Experimental Investigation of Control Updating Period Monitoring
In Industrial PLC-based Fast MPC:
Application to The Constrained Control of a Cryogenic Refrigerator

François Bonne¹, Mazen Alamir² and Patrick Bonnay¹

¹ Univ. Grenoble Alpes, INAC-SBT, F-38000 Grenoble, France CEA, INAC-SBT, F-38000 Grenoble, France
² CNRS, Control Systems Department, University of Grenoble, 11 rue des Mathématiques, Domaine Universitaire, 38402 Saint-Martin d’Hères, France

In this paper, a complete industrial validation of a recently published scheme for on-line adaptation of the control updating period in Model Predictive Control is proposed. The industrial process that serves in the validation is a cryogenic refrigerator that is used to cool the supra-conductors involved in particle accelerators or experimental nuclear reactors. Two recently predicted features are validated: the first states that it is sometimes better to use less efficient (per iteration) optimizer if the lack of efficiency is over-compensated by an increase in the updating control frequency. The second is that for a given solver, it is worth monitoring the control updating period based on the on-line measured behavior of the cost function.

Index Terms—Fast MPC, control updating period, real-time implementation, PLC-based MPC, cryogenic refrigerators.

I. INTRODUCTION

Model Predictive Control (MPC) is an attractive control design methodology because it offers a natural way to express optimal objective while handling constraints on both state and control variables [17]. MPC design is based on the repetitive on-line solution of finite-horizon open-loop optimal control problems that are parametrized by the state value. Once the optimal sequence of control inputs is obtained, the first control in the sequence is applied over some updating period \( \tau_u \) during which, the new problem (based on the next predicted state) is solved and the resulted solution is applied while the prediction horizon is shifted by \( \tau_u \) time units and the process is repeated yielding an implicit state feedback. The attractive features of MPC triggered attempts to apply it to increasingly fast systems. For such systems, the need for a high updating rate (small \( \tau_u \)) may be incompatible with a complete solution of the underlying optimization problem during a single updating period \( \tau_u \). This fact fired a rich and still active research area that is shortly referred to by "Fast MPC" (see [1, 2, 12, 13] and the reference therein).

Typical issues that are addressed in fast MPC literature concern the derivation of efficient computation of updating steps, reduction of the feedback delay, more or less rigorous computation of the Hessian, etc. Typical proofs of closed-loop stability in that context (see for instance [12]) depend on strong assumptions such as the proximity to the optimal solution, the quality of the Hessian matrix estimation, etc. With such assumptions, the corresponding stability proofs take the form of tautological assertions. In other words, when such assumptions are satisfied, the paradigm of fast MPC is less relevant since standard execution of efficient optimizers would anyway give satisfactory results.

When the effectively applied control is far from being optimal (which is the case for instance after a sudden change in the set-point value) the hot-start (initialization of the decision variable after horizon shift) it induces for the next horizon does not necessarily decrease the cost function before several iterations. This is because far from the ideal solution, the final stabilizing constraints invoked in the formal proof of [17] may be far from being satisfied. On the other hand, if a constant large control updating period is used in order to accommodate for such situations, the overall closed-loop performance would be badly affected.

In recent papers [2, 3], investigations have been conducted regarding the impact of the choice of the control updating period \( \tau_u \) on the behavior of the cost function. Simple algorithms have been also proposed to monitor on-line the updating period based on the on-line behavior of the cost function to be decreased. More recently [4], it has been shown that the control updating period choice is intimately linked to the basic iteration being used. The two major facts that come out from these investigations can be summarized as follows:

(Fact 1) In a constant updating period schemes, it could be interesting to use less efficient (per iteration) algorithms provided that a significantly shorter updating period can be used [4]. This fact enhanced the recent interest [6, 16] in fast gradient-like algorithms [18] as a simpler approach when compared to second order algorithms. The work in [4] gives a formal explanation for this intuitively accepted fact.

(Fact 2) For a given optimization algorithm, the closed-loop
performance can be enhanced by an almost computational-free on-line adaptation rule of the control updating period \[2, 3\].

Obviously, a combination of the preceding facts holds also, namely, in adaptive frameworks, it can be more efficient to use simpler optimization algorithms provided that the gain induced by a higher updating rate compensates for the lack of efficiency per iteration.

In view of the preceding discussion, the contribution of this paper is twofold:

**First contribution.** This paper gives the first industrial validation of the proposed on-line adaptation of the control updating period. The realistic PLC-based implementation framework being used enhances the sensitivity of the closed-loop performance to the adaptation mechanism since it is several orders of magnitude slower than nowadays desk computers. As such, this paper gives a complete and realistic layout to understand the chain of concepts and methods that underlie fast MPC paradigm.

**Second contribution.** Although simulation-based assessments have been proposed for Facts 1 and 2 mentioned above, these simulations always used first order gradient-based algorithms. Some promoters of second order algorithms may conjecture that such adaptation would be of no benefit since a second order scheme hardly needs more than a single iteration. This paper invalidates this conjecture by showing that 1) as far as the application at hand is concerned a first-order-like algorithm slightly outperforms a second order algorithm (in the realistic industrial hardware configuration at hand) strengthening Fact 1 in a constant updating period context. 2) the closed-loop performance of this first order algorithm can be improved by on-line adaptation of the control updating period. These two results put together infer that on-line adaptation is worth using even for second order algorithms and that a single iteration is not always sufficient for second order methods in realistic situations.

This paper is organized as follows: First, the problem is stated in section II by recalling the fast MPC implementation scheme and the main results of [3, 4]. In section III the two algorithms that are used in the validation section are presented which are the QPOASES solver [4] and an Ordinary-Differential-Equation (ODE)-based solver that is briefly presented and then applied in the experimental validation. This second algorithm can be viewed as a first-order algorithm since it is based on the definition of an ODE in which the vector field is linked to the steepest descent direction. In section IV the process is described, the control problem is stated and the computational PLC used in the implementation of the real-time MPC is presented in order to underline the computational limitation that qualifies the underlying problem as a fast MPC problem. The main contribution of the paper is given in section V namely, extensive simulations are first given using the two above cited algorithms and using different constant control updating periods in order to investigate the first fact discussed above. It is in particular shown that for both solvers, the locally (in time) optimal updating period changes dynamically depending on the context. Moreover, in order to draw conclusions that go beyond the specific case of the PLC at hand, several simulations are conducted using different conjectures regarding the PLC performances. This investigation shows that for the rather performant PLC we actually have today, the first order algorithm gives slightly better results, however, if faster future PLCs were to be used, QPOASES would give better results. This is the core message of the paper: the fast MPC paradigm is a matter of combined optimal choices involving the process bandwidth, the optimization algorithm, the available computational device, the control parametrization, etc. Finally, experimental results are shown under adaptive updating period. Section VII concludes the paper and gives hint for further investigation.

## II. Background

### A. Definitions and Notation

Consider a general nonlinear system with state vector \( \mathbf{x} \in \mathbb{R}^n \) and a control vector \( \mathbf{u} \in \mathbb{R}^{n_u} \). We consider a basic sampling period \( \tau > 0 \) used to define the piece-wise constant (pwc) control profiles (a sequence of control values in \( \mathbb{R}^{n_u} \) that are maintained constant during \( \tau \) time units). As far as the general presentation of concepts is concerned, the general control parametrization is adopted according to which the whole control sequence is defined by a vector of decision variables \( \mathbf{p} \in \mathbb{R}^{n_p} \) by:

\[
U_{pwc}(\mathbf{p}) := \{ u^{(1)}(\mathbf{p}) \ldots u^{(N)}(\mathbf{p}) \} \subset \mathbb{R}^{Nn_u}
\]

where \( u^{(i)}(\mathbf{p}) \in \mathbb{R}^{n_u} \) is the control to be applied during the future \( i \)-th sampling period of length \( \tau \) while \( \mathbb{U} \subset \mathbb{R}^{n_u} \) is some admissible set. At this stage, no specific form is required for the system equations describing the dynamic model. The state \( \mathbf{x}_{k+j} \) that is reached - according to the model - after \( j \) sampling periods, starting from some initial state \( \mathbf{x}_k \), under the sequence of control inputs \( U_{pwc}(\mathbf{p}) \) and some predicted disturbance \( \tilde{w}_k \) in \( \mathbb{R}^{n_w} \) is given by:

\[
\forall j \in \{1, \ldots, N\} \quad \mathbf{x}_{k+j} := X(j, \mathbf{x}_k, p, \tilde{w}_k)
\]

while the real state that is reached in the presence of true disturbances and/or model mismatched \( \tilde{w}_k \) (that takes place over the time interval \( [k\tau, (N+j)\tau) \)) is denoted by

\[
X^r(j, \mathbf{x}_k, p, \tilde{w}_k, \tilde{w}_k)
\]

In the sequel, explicit mentioning of \( \tilde{w} \) is sometimes omitted and the real state evolution is simply denoted by \( X^r(j, \mathbf{x}_k, p, \tilde{w}_k) \).

It is assumed that an MPC strategy is defined by the following optimization problem that depends on the current state \( \mathbf{x} \) according to:

\[
P(\mathbf{x}) : \min_{\mathbf{p} \in \mathbb{P}} J_0(p, \mathbf{x}) \quad \text{under} \quad g(p, \mathbf{x}) \leq 0
\]

where \( \mathbb{P} \subset \mathbb{R}^{n_p} \) is the admissible parameter set, \( J_0 \) is the cost function to be minimized while \( g(p, \mathbf{x}) \in \mathbb{R}^{n_c} \) defines
the set of inequality constraints.

Recall that in ideal MPC, the solution to (4), say $p_{opt}(x)$, is used to define the feedback
\[ K_{MPC}(x) := u^{(1)}(p_{opt}(x)) \]  
(5)

Indeed, ideal MPC frameworks assume that the optimal solution $p_{opt}(x)$ is instantaneously available. In reality, the optimization problem $P(x)$ is solved using an iterative solver that is denoted by:
\[ p^{(q)} = S^{(q)}(p^{(0)}, x) \]  
(6)

where $p^{(0)}$ stands for the initial guess while $p^{(q)}$ is the iterate that is delivered after $q$ successive iterations. In the sequel, the term iteration refers to the unbreakable set of operations (relative to $S$) that is necessary to deliver an update of $p$. In other words, if the time needed to perform a single iteration of $S$ on a given platform is denoted by $\tau_s > 0$, then no update can be given in less than $\tau_s$ time units. Based on this remark, it seems reasonable to adopt updating instants that are separated by multiples of $\tau_s$, namely:
\[ t_{k+1} = t_k + q(t_k) \cdot \tau_s^S \quad \text{with} \quad q(t_k) \in \mathbb{N} \]  
(7)

where the $t_k$s are the instants where updated values of $p$ can be delivered for use in the feedback control input. Moreover, we assume for simplicity that the basic sampling period $\tau$ involved in the definition of the control parametrization map $\mathcal{U}(p)$ is precisely $\tau_s^S$, namely:
\[ \tau = \tau_s^S \]  
(8)

Note that thanks to the flexibility of the parametrization, one can define pwc control profiles in which the control is maintained constant over multiples of $\tau_s$ while meeting (3) so that the latter condition is not really restrictive while it simplifies the description of the implementation framework.

Using the notation above, the real-life implementation scheme is defined as follows:

(1) $i \leftarrow 0$, $t_i \leftarrow 0$, some initial parameter vector $p(t_0)$ is chosen. An initial number of iterations $q(t_0) = q_0 \leq N$ is adopted.

(2) The first $q(t_i)$ elements of the control sequence $\mathcal{U}(p(t_i))$ are applied over the time interval $[t_i, t_{i+1} = t_i + q(t_i) \tau]$. Meanwhile, the computation unit performs the following tasks during $[t_i, t_{i+1}]$:

(3.1) Predict the future state $\hat{x}(t_{i+1})$ using the model and under the above mentioned sequence of controls. The time needed to achieve this very short time ahead prediction is assumed to be negligible for simplicity.

(3.2) Perform $q(t_i)$ iterations to get
\[ p(t_{i+1}) := S^{(q(t_i))}(p^+(t_i), \hat{x}(t_{i+1})) \]
where the initial guess $p^+(t_i)$ is either equal to $p(t_i)$ [cold start] or equal to some warm start that is derived from $p(t_i)$ by standard translation technique.

(4) At the updating instant $t_{i+1}$ compute the number $q(t_{i+1})$ of iterations to be performed during the next updating period $[t_{i+1}, t_{i+2} = t_{i+1} + q(t_{i+1}) \tau]$ using Algorithm [4] that is recalled in section II.B. As it has been shown in [4] and recalled hereafter, this updating costs no more than a dozen of elementary operations and can therefore be considered as instantaneous.

(5) $i \leftarrow i + 1$, Goto step (2).

In the next section, the updating rule for $q(t_{i+1})$ invoked in Step (4) of the implementation scheme is recalled. Note that by adapting $q(t_i)$, the control updating period $\tau_u = q(t_i) \cdot \tau$ is adapted.

B. Adaptation of the control updating period for a given solver $S$

The following definition specifies a class of solvers that is invoked in the sequel and for which the adaptation mechanism recalled in this section can be applied:

Definition 1: A solver $S$ is said to be monotonic w.r.t the cost function $J : \mathbb{R}^{n_p} \times \mathbb{R}^n \to \mathbb{R}$ if for all $x$, the iterations defined by (6) satisfies:
\[ J(p^{(i)}(x, x) \leq J(p^{(i-1)}(x, x) \]  
(9)

for all $i$. This function is called hereafter the augmented cost function.

Note that $J$ generally differs from $J_0$ involved in (4) because of the presence of constraints. A typical example of such $J$ is given by the norm of the nonlinear function that gathers the Karush-Kuhn-Tucker (KKT) necessary conditions of optimality and when the solver uses a descent approach such as projected gradient or a specific implementation of Sequential Quadratic Programming (SQP) approach with trust region mechanism. Interior point-based algorithm can also enter in this category under certain circumstances in which the map $J$ would be given by the penalized version of $J_0$ involving barrier functions.

Remark 1: The conditions of Definition 1 can be relaxed in the following sense: if a solver $S$ satisfies the following condition:
\[ J(p^{(i+\ell-1)}(x, x) \leq J(p^{(i-1)}(x, x) \]  
(10)

for some map $J$, then the solver $S'$ that is derived from $S$ by:
\[ S'(p, x) := S(\ell)(p, x) \]  
(11)

is monotonic in the sense of Definition 1 at the price of having a single iteration that takes $\ell$-times longer than $S$, namely $\tau_{s'}^S = \ell \cdot \tau_s^S$.

The following assumption is needed for the updating algorithm that can be stated as follows:

Assumption 1: The solver $S$ is monotonic and the corresponding map $J$ [see Definition 1] is bounded below by a
strictly positive real $\underline{J}$, namely:

$$\forall (p, x) \quad J(p, x) \geq \underline{J} > 0$$  \hspace{1cm} (12)

Note that the last condition (12) can always be satisfied by adding an appropriate positive constant to the original cost.

In order to recall the updating algorithm proposed in [3], the following notations are needed:

$$J_k^+ := J(p^+(t_k), \hat{x}(t_{k+1}))$$

the cost function value for the initial hot start $p^+(t_k)$ (before any iteration is performed) and based on the predicted state at the future updating instant $t_{k+1} = t_k + q(t_k) \cdot \tau$.

$$\hat{J}_{k+1} := J(p(t_k), \hat{x}(t_{k+1}))$$

the cost function value for the delivered value $p(t_{k+1})$ (after $q(t_k)$ iterations) and based on the predicted state at the future updating instant $t_{k+1} = t_k + q(t_k) \cdot \tau$.

$$J_{k+1} := J(p(t_k), x(t_{k+1}))$$

the effectively obtained cost function value for the delivered value $p(t_{k+1})$ and for the true state $x(t_{k+1})$ that is reached at instant $t_{k+1} = t_k + q(t_k) \cdot \tau$.

Based on these definitions, it comes out that the decrease of the augmented cost function can be studied by analyzing the behavior of the ratio $J_{k+1}/J_k$ which can be decomposed according to:

$$\frac{J_{k+1}}{J_k} = E_k^r(q(t_k)) \times D_k^r(q(t_k))$$  \hspace{1cm} (13)

where

$$E_k^r(q(t_k)) := \frac{\hat{J}_{k+1}}{J_k^+} \quad ; \quad D_k^r(q(t_k)) := \frac{J_{k+1}}{J_{k+1}^+} \times \frac{J^+}{J_k}$$  \hspace{1cm} (14)

A deep analysis of the above terms shows that $E_k^r(q)$ is linked to the current efficiency of the solver since it represents the ratio between the value of the augmented cost for the same predicted value $\hat{x}(t_{k+1})$ of the state before and after $q(t_k)$ iterations are performed. The first ratio $J_{k+1}/J_{k+1}^+$ in $D_k^r(q)$ is 1 if the model is perfect since it represents the ratio between two values of the augmented function for the same value $p(t_{k+1})$ of the parameter but for two different values $\hat{x}(t_{k+1})$ and $x(t_{k+1})$. Finally, the ratio $J^+ / J_k$ is linked to the quality of the hot start since it represents the predicted ratio between two values of the augmented function before and just after the horizon shift.

The algorithm proposed in [3] recalled hereafter updates the number of iterations $q(t_{k+1})$ to be performed during the next updating period so that the contraction ratio:

$$K_{k+1}^r(q(t_{k+1})) := E_{k+1}^r(q(t_{k+1})) \times D_{k+1}^r(q(t_{k+1}))$$  \hspace{1cm} (15)

is lower than 1 and if this is achievable, the updating rule tries to minimize the corresponding expected response time $t_r$ of the dynamics which is linked to the ratio $q / \log(K_{k+1}^r(q))$.

This leads to the following algorithm [3]:

**Algorithm 1** Updation rule for $q(t_{k+1})$

1. **Parameters** $q_{\max} \leq N, \delta \in \{1, \ldots, q_{\max}\}$
2. **Input data** (available after Step (3.2) page 5)
3. $q = q(t_k), p^{(0)} = p^+(t_k)$, $p^{(1)} = S^{(1)}(p^{(0)}, \hat{x}(t_{k+1}))$
4. $J_k, J_k^+, J_{k+1}, J_{k+1}^+$
5. Compute the following quantities:

   $$E^r \leftarrow \hat{J}_{k+1} / J_k^+$$
   $$D^r \leftarrow (\hat{J}_{k+1} / J_k^+) / (\hat{J}_{k+1} / J_k)$$
   $$\Delta D^r \leftarrow 1 - q / [D^r - 1]$$
   $$\Delta E^r \leftarrow (J(p^{(q)}, \hat{x}(t_{k+1})) - J(p^{(q-1)}, \hat{x}(t_{k+1}))) / J(p^{(q)}, \hat{x}(t_{k+1}))$$
   $$\Delta K^r \leftarrow E^r \cdot \Delta D^r \cdot \Delta E^r$$
   $$\Delta t_r \leftarrow - \log(K^r) / \Delta q$$

6. If $K^r > 1$ then $\Gamma \leftarrow \Delta K^r / \Delta q$ else $\Gamma \leftarrow \Delta t_r / \Delta q$

7. **Output** $q(t_{k+1}) \leftarrow \max\{2, \min\{q_{\max}, q - \delta \cdot \text{sign}(\Gamma)\}\}$

Roughly speaking, this algorithm implements a step of size $\delta$ in the descent direction defined by the sign of the approximated gradient $\Gamma$. The step is projected into the admissible domain of $q \in \{2, \ldots, q_{\max}\}$. More details regarding this algorithm are available in [3].

Section V shows the efficiency of the proposed algorithm when applied to a given solver for the PLC-based implementation of MPC to the cryogenic refrigerator. Before this, the next section gives a simple argumentation that underlines a fundamental trade-off between the efficiency (per iteration) of a solver and the basic corresponding unbreakable computation time $\tau^S_t$. This is done in adaptation-free context in order to decouple the analysis.

C. Fundamental trade-off in the choice of solvers

Let us consider a solver $\mathcal{S}$ and the corresponding time $\tau^S_t$ that is needed to perform the unbreakable amount of computations involved in a single iteration. Given a control updating period $\tau_u$, the number of iterations that can be performed is equal to $q = \lceil \tau_u / \tau^S_t \rceil$ and the corresponding variation of the augmented cost function would be given by:

$$J_{k+1} - J_k := J_{k+1} - J_k^+ - J_k + \frac{E_k^S(\tau_u)}{D_k(\tau_u)}$$  \hspace{1cm} (16)

where here again, $E_k^S(\tau_u)$ and $D_k(\tau_u)$ are linked to the current efficiency of the solver ($E_k^S$) and the combined
effect of model mismatch and the horizon shift effect on the cost function respectively. Both terms depend obviously on $\tau_u$. Indeed $E^S_k(\tau_u)$ depends on $\tau_u$ through the number of iterations while $D_k(\tau_u)$ depends on $\tau_u$ since when $\tau_u = 0$ then $D_k$ vanishes (no prediction error and no possible bad hot start). Note that $E^S_k$ and $D_k$ are absolute (non relative) versions of the relative maps $E^r_k$ and $D^r_k$ invoked in section II-B to introduce Algorithm 1. Note also that unlike the efficiency indicator $E^S_k(\tau_u)$ which heavily depends on the solver, the $D_k(\tau_u)$ term is solver-independent.

Figure 1 shows typical allures of these terms for two different solvers $S_1$ (most efficient) and $S_2$ (less efficient). It can be seen that the iterations of $S_1$ are more efficient at the price of longer computation time $\tau_1^{S_1} > \tau_1^{S_2}$. The dots on the right hand plot recall that the updating can be delivered only at quantized updating instants.

Now based on (16), the decrease of $J_k$ is conditioned by the inequality:

$$E^S_k(\tau_u) > D_k(\tau_u)$$

which expresses the need to have the $E^S_k$ curves above the $D_k$ curve for the adopted value of the updating period.

Consider the Quadratic Programming (QP) problem defined by:

$$\hat{P}(z) = \begin{cases} \min: J_0(z) = z^T \Phi z + z^T \phi \\ \text{under } \Gamma z - \gamma \leq 0 \\ \underline{z} \leq z \leq \overline{z} \end{cases}$$

where $z \in \mathbb{R}^{n_z}$ is the decision variable while $\Phi$ and $\phi$ are matrices of appropriate size. $\Gamma \in \mathbb{R}^{n_c \times n_z}$ and $\gamma \in \mathbb{R}^{n_c}$ are the matrices that define the set of $n_c$ inequality constraints while $\underline{z}$ and $\overline{z}$ are lower and upper bounds on the decision variables.

Based on the above formulation, the following augmented cost function can be defined:

$$J(z) := J_0(z) + \alpha \sum_{i=1}^{n_z} \max(\Gamma_i z - \gamma_i, 0)^\mu + \alpha \sum_{i=1}^{n_z} \max(z_i - \varpi_i, 0)^\mu + \alpha \sum_{i=1}^{n_z} \max(\varpi_i - z_i, 0)^\mu$$

where $\Gamma_i \in \mathbb{R}^{1 \times n_z}$ is the $i$-th line of $\Gamma$. Based on this augmented cost, the following Ordinary Differential Equation
(ODE) can be used to define a trajectory in the decision variable space along which the augmented cost decreases:

$$\dot{z} = -\frac{dJ}{dz}(z) \quad (20)$$

Note however that this ODE is generally stiff because of the high values of $\alpha$ one needs to use in order to enforce the constraints fulfillment. That is the reason why the one-step Backward-Differentiation-Formulae (TR-BDF2) described in \[5\] for stiff differential equations is used here.

Note also that after the computation of the TR-BDF2 step, all the decision variable that correspond to hard constraints (satisfaction on actuator for instance) are projected into the admissible box before a next iteration is computed. In addition to the integration scheme described in \[5\], the initial time step is defined by using the following expression:

$$\Delta t = \sqrt{\frac{1}{\|z(t)\|}} \quad (21)$$

In the case of the quadratic problem described in paragraph \[IV-C\] this method leads to fast convergence to the suboptimal solution $z^*$, being very close to the actual optimal solution of the original problem even with real-time constraints. The comparison between solvers \[III-A\] and \[III-B\] will be done in paragraph \[V-A\].

Note also that this solver fully satisfies the decrease condition \[9\] since it moves along the descent trajectory defined by \[20\]. Therefore, the adaptation mechanism of the control updating period can be applied.

**IV. PLANT DESCRIPTION**

**A. General presentation**

Fig. 3 shows an overview of the cryogenic plant of the CEA-INAC-SBT, Grenoble. This plant provides a nominal cooling capacity of 450 W at 4.4 K in the configuration in which this study have been done. It is dedicated to physical experiments (cryogenic component testing, turbulence and pulsed heat load studies, etc.).

The process flow diagram of the cryogenic plant is shown in Fig. 4. One may notice the following main elements:

- Two volumetric screw compressors ($NC_1$) and a set of control valves ($CV_{956}$),
- Several counterflow heat exchangers ($NEF_1$), a liquid nitrogen pre-cooler ($NEF_2$),
- A cold turbine expander which extracts work from the circulating gas ($Stt_{207}$),
- A so-called turbine valve ($CV_{156}$),
- A Joule-Thomson expansion valve for helium liquefaction ($CV_{155}$),
- A phase separator ($NS_1$), connected to the loads (simulated here by the heating device referred as $NCR_{22}$).

Note that the plant can be viewed as the interconnection of four elementary subsystems: the Warm Compression Station ($S_1$), the Nitrogen Pre-Cooler ($S_2$), the Brayton Cycle ($S_3$) and the JT cycle ($S_4$), delimited by dotted lines in Fig. 4. While constrained MPC is used in this study, the cryogenic system is classically controlled by three independent control loops:

- The output temperature of the turbine expander is controlled with a PI controller working with the turbine valve $CV_{156}$;
- the level of liquid helium in the tank is controller by a PI controller, working with the heating device $NCR_{22}$;
- the high and low pressures (in red and blue pipes,
respectively) is controlled by an LQ controller, like the one described in [8];
the valve \( CV_{155} \) being used at a constant opening set by the user, depending on the application. In this study, attention has been focused on subsystems 1 and 2, with are the coldest part of the refrigerator (from 80K to 4.4K). More informations about the plant can be found in [10].

B. Model derivation and properties

In order to derive the system model, several studies have been conducted [10, 11, 8, 7]. The Joule-Thompson cycle of this paper has been modelled in [9] while the Brayton cycle is presented in details in [7]. It is worth mentioning that heat exchangers involve models with coupled partial differential equations (PDEs) that have been spatially discretized, leading to rather large state space. In this study, the two models has been merged to obtain a state space model that takes the following form:

\[
\begin{align*}
\dot{x} &= f^1(x, u, w) \\ y &= f^2(x, u, w)
\end{align*}
\]  

(22a)

(22b)

where \( f^1 \) is the function that express the derivative of the state \( x \) while \( f^2 \) is the function that express the measured output vector \( y \). Both functions are continuously differentiable. State vector, input vector, and disturbance vector are expressed more precisely by

\[
\begin{pmatrix}
x_{n_{x1}} \\ x_{nef1} \\ x_{nef2}
\end{pmatrix}, \quad
\begin{pmatrix}
u
\end{pmatrix}, \quad
\begin{pmatrix}
w
\end{pmatrix}, \quad
\begin{pmatrix}
y
\end{pmatrix}, \quad
\begin{pmatrix}
CV_{155} \\ NCR_{22}^{A} \\ NCR_{22}^{HL}
\end{pmatrix}
\]

(23)

where \( x_{n_{x1}}, x_{nef1} \) and \( x_{nef2} \) depict the state vector of individual components, described in [7]. It has to be noted that \( NCR_{22}^{A} \) is used both to control the plant and to disturb it. That is why it has been named \( NCR_{22}^{HL} \) for the actuator part and \( NCR_{22}^{HL} \) for the heat load part. At the end, \( NCR_{22} = NCR_{22}^{HL} + NCR_{22}^{A} \). The vector of measured output is the following:

\[
y = (L_1 \quad V_1 \quad T_1 \quad \cdots \quad T_{10} \quad P_1 \quad \cdots \quad P_{10})^T
\]

(24)

It has been shown in [7] that the non-linear model expressed by (27) can be linearized around an operation point of interest defined by \( f^1(x_0, u_0, w_0) = 0 \). The linearized model is then discretized using Matlab function \( c2d(\cdot) \) with sampling period \( \tau = 5s \), leading to the following discrete LTI model:

\[
\begin{align*}
\dot{x}_{k+1} &= A \tilde{x}_k + B \tilde{u}_k + F \tilde{w}_k \\
y &= C \tilde{x}_k + D \tilde{u}_k + G \tilde{w}_k
\end{align*}
\]  

(25)

(26)

where variables with a tilde depict the deviation of the original variables around the operating point of interest:

\[
\begin{align*}
x_k &= x_k + x_0, \quad \tilde{u}_k &= u_k - u_0 \\
y_k &= \tilde{y}_k + y_0, \quad \tilde{w}_k &= w_k - w_0
\end{align*}
\]  

(27)

Note that the model defined by (26) stands for the prediction model invoked in the general presentation of MPC (section II-A). Following the same notation, the predicted output is denoted by \( y_{k+j} = Y(j, x_k, p, \tilde{w}_k) \) while the true measured output is denoted by \( Y^r(j, x_k, p, \tilde{w}_k) \).

C. Statement of the MPC-related optimisation problem

First of all, the following constraints have to be satisfied as far as possible:

\[
\begin{align*}
\underline{y}^c &\leq y_k^c \leq \overline{y}^c \\
\underline{u} &\leq u_k \leq \overline{u} \\
\delta \underline{u} &\leq \delta u_k \leq \delta \overline{u}
\end{align*}
\]  

(28a)

(28b)

(28c)

where \( \delta u_k \) stand for the increment \( u_{k} - u_{k-1} \) on the input vector. \( y_k^c \) denotes a subset of output components \( y_k \) which is constrained. This subset is composed of the helium bath level \( L_1 \) and the turbine output temperature \( T_5 \). Details regarding the variables involved in (28) are given in table I.

| Var. | Meaning | Value |
|------|---------|-------|
| \( u \) | min. control effort | \( (20 \quad 20 \quad 0)^T \) |
| \( \overline{u} \) | max. control effort | \( (60 \quad 60 \quad 150)^T \) |
| \( \underline{y}^c \) | low limit on the output | \( (59 \quad 16)^T \) |
| \( \delta \underline{y} \) | high limit on the output | \( (61 \quad 9)^T \) |
| \( \delta \underline{u} \) | max increment | \( (0.5 \quad 10 \quad 0.1)^T \) |
| \( \overline{\delta u} \) | min increment | \( (0.5 \quad 10 \quad 0.1)^T \) |

| TABLE I: The constraints bounds |

One of the specific feature of Output constraints is that they cannot be necessarily fully respected depending on the unpredictable thermal loads. That is why these constrained
are systematically relaxed. This is introduced through the constraint violation variable \( v_k \) that is defined as follows:

\[
 v_k = \max(y_k^c - y_k, 0) + \max(y_k^c - y_k^0, 0) \tag{29}
\]

while constraint violation prediction at sampling instant \( k + j \) is written:

\[
 V(j, x_k, p, \hat{w}_k) = \max(Y^c(j, x_k, p, \hat{w}_k) - y_k, 0) + \max(y_k^c - Y^c(j, x_k, p, \hat{w}_k), 0) \tag{30}
\]

where \( Y^c(j, x_k, p, W) \) is used to define the constrained subset of \( Y(j, x_k, p, W) \).

The sequence of control vectors \( u^{(i)}(p) \) is then obtained by minimizing the cost function:

\[
 J_k = \sum_{j=1}^{N_p} \|X(j, x_k, p, \hat{w}_k)\|_Q^2 + \|u^{(j)}(p)\|_R^2 + \|V(j, x_k, p, \hat{w}_k)\|_\rho^2
\]

where \( Q \) and \( R \) are weighting matrices on the state and input vectors while \( \rho \) defines the constraint violation-related penalty.

This cost function, together with the linear constrained and the linearized model \( \mathcal{L} \) lead to a constrained QP problem which if of the form \( \mathcal{L} \) in which the decision variable \( z \) is precisely the control profile parameter \( p \). Note also that the affine term \( \phi \) [see \( \mathcal{L} \)] does depend on the current value of the disturbance \( w = NCR_{22} \).

By choosing a sampling period \( \tau = 5 \) sec, preliminary simulations showed that a prediction horizon of at least \( N_p = 100 \) is required. This leads to an optimization problem that involves 700 decision variables and a total number of 1000 constraints to be satisfied if trivial pwc parametrization is adopted. Such problem are beyond the computational capacity of the targeted industrial PLC (see the performance of our PLC in the section IV-D).

To reduce the problem dimension, the control profile has been parametrized using classic piece-wise affine method that leaves as decision variables the values of the control inputs at 7 decisions instants\(^1\). Moreover, the constraints satisfaction is checked only at 14 future instants\(^2\). This finally leads to an optimization problem involving 49 decision variables (note that there are 7 control inputs, namely 3 physical input and 4 virtual input representing the constraints violation), with 56 (outputs) plus 38 rate saturation constraints to be satisfied.

To ensure that this scheme is appropriate to control the plant, the problem closed-loop system is first simulated using the qpOASES solver. Time results are presented in Fig. 5. Part (a) shows the thermal heat load that has been used in this simulation. Part (b) shows that the scheme is able to decrease the stage instantaneous cost define as:

\[
 J_{inst}^k = \|x_k^c\|_Q^2 + \|u_k\|_R^2 + \|v_k\|_\rho^2 \tag{32}
\]

\(^1\)decisions instants are chosen to be: \(1, 2, 4, 8, 16, 50, 50, 100\)

\(^2\)constraints verifications instants are chosen to be: \(1, 2, 3, 4, 6, 8, 16, 24, 32, 48, 60, 72, 84, 100\)

Parts (c) and (d) of Fig. 5 show that the constraints are violated within limited amplitude and duration. Part (e) shows the control effort. Part (f) shows the number of iterations of the qpOASES solver. It is worth mentioning that the number of iterations is important during heat load event. This has significant consequence on real-time feasibility of the qpOASES-based solution as it is examined in the sequel.

**D. Description of the PLC**

This section focuses on the Programmable Logic Controller (PLC) available to implement the QP-based constrained MPC. It is a Schneider TSX P574634M shown in Fig. 5. This PLC is fully dedicated for our application and it communicates optimisation results to another PLC that actually controls the plant.
According to the manufacturer, this PLC shows maximum computing capability of about 1.8 M\textit{flops} \cite{19}. In order to evaluate this claim, the Cholesky factorisation of increasing size matrices has been executed while monitored the computation times. Fig. 7 shows the results compares them to the performance of a nowadays DELL Latitude E6520 laptop with Intel I5-2520M CPU. This graph shows a slowing factor that lies around 4000. Note that the same graph shows the performance of the PLC in ms while the performance of the desk computer is shown in $\mu$s.

Note that the PLC is used with an external PCMCIA memory card of 2\textit{Mb}, shared for both code and variables. This makes memory also a crucial issue. Indeed without reduced parametrization, the Hessian of the QP problem would have just fit the memory size of the PLC, since it represents a total memory occupation $4 \times 700^2 = 1.96\text{Mb}$ in single precision representation.

Now since a single iteration of the qpOASES solver takes approximately 120\textit{$\mu$s}, the same iteration would take $0.12 \times 4000 = 0.48\text{s}$ when executed on the PLC. Therefore only 10 iterations of the qpOASES solver can be performed during the sampling period $\tau = 5$ sec. The scenario that has been shown in Fig. 5 with no bound on the number of iterations has been simulated with the qpOASES 'maxiter' option set to 10. The result is presented by Fig. 8 on which the unlimited case has been also reported for easiness of comparison.

Figure 8 shows that when the number of iterations of the hot-started qpOASES solver is limited to 10, the closed-loop performance as well as the constraints fulfillment are drastically affected. This is precisely for this reason that the ODE-based solver explained in section III-B has been developed since it corresponds to a less computation time per iteration and can therefore be potentially more suitable in presence of the limited performance available PLC following the discussion of section II-C.

V. FAST MPC-RELATED INVESTIGATION

A. Comparison of algorithms

The aim of the present section is to assess the first Fact mentioned above, namely that it is sometimes better to use a less efficient per iteration solver (the ODE-based solver in our case) provided that it corresponds to less computation time per iteration. In our case, as far as the above described PLC is used, it is possible to perform 20 iterations of the ODE-based solver against only 10 iterations of the qpOASES solver.

Eight hours simulations have been done with the two solvers, with a variable computational capability (i.e. a variable allowed number of iterations). Some relevant results are plotted, always as a function of the normalized computation capability $\bar{P} = P/P_0$ where $P_0$ is the computational capability of our device.

In order to support the comparison that can be difficult because of the presence of relaxed weighted constraints, the cost (31) to be minimized at each sampling period has been divided in two separated parts, in order to compare them separately. The first part represents the deviation cost:

$$J_k^{dev} = \sum_{j=1}^{N_p} \| X(j,x_k,p,\hat{w}_k) \|_Q^2 + \| u^{(j)}(p) \|_R^2$$ (33)
(a) $J_{\text{dev}}$ vs $J_{\text{cst}}$

(b) $J_{\text{dev}} + J_{\text{cst}}$

(c) $J_{\text{dev}}$

(d) $J_{\text{cst}}$

(e) $\bar{J}^\rho$

(f) $\bar{J}^{BF}$

Fig. 9: Performance indicators of the two solvers comparison vs the normalized computation power. The case $\bar{P} = 1$ corresponds to the PLC we dispose of and which is presented in section IV-D.

while the second part stands for the outputs constraints violation cost:

$$\bar{J}^c_k = \sum_{j=1}^{N_p} \| V(j, x_k, p, 0) \|_\rho^2$$ (34)

and then the sum of those two costs along the whole simulation is expressed:

$$\bar{J}^{\text{dev}} = \sum_{k=1}^{N_{\text{sim}}} \bar{J}^{\text{dev}}_k$$ (35)  
$$\bar{J}^{\text{cst}} = \sum_{k=1}^{N_{\text{sim}}} \bar{J}^{\text{cst}}_k$$ (36)

where $N_{\text{sim}}$ is the number of problems solved during the simulation.

Then, constraints respect is presented in two different manners:

$$c_1 = \max_{k\in\{1, ..., N_{\text{sim}}\}} \max_{j \in \{1, ..., n_c\}} \max \{ \Gamma_j p_k - \gamma_j, 0 \}$$ (37)

being the maximum predicted constraints violation during the simulation while

$$c_2 = \frac{1}{N_{\text{sim}}} \sum_{k=1}^{N_{\text{sim}}} \max_{j \in \{1, ..., n_c\}} \max \{ \Gamma_j p_k - \gamma_j, 0 \}$$ (38)

being the average predicted constraints violation during the simulation.

Finally, a closed-loop cost has been calculated according to:

$$\bar{J}^{BF} = \sum_{k=0}^{N_{\text{sim}}} \bar{J}_{\text{inst}}^k$$ (39)

The quantities (35), (36), (35)+(36), (37), (38) and (39) are shown in Fig. 9 against normalized computational performance $\bar{P}$. It can be noticed that the suboptimal ODE-based solver is behaving better than qpOASES in the case of low performance computation devices, while the qpOASES solver becomes clearly better beyond some hardware performance indicator. The trajectories of the two closed-loop results are shown in Fig. 10 comparing the two solvers for the nominal PLC performance $P_0$ against the result obtained with the qpOASES solver with limited number of iterations and with the 10 maximum number of iterations. It comes clearly that the use of the less efficient (per iteration) solver with 20 iterations outperform the use of 10 iterations of the qpOASES solver. Moreover, the use of the ODE-based solver enables the nominal qpOASES (without limitation) performance to be recovered.

B. Control updating period monitoring

In the section, attention is focused on the ODE-based solver. First of all, simulations will be done for updating period from one to five (i.e. a number of iterations from 4 to 20), and it will be shown that quadratic performances vary and there is an optimum to be found. Then the algorithm...
A six hour heat loads scenario presented by Fig. 12 will be divided in six one hour parts, to be simulated. Cost (35) + (36) defined in the previous section will be plotted against the chosen updating period. The result is presented by Fig. 11. It can be noted that the optimum updating period is different reading the scenario. It illustrates the fact that the updating period should be monitored to enhance performance.

The Fig. 11 also plots the obtained performance by monitoring the updating period using the algorithm described in [3] for \( \delta = 2 \) will be implemented to show its efficiency on the cryogenic plant.

VI. EXPERIMENTAL RESULTS

The control scheme derived in section IV-C and the solver depicted in section III-B has been implemented in the Schneider PLC described in section IV-D in structured language. The objective of the section is triple. First, we want to show that the problem we derive in section IV-C is relevant regarding the control of a cryo-refrigerator submitted to transient heat loads. Then, we want to emphasize that the algorithm described in III-B is PLC compliant, even with polyhedral constraints. Finally, we will see that monitoring the updating scheme is very useful in this particular cases.

A. Control result with real time PLC implementation

The plant has been submitted to a two hours scenario (first two hours of Fig. 12), starting from the equilibrium. The observed time per iteration is never longer than 500ms as expected and the problem preparation time do not exceed 500ms also. It allows the optimisation algorithm to iterate \( 4\tau_u - 1 \) time. For the first test, we chose to use a \( \tau_u = 5s \) updating period. Fig. 13 shows that the control scheme is able to stabilize the plant and make the constraints to be respected, even if the pant is submitted to transient variable loads.

B. Some leads on the updating scheme efficiency

The algorithm to update the updating period as been implemented on the PLC to show its efficiency. Unfortunately, the cost is not monitored but it is still possible to show result in the time domain. Fig 14 shows the difference between a constant updating period and a variable one. One can see that in the case of a serious change on the thermal load, the updating period is increasing to iterate more, while the algorithm is imposing a short updating period as soon as the problem is not changing much from an updating instant to another.

VII. CONCLUSION

In this paper, an efficient way to control a cryogenic plant submitted to variable heat loads using an industrial PLC with
Fig. 12: Six hours heat loads scenario

Fig. 14: Result with both constant (2.5s) and real-time updated updating period. It an be notice that in the case of a heat load disturbance, the updating period is increasing (in order to make the number of iteration to also increase), since the hot-starting solution is far from the actual solution. Period represent the updating period $\tau_u$. For actuators legend, please refer to Fig. 13.

reduced computing capabilities is proposed. It has been shown that in this application, an ODE-based solver gives robust sub optimal solution even in the case where the hot-start is far from the optimal solution (in the case of an unpredicted disturbance for instance). The control scheme and the solver has been both validated experimentally.

Moreover an algorithm that automatically monitors the control updating period has been implemented and experimentally successfully tested.

Future investigation will aim at developing an MPC control scheme for a refrigerator submitted to multiple different thermal loads, including at 1.8$K$ (super-fluid helium). Also, cryogenic systems could be very large (several buildings): the control scheme will be distributed in order to ensure a progressive integration to industrial systems.

ACKNOWLEDGMENT

Authors would like to thank every co-worker from the SBT for their kind help to improve models and control strategy and for their time to correct and discuss this paper. Authors give special thanks to Michel Bon-Mardion, Lionel Monteiro, François Millet, Christine Hoa, Bernard Roussel and Jean-Marc Poncet from SBT for their explanation about the process and their participation on experimental campaigns.

REFERENCES

[1] M. Alamir. Stabilization of nonlinear systems using receding-horizon control schemes: A parameterized approach for fast systems. Springer-Verlag, 2006.

[2] M. Alamir. A framework for monitoring control updating period in real-time NMPC schemes. In Lalo Magni, DavideMartino Raimondo, and Frank Allgöwer, editors, Nonlinear Model Predictive Control, volume 384 of Lecture Notes in Control and Information Sciences, pages 433–445. Springer Berlin Heidelberg, 2009.

[3] M. Alamir. Monitoring control updating period in fast gradient based NMPC. In Control Conference (ECC), 2013 European, pages 3621–3626, July 2013.

[4] M. Alamir. Fast NMPC, a reality-steered paradigm: Key properties of fast NMPC algorithms. In Control Conference (ECC), 2014 European, June 2014.

[5] R. E. Bank, William M. Coughran Jr., Wolfgang Fichtner, Eric Grosse, Donald J. Rose, and R. Kent Smith. Transient simulation of silicon devices and circuits. IEEE
[6] A. Bomporad and P. Patrinos. Simple and certifiable quadratic programming algorithms for embedded linear model predictive control. In *Proceeding of the IFAC Non-linear Predictive Control Conference*, Noordwijkerhout, NL, 2012.

[7] F. Bonne, M. Alamir, and P. Bonnay. Physical control oriented model of large scale refrigerators to synthesize advanced control schemes. Design, validation, and first control results. In *Proceedings of the Cryogenic Engineering Conference*, 2013.

[8] F. Bonne, M. Alamir, P. Bonnay, and B. Bradu. Model based multivariable controller for large scale compression stations. Design and experimental validation on the lhc 18kw cryorefrigerator. In *Proceedings of the Cryogenic Engineering Conference*, 2013.

[9] François Bonne, Mazen Alamir, and Patrick Bonnay. Nonlinear observers of the thermal loads applied to the helium bath of a cryogenic Joule-Thompson Cycle. *Journal of Process Control*, 24(3):73–80, January 2014.

[10] F. Clavel. Modélisation et contrôle d’un réfrigérateur cryogénique. Application à la station 800W à 4.5K du CEA Grenoble. PhD thesis, EEATS, 2011.

[11] F. Clavel, M. Alamir, P. Bonnay, A. Barraud, G. Bornard, and C. Deschildre. Multivariable control architecture for a cryogenic test facility under high pulsed loads: Model derivation, control design and experimental validation. *Journal of Process Control*, 21(7):1030–1039, 2011.

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

[13] M. Diehl, R. Findeisen, F. Allgower, H. G. Bock, and J. P. Schloder. Nominal stability of real-time iteration scheme for nonlinear model predictive control. *IEEE Proc.-Control Theory Appl.*, 152(3):296–308, 2005.

[14] 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*, 18(8):816–830, 2008.

[15] H.J. Ferreau, C. Kirches, A. Potschka, H.G. Bock, and M. Diehl. qpOASES: A parametric active-set algorithm for quadratic programming. *Mathematical Programming Computation*, 2014. (in print).

[16] C. N. Jones, A. Domahidi, M. Morari, S. Richter, F. Ullmann, and M. Zeilinger. Fast predictive control: Real-time computation and certification. In *Proceedings of the IFAC Nonlinear Predictive Control Conference*, Noordwijkerhout, NL, 2012.

[17] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36:789–814, 2000.

[18] Y. Nesterov. A method of solving a convex programming problem with convergence rate o (1/k2). *Soviet Mathematics Doklady*, 27(2):372–376, 1983.

[19] Schneider Electric, available at http://www.schneider-electric.com/products/. *Schneider PLC TSXP574634M Product data sheet.*

[20] V. M. Zavala, C. D. Laird, and L. T. Biegler. Fast implementation and rigorous models: can both be accommodated in NMPC? *International Journal of Robust and Nonlinear Control*, 18(8):800–815, 2008.