

## NUMERICAL STRUCTURE OF THE HESSIAN OF THE LAGRANGE DUAL FUNCTION FOR A CLASS OF CONVEX PROBLEMS\*

EMIL KLINTBERG<sup>†</sup> AND SEBASTIEN GROS<sup>†</sup>

**Abstract.** This paper considers a structured separable convex optimization problem, motivated by the deployment of model predictive control on multiagent systems that are interacting via nondelayed couplings. We show that the dual decomposition of this problem yields a numerical structure in the Hessian of the dual function. This numerical structure allows for deploying a quasi-Newton method in the dual space. For large problems, this approach yields a large reduction of the computational complexity of solving the problem, and for geographically distributed problems a reduction in the communication burden.

**Key words.** dual decomposition, Newton method, structure exploitation, dual Hessian

**AMS subject classifications.** 49M15, 49M20, 49M27, 49M29, 90C35

**DOI.** 10.1137/15M1024573

**1. Introduction.** Many problems in distributed control and networked systems can be formulated as separable convex optimization problems; see e.g. [23, 27, 21]. Problems of this form can be solved using standard optimization techniques once the data of the problem are centralized. However, when the problem is geographically distributed or when the problem comprises agents that do not want to share sensitive information, it can be highly impractical to centralize the data of the problem. Techniques from distributed optimization are then preferable.

Several decomposition methods have been proposed to solve separable optimization problems in a distributed fashion; see e.g. [3, 22]. A common technique is based on a Lagrangian relaxation, which allows for decomposing the optimization problem into low-dimensional subproblems that can be solved independently. The subproblems can then be coordinated by manipulating the dual variables, i.e., by solving the generally nonsmooth dual problem.

Lagrangian relaxations are used in many different contexts in order to solve large-scale convex optimization problems, e.g. the authors of [6], use a coordinate ascent method to solve matrix problems; in [18, 26, 11] a subgradient method is used to attain dual optimality, whereas in [24, 20, 10, 19] a fast gradient method is used. All these methods make use of only the first-order derivatives of the dual function to obtain a search direction and their theoretical and practical convergence can therefore not be faster than sublinear. The authors of, e.g., [21, 16, 25, 17, 14, 12, 9] overcome this limitation by using Newton strategies in the dual space.

The fast convergence rate of Newton strategies comes at the expense of factorizing the Hessian of the dual function at each iteration. For specific applications, the dual Hessian can be structured and the cost for its factorization can therefore be lowered; see e.g. [9]. However, most problems yield a dense dual Hessian, making its

---

\*Received by the editors June 5, 2015; accepted for publication (in revised form) November 28, 2016; published electronically February 28, 2017.

<http://www.siam.org/journals/sicon/55-1/M102457.html>

**Funding:** This research was supported by Chalmers Tekniska Högskola (Chalmers University of Technology).

<sup>†</sup>Signals and Systems, Chalmers University of Technology, Gothenberg, Sweden (kemil@chalmers.se, grosse@chalmers.se).

factorization computationally expensive, or even a bottleneck when solving problems that have a large number of complicating constraints between the subproblems.

In this paper, we show that for a class of separable optimization problems, the Hessian of the dual function is dense but numerically structured. The numerical structure implies that the Hessian contains negligible elements outside of a given band, such that a banded approximation of the dual Hessian can be used to deploy an inexact Newton method. The problem class is motivated by model predictive control of multiagent systems interacting via nondelayed couplings.

To avoid issues related to a nonsmooth dual function, the numerical structure is first analyzed when inequality constraints are omitted in the problem formulation. This allows for a straightforward and compact analysis of the convergence of the quasi-Newton strategy in the neighborhood of the solution. The results are then extended to inequality constrained problems, using the log-barrier method that was proposed in [21], and further developed in [25, 14, 12]. However, it should be emphasized that the numerical structure is a result of matrix properties and extends to the nonsmooth methods proposed in, e.g., [16, 9].

The paper is outlined as follows. Section 2 introduces the class of problems considered in this work. In section 3, dual decomposition is recalled, and the numerical structure of the Hessian is analyzed. Moreover, a quasi-Newton strategy is presented and its convergence is analyzed. In section 4, the results are extended to inequality constrained problems, using the log-barrier method proposed in [21]. In section 5, the structure of the Hessian, and the convergence of the quasi-Newton method are illustrated by a numerical example. The paper is concluded in section 6.

**1.1. Notations and terminology.** The notation  $\mathbb{S}^n$  defines the set of symmetric matrices of dimension  $n \times n$ . Moreover, we use the notation  $\mathbb{S}_{++}^n \subset \mathbb{S}^n$  for symmetric and positive definite matrices. For a scalar valued function  $f$  with two arguments,  $t \in \mathbb{R}$  and  $x \in \mathbb{R}^n$ , we denote the partial derivative of  $f(x, t)$  with respect to  $x$  as  $\nabla f(x, t)^T \in \mathbb{R}^n$ , and the partial derivative of  $\nabla f(x, t)$  with respect to  $x$  as  $\nabla^2 f(x, t) \in \mathbb{S}^n$ . The notation  $A_{i,j}$  is referring to element  $(i, j)$  of a matrix  $A$ . Furthermore, we say that a matrix  $A$  is banded with bandwidth  $m$  if  $A_{i,j} = 0$  for  $|i - j| > m$ . We use the notation  $x^{-1} \in \mathbb{R}^n$  for the elementwise inverse of a vector  $x \in \mathbb{R}^n$ . For a matrix  $A \in \mathbb{R}^{n \times n}$ , we will use the following matrix norms:

- The maximum row sum matrix norm is denoted as  $\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |A_{i,j}|$ .
- The Frobenius matrix norm is denoted as  $\|A\|_F = (\sum_{i,j=1}^n |A_{i,j}|^2)^{1/2}$ .
- The spectral norm is denoted as  $\|A\|_2 = \sigma_{max}(A)$ , where  $\sigma_{max}(A)$  denotes the largest singular value of  $A$ .

**2. Problem formulation.** In this paper, we consider the following separable optimization problem

$$(2.1a) \quad \min_z \sum_{k=1}^P f_k(z_k)$$

$$(2.1b) \quad \text{s.t.} \quad \sum_{k=1}^P F_k z_k = e,$$

$$(2.1c) \quad C_k z_k = d_k, \quad k = 1, \dots, P,$$

where  $z = [z_1^T, \dots, z_P^T]^T$ , with  $z_k \in \mathbb{R}^{\bar{n}}$ , is a vector of decision variables,  $F_k \in \mathbb{R}^{\bar{p} \times \bar{n}}$  and  $e \in \mathbb{R}^{\bar{p}}$  denote the coupling constraints,  $C_k \in \mathbb{R}^{\bar{m} \times \bar{n}}$  and  $d_k \in \mathbb{R}^{\bar{m}}$  yield local

equality constraints, and the  $f_k : \mathbb{R}^{\bar{n}} \rightarrow \mathbb{R}$  are functions of the form

$$(2.2) \quad f_k(z_k) = \sum_{i=1}^N f_{k,i}(z_{k,i}),$$

where we have introduced the partitioning  $z_k = [z_{k,1}^T, \dots, z_{k,N}^T]^T$ , with  $z_{k,i} \in \mathbb{R}^n$ , of the decision variables. Additionally, we consider the following structure of the constraints (2.1b) and (2.1c):

$$(2.3a) \quad C_k = \begin{bmatrix} C_{k,1} & D_{k,2} & & & \\ & \ddots & \ddots & & \\ & & & C_{k,N-1} & D_{k,N} \\ & & & & \end{bmatrix},$$

$$(2.3b) \quad F_k = \begin{bmatrix} F_{k,1} & & & \\ & \ddots & & \\ & & & F_{k,N} \end{bmatrix},$$

where  $C_{k,i} \in \mathbb{R}^{m \times n}$ ,  $D_{k,i} \in \mathbb{R}^{m \times n}$ , and  $F_{k,i} \in \mathbb{R}^{p \times n}$ . The uniform problem dimensions  $n$ ,  $m$ , and  $p$  are assumed for simplicity of presentation. Furthermore, for notational simplicity, we define the sets  $\mathcal{Z}_k = \{z_k \mid C_k z_k = d_k\}$  and  $\mathcal{Z} = \mathcal{Z}_1 \times \dots \times \mathcal{Z}_P$ .

Since local equality constraints of the form (2.1c) with matrices  $C_k$  of the form (2.3a) typically arise in the context of dynamical systems, problem (2.1) is motivated by the control of  $P$  interacting dynamical systems. Problems of the form (2.1) with matrices  $F_k$  of the form (2.3b) arise when multiagent systems interact via nondelayed couplings. In section 4, the problem formulation is extended to include local inequality constraints. Throughout the paper, we assume that the following assumptions hold.

*Assumption 2.1.*

1. For  $k = 1, \dots, P$ , the objective functions  $f_k$  are twice continuously differentiable and strictly convex.
2. The constraints (2.1b) and (2.1c) fulfill the linear independence constraint qualification.

Note that Assumption 2.1 implies that strong duality holds [4]. Additionally, we observe that Assumption 2.1 is not restrictive since it can be fulfilled via an ad hoc elimination of the redundant constraints.

**3. Dual decomposition.** In the context of Lagrangian dual decomposition, the dual variables  $\lambda \in \mathbb{R}^p$ , corresponding to the coupling constraints (2.1b), are introduced and the Lagrange function is defined as

$$(3.1) \quad \mathcal{L}(z, \lambda) = \sum_{k=1}^P f_k(z_k) + \lambda^T \left( \sum_{k=1}^P F_k z_k - e \right),$$

where we observe that  $\mathcal{L}(z, \lambda)$  is separable in  $z$ , i.e., we can write (3.1) as

$$(3.2) \quad \mathcal{L}(z, \lambda) = \sum_{k=1}^P \mathcal{L}_k(z_k, \lambda),$$

where

$$(3.3) \quad \mathcal{L}_k(z_k, \lambda) = f_k(z_k) + \lambda^T \left( F_k z_k - \frac{1}{P} e \right).$$

The dual function, defined here as  $d(\lambda) = -\min_{z \in \mathcal{Z}} \mathcal{L}(z, \lambda)$  (see e.g. [4]), can then be evaluated in parallel as

$$(3.4) \quad d(\lambda) = -\sum_{k=1}^P \min_{z_k \in \mathcal{Z}_k} \mathcal{L}_k(z_k, \lambda).$$

Because of strong duality the solution to problem (2.1) can be obtained via solving the dual problem

$$(3.5) \quad \min_{\lambda} d(\lambda).$$

The latter can be solved in a completely distributed or partly distributed fashion. Assumption 2.1 ensures the strict convexity of (2.1) such that  $d(\lambda)$  is convex, continuously differentiable, and has a gradient given by [3]:

$$(3.6) \quad \nabla d(\lambda) = -\sum_{k=1}^P F_k z_k^*(\lambda) + e,$$

where  $z_k^*(\lambda)$  denotes the solution to the subproblem

$$(3.7) \quad z_k^*(\lambda) = \arg \min_{z_k \in \mathcal{Z}_k} \mathcal{L}_k(z_k, \lambda).$$

Following directly from (3.6), the Hessian of the dual function is given by

$$(3.8) \quad \nabla^2 d(\lambda) = -\sum_{k=1}^P F_k \frac{\partial z_k^*(\lambda)}{\partial \lambda}.$$

In this paper, we aim at solving the dual problem (3.5), via a quasi-Newton method. To that end, we update the dual variables  $\lambda$  according to:

$$(3.9) \quad \lambda^+ = \lambda + t\Delta\lambda$$

for a  $t \in (0, 1]$ , where the search direction  $\Delta\lambda \in \mathbb{R}^{\bar{p}}$  is provided by the quasi-Newton system

$$(3.10) \quad \mathcal{M}(\lambda)\Delta\lambda = -\nabla d(\lambda),$$

where the matrix  $\mathcal{M}(\lambda) \in \mathbb{S}_{++}^{\bar{p}}$  is an approximation of the Hessian of the dual function (3.8). In the following, we will show that the Hessian of the dual function for problem (2.1) is dense, but numerically banded. This observation will motivate the use of a diagonal band of the dual Hessian as a Hessian approximation  $\mathcal{M}(\lambda)$  in (3.10), i.e., we propose to compute the quasi-Newton direction  $\Delta\lambda$  from (3.10) using

$$(3.11) \quad \mathcal{M}(\lambda) = [\nabla^2 d(\lambda)]_M,$$

where we use the notation  $[\nabla^2 d(\lambda)]_M$  for the diagonal  $M$ -band of  $\nabla^2 d(\lambda)$ , i.e.,

$$(3.12) \quad ([\nabla^2 d(\lambda)]_M)_{i,j} = \begin{cases} (\nabla^2 d(\lambda))_{i,j} & |i - j| \leq M, \\ 0 & \text{otherwise.} \end{cases}$$

The banded structure of  $[\nabla^2 d(\lambda)]_M$  allows for deploying structure-exploiting factorization techniques for solving the linear system (3.10), and therefore lower the computational cost of calculating the search direction.

In the following, it will be useful to introduce the matrix  $[\nabla^2 d(\lambda)]_M^C$  defined as

$$(3.13) \quad [\nabla^2 d(\lambda)]_M^C = \nabla^2 d(\lambda) - [\nabla^2 d(\lambda)]_M,$$

capturing the entries in the dual Hessian that are neglected in its banded approximation. Consequently,  $[\nabla^2 d(\lambda)]_M^C$  has a diagonal band of zeros, but is identical to the Hessian of the dual function outside of that band.

The rest of this section is organized as follows. In section 3.1, we show that the Hessian of the dual function is numerically structured by quantifying the Frobenius norm of  $[\nabla^2 d(\lambda)]_M^C$  in terms of the problem data and the bandwidth  $M$ . In section 3.2, we propose a quasi-Newton method, where the search directions are calculated from (3.10), and we analyze its convergence.

**3.1. Numerical structure of the dual Hessian.** In this section, we show that the Hessian of the dual function is dense but numerically structured. The numerical structure will allow us to use a banded approximation of the dual Hessian, and deploy it in the context of a dual quasi-Newton method.

We introduce here the notation  $H_k = \nabla^2 f_k(z_k^*(\lambda))$ . Because of the separable structure of  $f_k(z_k)$ , we observe that  $H_k$  is block diagonal:

$$(3.14) \quad H_k = \begin{bmatrix} H_{k,1} & & \\ & \ddots & \\ & & H_{k,N} \end{bmatrix}$$

with blocks  $H_{k,i} = \nabla^2 f_{k,i}(z_{k,i}^*(\lambda))$ . Additionally, Assumption 2.1 ensures that  $H_k$  is invertible. Using the notation (3.14), the analysis of  $[\nabla^2 d(\lambda)]_M^C$ , will be based on the following explicit expression for the dual Hessian.

PROPOSITION 3.1. *The Hessian of the dual function  $d(\lambda)$  is given by*

$$(3.15) \quad \nabla^2 d(\lambda) = \sum_{k=1}^P F_k (H_k^{-1} - H_k^{-1} C_k^T \Lambda_k^{-1} C_k H_k^{-1}) F_k^T,$$

where

$$(3.16) \quad \Lambda_k = C_k H_k^{-1} C_k^T.$$

*Proof.* Recall that the Hessian of the dual function is given by (3.8), where  $z_k^*(\lambda)$  is a solution to

$$(3.17a) \quad 0 = \nabla f_k(z_k^*(\lambda)) + F_k^T \lambda + C_k^T \mu_k^*(\lambda),$$

$$(3.17b) \quad 0 = C_k z_k^*(\lambda) - d_k,$$

where we have introduced the Lagrange multipliers  $\mu_k \in \mathbb{R}^m$  corresponding to (2.1c). Differentiating (3.17) yields the following linear system:

$$(3.18a) \quad 0 = H_k \frac{\partial z_k^*(\lambda)}{\partial \lambda} + F_k^T + C_k^T \frac{\partial \mu_k^*(\lambda)}{\partial \lambda},$$

$$(3.18b) \quad 0 = C_k \frac{\partial z_k^*(\lambda)}{\partial \lambda}.$$

Solving (3.18) for  $\frac{\partial z_k^*(\lambda)}{\partial \lambda}$  then delivers

$$(3.19) \quad \frac{\partial z_k^*(\lambda)}{\partial \lambda} = - (H_k^{-1} - H_k^{-1} C_k^T \Lambda_k^{-1} C_k H_k^{-1}) F_k^T,$$

where  $\Lambda_k$  is invertible due to Assumption 2.1. Using (3.8) and (3.19), the expression (3.15) for the dual Hessian follows.  $\square$

Note that all matrices in (3.15) are banded except  $\Lambda_k^{-1}$ , which is, in general, dense, and consequently also makes the dual Hessian dense. However, in the following we will show that the magnitude of the entries of matrix  $\Lambda_k^{-1}$  are decaying towards the off-diagonal corners, i.e., it is numerically structured. Since all other matrices in (3.15) are banded, the numerical structure of  $\Lambda_k^{-1}$  translates into a numerically structured dual Hessian. Specifically, we will show that  $[\nabla^2 d(\lambda)]_M^C$  is small for a sufficiently large bandwidth  $M$ . To characterize this structure, we start by observing that  $\Lambda_k$ , given by (3.16), is a block-tridiagonal matrix:

$$(3.20) \quad \Lambda_k = \begin{bmatrix} \Lambda_{k,(1,1)} & \Lambda_{k,(1,2)} & & & \\ \Lambda_{k,(1,2)}^T & \Lambda_{k,(2,2)} & & \ddots & \\ & \ddots & \ddots & \ddots & \\ & & & \Lambda_{k,(N-1,N)}^T & \Lambda_{k,(N-1,N)} \\ & & & \Lambda_{k,(N-1,N)} & \Lambda_{k,(N,N)} \end{bmatrix}$$

with blocks  $\Lambda_{k,(i,j)} \in \mathbb{R}^{m \times m}$  given by

$$(3.21a) \quad \Lambda_{k,(i,i)} = C_{k,i-1} H_{k,i-1}^{-1} C_{k,i-1}^T + D_{k,i} H_{k,i}^{-1} D_{k,i}^T,$$

$$(3.21b) \quad \Lambda_{k,(i,i+1)} = D_{k,i} H_{k,i}^{-1} C_{k,i}^T.$$

In consequence,  $\Lambda_k$  is a Hermitian, positive definite, banded matrix with bandwidth  $2m$ . To quantify the magnitude of matrix  $[\nabla^2 d(\lambda)]_M^C$ , we will rely on standard results regarding exponentially off-diagonally decaying matrices. We briefly recall them next.

DEFINITION 3.2. *A matrix  $A \in \mathbb{R}^{n \times m}$  is labeled an exponentially off-diagonally decaying matrix, if there exist constants  $K > 0$  and  $\omega \in (0, 1)$ , such that*

$$(3.22) \quad |A_{i,j}| \leq K \omega^{|i-j|}$$

for  $i = 1, \dots, n$  and  $j = 1, \dots, m$ .

Exponentially decaying matrices have been extensively studied in the context of smooth functions of sparse matrices; see, e.g., [2, 5, 1]. We will base our analysis on the following classical result regarding the decay rate for inverses of banded matrices.

LEMMA 3.3. *If  $A$  is a Hermitian positive definite and banded matrix, with bandwidth  $m$ , then  $A^{-1}$  is an exponentially off-diagonally decaying matrix, with constants*

$$(3.23a) \quad K = \max\{\sigma_{min}^{-1}(A), 1 + \sqrt{\kappa(A)}\},$$

$$(3.23b) \quad \omega = \left( \frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^{1/m},$$

where  $\sigma_{min}(A)$  and  $\kappa(A)$  denote the smallest singular value and the condition number of  $A$ , respectively.

*Proof.* For a proof, see, e.g., [8]. □

Lemma 3.3 essentially states that the entries of  $A^{-1}$  are bounded by an exponentially decaying function along each row or column, where the bound depends on the extreme singular values and the bandwidth of the matrix. Matrices with a high condition number and/or a large bandwidth can therefore result in a large  $K$  and  $\omega \approx 1$ , leading to a slow decay, whereas a low condition number and a small bandwidth result in a rapid decay.

As a direct consequence of Lemma 3.3, we can conclude that  $\Lambda_k^{-1}$  is an exponentially off-diagonally decaying matrix, i.e.,

$$(3.24) \quad \left| (\Lambda_k^{-1})_{i,j} \right| \leq K_k \omega_k^{|i-j|}$$

with constants

$$(3.25a) \quad K_k = \max\{\sigma_{min}^{-1}(\Lambda_k), 1 + \sqrt{\kappa(\Lambda_k)}\},$$

$$(3.25b) \quad \omega_k = \left( \frac{\sqrt{\kappa(\Lambda_k)} - 1}{\sqrt{\kappa(\Lambda_k)} + 1} \right)^{1/2m}.$$

From the structure of  $\Lambda_k$  as given in (3.20), we note that the bandwidth of  $\Lambda_k$  is small compared to the size of the matrix if  $mN$  is large compared to  $2m$ . This indicates that the problem dimension  $N$  is important for the decay rate of  $\Lambda_k$  and, consequently, for the numerical structure of the dual Hessian.

Let us now introduce the additional notations  $[\Lambda_k^{-1}]_{\tilde{M}}$  and  $[\Lambda_k^{-1}]_{\tilde{M}}^C$  defined as for the dual Hessian, but using another bandwidth  $\tilde{M}$ , i.e.,

$$(3.26) \quad ([\Lambda_k^{-1}]_{\tilde{M}})_{i,j} = \begin{cases} (\Lambda_k^{-1})_{i,j}, & |i-j| \leq \tilde{M}, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$(3.27) \quad [\Lambda_k^{-1}]_{\tilde{M}}^C = \Lambda_k^{-1} - [\Lambda_k^{-1}]_{\tilde{M}}.$$

Matrix  $[\Lambda_k^{-1}]_{\tilde{M}}$  is therefore banded, with bandwidth  $\tilde{M}$ , whereas  $[\Lambda_k^{-1}]_{\tilde{M}}^C$  has a diagonal band of zeros. In the following, we will use the exponentially off-diagonally decaying property of  $\Lambda_k^{-1}$  to bound the spectral norm of  $[\Lambda_k^{-1}]_{\tilde{M}}^C$ . This result is then used to quantify the numerical structure of the dual Hessian. Using (3.24), we can establish the following bound on  $[\Lambda_k^{-1}]_{\tilde{M}}^C$ .

**PROPOSITION 3.4.** *For the constants  $K_k > 0$  and  $\omega_k \in (0, 1)$  defined in (3.25), the following bound holds:*

$$(3.28) \quad \left\| [\Lambda_k^{-1}]_{\tilde{M}}^C \right\|_2 \leq K_k \frac{\omega_k^{\tilde{M}} - \omega_k^{mN}}{1 - \omega_k}.$$

*Proof.* Observe that the first and last rows in  $[\Lambda_k^{-1}]_{\tilde{M}}^C$  contain the largest number of nonzero elements. Considering the first row, we note that the elements 1 to  $\tilde{M}$  are zero. Accordingly, using (3.24), we observe that the maximum absolute row sum of  $[\Lambda_k^{-1}]_{\tilde{M}}^C$  is bounded by

$$(3.29) \quad \left\| [\Lambda_k^{-1}]_{\tilde{M}}^C \right\|_{\infty} \leq \sum_{j=\tilde{M}+1}^{mN} K_k \omega_k^{1-j}.$$

If we rewrite (3.29) as

$$(3.30) \quad \left\| [\Lambda_k^{-1}]_{\tilde{M}}^C \right\|_{\infty} \leq \sum_{j=1}^{mN} K_k \omega_k^{|1-j|} - \sum_{j=1}^{\tilde{M}} K_k \omega_k^{|1-j|} = \sum_{j=0}^{mN-1} K_k \omega_k^j - \sum_{j=0}^{\tilde{M}-1} K_k \omega_k^j,$$

we can then use the formula for geometric series to simplify (3.30) as

$$(3.31) \quad \left\| [\Lambda_k^{-1}]_{\tilde{M}}^C \right\|_{\infty} \leq K_k \frac{\omega_k^{\tilde{M}} - \omega_k^{mN}}{1 - \omega_k}.$$

Let us now make the following observations. First, since  $[\Lambda_k^{-1}]_{\tilde{M}}^C$  is symmetric, its singular values coincide with the magnitude of its eigenvalues. Second, from Gershgorin's circle theorem, all eigenvalues are located within a Gershgorin disc [13]. Since the diagonal elements of  $[\Lambda_k^{-1}]_{\tilde{M}}^C$  are zero, all Gershgorin discs are centered at the origin, and the magnitude of its eigenvalues are accordingly bounded by the maximum absolute row sum. We then can conclude that:

$$(3.32) \quad \left\| [\Lambda_k^{-1}]_{\tilde{M}}^C \right\|_2 \leq K_k \frac{\omega_k^{\tilde{M}} - \omega_k^{mN}}{1 - \omega_k}. \quad \square$$

Accordingly,  $[\Lambda]_{\tilde{M}}^C$  is small if the bandwidth  $\tilde{M}$  is large or the constants  $K_k$  and  $\omega_k$  are small, i.e., typically if  $\Lambda_k$  is well-conditioned. In the following we will show how these conditions translate into a bound on the norm of  $[\nabla^2 d(\lambda)]_M^C$ . We first establish the following result.

PROPOSITION 3.5. *If  $M$  and  $\tilde{M}$  fulfill*

$$(3.33a) \quad M > p,$$

$$(3.33b) \quad 0 \leq \tilde{M} < M - m - 2p + 2,$$

then the following holds:

$$(3.34) \quad \left( [\nabla^2 d(\lambda)]_M^C \right)_{i,j} = - \left( \sum_{k=1}^P F_k H_k^{-1} C_k^T [\Lambda_k^{-1}]_{\tilde{M}}^C C_k H_k^{-1} F_k^T \right)_{i,j}$$

for  $|i - j| > M$ .

*Proof.* For notational simplicity, we introduce

$$(3.35a) \quad \mathcal{H}_k = F_k H_k^{-1} F_k^T,$$

$$(3.35b) \quad \mathcal{M}_{k,\tilde{M}} = F_k H_k^{-1} C_k^T [\Lambda_k^{-1}]_{\tilde{M}} C_k H_k^{-1} F_k^T,$$

$$(3.35c) \quad \mathcal{M}_{k,\tilde{M}}^C = F_k H_k^{-1} C_k^T [\Lambda_k^{-1}]_{\tilde{M}}^C C_k H_k^{-1} F_k^T,$$

and observe that

$$(3.36) \quad \nabla^2 d(\lambda) = \sum_{k=1}^P \left( \mathcal{H}_k - \mathcal{M}_{k,\tilde{M}} - \mathcal{M}_{k,\tilde{M}}^C \right).$$

Because matrices  $F_k$  and  $H_k^{-1}$  are banded, matrix  $\mathcal{H}_k$  is banded with bandwidth  $p$ . Additionally, because matrices  $F_k$ ,  $H_k^{-1}$ ,  $C_k$ , and  $[\Lambda_k^{-1}]_{\tilde{M}}$  are banded, matrix  $\mathcal{M}_{k,\tilde{M}}$

is banded with bandwidth  $\tilde{M} + m + 2p - 2$ . Matrix  $\mathcal{M}_{k,\tilde{M}}^C$  is dense due to factor  $[\Lambda_k^{-1}]_{\tilde{M}}^C$ . Let us recall the definition of  $[\nabla^2 d(\lambda)]_M^C$ , which reads as

$$(3.37) \quad \left([\nabla^2 d(\lambda)]_M^C\right)_{i,j} = \begin{cases} 0, & |i - j| \leq M, \\ (\nabla^2 d(\lambda))_{i,j}, & \text{otherwise.} \end{cases}$$

Accordingly, if (3.33) holds, then

$$(3.38a) \quad M > p,$$

$$(3.38b) \quad M > \tilde{M} + m + 2p - 2,$$

and matrices  $\mathcal{H}_k$  and  $\mathcal{M}_{k,\tilde{M}}$  have bandwidths smaller than  $M$ , such that they do not contribute to  $[\nabla^2 d(\lambda)]_M^C$ . Equations (3.33) and (3.34) then follow directly from (3.38) and (3.36).  $\square$

Proposition 3.5 entails that if  $\tilde{M}$  is sufficiently small compared to  $M$ , the banded matrix  $[\Lambda_k^{-1}]_{\tilde{M}}^C$  does not contribute to  $[\nabla^2 d(\lambda)]_M^C$ . We can then use Proposition 3.4 in order to bound the Frobenius norm of  $[\nabla^2 d(\lambda)]_M^C$ , which is the main result of this section.

LEMMA 3.6. *For  $M > p$  and  $\theta = M - m - 2p + 1 \geq 0$ , the following bound holds:*

$$(3.39) \quad \left\| [\nabla^2 d(\lambda)]_M^C \right\|_F \leq \sum_{k=1}^P \|F_k H_k^{-1} C_k^T\|_F^2 \sqrt{\tilde{m}} K_k \frac{\omega_k^\theta - \omega_k^{mN}}{1 - \omega_k}$$

for the constants  $K_k > 0$  and  $\omega_k \in (0, 1)$  defined in (3.25).

*Proof.* Recall that  $\Lambda_k^{-1}$  is of dimension  $\tilde{m} \times \tilde{m}$ , implying that  $[\Lambda_k^{-1}]_{\tilde{M}}^C$  is of rank at most  $\tilde{m}$ . From the equivalence of norms and (3.28), it follows that

$$(3.40) \quad \left\| [\Lambda_k^{-1}]_{\tilde{M}}^C \right\|_F \leq \sqrt{\tilde{m}} \left\| [\Lambda_k^{-1}]_{\tilde{M}}^C \right\|_2 \leq \sqrt{\tilde{m}} K_k \frac{\omega_k^{\tilde{M}} - \omega_k^{mN}}{1 - \omega_k}.$$

Furthermore, since the Froebinius matrix norm is submultiplicative, we observe that

$$(3.41) \quad \left\| F_k H_k^{-1} C_k^T [\Lambda_k^{-1}]_{\tilde{M}}^C C_k H_k^{-1} F_k^T \right\|_F \leq \|F_k H_k^{-1} C_k^T\|_F^2 \left\| [\Lambda_k^{-1}]_{\tilde{M}}^C \right\|_F.$$

Using (3.40) in (3.41) yields

$$(3.42) \quad \left\| F_k H_k^{-1} C_k^T [\Lambda_k^{-1}]_{\tilde{M}}^C C_k H_k^{-1} F_k^T \right\|_F \leq \|F_k H_k^{-1} C_k^T\|_F^2 \sqrt{\tilde{m}} K_k \frac{\omega_k^{\tilde{M}} - \omega_k^{mN}}{1 - \omega_k}.$$

Now, assuming that (3.33) holds, and specifically that  $\tilde{M}$  takes the largest value such that (3.33b) is fulfilled, i.e.,

$$(3.43) \quad \tilde{M} = M - m - 2p + 1,$$

we note from (3.34) that

$$(3.44) \quad \left| \left([\nabla^2 d(\lambda)]_M^C\right)_{i,j} \right| = \left| \left( \sum_{k=1}^P F_k H_k^{-1} C_k^T [\Lambda_k^{-1}]_{\tilde{M}}^C C_k H_k^{-1} F_k^T \right)_{i,j} \right|$$

for  $|i - j| > M$ . Furthermore, since  $[\nabla^2 d(\lambda)]_M^C$  has a diagonal band of zeros whereas the matrix on the right-hand-side of (3.44) generally is dense, we note that

$$(3.45) \quad \left| \left( [\nabla^2 d(\lambda)]_M^C \right)_{i,j} \right| \leq \left| \left( \sum_{k=1}^P F_k H_k^{-1} C_k^T [\Lambda_k^{-1}]_M^C C_k H_k^{-1} F_k^T \right)_{i,j} \right|$$

for  $|i - j| \leq M$ . Hence, it follows that

$$(3.46) \quad \left\| [\nabla^2 d(\lambda)]_M^C \right\|_F \leq \left\| \sum_{k=1}^P F_k H_k^{-1} C_k^T [\Lambda_k^{-1}]_{M-m-2p+1}^C C_k H_k^{-1} F_k^T \right\|_F.$$

The triangular inequality and (3.42) is then used to obtain the bound

$$(3.47) \quad \left\| [\nabla^2 d(\lambda)]_M^C \right\|_F \leq \sum_{k=1}^P \|F_k H_k^{-1} C_k^T\|_F^2 \sqrt{m} K_k \frac{\omega_k^{M-m-2p+1} - \omega_k^{mN}}{1 - \omega_k}$$

which yields (3.39). □

From (3.39), we observe that regardless of the size of the full Hessian,  $[\nabla^2 d(\lambda)]_M^C$  is small if the constants  $K_k$  and  $\omega_k$  are small, i.e., essentially if the problem dimension  $m$  is small and the matrix  $\Lambda_k$  is well-conditioned. For a problem originating from the control of dynamical systems, note that the problem dimension  $m$  represents the number of states. In section 5, the practical implications of Lemma 3.6 are evaluated via a numerical example. Finally, we extend Lemma 3.6 to the spectral norm, due to its usefulness in a convergence analysis.

**PROPOSITION 3.7.** *For  $M > p$  and  $\theta = M - m - 2p + 1 \geq 0$ , the following bound holds:*

$$(3.48) \quad \left\| [\nabla^2 d(\lambda)]_M^C \right\|_2 \leq \sum_{k=1}^P \|F_k H_k^{-1} C_k^T\|_F^2 \sqrt{m} K_k \frac{\omega_k^\theta - \omega_k^{mN}}{1 - \omega_k}$$

for the constants  $K_k$  and  $\omega_k$  defined in (3.25).

*Proof.* Since the euclidean norm is bounded by the Frobenius norm, (3.48) follows directly from (3.39). □

**3.2. A distributed quasi-Newton strategy.** In this section, we propose a quasi-Newton strategy based on the results presented in section 3.1.

**3.2.1. Algorithm.** Based on the numerical structure in the Hessian of the dual function, we propose to solve (2.1) using search directions provided by the quasi-Newton system (3.10); see Algorithm 3.1.

The quasi-Newton strategy is to a large extent separable, and can accordingly be performed in a distributed or parallel fashion. More specifically, evaluating the dual function and calculating its gradient and Hessian are separable, because of the decomposable structures of (3.4), (3.6), and (3.8). This implies that the quasi-Newton system can be formed by separable calculations, i.e., in a parallel manner. Depending on the structure of the coupling constraints, it can be viable to solve the quasi-Newton system using parallel or distributed factorization techniques; see, e.g., [7]. Additionally, note that for a geographically distributed problem, a banded approximation does not only reduce the computational complexity of one iteration, it also alleviates the

communication burden, since the corners of the Hessian can be omitted in the communication.

To ensure convergence, we choose the step size  $t \in (0, 1]$ . The choice is commonly based on strategies that involves evaluating the dual function, implying that the decision can be based on parallel or distributed calculations. In our implementation, we use an Armijo line search with backtracking [4]. An analysis of the convergence of Algorithm 3.1, in the neighborhood of the solution  $\lambda^*$ , is provided in section 3.2.2.

**Algorithm 3.1:** DUAL QUASI-NEWTON STRATEGY( $\lambda, M, \epsilon$ ).

**comment:** Solves (2.1) using the dual quasi-Newton strategy

**while** No convergence

**do**  $\left\{ \begin{array}{l} \text{Calculate } [\nabla^2 d(\lambda)]_M \text{ and } \nabla d(\lambda) \text{ in parallel} \\ \text{if } \|\nabla d(\lambda)\| \leq \epsilon \\ \quad \text{then break} \\ \text{Solve } [\nabla^2 d(\lambda)]_M \Delta\lambda = -\nabla d(\lambda) \\ \text{Find proper step size } t \in (0, 1] \\ \text{Update dual variables } \lambda \leftarrow \lambda + t\Delta\lambda \end{array} \right.$

**return** ( $\lambda$ )

**3.2.2. Convergence of the quasi-Newton strategy.** In this subsection, we investigate the impact of the Hessian approximation on the convergence of the quasi-Newton method.

In the following, we assume that  $M$  is chosen such that  $[\nabla^2 d(\lambda)]_M$  is positive definite. Note that positive definiteness for a sufficiently large  $M$  follows from positive definiteness of the full dual Hessian and Weyl's inequality for eigenvalues of Hermitian matrices [13]. However, the result turns out to be conservative in practice, whereas it appears to be difficult to find a practical example where the positiveness of  $[\nabla^2 d(\lambda)]_M$  is not fulfilled. Thus, we omit the result, and directly turn to analyze the behavior of the method in the neighborhood of the solution  $\lambda^*$ , i.e., where undamped Newton steps yield a quadratic convergence.

LEMMA 3.8. *If the dual Hessian is Lipschitz continuous with constant  $L$  in a neighborhood of the solution  $\lambda^*$ , then the following bound holds:*

$$(3.49) \quad \|\nabla d(\lambda + \Delta\lambda)\|_2 \leq \alpha \|\nabla d(\lambda)\|_2^2 + g(M) \|\nabla d(\lambda)\|_2$$

for  $\lambda$  in the neighborhood, where

$$(3.50a) \quad g(M) = \|\nabla d(\lambda)]_M^C [\nabla d(\lambda)]_M^{-1}\|_2,$$

$$(3.50b) \quad \alpha = \frac{1}{2}L \|\nabla^2 d(\lambda)]_M\|_2^2.$$

*Proof.* Recall that a step in the quasi-Newton method is given by the linear system

$$(3.51) \quad [\nabla^2 d(\lambda)]_M \Delta\lambda + \nabla d(\lambda) = 0.$$

Using (3.51), we observe that

$$\begin{aligned}
 \|\nabla d(\lambda + \Delta\lambda)\|_2 &= \|\nabla d(\lambda + \Delta\lambda) - \nabla d(\lambda) - [\nabla^2 d(\lambda)]_M \Delta\lambda\|_2 \\
 &= \left\| \int_0^1 \nabla^2 d(\lambda + t\Delta\lambda) \Delta\lambda dt - [\nabla^2 d(\lambda)]_M \Delta\lambda \right\|_2 \\
 (3.52) \quad &= \left\| \left( \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda) - \nabla^2 d(\lambda)) dt + \nabla^2 d(\lambda) - [\nabla^2 d(\lambda)]_M \right) \Delta\lambda \right\|_2 \\
 &= \left\| \left( \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda) - \nabla^2 d(\lambda)) dt + [\nabla^2 d(\lambda)]_M^C \right) [\nabla^2 d(\lambda)]_M^{-1} \nabla d(\lambda) \right\|_2.
 \end{aligned}$$

Consequently, using (3.52), the Cauchy–Schwarz inequality, and the triangle inequality, we can bound  $\|\nabla d(\lambda + \Delta\lambda)\|_2$  according to

$$(3.53) \quad \|\nabla d(\lambda + \Delta\lambda)\|_2 \leq \alpha \|\nabla d(\lambda)\|_2^2 + \|[\nabla^2 d(\lambda)]_M^C [\nabla^2 d(\lambda)]_M^{-1}\|_2 \|\nabla d(\lambda)\|_2,$$

where  $\alpha$  fulfills

$$(3.54) \quad \left\| \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda) - \nabla^2 d(\lambda)) dt [\nabla^2 d(\lambda)]_M^{-1} \right\|_2 \leq \alpha \|\nabla d(\lambda)\|_2.$$

By the definition of Lipschitz continuity, the following criterion holds:

$$(3.55) \quad \|\nabla^2 d(\lambda + \Delta\lambda) - \nabla^2 d(\lambda)\|_2 \leq L \|\Delta\lambda\|_2.$$

Moreover, observe that

$$\begin{aligned}
 &\left\| \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda) - \nabla^2 d(\lambda)) dt [\nabla^2 d(\lambda)]_M^{-1} \right\|_2 \\
 (3.56) \quad &\leq \left\| \int_0^1 (\nabla^2 d(\lambda + t\Delta\lambda) - \nabla^2 d(\lambda)) dt \right\|_2 \left\| [\nabla^2 d(\lambda)]_M^{-1} \right\|_2 \\
 &\leq \int_0^1 \|\nabla^2 d(\lambda + t\Delta\lambda) - \nabla^2 d(\lambda)\|_2 dt \left\| [\nabla^2 d(\lambda)]_M^{-1} \right\|_2 \\
 &\leq \frac{1}{2} L \|\Delta\lambda\|_2 \left\| [\nabla^2 d(\lambda)]_M^{-1} \right\|_2 \leq \frac{1}{2} L \left\| [\nabla^2 d(\lambda)]_M^{-1} \right\|_2^2 \|\nabla d(\lambda)\|_2.
 \end{aligned}$$

From (3.56), we can identify that the condition (3.54) is fulfilled for  $\alpha$  according to (3.50b). □

Observe that the (local) Lipschitz continuity of the Hessian follows from self-concordance of the dual function. A detailed description regarding self-concordance of the dual function can be found in [21].

Lemma 3.8 entails that choosing a small  $M$ , resulting in a nonnegligible value of  $g(M)$ , results in a degraded local linear convergence. The degradation is at worst linear in the norm of the part of the Hessian that is neglected, and inversely proportional to the lowest singular value of the dual Hessian. This implies that there is a trade-off between using a large  $M$  to reduce the number of iterations and a small  $M$  for reducing the computational cost for each iteration. A study of the trade-off is, however, beyond the scope of this paper.

Next, we provide a bound on  $g(M)$ , following directly from Lemma 3.6, showing that  $g(M)$  is small when the constants  $K_k$  and  $\omega_k$  are small, i.e., when  $\Lambda_k^{-1}$  is decaying rapidly.

COROLLARY 3.9. For  $M > p$  and  $\theta = M - m - 2p + 1 \geq 0$ , the following bound holds:

$$(3.57) \quad g(M) \leq \left\| [\nabla^2 d(\lambda)]_M^{-1} \right\|_2 \sum_{k=1}^M \|F_k H_k^{-1} C_k^T\|_F^2 \sqrt{m} K_k \frac{\omega_k^\theta - \omega_k^{mN}}{1 - \omega_k}.$$

*Proof.* By using the Cauchy–Schwarz inequality and (3.39), (3.57) follows directly from (3.50a).  $\square$

**4. Extension to inequality constrained problems.** In many practical applications, the decision variables are subject to inequality constraints. In this section, we consider the extension of (2.1), where local inequality constraints are included in the problem formulation, i.e., we consider the problem

$$(4.1a) \quad \min_z \sum_{k=1}^P f_k(z_k)$$

$$(4.1b) \quad \text{s.t.} \quad \sum_{k=1}^P F_k z_k = e,$$

$$(4.1c) \quad C_k z_k = d_k, \quad k = 1, \dots, P,$$

$$(4.1d) \quad G_k z_k \leq h_k, \quad k = 1, \dots, P,$$

where we have introduced  $G_k \in \mathbb{R}^{\bar{l} \times \bar{n}}$ , with the following structure:

$$(4.2) \quad G_k = \begin{bmatrix} G_{k,1} & & & \\ & \ddots & & \\ & & & G_{k,N} \end{bmatrix}.$$

Problems of the form (4.1), arise, e.g., when model predictive control is used to control multiagent systems, that are interacting via nondelayed couplings. In that context, the local equality constraints (4.1c) represent simulations of the agents dynamics, the local inequality constraints (4.1d) describe bounds on state and control variables, whereas the coupling constraints (4.1b) describe the nondelayed interactions between the agents.

Since the active inequality constraints can be concatenated with the local equality constraints while preserving a block-diagonal structure, the results from section 3.1 trivially extend to problems of the form (4.1). Consequently, by dualizing (4.1b), the Hessian of the dual function is numerically structured.

However, the resulting dual function is nonsmooth due to the local inequality constraints (4.1d) (see, e.g., [9]), and the results in section 3.2.2 do not hold in this context. To tackle the nonsmoothness, many methods have been proposed to obtain a smooth dual function, including strategies where a quadratic penalty term is introduced to form the augmented Lagrangian. The quadratic penalty, however, destroys the separability of the Lagrange function. In contrast to methods based on the augmented Lagrangian, it was proposed in [21] to add self-concordant barrier terms to the Lagrange function. The barrier terms preserve the separability of the problem, and results in a self-concordant dual function. Newton steps can then be used to efficiently trace the central path. Similar methods were proposed in, e.g., [25, 14, 12]. In this section, we extend the results in section 3 to the method proposed in [21].

**4.1. Dual decomposition using the relaxed dual function.** We add a self-concordant log-barrier to the Lagrange function, and define the relaxed dual function as

$$(4.3) \quad d(\lambda, \tau) = - \sum_{k=1}^P \min_{(z_k, s_k) \in \bar{\mathcal{Z}}_k} \left( \mathcal{L}_k(z_k, \lambda) - \tau \sum_{i=1}^{\bar{l}} \log((s_k)_i) \right),$$

where we have introduced the slack variables  $s_k \in \mathbb{R}^{\bar{l}}$  corresponding to (4.1d), the barrier parameter  $\tau > 0$ , and  $\bar{\mathcal{Z}}_k = \{(z_k, s_k) | C_k z_k = d_k, G_k z_k + s_k = h_k, s_k > 0\}$ . Note that the separability of the problem is preserved, and that the relaxed dual function can be evaluated in a distributed fashion. Additionally, observe that  $d(\lambda, \tau)$  can be made to be an arbitrary good approximation of  $d(\lambda)$  by making  $\tau$  small. We define the *relaxed dual problem* as

$$(4.4) \quad \min_{\lambda} d(\lambda, \tau)$$

and note that a sequence of problems  $\{\min_{\lambda} d(\lambda, \tau)\}_{\tau \rightarrow 0}$  can be solved, in order to find an arbitrary good approximation of the solution to (4.1). In particular, the relaxed dual problem is the dual problem of

$$(4.5) \quad \min_{(z, s) \in \bar{\mathcal{Z}}} \sum_{k=1}^P f_k(z_k) - \tau \sum_{k=1}^P \sum_{i=1}^{\bar{l}} \log((s_k)_i),$$

where  $\bar{\mathcal{Z}} = \bar{\mathcal{Z}}_1 \times \dots \times \bar{\mathcal{Z}}_P$ , i.e., the relaxed dual problem corresponds to the centering problem in an interior point method. Observe that (4.5) corresponds to a subclass of problems (2.1), implying that the results regarding the dual function extend to the relaxed dual function. Due to the importance of inequality constrained problems, this is detailed in the following.

In this section, we aim at solving (4.4) by updating the dual variables according to

$$(4.6) \quad \lambda^+ = \lambda + t\Delta\lambda,$$

where  $t \in (0, 1]$ , and the search direction  $\Delta\lambda$  is given by the quasi-Newton system

$$(4.7) \quad [\nabla^2 d(\lambda, \tau)]_M \Delta\lambda = -\nabla d(\lambda, \tau),$$

where we define  $[\nabla^2 d(\lambda, \tau)]_M$  as

$$(4.8) \quad ([\nabla^2 d(\lambda, \tau)]_M)_{i,j} = \begin{cases} (\nabla^2 d(\lambda, \tau))_{i,j}, & |i - j| \leq M, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly as (3.6), the gradient of the relaxed dual problem is given by

$$(4.9) \quad \nabla d(\lambda, \tau) = - \sum_{k=1}^P F_k z_k^*(\lambda, \tau) + e,$$

where  $z_k^*(\lambda, \tau)$  is defined by

$$(4.10) \quad z_k^*(\lambda, \tau) = \arg \min_{(z_k, s_k) \in \bar{\mathcal{Z}}_k} \mathcal{L}_k(z_k, \lambda) - \tau \sum_{i=1}^{\bar{l}} \log((s_k)_i).$$

Directly from (4.9), it follows that the Hessian of the relaxed dual function is given by

$$(4.11) \quad \nabla^2 d(\lambda, \tau) = - \sum_{k=1}^P F_k \frac{\partial z_k^*(\lambda, \tau)}{\partial \lambda}.$$

The procedure is summarized in Algorithm 4.1. Additionally, a strategy for solving a sequence of relaxed dual problems in order to find an arbitrary good approximation of the solution to (4.1) is provided in Algorithm 4.2; for a detailed description see, e.g., [21, 25, 14].

**Algorithm 4.1:** RELAXED DUAL QUASI-NEWTON STRATEGY( $\lambda, \tau$ ).

**comment:** Solves (4.4) using the dual quasi-Newton strategy

**while** No convergence  
     **do**  $\left\{ \begin{array}{l} \text{Calculate } [\nabla^2 d(\lambda, \tau)]_M \text{ and } \nabla d(\lambda, \tau) \\ \text{if } \|\nabla d(\lambda, \tau)\| \leq \epsilon_2 \\ \quad \text{then break} \\ \text{Solve } [\nabla^2 d(\lambda, \tau)]_M \Delta \lambda = -\nabla d(\lambda, \tau) \\ \text{Find proper step size } t \\ \text{Update dual variables } \lambda \leftarrow \lambda + t\Delta \lambda \end{array} \right.$   
**return** ( $\lambda$ )

**Algorithm 4.2:** SUCCESSIVE DUAL QUASI-NEWTON STRATEGY( $\lambda, \tau$ ).

**comment:** Solves (4.1) by successively calling Algorithm 4.1

**while** No convergence  
      $\lambda \leftarrow$  Relaxed dual quasi-Newton strategy( $\lambda, \tau$ )  
     **do**  $\left\{ \begin{array}{l} \text{if } \tau \leq \epsilon_2 \\ \quad \text{then break} \\ \tau \leftarrow \beta \tau \end{array} \right.$   
**return** ( $\lambda$ )

Since the results in section 3.1 are based on matrix properties, we will in the following show that the Hessian of the relaxed dual function is structurally equivalent to  $\nabla^2 d(\lambda)$ , implying that it is numerically structured. To that end, we introduce the notation  $y_k = \tau s_k^{-1} \in \mathbb{R}^l$  and find an explicit expression for the relaxed dual Hessian.

PROPOSITION 4.1. *The Hessian of  $d(\lambda, \tau)$  is given by*

$$(4.12) \quad \nabla^2 d(\lambda, \tau) = \sum_{k=1}^P F_k (\Phi_k^{-1} - \Phi_k^{-1} C_k^T \bar{\Lambda}_k^{-1} C_k \Phi_k^{-1}) F_k^T,$$

where

$$(4.13a) \quad \Phi_k = H_k + G_k^T S_k^{-1} Y_k G_k,$$

$$(4.13b) \quad \bar{\Lambda}_k = C_k \Phi_k^{-1} C_k^T,$$

where  $S_k = \text{diag}(s_k)$  and  $Y_k = \text{diag}(y_k)$ .

*Proof.* The optimality conditions of (4.10) can be stated as

$$\begin{aligned}
 (4.14a) \quad & 0 = \nabla f_k(z_k^*(\lambda, \tau)) + F_k^T \lambda + C_k^T \mu_k^*(\lambda, \tau) + G_k^T y_k^*(\lambda, \tau), \\
 (4.14b) \quad & 0 = C_k z_k^*(\lambda, \tau) - d_k, \\
 (4.14c) \quad & 0 = G_k z_k^*(\lambda, \tau) + s_k^*(\lambda, \tau) - h_k, \\
 (4.14d) \quad & 0 = Y_k s_k^*(\lambda, \tau) - \tau \mathbf{1}, \\
 (4.14e) \quad & s_k^*(\lambda, \tau) > 0, \quad y_k^*(\lambda, \tau) > 0.
 \end{aligned}$$

where  $Y_k = \text{diag}(y_k^*(\lambda, \tau))$  and  $\mathbf{1}$  represents a vector of ones. By differentiating (4.14), the following linear system is obtained:

$$\begin{aligned}
 (4.15a) \quad & 0 = H_k \frac{\partial z_k^*(\lambda, \tau)}{\partial \lambda} + F_k^T + C_k^T \frac{\partial \mu_k^*(\lambda, \tau)}{\partial \lambda} + G_k^T \frac{\partial y_k^*(\lambda, \tau)}{\partial \lambda}, \\
 (4.15b) \quad & 0 = C_k \frac{\partial z_k^*(\lambda, \tau)}{\partial \lambda}, \\
 (4.15c) \quad & 0 = G_k \frac{\partial z_k^*(\lambda, \tau)}{\partial \lambda} + \frac{\partial s_k^*(\lambda, \tau)}{\partial \lambda}, \\
 (4.15d) \quad & 0 = Y_k \frac{\partial s_k^*(\lambda, \tau)}{\partial \lambda} + S_k \frac{\partial y_k^*(\lambda, \tau)}{\partial \lambda},
 \end{aligned}$$

where  $S_k = \text{diag}(s_k^*(\lambda, \tau))$ . The expression (4.12) follows directly from (4.11) and (4.15).  $\square$

Let us now observe the following. Since  $Y_k$  and  $S_k^{-1}$  are positive definite,  $\Phi_k$  is positive definite. Additionally, note the structural equivalences between  $\Phi_k$  and  $H_k$  (see (3.14)), and consequently between  $\bar{\Lambda}_k$  and  $\Lambda_k$  (see (3.16)). Since the results of section 3.1 are based on properties of matrices, a straightforward extension can be made for the relaxed dual function, i.e., the Hessian of the relaxed dual function is numerically structured. Here, we only extend the main result of section 3.1.

**COROLLARY 4.2.** *For  $M > p$  and  $\theta = M - m - 2p + 1 \geq 0$ , the following bound holds:*

$$(4.16) \quad \left\| [\nabla^2 d(\lambda, \tau)]_M^C \right\|_F \leq \sum_{k=1}^P \|F_k \Phi_k^{-1} C_k^T\|_F^2 \sqrt{m} \bar{K}_k \frac{\bar{\omega}_k^\theta - \bar{\omega}_k^{mN}}{1 - \bar{\omega}_k}$$

for the constants

$$\begin{aligned}
 (4.17a) \quad & \bar{K}_k = \max\{\sigma_{\min}^{-1}(\bar{\Lambda}_k), 1 + \sqrt{\kappa(\bar{\Lambda}_k)}\}, \\
 (4.17b) \quad & \bar{\omega}_k = \left( \frac{\sqrt{\kappa(\bar{\Lambda}_k)} - 1}{\sqrt{\kappa(\bar{\Lambda}_k)} + 1} \right)^{1/2m}.
 \end{aligned}$$

*Proof.* See Lemma 3.6 for the proof.  $\square$

Accordingly, we observe that  $[\nabla^2 d(\lambda, \tau)]_M^C$  is small if the constants  $\bar{K}_k$  and  $\bar{\omega}_k$  are small, i.e., essentially if the problem dimension  $m$  is small and the matrix  $\bar{\Lambda}_k$  is well-conditioned. It should, however, be emphasized that the condition number of  $\bar{\Lambda}_k$  can be high when some elements in the factor  $S_k^{-1} Y_k$  are large compared to  $H_k$ . Typically, this happens when the barrier parameter is small. In our practical experiments, however, we have observed a rapid and consistent decay in  $\bar{\Lambda}_k$  for all problems and all values of the barrier parameter  $\tau$ .

Similarly, the convergence results of section 3.2.2 are based on properties of matrices and on Lipschitz continuity of the dual Hessian. From self-concordance of the relaxed dual function, it follows that the relaxed dual Hessian is at least locally Lipschitz continuous. This implies that a straightforward extension can be made of the results in section 3.2.2 to the method presented in Algorithm 4.1.

As a final remark on the results in this section, note that the method proposed in [21] closely resembles an interior point method. In particular, for a primal-dual interior point method using the so-called normal equation form to calculate search directions, the normal equation corresponding to (4.1b) is closely related to the dual Newton system. This suggests that a similar numerical structure can be found in the normal equations if problems of the form (4.1) are tackled via an interior point method. In [15], this is analyzed for an interior point method used to solve problems originating from direct collocation.

**5. Numerical experiments.** In this section, we give a numerical example to illustrate the proposed results. Consider a quadratic program of the form (4.1), with  $P = 2$  and  $N = 30$ , i.e.,

$$(5.1) \quad f_k(z_k) = \frac{1}{2} z_k^T H_k z_k + c_k^T z_k, \quad k = 1, 2,$$

with the following data:

$$(5.2) \quad \begin{aligned} H_{k,1} = \cdots = H_{k,30} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & k = 1, 2, \\ F_{k,1} = \cdots = F_{k,30} &= \begin{bmatrix} 0 & 1 \end{bmatrix}, & k = 1, 2, \\ C_{k,1} = \cdots = C_{k,29} &= \begin{bmatrix} 1 & 0 \end{bmatrix}, & k = 1, 2, \\ D_{k,2} = \cdots = D_{k,30} &= \begin{bmatrix} 1 & 1 \end{bmatrix}, & k = 1, 2, \\ G_{k,1} = \cdots = G_{k,30} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & k = 1, 2, \\ h_{k,1} = \cdots = h_{k,30} &= \begin{bmatrix} 2 \\ 2 \end{bmatrix}, & k = 1, 2, \end{aligned}$$

and  $c_k$ ,  $e$ , and  $d_k$  are zero vectors. Additionally, we consider the relaxed dual problem in the case where  $\tau = 1$ . For  $\lambda = \mathbf{1} \in \mathbb{R}^{30}$ , the absolute value of the elements in the relaxed dual Hessian, are visualized in Figure 1. Observe that elements with a large absolute value are located close to the diagonal, whereas elements with a small absolute value are located close to the off-diagonal corners.

Let us now evaluate the convergence of Algorithm 4.1. In our implementation, we use an Armijo line search with backtracking to choose the step size  $t \in (0, 1]$  [4]. Moreover, the algorithm is initialized far from the solution at  $\lambda = 50 \cdot \mathbf{1}$ . By using the full Hessian, i.e.,  $M = 29$ , the convergence is visualized in Figure 2(a). Note that the method is converging linearly in its damped phase, and quadratically in its undamped phase, i.e., from iteration 9. For smaller bandwidths, the convergence is visualized in Figures 2(b) and 2(c). For small bandwidths it can be observed that the quadratic term in (3.49) is dominating far from the solution, whereas the linear term is more prominent close to the solution. Note that a small  $M$  can be used before the convergence is seriously hindered.

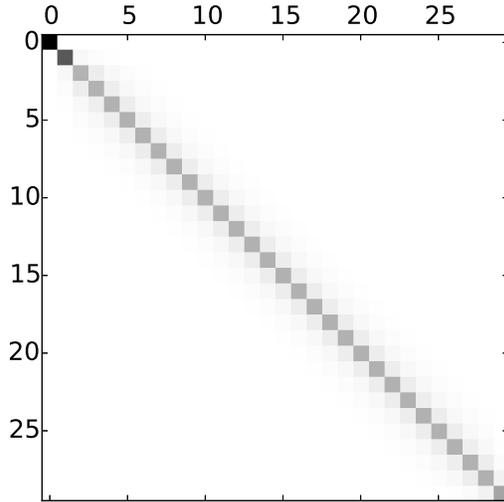


FIG. 1. A heat map illustrating the numerical structure of the relaxed dual Hessian evaluated at  $\lambda = \mathbf{1}$ ,  $\tau = 1$ . The intensity represents the magnitude of the elements in  $\nabla^2 d(\lambda, \tau) \in \mathbb{S}_{++}^{30}$ .

TABLE 1  
Evaluation of the Hessian approximation, evaluated at  $\lambda = \mathbf{1}$ ,  $\tau = 1$ .

| $M$ | $\ [\nabla^2 d(\lambda, \tau)]_M\ _F$ | $\ [\nabla^2 d(\lambda, \tau)]_M^C\ _F$ | Bound in (4.16)        |
|-----|---------------------------------------|---|------------------------|
| 20  | 4.0244                                | $3.1554 \cdot 10^{-13}$                 | $2.8730 \cdot 10^{-3}$ |
| 17  | 4.0244                                | $2.7623 \cdot 10^{-11}$                 | $2.3436 \cdot 10^{-2}$ |
| 15  | 4.0244                                | $5.2877 \cdot 10^{-10}$                 | $9.5057 \cdot 10^{-2}$ |
| 10  | 4.0244                                | $8.0137 \cdot 10^{-7}$                  | 3.1605                 |
| 9   | 4.0244                                | $3.4445 \cdot 10^{-6}$                  | 6.3736                 |
| 7   | 4.0244                                | $6.3346 \cdot 10^{-5}$                  | 25.937                 |
| 5   | 4.0244                                | $1.1589 \cdot 10^{-3}$                  | 105.64                 |
| 3   | 4.0243                                | $2.1113 \cdot 10^{-2}$                  | 430.61                 |

In Table 1, the Frobenius norms of  $[\nabla^2 d(\lambda, \tau)]_M$  and  $[\nabla^2 d(\lambda, \tau)]_M^C$  are summarized at  $\lambda = \mathbf{1}$ ,  $\tau = 1$ , for the bandwidths in Figures 2(b) and 2(c). Note that the neglected elements in the relaxed dual Hessian, are small compared to elements located close to the diagonal, even for small bandwidths. Additionally, the bound (4.16) on  $[\nabla^2 d(\lambda, \tau)]_M^C$  (see Corollary 4.2), is provided in Table 1. Note that the bound is nontight in practice.

**6. Conclusions and future research directions.** In this paper, we show that the Hessian of the Lagrange dual function, originating from a class of separable optimization problems, is numerically structured. The problem class is motivated by distributed model predictive control. Because of the numerical structure, a banded approximation of the Hessian can be used to form a quasi-Newton method. Using structure-exploiting factorization techniques for banded matrices, the quasi-Newton method can be used to solve problems at a lower computational cost compared to an exact Newton method.

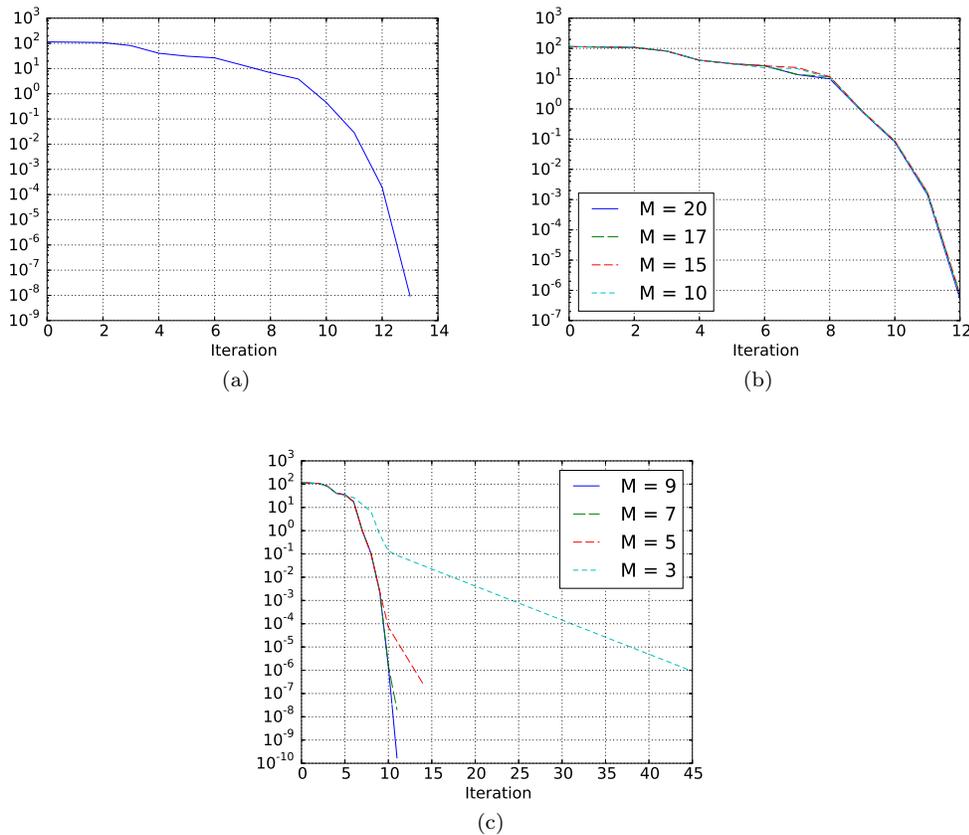


FIG. 2. Convergence of Algorithm 4.1, initialized at  $\lambda = 50 \cdot \mathbf{1}$  with  $\tau = 1$ , for (a): full Hessian; (b)–(c): banded approximations of the Hessian.

The results were presented for problems without inequality constraints, and extended to the methods presented in [21, 25, 14]. The results were illustrated via a numerical example.

It should be emphasized here that there are several directions for future research. An analysis of the trade-off between the computational effort for one iteration of the quasi-Newton method and the increased number of iterations required to solve a particular problem should be explored. Alternatively, an online adaptation of the bandwidth, e.g., along the lines proposed in [15], could be considered. Additionally, the authors of [16, 17] propose to solve the Newton system using a conjugate-gradient method, in contrast to the direct method which is proposed in this paper. In that context, the numerically structured Hessian could serve as a basis for forming a practical and efficient preconditioner to ease the solution of the Newton system.

#### REFERENCES

- [1] M. BENZI, P. BOITO, AND N. RAZOUK, *Decay properties of spectral projectors with applications to electronic structure*, SIAM Rev., 55 (2013), pp. 3–64.
- [2] M. BENZI AND G. H. GOLUB, *Bounds for the entries of matrix functions with applications to preconditioning*, BIT, 39 (1999), pp. 417–438.
- [3] D. BERTSEKAS AND J. N. TSITSIKLIS, *Parallel and Distributed Computation: Numerical Methods*, Prentice Hall, Englewood Cliffs, NJ, 1989.

- [4] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge, University Press, Cambridge, 2004.
- [5] F. BRANDAO AND M. HORODECKI, *An area law for entanglement from exponential decay of correlations*, *Nature Phys.*, 9 (2013), pp. 721–726.
- [6] R. W. COTTLE, S. G. DUVALL, AND K. ZIKAN, *A Lagrangean relaxation algorithm for the constrained matrix problem*, *Naval Res. Logist. Quart.*, 33 (1986), pp. 55–76.
- [7] T. DAVIS, *Direct Methods for Sparse Linear Systems*, Fundam, Algorithms, SIAM, Philadelphia, 2006.
- [8] S. DEMKO, W. MOSS, AND P. SMITH, *Decay rates for inverses of band matrices*, *Math. Comp.*, 43 (1986), pp. 491–499.
- [9] J. FRASCH, S. SAGER, AND M. DIEHL, *A parallel quadratic programming method for dynamic optimization problems*, *Math. Program. Comput.*, 7 (2015), pp. 289–329.
- [10] P. GISELSSON, *Improved dual decomposition for distributed model predictive control*, in IFAC World Congress, Curran, Red Hook, NY, 2014 pp. 1203–1209.
- [11] P. GISELSSON AND A. RANTZER, *Distributed model predictive control with suboptimality and stability guarantees*, in Conference on Decision and Control, IEEE, Piscataway, NJ, 2010, pp. 7272–7277.
- [12] S. GROS, *A Newton algorithm for distributed semi-definite programs using the primal-dual interior-point method*, in Conference on Decision and Control, IEEE, Piscataway, NJ, 2014, PP.3222–3227.
- [13] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, 2nd ed., Cambridge University Press, New York, 2013.
- [14] E. KLINTBERG AND S. GROS, *A primal-dual Newton method for distributed quadratic programming*, in Conference on Decision and Control, IEEE, Piscataway, NJ, 2014, pp. 5843–5848.
- [15] E. KLINTBERG AND S. GROS, *An inexact interior point method for optimization of differential algebraic systems*, *Comput. Chem. Eng.*, 92 (2016), pp. 163–171.
- [16] A. KOZMA, J. V. FRASCH, AND M. DIEHL, *A Distributed Method for Convex Quadratic Programming Problems Arising in Optimal Control of Distributed Systems*, in Conference on Decision and Control, IEEE, Piscataway, NJ, 2013, pp. 1526–1531.
- [17] A. KOZMA, E. KLINTBERG, S. GROS, AND M. DIEHL, *An improved distributed dual Newton-CG method for convex quadratic programming problems*, in American Control Conference, IEEE, Piscataway, NJ, 2014, pp. 2324–2329.
- [18] L. S. LASDON, *Optimization Theory for Large Systems*, Macmillan, New York, 1970.
- [19] I. NECOARA AND V. NEDELCU, *Rate analysis of inexact dual first-order methods application to dual decomposition*, *IEEE Trans. Automat. Control*, 59 (2014), pp. 1232–1243.
- [20] I. NECOARA, V. NEDELCU, AND I. DUMITRACHE, *Parallel and distributed optimization methods for estimation and control in networks*, *J. Process Control*, 21 (2011), pp. 756 – 766.
- [21] I. NECOARA AND J. SUYKENS, *An interior-point Lagrangian decomposition method for separable convex optimization*, *J. Optim. Theory Appl.*, 143 (2009), pp. 567–588.
- [22] A. NEDIĆ, A. OZDAGLAR, AND P. PARRILO, *Constrained consensus and optimization in multi-agent networks*, *IEEE Trans. Automat. Control*, 55 (2010), pp. 922–938.
- [23] N. GARG AND J. KÖNEMANN, *Faster and simpler algorithms for multicommodity flow and other fractional packing problems*, *SIAM J. Comput.*, 37 (2007), pp. 630–652.
- [24] S. RICHTER, M. MORARI, AND C. JONES, *Towards computational complexity certification for constrained MPC based on Lagrange relaxation and the fast gradient method*, in Conference on Decision and Control and European Control Conference, IEEE, Piscataway, NJ, 2011, pp. 5223–5229.
- [25] Q. T. DINH, I. NECOARA, C. SAVORGNAN, AND M. DIEHL, *An inexact perturbed path-following method for Lagrangian decomposition in large-scale separable convex optimization*, *SIAM J. Optim.*, 23 (2013), pp. 95–125.
- [26] P. TSENG, *Dual ascent methods for problems with strictly convex costs and linear constraints: A unified approach*, *SIAM J. Control Optim.*, 28 (1990), pp. 214–242, <https://dx.doi.org/10.1137/0328011>.
- [27] A. VENKAT, I. HISKENS, J. RAWLINGS, AND S. WRIGHT, *Distributed MPC strategies with application to power system automatic generation control*, *IEEE Trans. Control Sys. Tech.*, 16 (2008), pp. 1192–126.