters to be independent of both horizon length and system parameters, so that a single set of values can be used in all conditions. Here we demonstrate the effect of ρ_1 and ρ_2 on the convergence properties of the algorithm for 20 randomly generated systems. Figure 1 shows the average numbers of iterations required for completion of the algorithm for horizon lengths of 50, 100, 200, and 400 with $\epsilon = 0.001$, $10^{-1} \leq \rho_1 \leq 10^2$ and $10^{-2} \leq \rho_2 \leq 10$.

For each horizon length, the minimum number of iterations for convergence is obtained with $1 \leq \rho_1 \leq 10$ and $0.1 \leq \rho_2 \leq 1$. Moreover, variations in ρ within an order of magnitude of this minimum do not produce large increases in total number of iterations, implying a degree of robustness. Therefore ρ values within these ranges provide close to the minimum number of iterations for horizon lengths between 50 and 400, suggesting that a single set of values is appropriate for the PHEV problem. Although the optimal ρ values are similar, the minimum number of iterations for completion increases significantly with horizon length. This is explored further in the following sections.

B. Variation of required number of iterations with ϵ

The threshold ϵ determines the degree of optimality at termination, since $(u, x) \rightarrow (u^*, x^*)$ as $\epsilon \rightarrow 0$ (see Appendix A). However, a trade-off exists between accuracy and computation since small values of ϵ require many iterations. Here we investigate how the number of iterations before termination varies with ϵ for 200 randomly generated problems with

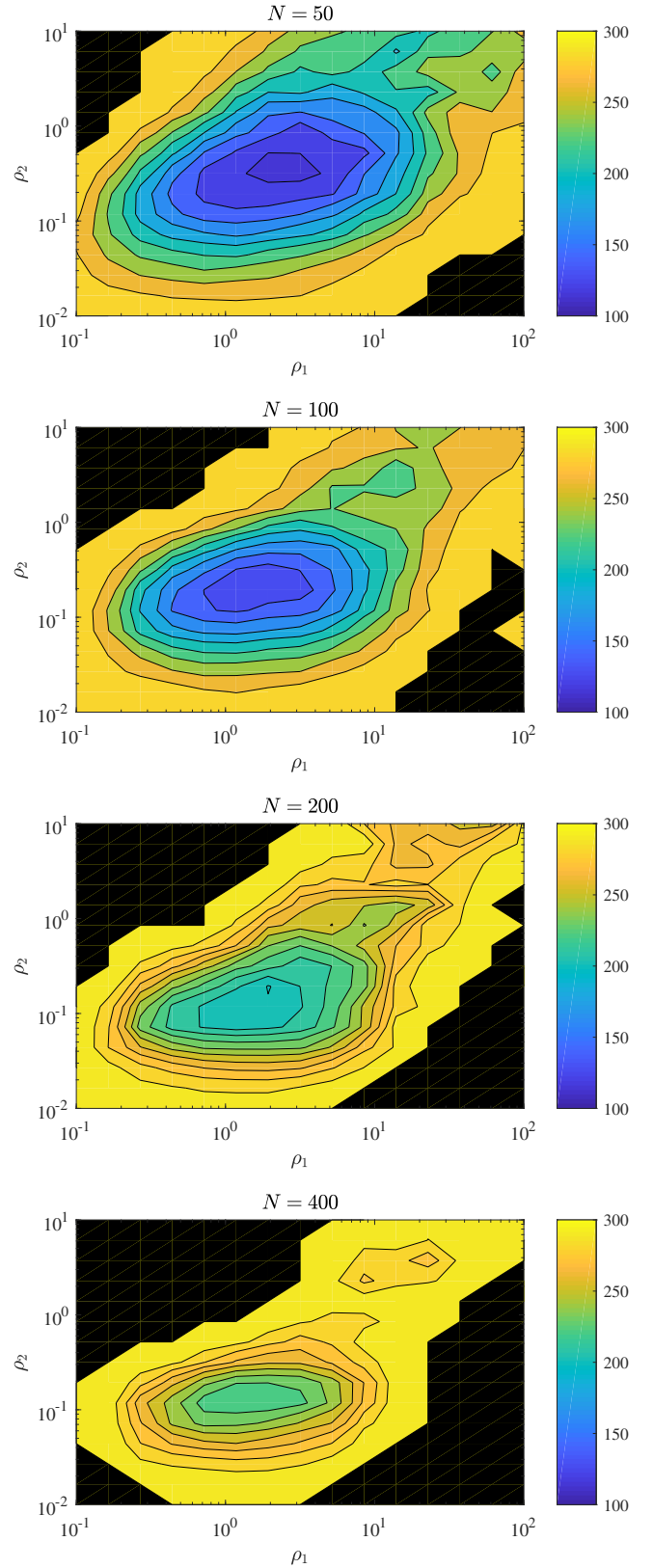


Fig. 1. Variation with ρ_1, ρ_2 of the number of iterations before termination. The data presented is the arithmetic mean for 20 generated systems.

$N = 100$. Figure 2 shows the number of iterations required with $10^{-4} \leq \epsilon \leq 10^0$, $\rho_1 = 1$, and $\rho_2 = 0.2$.

Clearly, for a low-accuracy solution only a few tens of iterations are required, but this rises rapidly for high-accuracy

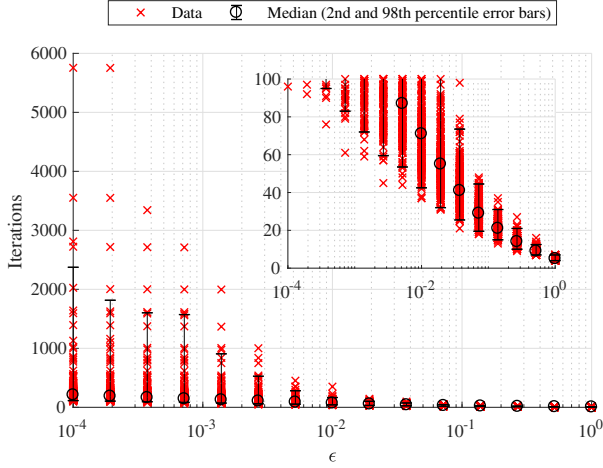


Fig. 2. Number of iterations required for completion as ϵ varies.

solutions. For example, with $\epsilon = 0.14$ the median iteration count is 29, but $\epsilon = 10^{-4}$ requires a median iteration count of 210. Furthermore, the increase in iteration count is highly nonlinear as ϵ is reduced, and for small ϵ modest further reductions have a large effect on the iteration count.

The variation in the number of iterations required increases as ϵ is reduced (Figure 2): the 98th percentile is 45 for $\epsilon = 0.14$ (~ 2 times the median), rising to 2374 (~ 10 times the median) for $\epsilon = 10^{-4}$. Again, this increase in uncertainty increases nonlinearly as ϵ is reduced. Furthermore, the distribution of the data is heavily skewed, and outliers become more extreme as ϵ is reduced. This is problematic as the MPC optimization algorithm must be designed for the worst case, and at low values of ϵ such a high potential iteration count could be extremely prohibitive.

These properties suggest that the ADMM algorithm may be appropriate for finding an approximate solution to the PHEV problem that is then used to initialise another method, but inappropriate for obtaining an exact solution. For example, the ADMM algorithm could be used to obtain the active set, then an alternative method (such as that proposed in [13]) used to obtain the corresponding optimum. This observation is application-specific as large numbers of iterations may not be problematic if the system dynamics are sufficiently slow or computational resources are sufficiently high.

C. Variation of computation with horizon length

The MPC framework can be implemented in the PHEV energy management problem with either a receding horizon (with fixed N) or a shrinking horizon (with N reducing as the vehicle progresses through a journey). Although the algorithm can be tuned for a specific horizon length in the receding horizon case, for a shrinking horizon implementation it must be robust to a range of horizon lengths. Thus, for a 1 hour journey sampled at 1 s, N varies between 1 and 3600. Here we demonstrate the effect of horizon length on number of iterations and time to termination for 200 systems, with N varying between 10 and 400. We choose $\epsilon = 10^{-2}$, and $\rho_1 = 1$, $\rho_2 = 0.2$. The results are shown in Figure 3.

For small N , increases in horizon length produce large increases in the number of iterations (the median for $N = 25$

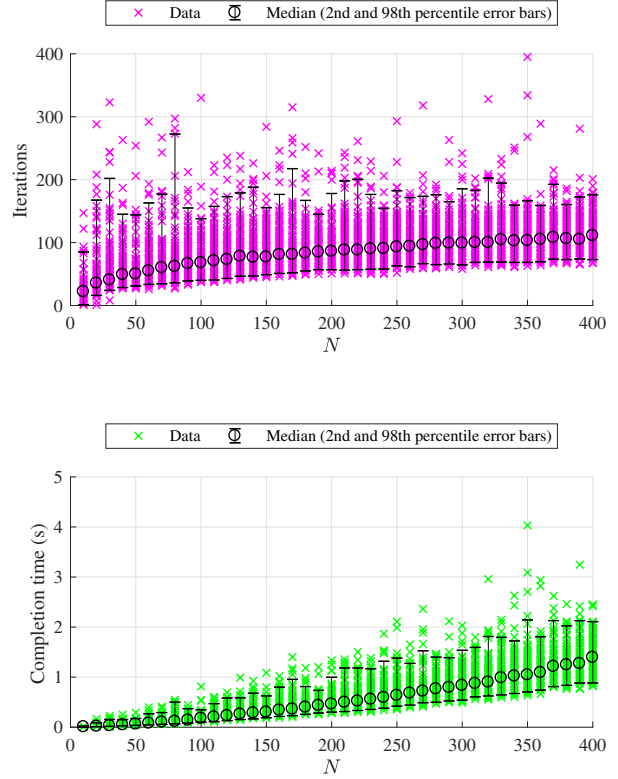


Fig. 3. Iterations and time required for completion against N .

is 22, increasing to 40 when $N = 75$), but this effect tapers off quickly, and the median iteration count is 105 for $N = 350$. There is no clear dependence on N of the uncertainty in the required number of iterations; the 98th percentile has a maximum value of 273 iterations at $N = 80$, a minimum of 85 at $N = 10$, and large variations between these values.

Although the iteration count does not increase significantly with longer horizons, the total time taken to find the solution increases roughly linearly with horizon length, from a median of 0.0075 s for $N = 10$, to a median of 1.39 s for $N = 400$. The uncertainty also appears to increase within a near-linear envelope. This is due to the computation required to update u^{j+1} , for which N cubic equations must be solved at each iteration. This suggests that for long prediction horizons, improving the solution times for u^{j+1} and x^{j+1} would have a significant effect on total solution time. Note that the absolute computation times presented here demonstrate the relationship between horizon length and computation, but the hardware used for the simulations does not correlate to that typically found in a PHEV, and Matlab is not typically used for embedded computing.

V. CONCLUSIONS

A general MPC optimization problem is considered for systems with separable cost functions, nonlinear input map and simple constraints, motivated by the blended mode energy management problem for PHEVs. We propose an ADMM algorithm for solving this problem and demonstrate conditions for convergence to a solution satisfying first order necessary conditions for optimality. Numerical experiments show that an approximate solution is reached within a few

tens of iterations, while significantly more are required for an exact solution. The effect of algorithm parameters on the number of iterations and solution times is also investigated. Future work will investigate methods for implementing the proposed algorithm in a PHEV, to obtain globally optimal control inputs that are implementable in real time.

APPENDIX

A. Optimality

We use a similar method to that given in Section 3 of [9], however we make use of an alternative Lyapunov function to demonstrate convergence, as required by the nonconvexity of the problem. Problem (4) is first rewritten as

$$\begin{aligned} & \underset{u, \hat{x}}{\text{minimize}} && \hat{f}(\hat{x}) + \hat{g}(u) \\ & \text{subject to} && \hat{b}(u) + B\hat{x} = 0 \end{aligned} \quad (7)$$

where

$$\begin{aligned} \hat{f}(\hat{x}) &= f(x) + h^x(x), \quad \hat{g}(u) = g(u) + h^u(u) \\ \hat{b}(u) &= \begin{bmatrix} b(u) \\ \Phi x_0 \end{bmatrix}, \quad B = \begin{bmatrix} -I & 0 \\ \Psi & -I \end{bmatrix}, \quad \hat{x} = \begin{bmatrix} v \\ x \end{bmatrix}. \end{aligned}$$

Then, by expressing the augmented Lagrangian (5) as

$$L(u, \hat{x}, \hat{y}) = \hat{f}(\hat{x}) + \hat{g}(u) + \hat{y}^\top (\hat{b}(u) + B\hat{x}) + \frac{1}{2} \|\hat{b}(u) + B\hat{x}\|_R^2$$

where $\|w\|_R^2 = w^\top R w$ for any vector w of conformal dimensions, and defining $\hat{y} = [\hat{y}^\top \ z^\top]^\top$, $R = \text{diag}\{\rho_1 I, \rho_2 I\}$, the ADMM iteration becomes

$$\begin{aligned} u^{j+1} &= \arg \min_u \left\{ \hat{g}(u) + (\hat{y}^j)^\top (\hat{b}(u) + B\hat{x}^j) \right. \\ & \quad \left. + \frac{1}{2} \|\hat{b}(u) + B\hat{x}^j\|_R^2 \right\} \end{aligned} \quad (8a)$$

$$\begin{aligned} \hat{x}^{j+1} &= \arg \min_{\hat{x}} \left\{ \hat{f}(\hat{x}) + (\hat{y}^j)^\top (\hat{b}(u^{j+1}) + B\hat{x}) \right. \\ & \quad \left. + \frac{1}{2} \|\hat{b}(u^{j+1}) + B\hat{x}\|_R^2 \right\} \end{aligned} \quad (8b)$$

$$\hat{y}^{j+1} = \hat{y}^j + R(\hat{b}(u^{j+1}) + B\hat{x}^{j+1}) \quad (8c)$$

Assumption 1: Slater's condition holds, i.e. u, v, x exist such that $\underline{u}_k < u_k < \bar{u}_k$, $\underline{x}_{k+1} < x_{k+1} < \bar{x}_{k+1}$ for all k , where $x = \Phi x_0 + \Psi v$ and $v = b(u)$.

The first order conditions satisfied by a (possibly local, possibly maximal) optimal solution $(u^*, \hat{x}^*, \hat{y}^*)$ are

$$0 = \hat{b}(u^*) + B\hat{x}^* \quad (9a)$$

$$0 = \nabla \hat{g}(u^*) + [\partial \hat{b}(u^*)]^\top \hat{y}^* \quad (9b)$$

$$0 = \nabla \hat{f}(\hat{x}^*) + B^\top \hat{y}^* \quad (9c)$$

where $\nabla \hat{g}$, $\nabla \hat{f}$ are the gradients of \hat{g} , \hat{f} , and $\partial \hat{b}$ is the Jacobian matrix of \hat{b} .

Lemma 1: $(u^{j+1}, \hat{x}^{j+1}, \hat{y}^{j+1})$ satisfies the first order necessary optimality conditions (9) iff $r^{j+1} = 0$ and $s^{j+1} = 0$.

Proof: First consider condition (9c). Since $\hat{x} = \hat{x}^{j+1}$ is by definition the minimizer of $L(u^{j+1}, \hat{x}, \hat{y}^j)$, we have

$$\nabla \hat{f}(\hat{x}^{j+1}) + B^\top \hat{y}^j + B^\top R(\hat{b}(u^{j+1}) + B\hat{x}^{j+1}) = 0$$

and hence the update law $\hat{y}^{j+1} = \hat{y}^j + R(\hat{b}(u^{j+1}) + B\hat{x}^{j+1})$ implies $\nabla \hat{f}(\hat{x}^{j+1}) + B^\top \hat{y}^{j+1} = 0$ for all j . Next, consider

(9b). By definition $u = u^{j+1}$ minimizes $L(u, \hat{x}^j, \hat{y}^j)$ and therefore u^{j+1} necessarily satisfies

$$\nabla \hat{g}(u^{j+1}) + [\partial \hat{b}(u^{j+1})]^\top (\hat{y}^j + R(\hat{b}(u^{j+1}) + B\hat{x}^j)) = 0$$

Hence the update law for \hat{y}^{j+1} gives

$$\nabla \hat{g}(u^{j+1}) + [\partial \hat{b}(u^{j+1})]^\top (\hat{y}^{j+1} + RB(\hat{x}^j - \hat{x}^{j+1})) = 0$$

and we have, for all j ,

$$\nabla \hat{g}(u^{j+1}) + [\partial \hat{b}(u^{j+1})]^\top \hat{y}^{j+1} + s^{j+1} = 0$$

Therefore, the ADMM iteration converges to a point satisfying the first order necessary conditions (9a)-(9c) if and only if the residual variables defined by

$$\begin{aligned} r^{j+1} &= \hat{b}(u^{j+1}) + B\hat{x}^{j+1} \\ s^{j+1} &= [\partial \hat{b}(u^{j+1})]^\top RB(\hat{x}^j - \hat{x}^{j+1}) \end{aligned}$$

converge to zero. ■

B. Convergence

Let $(u^\dagger, \hat{x}^\dagger, \hat{y}^\dagger)$ denote a solution of (7) that achieves the globally minimum value of the objective. Let $L_0(u, \hat{x}, \hat{y}) = \hat{f}(\hat{x}) + \hat{g}(u) + \hat{y}^\top (\hat{b}(u) + B\hat{x})$ and define \mathcal{S}_c as the set

$$\mathcal{S}_c = \{(u, \hat{x}, \hat{y}) : \|\hat{y} - \hat{y}^\dagger\|_{R^{-1}}^2 + \|B(\hat{x} - \hat{x}^\dagger)\|_R^2 + \|r(u, \hat{x})\|_R^2 \leq c\}$$

for $c > 0$, where $r(u, \hat{x}) = \hat{b}(u) + B\hat{x}$.

Assumption 2: \mathcal{S}_c , $(u^\dagger, \hat{x}^\dagger, \hat{y}^\dagger)$, $(u^j, \hat{x}^j, \hat{y}^j)$ are such that:

- iteration (8) is initialised at $(u^0, \hat{x}^0, \hat{y}^0) \in \mathcal{S}_c$, i.e. $c \geq \|\hat{y}^0 - \hat{y}^\dagger\|_{R^{-1}}^2 + \|B(\hat{x}^0 - \hat{x}^\dagger)\|_R^2 + \|r(u^0, \hat{x}^0)\|_R^2$,
- $u = u^{j+1}$ in (8a) is the global minimiser of $L(u, \hat{x}^j, \hat{y}^j)$.

Theorem 1: $(u^j, \hat{x}^j, \hat{y}^j) \in \mathcal{S}_c$ for all $j \in \mathbb{Z}^+$, and $r^j \rightarrow 0$ and $s^j \rightarrow 0$ as $j \rightarrow \infty$.

Proof: We first show that $V^j = V(u^j, \hat{x}^j, \hat{y}^j)$ is a Lyapunov function, where V is defined by

$$V(u, \hat{x}, \hat{y}) = \|\hat{y} - \hat{y}^\dagger\|_{R^{-1}}^2 + \|B(\hat{x} - \hat{x}^\dagger)\|_R^2 + \|r(u, \hat{x})\|_R^2.$$

This part of the proof is split into three steps:

- The update law for u^{j+1} can be equivalently written as

$$u^{j+1} = \arg \min_u \left\{ \hat{g}(u) + (\hat{y}^{j+1} + RB(\hat{x}^j - \hat{x}^{j+1}))^\top \hat{b}(u) \right\}$$

and, from Assumption 3(b) it follows that

$$\begin{aligned} \hat{g}(u^{j+1}) + (\hat{y}^{j+1} + RB(\hat{x}^j - \hat{x}^{j+1}))^\top \hat{b}(u^{j+1}) \\ \leq \hat{g}(u^\dagger) + (\hat{y}^{j+1} + RB(\hat{x}^j - \hat{x}^{j+1}))^\top \hat{b}(u^\dagger). \end{aligned}$$

Similarly, $\nabla \hat{f}(\hat{x}^{j+1}) + B^\top \hat{y}^{j+1} = 0$ implies that

$$\hat{f}(\hat{x}^{j+1}) + (\hat{y}^{j+1})^\top B\hat{x}^{j+1} \leq \hat{f}(\hat{x}^\dagger) + (\hat{y}^{j+1})^\top B\hat{x}^\dagger.$$

Therefore, defining $p^{j+1} = \hat{f}(\hat{x}^{j+1}) + \hat{g}(u^{j+1})$ and $p^\dagger = \hat{f}(\hat{x}^\dagger) + \hat{g}(u^\dagger)$, we obtain

$$\begin{aligned} p^{j+1} - p^\dagger &= \hat{f}(\hat{x}^{j+1}) - \hat{f}(\hat{x}^\dagger) + \hat{g}(u^{j+1}) - \hat{g}(u^\dagger) \\ &\leq (\hat{y}^{j+1})^\top (B(\hat{x}^\dagger - \hat{x}^{j+1}) + \hat{b}(u^\dagger) - \hat{b}(u^{j+1})) \\ &\quad + (\hat{x}^j - \hat{x}^{j+1})^\top B^\top R(\hat{b}(u^\dagger) - \hat{b}(u^{j+1})) \\ &= -(\hat{y}^{j+1})^\top r^{j+1} \\ &\quad + (\hat{x}^j - \hat{x}^{j+1})^\top B^\top R(-r^{j+1} + B(\hat{x}^{j+1} - \hat{x}^\dagger)) \end{aligned}$$

ii) Since $(u^\dagger, \hat{x}^\dagger, \hat{y}^\dagger)$ must satisfy the first order necessary conditions (9), it follows that $u = u^\dagger$ is the minimiser of $\hat{g}(u) + (\hat{y}^\dagger)^\top \hat{b}(u)$ and $\hat{x} = \hat{x}^\dagger$ is the minimiser of $\hat{f}(\hat{x}) + (\hat{y}^\dagger)^\top B\hat{x}$. Therefore, we have, for all u and \hat{x} ,

$$\begin{aligned}\hat{g}(u^\dagger) + (\hat{y}^\dagger)^\top \hat{b}(u^\dagger) &\leq \hat{g}(u) + (\hat{y}^\dagger)^\top \hat{b}(u) \\ \hat{f}(\hat{x}^\dagger) + (\hat{y}^\dagger)^\top B\hat{x}^\dagger &\leq \hat{f}(\hat{x}) + (\hat{y}^\dagger)^\top B\hat{x},\end{aligned}$$

so that $\hat{f}(\hat{x}^\dagger) + \hat{g}(u^\dagger) \leq \hat{f}(\hat{x}) + \hat{g}(u) + (\hat{y}^\dagger)^\top (B\hat{x} + \hat{b}(u))$ and hence $p^\dagger - p^{j+1} \leq (\hat{y}^\dagger)^\top r^{j+1}$.

iii) Combining the bounds on $p^\dagger - p^{j+1}$ in i) and ii) yields

$$\begin{aligned}(\hat{y}^{j+1} - \hat{y}^\dagger)^\top r^{j+1} - (\hat{x}^{j+1} - \hat{x}^j)^\top B^\top Rr^{j+1} \\ + (\hat{x}^{j+1} - \hat{x}^j)^\top B^\top RB(\hat{x}^{j+1} - \hat{x}^\dagger) \leq 0.\end{aligned}\quad (10)$$

The first term in (10) can be simplified using the update law $\hat{y}^{j+1} = \hat{y}^j + Rr^{j+1}$ and completing the square:

$$\begin{aligned}2(\hat{y}^{j+1} - \hat{y}^\dagger)^\top r^{j+1} &= 2(\hat{y}^j - \hat{y}^\dagger)^\top r^{j+1} + 2\|r^{j+1}\|_R^2 \\ &= \|\hat{y}^j - \hat{y}^\dagger + Rr^{j+1}\|_{R^{-1}}^2 - \|\hat{y}^j - \hat{y}^\dagger\|_{R^{-1}}^2 + \|r^{j+1}\|_R^2 \\ &= \|\hat{y}^j - \hat{y}^\dagger + \hat{y}^{j+1} - \hat{y}^j\|_{R^{-1}}^2 - \|\hat{y}^j - \hat{y}^\dagger\|_{R^{-1}}^2 + \|r^{j+1}\|_R^2 \\ &= \|\hat{y}^{j+1} - \hat{y}^\dagger\|_{R^{-1}}^2 - \|\hat{y}^j - \hat{y}^\dagger\|_{R^{-1}}^2 + \|r^{j+1}\|_R^2.\end{aligned}$$

To simplify the 2nd and 3rd terms in (10) we use

$$\begin{aligned}\|r^{j+1}\|_R^2 - 2(\hat{x}^{j+1} - \hat{x}^j)^\top B^\top Rr^{j+1} \\ + 2(\hat{x}^{j+1} - \hat{x}^j)^\top B^\top RB(\hat{x}^{j+1} - \hat{x}^\dagger) \\ = \|r^{j+1} - B(\hat{x}^{j+1} - \hat{x}^j)\|_R^2 \\ + \|B(\hat{x}^{j+1} - \hat{x}^\dagger)\|_R^2 - \|B(\hat{x}^j - \hat{x}^\dagger)\|_R^2.\end{aligned}$$

Therefore (10) is equivalent to

$$\begin{aligned}\|\hat{y}^{j+1} - \hat{y}^\dagger\|_{R^{-1}}^2 - \|\hat{y}^j - \hat{y}^\dagger\|_{R^{-1}}^2 \\ + \|r^{j+1} - B(\hat{x}^{j+1} - \hat{x}^j)\|_R^2 + \|B(\hat{x}^{j+1} - \hat{x}^\dagger)\|_R^2 \\ - \|B(\hat{x}^j - \hat{x}^\dagger)\|_R^2 \leq 0\end{aligned}$$

which, using the definition of $V^j = V(u^j, \hat{x}^j, \hat{y}^j)$, is equivalent to

$$\begin{aligned}V^{j+1} - V^j &\leq -\|r^{j+1} - B(\hat{x}^{j+1} - \hat{x}^j)\|_R^2 \\ &\quad - \|r^j\|_R^2 + \|r^{j+1}\|_R^2 \\ &= -\|r^j\|_R^2 - \|B(\hat{x}^{j+1} - \hat{x}^j)\|_R^2 \\ &\quad + 2(r^{j+1})^\top RB(\hat{x}^{j+1} - \hat{x}^j) \\ &= -\|r^j\|_R^2 - \|B(\hat{x}^{j+1} - \hat{x}^j)\|_R^2 \\ &\quad + 2(\hat{y}^{j+1} - \hat{y}^j)^\top B(\hat{x}^{j+1} - \hat{x}^j).\end{aligned}$$

To bound the RHS of this expression, we recall from i) that $\hat{x} = \hat{x}^j$ and $\hat{x} = \hat{x}^{j+1}$ are the minimizers of $\hat{f}(\hat{x}) + (\hat{y}^j)^\top B\hat{x}$ and $\hat{f}(\hat{x}) + (\hat{y}^{j+1})^\top B\hat{x}$, so that

$$\hat{f}(\hat{x}^j) + (\hat{y}^j)^\top B\hat{x}^j \leq \hat{f}(\hat{x}^{j+1}) + (\hat{y}^j)^\top B\hat{x}^{j+1}$$

and

$$\hat{f}(\hat{x}^{j+1}) + (\hat{y}^{j+1})^\top B\hat{x}^{j+1} \leq \hat{f}(\hat{x}^j) + (\hat{y}^{j+1})^\top B\hat{x}^j$$

By summing these two inequalities we obtain $(\hat{y}^{j+1} - \hat{y}^j)^\top B(\hat{x}^{j+1} - \hat{x}^j) \leq 0$, and it follows that

$$V^{j+1} - V^j \leq -\|r^j\|_R^2 - \|B(\hat{x}^{j+1} - \hat{x}^j)\|_R^2.$$

The positive invariance of \mathcal{S}_c follows from its definition as a level set of $V(u, \hat{x}, \hat{y})$ and from the fact that V^j is monotonically non-increasing with j . Asymptotic convergence of the residuals r^j and s^j can be established by summing both sides of the inequality satisfied by $V^{j+1} - V^j$ over all $j \geq 0$:

$$\sum_{j=0}^{\infty} (\|r^j\|_R^2 + \|B(\hat{x}^{j+1} - \hat{x}^j)\|_R^2) \leq V^0.$$

From this bound we can conclude that, as $j \rightarrow \infty$, $r^j \rightarrow 0$ and $B(\hat{x}^{j+1} - \hat{x}^j) \rightarrow 0$, and hence also $s^j \rightarrow 0$. ■

Lemma 1 and Theorem 1 imply that the ADMM iteration converges to a point satisfying the first order necessary optimality conditions, and this will be the global minimum if the problem is convex. The invariant nature of \mathcal{S}_c also implies that even when the problem is nonconvex, the algorithm will converge to the global minimum if it is initialised within a sub-level set of V that contains no local minima or maxima. This condition is, however, impossible to demonstrate *a priori*, and if the algorithm is initialised outside of this set it could converge to a local minimum or even a local maximum point. We note finally that the Theorem 1 also demonstrates that the iteration necessarily terminates after a finite number of steps whenever the threshold ϵ on r^{j+1} and s^{j+1} is set to a fixed non-zero value.

REFERENCES

- [1] J.H. Lee. Model predictive control: Review of the three decades of development. *Int. J. Control Autom. Sys.*, 9(3):415–424, 2011.
- [2] H. Ferreau, G. Bock, and M. Diehl. An online active set strategy to overcome the limitations of explicit MPC. *Int. J. Robust Nonlin. Control*, 18:816–830, 2008.
- [3] A. Domahidi, A.U. Zraggen, M.N. Zeilinger, M. Morari, and C.N. Jones. Efficient interior point methods for multistage problems arising in receding horizon control. In *51st IEEE Conf. Decision Control*, pages 668–674, 2012.
- [4] E. Klintberg and S. Gros. Approximate inverses in preconditioned fast dual gradient methods for MPC. In *20th IFAC World Congress*, pages 6075–6080, 2017.
- [5] T.V. Dang, K.V. Ling, and J.M. Maciejowski. Banded null basis and ADMM for embedded MPC. In *20th IFAC World Congress*, pages 13712–13717, 2017.
- [6] T.V. Dang, K.V. Ling, and J.M. Maciejowski. Embedded ADMM-based QP solver for MPC with polytopic constraints. In *European Control Conference*, pages 3446–3451, 2015.
- [7] J.F.C. Mota, J.M.F. Xavier, P.M.Q. Aguiar, and M. Puschel. D-ADMM: A communication-efficient distributed algorithm for separable optimization. *IEEE Trans. Sig. Proc.*, 61(10):2718–2723, 2013.
- [8] R.P. Costa, J.M. Lemos, J.F.C. Mota, and J.M.F. Xavier. D-ADMM based distributed MPC with input-output models. In *IEEE Conf. Control Appl.*, pages 699–704, 2014.
- [9] S.P. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(11):1–122, 2011.
- [10] A. Sciarretta and L. Guzzella. Control of hybrid electric vehicles. *IEEE Control Systems Magazine*, 27(2):60–70, 2007.
- [11] M. Back, M. Simons, F. Kirschaum, and V. Krebs. Predictive control of drivetrains. In *15th IFAC World Congress*, pages 241–246, 2002.
- [12] V. Larsson, L. Johannesson, and B. Egardt. Analytic solutions to the dynamic programming sub-problem in hybrid vehicle energy management. *IEEE Trans. Veh. Tech.*, 64(4):1458–1467, 2015.
- [13] J. Buerger, S. East, and M. Cannon. Fast dual-loop nonlinear receding horizon control for energy management in hybrid electric vehicles. *IEEE Trans. Control Sys. Tech.*, PP:1–11, 2018.
- [14] S. Demko, W.F. Moss, and P.W. Smith. Decay rates for inverses of band matrices. *Math. Comput.*, 43(168):491–499, 1984.