A Parametric Multi-Convex Splitting Technique with Application to Real-Time NMPC

Jean-Hubert Hours   Colin N. Jones

***Abstract*— A novel splitting scheme to solve parametric multi-convex programs is presented. It consists of a fixed number of proximal alternating minimisations and a dual update per time step, which makes it attractive in a real-time *Nonlinear Model Predictive Control* (NMPC) framework and for distributed computing environments. Assuming that the parametric program is semi-algebraic and that its critical points are strongly regular, a contraction estimate is derived and it is proven that the sub-optimality error remains stable under some mild assumptions. Efficacy of the method is demonstrated by solving a bilinear NMPC problem to control a DC motor. In particular, the effect of the sampling period on the optimality tracking error is analysed for a fixed computational power.**

I. Introduction

The applicability of NMPC to fast and complex dynamics is hampered by the fact that a nonlinear program (NLP), which is generally non-convex, is to be solved at every sampling time. Solving an NLP to full accuracy is not tractable when the system's sampling frequency is high, which is the case for many mechanical or electrical systems. This difficulty is enhanced when dealing with distributed systems, as they typically lead to large-scale NLPs. Several techniques have been proposed in order to improve the computational efficacy of NMPC schemes by avoiding solving with more accuracy than needed. Most of them rely on the parametric nature of the NLP [17,16,10]. Most of the existing approaches to real-time NMPC are based on Newton type methods, which benefit from local quadratic convergence, but are not easily applicable in a distributed context.

In this paper, a parametric optimisation scheme based on augmented Lagrangian techniques [7,3] is proposed. In an NMPC context, such an alternative has already been explored in [16]. In [16], the theoretical analysis relies on the fact that the primal quadratic program is solved to a given accuracy and the influence of the number of iterations of the suggested projected successive over-relaxation (PSOR) method on the sub-optimality error is not examined. Moreover, the efficacy of the proposed algorithm strongly relies on the fact that the current iterate is close to the optimal solution, as this is required to guarantee convexity of the quadratic program. Finally, due to the dual update, the tracking error is only first-order in the parameter difference. Therefore, the augmented Lagrangian approach may not be very competitive as a fast local method for NMPC, compared to Newton strategies. Yet it is an interesting direction for parallel computing environments or in a distributed NMPC context, assuming that one is able to decompose the evaluation of the primal iterates, as shown in [6] for convex problems.

The main idea of our algorithm is to apply a truncated version of the proximal alternating minimisation method [2] to solve the primal program approximately. Alternating minimisation strategies are known to lead to 'easily' solvable sub-problems, which can potentially be parallelised and are well-suited to distributed computing platforms [4]. At the expense of a few assumptions on the parametric program, we provide an analysis of the stability of the sub-optimality error. In particular, we give new insights on how the penalty parameter and the number of primal iterations should be tuned in order to ensure boundedness of the tracking error over time. The proposed framework can also address more general problem formulations than [16], where the PSOR strategy is restricted to quadratic objectives subject to non-negativity constraints.

In Section III, the parametric optimisation scheme is presented. In Section IV, some key theoretical tools are introduced. Then, in Section V, a contraction estimate for the primal-dual iterates is derived, ensuring stability of the optimality tracking error. Finally, in Section VI, the conditions derived in Section V are verified on a numerical example, which consists in controlling the speed of a DC motor to track a piecewise constant reference. An analysis of the evolution of the tracking error as a function of the sampling period for a fixed computational power is also presented.

II. Background definitions

Definition 1 (Critical point): Let f be a proper lower semicontinuous function. A necessary condition for x^* to be a minimiser of f is that

$$0 \in \partial f(x^*) \ , \tag{1}$$

where $\partial f(x^*)$ is the sub-differential of f at x^* [14]. Points satisfying (1) are called *critical points*.

Definition 2 (Normal cone to a convex set): Let Ω be a convex set in $\mathbb{R}^n$ and $\bar{x} \in \Omega$. The *normal cone* to Ω at $\bar{x}$ is the set

$$\mathcal{N}_\Omega(\bar{x}) := \left\{ v \in \mathbb{R}^n \;\middle|\; \forall x \in \Omega,\ v^\top (x - \bar{x}) \leq 0 \right\} \ . \tag{2}$$

The indicator function of a closed subset Ω of $\mathbb{R}^n$ is denoted by ι_Ω and is defined as

$$\iota_\Omega(x) = \begin{cases} 0 & \text{if } x \in \Omega \\ +\infty & \text{if } x \notin \Omega \ . \end{cases} \tag{3}$$

Jean-Hubert Hours and Colin N. Jones are with the Laboratoire d'Automatique, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne (e-mail:`jean-hubert.hours, colin.jones@epfl.ch`). The work of Jean-Hubert Hours is supported by the Swiss National Science Foundation (SNSF) under grant 200021-135218. The work of Colin N. Jones is supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013), ERC Grant Agreement 307608.

Lemma 1 (Sub-differential of indicator function [14]): Given a convex set Ω, for all $x \in \Omega$,

$$\partial\iota_{\Omega}(x) = \mathcal{N}_{\Omega}(x) \ . \quad (4)$$

The distance of a point $x \in \mathbb{R}^n$ to a subset Σ of $\mathbb{R}^n$ is defined by

$$d(x, \Sigma) := \inf_{y\in\Sigma} \left\|x - y\right\|_2 \ . \quad (5)$$

A function $h : \left(z_1, \ldots, z_P\right) \mapsto h\left(z_1, \ldots, z_P\right)$ is said to be *multi-convex* if for all $i \in \{1, \ldots, P\}$, by fixing variables z_j with $j \neq i$, the resulting function is convex in z_i. The open ball with center x and radius r is denoted by $\mathcal{B}\left(x, r\right)$.

III. Solving time-dependent multi-convex parametric programs

A. Problem formulation

We consider multi-convex parametric programs

$$\begin{aligned} &\text{minimise} \ \ f(z_1, \ldots, z_P) \\ &\text{s.t.} \quad g(z_1, \ldots, z_P, s_k) = 0 \\ &\qquad z_i \in \mathcal{Z}_i, \ \forall i \in \{1, \ldots, P\} \ , \end{aligned} \quad (6)$$

where f is multi-convex in $z := (z_1^\top, \ldots, z_P^\top)^\top \in \mathbb{R}^{n_z}$ with $z_i \in \mathbb{R}^{n_i}$ and $n_z := \sum_{i=1}^P n_i$, $g(\cdot, s_k)$ is a multi-linear function mapping $\mathbb{R}^{n_z}$ into $\mathbb{R}^m$, the constraint sets $\mathcal{Z}_i$ are compact convex and k is a time-index. The time-dependent parameter s_k is assumed to lie in a subset $\mathcal{S} \subset \mathbb{R}^p$. Critical points of the parametric nonlinear program (6) are denoted by z_k^* or $z^*(s_k)$ without distinction.

Assumption 1 (Smoothness and semi-algebraicity): The functions f and g are twice continuously differentiable and semi-algebraic.

B. A truncated multi-convex splitting scheme

The basic idea of the proposed algorithm is to track time-dependent local optima z_k^* of (6) by approximately computing saddle points of the augmented Lagrangian

$$L_\rho\left(z, \mu, s_k\right) := f\left(z\right) + \left(\mu + \frac{\rho}{2} g\left(z, s_k\right)\right)^\top g\left(z, s_k\right) \quad (7)$$

subject to $z \in \mathcal{Z}$, where $\mathcal{Z} := \mathcal{Z}_1 \times \ldots \times \mathcal{Z}_P$, $\mu \in \mathbb{R}^m$ is a multiplier associated with the equality constraint $g(z, s_k) = 0$ and $\rho > 0$ is a well-chosen fixed penalty parameter. In Algorithm 1 below, the coefficients $\alpha_i > 0$ are regularisation parameters. Algorithm 1 builds a suboptimal solution $\bar{z}_{k+1}$ by applying M iterations of the proximal alternating minimisation method proposed in [2], to evaluate the primal iterates approximately. The dual variable μ is then updated in a (non-smooth) gradient ascent fashion.

Remark 1: Note that each of the subproblems in Algorithm 1 is strongly convex, hence uniquely solvable.

IV. Theoretical tools

In order to analyse the truncated augmented Lagrangian scheme, we use the concept of generalised equation, which has been introduced in real-time NMPC by [16]. The stability analysis of the sub-optimality error is also based on the convergence rate of the proximal Gauss-Seidel method in Algorithm 1.

Algorithm 1 Optimality tracking splitting algorithm

Input: Suboptimal primal-dual solution $(\bar{z}_k^\top, \bar{\mu}_k^\top)^\top$, parameter s_{k+1}, augmented Lagrangian $L_\rho\left(\cdot, \bar{\mu}_k, s_{k+1}\right)$.

$z^{(0)} \leftarrow \bar{z}_k$

for $l = 0 \ldots M-1$ **do**

  for $i = 1 \ldots P$ **do**

$$z_i^{(l+1)} \leftarrow \operatorname*{argmin}_{z_i \in \mathcal{Z}_i} \ L_\rho\Big(z_1^{(l+1)}, \ldots, z_{i-1}^{(l+1)}, z_i, z_{i+1}^{(l)}, \ldots, z_P^{(l)}, \bar{\mu}_k, s_{k+1}\Big) + \frac{\alpha_i}{2}\left\|z_i - z_i^{(l)}\right\|_2^2$$

  end for

end for

$\bar{z}_{k+1} \leftarrow z^{(M)}$; $\bar{\mu}_{k+1} \leftarrow \bar{\mu}_k + \rho g\left(\bar{z}_{k+1}, s_{k+1}\right)$

A. Parametric generalised equations

Critical points $w^*\left(s_k\right)$ of the parametric nonlinear program (6) satisfy the generalised equation

$$0 \in F\left(w, s_k\right) + \mathcal{N}_{\mathcal{Z}\times\mathbb{R}^m}\left(w\right), \quad (8)$$

where

$$F\left(w, s_k\right) := \begin{bmatrix} \nabla_z f\left(z\right) + \nabla_z g\left(z, s_k\right)^\top \mu \\ g\left(z, s_k\right) \end{bmatrix} \ , \quad (9)$$

and $w = \left(z^\top, \mu^\top\right)^\top$. A central concept of our analysis is the *strong regularity* of the generalised equation (8). As addressed in the sequel, strong regularity provides a measure of how close two time-dependent parameters need to be in order to guarantee recursive stability of the sub-optimality error.

Definition 3 (Strong regularity, [13]): Given a closed convex set C in $\mathbb{R}^n$ and a differentiable mapping $F : \mathbb{R}^n \to \mathbb{R}^n$, a generalised equation $0 \in F(x) + \mathcal{N}_C\left(x\right)$ is said to be strongly regular at a solution $x^* \in C$ if there exists radii $\eta > 0$ and $\kappa > 0$ such that for all $r \in \mathcal{B}\left(0, \eta\right)$, there exists a unique $x \in \mathcal{B}\left(x^*, \kappa\right)$ such that

$$r \in F(x^*) + \nabla F(x^*)(x - x^*) + \mathcal{N}_C\left(x\right) \quad (10)$$

and the inverse mapping from $\mathcal{B}\left(0, \eta\right)$ to $\mathcal{B}\left(x^*, \kappa\right)$ is Lipschitz continuous.

Assumption 2 (Strong regularity of (8)): For all time instants k and associated parameters $s_k \in \mathcal{S}$, the generalised equation (8) is strongly regular at a solution $w^*\left(s_k\right)$ in the sense of [13].

From Assumption 2, the following Lemma can be proven [13], guaranteeing local Lipschitz continuity of the primal-dual solution to (8).

Lemma 2 (Theorem 2.1 in [13]): There exists radii $\delta_A > 0$ and $r_A > 0$ such that for all $k \in \mathbb{N}$, for all $s \in \mathcal{B}\left(s_k, r_A\right)$, there exists a unique $w^*\left(s\right) \in \mathcal{B}\left(w_k^*, \delta_A\right)$ such that

$$0 \in F(w^*(s), s) + \mathcal{N}_{\mathcal{Z}\times\mathbb{R}^m}(w^*(s)) \quad (11)$$

and for all $s, s' \in \mathcal{B}(s_k, r_A)$,

$$\|w^*(s) - w^*(s')\|_2 \leq \lambda_A \left\|F\left(w^*(s'), s\right) - F\left(w^*(s'), s'\right)\right\|_2,$$

where $\lambda_A > 0$ is a Lipschitz constant associated with (8).

Remark 2: Without loss of generality, the radii δ_A and r_A are assumed not to depend on the parameter s_k.

Assumption 3: There exists $\lambda_F > 0$ such that for all $w \in \mathcal{Z} \times \mathbb{R}^m$,

$$\forall s, s' \in \mathcal{S}, \|F(w,s) - F(w,s')\|_2 \leq \lambda_F \|s - s'\|_2 \ . \quad (12)$$

Such an assumption is valid if, for instance, the parameter s enters the equality constraint $g(z,s) = 0$ linearly.

B. Kurdyka-Lojasiewicz property and convergence rate

The convergence properties of the proximal Gauss-Seidel scheme of Algorithm 1 have been analysed in the case of two alternations [2] under fairly general assumptions, the main one being the Kurdyka-Lojasiewicz (KL) property.

Property 1 (KL property): A lower semi-continuous function f satisfies the KL property at a point x^* in its domain if there exists a neighbourhood U of x^*, $\eta \in (0, +\infty]$ and $\phi : [0, \eta) \to \mathbb{R}_+$ such that $\phi(0) = 0$, ϕ is C^1 on $(0, \eta)$ with $\phi' > 0$ and

$$\phi'(f(x) - f(x^*))\, d(0, \partial f(x)) \geq 1 \ , \quad (13)$$

for all $x \in U \cap \{f(x^*) < f(x) < f(x^*) + \eta\}$.
Given a semi-algebraic function $L : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, it can actually be shown that L satisfies the KL property at a given critical point x^* with $\phi(t) = ct^{1-\theta}$ [5], that is there exists $\delta > 0$, $c > 0$ and $\theta \in [0, 1)$ such that for all $x \in \mathcal{B}(x^*, \delta) \cap \{x \in \mathbb{R}^n \mid L(x) > L(x^*)\}$,

$$d(0, \partial L(x)) \geq c\, (L(x) - L(x^*))^\theta \ , \quad (14)$$

where θ is taken as the smallest possible exponent satisfying (14). The parameter θ can be seen as a shape parameter of the graph of L around a critical point x^*. When θ is close to 0, the graph is sharp at x^*. When θ is close to 1, the graph is flat around x^*.

Assumption 4: The augmented Lagrangian (7) satisfies the KL property for all $\mu \in \mathbb{R}^m$ and $s \in \mathcal{S}$ with Lojasiewicz exponents $\theta(\mu, s) \in (1/2, 1)$ and radius $\delta > 0$ at its critical points. The exponents $\theta(\mu, s)$ can be upper bounded by $\hat{\theta} \in (1/2, 1)$.

Remark 3: Such an assumption is reasonable. It can be proven that for real analytic functions, the exponent θ lies within $[1/2, 1)$ [12]. Moreover, for multivariate polynomials of degree higher than two, such as $L_\rho(\cdot, \mu, s)$, an upper bound on θ can be computed, which depends only on the number of variables and the degree [8]. In many cases, the radius δ is large. For instance, in the case of strongly convex functions, $\delta = +\infty$.

Lemma 3 (Theorem 3.2 in [2]): Assuming that $M = \infty$, the sequence $\{z^{(l)}\}$ generated by the inner loop of Algorithm 1 converges to a critical point $z^\infty(\bar{\mu}_k, s_{k+1})$ of $L_\rho(\cdot, \bar{\mu}_k, s_{k+1}) + \iota_{\mathcal{Z}}(\cdot)$.
This convergence result comes with a local sub-linear R-convergence rate estimate.

Lemma 4 (Local R-convergence rate estimate): There exists a constant $C > 0$ such that, assuming $\bar{z}_k \in \mathcal{B}(z^\infty(\bar{\mu}_k, s_{k+1}), \delta)$,

$$\begin{aligned}\|\bar{z}_{k+1} - z^\infty(\bar{\mu}_k, s_{k+1})\|_2 \leq \quad & \quad (15)\\ CM^{-\psi(\hat{\theta})}\|\bar{z}_k - z^\infty(\bar{\mu}_k, s_{k+1})\|_2 \ , &\end{aligned}$$

where, given $\theta \in (1/2, 1)$,

$$\psi(\theta) := \frac{1-\theta}{2\theta - 1} \ . \quad (16)$$

Proof: From [1], as the Lojasiewicz exponent $\theta(\bar{\mu}_k, s_{k+1})$ associated with $z^\infty(\bar{\mu}_k, s_{k+1})$ lies in $(1/2, 1)$, by Assumption 4, and $\bar{z}_k \in \mathcal{B}(z^\infty(\bar{\mu}_k, s_{k+1}), \delta)$, it can be shown that, given $\bar{\mu}_k \in \mathbb{R}^m$ and $s_{k+1} \in \mathcal{S}$, there exists $C(\bar{\mu}_k, s_{k+1}) > 0$ such that

$$\|\bar{z}_{k+1} - z^\infty(\bar{\mu}_k, s_{k+1})\|_2 \leq C(\bar{\mu}_k, s_{k+1})\, M^{-\psi(\theta(\bar{\mu}_k, s_{k+1}))} \ .$$

Note that $\theta \mapsto M^{-\psi(\theta)}$ is strictly increasing on $(1/2, 1)$. Hence, from Assumption 4,

$$M^{-\psi(\theta(\bar{\mu}_k, s_{k+1}))} \leq M^{-\psi(\hat{\theta})} \ .$$

Clearly, as $\bar{z}_k$ is the suboptimal primal solution of (6) at time k, there exists $\kappa \in (0, \delta)$ such that for all $k \geq 0$,

$$\|\bar{z}_k - z^\infty(\bar{\mu}_k, s_{k+1})\|_2 \geq \kappa \ .$$

Hence there exists $C'(\bar{\mu}_k, s_{k+1}) > 0$ such that

$$\begin{aligned}&\|\bar{z}_{k+1} - z^\infty(\bar{\mu}_k, s_{k+1})\|_2 \leq \\ &\quad C'(\bar{\mu}_k, s_{k+1})\, M^{-\psi(\hat{\theta})}\, \|\bar{z}_k - z^\infty(\bar{\mu}_k, s_{k+1})\|_2 \ .\end{aligned}$$

Without loss of generality, one can assume that the constants $C'(\bar{\mu}_k, s_{k+1})$ are upper bounded, which yields (15). ∎

Remark 4: Note that the R-convergence rate of Lemma 4 shows that convergence of the multi-convex alternations is theoretically quite slow. Yet, the algorithm is observed to be quite efficient in practice, as shown in Section VI. We insist on the fact that it is an upper bound, which is used for theoretical purpose only.

V. CONTRACTION OF THE PRIMAL-DUAL SEQUENCE

As Algorithm 1 is a truncated scheme applied online for varying values of the parameters $s \in \mathcal{S}$, a natural question is the following: under which conditions does the sub-optimal primal-dual solution computed via Algorithm 1 remain close to a solution of (6) as the parameter s changes ? In the sequel, we show that if ρ and M are carefully chosen, a contraction property is satisfied by the primal-dual iterates, and the error sequence $\|\bar{w}_k - w_k^*\|_2$ remains bounded, assuming that the parameter difference $\|s_{k+1} - s_k\|_2$ is small enough.

A. Existence and uniqueness of critical points

Given a critical point w_k^* of problem (6), strong regularity of (8) implies that a critical point of (6) exists for $s = s_{k+1}$ and is unique in a neighbourhood of w_k^*, assuming that s_{k+1} is in a well-chosen neighbourhood of s_k.

Assumption 5: For all $k \in \mathbb{N}$, $\|s_{k+1} - s_k\|_2 \leq r_A$.

Lemma 5: For all $k \in \mathbb{N}$ and $s_k \in \mathcal{S}$, given w_k^* satisfying (8), there exists a unique $w_{k+1}^* \in \mathcal{B}(w_k^*, \delta_A)$ such that

$$0 \in F\left(w_{k+1}^*, s_{k+1}\right) + \mathcal{N}_{\mathcal{Z}\times\mathbb{R}^m}\left(w_{k+1}^*\right) \ . \quad (17)$$

Proof: Immediate from Assumption 5 and strong regularity of (8). ∎

B. An auxiliary generalised equation

In Algorithm 1, the proximal alternating loop, warm-started at $\bar{z}_k$, converges to $z^{\infty}(\bar{\mu}_k, s_{k+1})$, which is a critical point of $L_\rho(\cdot, \bar{\mu}_k, s_{k+1}) + \iota_{\mathcal{Z}}(\cdot)$, by Lemma 3. The following generalised equation characterises critical points of the augmented Lagrangian function $L_\rho(\cdot, \bar{\mu}, s) + \iota_{\mathcal{Z}}(\cdot)$ in a primal-dual manner, which is helpful in our analysis:

$$0 \in G_\rho(w, d_\rho(\bar{\mu}), s) + \mathcal{N}_{\mathcal{Z}\times\mathbb{R}^m}(w) \quad , \tag{18}$$

where $d_\rho(\bar{\mu}) := (\bar{\mu} - \mu_k^*)/\rho$ and

$$G_\rho(w, d_\rho(\bar{\mu}), s) := \begin{bmatrix} \nabla_z f(z) + \nabla_z g(z,s)^\top \mu \\ g(z,s) + d_\rho(\bar{\mu}) + \dfrac{\mu_k^* - \mu}{\rho} \end{bmatrix} \quad . \tag{19}$$

In the sequel, a primal-dual point satisfying (18) is denoted by $w^*(d_\rho(\bar{\mu}), s)$ or $w^*(\bar{\mu}, s)$ without distinction.

Lemma 6: Let $\bar{\mu} \in \mathbb{R}^m$, $\rho > 0$ and $s \in \mathcal{S}$. The primal point $z^*(\bar{\mu}, s)$ is a critical point of $L_\rho(\cdot, \bar{\mu}, s) + \iota_{\mathcal{Z}}(\cdot)$ if and only if the primal-dual point

$$w^*(\bar{\mu}, s) = \begin{pmatrix} z^*(\bar{\mu}, s) \\ \bar{\mu}_k + \rho g(z^*(\bar{\mu}, s), s) \end{pmatrix} \tag{20}$$

is a solution of (18).

Proof: The necessary condition is clear. To prove the sufficient condition, assume that $w^*(\bar{\mu}, s) = \left(z^*(\bar{\mu}, s)^\top, \mu^*(\bar{\mu}, s)^\top\right)^\top$ satisfies (18). The second half of (18) implies that $\mu^*(\bar{\mu}, s) = \bar{\mu} + \rho g(z^*(\bar{\mu}, s), s)$. Putting this expression in the first part of (18), this implies that $z^*(\bar{\mu}, s)$ is a critical point of $L_\rho(\cdot, \bar{\mu}, s) + \iota_{\mathcal{Z}}(\cdot)$. ∎

As $z^{\infty}(\bar{\mu}_k, s_{k+1})$ is a critical point of $L_\rho(\cdot, \bar{\mu}_k, s_{k+1}) + \iota_{\mathcal{Z}}(\cdot)$, one can define

$$w^{\infty}(d_\rho(\bar{\mu}_k), s_{k+1}) := \begin{pmatrix} z^{\infty}(\bar{\mu}_k, s_{k+1}) \\ \bar{\mu}_k + \rho g(z^{\infty}(\bar{\mu}_k, s_{k+1}), s_{k+1}) \end{pmatrix} \quad , \tag{21}$$

which satisfies (18). Note that the generalised equation (18) is parametric in s and $d_\rho(\cdot)$, which represents the normalised distance between the sub-optimal dual and the optimal dual parameters. Assuming that the penalty parameter ρ is well-chosen, the generalised equation (18) can be proven to be strongly regular at a given solution.

Lemma 7 (Strong regularity of (18)*):* There exists $\tilde{\rho} > 0$ such that for all $\rho > \tilde{\rho}$ and $k \in \mathbb{N}$, (18) is strongly regular at $w_k^* = w^*(0, s_k)$.

Proof: This follows from the reduction procedure described in [13], the arguments developed in Proposition 2.4 in [3] and strong regularity of (8) for all $k \in \mathbb{N}$. ∎

Assumption 6: The penalty parameter satisfies $\rho > \tilde{\rho}$.

From the strong regularity of (18) at w_k^*, using Theorem 2.1 in [13], one obtains the following local Lipschitz property of a solution $w(\cdot)$ to (18).

Lemma 8: There exists radii $\delta_B > 0$, $r_B > 0$ and $q_B > 0$ such that for all $k \in \mathbb{N}$,

$$\begin{aligned} &\forall d \in \mathcal{B}(0, q_B), \forall s \in \mathcal{B}(s_k, r_B), \exists! w^*(d,s) \in \mathcal{B}(w_k^*, \delta_B), \\ &\qquad 0 \in G_\rho(w^*(d,s), d, s) + \mathcal{N}_{\mathcal{Z}\times\mathbb{R}^m}(w^*(d,s)) \end{aligned}$$

and for all $d, d' \in \mathcal{B}(0, q_B)$ and all $s, s' \in \mathcal{B}(s_k, r_B)$,

$$\begin{aligned} &\|w^*(d,s) - w^*(d',s')\|_2 \leq \\ &\qquad \lambda_B \|G_\rho(w^*(d',s'), d, s) - G_\rho(w^*(d',s'), d', s')\|_2 \quad , \end{aligned}$$

where $\lambda_B > 0$ is a Lipschitz constant associated with (18). Note that, given $w \in \mathcal{Z} \times \mathbb{R}^m$, $d, d' \in \mathbb{R}^m$ and $s, s' \in \mathcal{S}$, one can write

$$\begin{aligned} G_\rho(w,d,s) - G_\rho(w,d',s') =& F(w,s) - F(w,s') \\ &+ \begin{bmatrix} 0 \\ d - d' \end{bmatrix} \quad , \end{aligned} \tag{22}$$

which, from Assumption 3, implies the following Lemma.

Lemma 9: There exists $\lambda_G > 0$ such that for all $w \in \mathcal{Z} \times \mathbb{R}^m$, for all $d, d' \in \mathbb{R}^m$ and all $s, s' \in \mathbb{R}^m$,

$$\|G_\rho(w,d,s) - G_\rho(w,d',s')\|_2 \leq \lambda_G \left\| \begin{pmatrix} d \\ s \end{pmatrix} - \begin{pmatrix} d' \\ s' \end{pmatrix} \right\|_2 \quad .$$

Proof: This follows from straightforward computations. ∎

C. Contraction estimate

In this paragraph, it is proven that under some conditions, the optimality tracking error $\|\bar{w}_k - w_k^*\|_2$ of Algorithm 1 decreases as the parameter s varies slowly. First, note that given a sub-optimal primal-dual solution $\bar{w}_{k+1}$ and a critical point w_{k+1}^*,

$$\begin{aligned} \left\|\bar{w}_{k+1} - w_{k+1}^*\right\|_2 \leq& \|\bar{w}_{k+1} - w^{\infty}(d_\rho(\bar{\mu}_k), s_{k+1})\|_2 \\ &+ \left\|w^{\infty}(d_\rho(\bar{\mu}_k), s_{k+1}) - w_{k+1}^*\right\|_2 \quad , \end{aligned} \tag{23}$$

where $w^{\infty}(d_\rho(\bar{\mu}_k), s_{k+1})$ has been defined in (21). The analysis then consists in bounding the two right hand side terms in (23), for the first term using strong regularity of (18) and for the second one using the convergence rate of the primal loop in Algorithm 1.

Lemma 10: If $\|s_{k+1} - s_k\|_2$ satisfies

$$\|s_{k+1} - s_k\|_2 < \min\left\{ r_B, \frac{q_B \rho}{\lambda_A \lambda_F} \right\} \quad ,$$

and $\|\bar{w}_k - w_k^*\|_2 < q_B \rho$,

$$\begin{aligned} \left\|w^{\infty}(d_\rho(\bar{\mu}_k), s_{k+1}) - w_{k+1}^*\right\|_2 \leq& \frac{\lambda_B \lambda_G}{\rho} \big(\|\bar{w}_k - w_k^*\|_2 \\ &+ \lambda_A \lambda_F \|s_{k+1} - s_k\|_2 \big) \quad . \end{aligned}$$

Proof: The proof can be found in [11]. ∎

In the following Lemma, using the convergence rate estimate presented in Section IV, we derive a bound on the first summand $\|\bar{w}_{k+1} - w^{\infty}(d_\rho(\bar{\mu}_k), s_{k+1})\|_2$.

Lemma 11: If $\|s_{k+1} - s_k\|_2 < r_B$, $\|\bar{w}_k - w_k^*\|_2 < q_B \rho$ and

$$\left(1 + \frac{\lambda_G \lambda_B}{\rho}\right) q_B \rho + \lambda_G \lambda_B r_B < \delta \quad , \tag{24}$$

then

$$\begin{aligned} &\left\|\bar{w}_{k+1} - w^{\infty}(d_\rho(\bar{\mu}_k), s_{k+1})\right\|_2 \leq \\ &\qquad C(1 + \rho\lambda_g) M^{-\psi(\hat{\theta})} \Big(\lambda_B \lambda_G \|s_{k+1} - s_k\|_2 \\ &\qquad + \|\bar{w}_k - w_k^*\|_2 \Big(1 + \frac{\lambda_B \lambda_G}{\rho}\Big)\Big) \quad , \end{aligned} \tag{25}$$

where $\lambda_g > 0$ is the Lipschitz constant of $g(\cdot, s)$ on $\mathcal{Z}$ (well-defined as $\mathcal{Z}$ is bounded).

Proof: Similar to the proof of Lemma 12 in [11]. ■

Gathering the results of Lemmas 10 and 11, one can formalise the following theorem.

Theorem 1 (Contraction): Given a time instant k, if the primal-dual error $\|\bar{w}_k - w_k^*\|_2$, the number of primal iterations M, the penalty parameter ρ and the parameter difference $\|s_{k+1} - s_k\|_2$ satisfy

- $\|s_{k+1} - s_k\|_2 < \min\left\{r_A, r_B, \dfrac{q_B \rho}{\lambda_A \lambda_F}\right\}$,
- $\|\bar{w}_k - w_k^*\|_2 < q_B \rho$,
- $\rho > \tilde{\rho}$,
- $$\left(1 + \frac{\lambda_G \lambda_B}{\rho}\right) \|\bar{w}_k - w_k^*\|_2 + \lambda_G \lambda_B \|s_{k+1} - s_k\|_2 < \delta \ , \tag{26}$$

then

$$\left\|\bar{w}_{k+1} - w_{k+1}^*\right\|_2 \leq \beta_w(\rho, M) \|\bar{w}_k - w_k^*\|_2 + \beta_s(\rho, M) \|s_{k+1} - s_k\|_2 \ ,$$

where

$$\beta_w(\rho, M) := C(1 + \rho\lambda_g)\left(1 + \frac{\lambda_B \lambda_G}{\rho}\right) M^{-\psi(\hat{\theta})} + \frac{\lambda_B \lambda_G}{\rho} \ , \tag{27}$$

and

$$\beta_s(\rho, M) := C(1 + \rho\lambda_g) \lambda_B \lambda_G M^{-\psi(\hat{\theta})} + \frac{\lambda_B \lambda_G \lambda_A \lambda_F}{\rho} \ .$$

Proof: This is a direct consequence of Lemmas 10 and 11. ■

Remark 5: Note that the last hypothesis (26) may be quite restrictive, since $\|\bar{w}_k - w_k^*\|_2$ needs to be small enough for it to be satisfied. However, in many cases the radius δ is large ($+\infty$ for strongly convex functions).

In order to ensure stability of the sequence of sub-optimal iterates $\bar{w}_k$, the parameter difference $\|s_{k+1} - s_k\|_2$ has to be small enough and the coefficient $\beta_w(\rho, M)$ needs to be strictly less than 1. This last requirement is clearly satisfied if ρ is large enough to make $\lambda_B\lambda_G/\rho$ small in (27). Yet ρ also appears in $1 + \rho\lambda_g$. Hence it needs to be balanced by a large enough number of primal iterations M in order to make the first summand in (27) small. The same analysis applies to the second coefficient $\beta_s(\rho, M)$ in order to mitigate the effect of the parameter difference $\|s_{k+1} - s_k\|_2$.

Corollary 1 (Boundedness of the error sequence): Assume that ρ and M have been chosen so that $\beta_w(\rho, M)$ and $\beta_s(\rho, M)$ are strictly less than 1, and $\rho > \tilde{\rho}$. Let $r_w > 0$ such that $\delta - \left(1 + \lambda_G\lambda_B/\rho\right) r_w > 0$ and $r_w < q_B \rho$. Let $r_s > 0$ such that $r_s < (1-\beta_w(\rho,M)) r_w / \beta_s(\rho, M)$.
If $\|\bar{w}_0 - w_0^*\|_2 < r_w$ and for all $k \geq 0$,

$$\|s_{k+1} - s_k\|_2 \leq \min\left\{r_s, r_A, r_B, \frac{q_B \rho}{\lambda_A \lambda_F}\right\} \ , \tag{28}$$

then for all $k \geq 0$, the error sequence satisfies

$$\|\bar{w}_k - w_k^*\|_2 < r_w \ . \tag{29}$$

Proof: Similar to the proof of Corollary 1 in [11]. ■

In the remainder, we show that a tuning of ρ, M and $\|s_{k+1} - s_k\|_2$, which ensures stability of the error sequence, is achievable on a realistic numerical example and that good tracking performance can be obtained.

VI. Numerical Example

The efficacy of Algorithm 1 is demonstrated at controlling a simple bilinear system, namely a DC motor. The discrete-time dynamics are

$$x_{l+1} = A_d x_l + B_d x_l u_l + c_d \ , \tag{30}$$

where

$$A_d := \begin{pmatrix} 1 - \frac{R_a \Delta t}{L_a} & 0 \\ 0 & 1 - \frac{B \Delta t}{J} \end{pmatrix}, \ B_d := \begin{pmatrix} 0 & -\frac{k_m \Delta t}{L_a} \\ \frac{k_m \Delta t}{J} & 0 \end{pmatrix} \ ,$$
$$c_d := \Delta t \begin{pmatrix} \frac{u_a}{L_a} \\ -\frac{\tau_l}{J} \end{pmatrix} \ , \tag{31}$$

with Δt the sampling period. The parameters values are taken from [9]. In the state variable, $x_k(1)$ is the armature current, while $x_k(2)$ is the angular speed. The control input is the field current of the machine. The control objective is to make the angular speed track a piecewise constant reference ± 2 rad/sec, while satisfying the following state and input constraints:

$$\underline{x} = \begin{pmatrix} -2 \text{ A} \\ -8 \text{ rad/sec} \end{pmatrix}, \ \overline{x} = \begin{pmatrix} 5 \text{ A} \\ 1.5 \text{ rad/sec} \end{pmatrix} \ ,$$
$$\underline{u} = 1.27 \text{ A}, \ \overline{u} = 1.4 \text{ A} \ . \tag{32}$$

The corresponding NMPC problem is solved via Algorithm 1. The tracking algorithm 1 is tested for different sampling periods, while initialised at a perturbed solution $5 \cdot w_0^*$, where w_0^* has been computed using IPOPT [15]. The speed trajecto-

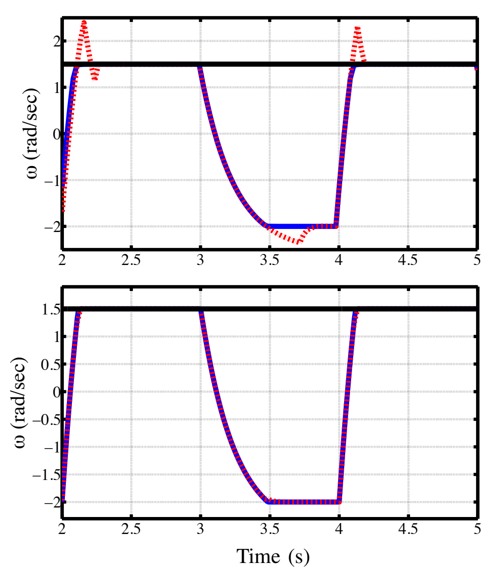

Fig. 1. Speed responses for $\Delta t = 0.026$ sec (top) and $\Delta t = 0.01$ sec (bottom): full NMPC solved using IPOPT in blue, using Algorithm 1 in dashed red.

ries are plotted in Fig. 1 and the input in Fig. 2. It clearly appears that as the sampling period is low, the tracking performance is better, the full NMPC trajectory and the sub-optimal one are almost the same. For a larger sampling period, the state constraints may be violated, as illustrated in Fig. 1, while the input constraints are always satisfied, as shown on Fig. 2, due to the formulation of Algorithm 1. Finally, the

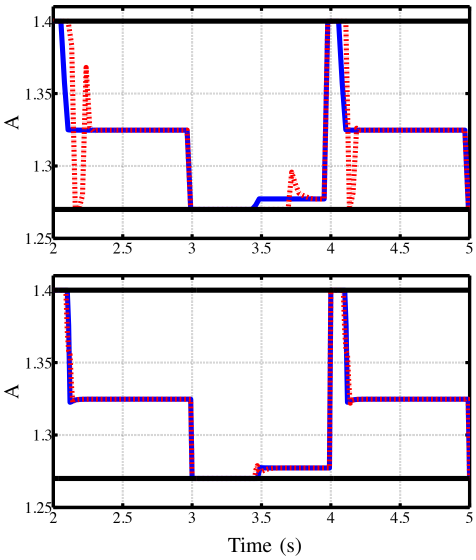

Fig. 2. Input for $\Delta t = 0.026$ sec (top) and $\Delta t = 0.01$ sec (bottom): full NMPC solved using IPOPT in blue, using Algorithm 1 in dashed red.

computational power is fixed artificially, that is a maximum number of iterations per second is given a priori. The sampling period is then made vary within a fixed range and the performance of Algorithm 1 is measured using the normalised L2-norm of the difference between the full NMPC trajectory and the sub-optimal one obtained by tracking at the given time period. As the sampling period increases, more

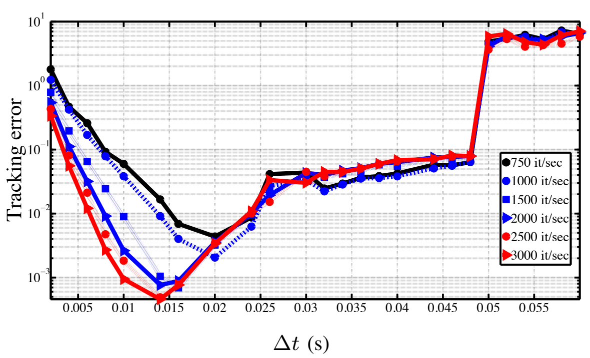

Fig. 3. Evolution of the tracking error (normalised L2-norm) versus sampling period for different computational powers.

iterations are allowed, so the tracking error decreases, as pictured on Fig. 3. If the sampling period is too large, the warm-start is too far from the optimal solution and increasing the number of iterations does not allow to reduce the error, as only one dual update is performed at each time step. As a result, the tracking error explodes for large sampling periods.

VII. Conclusion

A parametric splitting technique has been presented in order to solve time-dependent multi-convex parametric problems. A contraction estimate has been derived, which guarantees boundedness of the error sequence assuming the parameter difference is small enough. Finally, efficacy of our approach has been assessed on a realistic example consisting in speed control of a DC motor using NMPC. Our algorithm seems to be well-adapted to parallel computational environments and can be further extended to solve distributed NMPC problems in a real-time framework.

References

[1] H. Attouch and J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Mathematical Programming, 116:5–16, 2009.

[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimisation and projection methods for non-convex problems: an approach based on the Kurdyka-Lojasiewicz inequality. Mathematics of Operations Research, 35:438–457, 2010.

[3] D.P. Bertsekas. Constrained optimization and Lagrange multiplier methods. Athena Scientific, 1982.

[4] D.P. Bertsekas and J.N. Tsitsiklis. Parallel and distributed computation: numerical methods. Athena Scientific, 1997.

[5] J. Bolte, A. Daniilidis, and A. Lewis. The Lojasiewicz inequality for nonsmooth sub-analytic functions with applications to subgradient dynamical systems. SIAM Journal on Optimization, 17:1205–1223, 2007.

[6] S. 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(1):1–122, 2010.

[7] A.R. Conn, N. Could, A. Sartenaer, and P.L. Toint. Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. SIAM Journal on Optimization, 6(3):674–703, 1996.

[8] D. D'Acunto and K. Kurdyka. Explicit bounds for the Lojasiewicz exponent in the gradient inequality for polynomials. Annales Polonici Mathematici, 87, 2005.

[9] S. Daniel-Berhe and H. Unbehauen. Experimental physical parameter estimation of a thyristor driven DC-motor using the HMF-method. Control Engineering Practice, 6:615–626, 1998.

[10] M. Diehl, H.G. Bock, and J.P. Schloeder. A real-time iteration scheme for nonlinear optimization in optimal feedback control. SIAM Journal on Control and Optimization, 43(5):1714–1736, 2005.

[11] J.-H. Hours and C.N. Jones. A parametric non-convex decomposition algorithm for real-time and distributed NMPC. Technical report, EPFL, 2014. Also available at `http://arxiv-web3.library.cornell.edu/abs/1408.5120` and `http://infoscience.epfl.ch/record/201489`.

[12] S. Lojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. In Les Équations aux Dérivées Partielles, pages 87–89. Éditions du CNRS, 1963.

[13] Stephen M. Robinson. Strongly regular generalised equations. Mathematics of Operations Research, 5(1):43–62, 1980.

[14] R.T. Rockafellar and R. J.-B. Wets. Variational analysis. Springer, 2009.

[15] A. Waechter and L.T. Biegler. On the implementation of a primal-dual interior point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming, 106(1):25–57, 2006.

[16] V.M. Zavala and M. Anitescu. Real-time nonlinear optimization as a generalised equation. SIAM Journal on Control and Optimization, 48(8):5444–5467, 2010.

[17] V.M. Zavala and L.T. Biegler. The advanced-step NMPC controller: optimality, stability and robustness. Automatica, 45(1):86–93, 2009.