

An improved primal-dual interior-point solver for model predictive control

Zhang, Xi; Ferranti, Laura; Keviczky, Tamas

DOI

[10.1109/CDC.2017.8263807](https://doi.org/10.1109/CDC.2017.8263807)

Publication date

2017

Document Version

Accepted author manuscript

Published in

Proceedings of the 2017 56th IEEE Annual Conference on Decision and Control

Citation (APA)

Zhang, X., Ferranti, L., & Keviczky, T. (2017). An improved primal-dual interior-point solver for model predictive control. In A. Astolfi (Ed.), *Proceedings of the 2017 56th IEEE Annual Conference on Decision and Control* (pp. 1126-1131). IEEE. <https://doi.org/10.1109/CDC.2017.8263807>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

An Improved Primal-Dual Interior Point Solver for Model Predictive Control*

Xi Zhang¹, Laura Ferranti² and Tamás Keviczky²

Abstract—We propose a primal-dual interior-point (PDIP) method for solving quadratic programming problems with linear inequality constraints that typically arise from MPC applications. We show that the solver converges (locally) quadratically to a suboptimal solution of the MPC problem. PDIP solvers rely on two phases: the damped and the pure Newton phases. Compared to state-of-the-art PDIP methods, our solver replaces the initial damped Newton phase (usually used to compute a medium-accuracy solution) with a dual solver based on Nesterov’s fast gradient scheme (DFG) that converges with a sublinear convergence rate of order $\mathcal{O}(1/k^2)$ to a medium-accuracy solution. The switching strategy to the pure Newton phase, compared to the state of the art, is computed in the dual space to exploit the dual information provided by the DFG in the first phase. Removing the damped Newton phase has the additional advantage that our solver saves the computational effort required by backtracking line search. The effectiveness of the proposed solver is demonstrated on a 2-dimensional discrete-time unstable system and on an aerospace application.

I. INTRODUCTION

Model predictive control (MPC) is an advanced control technique that offers an elegant framework to solve a wide range of control problems (regulation, tracking, supervision, etc.) and handle constraints on the plant. The control objectives and the constraints are usually formulated as an optimization problem that the MPC controller has to solve (either offline or online) to return the control command for the plant. In this work, we focus on MPC problems with quadratic cost and linear constraints that can be formulated as quadratic programming (QP) problems (which cover a large range of practical applications).

The presence of this optimization problem has traditionally limited the use of MPC to *slow* processes, that is, processes with no hard real-time constraints. Recently, MPC has received increasing attention in fields, such as aerospace and automotive, where the real-time aspects are critical and the computation time for the controller is limited. Hence, offline (e.g., the explicit MPC proposed by [1]) and online (using solvers tailored for MPC applications) solutions have been investigated to overcome the computational issues related to the MPC controller. In this work, we focus on online solutions that allow one to handle a wider range of problems. Online optimization algorithms can be divided in

two main families: first- and second-order methods. First-order methods, such as gradient or splitting methods and their accelerated versions [2], [3], [4], [5], have simpler theoretical requirements (such as, Lipschitz continuity only on the first derivative of the cost) and converge to a medium-accuracy solution within few iterations. Second-order methods, such as active-set [6], [7], [8] and interior-point methods [9], [10], [11], have more strict theoretical requirements (such as, Lipschitz continuity on the first and second derivative of the cost), but are more suitable when a high-accuracy solution is required. In this work, we are mainly interested in primal-dual interior-point (PDIP) solvers [12]. As detailed for example in [12], [13], their convergence can be divided in two phases: (i) *damped Newton phase* characterized by a linear convergence rate and used to reach a medium-accuracy solution; (ii) *pure Newton phase* characterized by a quadratic convergence rate and used to improve the accuracy of the solution obtained from phase (i).

Contribution. The main contribution of this paper is a PDIP solver for solving inequality-constrained QP problems that commonly arise from MPC applications. The proposed solver combines the advantages of Nesterov’s fast gradient (FG) method [4] and of PDIP solvers. In particular, we exploit the ability of the fast gradient method to converge to medium-accuracy solutions within few iterations (that can be performed efficiently) and the ability of PDIP solvers to converge to high-accuracy solutions in Phase (ii). The proposed combination aims to improve the convergence of the PDIP by replacing Phase (i) with the FG method that has a sublinear convergence rate of order $\mathcal{O}(1/k^2)$. We modify the classical analysis of the PDIP [13] to take into account the presence of (active) inequality constraints and allow the switch between the two solvers. This is done by moving the convergence analysis of the solver and the formulation of the switching strategy to the dual framework. As a consequence, we provide bounds on the level of primal suboptimality and infeasibility achieved with our proposed algorithm. An additional feature of the proposed solver is that the computational effort related to the backtracking line search (required in Phase (i), refer to [13, Chapter 9] in details) is removed, given that Phase (ii) uses a unit step size. Finally, the proposed approach is tested on two MPC applications, that is, the regulation of an unstable input- and output- constrained planar system and the Cessna Citation Aircraft system [14].

Notation. We work in the space \mathbb{R}^n composed of column vectors. $\|\mathbf{z}\|_1$ and $\|\mathbf{z}\|_2$ indicate the 1-norm and 2-norm, respectively. $[\mathbf{z}]_+$ and $[\mathbf{z}]_-$ indicate the projection

*This research is supported by the European Union’s Seventh Framework Programme FP7/2007-2013 under grant agreement n. AAT-2012-RTD-2314544 (RECONFIGURE) and by the TU Delft Space Institute.

¹ cindyzhang0919@gmail.com

²Delft University of Technology, Delft, 2628 CD, The Netherlands {l.ferranti,t.keviczky}@tudelft.nl

onto nonnegative orthant and negative orthant, respectively. Furthermore, $\text{diag}(\mathbf{z})$ indicates a matrix that has the elements of \mathbf{z} on its main diagonal. Finally, $\mathbf{1}$ is the vector of ones.

II. PROBLEM FORMULATION

In this work we focus on the control of *discrete, linear, time-invariant* systems, which can be described as follows:

$$x(t+1) = Ax(t) + Bu(t) \quad (1a)$$

$$y(t) = Cx(t) + Du(t) \quad (1b)$$

where $x \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$, $u \in \mathcal{U} \subseteq \mathbb{R}^{n_u}$, $y \in \mathcal{Y} \subseteq \mathbb{R}^{n_y}$ denote the state, control command, output, respectively. The sets \mathcal{U} , \mathcal{X} , \mathcal{Y} are closed convex sets that contain the origin in their interior. In addition, $t \geq 0$ denotes the sampling instant. Furthermore, the system matrices A , B , C , and D are constant matrices of fixed dimension. In the remainder of the paper, we assume that the pair (A, B) is stabilizable.

In this work, we focus on regulation problems that can be addressed by solving online (i.e., every time new measurements $x(t)$ are available from the plant) the following optimization problem:

$$J^* = \underset{x_k \in \mathcal{X}, u_k \in \mathcal{U}}{\text{minimize}} \sum_{k=0}^N J_k(x_k, u_k) \quad (2a)$$

$$\text{subject to: } x_{k+1} = Ax_k + Bu_k \quad (2b)$$

$$y_k \in \mathcal{Y}, \quad k = 0, \dots, N, \quad (2c)$$

$$x_0 = x(t), \quad (2d)$$

where x_k , u_k , and y_k represent the predicted evolution of the state, control command, and output, respectively. Furthermore, $J_k(x_k, u_k)$ denotes the stage cost and is defined as follows:

$$J_k(x_k, u_k) = \frac{1}{2} \begin{cases} (x_k^T Q x_k + u_k^T R u_k) & k = 0, \dots, N-1, \\ x_N^T Q x_N & k = N, \end{cases}$$

where $Q = Q^T \succeq 0$ and $R = R^T \succ 0$ weigh the state and input, respectively.

Condensed Formulation. We exploit the plant dynamics (2b) to eliminate the states from the decision variables and express them as an explicit function of the current measured state $x(t)$ and future control inputs (refer to [14] for more details). This is known as *condensed* QP formulation, which leads to compact and dense QPs with the control inputs as decision variables:

$$\mathbf{z} := [u_0^T \quad u_1^T \quad \dots \quad u_{N-1}^T]^T. \quad (3)$$

This leads to a set of equalities with dense matrices expressing the decision variables as a function of the current state and input sequence:

$$\mathbf{x} = A_N x(t) + B_N \mathbf{z} \quad (4a)$$

$$\mathbf{y} = C_N x(t) + D_N \mathbf{z} \quad (4b)$$

where $\mathbf{x} = [x_0^T \quad x_1^T \quad \dots \quad x_N^T]$ and $\mathbf{y} = [y_0^T \quad y_1^T \quad \dots \quad y_N^T]$. For details on the structure of A_n , B_n , C_n and D_n refer to [14].

Further, define $\mathcal{Z} := \mathcal{U} \times \dots \times \mathcal{U}$ (N times). Finally, the optimization problem in (2) can be written in the following standard form which only consists of inequality constraints:

$$\min_{\mathbf{z}} f_0(\mathbf{z}) := \frac{1}{2} \mathbf{z}^T H \mathbf{z} + (hx(t))^T \mathbf{z} \quad (5a)$$

$$\text{s.t. } g(\mathbf{z}) := G \mathbf{z} + Ex(t) + g \leq 0 \quad (5b)$$

where $H \succ 0$ (given $R \succ 0$), h , G , E , and g are given matrices (more details on their structure can be found in [14]) and (5b) encodes \mathcal{U} , \mathcal{X} , and \mathcal{Y} .

In this paper, we consider the following assumptions:

Assumption 1. Function $f_0(\mathbf{z})$ is m_p -strongly convex and twice differentiable, m_p is the convexity parameter.

Assumption 2. The Slater condition holds for Problem (5), i.e., there exists $\mathbf{z} \in \text{relint}(\mathcal{Z})$ with $g(\mathbf{z}) < 0$.

III. PRELIMINARIES

A. Dual Fast Gradient Method

In the following, we provide a high level description of the fast gradient method proposed by Nesterov (refer to [2], [15], [4] for more details) with a focus on its application to the dual of Problem (5), as Algorithm 1 details. In particular, notice that Algorithm 1 exploits (steps 5 and 6) a projection step at each iteration. If this projection is hard to compute, this operation can be challenging to solve efficiently online. This is the case, if we apply Algorithm 1 directly to Problem (5), in which the set $g(\mathbf{z}) \leq 0$ consists of so-called complicating constraints. Hence, we operate on the dual of Problem (5) that only requires the computation of the projection on the positive orthant, which can be computed efficiently. The dual of Problem (5) can be computed as follows. Given the Lagrangian described below:

$$\mathcal{L}(\mathbf{z}, \lambda) = f_0(\mathbf{z}) + \lambda^T g(\mathbf{z}) \quad (6)$$

where $\lambda \in \mathbb{R}^m$ is the Lagrange multiplier, the dual function is described as follows:

$$d(\lambda) = \min_{\mathbf{z}} \mathcal{L}(\mathbf{z}, \lambda) \quad (7)$$

which first-order derivative is given as follows

$$\nabla d(\lambda) = g(\mathbf{z}).$$

The dual of Problem (5) is given by

$$f_0(\mathbf{z}^*) = d(\lambda^*) = \max_{\lambda \geq 0} d(\lambda), \quad (8)$$

where the $f_0(\mathbf{z}^*) = d(\lambda^*)$ follows from Assumptions 1 and 2.

Algorithm 1 uses $1/L_d$ ($L_d := \|GH^{-1}G^T\|_2$) as step size. T his step size is optimal for the proposed algorithm [16]. In particular, as detailed in [16], $d(\lambda)$ is Lipschitz continuous with constant L_d , leading to a tighter upper bound than the one provided for example in [4], [17]. Furthermore, the algorithm initializes $\lambda_0 = \hat{\lambda} \geq 0$ (step 2). A complete analysis for Algorithm 1 starting from $\hat{\lambda} \neq 0$ can be found in [17]. For simplicity, we start Algorithm 1 with $\hat{\lambda} = 0$. At

Algorithm 1 Dual Fast Gradient Method.

- 1: Given $H, h, g, E, G, x(t), \hat{\lambda}, k_{\max}, L_d$.
- 2: Initialize $\lambda_0 = \hat{\lambda}$.
- 3: **for** $k = 0, \dots, k_{\max}$ **do**
- 4: Compute $\mathbf{z}_k = \arg \min_{\mathbf{z}} \mathcal{L}(\mathbf{z}, \lambda_k)$.
- 5: Compute $\hat{\lambda}_k = \left[\lambda_k + \frac{1}{L_d} \nabla d(\lambda_k) \right]_+$.
- 6: Compute $\lambda_{k+1} = \frac{k+1}{k+3} \hat{\lambda}_k + \frac{2}{L_d(k+3)} \left[\sum_{j=0}^k \frac{j+1}{2} \nabla d(\lambda_j) \right]_+$.
- 7: **end for**

Algorithm 2 Primal-Dual Interior Point Method.

- 1: Given $H, h, g, E, G, x(t)$.
- 2: Initialize $\mathbf{z}_0, \lambda_0 > 0, s_0 > 0$, centering parameter $\kappa \in (0, 1)$, backtracking line search parameters $\alpha \in (0, 0.5), \beta \in (0, 1)$, tolerance $\varepsilon > 0$, and $k = 0$.
- 3: **repeat**
- 4: Determine $\tau_{k+1} = \mu_{k+1} = \kappa \mu_k$.
- 5: Compute search direction $\Delta \zeta_{\text{pd}}$ by solving (11).
- 6: Backtracking line search $\rho_k := 1$
- 7: **while** $f_0(\mathbf{z} + \rho_k \Delta \mathbf{z}) > f_0(\mathbf{z}) + \alpha \rho_k \nabla f_0(\mathbf{z})^T \Delta \mathbf{z}$ **do**
- 8: $\rho_k := \beta \rho_k$.
- 9: **end while**
- 10: Update $\zeta_{k+1} = \zeta_k + \rho_k \Delta \zeta_{\text{pd}}$ where $\rho_k > 0$ is the step size.
- 11: $k = k + 1$.
- 12: **until** stopping criterion $\mu_k \leq \varepsilon$.
- 13: **return** Point close to \mathbf{z}^* from $\zeta_k = (\mathbf{z}_k, \lambda_k, s_k)$.

every iteration, the algorithm first computes a minimizer for Problem (7) (step 4). Then, it performs a linear update of the dual variables (steps 5-6).

Convergence Analysis. Algorithm 1 has a convergence rate of $\mathcal{O}(1/k^2)$, as demonstrated in [2], [4]. In addition, as discussed for example in [17], Algorithm 1 converges to a suboptimal solution of MPC problem after k_{\max} iterations (refer to [17] for more details). In the remainder of the paper, we make the following assumption:

Assumption 3. *There exists a $k_{\max} \geq 0$ such that after k_{\max} iterations, Algorithm 1 is able to find a solution $\mathbf{z}_{k_{\max}}$ close to the central path, such that, for $\eta_d \geq 0$, the following holds:*

$$\| [g(\mathbf{z}_{k_{\max}})]_+ \|_2 \leq \eta_d.$$

B. Primal-Dual Interior Point Method

In the following, we present a version of the primal-dual interior point (PDIP) method proposed by [12] and described in Algorithm 2.

The general idea of primal-dual interior point methods is to solve the KKT conditions by using a modified version of Newton's method. In this respect, recall the Lagrangian defined in (6). PDIP solves the following *relaxed* KKT conditions:

$$\nabla f_0(\mathbf{z}) + Dg(\mathbf{z})^T \lambda = 0 \quad (9a)$$

$$g(\mathbf{z}) + s = 0 \quad (9b)$$

$$S\lambda = \tau \mathbf{1}_m, \quad (9c)$$

$$(s, \lambda) > 0, \quad (9d)$$

where $S := \text{diag}(s)$, $s \in \mathbb{R}^m$ is the slackness variable, $Dg(\mathbf{z})$ is the derivative matrix of the inequality constraint

function $g(\mathbf{z})$, and $\tau \in [0, \mu]$, where $\mu = s^T \lambda / m$ denotes the average duality gap.

From (9), the residual variable is defined as follows:

$$r_\tau(\mathbf{z}, \lambda, s) = \begin{bmatrix} \nabla f_0(\mathbf{z}) + Dg(\mathbf{z})^T \lambda \\ g(\mathbf{z}) + s \\ S\lambda - \tau \mathbf{1} \end{bmatrix} = \begin{bmatrix} r_{\text{dual}} \\ r_{\text{pri}} \\ r_{\text{cent}} \end{bmatrix}. \quad (10)$$

The search direction can be obtained by linearizing (10) at the current iterate $\zeta_k = (\mathbf{z}_k, \lambda_k, s_k)$:

$$\underbrace{\begin{bmatrix} \nabla^2 f_0(\mathbf{z}_k) & Dg(\mathbf{z}_k)^T & 0 \\ Dg(\mathbf{z}_k) & 0 & I \\ 0 & S_k & \Lambda_k \end{bmatrix}}_{Dr_\tau(\zeta)} \underbrace{\begin{bmatrix} \Delta \mathbf{z}_{\text{pd}} \\ \Delta \lambda_{\text{pd}} \\ \Delta s_{\text{pd}} \end{bmatrix}}_{\Delta \zeta_{\text{pd}}} = - \begin{bmatrix} r_{\text{dual}} \\ r_{\text{pri}} \\ r_{\text{cent}} \end{bmatrix} \quad (11)$$

where $\Lambda_k = \mathbf{diag}(\lambda_k)$ and we use $\lambda^T \nabla^2 g(\mathbf{z}) = 0$.

Convergence Analysis. The convergence of PDIP has been shown in [13, Chapter 10]. The convergence analysis is based on the convergence of the residual variable r_τ and can be divided into two phases. In the first phase, that is, the *damped Newton phase*, Algorithm 2 converges *linearly*, while in the second phase, that is, the *pure Newton phase*, the backtracking line search selects unit step size and *quadratic* convergence rate can be achieved. In order to enter pure Newton phase, the 2-norm of the residual variable has to satisfy the following condition

$$\|r_\tau(\mathbf{z}, \lambda, s)\| \leq \eta_p \quad (12)$$

where $0 < \eta_p \leq \frac{m_{Dr}}{L_{Dr}}$ with m_{Dr} the lower bound on $Dr_\tau(\zeta)$ and L_{Dr} denotes the Lipschitz constant of $Dr_\tau(\zeta)$.

Loosely speaking, the condition above states that if the algorithm is far from the optimal solution it converges more slowly (damped Newton phase), while when it is close to the optimal solution, it converges faster (pure Newton phase). We rely on this observation to design an improved PDIP method that fully replaces the (slow) damped Newton phase with the dual fast gradient (DFG). In particular, we propose to exploit the DFG until (12) is satisfied, that is, we completely remove the damped Newton phase in Algorithm 2. Then, we initialize the PDIP with the solution returned by DFG to enter directly into the pure Newton phase, in which PDIP converges quadratically. As shown in the next section, combining these two methods is not trivial. One of the main issues is related to the presence of inequality constraints, compared to [13, Chapter 10], which only takes into account equality constraints. In the presence of inequality, (12) (in the primal space) becomes too conservative to be used in practice (especially in the presence of active constraints), because the condition numbers of $Dr_\tau(\zeta)$ are hard to derive due to asymmetry and variance of $Dr_\tau(\zeta)$. Therefore, to overcome this issue, we propose, in the next section, a new switching condition (in the dual space) to enter the pure Newton phase.

IV. PROPOSED SOLVER

As described in Section III-B, we would like to eliminate the damped Newton phase from PDIP and preserve the pure Newton phase, which allows the algorithm to converge

quadratically to the optimal solution of Problem (5). Compared to the state of the art that analyzes the convergence of the algorithm in the presence of equality constraints and in primal space, we move our analysis to the dual framework. This choice is strengthened by the decision of using the DFG to replace the damped Newton phase. In particular, building on the convergence analysis in [13, Chapter 9] of the standard Newton's method, we derive an estimate on primal suboptimality and feasibility violation achieved with the proposed solver.

The following lemma (from [17, Theorem 2.1]) is useful to prove the convergence of our proposed solver.

Lemma 1. *Under Assumption 1 and 2, the dual function in (7) is twice differentiable. Then the gradient is given by: $\nabla d(\lambda) = g(\mathbf{z}(\lambda))$ and the dual Hessian is given by:*

$$\begin{aligned} \nabla^2 d(\lambda) &= -\nabla g(\mathbf{z}(\lambda)) [\nabla^2 f_0(\mathbf{z})]^{-1} \nabla g(\mathbf{z}(\lambda))^T \\ &= -GH^{-1}G^T \end{aligned} \quad (13)$$

Furthermore, the 2-norm of dual Hessian is bounded as follows:

$$m_d = \frac{\|G\|_2^2}{\sigma_{\max}(H)} \leq \|\nabla^2 d(\lambda)\|_2 \leq \frac{\|G\|_2^2}{\sigma_{\min}(H)} = M_d \quad (14)$$

where $\sigma_{\max}(H)$ and $\sigma_{\min}(H)$ denote the maximum and minimum eigenvalues of $H \succ 0$, respectively.

Proof. According to Assumption 1, $f_0(\mathbf{z})$ is m_p -strongly convex, that is, H is positive definite, then the following holds:

$$\begin{aligned} \|\nabla^2 d(\lambda)\|_2 &= \|GH^{-1}G^T\|_2 \\ &\geq \frac{1}{\sigma_{\max}(H)} \|GIG^T\|_2 \end{aligned} \quad (15)$$

using the fact that $H^{-1} = U\Sigma^{-1}U^T$ where Σ is the singular value of H and U is an unitary matrix from singular value decomposition (SVD). The proof of the upper bound is similar to the one in [17, Theorem 2.1]. In particular, we consider the derivative of the first-order optimality condition of Problem (5) with respect to λ to obtain $\nabla \mathbf{z}(\lambda)$. Then, (13) follows by using $\lambda^T \nabla^2 g(\mathbf{z}) = 0$, given that only linear inequality constraints are involved in (5). \square

From (13), it follows that $\nabla^2 d(\lambda)$ is Lipschitz continuous with any $L_{dH} \geq 0$.

Theorem 1. *Under Assumptions 1-3, there exist $0 < \eta_d \leq m_d^2/L_{dH}$, $\gamma_d > 0$ and $L_d > 0$ such that the following holds:*

- DFG Phase. *If $\|[g(\mathbf{z})]_+\|_2 \geq \eta_d$, DFG runs until the η_d -solution is achieved (after k_{\max} iterations).*
- Pure Newton Phase. *If $\|[g(\mathbf{z})]_+\|_2 \leq \eta_d$, for all $k \geq 0$, with step size $\rho_k = 1$, the following holds:*

$$\|\nabla d(\lambda_{k+1})\|_2 \leq \frac{L_{dH}}{2m_d^2} \|\nabla d(\lambda_k)\|_2^2. \quad (16)$$

Hence, the following holds for $\bar{k} > 0$:

$$0 \geq f(\mathbf{z}_{\bar{k}}) - f_0(\mathbf{z}^*) \geq -\left(\frac{m_d}{4L_{dH}}\right)^{\bar{k}}. \quad (17)$$

Proof. We summarize the proof here into the following steps:

- **Newton Increment.** Similar to Newton decrement in primal space, in dual space, we define Newton increment for dual problem as following:

$$\begin{aligned} \nu(\lambda) &= [-\nabla d(\lambda)^T \nabla^2 d(\lambda)^{-1} \nabla d(\lambda)]^{\frac{1}{2}} \\ &= [-\Delta \lambda_{\text{nt}}^T \nabla^2 d(\lambda) \Delta \lambda_{\text{nt}}]^{\frac{1}{2}} \end{aligned} \quad (18)$$

where $\Delta \lambda_{\text{nt}} = -\nabla^2 d(\lambda)^{-1} \nabla d(\lambda) \geq 0$ is the Newton search direction in dual space.

- **Unit Step Size.** In this step, we show that the following condition leads to a unit step size in backtracking line search in dual space:

$$\|\nabla d(\lambda_k)\|_2 \leq \eta_d. \quad (19)$$

Note that due to space limitations, we omit some mathematical computations, which are similar to the convergence analysis proposed in [13, Chapter 9] for the proofs in primal space.

By Lipschitz continuity of dual Hessian $\nabla^2 d(\lambda)$, for $\rho \geq 0$, the following relation holds:

$$\|\nabla^2 d(\lambda + \rho \Delta \lambda_{\text{nt}}) - \nabla^2 d(\lambda)\|_2 \leq \rho L_{dH} \|\Delta \lambda_{\text{nt}}\|_2 \quad (20)$$

Multiply (20) by $\Delta \lambda_{\text{nt}}$ on both sides and obtain:

$$\begin{aligned} \|\Delta \lambda_{\text{nt}}^T [\nabla^2 d(\lambda + \rho \Delta \lambda_{\text{nt}}) - \nabla^2 d(\lambda)] \Delta \lambda_{\text{nt}}\|_2 \\ \leq \rho L_{dH} \|\Delta \lambda_{\text{nt}}\|_2^3 \end{aligned} \quad (21)$$

Define $\tilde{d}(\rho) = d(\lambda + \rho \Delta \lambda_{\text{nt}})$. Then $\nabla_\rho^2 \tilde{d}(\rho) = \Delta \lambda_{\text{nt}}^T \nabla^2 d(\lambda + \rho \Delta \lambda_{\text{nt}}) \Delta \lambda_{\text{nt}}$. Therefore, (21) can be written as:

$$\left| \nabla_\rho^2 \tilde{d}(\rho) - \nabla_\rho^2 \tilde{d}(0) \right| \leq \rho L_{dH} \|\Delta \lambda_{\text{nt}}\|_2^3 \quad (22)$$

Considering (19) and the strong convexity of function $-d(\lambda)$, by setting backtracking line search parameter $\alpha = \frac{1}{3}$ and integrating (22) three times, we can prove the following:

$$d(\lambda + \Delta \lambda_{\text{nt}}) \geq d(\lambda) + \alpha \nabla^2 d(\lambda)^T \Delta \lambda_{\text{nt}} \quad (23)$$

which shows that with the unit step $\rho = 1$ is accepted by backtracking line search in dual space.

- **Modified Condition.** Under Assumption 3, (7) can be modified as follows:

$$d(\lambda) = \min_{\mathbf{z}} f_0(\mathbf{z}) + \lambda^T g(\mathbf{z}) = \min_{\mathbf{z}} f_0(\mathbf{z}) + \lambda^T [g(\mathbf{z})]_+ \quad (24)$$

where we use the fact that in λ the entries corresponding to inactive constraints are zero. Therefore, the first-order derivative of the dual function $d(\lambda)$ for Problem (24) is:

$$\nabla d(\lambda) = [g(\mathbf{z})]_+ \quad (25)$$

Thus, the switching condition (19) becomes:

$$\|\nabla d(\lambda_k)\|_2 = \|[g(\mathbf{z})]_+\|_2 \leq \eta_d \quad (26)$$

- **Upper Bound on Duality Gap.** With the backtracking line search selecting unit step, by Lipschitz continuity,

we can derive the following relation of the first-order derivative of the dual function between two iterations:

$$\begin{aligned} \|\nabla d(\lambda_+)\|_2 &= \left\| \int_0^1 [\nabla^2 d(\lambda + \rho \Delta \lambda_{nt}) - \nabla^2 d(\lambda)] \Delta \lambda_{nt} d\rho \right\|_2 \\ &\leq \frac{L_{dH}}{2} \|\Delta \lambda_{nt}\|_2^2 = \frac{L_{dH}}{2} \|\nabla^2 d(\lambda)^{-1} \nabla d(\lambda)\|_2^2 \\ &\leq \frac{L_{dH}}{2m_d^2} \|\nabla d(\lambda)\|_2^2 \end{aligned} \quad (27)$$

Therefore, for $\bar{k} > 0$, recursively we can have

$$\frac{L_{dH}}{2m_d^2} \|\nabla d(\lambda_{\bar{k}})\|_2 \leq \left(\frac{L_{dH}}{2m_d^2} \|\nabla d(\lambda)\|_2 \right)^2 \quad (28)$$

By relying on strong convexity property of function $-d(\lambda)$ (which follows from the assumptions of the theorem), a crucial result on the duality gap can be obtained, that is,

$$f_0(\mathbf{z}^*) - f(\mathbf{z}_{\bar{k}}) \leq f_0(\mathbf{z}^*) - d(\lambda_{\bar{k}}) \leq \frac{1}{2m_d} \|\nabla d(\lambda_{\bar{k}})\|_2^2 \quad (29)$$

Finally, substituting (31) into (19), (19) into (28) and (28) into (29) concludes the proof. \square

Remark 1. The theorem above states that by relying on the DFG we can initialize the PDIP η_d -close to the central path. This allows the PDIP to "skip" the damped Newton phase and enter directly to pure Newton phase.

Remark 2. Algorithm 3 does not require backtracking line search given that the damped Newton phase has been replaced by a less computational demanding DFG phase. In this respect, note that the most computationally demanding step of the DFG is the (inner) minimization of the Lagrangian, which can be computed efficiently and up to a given accuracy (for details refer, for example, to [18] and the references within).

Algorithm 3 summarizes the proposed strategy to compute a solution for the MPC problem (2), in which the DFG (steps 3-9) is used to compute an η_d -solution that allows the PDIP (steps 10-17) to enter directly in the pure Newton phase and converge quadratically to the optimal solution of the MPC problem. When Algorithm 3 switches from the DFG phase to the pure Newton phase (step 5), we have to make sure that we preserve the information already computed by the DFG. Hence, the initialization strategy for the PDIP is important to ensure a successful switch. In this respect, step 5 of Algorithm 3 relies on the solution $(\mathbf{z}_{\text{DFG}}, \lambda_{\text{DFG}})$ returned by DFG as follows:

$$\zeta_0 = (\mathbf{z}_{\text{DFG}}, \lambda_{\text{DFG}}, s_{0,\text{PDIP}}), \quad (30a)$$

$$s_{0,\text{PDIP}} = -[g(\mathbf{z}_{\text{DFG}})]_- + [g(\mathbf{z}_{\text{DFG}})]_+. \quad (30b)$$

Equation (30b) provides an initialization for s that guarantees $s > 0$. Note that the information on the duality gap is fully fed into the pure Newton phase by (30b), according to the following

$$\begin{aligned} (s_{0,\text{PDIP}})^T \lambda_{0,\text{PDIP}} &= [-[g(\mathbf{z}_{\text{DFG}})]_- + [g(\mathbf{z}_{\text{DFG}})]_+]^T \lambda_{0,\text{PDIP}} \\ &= [g(\mathbf{z}_{\text{DFG}})]_+^T \lambda_{0,\text{PDIP}}, \end{aligned}$$

Algorithm 3 Proposed Solver.

- 1: Given $H, h, g, E, G, x(t), \hat{\lambda}, L_d, \eta_d, \varepsilon < \eta_d$.
 - 2: Initialize $\lambda_0 = \hat{\lambda}, k = 0$.
 - 3: **repeat**
 - 4: Compute $\mathbf{z}_k = \arg \min_{\mathbf{z}} \mathcal{L}(\mathbf{z}, \lambda_k)$.
 - 5: Compute $\hat{\lambda}_k = \left[\lambda_k + \frac{1}{L_d} \nabla d(\lambda_k) \right]_+$.
 - 6: Compute $\lambda_{k+1} = \frac{k+1}{k+3} \hat{\lambda}_k + \frac{2}{L_d(k+3)} \left[\sum_{j=0}^k \frac{j+1}{2} \nabla d(\lambda_j) \right]_+$.
 - 7: $k = k + 1$.
 - 8: **until** $\| [g(\mathbf{z})]_+ \|_2 \leq \eta_d$ (31)
 - 9: **return** $(\mathbf{z}_{\text{DFG}}, \lambda_{\text{DFG}})$
 - 10: Initialize ζ_0 according to (30) and $k = 0$.
 - 11: **repeat**
 - 12: Determine $\tau_{k+1} = \mu_{k+1} = \kappa \mu_k = (s_k)^T \lambda_k / m$.
 - 13: Compute search direction $\Delta \zeta_{\text{pd}}$ by solving (11).
 - 14: Update $\zeta_{k+1} = \zeta_k + \Delta \zeta_{\text{pd}}$.
 - 15: $k = k + 1$.
 - 16: **until** stopping criterion $\mu_k \leq \varepsilon$.
 - 17: **return** Point close to \mathbf{z}^*
-

where we use the fact that $[g(\mathbf{z}_{\text{DFG}})]_-^T \lambda_{0,\text{PDIP}} = 0$.

Remark 3. One could also try to derive the condition for PDIP to enter pure Newton phase in primal space directly when solving Problem (5). In this respect, in primal space, the convergence of PDIP can be analyzed using its similarity to the barrier method [13, Chapter 11]. The main limitation, however, is that the algorithm has to be initialized with $r_{\text{pri}} = 0$, that is, a strictly feasible starting point is required. This difficulty can be handled by either softening the constraints [19] in the problem solved by the PDIP or by tightening the constraints [18] in the problem solved by the DFG. These solutions are, however, more conservative. Furthermore, the switching condition in primal space involves the inverse of $g(\mathbf{z})$ that becomes undefined when there are constraints active at the optimum.

V. NUMERICAL RESULTS

The proposed solver (Algorithm 3) is tested both on the planar discrete-time linear unstable system in [20] and the open-loop unstable Cessna Citation Aircraft system in [14] (the descriptions of the two systems are omitted here for space limitations).

We tested the algorithm on a Windows OS, using an Intel(R) Core(TM) i7-4550 CPU (1.50-2.10 GHz) and RAM 8.00GB. The algorithms are implemented in Matlab.

A. Planar Linear System

The input and output of the system are subjected to the following constraints: $\|u\|_\infty \leq 1, \|y\|_\infty \leq 1$. Furthermore, Q and R are defined according to [20]. As mentioned in Section IV, L_{dH} can be selected as any real number greater than zero. In this respect, we chose $L_{dH} = 200$. Furthermore, $m_d = 1.4529$ and, consequently $\eta_d = 0.0106$. Finally, the backtracking line search parameters for Algorithm 2 are chosen as $\alpha = \frac{1}{3}, \beta = 0.5$.

We compared the following 4 scenarios:

- 1) Algorithm 2 is *warm-started* with $\mathbf{z}_0 := \mathbf{z}_{\text{LS}}$, which is the optimal solution of the unconstrained problem,

TABLE I
ITERATIONS OF ALGORITHMS 2 (COLUMNS 1-2) AND 3 (COLUMN 3).

	Scenario 1	Scenario 2	Scenario 3
Damped Newton Phase	6	16	–
Pure Newton Phase	20	8	6

that is, a least-squares (LS) problem, associated with Problem (5), and $\lambda_0 = \mathbf{1}_m$.

- 2) Algorithm 2 is *warm-started* with $\mathbf{z}_0 := \mathbf{z}_{LS}$ as in Scenario 1 and $\lambda_0 = 10^{-6}\mathbf{1}_m$.
- 3) Algorithm 3.
- 4) Algorithm 1.

Remark 4. Scenarios 1 and 2 rely on two different initializations of the dual variables to show the impacts that their initialization have on the behavior of the solver. To the best of our knowledge, many off-the-shelf interior-point solvers do not allow the user to access the dual variables for their initialization (e.g., MATLAB’s *quadprog*). A default choice is the one proposed in Scenario 1 in which the dual variables are initialized to $\mathbf{1}$. The warm-starting strategy proposed in Scenario 2 considers λ_0 close to zero, which is the optimal value of the multipliers associated with the unconstrained LS problem.

Table I compares Algorithm 3 (Scenario 3) with Algorithm 2 (Scenarios 1 and 2) in terms of Newton iterations. Algorithm 2 requires 26 iterations in Scenario 1, while it requires 24 iterations in Scenario 2. Notice, however, that, while the total number of iterations in Scenario 2 is reduced, Algorithm 2 requires more damped-Newton-phase iterations. The main reason is that the initialization in Scenario 2 is farther from the optimal value, given that it is initialized with the pair $(\mathbf{z}_{LS}, \lambda_0 \approx 0)$, which assumes no active constraints (while constraints are active at the optimum). As Table I shows (Scenario 3), Algorithm 3 reduces the number of iterations to 6 and completely eliminates the damped Newton phase, thanks to the use of the DFG to initialize the interior-point iterates. This leads to significant improvements also from the computation point of view as Figure 1 depicts. In particular, Figure 1 shows the computation time required to solve Problem (2) online using the proposed algorithm. In Figure 1, $f_0(\mathbf{z}^*)$ is the optimal cost corresponding to \mathbf{z}^* , computed by MATLAB’s *quadprog*. PDIP terminates when the duality gap satisfies $\mu \leq \varepsilon = 10^{-6}$. The blue solid line shows the computation time of Algorithm 2 in Scenario 1. The light-blue solid line is the computation time of Algorithm 2 in Scenario 2. The red solid line shows the computation time of Algorithm 1 (Scenario 4). The black bold line shows our proposed solver. The green star indicates the switching point when the solution returned by DFG satisfies condition (31) ($\approx 10^{-3}$). Note that, Scenario 4 requires 591 iterations of Algorithm 1 to reach a high accuracy ($\approx 10^{-6}$), while Scenario 3 only requires 65 DFG iterates to reach the medium accuracy needed and enter the pure Newton phase in Algorithm 3. Furthermore, note that the improvements in terms of computation time compared

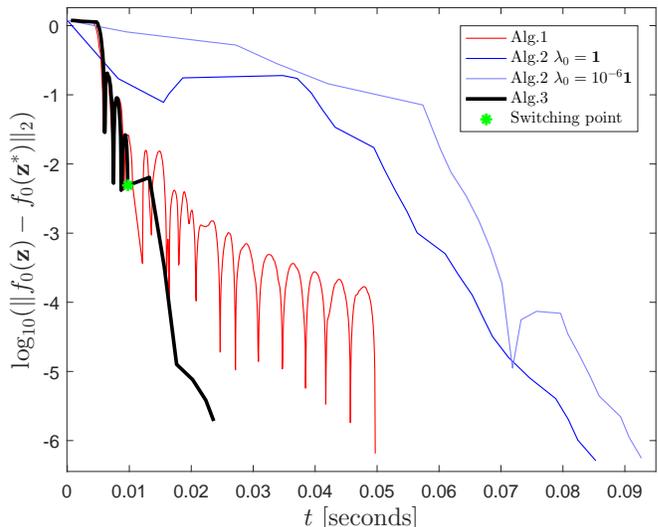


Fig. 1. Time required for online optimization using Algorithms 1, 2, and 3 for the 2-dimensional planar system.

to Scenarios 1 and 2 are related to the following facts: (i) our algorithm does not require backtracking line search (computationally costly), and (ii) the damped Newton phase is replaced by DFG iterates.

Remark 5. Figure 1 monitors $\|f_0(\mathbf{z}_k) - f_0(\mathbf{z}^*)\|_2$ (in logarithmic scale). When Algorithm 2 is initialized with the optimal solution $\mathbf{z}_0 := \mathbf{z}_{LS}$ (Scenarios 1 and 2), $f_0(\mathbf{z}_0) < f_0(\mathbf{z}^*)$, given that it does not account for the presence of (active) constraints at the optimum. This leads to the nonmonotonic behavior at 0.02 sec (blue line) and 0.08 sec (light-blue line) in Figure 1, when PDIP enters the feasible region (and in pure Newton phase).

B. Cessna Citation Aircraft [14]

The aircraft is flying at an altitude of 5000 km and a speed of 128.2 m/sec. The system has 4 states, 1 input, 2 outputs and is subject to the following constraints: elevator angle $\pm 15^\circ$, elevator rate $\pm 30^\circ/\text{sec}$ and pitch angle $\pm 30^\circ$. To discretize the system, we use a sampling time of $T = 0.25$ sec. Moreover, Q is identity matrix, $R = 10$ and $N = 10$. For this problem, $m_d = 1.1394 \times 10^{-4}$ and $L_d = 5 \times 10^{-7}$. Thus, $\eta_d = 2.6 \times 10^{-2}$. The backtracking line search parameters for Algorithm 2 are chosen as $\alpha = \frac{1}{3}$, $\beta = 0.5$.

We test the proposed solver at 30 different initial states sampled uniformly in the feasible region $[0, 60\text{km}]$ of altitude (other states fixed). For each state, the solver is run for 11 times and the median of the computational time is computed. In order to examine how much noise is in the measurements of computational time, the standard deviation is computed for each initial condition.

Table II compares the three algorithms in terms of computational time. Each solver is supposed to reach the desired accuracy on duality gap $\mu \leq \varepsilon = 10^{-6}$. In the best case scenario, Algorithm 3 (15 iterations in DFG phase and 12 iterations in the pure Newton phase) only consumes 26.95% of

TABLE II
COMPUTATIONAL TIME (SEC) OF ALGORITHMS 1, 2, AND 3

	Alg. 1	Alg. 2	Alg. 3
<i>Best</i>	0.0063	0.0141	0.0024
<i>Worst</i>	–	0.0398	0.0214
<i>Average</i>	–	0.0264	0.0120
<i>Average Deviation</i>	1.4×10^{-3}	3.7×10^{-4}	3.6×10^{-4}

the computational time of Algorithm 2 (22 Newton iterations, 10 in the damped Newton phase), and 60.32% of Algorithm 1 (280 iterations). In the worst case, Algorithm 3 (352 iterations in DFG phase and 11 iterations in the pure Newton phase) saves 46% of the computational time consumed by Algorithm 2 (36 Newton iterations, 25 in damped Newton phase). Note that the execution of Algorithm 1 is terminated when it exceeded a given threshold (such as maximal number of iterations). On average, Algorithm 3 saves more than 54% of the computational time required by Algorithm 2 thanks to elimination of damped Newton phase.

VI. CONCLUSIONS AND FUTURE WORK

This paper proposes an improved primal-dual interior point (PDIP) method for optimization problems that typically arise from model predictive control applications, that is, quadratic programming problems with linear inequality constraints. The proposed solver improves the convergence of state-of-the-art PDIP methods by replacing the damped Newton phase with a dual fast gradient method. This result is obtained by working in the dual space and modifying the condition to enter the pure Newton phase. Finally, we showed the benefits of the proposed algorithm on a discrete unstable planar system and the Cessna Citation Aircraft system.

As part of the future work, we plan to test the technique by comparing with other solvers, such as, QPOASES, CPLEX, GUROBI.

REFERENCES

- [1] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, “The explicit linear quadratic regulator for constrained systems,” *Automatica*, vol. 38, no. 1, pp. 3–20, 2002.
- [2] Y. Nesterov, “A method of solving a convex programming problem with convergence rate $O(1/k^2)$,” in *Soviet Mathematics Doklady*, vol. 27, no. 2, 1983, pp. 372–376.
- [3] M. Kogel and R. Findeisen, “Fast predictive control of linear systems combining Nesterov’s gradient method and the method of multipliers,” in *Proc. of the 50th IEEE CDC*. IEEE, 2011, pp. 501–506.
- [4] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*. Springer Science & Business Media, 2013, vol. 87.
- [5] G. Stathopoulos, H. Shukla, A. Szucs, Y. Pu, and C. N. Jones, “Operator splitting methods in control,” *Foundations and Trends® in Systems and Control*, 2016.
- [6] N. L. Ricker, “Use of quadratic programming for constrained internal model control,” *Industrial & Engineering Chemistry Process Design and Development*, vol. 24, no. 4, pp. 925–936, 1985.
- [7] C. Schmid and L. T. Biegler, “Quadratic programming methods for reduced hessian SQP,” *Computers & chemical engineering*, vol. 18, no. 9, pp. 817–832, 1994.
- [8] H. J. Ferreau, “An online active set strategy for fast solution of parametric quadratic programs with applications to predictive engine control,” *University of Heidelberg*, 2006.
- [9] C. V. Rao, S. J. Wright, and J. B. Rawlings, “Application of interior-point methods to model predictive control,” *Journal of optimization theory and applications*, vol. 99, no. 3, pp. 723–757, 1998.

- [10] Y. Wang and S. Boyd, “Fast model predictive control using online optimization,” *IEEE Transactions on CST*, vol. 18, no. 2, pp. 267–278, 2010.
- [11] M. N. Zeilinger, D. M. Raimondo, A. Domahidi, M. Morari, and C. N. Jones, “On real-time robust model predictive control,” *Automatica*, vol. 50, no. 3, pp. 683–694, 2014.
- [12] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems*, 2015, available at www.mpc.berkeley.edu/mpc-course-material.
- [13] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [14] J. M. Maciejowski, *Predictive control: with constraints*. Pearson education, 2002.
- [15] Y. Nesterov, “Smooth minimization of non-smooth functions,” *Mathematical programming*, vol. 103, no. 1, pp. 127–152, 2005.
- [16] P. Giselsson, “Improved fast dual gradient methods for embedded model predictive control,” *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 2303–2309, 2014.
- [17] I. Necoara and V. Nedelcu, “Rate analysis of inexact dual first-order methods application to dual decomposition,” *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1232–1243, 2014.
- [18] I. Necoara, L. Ferranti, and T. Keviczky, “An adaptive constraint tightening approach to linear model predictive control based on approximation algorithms for optimization,” *OCAM*, vol. 36, no. 5, pp. 648–666, 2015.
- [19] E. C. Kerrigan and J. M. Maciejowski, “Soft constraints and exact penalty functions in model predictive control,” in *International Conference (Control 2000)*, Cambridge, 2000.
- [20] M. Rubagotti, P. Patrinos, and A. Bemporad, “Stabilizing linear model predictive control under inexact numerical optimization,” *IEEE TAC*, vol. 59, no. 6, pp. 1660–1666, 2014.