Noname manuscript No.
(will be inserted by the editor)

Exploitation of the Quasi-structure of the Dual Hessian for distributed MPC with Non-delayed Couplings

Emil Klintberg · **Sebastien Gros**

Received: date / Accepted: date

Abstract Bla bla bla

Keywords First keyword · Second keyword · More

1 Introduction

Recently there has been a great interest invested in controlling large-scale, networked systems with Model Predictive Control (MPC). Many approaches have emerged, with the common feature that they allow for distributed computations with a limited amount of communication. This is indeed an important property since a large-scale problem easily can be intractable for a single computer. Moreover, when the system is geographically distributed, or when subsystems do not want to share sensitive information, it can be highly impractical to centralize the data of the problem.

Distributed MPC schemes have been proposed for various classes of control problems. In [19], a distributed multiple shooting method is proposed to tackle large-scale nonlinear MPC problems, in [27], a method to tackle distributed stochastic MPC problems is proposed, whereas in [3] distributed robust MPC is discussed.

Formulations of distributed MPC schemes can be divided into categories depending on the degree of communication between subsystems. Methods where no communication between the local controllers are allowed are commonly denoted as decentralized MPC. Methods belonging to this class offer

Emil Klintberg
first address
Tel.: +123-45-678910
Fax: +123-45-678910
E-mail: fauthor@example.com

Sebastien Gros
second address

great scalability, and can give a good performance when the subsystems are weakly coupled. However, when this is not fulfilled, the performance can be far from optimal.

Methods where communication between subsystems are allowed, are commonly referred to as cooperative/non-cooperative MPC depending on the cost function used in the local MPC controllers. In [14], autonomous vehicles are controlled through traffic intersections.

Another promising approach to achieve distributed MPC is to formulate the MPC problem centrally and solve the underlaying optimization problem with techniques from the field of distributed optimization. A commonly used technique for this purpose is to use a *Lagrangian relaxation* to decompose the optimization problem into low-dimensional subproblems that can be solved independently.

Lagrangian relaxation is used in many different contexts to tackle convex large-scale problems, e.g. the authors in [4] propose a coordinate ascent approach to solve matrix problems. In [20], [11] a gradient method, whereas in [28], [22], [10], [9] a fast gradient method is used in order to attain dual optimality. Moreover, in [21], [23], convergence rates for first-order methods are studied. All these methods make use of only first-order derivatives when finding a search direction.

Methods using second-order information offer advantages in terms of convergence speed. In the context of active-set methods, this is exploited by the authors in [7], [17], [18], [8]. In [12], it is shown how methods in this class can be used to control wind farms with nonlinear MPC. However, the dual Hessian can be singular if a poor initial guess is used for the dual variables, and hence possibly leading to an inconsistent Newton system. In [18], this issue was avoided by relaxing the local inequality constraints with an L_2-penalty.

In [24], it was proved that adding self-concordant barrier terms to the Lagrange function of a generic convex problem render a self-concordant dual function. Hence, in the context of interior-point methods, it is possible to use the Newton method to efficiently trace the central path. This result was used in [29], where a method based on inexact solutions of the local subproblems was proposed. Moreover, in [15], a primal-dual interior-point method was proposed, which supports linear predictors and inexact solutions and showed a fast and consistent practical convergence.

Methods using second-order information offer a fast convergence in terms of iterations. However, this comes to the cost of factorizing the Hessian of the dual function at each iteration. For a high-dimensional dual space, this can be a limitation for second-order methods.

In this paper, we aim at substantially lower the complexity of factorizing the dual Hessian by exploiting its inherent quasi-structure originating from certain MPC schemes. The quasi structure arise when distributed systems, with non-delayed interactions between subsystems, are controlled with MPC. As an illustrative example, the structure will be analyzed in the context of the method described in [24]. It should however be understood that the dual Hessian of the methods described in [29] and [15] take the same form.

1.1 Contributions

Bla bla bla

1.2 Notations

We note the space of symmetric positive (semi) definite matrices by ($\mathbb{S}^n_+$) $\mathbb{S}^n_{++} \subset \mathbb{R}^{n \times n}$.

2 Prelimenaries

2.1 Problem formulation

In this section, we present the considered class of distributed Quadratic Programs (QPs). Moreover, we show how problems belonging to this class arise when MPC is used to control several linear systems interacting via non-delayed couplings.

Let us consider M discrete time, possibly time varying, linear systems:

$$x_{k,i+1} = A_{k,i}x_{k,i} + B_{k,i}u_{k,i}, \quad k = 1, \ldots, M \tag{1}$$

where $x_{k,i} \in \mathbb{R}^{p_k}$ and $u_{k,i} \in \mathbb{R}^{m_k}$ represents the state and input of system k at time instance i. Moreover, each system are subject to hard state and input constraints:

$$x_{k,i} \in \mathcal{X}_{k,i} \subseteq \mathbb{R}^{p_k}, \quad k = 1, \ldots, M \tag{2a}$$

$$u_{k,i} \in \mathcal{U}_{k,i} \subseteq \mathbb{R}^{m_k}, \quad k = 1, \ldots, M \tag{2b}$$

Let us now also assume that the systems interact with static, non-delayed couplings:

$$h_i(x_{1,i}, u_{1,i}, \ldots, x_{M,i}, u_{M,i}) = 0 \tag{3}$$

Observe that with non-delayed couplings, we refer to interactions where states and inputs for a subsystem at time i can only have a direct impact on states and inputs of another subsystem at time i. The interactions between the systems can accordingly not include dynamical relationships or time-delays. However, it should be understood that this can always be fulfilled by introducing extra states to handle the delays locally, even when delayed couplings are present in the original formulation.

Furthermore, let us assume that the the constraints (2) are convex polytopes, and that the non-delayed couplings (3) are affine. Moreover, we introduce the notation: $z_{k,i} = [x_{k,i}^T \quad u_{k,i}^T]^T \in \mathbb{R}^{n_k}$, representing the optimization variables. Using a quadratic objective function, the *finite-time optimal control problem* over the horizon N, can then be stated as:

$$\begin{aligned}
\min_{z} \quad & \sum_{k=1}^{M}\sum_{i=0}^{N} \frac{1}{2} z_{k,i}^T H_{k,i} z_{k,i} + c_{k,i}^T z_{k,i} && \text{(4a)}\\
\text{s.t.} \quad & \sum_{k=1}^{M} F_{k,i} z_{k,i} = e_i && \text{(4b)}\\
& C_{k,i} z_{k,i} + D_{k,i+1} z_{k,i+1} = d_{k,i} && \text{(4c)}\\
& G_{k,i} z_{k,i} \le f_{k,i} && \text{(4d)}
\end{aligned}$$

where $H_{k,i} \in \mathbb{S}_{++}^{n_k}$ and $c_{k,i} \in \mathbb{R}^{n_k}$, the dynamics (1) are described by $C_{k,i} \in \mathbb{R}^{l_k \times n_k}$, $D_{k,i+1} \in \mathbb{R}^{l_k \times n_k}$ and $d_{k,i} \in \mathbb{R}^{l_k}$, $F_{k,i} \in \mathbb{R}^{r_i \times n_k}$ and $e_i \in \mathbb{R}^{r_i}$ yield the non-delayed coupling constraints (3), $G_{k,i} \in \mathbb{R}^{t_k \times n_k}$ and $f_{k,i} \in \mathbb{R}^{t_k}$ form the local state and input constraints (2).

Additionaly, to avoid unnecessarily heavy notation, we introduce the following augmented notations: $z_k = [z_{k,0}^T \dots, z_{k,N}^T]^T \in \mathbb{R}^{(N+1)n_k}$, which represents the collection of states and inputs of system k over the horizon N. The MPC problem (4) can then be expressed as:

$$\begin{aligned}
\min_{z} \quad & \sum_{k=1}^{M} \frac{1}{2} z_k^T H_k z_k + c_k^T z_k && \text{(5a)}\\
\text{s.t.} \quad & \sum_{k=1}^{M} F_k z_k = e && \text{(5b)}\\
& C_k z_k = d_k && \text{(5c)}\\
& G_k z_k \le f_k && \text{(5d)}
\end{aligned}$$

where $H_k \in \mathbb{S}_{++}^{(N+1)n_k \times (N+1)n_k}$, $C_k \in \mathbb{R}^{Nl_k \times (N+1)n_k}$ and $d_k \in \mathbb{R}^{Nl_k}$, $F_k \in \mathbb{R}^{(N+1)r \times (N+1)n_k}$ and $e \in \mathbb{R}^{(N+1)r}$, $G_k \in \mathbb{R}^{(N+1)t_k \times (N+1)n_k}$ and $f_k \in \mathbb{R}^{(N+1)t_k}$.

Observe that the matrices in (5) possess banded structures given by:

$$H_k = \begin{bmatrix} H_{k,0} & & \\ & \ddots & \\ & & H_{k,N} \end{bmatrix},$$

$$F_k = \begin{bmatrix} F_{k,0} & & \\ & \ddots & \\ & & F_{k,N} \end{bmatrix},$$

$$C_k = \begin{bmatrix} C_{k,0} & D_{k,1} & & & \\ & C_{k,1} & D_{k,2} & & \\ & & \ddots & \ddots & \\ & & & C_{k,N-1} & D_{k,N} \end{bmatrix},$$

$$A_k = \begin{bmatrix} A_{k,0} & & \\ & \ddots & \\ & & A_{k,N} \end{bmatrix}.$$

The banded structures will be key elements for our results.

2.2 Dual decomposition with second-order information

In this section, we consider dual decomposition. The aim is to present a constraint relaxation, first described in [24], in order to make the dual function self-concordant. Self-concordance is an important property for constructing polynomial time interior-point methods [25], and is hence a desirable feature since Newton steps in the dual space are considered. Moreover, since constructing the Hessian is crucial to compute a Newton step, we will detail how the dual Hessian can be computed.

We introduce the dual variables $\lambda \in \mathbb{R}^{(N+1)r}$ corresponding to the coupling constraints (5b) and define the Lagrange function as:

$$\mathcal{L}(z, \lambda) = \sum_{k=1}^{M} (\frac{1}{2} z_k^T H_k z_k + c_k^T z_k) + \lambda^T (\sum_{k=1}^{M} F_k z_k - e) \tag{7}$$

We observe that $\mathcal{L}(z, \lambda)$ is separable in z, i.e.

$$\mathcal{L}(z, \lambda) = \sum_{k=1}^{M} \mathcal{L}_k(z_k, \lambda), \tag{8}$$

with

$$\mathcal{L}_k(z_k, \lambda) = \frac{1}{2} z_k^T H_k z_k + c_k^T z_k + \lambda^T (F_k z_k - \frac{1}{M} e), \tag{9}$$

such that the dual function $d(\lambda) = -\min_{z\in\mathcal{Z}} \mathcal{L}(z,\lambda)$ can be evaluated in parallel as:

$$d(\lambda) = -\sum_{k=1}^{M} \min_{z_k\in\mathcal{Z}_k} \mathcal{L}_k(z_k,\lambda) \tag{10}$$

Since (5) is strictly convex, $d(\lambda)$ is convex and continuously differentiable, but not twice differentiable [18]. However, the Hessian of $d(\lambda)$ is a constant matrix over any fixed active-set of (4) [18].

As a result, $d(\lambda)$ is not self-concordant and the solution to the dual problem is not easily tracked with the Newton method. However, if the inequality constraints (5d) are relaxed with a self-concordant log-barrier, according to:

$$\begin{aligned}
\min_{z} \quad & \sum_{k=1}^{M} \frac{1}{2} z_k^T H_k z_k + c_k^T z_k - \tau \sum_{i=1}^{(N+1)t_k} \log([s_k]_i) && \text{(11a)}\\
\text{s.t.} \quad & \sum_{k=1}^{M} F_k z_k = e && \text{(11b)}\\
& C_k z_k = d_k && \text{(11c)}\\
& G_k z_k + s_k = f_k && \text{(11d)}
\end{aligned}$$

where $\tau > 0$ will be referred to as the *barrier parameter*, the resulting *relaxed dual function* $d(\lambda,\tau)$ is self-concordant [24]. As a result, a sequence of dual problems $\{\min_\lambda d(\lambda,\tau)\}_{\tau\to 0}$ can be solved, where each problem is self-concordant and therefore easily solved with the Newton method.

In contrast to smoothing techniques based on the *augmented Lagrangian* [16], the separability of the dual function is preserved, and the relaxed dual function can be computed in parallel using:

$$d(\lambda,\tau) = -\sum_{k=1}^{M} \min_{z_k\in\mathcal{Z}_k} \left(\mathcal{L}_k(z_k,\lambda) - \tau \sum_{i=1}^{m_k} \log([s_k]_i) \right) \tag{12}$$

Observe that evaluating (12) involves solving M independent subproblems of the form:

$$\begin{aligned}
\min_{z_k} \quad & \frac{1}{2} z_k^T H_k z_k + c_k^T z_k + \lambda^T F_k z_k - \tau \sum_{i=1}^{m_k} \log([s_k]_i)\\
\text{s.t.} \quad & C_k z_k = d_k\\
& G_k z_k + s_k = f_k\\
& s_k \geq 0
\end{aligned} \tag{13}$$

Strict convexity of (11) implies that the gradient of $d(\lambda,\tau)$ with respect to λ is given by the residual of the coupling constraints [2], i.e.

$$\nabla d(\lambda,\tau) = -\sum_{k=1}^{M} F_k z_k^*(\lambda,\tau) + e \tag{14}$$

where $z_k^*(\lambda, \tau)$ is the solution to (13). The dual Hessian is then given by:

$$\nabla^2 d(\lambda, \tau) = -\sum_{k=1}^{M} F_k \frac{\partial z_k^*(\lambda, \tau)}{\partial \lambda} \tag{15}$$

Recall that in order to find the solution $\lambda^*(\tau)$ to the dual problem $\min_\lambda d(\lambda, \tau)$, λ is updated by taking Newton steps $\Delta\lambda$. A Newton step in the dual space can be obtained as a solution to the Newton system:

$$\nabla^2 d(\lambda, \tau)\Delta\lambda + \nabla d(\lambda, \tau) = 0 \tag{16}$$

Observe that finding the Newton step requires knowledge of the gradient and the Hessian of the dual function $d(\lambda, \tau)$. Note that the dual gradient (14) can be easily computed from the solutions $z_k^*(\lambda, \tau)$ to (13), while the Hessian (15) requires the sensitivities $\frac{\partial z_k^*(\lambda,\tau)}{\partial \Lambda}$. In the following, we will describe how $\frac{\partial z_k^*(\lambda,\tau)}{\partial \Lambda}$ can be computed, for a detailed description we refer to [15].

By introducing $y_k = \tau s_k^{-1} \in \mathbb{R}^{t_k(N+1)}$, the primal-dual interior point KKT conditions to (13) can be stated as:

$$0 = H_k z_k^* + c_k + F_k^T \lambda + C_k^T \mu_k^* + G_k^T y_k^* \tag{17a}$$

$$0 = C_k z_k^* - d_k \tag{17b}$$

$$0 = G_k z_k^* + s_k^* - f_k \tag{17c}$$

$$0 = Y_k^* s_k^* - \tau \mathbf{1} \tag{17d}$$

$$s_k^* > 0, \quad y_k^* > 0 \tag{17e}$$

By differentiating (17), the following linear system is obtained [15]:

$$\begin{bmatrix} H_k & C_k^T & G_k^T & 0 \\ C_k & 0 & 0 & 0 \\ G_k & 0 & 0 & I \\ 0 & 0 & S_k & Y_k \end{bmatrix} \begin{bmatrix} \frac{\partial z_k^*}{\partial \lambda} \\ \frac{\partial \mu_k^*}{\partial \lambda} \\ \frac{\partial y_k^*}{\partial \lambda} \\ \frac{\partial s_k^*}{\partial \lambda} \end{bmatrix} = - \begin{bmatrix} F_k^T \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{18}$$

from which solution, $\frac{\partial z_k^*(\lambda,\tau)}{\partial \lambda}$, the dual Hessian can be formed according to (15).

3 Quasi-structure of the dual Hessian

3.1 Quasi-banded

In Section 2.2, we presented a way of computing the dual Hessian. In this section, we will detail its properties by forming an explicit expression and characterizing a quasi-banded structure. The quasi-banded structure is important since it implies that the dual Hessian can be approximated by a banded matrix, which can be used to form a quasi-Newton method.

The analysis of the structure will be based on the explicit expression of the dual Hessian in the following minor propsition:

Proposition 1 *The Hessian of the dual function $d(\lambda, \tau)$ is given by:*

$$\nabla^2 d(\lambda, \tau) = \sum_{k=1}^{M} F_k (\Phi_k^{-1} - \Phi_k^{-1} C_k^T \Lambda_k^{-1} C_k \Phi_k^{-1}) F_k^T \tag{19}$$

where

$$\Phi_k = H_k + G_k^T S_k^{-1} Y_k G_k \tag{20}$$

$$\Lambda_k = C_k \Phi_k^{-1} C_k^T \tag{21}$$

Proof By using block elimination of (18), it can be obtained that:

$$\Lambda_k(s_k, y_k) \frac{\partial \mu_k^*}{\partial \lambda} = -C_k \Phi_k^{-1} F_k^T \tag{22a}$$

$$\Phi_k(s_k, y_k) \frac{\partial z_k^*}{\partial \lambda} = -F_k^T - C_k^T \frac{\partial \mu_k^*}{\partial \lambda} \tag{22b}$$

where

$$\Phi_k(s_k, y_k) = H_k + G_k^T S_k^{-1} Y_k G_k \in \mathbb{S}_{++}^{(N+1)n_k \times (N+1)n_k} \tag{23a}$$

$$\Lambda_k(s_k, y_k) = C_k \Phi_k^{-1}(s_k, y_k) C_k^T \in \mathbb{S}_{++}^{Nl_k \times Nl_k} \tag{23b}$$

By using (22a) to eliminate $\frac{\partial \mu_k^*}{\partial \lambda}$ in (22b), the following is obtained:

$$F_k \frac{\partial z_k^*}{\partial \lambda} = -F_k (\Phi_k^{-1} - \Phi_k^{-1} C_k^T \Lambda_k^{-1} C_k \Phi_k^{-1}) F_k^T \tag{24}$$

where we, for the sake of brevity, have omitted the arguments and used $\Phi_k = \Phi_k(s_k, y_k)$ and $\Lambda_k = \Lambda_k(s_k, y_k)$. As a consequence, the dual Hessian (15) can be written as

$$\nabla^2 d(\lambda, \tau) = \sum_{k=1}^{M} F_k (\Phi_k^{-1} - \Phi_k^{-1} C_k^T \Lambda_k^{-1} C_k \Phi_k^{-1}) F_k^T \tag{25}$$

□

Remark 1 Observe, that the relaxed dual Hessian of a problem distributed in space according to (5) can always be written in the form (19), regardless of the additional structure in time described by (4). However, by considering the structure in time, it can be observed from (20) that Φ_k is block diagonal with blocks $\Phi_{k,i} = H_{k,i} + G_{i,k}^T S_{k,i}^{-1} Y_{k,i} G_{k,i}$, and as a consequence, Λ_k has a block tridiagonal structure given by:

$$\Lambda_k = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & & \\ \Lambda_{12}^T & \Lambda_{22} & \ddots & \\ & \ddots & \ddots & \Lambda_{N-1,N} \\ & & \Lambda_{N-1,N}^T & \Lambda_{N,N} \end{bmatrix} \in \mathbb{S}_{++}^{Nl_k \times Nl_k} \tag{26}$$

where

$$\Lambda_{i,i} = C_{k,i-1}\Phi_{k,i-1}^{-1}C_{k,i-1}^T + D_{k,i}\Phi_{k,i}^{-1}D_{k,i}^T \in \mathbb{R}^{l_k, \times l_k} \tag{27a}$$

$$\Lambda_{i,i+1} = D_{k,i}\Phi_{k,i}^{-1}C_{k,i}^T \in \mathbb{R}^{l_k \times l_k} \tag{27b}$$

Note that all matrices in (19) are banded except Λ_k^{-1} which in general is dense, and consequently, makes also the dual Hessian dense. However, inverses of sparse matrices typically contain many elements that are small in magnitude [1]. In the following, we will show that Λ_k^{-1} possess a decaying property which will translate to a *quasi-banded structure* in $\nabla^2 d(\lambda, \tau)$. In the analysis we consider two assumptions. First, we assume boundedness of the problem data:

Assumption 1 *The row and column absolute sums of Jacobians of equality constraints (i.e. (5b) and (5c)) are bounded. Hence:*

1. $\|C_{k,i}\|_\bullet \leq \gamma$
2. $\|D_{k,i}\|_\bullet \leq \gamma$
3. $\|F_{k,i}\|_\bullet \leq \gamma$

where $\bullet$ *represent* ∞ *and* 1*.*

For the sake of simplicity, a single constant γ is chosen to bound all problem data. Moreover, it should be observed that Assumption 1 is not by any means restrictive, since any solver would struggle with a problem where it is not fulfilled []. Furthermore, we assume boundedness of Φ_k^{-1}:

Assumption 2 *The row and column absolute sums of* Φ_k^{-1} *are bounded. Hence:*

$$\|\Phi_{k,i}^{-1}\|_\bullet \leq \gamma_{\Phi_k^{-1}}, \quad \forall i \tag{28}$$

where $\bullet$ *represents* ∞ *and* 1*.*

Remark 2 As a result of Φ_k being symmetric, $\|\Phi_{k,i}^{-1}\|_\infty = \|\Phi_{k,i}^{-1}\|_1$.

Remark 3 A possible difficulty with interior-point methods stems from numerical difficulties due to ill-conditioning of the normal equations [30]. Indeed, since elements in $S_k^{-1}Y_k$ can be close to zero in late iterations, i.e. when τ is small. However, since we assume that (4) is strongly convex, the eigenvalues of Φ_k are lower bounded by the smallest eigenvalue of H_k, even when $S_k^{-1}Y_k$ is singular. According to this reasoning it is not restrictive to assume boundedness of Φ_k^{-1}.

Let us start by characterizing the structure of Λ_k^{-1}. Indeed, inverses of sparse matrices are in general dense, but they typically contain many elements that are small in magnitude [1]. Since Λ_k is banded, symmetric and positive definite, we will use the following classical result:

Lemma 1 *If* A *is Hermitian positive definite and* m*-banded, i.e.* $[A]_{ij} = 0$ *if* $|i-j| > m$*, the entries of* A^{-1} *satisfy the following bound:*

$$|[A^{-1}]_{ij}| < K\omega^{|i-j|}, \quad \forall i, j \tag{29}$$

where $\sigma_{\min}(A)$ *and* $\kappa(A)$ *are the the smallest singular value and the condition number of* A *respectively,* $K = \max\{\sigma_{\min}^{-1}(A), K_0\}$, $K_0 = (1 + \sqrt{\kappa(A)})$, $\omega = \left(\frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1}\right)^{1/m}$.

Proof A proof is given in [5]. □

Remark 4 The entries in A^{-1} are bounded by an exponentially decaying function along each row or column. However, the bound depends on the condition number and the bandwidth of the matrix. Matrices with a high condition number and/or a high bandwidth can therefore result in a large K and $\omega \approx 1$, leading to a slow decay. A low condition number and a small band results in a rapid decay.

Remark 5 Observe that due to its block tridiagonal structure, Λ_k is $3l_k$-banded, i.e. $m = 3l_k$ for Λ_k.

Accordingly, if we introduce the notation

$$\Lambda_k^{-1} = \begin{bmatrix} T_{11} & T_{21}^T & \dots & T_{N+1,1}^T \\ T_{21} & T_{22} & \dots & T_{N+1,2}^T \\ \vdots & \vdots & \ddots & \vdots \\ T_{N+1,1} & T_{N+1,2} & \dots & T_{N+1,N+1} \end{bmatrix} \tag{30}$$

where $T_{i,j} \in \mathbb{R}^{l_k \times l_k}$, we can establish the following proposition:

Proposition 2 *The off-diagonal blocks (i.e.* $T_{i,j}$ *where* $i - j > 0$*) in* Λ_k^{-1} *satisfy the following bounds:*

$$\|T_{i,j}\|_\bullet \leq K_{\Lambda_k} \omega_{\Lambda_k}^{i-j} \tag{31}$$

where $\bullet$ *represents* ∞ *and* 1, $\sigma_{\min}(\Lambda_k)$ *and* $\kappa(\Lambda_k)$ *are the the smallest singular value and the condition number of* Λ_k *respectively,* $K_{\Lambda_k} = \max\{\sigma_{\min}(\Lambda_k)^{-1}, 1 + \sqrt{\kappa(\Lambda_k)}\} l_k \omega_{\Lambda_k}^{1/l_k}$ *and* $\omega_{\Lambda_k} = \left(\frac{\sqrt{\kappa(\Lambda_k)}-1}{\sqrt{\kappa(\Lambda_k)}+1}\right)^{\frac{1}{3}}$.

Proof According to Lemma 1, the element in $T_{i,j}$ with the largest bound is located in the top-right corner, and is hence the element $[\Lambda_k^{-1}]_{il_k+1, jl_k}$ in Λ_k^{-1}. By directly applying Lemma 1 it follows that:

$$\max |[T_{ij}]| \leq \max\{\sigma_{min}(\Lambda_k)^{-1}, 1 + \sqrt{\kappa(\Lambda_k)}\} \left(\frac{\sqrt{\kappa(\Lambda_k)} - 1}{\sqrt{\kappa(\Lambda_k)} + 1}\right)^{\frac{1}{3l_k}((i-j)l_k+1)} \tag{32}$$

where $\max |[T_{ij}]|$ refers to the maximum absolute value of the components in T_{ij}. Moreover, since there are l_k elements in each row or column of a block $T_{i,j}$, we obtain:

$$\|T_{i,j}\|_\bullet \leq l_k \max\{\sigma_{min}(\Lambda_k)^{-1}, 1 + \sqrt{\kappa(\Lambda_k)}\} \left(\frac{\sqrt{\kappa(\Lambda_k)} - 1}{\sqrt{\kappa(\Lambda_k)} + 1}\right)^{\frac{1}{3l_k}((i-j)l_k+1)} \tag{33}$$

where $\bullet$ represents ∞ and 1. Observe that (33) is identical to (31). □

Remark 6 The constants K_{Λ_k} and ω_{Λ_k} in (31) depend heavily on the conditioning of Λ_k, which for interior point methods can in general be very high for small values of τ. Our experience is however that this is not a major issue in practice, which is also supported by commercial interior-point implementations [30] [6]. Moreover, it should be understood that strong convexity of (4), should improve the worst case conditioning of Φ_k and hence also of Λ_k.

We have established that Λ_k^{-1} is decaying exponentially towards its off-diagonal corners. Moreover, since C_k is block bi-banded, we will see next that also $C_k^T \Lambda_k^{-1} C_k$, which enters in the expression of $\nabla^2 d(\lambda, \tau)$, decays towards the off-diagonal corners. Let us introduce the notation:

$$C_k^T \Lambda_k^{-1} C_k = \begin{bmatrix} V_{11} & V_{21}^T & \dots & V_{N+1,1}^T \\ V_{21} & V_{22} & \dots & V_{N+1,2}^T \\ \vdots & \vdots & \ddots & \vdots \\ V_{N+1,1} & V_{N+1,2} & \dots & V_{N+1,N+1} \end{bmatrix} \tag{34}$$

where $V_{ij} \in \mathbb{R}^{l_k \times l_k}$, and look at the decay of $C_k^T \Lambda_k^{-1} C_k$.

Proposition 3 *The off diagonal blocks (i.e. $V_{i,j}$ where $i - j > 0$) of $C_k^T \Lambda_k^{-1} C_k$ are bounded by:*

$$\|V_{i,j}\|_\bullet \leq \gamma^2 \bar{K}_{\Lambda_k} \omega_{\Lambda_k}^{i-j} \tag{35a}$$

$$\tag{35b}$$

where $\bar{K}_{\Lambda_k} = (2 + \omega_{\Lambda_k} + \omega_{\Lambda_k}^{-1}) K_{\Lambda_k}$ and $\bullet$ represents ∞ and 1.

Proof By performing the multiplication $C_k^T \Lambda_k^{-1} C_k$, we find that:

$$\begin{aligned} V_{i,j} = D_{k,i-1}^T T_{i,j-1} D_{k,j-1} + & C_{k,i-1}^T T_{i-1,j-1} D_{k,j-1} + \\ & + D_{k,i-1}^T T_{i-1,j} C_{k,j-1} + C_{k,i-1}^T T_{i,j} C_{k,j-1} \end{aligned} \tag{36}$$

for $i - j > 0$. Using the Cauchy-Schwarz inequality, Assumption 1 and (31), each term in (36) can be bounded according to:

$$\|D_{k,i-1}^T T_{i,j-1} D_{k,j-1}\|_\bullet \leq \gamma^2 K_{\Lambda_k} \omega_{\Lambda_k}^{i-j+1} \tag{37a}$$

$$\|C_{k,i-1}^T T_{i-1,j-1} D_{k,j-1}\|_\bullet \leq \gamma^2 K_{\Lambda_k} \omega_{\Lambda_k}^{i-j} \tag{37b}$$

$$\|D_{k,i-1}^T T_{i-1,j} C_{k,j-1}\|_\bullet \leq \gamma^2 K_{\Lambda_k} \omega_{\Lambda_k}^{i-j-1} \tag{37c}$$

$$\|C_{k,i-1}^T T_{i,j} C_{k,j-1}\|_\bullet \leq \gamma^2 K_{\Lambda_k} \omega_{\Lambda_k}^{i-j} \tag{37d}$$

where $\bullet$ represents ∞ and 1. Hence, we can establish the following bound:

$$\|V_{i,j}\|_\bullet \leq \gamma^2 (2 + \omega_{\Lambda_k} + \omega_{\Lambda_k}^{-1}) K_{\Lambda_k} \omega_{\Lambda_k}^{i-j} \tag{38}$$

□

Let us now introduce the notation:

$$\nabla^2 d(\lambda,\tau) = \begin{bmatrix} W_{11} & W_{21}^T & \dots & W_{N+1,1}^T \\ W_{21} & W_{22} & \dots & W_{N+1,2}^T \\ \vdots & \vdots & \ddots & \vdots \\ W_{N+1,1} & W_{N+1,2} & \dots & W_{N+1,N+1} \end{bmatrix} \tag{39}$$

where $W_{i,j} \in \mathbb{R}^{r_i \times r_j}$. We can now establish a decay towards the off-diagonal corners of $\nabla^2 d(\lambda,\tau)$.

Proposition 4 *The off-diagonal blocks (i.e. $W_{i,j}$ where $i-j>0$) of $\nabla^2 d(\lambda,\tau)$ satisfy the following bounds:*

$$\|W_{i,j}\|_\bullet \le \sum_{k=1}^{M} \gamma^4 \gamma^2_{\Phi_k^{-1}} \bar{K}_{\Lambda_k} \omega_{\Lambda_k}^{i-j} \tag{40}$$

where $\bullet$ represents ∞ and 1.

Proof Recall that the dual Hessian is given by:

$$\nabla^2 d(\lambda,\tau) = \sum_{k=1}^{M} F_k(\Phi_k^{-1} - \Phi_k^{-1} C_k^T \Lambda_k^{-1} C_k \Phi_k^{-1}) F_k^T \tag{41}$$

The matrices $F_k \Phi_k^{-1} F_k^T$ are block-diagonal and do not contribute to the off-diagonal blocks in $\nabla^2 d(\lambda,\tau)$. Therefore using (39) and (41), for $i-j>0$:

$$\|W_{i,j}\|_\bullet = \left\| \sum_{k=1}^{M} F_{k,i} \Phi_{k,i}^{-1} V_{i,j} \Phi_{k,j}^{-1} F_{k,j} \right\|_\bullet \le \sum_{k=1}^{M} \left\| F_{k,i} \Phi_{k,i}^{-1} V_{i,j} \Phi_{k,j}^{-1} F_{k,j} \right\|_\bullet \tag{42}$$

where $\bullet$ represents ∞ and 1. Moreover, by using Cauchy-Schwarz inequality, Assumption 1, 2 and Proposition 3, it can be verified that:

$$\|W_{i,j}\|_\bullet \le \sum_{k=1}^{M} \gamma^2 \gamma^2_{\Phi_k^{-1}} \|V_{i,j}\|_\bullet \le \sum_{k=1}^{M} \gamma^4 \gamma^2_{\Phi_k^{-1}} \bar{K}_{\Lambda_k} \omega_{\Lambda_k}^{i-j} \tag{43}$$

□

Proposition 4 concludes the characterization of the quasi-banded structure of $\nabla^2 d(\lambda,\tau)$.

3.2 A banded approximation of the dual Hessian

In Section 3.1, we established that the dual Hessian decays exponentially towards its off-diagonal corners. This suggests that a banded matrix can be used to form a good approximation of the dual Hessian, and accordingly be a basis for a quasi-Newton method. In this section, we will approximate the dual Hessian with a diagonal band of itself, and characterize the error in the approximation.

In the following, $\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}}$ represents the diagonal $\mathcal{M}$-block band of (39), which will be referred to as the *clipped dual Hessian*. We recall next Gershgorin's circle theorem, which will be instrumental in the following development:

Theorem 1 *For $A \in \mathbb{R}^{n\times n}$ with elements a_{ij}, let $R_i = \sum_{j\neq i} |a_{ij}|$ be the sum of the absolute values of the non-diagonal entries in the row i. Let $D(a_{ii}, R_i)$ be the closed disc centered in a_{ii} with radius R_i, then every eigenvalue of A lies within at least one of the discs $D(a_{ii}, R_i)$.*

Proof A proof is given in [13]. □

Using Gershgorin's theorem, we can establish a bound on the euclidean distance between the dual Hessian and its clipped counterpart:

Lemma 2 *The following bound holds:*

$$\| \lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau) \|_2 \leq \left(N - \frac{\mathcal{M}-1}{2} \right) \sum_{k=1}^{M} \gamma^4 \gamma^2_{\Phi_k^{-1}} \tilde{K}_{\Lambda_k} \omega_{\Lambda_k}^{\frac{\mathcal{M}-1}{2}} \tag{44}$$

where $\tilde{K}_{\Lambda_k} = \bar{K}_{\Lambda_k} \omega_{\Lambda_k} = (2 + \omega_{\Lambda_k} + \omega_{\Lambda_k}^{-1}) K_{\Lambda_k}$.

Proof Note that $\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau)$ is symmetric, which implies that:

$$\| \lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau) \|_2 = |\lambda_{\max}(\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau))| \tag{45}$$

where $\lambda_{\max}(\bullet)$ represents the largest eigenvalue of $\bullet$. We then us use Gershgorin's circle theorem to bound the magnitude of the largest eigenvalue in (45). Since, the diagonal elements of $\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau)$ are zero, all Gershgorin discs are centered at the origin. Since the maximum absolute row sum is equal to the largest Gershgorin radius, the following holds:

$$|\lambda_{\max}(\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau))| \leq \| \lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau) \|_\infty \tag{46}$$

Observe that each block row in $\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau)$ has not more than $N - \frac{\mathcal{M}-1}{2}$ nonzero blocks. Moreover, according to (40), the blocks $W_{i,j}$ in $\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau)$ with the largest bound on the maximum absolute row sum are located next to the diagonal $\mathcal{M}$-block band of zeros, i.e. where

$i - j = \frac{\mathcal{M}-1}{2} + 1$. Hence, by using the width of the band and (40), we can establish that:

$$\| \lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau) \|_\infty \le \left(N - \frac{\mathcal{M}-1}{2} \right) \sum_{k=1}^{M} \gamma^4 \gamma^2_{\Phi_k^{-1}} \bar{K}_{\Lambda_k} \omega_{\Lambda_k}^{\frac{\mathcal{M}-1}{2}+1} \tag{47}$$

Accordingly, by using (45), (46) and (47) ,we can establish that:

$$\| \lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau) \|_2 \le \left(N - \frac{\mathcal{M}-1}{2} \right) \sum_{k=1}^{M} \gamma^4 \gamma^2_{\Phi_k^{-1}} \bar{K}_{\Lambda_k} \omega_{\Lambda_k}^{\frac{\mathcal{M}-1}{2}+1} \tag{48}$$

which is identical to (44). □

It is well known that the use of an inexact Hessian will degrade the convergence of the Newton method to a linear rate provided that the inexact Hessian has full rank [26]. In the following, we will establish a criterion that guarantees a positive definite clipped dual Hessian. The criterion is based on the following proposition that follows from Weyl's inequality [13]:

Proposition 5 *Let* $A, B \in \mathbb{S}^{n \times n}$, *then*

$$\lambda_{\min}(A) + \lambda_{\min}(B) \le \lambda_{\min}(A + B) \tag{49}$$

where $\lambda_{\min}$ *represents the minimum eigenvalue.*

Proof A proof is given in [13].

We can now establish a criterion for the non-singularity of $\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}}$:

Lemma 3 *If* $\mathcal{M}$ *is chosen such that:*

$$\left(N - \frac{\mathcal{M}-1}{2} \right) \sum_{k=1}^{M} \gamma^4 \gamma^2_{\Phi_k^{-1}} \tilde{K}_{\Lambda_k} \omega_{\Lambda_k}^{\frac{\mathcal{M}-1}{2}} < \lambda_{\min}(\nabla^2 d(\lambda, \tau)) \tag{50}$$

then $\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}}$ *is positive definite.*

Proof Observe that

$$\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} = \nabla^2 d(\lambda, \tau) + \left(\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau) \right) \tag{51}$$

Proposition 5 entails that:

$$\lambda_{\min}(\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}}) \ge \lambda_{\min}(\nabla^2 d(\lambda, \tau)) + \lambda_{min}(\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau)) \tag{52}$$

Accordingly, $\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}}$ is nonsingular if:

$$\lambda_{\min}(\nabla^2 d(\lambda, \tau)) > |\lambda_{min}(\lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda, \tau))| \tag{53}$$

We have already established that the eigenvalues of $\lfloor \nabla^2 d(\lambda,\tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda,\tau))$ are located close to the origin and are not more negative then the largest Gershgorin radius given by (47). Hence, we realize that

$$\lambda_{min}(\lfloor \nabla^2 d(\lambda,\tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda,\tau)) \geq -\left(N - \frac{\mathcal{M}-1}{2}\right)\sum_{k=1}^{M} \gamma^4 \gamma^2_{\Phi_k^{-1}} \tilde{K}_{\Lambda_k} \omega_{\Lambda_k}^{\frac{\mathcal{M}-1}{2}} \tag{54}$$

Using (53) and (54) gives (50). □

Remark 7 It should be noted that $\lambda_{\min}(\nabla^2 d(\lambda,\tau)) > 0, \ \forall \lambda$ [15], such that (50) can always be satisfied.

Remark 8 Observe that (50) is a sufficient and not a necessary condition. According to our experiments, singularity is not a problem in practice, and (50) is hence very conservative.

3.3 Convergence of the quasi-Newton method

In the previous section, we have seen that the clipped dual Hessian is positive definite and close, in an euclidean sense, to the exact Hessian provided that the band $\mathcal{M}$ is wide enough. In this section, we will consider issues related to the convergence and complexity of a quasi-Newton method based on the *clipped dual Hessian.*

Let us start by detailing how the clipped dual Hessian affect the convergence of the Newton method:

Theorem 2 *If the dual function fulfills:*

$$\left\| \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda,\tau) - \nabla^2 d)\, dt \, \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \right\|_2 \leq \omega \|\nabla d\|_2 \tag{55}$$

then the following bound holds:

$$\|\nabla d(\lambda + \Delta\lambda,\tau)\|_2 \leq \omega \|\nabla d(\lambda,\tau)\|_2^2 + f(\mathcal{M})\|\nabla d(\lambda,\tau)\|_2 \tag{56}$$

where

$$f(\mathcal{M}) = \|(\lfloor \nabla^2 d(\lambda,\tau) \rceil_{\mathcal{M}} - \nabla^2 d(\lambda,\tau))\lfloor \nabla^2 d(\lambda,\tau) \rceil_{\mathcal{M}}^{-1}\|_2 \tag{57}$$

Remark 9 We will further discuss assumption (55) in the next proposition.

Proof An approximate Newton step based on the clipped dual Hessian is given by the following linear system:

$$\lfloor \nabla^2 d(\lambda,\tau) \rceil_{\mathcal{M}} \Delta\lambda + \nabla d(\lambda,\tau) = 0 \tag{58}$$

In order to avoid heavy notations, we will omit arguments when they are obvious from the context and introduce the notation $\nabla d^+ = \nabla d(\lambda + \Delta\lambda, \tau)$. We then have:

$$\begin{aligned}\|\nabla d^+\|_2 &= \|\nabla d^+ - \nabla d - \lfloor \nabla^2 d \rceil_{\mathcal{M}} \Delta\lambda\|_2 = \\ &= \left\| \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda, \tau)\Delta\lambda)\mathrm{dt} - \lfloor \nabla^2 d \rceil_{\mathcal{M}} \Delta\lambda \right\|_2 = \\ &= \left\| \left(\int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda, \tau) - \nabla^2 d)\mathrm{dt} + \nabla^2 d - \lfloor \nabla^2 d \rceil_{\mathcal{M}} \right) \Delta\lambda \right\|_2 = \\ &= \left\| \left(\int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda, \tau) - \nabla^2 d)\mathrm{dt} + \nabla^2 d - \lfloor \nabla^2 d \rceil_{\mathcal{M}} \right) \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \nabla d \right\|_2\end{aligned} \tag{59}$$

where we have used (58) and the following result from calculus:

$$\int_0^1 \nabla^2 d(\lambda + t\Delta\lambda, \tau)\Delta\lambda \ \mathrm{dt} = \nabla d^+ - \nabla d \tag{60}$$

Consequently, we can bound (59) according to:

$$\|\nabla d^+\|_2 \leq \omega \|\nabla d\|_2^2 + f(\mathcal{M}) \|\nabla d\|_2 \tag{61}$$

where we have introduced:

$$\left\| \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda, \tau) - \nabla^2 d) \ \mathrm{dt} \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \right\|_2 \leq \omega \|\nabla d\|_2 \tag{62}$$

and

$$f(\mathcal{M}) = \left\| \left(\nabla^2 d - \lfloor \nabla^2 d \rceil_{\mathcal{M}} \right) \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \right\|_2 \tag{63}$$

□

Remark 10 Observe the $f(\mathcal{M}) \geq 0$ is a function which value we in principle can choose.

The condition (55) can be related to the Lipschitz continuity of the dual Hessian. This is established in the following:

Proposition 6 *Let the dual Hessian be Lipschitz continuous with constant L, then:*

$$\omega = \frac{1}{2} L \| \lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}}^{-1} \|_2^2 \tag{64}$$

fulfills (55).

Proof By definition of Lipschitz continuity, the following criteria holds:

$$\|\nabla^2 d(\lambda + \Delta\lambda, \tau) - \nabla^2 d(\lambda, \tau)\|_2 \leq L \|\Delta\lambda\|_2 \tag{65}$$

Observe that:

$$\begin{aligned}
&\left\| \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda, \tau) - \nabla^2 d)\mathrm{dt} \; \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \right\|_2 \leq \\
&\leq \left\| \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda, \tau) - \nabla^2 d)\mathrm{dt} \; \right\|_2 \left\| \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \right\|_2 \leq \\
&\leq \int_0^1 \left\| (\nabla^2 d(\lambda + t\Delta\lambda, \tau) - \nabla^2 d) \right\|_2 \mathrm{dt} \left\| \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \right\|_2 \leq \\
&\leq \frac{1}{2} L \left\| \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \right\|_2 \|\Delta\lambda\|_2 \leq \frac{1}{2} L \left\| \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \right\|_2^2 \|\nabla d(\lambda, \tau)\|_2
\end{aligned} \tag{66}$$

□

In terms of iterations, the optimal choice would be to use the full dual Hessian and hence make $f(\mathcal{M}) = 0$. However, since a dense factorization is more costly than its banded counterpart, the overall complexity of solving a problem can be lower if a nonzero $f(\mathcal{M})$ is used. As an illustrative example, the complexity of finding a Cholesky factorization of a dense matrix $A \in \mathbb{S}_+^n$ is $\frac{1}{3}n^3$ flops. On the other hand, if A is $k-$banded (where $k \ll n$), the complexity is reduced to nk^2 flops.

To further specify properties of $f(\mathcal{M})$, we state the following corollary that follows directly from Lemma 2:

Corollary 1 *The following bound holds:*

$$f(\mathcal{M}) \leq \left\| \lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}}^{-1} \right\|_2 \left(N - \frac{\mathcal{M} - 1}{2} \right) \sum_{k=1}^{M} \gamma^4 \gamma_{\Phi_k^{-1}}^2 \tilde{K}_{\Lambda_k} \omega_{\Lambda_k}^{\frac{\mathcal{M}-1}{2}} \tag{67}$$

Proof By using Cauchy Schwarz inequality, we have that:

$$f(\mathcal{M}) \leq \left\| \nabla^2 d - \lfloor \nabla^2 d \rceil_{\mathcal{M}} \right\|_2 \left\| \lfloor \nabla^2 d \rceil_{\mathcal{M}}^{-1} \right\|_2 \tag{68}$$

Moreover, by using (44), we can establish that

$$f(\mathcal{M}) \leq \left\| \lfloor \nabla^2 d(\lambda, \tau) \rceil_{\mathcal{M}}^{-1} \right\|_2 \left(N - \frac{\mathcal{M} - 1}{2} \right) \sum_{k=1}^{M} \gamma^4 \gamma_{\Phi_k^{-1}}^2 \tilde{K}_{\Lambda_k} \omega_{\Lambda_k}^{\frac{\mathcal{M}-1}{2}} \tag{69}$$

□

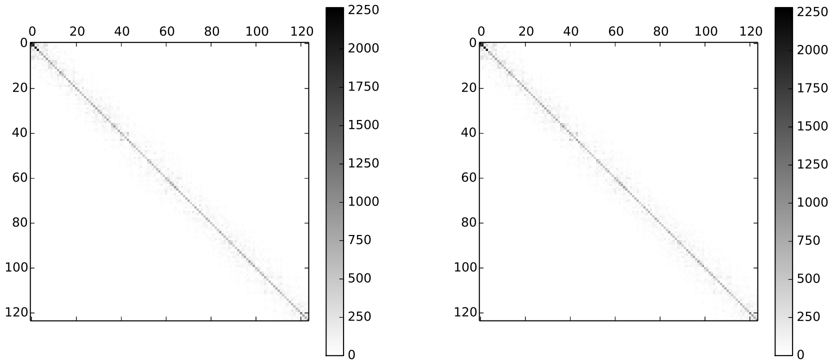
4 Numerical experiments

4.1 An example

To investigate the consistency of the off-diagonal decay for a given problem, we look at a randomly generated problem, of the form (4), with $M = 4$ subproblems each with a horizon $N = 30$, 6 states, 4 controls, and $4(N+1) = 124$ inequality constraints. At each time instance the subproblems are coupled via

4 coupling constraints. Moreover, the elements are generated with an even distribution in the interval $[-1000, 1000]$.

The absolute values of the elements in the dual Hessian evaluated at the solution $\lambda^*(\tau)$ for $\tau = 1$, are visualized in Figure 1a. It can be seen that elements with a large magnitude are present along the diagonal, while elements towards the off-diagonal corners are small in absolute value.



a) $\tau = 1$ b) $\tau = 10^{-6}$

Fig. 1: Illustration of the dual Hessian evaluated at $\lambda^*(\tau)$.

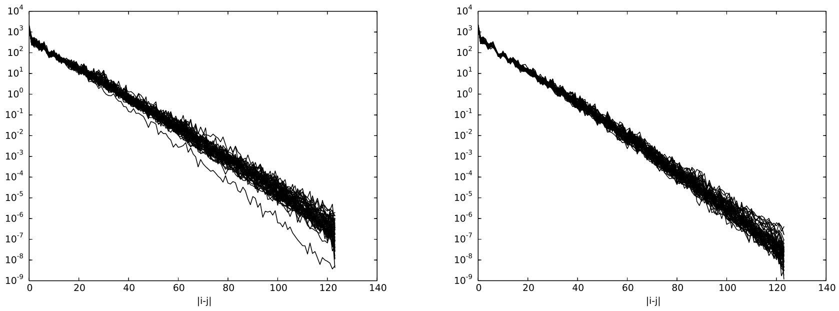
Let us investigate the consistency of the decay with respect to λ. The maximal absolute value of the elements in $\nabla^2 d(\lambda, \tau)$ for given distances from its diagonal, are visualized in Figure 2a for 50 randomly selected values of λ. More specifically, the measure used in Figure 2a is:

$$\begin{aligned} \max_{i,j} \quad & |[\nabla^2 d(\lambda, \tau)]_{i,j}| \\ \text{s.t.} \quad & |i - j| = \text{constant} \end{aligned} \tag{70}$$

Observe that for this example, at all evaluated points λ, the dual Hessian has a clear decaying trend towards its off-diagonal corners.

As discussed previously, the conditioning of Λ_k affects the decaying property and small values of τ would therefore possibly lead to a slower decay. Corresponding figures to Figure 1a and Figure 2a, for $\tau = 10^{-6}$, are shown in Figure 1b and Figure 2b respectively. Observe that, for this example, there is no significant difference in the decay between the different values of τ. This observation pertain to all experiments we have seen so far.

The parameter that seems to have the strongest impact on the decaying is the horizon length. To illustrate this, we generate another random problem with $N = 10$, and all other dimensions being the same as before. The dual Hessian at the solution $\lambda^*(\tau)$, for $\tau = 1$ is visualized in Figure 3. Observe



a) $\tau = 1$ b) $\tau = 10^{-6}$

Fig. 2: Illustration of the off-diagonal decay. Each line corresponds to different λ.

that the dual Hessian also in this case has a clear decay towards the off-diagonal corners. The band with elements of a significant magnitude is however wider compared to the size of the matrix. Moreover, the decaying property is visualized for 50 different values of λ in Figure 4. Note that in this figure it is easier to see that there is a smaller difference between the diagonal and the off-diagonal elements.

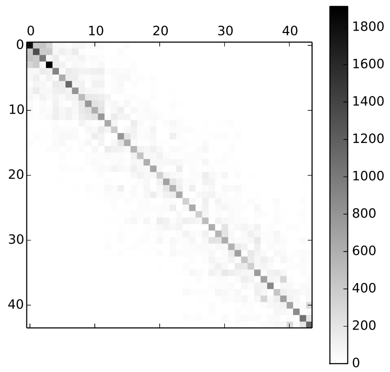


Fig. 3: Illustration of the dual Hessian evaluated at $\lambda^*(\tau)$.

5 Conclusions

Bla bla bla

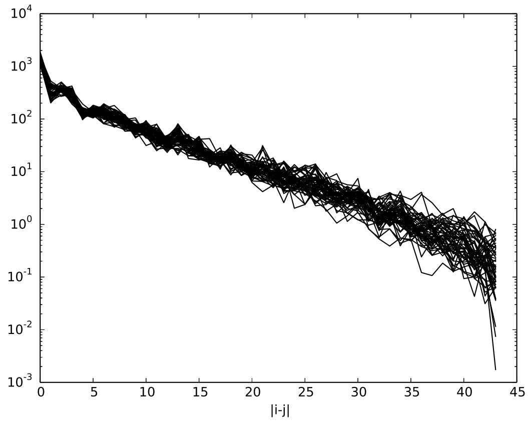


Fig. 4: Illustration of the off-diagonal decay. Each line corresponds to different λ.

References

1. Benzi, M.: Preconditioning techniques for large linear systems: a survey. Journal of Computational Physics **182**, 418–477 (2002)
2. Bertsekas, D., Tsitsiklis, J.N.: Parallel and distributed computation: Numerical methods. Prentice Hall (1989)
3. Conte, C., Zeilinger, M., Morari, M., Jones, C.: Robust distributed model predictive control of linear systems. In: European Control Conference (2013)
4. Cottle, R.W., Duvall, S.G., Zikan, K.: A lagrangean relaxation algorithm for the constrained matrix problem. Naval Research Logistics Quarterly **33**(1), 55–76 (1986). DOI 10.1002/nav.3800330106. URL http://dx.doi.org/10.1002/nav.3800330106
5. Demko, S., Moss, W., Smith, P.: Decay rates for inverses of band matrices. Mathemathics of Computation **43**, 491–499 (1986)
6. Domahidi, A., Zgraggen, A., Zeilinger, M., Morari, M., Jones, C.: Efficient Interior Point Methods for Multistage Problems Arising in Receding Horizon Control. In: IEEE Conference on Decision and Control (CDC), pp. 668 – 674. Maui, HI, USA (2012)
7. Ferreau, H., Kozma, A., Diehl, M.: A parallel active-set strategy to solve sparse parametric quadratic programs arising in MPC. In: Proceedings of the 4th IFAC Nonlinear Model Predictive Control Conference, Noordwijkerhout, The Netherlands (2012)
8. Frasch J.and Vukov, M., Ferreau H.and Diehl, M.: A new quadratic programming strategy for efficient sparsity exploitation in sqp-based nonlinear mpc and mhe. In: Proceedings of the 19th IFAC World Congress (2014)
9. Giselsson, P.: Improved dual decomposition for distributed model predictive control. In: Proceedings of the 19th IFAC World Congress (2014)
10. Giselsson, P., Doan, M.D., Keviczky, T., Schutter, B.D., Rantzer, A.: Accelerated gradient methods and dual decomposition in distributed model predictive control. Automatica **49**(3), 829 – 833 (2013). DOI http://dx.doi.org/10.1016/j.automatica.2013.01.009. URL http://www.sciencedirect.com/science/article/pii/S0005109813000101
11. Giselsson, P., Rantzer, A.: Distributed model predictive control with suboptimality and stability guarantees. In: Decision and Control (CDC), 2010 49th IEEE Conference on, pp. 7272–7277 (2010). DOI 10.1109/CDC.2010.5717026
12. Gros, S.: A distributed algorithm for nmpc-based wind farm control. In: Conference on Decision and Control (2014)
13. Horn, R.A., Johnson, C.R.: Matrix Analysis - 2nd ed. Cambridge University Press (2013)

14. Hult, R., Campos, G., Falcone, P., Wymeersch, H.: An approximate solution to the optimal coordination problem for autonomous vehicles at intersections. In: American Control Conference (2015)
15. Klintberg, E., Gros, S.: A primal-dual newton method for distributed quadratic programming. In: Conference on Decision and Control (2014)
16. Kozma, A., Conte, C., Diehl, M.: Benchmarking large scale distributed convex quadratic programming algorithms. Optimization Methods & Software (2014). (submitted for publication)
17. Kozma, A., Frasch, J.V., Diehl, M.: A Distributed Method for Convex Quadratic Programming Problems Arising in Optimal Control of Distributed Systems. In: Proceedings of the 52nd Conference on Decision and Control (CDC) (2013)
18. Kozma, A., Klintberg, E., Gros, S., Diehl, M.: An improved distributed dual Newton-CG method for convex quadratic programming problems. In: American Control Conference (2014)
19. Kozma, A., Savorgnan, C., Diehl, M.: Distributed Model Predictive Control Made Easy, *Intelligent Systems, Control and Automation: Science and Engineering*, vol. Vol 69., chap. Distributed multiple shooting for large scale nonlinear systems, p. 604. Springer (2014)
20. Lasdon, L.S.: Optimization theory for Large Systems. Dover (1970)
21. Necoara, I., Nedelcu, V.: Rate analysis of inexact dual first-order methods application to dual decomposition. IEEE Trans. Automat. Control **59**, 1232–1243 (2014)
22. Necoara, I., Nedelcu, V., Dumitrache, I.: Parallel and distributed optimization methods for estimation and control in networks. Journal of Process Control **21**(5), 756 – 766 (2011). DOI http://dx.doi.org/10.1016/j.jprocont.2010.12.010. URL http://www.sciencedirect.com/science/article/pii/S095915241000257X. Special Issue on Hierarchical and Distributed Model Predictive Control
23. Necoara, I., Patrascu, A.: Iteration complexity analysis of dual first order methods for convex programming. Journal of Optimization Theory and Applications (Submitted) (2014)
24. Necoara, I., Suykens, J.: An interior-point Lagrangian decomposition method for separable convex optimization. J. Optim. Theory and Appl. **143**(3), 567–588 (2009)
25. Nesterov, Y., Nemirovski, A.: Interior-point Polynomial Algorithms in Convex Programming. Society for Industrial Mathematics (1994)
26. Nocedal, J., Wright, S.: Numerical Optimization, 2 edn. Springer Series in Operations Research and Financial Engineering. Springer (2006)
27. Perizzato, A., Farina, M., Scattolini, R.: Stochastic distributed predictive control of independent systems with coupling constraints. In: Conference on Decision and Control (2014)
28. Richter, S., Morari, M., Jones, C.: Towards computational complexity certification for constrained MPC based on Lagrange Relaxation and the fast gradient method. In: Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on, pp. 5223 –5229 (2011). DOI 10.1109/CDC.2011.6160931
29. Tran, D.Q., Necoara, I., Savorgnan, C., Diehl, M.: An inexact perturbed path-following method for lagrangian decomposition in large-scale separable convex optimization. SIAM Journal on Optimization **23**, 95–125 (2013)
30. Wright, M.: Ill-conditioning and computational error in interior methods for nonlinear programming. SIAM Journal on Optimization **9**, 84–111 (1998)