

IMPROVED AUTOMATIC COMPUTATION OF HESSIAN MATRIX SPECTRAL BOUNDS

MORITZ SCHULZE DARUP AND MARTIN MÖNNIGMANN

Abstract. This paper presents a fast and powerful method for the computation of eigenvalue bounds for Hessian matrices $\nabla^2\varphi(x)$ of nonlinear functions $\varphi : \mathcal{U} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ on hyperrectangles $\mathcal{B} \subset \mathcal{U}$. The method is based on a recently proposed procedure [9] for an efficient computation of spectral bounds using extended codelists. Both the approach from [9] and the one presented here substantially differ from established methods in that they do deliberately not use any interval matrices and thus result in a favorable numerical complexity of order $\mathcal{O}(n)N(\varphi)$, where $N(\varphi)$ denotes the number of operations needed to evaluate φ at a point in its domain.

We improve the method presented in [9] by exploiting sparsity, which naturally arises in the underlying codelists. The new method provides bounds that are as good as, or better than those from the most accurate existing method in about 82% of the test cases.

1. Introduction. We present important improvements for a recently proposed method (see [9]) for the efficient calculation of spectral bounds for Hessian matrices on hyperrectangles. The improvements build on a systematic treatment of sparsity of the involved matrices, which will be shown to result in significantly tighter eigenvalue bounds. The problem can concisely be summarized as follows. Let $\varphi : \mathcal{U} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice continuously differentiable function on an open set $\mathcal{U} \subseteq \mathbb{R}^n$ and let $\mathcal{B} = [\underline{x}_1, \bar{x}_1] \times \cdots \times [\underline{x}_n, \bar{x}_n]$ be a closed hyperrectangle in \mathcal{U} . We seek bounds $\underline{\lambda}, \bar{\lambda} \in \mathbb{R}$ such that the relations $\underline{\lambda} \leq \lambda \leq \bar{\lambda}$ hold for all eigenvalues λ of all matrices $H \in \{\nabla^2\varphi(x) \mid x \in \mathcal{B}\}$. More precisely, we solve the following problem:

$$\text{Find } \underline{\lambda}, \bar{\lambda} \in \mathbb{R} \text{ such that } \underline{\lambda} \leq \min_{x \in \mathcal{B}} \lambda_{\min}(\nabla^2\varphi(x)) \text{ and } \max_{x \in \mathcal{B}} \lambda_{\max}(\nabla^2\varphi(x)) \leq \bar{\lambda}, \quad (1.1)$$

where $\lambda_{\min}(H)$ and $\lambda_{\max}(H)$ denote the smallest and largest eigenvalue, respectively, of the symmetric matrix $H \in \mathbb{R}^{n \times n}$. A bound $\underline{\lambda}$ (resp. $\bar{\lambda}$) is called *tight* if there exists at least one $x \in \mathcal{B}$ such that $\underline{\lambda} = \lambda_{\min}(\nabla^2\varphi(x))$ (resp. $\bar{\lambda} = \lambda_{\max}(\nabla^2\varphi(x))$). Note that the problem statement (1.1) does not necessarily imply that $\underline{\lambda}$ and $\bar{\lambda}$ are tight.

Eigenvalue bounds $\underline{\lambda}, \bar{\lambda}$ are used, for example, in numerical optimization methods to detect convexity, or to construct convex underestimators of nonconvex functions [1, 2, 3]. If (1.1) yields $\underline{\lambda} \geq 0$ then φ is convex on the interior of the hyperrectangle \mathcal{B} . While no conclusion on the convexity can be drawn if (1.1) results in $\underline{\lambda} < 0$, the bound $\underline{\lambda}$ can still be used to construct a convex underestimator for φ on \mathcal{B} . Specifically,

$$\check{\varphi}(x) = \varphi(x) - \frac{1}{2} \underline{\lambda} \sum_{i=1}^n (x_i - x_i)(\bar{x}_i - x_i) \quad (1.2)$$

is convex, coincides with φ at the vertices of \mathcal{B} , and bounds φ from below everywhere else in \mathcal{B} . Since a large fraction of the total computation time is spent on the calculation of convex underestimators in global optimization methods [1], fast methods for solving (1.1) are of interest. We briefly note that (1.1) must also be solved in certain problems in automatic control and systems theory. An illustrative example is given in [8].

Existing approaches to solving (1.1) proceed in two steps: First, a symmetric interval matrix (also called *interval Hessian*) that contains all Hessians $\nabla^2\varphi(x)$ on \mathcal{B} is calculated:

$$\text{Find } \underline{H} = \underline{H}^T, \bar{H} = \bar{H}^T \in \mathbb{R}^{n \times n} \text{ such that } \underline{H}_{ij} \leq (\nabla^2\varphi(x))_{ij} \leq \bar{H}_{ij} \quad (1.3)$$

for every $i, j \in \{1, \dots, n\}$ and every $x \in \mathcal{B}$.

This task can efficiently be carried out by combining interval arithmetics (IA, see [10], for example) and automatic differentiation (AD, see [4, 11], for example). In the second step, spectral bounds can be found by solving the following problem, which is similar to, but different from (1.1):

$$\text{Find } \underline{\lambda}, \overline{\lambda} \in \mathbb{R} \text{ such that } \underline{\lambda} \leq \min_{H \in \mathcal{H}} \lambda_{\min}(H) \text{ and } \max_{H \in \mathcal{H}} \lambda_{\max}(H) \leq \overline{\lambda}, \quad (1.4)$$

where $\mathcal{H} = \{H \in \mathbb{R}^{n \times n} \mid H_{ij} \in [\underline{H}_{ij}, \overline{H}_{ij}], H = H^T\}$ is the set of all symmetric matrices that respect the bounds \underline{H} and \overline{H} . Various approaches exist to solving (1.4) (see, e.g., [2, 5, 6, 12]). However, since $\{\nabla^2 \varphi(x) \mid x \in \mathcal{B}\} \subseteq \mathcal{H}$, problem (1.4) is conservative compared to the original problem (1.1). In fact it is the very point of the method introduced in [9] and refined in the present paper to avoid computing interval matrices of the form (1.3) when solving (1.1) in order to avoid this conservatism.

We briefly summarize the computational complexity of the existing methods. The computation of the matrices \underline{H} and \overline{H} in (1.3) requires $\mathcal{O}(n^2) N(\varphi)$ (resp. $\mathcal{O}(n) N(\varphi)$) operations if the forward (resp. backward) mode of AD is used, where $N(\varphi)$ denotes the number of operations needed to evaluate φ at a point in its domain [4]. After the interval Hessian has been calculated, solving (1.4) requires between $\mathcal{O}(n^2)$ operations for the interval variant of Gershgorin's circle criterion [2, 5] and $\mathcal{O}(2^n n^3)$ operations for Hertz and Rohn's method [6, 12]. The latter method is an important benchmark in that it yields *tight* spectral bounds for the matrix set \mathcal{H} . Albeit the conservatism in (1.3), Hertz and Rohn's method therefore provides the best possible option to solve (1.1) via (1.3) and (1.4).

The total numerical effort of any approach that uses (1.3) and (1.4) corresponds to the sum of the efforts for calculating $\underline{H}, \overline{H}$ and solving (1.4). Thus, the numerical effort for the established methods varies between $\mathcal{O}(n) N(\varphi) + \mathcal{O}(n^2)$ (backward mode AD combined with Gershgorin's circle criterion) and $\mathcal{O}(n^2) N(\varphi) + \mathcal{O}(2^n n^3)$ operations (forward mode AD combined with Hertz and Rohn's method). The major advantage of the direct method presented in [9] is its low computational complexity, which was shown to be of order $\mathcal{O}(n) N(\varphi)$.

It is the purpose of this paper to improve the method introduced in [9] such that sparsity can be exploited to find tighter eigenvalue bounds. The improvements do not increase the numerical effort compared to the original method in [9]. In fact, sparsity needs to be investigated once during the automatic generation of the extended codelist. The computations required to evaluate the codelist to obtain eigenvalue bounds on a specific hyperrectangle \mathcal{B} are no more expensive than those for the non-sparse case treated in [9]. While the computational effort remains the same, the improved method results in significantly tighter eigenvalue bounds than the original procedure from [9]. To show this, we investigate 1522 examples (taken from the COCONUT collection [14], see [13] for details) and compare the eigenvalue bounds resulting from the improved procedure to those obtained with the original one [9] and to bounds obtained with the interval Hessian (1.3) and Gershgorin's circle criterion and Hertz and Rohn's method.

We summarize the major aspects of the direct method for the computation of eigenvalue bounds from [9] in Sect. 3. Our main result, the exploitation of sparsity for the improvement of the eigenvalue bounds from [9], is stated in Sect. 4. We analyze 1522 numerical examples from [14, 13] in Sect. 5¹. Conclusions are given in Sect. 6.

¹Results were obtained with `Jcodegen`, a code generator available from the authors on request, or to be used online on www.rus.rub.de/software/jcodegen. `Jcodegen` generates ANSI-C code for the algorithm described in Prop. 4.17 for a given function φ . In particular sparsity is treated automatically. The specific hyperrectangle \mathcal{B} is passed to the resulting code as a runtime parameter.

2. Notation and Preliminaries. We frequently use index sets $\mathcal{J} \subseteq \mathcal{N}$, where $\mathcal{N} := \mathbb{N}_{1,n}$ and where $\mathbb{N}_{m,n} := \{i \in \mathbb{N} \mid m \leq i \leq n\}$. The complement of an index set \mathcal{J} is defined as $\mathcal{J}^c := \mathcal{N} \setminus \mathcal{J}$. The cardinality of an index set \mathcal{J} is denoted by $|\mathcal{J}|$.

It is convenient to state eigenvalue bounds as intervals (e.g. $\lambda \in [\underline{\lambda}, \overline{\lambda}]$). Intervals $[\underline{a}, \overline{a}] \subset \mathbb{R}$ with $\underline{a} \leq \overline{a}$ are further abbreviated by $[a] := [\underline{a}, \overline{a}]$ whenever appropriate. Interval equality $[\underline{a}, \overline{a}] = [\underline{b}, \overline{b}]$, is understood to mean $\underline{a} = \underline{b}$ and $\overline{a} = \overline{b}$. We frequently carry out calculations on intervals with standard interval arithmetics (IA) rules. The required rules are summarized in Lem. 2.1 and Tab. 2.1.

Lemma 2.1 (basic interval operations [10]). *Let $[a] = [\underline{a}, \overline{a}]$ and $[b] = [\underline{b}, \overline{b}]$ be intervals in \mathbb{R} . Let $a \in [\underline{a}, \overline{a}]$, $b \in [\underline{b}, \overline{b}]$, and $c \in \mathbb{R}$ be arbitrary real numbers. Then, the relations in the second column of Tab. 2.1 hold under the additional restrictions stated in the last column.*

TABLE 2.1
Basic interval arithmetic.

no.	operation / bounds	definition	restriction
1	$a + b \in [a] + [b]$	$:= [\underline{a} + \underline{b}, \overline{a} + \overline{b}]$	
2	$ab \in [a][b]$	$:= [\min\{\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}\}, \max\{\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}\}]$	
3	$1/a \in 1/[a]$	$:= [1/\overline{a}, 1/\underline{a}]$	$0 \notin [a]$
4	$a^m \in [a]^m$	$:= [\underline{a}^m, \overline{a}^m]$ if $\underline{a} > 0$ or m odd $:= [\overline{a}^m, \underline{a}^m]$ if $\overline{a} < 0$ and m even $:= [0, \max\{\underline{a}^m, \overline{a}^m\}]$ otherwise	
5	$\sqrt{a} \in \sqrt{[a]}$	$:= [\sqrt{\underline{a}}, \sqrt{\overline{a}}]$	$\underline{a} \geq 0$
6	$\exp a \in \exp([a])$	$:= [\exp(\underline{a}), \exp(\overline{a})]$	
7	$\ln(a) \in \ln([a])$	$:= [\ln(\underline{a}), \ln(\overline{a})]$	$\underline{a} > 0$
8	$a + c \in [a] + c$	$:= [\underline{a} + c, \overline{a} + c]$	
9	$ca \in c[a]$	$:= [c\underline{a}, c\overline{a}]$ if $c \geq 0$ $:= [c\overline{a}, c\underline{a}]$ otherwise	

A lower case letter surrounded by brackets may refer to a real interval $[x] = [\underline{x}, \overline{x}] \subset \mathbb{R}$ (as introduced above) or a hyperrectangle $[x] = [\underline{x}_1, \overline{x}_1] \times \cdots \times [\underline{x}_n, \overline{x}_n] \subset \mathbb{R}^n$ (with $n \geq 2$). In the latter case, the interval operations listed in Tab. 2.1 are understood to apply to every component. For a hyperrectangle $[x] \subset \mathbb{R}^n$ and a nonempty index set $\mathcal{J} \subseteq \mathcal{N}$ with the m elements $j_1 < \cdots < j_m$, the term $[x_{\mathcal{J}}]$ refers to the hyperrectangle $[x_{\mathcal{J}}] = [x_{j_1}] \times \cdots \times [x_{j_m}]$.

It is furthermore convenient to use *null matrices* in $\mathbb{R}^{m \times r}$, which we denote by $0_{m,r}$, when dealing with sparsity. For the special cases $m = 0$ or $r = 0$ we obtain an *empty matrix*. Formally, the empty square matrix $0_{0,0}$ has no eigenvalues. It proves useful to assign the eigenvalue bounds $[\lambda] = [0, 0]$ to it. Finally, the Cartesian unit vector along the k -th direction is denoted by $e_k \in \mathbb{R}^n$.

3. Direct computation of eigenvalue bounds for Hessian matrices on hyperrectangles. We summarize the method introduced in [9] for the direct solution of (1.1) as needed in the present paper. We assume the function φ can be evaluated

at an arbitrary point $x \in \mathcal{U}$ by carrying out a finite sequence of operations of the form

$$\begin{aligned}
y_1 &= x_1 \\
&\vdots \\
y_n &= x_n \\
y_{n+1} &= \Phi_{n+1}(y_1, \dots, y_n) \\
y_{n+2} &= \Phi_{n+2}(y_1, \dots, y_n, y_{n+1}) \\
&\vdots \\
y_{n+t} &= \Phi_{n+t}(y_1, \dots, y_n, y_{n+1}, \dots, y_{n+t-1}) \\
\varphi &= y_{n+t}
\end{aligned} \tag{3.1}$$

where each Φ_{n+k} , $k = 1, \dots, t$, represents one of the elementary operations listed in the first column of Tab. 3.1. We treat the same operations as in [9] for ease of comparison. Note that additional unary operations can be added according to the rules given in [9]. We refer to (3.1) as the *codelist* of the function φ .

The codelist (3.1) can be used to evaluate the function value $\varphi(x)$ at a specific point x in its domain. Using automatic differentiation (AD) [11] the codelist (3.1) can be extended in such a way that the gradient $\nabla\varphi(x)$ or the Hessian $\nabla^2\varphi(x)$ at the point x are calculated. Moreover, using AD and interval arithmetic (IA), (3.1) can be modified such that interval extensions, interval gradients or interval Hessians of φ on hyperrectangles $\mathcal{B} \subset \mathcal{U}$ are computed. In fact, extended codelists are commonly used to solve problem (1.3) as part of the established procedures for the computation of eigenvalue bounds (see, e.g., [2]). In contrast, the method introduced in [9] only requires the interval gradient, but not the interval Hessian. Essentially, the codelist is extended by arithmetic operations that compute the eigenvalue bounds for the Hessian of the intermediate function in every codelist line. Formally, this leads to the extended codelist which we introduce in the following theorem.

Theorem 3.1 (algorithm for direct eigenvalue bound computation [9, Prop. 4.2]). *Assume φ is twice continuously differentiable on \mathcal{U} and can be written as a codelist (3.1). Let $\mathcal{B} = [x_1] \times \dots \times [x_n] \subset \mathcal{U}$ be arbitrary. Then, for all $x \in \mathcal{B}$, we have $\varphi(x) \in [\varphi]$, $\nabla\varphi(x) \in [\nabla\varphi]$, and $[\lambda_{\min}(\nabla^2\varphi(x)), \lambda_{\max}(\nabla^2\varphi(x))] \subseteq [\lambda_\varphi]$, where $[\varphi]$, $[\nabla\varphi]$, and $[\lambda_\varphi]$ are calculated by the following algorithm.*

1. For $k = 1, \dots, n$, set $[y_k] = [\underline{x}_k, \bar{x}_k]$, $[\nabla y_k] = [e_k, e_k]$, and $[\lambda_k] = [0, 0]$.
2. For $k = n + 1, \dots, n + t$, calculate $[y_k]$, $[\nabla y_k]$ and $[\lambda_k]$ according to the third, fourth, and fifth column of Tab. 3.1.
3. Set $[\varphi] = [y_{n+t}]$, $[\nabla\varphi] = [\nabla y_{n+t}]$, and $[\lambda_\varphi] = [\lambda_{n+t}]$.

In Tab. 3.1, we use the interval operators $[\Lambda_s([a])]$ and $[\Lambda_t([a], [b])]$, which are defined according to

$$[\Lambda_s([a])] = \begin{cases} [a]^2 & \text{if } m = 1, \\ [0, \sum_{i=1}^m \max\{\underline{a}_i^2, \bar{a}_i^2\}] & \text{otherwise,} \end{cases} \tag{3.2}$$

$$\text{and } [\Lambda_t([a], [b])] = \begin{cases} 2[a][b] & \text{if } m = 1, \\ [-\beta, \beta] + \sum_{i=1}^m [\underline{a}_i, \bar{a}_i][\underline{b}_i, \bar{b}_i] & \text{otherwise} \end{cases} \tag{3.3}$$

for hyperrectangles $[a], [b] \subset \mathbb{R}^m$, where $\beta = \sqrt{(\sum_{i=1}^m \max\{\underline{a}_i^2, \bar{a}_i^2\})(\sum_{i=1}^m \max\{\underline{b}_i^2, \bar{b}_i^2\})}$. We refer to [9, Lems. 2.2 and 2.3] for details on $[\Lambda_s([a])]$ and $[\Lambda_t([a], [b])]$.

4. Improved computation of eigenvalue bounds using sparsity. If sparsity is exploited, tighter eigenvalue bounds can be obtained than those that result from

TABLE 3.1

Rules for the calculation of y_k , $[y_k]$, $[\nabla y_k]$ and $[\lambda_k]$ in the k -th line of the codelist (3.1). $[\nabla y_k]$ refers to the interval gradient of line k with respect to x . The interval operators $[\Lambda_s([a])]$ and $[\Lambda_t([a], [b])]$ are defined in (3.2) and (3.3).

op Φ_k	y_k	$[y_k]$	$[\nabla y_k]$	$[\lambda_k]$
var	x_k	$[x_k]$	$[e_k, e_k]$	$[0, 0]$
add	$y_i + y_j$	$[y_i] + [y_j]$	$[\nabla y_i] + [\nabla y_j]$	$[\lambda_i] + [\lambda_j]$
mul	$y_i y_j$	$[y_i][y_j]$	$[y_j][\nabla y_i] + [y_i][\nabla y_j]$	$[y_j][\lambda_i] + [y_i][\lambda_j] + [\Lambda_t([\nabla y_i], [\nabla y_j])]$
powNat	y_i^m	$[y_i]^m$	$m [y_i]^{m-1} [\nabla y_i]$	$m [y_i]^{m-2} ((m-1)[\Lambda_s([\nabla y_i])] + [y_i][\lambda_i])$
oneOver	$1/y_i$	$1/[y_i]$	$-[y_k]^{-2} [\nabla y_i]$	$[y_k]^{-2} (2 [y_k] [\Lambda_s([\nabla y_i])] - [\lambda_i])$
sqrt	$\sqrt{y_i}$	$[\sqrt{y_i}]$	$1/(2 [y_k]) [\nabla y_i]$	$1/(2 [y_k]) ([\lambda_i] + 1/(-2 [y_i]) [\Lambda_s([\nabla y_i])])$
exp	$\exp(y_i)$	$[\exp([y_i])]$	$[y_k] [\nabla y_i]$	$[y_k] ([\Lambda_s([\nabla y_i])] + [\lambda_i])$
ln	$\ln(y_i)$	$[\ln([y_i])]$	$1/[y_i] [\nabla y_i]$	$1/[y_i] ([\lambda_i] - 1/[y_i] [\Lambda_s([\nabla y_i])])$
addC	$y_i + c$	$[y_i] + [c, c]$	$[\nabla y_i]$	$[\lambda_i]$
mulByC	$c y_i$	$c [y_i]$	$c [\nabla y_i]$	$c [\lambda_i]$

the method summarized in Section 3. This is evident from the following motivating example.

Example 1 (method from [9] applied to $\varphi(x) = x_1^2 + x_2^2$). Consider the function $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ with $\varphi(x) = x_1^2 + x_2^2$. Theorem 3.1 results in the following extended codelist. Note that the expressions for y_k listed in (4.1) do not result from Thm. 3.1, but are only given for illustration of the codelist (3.1) of φ .

k	y_k	$[y_k]$	$[\nabla y_k]$	$[\lambda_k]$
1	x_1	$[x_1]$	$[e_1, e_1]$	$[0, 0]$
2	x_2	$[x_2]$	$[e_2, e_2]$	$[0, 0]$
3	y_1^2	$[y_1]^2$	$2 [y_1] [\nabla y_1]$	$2([\Lambda_s([\nabla y_1])] + [y_1] [\lambda_1])$
4	y_2^2	$[y_2]^2$	$2 [y_2] [\nabla y_2]$	$2([\Lambda_s([\nabla y_2])] + [y_2] [\lambda_2])$
5	$y_3 + y_4$	$[y_3] + [y_4]$	$[\nabla y_3] + [\nabla y_4]$	$[\lambda_3] + [\lambda_4]$
	$\varphi = y_5$	$[\varphi] = [y_5]$	$[\nabla \varphi] = [\nabla y_5]$	$[\lambda_\varphi] = [\lambda_5]$

Evaluating the extended codelist (4.1) for the hyperrectangle $\mathcal{B} = [0, 1] \times [0, 1]$ by computing $[y_k]$, $[\nabla y_k]$, and $[\lambda_k]$ and storing the results line by line yields

$$\begin{aligned}
[y_1] &= [0, 1], & [\nabla y_1] &= ([1, 1], [0, 0])^T, & [\lambda_1] &= [0, 0], \\
[y_2] &= [0, 1], & [\nabla y_2] &= ([0, 0], [1, 1])^T, & [\lambda_2] &= [0, 0], \\
[y_3] &= [0, 1], & [\nabla y_3] &= ([0, 2], [0, 0])^T, & [\lambda_3] &= [0, 2], \\
[y_4] &= [0, 1], & [\nabla y_4] &= ([0, 0], [0, 2])^T, & [\lambda_4] &= [0, 2], \\
[y_5] &= [0, 2], & [\nabla y_5] &= ([0, 2], [0, 2])^T, & [\lambda_5] &= [0, 4],
\end{aligned} \tag{4.2}$$

where $[\Lambda_s([\nabla y_1])] = [0, 1]$ and $[\Lambda_s([\nabla y_2])] = [0, 1]$ according to Eq. (3.3). Thus, we obtain the eigenvalue bounds $[\lambda_\varphi] = [\lambda_5] = [0, 4]$ for $\nabla^2(\varphi(x))$ on \mathcal{B} . Now, consider the functions $g, h : \mathbb{R}^2 \rightarrow \mathbb{R}$ with $g(x) = x_1^2$ and $h(x) = x_2^2$. From

$$\nabla^2 g(x) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \nabla^2 h(x) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix},$$

we infer that both $\nabla^2 g(x)$ and $\nabla^2 h(x)$ have the eigenvalues 0 and 2 for every $x \in \mathcal{B}$. Hence, the eigenvalue bounds $[\lambda_g] = [\lambda_3] = [0, 2]$ and $[\lambda_h] = [\lambda_4] = [0, 2]$ that result in line 3 and 4 of extended codelist (4.1) are tight. The eigenvalue bounds $[\lambda_5] = [\lambda_3] + [\lambda_4] = [0, 4]$ that result in the subsequent line are conservative, however. In fact,

the Hessian of φ reads

$$\nabla^2\varphi(x) = \nabla^2g(x) + \nabla^2h(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

for all $x \in \mathcal{B}$ and the tight eigenvalue bounds obviously read $[\lambda_\varphi^*] = [2, 2]$.

The Hessian matrices $\nabla^2g(x)$ and $\nabla^2h(x)$ in Exmp. 1 have zero eigenvalues which disappear when adding the two functions to $f(x) = g(x) + h(x)$. The situation illustrated in Example 1 arises naturally in the codelists introduced in Section 3, because codelists build up functions of many variables from functions of very few of these variables. In order to mitigate eigenvalue bound overestimation in these cases, we need to consider functions like $g(x)$ and $h(x)$ in Exmp. 1 as functions of only those variables that they actually depend on nonlinearly. To this end, some simple terminology and intermediate results are introduced in Sect. 4.1. Subsequently, sparse sums, products, and compositions are treated in Sects. 4.2, 4.3, and 4.4, respectively. Section 4.5 summarizes how to compute the improved eigenvalue bounds based on the rules introduced in Sects. 4.2–4.4.

4.1. Sparsity handling using reduced Hessians and reduced gradients.

As pointed out in Exmp. 1, sparsity occurs if functions depend at most linearly on some variables x_i , where *at most linear dependence* is defined as follows.

Definition 4.1 (at most linear dependence). *Let $f : \mathcal{U} \rightarrow \mathbb{R}$ be a continuously differentiable function on an open set $\mathcal{U} \subseteq \mathbb{R}^n$. Let $i \in \mathcal{N}$. The function f is said to depend at most linearly on x_i if there exists a $c \in \mathbb{R}$ such that*

$$\frac{\partial f}{\partial x_i}(x) = c \text{ for all } x \in \mathcal{U}. \quad (4.3)$$

The function f obviously is independent of x_i if (4.3) holds with $c = 0$.

Assume a function f is known to depend at most linearly on x_j for all $j \in \mathcal{L}_f$, where $\mathcal{L}_f \subseteq \mathcal{N}$ is a given index set. Then, only the eigenvalues of the *reduced Hessian* (see Def. 4.2) associated with the index set $\mathcal{J} = \mathcal{L}_f^c$ are nontrivial, i.e., not necessarily equal to zero.

Definition 4.2 (reduced Hessian $\nabla_{\mathcal{J}}^2 f(x)$). *Let $f : \mathcal{U} \rightarrow \mathbb{R}$ be a twice continuously differentiable function on an open set $\mathcal{U} \subseteq \mathbb{R}^n$. Let $\mathcal{J} \subseteq \mathcal{N}$ be an index set and let $m = |\mathcal{J}|$. If $m = 0$ set $\nabla_{\mathcal{J}}^2 f(x) = 0_{0,0}$, otherwise denote the m elements of \mathcal{J} by $j_1 < \dots < j_{n-m}$ in ascending order and define the reduced Hessian $\nabla_{\mathcal{J}}^2 f(x) \in \mathbb{R}^{(n-m) \times (n-m)}$ by its elements*

$$(\nabla_{\mathcal{J}}^2 f(x))_{ik} = \frac{\partial^2 f(x)}{\partial x_{j_i} \partial x_{j_k}},$$

where $i, k \in \mathbb{N}_{1,m}$.

We also need to consider reduced gradient vectors.

Definition 4.3 (reduced gradient $\nabla_{\mathcal{J}} f(x)$). *Let $f : \mathcal{U} \rightarrow \mathbb{R}$ be a continuously differentiable function on an open set $\mathcal{U} \subseteq \mathbb{R}^n$. Let $\mathcal{J} \subseteq \mathcal{N}$ be a nonempty index set and let $m = |\mathcal{J}|$. Denote the m elements of \mathcal{J} by $j_1 < \dots < j_{n-m}$ in ascending order and define the reduced gradient $\nabla_{\mathcal{J}} f(x) \in \mathbb{R}^{n-m}$ by its elements*

$$(\nabla_{\mathcal{J}} f(x))_i = \frac{\partial f(x)}{\partial x_{j_i}},$$

where $i \in \mathbb{N}_{1,m}$.

Note that \mathcal{J} may be empty in Def. 4.2, while there must exist at least one element in \mathcal{J} in Def. 4.3. This difference arises since codelist lines may depend at most linearly on all variables x_i but they are never independent of all x_i .

We can easily evaluate eigenvalue bounds for the Hessian of a function from eigenvalue bounds for its reduced Hessian. This is stated precisely in Lem. 4.4.

Lemma 4.4 (spectral bounds for Hessian from reduced Hessian). *Let f denote a twice continuously differentiable function $f : \mathcal{U} \rightarrow \mathbb{R}$ on an open set \mathcal{U} . Let the index set $\mathcal{L}_f \subseteq \mathcal{N}$ be such that f depends at most linearly on x_i for all $i \in \mathcal{L}_f$. Let $\mathcal{B} \subset \mathcal{U}$ and let the interval $[\lambda_f^\dagger] \subset \mathbb{R}$ be such that*

$$\Delta_f^\dagger \leq \min_{x \in \mathcal{B}} \lambda_{\min}(\nabla_{\mathcal{L}_f^c}^2 f(x)) \quad \text{and} \quad \max_{x \in \mathcal{B}} \lambda_{\max}(\nabla_{\mathcal{L}_f^c}^2 f(x)) \leq \bar{\lambda}_f^\dagger. \quad (4.4)$$

Then, the eigenvalues of the Hessian $\nabla^2 f(x)$ on \mathcal{B} lie in the interval

$$[\lambda_f] = \begin{cases} [\lambda_f^\dagger] & \text{if } \mathcal{L}_f = \emptyset, \\ [0, 0] & \text{if } \mathcal{L}_f = \mathcal{N}, \\ [\min\{\Delta_f^\dagger, 0\}, \max\{\bar{\lambda}_f^\dagger, 0\}] & \text{otherwise.} \end{cases} \quad (4.5)$$

Proof. We consider the cases in (4.5) separately. $\mathcal{L}_f = \emptyset$ implies $\mathcal{L}_f^c = \mathcal{N}$ and consequently $\nabla^2 f(x) = \nabla_{\mathcal{L}_f^c}^2 f(x)$, which proves the first case. In the second case, i.e., $\mathcal{L}_f = \mathcal{N}$, we have $\nabla^2 f(x) = \nabla_{\mathcal{L}_f}^2 f(x)$. Since f depends at most linearly on x_i for all $i \in \mathcal{L}_f$, we find $\nabla_{\mathcal{L}_f}^2 f(x) = 0_{n,n}$. Thus, the eigenvalue bounds $[\lambda_f] = [0, 0]$ hold. Regarding the third case, we note that $\emptyset \subset \mathcal{L}_f \subset \mathcal{N}$ implies $\emptyset \subset \mathcal{L}_f^c \subset \mathcal{N}$. Thus, $m = |\mathcal{L}_f|$ satisfies $0 < m < n$. Without loss of generality we assume $\mathcal{L}_f^c = \mathbb{N}_{1,m}$. Then

$$\nabla^2 f(x) = \begin{pmatrix} \nabla_{\mathcal{L}_f^c}^2 f(x) & 0_{m,n-m} \\ 0_{n-m,m} & \nabla_{\mathcal{L}_f}^2 f(x) \end{pmatrix} = \begin{pmatrix} \nabla_{\mathcal{L}_f^c}^2 f(x) & 0_{m,n-m} \\ 0_{n-m,m} & 0_{n-m,n-m} \end{pmatrix}.$$

Now consider an arbitrary but fixed $x \in \mathcal{B}$. We obtain

$$\lambda_{\min}(\nabla^2 f(x)) = \min\{\lambda_{\min}(\nabla_{\mathcal{L}_f^c}^2 f(x)), 0\} \quad \text{and} \quad (4.6)$$

$$\lambda_{\max}(\nabla^2 f(x)) = \max\{\lambda_{\max}(\nabla_{\mathcal{L}_f^c}^2 f(x)), 0\} \quad (4.7)$$

based on the block-diagonal structure of $\nabla_{\mathcal{L}_f^c}^2 f(x)$. Bounding (4.6) below and bounding (4.7) above for all $x \in \mathcal{B}$ yields

$$[\min_{x \in \mathcal{B}} \lambda_{\min}(\nabla^2 f(x)), \max_{x \in \mathcal{B}} \lambda_{\max}(\nabla^2 f(x))] \subseteq [\min\{\Delta_f^\dagger, 0\}, \max\{\bar{\lambda}_f^\dagger, 0\}]$$

according to Eqs. (4.6) and (4.7) and condition (4.4). \square

4.2. Improved eigenvalue bounds for the sum of two functions. We collect some recurring conditions first.

Conditions 4.5. *Let g and h denote twice continuously differentiable functions $g : \mathcal{U} \rightarrow \mathbb{R}$ and $h : \mathcal{U} \rightarrow \mathbb{R}$ on an open set $\mathcal{U} \subset \mathbb{R}^n$. Let the index sets $\mathcal{L}_g \subseteq \mathcal{N}$ and $\mathcal{L}_h \subseteq \mathcal{N}$ be such that g (resp. h) depends at most linearly on x_i for all $i \in \mathcal{L}_g$ (resp. all $i \in \mathcal{L}_h$). Moreover, let the index sets $\mathcal{I}_g \subseteq \mathcal{L}_g$ and $\mathcal{I}_h \subseteq \mathcal{L}_h$ with $\mathcal{I}_g \subset \mathcal{N}$ and $\mathcal{I}_h \subset \mathcal{N}$*

be such that g (resp. h) is independent of x_i for all $i \in \mathcal{I}_g$ (resp. all $i \in \mathcal{I}_h$). Let $\mathcal{B} \subset \mathcal{U}$ and assume there exist intervals $[\lambda_g^\dagger] \subset \mathbb{R}$ and $[\lambda_h^\dagger] \subset \mathbb{R}$ such that

$$\underline{\lambda}_g^\dagger \leq \min_{x \in \mathcal{B}} \lambda_{\min}(\nabla_{\mathcal{L}_g^c}^2 g(x)) \quad \text{and} \quad \max_{x \in \mathcal{B}} \lambda_{\max}(\nabla_{\mathcal{L}_g^c}^2 g(x)) \leq \bar{\lambda}_g^\dagger, \quad (4.8)$$

$$\underline{\lambda}_h^\dagger \leq \min_{x \in \mathcal{B}} \lambda_{\min}(\nabla_{\mathcal{L}_h^c}^2 h(x)) \quad \text{and} \quad \max_{x \in \mathcal{B}} \lambda_{\max}(\nabla_{\mathcal{L}_h^c}^2 h(x)) \leq \bar{\lambda}_h^\dagger. \quad (4.9)$$

Now assume Conds. 4.5 hold and we intend to calculate eigenvalue bounds for $\nabla^2 f(x)$ on a hyperrectangle \mathcal{B} for $f(x) = g(x) + h(x)$. We could determine eigenvalue bounds for the full Hessians $\nabla^2 g(x)$ and $\nabla^2 h(x)$ with Lem. 4.4 and apply the rule for the eigenvalue bounds of the sum of full Hessians (line **add** in Tab. 3.1 reproduced from [9]). However, we show in Lem. 4.8 below that it is advantageous to, roughly speaking, carry out calculations with the sparse Hessians as long as possible and to apply Lem. 4.4 as late as possible. We first state the rules for determining \mathcal{L}_f , \mathcal{I}_f and the eigenvalues of the reduced Hessian of f in Lems. 4.6 and 4.7, respectively. The trivial proof of Lemma 4.6 is omitted for brevity.

Lemma 4.6 (index sets for sums). *Assume Conds. 4.5 hold and consider the function $f : \mathcal{U} \rightarrow \mathbb{R}$ with $f(x) = g(x) + h(x)$. Let $\mathcal{L}_f = \mathcal{L}_g \cap \mathcal{L}_h$ and $\mathcal{I}_f = \mathcal{I}_g \cap \mathcal{I}_h$. Then, f depends at most linearly on x_i for all $i \in \mathcal{L}_f$ and f is independent of x_i for all $i \in \mathcal{I}_f$.*

Lemma 4.7 (spectral bounds for reduced Hessian of sums). *Assume Conds. 4.5 hold and consider the function $f : \mathcal{U} \rightarrow \mathbb{R}$, $f(x) = g(x) + h(x)$. Let $\mathcal{L}_f = \mathcal{L}_g \cap \mathcal{L}_h$. Then,*

$$\underline{\lambda}_f^\dagger \leq \min_{x \in \mathcal{B}} \lambda_{\min}(\nabla_{\mathcal{L}_f^c}^2 f(x)) \quad \text{and} \quad \max_{x \in \mathcal{B}} \lambda_{\max}(\nabla_{\mathcal{L}_f^c}^2 f(x)) \leq \bar{\lambda}_f^\dagger, \quad (4.10)$$

where $[\lambda_f^\dagger]$ is computed according to the rules listed in Tab. 4.1.

TABLE 4.1

Rules for the computation of eigenvalue bounds $[\lambda_f^\dagger]$ for the reduced Hessian $\nabla_{\mathcal{L}_f^c}^2 f(x)$ of a sum $f(x) = g(x) + h(x)$. Let \mathcal{L}_\cup be short for $\mathcal{L}_\cup := \mathcal{L}_g \cup \mathcal{L}_h$. See the end of Sect. 4.2 for a discussion of the eight cases.

case	$[\lambda_f^\dagger]$	condition
1	$[0, 0]$	$\mathcal{L}_g = \mathcal{N} \wedge \mathcal{L}_h = \mathcal{N}$
2	$[\lambda_g^\dagger]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_h = \mathcal{N}$
3	$[\lambda_h^\dagger]$	$\mathcal{L}_g = \mathcal{N} \wedge \mathcal{L}_h \subset \mathcal{N}$
4	$[\min\{\underline{\lambda}_g^\dagger, \underline{\lambda}_h^\dagger\}, \max\{\bar{\lambda}_g^\dagger, \bar{\lambda}_h^\dagger\}]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_h \subset \mathcal{N} \wedge \mathcal{L}_\cup = \mathcal{N}$
5	$[\lambda_g^\dagger] + [\lambda_h^\dagger]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_g = \mathcal{L}_h$
6	$[\lambda_g^\dagger] + [\min\{\underline{\lambda}_h^\dagger, 0\}, \max\{\bar{\lambda}_h^\dagger, 0\}]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_g \subset \mathcal{L}_h$
7	$[\min\{\underline{\lambda}_g^\dagger, 0\}, \max\{\bar{\lambda}_g^\dagger, 0\}] + [\lambda_h^\dagger]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_h \subset \mathcal{L}_g$
8	$[\min\{\underline{\lambda}_g^\dagger, 0\}, \max\{\bar{\lambda}_g^\dagger, 0\}] + [\min\{\underline{\lambda}_h^\dagger, 0\}, \max\{\bar{\lambda}_h^\dagger, 0\}]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_g \not\subset \mathcal{L}_h \wedge \mathcal{L}_h \not\subset \mathcal{L}_g$

Proof. We prove the fourth case in Tab. 4.1 since it will be instrumental for Exmp. 2. All other cases in Tab. 4.1 can be proven analogously. The reduced Hessian of f reads $\nabla_{\mathcal{L}_f^c}^2 f(x) = \nabla_{\mathcal{L}_g^c}^2 g(x) + \nabla_{\mathcal{L}_h^c}^2 h(x)$. From $\mathcal{L}_g \subset \mathcal{N}$, $\mathcal{L}_h \subset \mathcal{N}$, and $\mathcal{L}_\cup = \mathcal{L}_g \cup \mathcal{L}_h = \mathcal{N}$, we infer $\emptyset \subset \mathcal{L}_g \subset \mathcal{N}$ and $\emptyset \subset \mathcal{L}_h \subset \mathcal{N}$ and consequently $\emptyset \subset \mathcal{L}_g^c \subset \mathcal{N}$ and $\emptyset \subset \mathcal{L}_h^c \subset \mathcal{N}$. Thus, the cardinalities $r = |\mathcal{L}_g^c|$ and $s = |\mathcal{L}_h^c|$ satisfy $0 < r < n$ and $0 < s < n$. Moreover, $\mathcal{L}_g \cup \mathcal{L}_h = \mathcal{N}$ implies $\mathcal{L}_g^c \cap \mathcal{L}_h^c = \emptyset$. Hence, there does not

exist any index $i \in \mathcal{N}$ such that both $i \in \mathcal{L}_g^c$ and $i \in \mathcal{L}_h^c$. We assume $\mathcal{L}_g^c = \mathbb{N}_{1,r}$ and $\mathcal{L}_h^c = \mathbb{N}_{r+1,r+s}$ without loss of generality. Note that $\mathcal{L}_f^c = \mathcal{L}_g^c \cup \mathcal{L}_h^c$ implies $\mathcal{L}_f^c = \mathbb{N}_{1,r+s}$ and $m = |\mathcal{L}_f^c| = r + s$ under this assumption. Thus, $\nabla_{\mathcal{L}_f^c}^2 f(x)$ equals

$$\begin{pmatrix} \nabla_{\mathcal{L}_g^c}^2 g(x) & 0_{r,s} \\ 0_{s,r} & 0_{s,s} \end{pmatrix} + \begin{pmatrix} 0_{r,r} & 0_{r,s} \\ 0_{s,r} & \nabla_{\mathcal{L}_h^c}^2 h(x) \end{pmatrix} = \begin{pmatrix} \nabla_{\mathcal{L}_g^c}^2 g(x) & 0_{r,s} \\ 0_{s,r} & \nabla_{\mathcal{L}_h^c}^2 h(x) \end{pmatrix}. \quad (4.11)$$

The block-diagonal structure implies

$$\lambda_{\min}(\nabla_{\mathcal{L}_f^c}^2 f(x)) = \min\{\lambda_{\min}(\nabla_{\mathcal{L}_g^c}^2 g(x)), \lambda_{\min}(\nabla_{\mathcal{L}_h^c}^2 h(x))\} \quad \text{and} \quad (4.12)$$

$$\lambda_{\max}(\nabla_{\mathcal{L}_f^c}^2 f(x)) = \max\{\lambda_{\max}(\nabla_{\mathcal{L}_g^c}^2 g(x)), \lambda_{\max}(\nabla_{\mathcal{L}_h^c}^2 h(x))\} \quad (4.13)$$

for an arbitrary but fixed $x \in \mathcal{B}$. Bounding (4.12) below and bounding (4.13) above for all $x \in \mathcal{B}$ yields

$$[\min_{x \in \mathcal{B}} \lambda_{\min}(\nabla^2 f(x)), \max_{x \in \mathcal{B}} \lambda_{\max}(\nabla^2 f(x))] \subseteq [\min\{\underline{\Delta}_g^\dagger, \underline{\Delta}_h^\dagger\}, \max\{\overline{\lambda}_g^\dagger, \overline{\lambda}_h^\dagger\}]$$

where we used Eqs. (4.12) and (4.13) and Conds. 4.5. Thus, the eigenvalues of $\nabla_{\mathcal{L}_f^c}^2 f(x)$ on \mathcal{B} lie in the interval $[\lambda_f^\dagger] = [\min\{\underline{\Delta}_g^\dagger, \underline{\Delta}_h^\dagger\}, \max\{\overline{\lambda}_g^\dagger, \overline{\lambda}_h^\dagger\}]$ as claimed in Tab. 4.1. \square

We anticipated the bounds from Lem. 4.7 can be shown to be as tight as or tighter than those from the original method proposed in [9] that does not account for sparsity. This can now be shown in Lemma 4.8 below. Recall the bounds in [9] result in $[\lambda_f] = [\lambda_g] + [\lambda_h]$ for $f(x) = g(x) + h(x)$ according to [9, Prop 3.2.(iii)].

Lemma 4.8 (improved bounds for sums). *Assume Conds. 4.5 hold and let f , $[\lambda_f^\dagger]$, and \mathcal{L}_f be as in Lem. 4.7. Let $[\lambda_f]$, $[\lambda_g]$, and $[\lambda_h]$ be the eigenvalue bounds for the Hessians $\nabla^2 f(x)$, $\nabla^2 g(x)$, and $\nabla^2 h(x)$ on \mathcal{B} , calculated according to Eq. (4.5). Then,*

$$[\lambda_f] \subseteq [\lambda_g] + [\lambda_h]. \quad (4.14)$$

Proof. We prove the relation for the fourth case in Tab. 4.1. The remaining cases can be proven analogously. As pointed out in the proof of Lem. 4.7, we have $\emptyset \subset \mathcal{L}_g \subset \mathcal{N}$ and $\emptyset \subset \mathcal{L}_h \subset \mathcal{N}$. Thus, the r.h.s. in (4.14) yields

$$\begin{aligned} [\lambda_g] + [\lambda_h] &= [\min\{\underline{\Delta}_g^\dagger, 0\}, \max\{\overline{\lambda}_g^\dagger, 0\}] + [\min\{\underline{\Delta}_h^\dagger, 0\}, \max\{\overline{\lambda}_h^\dagger, 0\}] \\ &= [\min\{\underline{\Delta}_g^\dagger, 0\} + \min\{\underline{\Delta}_h^\dagger, 0\}, \max\{\overline{\lambda}_g^\dagger, 0\} + \max\{\overline{\lambda}_h^\dagger, 0\}], \\ &= [\min\{\underline{\Delta}_g^\dagger + \underline{\Delta}_h^\dagger, \underline{\Delta}_g^\dagger, \underline{\Delta}_h^\dagger, 0\}, \max\{\overline{\lambda}_g^\dagger + \overline{\lambda}_h^\dagger, \overline{\lambda}_g^\dagger, \overline{\lambda}_h^\dagger, 0\}], \end{aligned} \quad (4.15)$$

where the equations hold according to the third case in (4.5), by definition of the sum of two intervals (see Tab. 2.1), and by definition of $\min\{\cdot\}$ and $\max\{\cdot\}$, respectively. To evaluate the l.h.s. in (4.14), we have to analyze the index set \mathcal{L}_f . We obviously have $\mathcal{L}_f = \mathcal{L}_g \cap \mathcal{L}_h \subset \mathcal{N}$. Thus, the second case in Eq. (4.5) does not apply. However, from the conditions characterizing the fourth case in Tab. 4.1, it is not clear whether $\mathcal{L}_f = \emptyset$ or $\mathcal{L}_f \supset \emptyset$. Thus, according to (4.5), the l.h.s. in (4.14) results in

$$[\lambda_f] = \begin{cases} [\lambda_f^\dagger] & \text{if } \mathcal{L}_f = \emptyset, \\ [\min\{\underline{\Delta}_f^\dagger, 0\}, \max\{\overline{\lambda}_f^\dagger, 0\}] & \text{if } \emptyset \subset \mathcal{L}_f \subset \mathcal{N}. \end{cases}$$

However, since $[a] \subseteq [\min\{\underline{a}, 0\}, \max\{\bar{a}, 0\}]$, the relation $[\lambda_f] \subseteq [\min\{\underline{\lambda}_f^\dagger, 0\}, \max\{\bar{\lambda}_f^\dagger, 0\}]$ holds in both cases. Since $[\lambda_f^\dagger] = [\min\{\underline{\lambda}_g^\dagger, \underline{\lambda}_h^\dagger\}, \max\{\bar{\lambda}_g^\dagger, \bar{\lambda}_h^\dagger\}]$ according to Lem. 4.7 (resp. Tab. 4.1), we obtain

$$\begin{aligned} [\lambda_f] &\subseteq [\min\{\min\{\underline{\lambda}_g^\dagger, \underline{\lambda}_h^\dagger\}, 0\}, \max\{\max\{\bar{\lambda}_g^\dagger, \bar{\lambda}_h^\dagger\}, 0\}], \\ &= [\min\{\underline{\lambda}_g^\dagger, \underline{\lambda}_h^\dagger, 0\}, \max\{\bar{\lambda}_g^\dagger, \bar{\lambda}_h^\dagger, 0\}]. \end{aligned} \quad (4.16)$$

Comparing Eqs. (4.15) and (4.16) yields

$$[\min\{\underline{\lambda}_g^\dagger, \underline{\lambda}_h^\dagger, 0\}, \max\{\bar{\lambda}_g^\dagger, \bar{\lambda}_h^\dagger, 0\}] \subseteq [\min\{\underline{\lambda}_g^\dagger + \underline{\lambda}_h^\dagger, \underline{\lambda}_g^\dagger, \underline{\lambda}_h^\dagger, 0\}, \max\{\bar{\lambda}_g^\dagger + \bar{\lambda}_h^\dagger, \bar{\lambda}_g^\dagger, \bar{\lambda}_h^\dagger, 0\}],$$

which proves (4.14). \square

TABLE 4.2
Fundamental cases underlying Tab. 4.1.

case	reduced Hessian $\nabla_{\mathcal{L}_f^c}^2 f(x)$	contribution	condition
(i)	$0_{0,0}$	none	$\mathcal{L}_g^c = \emptyset \wedge \mathcal{L}_h^c = \emptyset$
(ii)	$\nabla_{\mathcal{L}_f^c}^2 g(x)$	first Hessian	$\mathcal{L}_g^c \supset \emptyset \wedge \mathcal{L}_h^c = \emptyset$
(iii)	$\nabla_{\mathcal{L}_f^c}^2 h(x)$	second Hessian	$\mathcal{L}_g^c = \emptyset \wedge \mathcal{L}_h^c \supset \emptyset$
(iv)	$\nabla_{\mathcal{L}_f^c}^2 g(x) + \nabla_{\mathcal{L}_f^c}^2 h(x)$	both Hessians	$\mathcal{L}_g^c \supset \emptyset \wedge \mathcal{L}_h^c \supset \emptyset$

Lemmas 4.7 and 4.8 are based on the eight cases listed in Tab. 4.1. Since they are not obvious at first sight, it is instructive to see how these eight cases arise from the four simpler ones listed in Tab. 4.2. The first case in Tab. 4.2 applies if both $\nabla_{\mathcal{L}_f^c}^2 g(x)$ and $\nabla_{\mathcal{L}_f^c}^2 h(x)$ vanish because of $\mathcal{L}_g^c = \mathcal{L}_h^c = \emptyset$. Since these two conditions, i.e. $\mathcal{L}_g^c = \emptyset \wedge \mathcal{L}_h^c = \emptyset$, are equivalent to $\mathcal{L}_g = \mathcal{N} \wedge \mathcal{L}_h = \mathcal{N}$, case (i) in Tab. 4.2 is equivalent to case 1 in Tab. 4.1. Analogously, cases (ii) and (iii) in Tab. 4.2, where either $\nabla_{\mathcal{L}_f^c}^2 g(x)$ or $\nabla_{\mathcal{L}_f^c}^2 h(x)$ contribute to $\nabla_{\mathcal{L}_f^c}^2 f(x)$, are equivalent to cases 2 and 3 in Tab. 4.1, respectively. It remains to relate case (iv) in Tab. 4.2 to cases 4–8 in Tab. 4.1. In fact, the conditions of the cases 4–8 in Tab. 4.1 all imply $\mathcal{L}_g^c \supset \emptyset$ and $\mathcal{L}_h^c \supset \emptyset$, which are the defining conditions for case (iv) in Tab. 4.2. Figure 4.1 illustrates that every instance of case (iv) from Tab. 4.2 actually uniquely belongs to one of the cases 4–8 from Tab. 4.1.



FIG. 4.1. Illustration of the sparsity patterns of the reduced Hessians $\nabla_{\mathcal{L}_f^c}^2 g(x)$ and $\nabla_{\mathcal{L}_f^c}^2 h(x)$ for the cases 4–8 in Tab. 4.1 and Lem. 4.7. White areas correspond to zero blocks in the Hessians, black areas to non-trivial blocks. The black areas in case 4 correspond to the non-trivial submatrices $\nabla_{\mathcal{L}_g^c}^2 g(x)$ and $\nabla_{\mathcal{L}_h^c}^2 h(x)$ on the l.h.s. of Eq. (4.11).

4.3. Improved eigenvalue bounds for the composition of two functions.

We collect some recurring conditions again first.

Conditions 4.9. Assume Conds. 4.5 hold. Let $r : \mathcal{V} \rightarrow \mathbb{R}$ be a twice differentiable function on an open set $\mathcal{V} \subset \mathbb{R}$ and assume $g(x) \in \mathcal{V}$ for every $x \in \mathcal{U}$. Moreover,

assume there exist intervals $[r'] \subset \mathbb{R}$ and $[r''] \subset \mathbb{R}$ and a hyperrectangle $[\nabla g] \subset \mathbb{R}^n$ such that

$$\underline{r}' \leq r'(g(x)) \leq \overline{r}', \quad \underline{r}'' \leq r''(g(x)) \leq \overline{r}'', \quad \text{and} \quad \underline{\nabla g}_i \leq (\nabla g(x))_i \leq \overline{\nabla g}_i \quad (4.17)$$

for every $x \in \mathcal{B}$ and every $i \in \mathcal{N}$, where $r'(z)$ and $r''(z)$ refer to the first and the second derivative of $r(z)$, respectively.

The following lemma, which we state without proof, provides rules for the identification of at most linear dependencies and independencies of compositions.

Lemma 4.10 (index sets for compositions). *Assume Conds. 4.9 hold and consider the function $f : \mathcal{U} \rightarrow \mathbb{R}$, $f(x) = r(g(x))$. Let*

$$\mathcal{I}_f = \mathcal{I}_g \quad \text{and} \quad \mathcal{L}_f = \begin{cases} \mathcal{L}_g & \text{if } r \text{ is an affine function,} \\ \mathcal{I}_g & \text{otherwise.} \end{cases} \quad (4.18)$$

Then, f depends at most linearly on x_i for all $i \in \mathcal{L}_f$ and f is independent of x_i for all $i \in \mathcal{I}_f$.

Bounds for compositions can now be calculated as follows.

Lemma 4.11 (spectral bounds for reduced Hessian of compositions). *Assume Conds. 4.9 hold and consider the function $f : \mathcal{U} \rightarrow \mathbb{R}$, $f(x) = r(g(x))$. Let \mathcal{L}_f be defined as in Lem. 4.10. Then the bounds (4.10) hold for $[\lambda_f^\dagger]$ computed according to the rules listed in Tab. 4.3.*

TABLE 4.3

Rules for the computation of eigenvalue bounds $[\lambda_f^\dagger]$ for the reduced Hessian $\nabla_{\mathcal{L}_f^c}^2 f(x)$ of compositions $f(x) = r(g(x))$.

case	$[\lambda_f^\dagger]$	condition
1	$[r''] [\Lambda_s([\nabla_{\mathcal{L}_f^c} g])]$	$\mathcal{L}_g = \mathcal{N}$
2	$[r''] [\Lambda_s([\nabla_{\mathcal{L}_f^c} g])] + [r'] [\lambda_g^\dagger]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_f = \mathcal{L}_g$
3	$[r''] [\Lambda_s([\nabla_{\mathcal{L}_f^c} g])] + [r'] [\min\{\underline{\lambda}_g^\dagger, 0\}, \max\{\overline{\lambda}_g^\dagger, 0\}]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_f \subset \mathcal{L}_g$

Proof. We prove the last case in Tab. 4.3. The remaining cases can be proven analogously. The reduced Hessian of f reads $\nabla_{\mathcal{L}_f^c}^2 f(x) = r''(g(x)) \nabla_{\mathcal{L}_f^c} g(x) \nabla_{\mathcal{L}_f^c}^T g(x) + r'(g(x)) \nabla_{\mathcal{L}_f^c}^2 g(x)$. Combining the two conditions of case 3 in Tab. 4.3 yields $\mathcal{L}_f \subset \mathcal{L}_g \subset \mathcal{N}$, which implies $\mathcal{L}_f^c \supset \mathcal{L}_g^c \supset \emptyset$. Thus, $m = |\mathcal{L}_f^c|$ and $r = |\mathcal{L}_g^c|$ satisfy $m > r > 0$. We assume $\mathcal{L}_f^c = \mathbb{N}_{1,m}$ and $\mathcal{L}_g^c = \mathbb{N}_{1,r}$ without loss of generality. Under these assumptions, we obtain

$$\nabla_{\mathcal{L}_f^c}^2 f(x) = r''(g(x)) \nabla_{\mathcal{L}_f^c} g(x) \nabla_{\mathcal{L}_f^c}^T g(x) + r'(g(x)) \begin{pmatrix} \nabla_{\mathcal{L}_g^c}^2 g(x) & 0_{r,s} \\ 0_{s,r} & 0_{s,s} \end{pmatrix}, \quad (4.19)$$

where $s = m - r > 0$. Since $\nabla g(x) \in [\nabla g]$ for every $x \in \mathcal{B}$, we find

$$\underline{\Lambda}_s([\nabla_{\mathcal{L}_f^c} g]) \leq \lambda_{\min}(\nabla_{\mathcal{L}_f^c} g(x) \nabla_{\mathcal{L}_f^c}^T g(x)) \quad \text{and} \quad \lambda_{\max}(\nabla_{\mathcal{L}_f^c} g(x) \nabla_{\mathcal{L}_f^c}^T g(x)) \leq \overline{\Lambda}_s([\nabla_{\mathcal{L}_f^c} g]),$$

for every $x \in \mathcal{B}$ according to [9, Lem. 2.2]. Combining this intermediate result with the bounds on $r'(g(x))$ and $r''(g(x))$ from Conds. 4.9 yields the eigenvalue bounds

$$[\lambda_f^\dagger] = [r''] [\Lambda_s([\nabla_{\mathcal{L}_f^c} g])] + [r'] [\min\{\underline{\lambda}_g^\dagger, 0\}, \max\{\overline{\lambda}_g^\dagger, 0\}]$$

on \mathcal{B} . \square

Lemma 4.12 below shows that the bounds from Lem. 4.11 are as tight as or tighter than those from the original method proposed in [9]. Recall the bounds in [9] result in $[\lambda_f] = [r''] [\Lambda_s([\nabla g])] + [r'] [\lambda_g]$ for $f(x) = r(g(x))$ according to [9, Prop 3.4].

Lemma 4.12 (improved bounds for compositions). *Assume Conds. 4.9 hold and let f , $[\lambda_f^\dagger]$, and \mathcal{L}_f be as in Lem. 4.11. Let $[\lambda_f]$ and $[\lambda_g]$ be the eigenvalue bounds for the Hessians $\nabla^2 f(x)$ and $\nabla^2 g(x)$ on \mathcal{B} , calculated according to Eq. (4.5). Then,*

$$[\lambda_f] \subseteq [r''] [\Lambda_s([\nabla g])] + [r'] [\lambda_g].$$

Since the proof is very similar to the proof of Lem. 4.8, we omit it.

4.4. Improved eigenvalue bounds for the product of two functions. We begin by collecting recurring conditions again.

Conditions 4.13. *Assume Conds. 4.5 hold and assume there exist intervals $[g] \subset \mathbb{R}$ and $[h] \subset \mathbb{R}$ and hyperrectangles $[\nabla g] \subset \mathbb{R}^n$ and $[\nabla h] \subset \mathbb{R}^n$ such that*

$\underline{g} \leq g(x) \leq \bar{g}$, $\underline{h} \leq h(x) \leq \bar{h}$, $\underline{\nabla g}_i \leq (\nabla g(x))_i \leq \bar{\nabla g}_i$, and $\underline{\nabla h}_i \leq (\nabla h(x))_i \leq \bar{\nabla h}_i$ for every $x \in \mathcal{B}$ and every $i \in \mathcal{N}$.

The following lemma provides rules for the identification of at most linear dependencies and independences of products.

Lemma 4.14 (index sets for products). *Assume Conds. 4.13 hold and consider the function $f : \mathcal{U} \rightarrow \mathbb{R}$, $f(x) = g(x) h(x)$. Let $\mathcal{I}_f = \mathcal{I}_g \cap \mathcal{I}_h$ and $\mathcal{L}_f = \mathcal{I}_g \cap \mathcal{I}_h$. Then, f depends at most linearly on x_i for all $i \in \mathcal{L}_f$ and f is independent of x_i for all $i \in \mathcal{I}_f$.*

Based on Conds. 4.13 and Lem. 4.14, we are able to compute bounds on the spectrum of $\nabla_{\mathcal{L}_f}^2 f(x)$ according to the rules summarized in Lem. 4.15 and Tab. 4.4. As a preparation, we introduce the interval operators

$$[\Lambda_r([a], [b])] := [\min\{\underline{a}, \underline{b}\}, \max\{\bar{a}, \bar{b}\}] \quad \text{and} \quad (4.20)$$

$$[\Lambda_\star([a], [b], [c])] := \frac{1}{2} \left[\underline{a} + \underline{b} - \sqrt{(\underline{a} - \underline{b})^2 + d}, \bar{a} + \bar{b} + \sqrt{(\bar{a} - \bar{b})^2 + d} \right] \quad (4.21)$$

for real intervals $[a], [b], [c] \subset \mathbb{R}$, where $d = 4 \max\{\underline{c}^2, \bar{c}^2\}$. Definition (4.20) is only introduced for the sake of a compact notation. Whenever it is more instructive, we use the notation on the r.h.s. of (4.20).

Lemma 4.15 (spectral bounds for reduced Hessian of products). *Assume Conds. 4.13 hold and consider the function $f : \mathcal{U} \rightarrow \mathbb{R}$ with $f(x) = g(x) h(x)$. Let \mathcal{L}_f be defined as in Lem. 4.14. Then the bounds (4.10) hold for $[\lambda_f^\dagger]$ computed according to the rules listed in Tab. 4.4.*

Proof. We prove case 10 from Tab. 4.4. Cases 4 and 7 can be shown analogously. The remaining cases can be proven in the same fashion as those treated in the proofs of Lems. 4.7 and 4.11. The reduced Hessian of f , which reads

$$\nabla_{\mathcal{L}_f}^2 f(x) = \nabla_{\mathcal{L}_f^c} g(x) \nabla_{\mathcal{L}_f^c}^T h(x) + \nabla_{\mathcal{L}_f^c} h(x) \nabla_{\mathcal{L}_f^c}^T g(x) + h(x) \nabla_{\mathcal{L}_f^c}^2 g(x) \quad (4.22)$$

in all cases, is a two-by-two matrix with a particularly simple block structure in case 10. To see this, first note that g and h are independent of all but one variable each (the conditions $|\mathcal{I}_g| = n - 1$ and $|\mathcal{I}_h| = n - 1$ imply $|\mathcal{I}_g^c| = 1$ and $|\mathcal{I}_h^c| = 1$).

TABLE 4.4

Rules for the computation of eigenvalue bounds $[\lambda_f^\dagger]$ for the reduced Hessian $\nabla_{\mathcal{L}_f^c}^2 f(x)$ of a product $f(x) = g(x)h(x)$. The expressions $[\lambda_t]$, $[\lambda_{g,0}]$, $[\lambda_{h,0}]$, \mathcal{L}_\cup , and \mathcal{L}_\cap are short for $[\lambda_t] = [\Lambda_t([\nabla_{\mathcal{L}_f^c} g], [\nabla_{\mathcal{L}_f^c} h])]$, $[\lambda_{g,0}] = [\min\{\underline{\lambda}_g^\dagger, 0\}, \max\{\bar{\lambda}_g^\dagger, 0\}]$, $[\lambda_{h,0}] = [\min\{\underline{\lambda}_h^\dagger, 0\}, \max\{\bar{\lambda}_h^\dagger, 0\}]$, $\mathcal{L}_\cup = \mathcal{L}_g \cup \mathcal{L}_h$ and $\mathcal{L}_\cap = \mathcal{L}_g \cap \mathcal{L}_h$. Condition C_\star reads $(\mathcal{I}_g \cup \mathcal{I}_h = \mathcal{N}) \wedge (|\mathcal{I}_g| = n-1) \wedge (|\mathcal{I}_h| = n-1)$.

case	$[\lambda_f^\dagger]$	condition
1	$[\lambda_t]$	$\mathcal{L}_g = \mathcal{N} \wedge \mathcal{L}_h = \mathcal{N}$
2	$[\lambda_t] + [h] [\lambda_g^\dagger]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_h = \mathcal{N} \wedge \mathcal{L}_f = \mathcal{L}_g$
3	$[\lambda_t] + [h] [\lambda_{g,0}]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_h = \mathcal{N} \wedge \mathcal{L}_f \subset \mathcal{L}_g \wedge \neg C_\star$
4	$[\Lambda_\star([h] [\lambda_g^\dagger], [0, 0], [\nabla_{\mathcal{I}_g^c} g][\nabla_{\mathcal{I}_h^c} h])]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_h = \mathcal{N} \wedge C_\star$
5	$[\lambda_t] + [g] [\lambda_h^\dagger]$	$\mathcal{L}_g = \mathcal{N} \wedge \mathcal{L}_h \subset \mathcal{N} \wedge \mathcal{L}_f = \mathcal{L}_h$
6	$[\lambda_t] + [g] [\lambda_{h,0}]$	$\mathcal{L}_g = \mathcal{N} \wedge \mathcal{L}_h \subset \mathcal{N} \wedge \mathcal{L}_f \subset \mathcal{L}_h \wedge \neg C_\star$
7	$[\Lambda_\star([0, 0], [g] [\lambda_h^\dagger], [\nabla_{\mathcal{I}_g^c} g][\nabla_{\mathcal{I}_h^c} h])]$	$\mathcal{L}_g = \mathcal{N} \wedge \mathcal{L}_h \subset \mathcal{N} \wedge C_\star$
8	$[\lambda_t] + [\Lambda_r([\Lambda_r([h] [\lambda_g^\dagger], [g] [\lambda_h^\dagger]), [0, 0])]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_h \subset \mathcal{N} \wedge \mathcal{L}_\cup = \mathcal{N} \wedge \mathcal{L}_f \subset \mathcal{L}_\cap$
9	$[\lambda_t] + [\Lambda_r([h] [\lambda_g^\dagger], [g] [\lambda_h^\dagger])]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_h \subset \mathcal{N} \wedge \mathcal{L}_\cup = \mathcal{N} \wedge \mathcal{L}_f = \mathcal{L}_\cap \wedge \neg C_\star$
10	$[\Lambda_\star([h] [\lambda_g^\dagger], [g] [\lambda_h^\dagger], [\nabla_{\mathcal{I}_g^c} g][\nabla_{\mathcal{I}_h^c} h])]$	$\mathcal{L}_g \subset \mathcal{N} \wedge \mathcal{L}_h \subset \mathcal{N} \wedge C_\star$
11	$[\lambda_t] + [h] [\lambda_g^\dagger] + [g] [\lambda_h^\dagger]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_f = \mathcal{L}_g = \mathcal{L}_h$
12	$[\lambda_t] + [h] [\lambda_g^\dagger] + [g] [\lambda_{h,0}]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_f = \mathcal{L}_g \subset \mathcal{L}_h$
13	$[\lambda_t] + [h] [\lambda_{g,0}] + [g] [\lambda_h^\dagger]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_f = \mathcal{L}_h \subset \mathcal{L}_g$
14	$[\lambda_t] + [\Lambda_r([h] [\lambda_g^\dagger] + [g] [\lambda_h^\dagger], [0, 0])]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_f \subset \mathcal{L}_g = \mathcal{L}_h$
15	$[\lambda_t] + [\Lambda_r([h] [\lambda_g^\dagger] + [g] [\lambda_{h,0}], [0, 0])]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_f \subset \mathcal{L}_g \subset \mathcal{L}_h$
16	$[\lambda_t] + [\Lambda_r([h] [\lambda_{g,0}] + [g] [\lambda_h^\dagger], [0, 0])]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_f \subset \mathcal{L}_h \subset \mathcal{L}_g$
17	$[\lambda_t] + [h] [\lambda_{g,0}] + [g] [\lambda_{h,0}]$	$\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_g \not\subseteq \mathcal{L}_h \wedge \mathcal{L}_h \not\subseteq \mathcal{L}_g$

Moreover, $\mathcal{I}_g \cup \mathcal{I}_h = \mathcal{N}$ implies $\mathcal{I}_g^c \cap \mathcal{I}_h^c = \emptyset$, which implies g and h depend on two different variables. Without loss of generality we assume g depends on x_1 , and h depends on x_2 , i.e., $\mathcal{I}_g^c = \{1\}$ and $\mathcal{I}_h^c = \{2\}$. As a further preparation note that $\mathcal{L}_f = \mathcal{I}_g \cap \mathcal{I}_h$, which holds according to Lem. 4.14, implies $\mathcal{L}_f^c = \mathcal{I}_g^c \cup \mathcal{I}_h^c$, which evaluates to $\mathcal{L}_f^c = \{1, 2\}$. Since $\mathcal{L}_f^c = \{1, 2\}$ and g only depends on x_1 (resp. h only depends on x_2), we have

$$\nabla_{\mathcal{L}_f^c} g(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} g(x) \\ \frac{\partial}{\partial x_2} g(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1} g(x) \\ 0 \end{pmatrix}, \quad \nabla_{\mathcal{L}_f^c}^2 g(x) = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} g(x) & 0 \\ 0 & 0 \end{pmatrix} \quad (4.23)$$

respectively

$$\nabla_{\mathcal{L}_f^c} h(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} h(x) \\ \frac{\partial}{\partial x_2} h(x) \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{\partial}{\partial x_2} h(x) \end{pmatrix}, \quad \nabla_{\mathcal{L}_f^c}^2 h(x) = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial^2}{\partial x_2^2} h(x) \end{pmatrix}. \quad (4.24)$$

Substituting (4.23) and (4.24) into (4.22) yields

$$\nabla_{\mathcal{L}_f^c}^2 f(x) = \begin{pmatrix} h(x) \frac{\partial^2}{\partial x_1^2} g(x) & \frac{\partial}{\partial x_1} g(x) \frac{\partial}{\partial x_2} h(x) \\ \frac{\partial}{\partial x_1} g(x) \frac{\partial}{\partial x_2} h(x) & g(x) \frac{\partial^2}{\partial x_2^2} h(x) \end{pmatrix},$$

where all entries are scalars. Now, consider the matrix set

$$\mathcal{H} = \{H \in \mathbb{R}^{2 \times 2} \mid H_{11} \in [h] [\lambda_g^\dagger], H_{22} \in [g] [\lambda_h^\dagger], H_{12} \in [\nabla_{\mathcal{I}_g^c} g][\nabla_{\mathcal{I}_h^c} h], H = H^T\}$$

and observe $\{\nabla_{\mathcal{L}_f^c}^2 f(x) \in \mathbb{R}^{2 \times 2} \mid x \in \mathcal{B}\} \subseteq \mathcal{H}$. To see this, note that $\{\nabla_{\mathcal{L}_f^c}^2 g(x) \in \mathbb{R} \mid x \in \mathcal{B}\} \subseteq [\lambda_g^\dagger]$ and $\{\nabla_{\mathcal{L}_f^c}^2 h(x) \in \mathbb{R} \mid x \in \mathcal{B}\} \subseteq [\lambda_h^\dagger]$, since the eigenvalue of a

matrix $M \in \mathbb{R}^{1 \times 1}$ is $\lambda = M_{1,1}$. According to Lem. A.1 stated in the appendix, eigenvalue bounds for the matrix set \mathcal{H} and consequently for $\nabla_{\mathcal{L}_f^c}^2 f(x)$ on \mathcal{B} read

$[\Lambda_\star([h][\lambda_g^\dagger], [g][\lambda_h^\dagger], [\nabla_{\mathcal{I}_g^c} g][\nabla_{\mathcal{I}_h^c} h])]$ as claimed in Tab. 4.4. \square

Lemma 4.16 shows that the bounds from Lem. 4.15 are as tight as or tighter than those from the original method proposed in [9]. Recall the bounds in [9] result in $[\lambda_f] = [\Lambda_t([\nabla g], [\nabla h])] + [h][\lambda_g] + [g][\lambda_h]$ for $f(x) = g(x)h(x)$ according to [9, Prop 3.2.(iv)]. We omit the proof of Lem. 4.16, since it is similar to its counterparts in Sect. 4.2.

Lemma 4.16 (improved bounds for products). *Assume Conds. 4.9 hold and let f , $[\lambda_f^\dagger]$, and \mathcal{L}_f be as in Lem. 4.11. Let $[\lambda_f]$, $[\lambda_g]$, and $[\lambda_h]$ be the eigenvalue bounds for the Hessians $\nabla^2 f(x)$, $\nabla^2 g(x)$, and $\nabla^2 h(x)$ on \mathcal{B} , calculated according to Eq. (4.5). Then,*

$$[\lambda_f] \subseteq [\Lambda_t([\nabla g], [\nabla h])] + [h][\lambda_g] + [g][\lambda_h].$$

4.5. Numerical computation of improved eigenvalue bounds. In this section, we combine the results from Sects. 4.1 through 4.4 in order to compute improved eigenvalue bounds using a codelist. Formally, this leads to the extended codelist in Prop. 4.17.

Proposition 4.17 (algorithm for the computation of eigenvalue bounds using sparsity). *Assume φ is twice continuously differentiable on \mathcal{U} and can be written as a codelist (3.1) with $t \in \mathbb{N}$ operations. Let $\mathcal{B} = [x_1] \times \dots \times [x_n] \subset \mathcal{U}$ be arbitrary. Then, for all $x \in \mathcal{B}$, we have $\varphi(x) \in [\varphi]$, $\nabla \varphi(x) \in [\nabla \varphi]$, and $[\lambda_{\min}(\nabla^2 \varphi(x)), \lambda_{\max}(\nabla^2 \varphi(x))] \subseteq [\lambda_\varphi]$, where $[\varphi]$, $[\nabla \varphi]$, and $[\lambda_\varphi]$ are calculated by the following algorithm.*

1. For $k = 1, \dots, n$, set $\mathcal{I}_k = \mathcal{N} \setminus \{k\}$, $\mathcal{L}_k = \mathcal{N}$, $[y_k] = [\underline{x}_k, \overline{x}_k]$, $[\nabla y_k] = [e_k, e_k]$, and $[\lambda_k^\dagger] = [0, 0]$.
2. For $k = n+1, \dots, n+t$, evaluate \mathcal{I}_k and \mathcal{L}_k according to the third and fourth column of Tab. 4.5, respectively. Calculate $[y_k]$ and $[\nabla y_k]$ according to the third and fourth column of Tab. 3.1, respectively. Compute $[\lambda_k^\dagger]$ depending on \mathcal{L}_i , \mathcal{L}_j , and \mathcal{L}_k according to the second column of Tab. A.1 in the appendix.
3. Compute $[\lambda_{n+t}]$ from $[\lambda_{n+t}^\dagger]$ and \mathcal{L}_{n+t} according to Eq. (4.5) and set $[\varphi] = [y_{n+t}]$, $[\nabla \varphi] = [\nabla y_{n+t}]$, and $[\lambda_\varphi] = [\lambda_{n+t}]$.

TABLE 4.5

Rules for the computation of the sets \mathcal{I}_k and \mathcal{L}_k in the k -th line of the codelist (3.1) for variables and binary operations (left) and compositions (right). Rules for y_k are repeated here for convenience.

op Φ_k	y_k	\mathcal{I}_k	\mathcal{L}_k	op Φ_k	y_k	\mathcal{I}_k	\mathcal{L}_k
var	x_k	$\mathcal{N} \setminus \{k\}$	\mathcal{N}	powNat	y_i^m	\mathcal{I}_i	\mathcal{I}_i
add	$y_i + y_j$	$\mathcal{I}_i \cap \mathcal{I}_j$	$\mathcal{L}_i \cap \mathcal{L}_j$	oneOver	$1/y_i$	\mathcal{I}_i	\mathcal{I}_i
mul	$y_i y_j$	$\mathcal{I}_i \cap \mathcal{I}_j$	$\mathcal{I}_i \cap \mathcal{I}_j$	sqrt	$\sqrt{y_i}$	\mathcal{I}_i	\mathcal{I}_i
				exp	$\exp(y_i)$	\mathcal{I}_i	\mathcal{I}_i
				ln	$\ln(y_i)$	\mathcal{I}_i	\mathcal{I}_i
				addC	$y_i + c$	\mathcal{I}_i	\mathcal{L}_i
				mulByC	$c y_i$	\mathcal{I}_i	\mathcal{L}_i

Proof. The claims $\varphi(x) \in [\varphi]$ and $\nabla \varphi(x) \in [\nabla \varphi]$ for all $x \in \mathcal{B}$ are covered by Thm. 3.1. It remains to prove that $[\lambda_{\min}(\nabla^2 \varphi(x)), \lambda_{\max}(\nabla^2 \varphi(x))] \subseteq [\lambda_\varphi]$ for all $x \in \mathcal{B}$. Since $y_k(x) = x_k$ for $k = 1, \dots, n$, the functions $y_k(x)$, $k \in \mathcal{N}$, are

independent of x_j for every $j \in \mathcal{I}_k = \mathcal{N} \setminus \{k\}$ and at most linearly dependent on x_j for every $j \in \mathcal{L}_k = \mathcal{N}$. Thus, the reduced Hessian reads $\nabla_{\mathcal{L}_k}^2 y_k(x) = \nabla_{\emptyset}^2 y_k(x) = 0_{0,0}$ and $[\lambda_k^\dagger] = [0, 0]$ for every $k \in \mathcal{N}$. Now assume eigenvalue bounds $[\lambda_1^\dagger], \dots, [\lambda_l^\dagger]$ for the reduced Hessians $\nabla_{\mathcal{L}_1}^2 y_1(x), \dots, \nabla_{\mathcal{L}_l}^2 y_l(x)$ and index sets $\mathcal{I}_1, \dots, \mathcal{I}_l$ and $\mathcal{L}_1, \dots, \mathcal{L}_l$ have been calculated for some $l \in \mathbb{N}_{n, n+t-1}$, and let $k = l + 1$. Since $\Phi_k(y_1, \dots, y_{k-1})$ is one of the unary or binary functions listed in Tab. 3.1 (and therefore Tabs. 4.5 and A.1), it depends on either one (say y_i) or two (say y_i and y_j) of the intermediate variables y_1, \dots, y_{k-1} . The remainder of the proof must be carried out for each type of operation Φ_k separately. We state the proof for one of the `mul` cases and claim the remaining cases can be shown accordingly. Let $g(x) = y_i(x)$, $h(x) = y_j(x)$, and $f(x) = y_k(x)$, which implies $f(x) = g(x)h(x)$, since the operation in the k -th line is of type `mul`. In order to compute eigenvalue bounds $[\lambda_f^\dagger]$ for the reduced Hessian $\nabla_{\mathcal{L}_f}^2 f(x)$, we first evaluate the index sets \mathcal{I}_f and \mathcal{L}_f . According to Lem. 4.14, we obtain

$$\mathcal{I}_f = \mathcal{I}_g \cap \mathcal{I}_h = \mathcal{I}_i \cap \mathcal{I}_j \quad \text{and} \quad \mathcal{L}_f = \mathcal{I}_g \cap \mathcal{I}_h = \mathcal{I}_i \cap \mathcal{I}_j,$$

where we used $\mathcal{I}_g = \mathcal{I}_i$ and $\mathcal{I}_h = \mathcal{I}_j$, which hold by construction. Assuming we have $\mathcal{L}_g \cup \mathcal{L}_h \subset \mathcal{N}$ and $\mathcal{L}_f = \mathcal{L}_g \subset \mathcal{L}_h$, applying Lem. 4.15 (specifically, rule 12 in Tab. 4.4) results in

$$\begin{aligned} [\lambda_f^\dagger] &= [\Lambda_t([\nabla_{\mathcal{L}_f} g], [\nabla_{\mathcal{L}_f} h])] + [h] [\lambda_g^\dagger] + [g] [\min\{\Delta_h^\dagger, 0\}, \max\{\Delta_h^\dagger, 0\}], \\ &= [\Lambda_t([\nabla_{\mathcal{L}_f} y_i], [\nabla_{\mathcal{L}_f} y_j])] + [y_j] [\lambda_i^\dagger] + [y_i] [\min\{\Delta_j^\dagger, 0\}, \max\{\Delta_j^\dagger, 0\}], \end{aligned} \quad (4.25)$$

where the second equation results from substituting the codelist notation $[g] = [y_i]$, $[h] = [y_j]$, $[\nabla g] = [\nabla y_i]$, $[\nabla h] = [\nabla y_j]$, $[\lambda_g^\dagger] = [\lambda_i^\dagger]$, and $[\lambda_h^\dagger] = [\lambda_j^\dagger]$. Finally, since $\mathcal{L}_k = \mathcal{L}_f$ and consequently $\nabla_{\mathcal{L}_k}^2 y_k(x) = \nabla_{\mathcal{L}_f}^2 f(x)$, the eigenvalues of $\nabla_{\mathcal{L}_k}^2 y_k(x)$ are confined to $[\lambda_k^\dagger] = [\lambda_f^\dagger]$ for all $x \in \mathcal{B}$. Since the second equation in (4.25) is equal to the rule in Tab. A.1 for the case `mul` and $\mathcal{L}_i \cup \mathcal{L}_j \subset \mathcal{N}$ and $\mathcal{L}_k = \mathcal{L}_i \subset \mathcal{L}_j$, this proves the claim for the selected case. \square

Proposition 4.17 is illustrated with two examples. First, we revisit the motivating Exmp. 1. Recall that we evaluated the conservative eigenvalue bounds $[\lambda_\varphi] = [0, 4]$ using the original method from [9].

Example 2 (improved method applied to $\varphi(x) = x_1^2 + x_2^2$ from Exmp. 1). *Consider the function φ from Exmp. 1 again. Proposition 4.17 results in the following extended codelist. Note that we do not list the expressions for $[y_k]$ and $[\nabla y_k]$ in (4.26) since they are identical to the corresponding expressions in (4.1). Further note that \mathcal{I}_k and \mathcal{L}_k are independent of \mathcal{B} .*

k	\mathcal{I}_k	\mathcal{L}_k	$[\lambda_k^\dagger]$
1	$\mathcal{N} \setminus \{1\} = \{2\}$	$\mathcal{N} = \{1, 2\}$	$[0, 0]$
2	$\mathcal{N} \setminus \{2\} = \{1\}$	$\mathcal{N} = \{1, 2\}$	$[0, 0]$
3	$\mathcal{I}_1 = \{2\}$	$\mathcal{I}_1 = \{2\}$	$2[\Lambda_s([\nabla_{\mathcal{L}_3} y_1])] = 2[\Lambda_s([\nabla_{\{1\}} y_1)]]$
4	$\mathcal{I}_2 = \{1\}$	$\mathcal{I}_2 = \{1\}$	$2[\Lambda_s([\nabla_{\mathcal{L}_4} y_2])] = 2[\Lambda_s([\nabla_{\{2\}} y_2)]]$
5	$\mathcal{I}_3 \cap \mathcal{I}_4 = \emptyset$	$\mathcal{L}_3 \cap \mathcal{L}_4 = \emptyset$	$[\min\{\Delta_3^\dagger, \Delta_4^\dagger\}, \max\{\Delta_3^\dagger, \Delta_4^\dagger\}]$
			$[\lambda_\varphi] = [\lambda_5^\dagger]$

The expressions for $[\lambda_k^\dagger]$ in lines 3 and 4 of the extended codelist in (4.26) refer to the first rule associated with the `powNat`-operation in Tab. A.1 since $\mathcal{L}_1 = \mathcal{N}$ and

$\mathcal{L}_2 = \mathcal{N}$, respectively. Since $\mathcal{L}_3 = \{2\} \subset \mathcal{N}$, $\mathcal{L}_4 = \{1\} \subset \mathcal{N}$, and $\mathcal{L}_3 \cup \mathcal{L}_4 = \mathcal{N}$, we obtain the bounds $[\lambda_5^\dagger] = [\min\{\underline{\lambda}_3^\dagger, \underline{\lambda}_4^\dagger\}, \max\{\overline{\lambda}_3^\dagger, \overline{\lambda}_4^\dagger\}]$ according to the last rule for the **add**-operation in Tab. A.1. Finally, since $\mathcal{L}_5 = \emptyset$, we have $[\lambda_\varphi] = [\lambda_5] = [\lambda_5^\dagger]$ according to Eq. (4.5).

Evaluating the extended codelist (4.26) for the hyperrectangle $\mathcal{B} = [0, 1] \times [0, 1]$ (as in Exmp. 1) by computing $[y_k]$ and $[\nabla y_k]$ according to (4.1) and $[\lambda_k]$ according to (4.26) yields $[\lambda_1^\dagger] = [\lambda_2^\dagger] = [0, 0]$ and $[\lambda_3^\dagger] = [\lambda_4^\dagger] = [\lambda_5^\dagger] = [2, 2]$, where we used $[\Lambda_s([\nabla_{\{1\}}y_1])] = [\Lambda_s([1, 1])] = [1, 1]$ and $[\Lambda_s([\nabla_{\{2\}}y_2])] = [\Lambda_s([1, 1])] = [1, 1]$ (see Eq. (4.2) for numerical results on $[y_k]$ and $[\nabla y_k]$). Thus, using the improved method, we obtain the tight eigenvalue bounds $[\lambda_\varphi] = [\lambda_5^\dagger] = [\lambda_\varphi^*] = [2, 2]$.

We analyze another example to demonstrate that the new method results in considerable improvements for all functions that involve multiplications. In fact, we know from [9, Rem. 4.3] that $0 \in [\lambda_\varphi]$ for the original method if the **mul**-operation is required in the codelist of any φ with $n \geq 2$. This is a severe drawback of the original method, since it implies that any convex (resp. concave) function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ involving **mul**-operations will never be identified to be convex (resp. concave) using the method from [9]. The following example shows that this restriction does not apply for the improved method.

Example 3 (comparison of [9] and improved method for $\varphi(x) = x_1^2 + x_2 \exp(x_2)$). Consider the function $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ with $\varphi(x) = x_1^2 + x_2 \exp(x_2)$ on a $\mathcal{B} \subset \mathbb{R}^2$. Theorem 3.1 (i.e., the original method from [9]) results in the following extended codelist, where the expressions for y_k are only listed for illustration of the codelist (3.1) of φ . We skip the first three lines, since they are identical to those in (4.1).

k	y_k	$[y_k]$	$[\nabla y_k]$	$[\lambda_k]$
4	$\exp(y_2)$	$[\exp([y_2])]$	$[y_4] [\nabla y_2]$	$[y_4] ([\Lambda_s([\nabla y_2])) + [\lambda_2]$
5	$y_2 y_4$	$[y_2] [y_4]$	$[y_4] [\nabla y_2] + [y_2] [\nabla y_4]$	$[y_4] [\lambda_2] + [y_2] [\lambda_4] + [\Lambda_t([\nabla y_2], [\nabla y_4])]$ (4.27)
6	$y_3 + y_5$	$[y_3] + [y_5]$	$[\nabla y_3] + [\nabla y_5]$	$[\lambda_3] + [\lambda_5]$
	$\varphi = y_6$	$[\varphi] = [y_6]$	$[\nabla \varphi] = [\nabla y_6]$	$[\lambda_\varphi] = [\lambda_6]$

Evaluating this codelist for $\mathcal{B} = [0, 1] \times [0, 1]$ yields

$$[\lambda_\varphi] = [\lambda_6] = [-\exp(1) + 1, 3 \exp(1) + 2] \approx [-1.7183, 10.1548].$$

Proposition 4.17 (i.e., the improved method) results in the following extended codelist. The first three lines are identical to those in (4.26) in this case. Further note that the expressions for $[y_k]$ and $[\nabla y_k]$ can be found in Eq. (4.1) (lines 1-3) and Eq. (4.27) (lines 4-6).

k	\mathcal{I}_k	\mathcal{L}_k	$[\lambda_k^\dagger]$
4	$\mathcal{I}_2 = \{1\}$	$\mathcal{I}_2 = \{1\}$	$[y_4] [\Lambda_s([\nabla_{\mathcal{L}_4^c} y_2])] = [y_4] [\Lambda_s([\nabla_{\{2\}} y_2])]$
5	$\mathcal{I}_2 \cap \mathcal{I}_4 = \{1\}$	$\mathcal{I}_2 \cap \mathcal{I}_4 = \{1\}$	$[\Lambda_t([\nabla_{\mathcal{L}_5^c} y_2], [\nabla_{\mathcal{L}_5^c} y_4])] + [y_4] [\lambda_2^\dagger] + [y_2] [\lambda_4^\dagger]$ $= [\Lambda_t([\nabla_{\{2\}} y_2], [\nabla_{\{2\}} y_4])] + [y_4] [\lambda_2^\dagger] + [y_2] [\lambda_4^\dagger]$ (4.28)
6	$\mathcal{I}_3 \cap \mathcal{I}_5 = \emptyset$	$\mathcal{L}_3 \cap \mathcal{L}_5 = \emptyset$	$[\min\{\underline{\lambda}_3^\dagger, \underline{\lambda}_5^\dagger\}, \max\{\overline{\lambda}_3^\dagger, \overline{\lambda}_5^\dagger\}]$
			$[\lambda_\varphi] = [\lambda_6^\dagger]$

Evaluating (4.28) for $\mathcal{B} = [0, 1] \times [0, 1]$ yields

$$[\lambda_\varphi] = [\lambda_6^\dagger] = [2, 3 \exp(1)] \approx [2, 8.1548],$$

Just as in Exmp. 3, the improved method results in tight spectral bounds while the original method from [9] provides loose outer approximations. In particular, $0 \in$

$[-1.7183, 10.1548]$ for the original method as predicted by [9, Rem. 4.3] but $0 \notin [2, 8.1548]$ for the improved method presented here. Convexity of φ on \mathcal{B} can therefore be established with the improved but not with the original method.

More generally, the improved method results in eigenvalue bounds that are always as tight as, or tighter than, the original method from [9], as stated in the following proposition.

Proposition 4.18 (accuracy of the improved method). *Assume φ is twice continuously differentiable on \mathcal{U} and can be written as a codelist (3.1). Let $\mathcal{B} = [x_1] \times \cdots \times [x_n] \subset \mathcal{U}$ be arbitrary and denote the eigenvalue bounds for $\nabla^2\varphi(x)$ on \mathcal{B} computed according to Thm. 4.17 and Prop. 4.17 by $[\lambda_\varphi^{(\text{Thm. 3.1})}]$ and $[\lambda_\varphi^{(\text{Prop. 4.17})}]$, respectively. Then,*

$$[\lambda_\varphi^{(\text{Prop. 4.17})}] \subseteq [\lambda_\varphi^{(\text{Thm. 3.1})}].$$

Proof. The proof immediately follows from Lems. 4.8, 4.11, and 4.15. \square

In [9, Prop. 4.4] it was shown that the numerical complexity for evaluating the extended codelist resulting from Thm. 3.1, is of order $\mathcal{O}(n)N(\varphi)$, where $N(\varphi)$ denotes the number of operations needed to evaluate φ at a point in its domain. It is remarkable that this order of complexity can be maintained for the improved method. This is summarized in the following proposition.

Proposition 4.19 (numerical complexity of the improved method). *Assume φ is twice continuously differentiable on \mathcal{U} and can be written as a codelist (3.1) with $t = N(\varphi)$ operations. Let $N([\varphi], [\nabla\varphi], [\lambda_\varphi])$ denote the number of operations that are necessary to calculate the bounds $[\varphi] \subset \mathbb{R}$, $[\nabla\varphi] \subset \mathbb{R}^n$, and $[\lambda_\varphi] \subset \mathbb{R}$ for a given hyperrectangle $\mathcal{B} \subset \mathcal{U}$ using the extended codelist from Prop. 4.17. Then,*

$$N([\varphi], [\nabla\varphi], [\lambda_\varphi]) = \mathcal{O}(n)N(\varphi).$$

Since the proof of Prop. 4.19 is very similar to that of [9, Prop. 4.4], we only sketch it. The extended codelist that results from Prop. 4.17 involves the index sets \mathcal{I}_k and \mathcal{L}_k , which were not required in the original method. These index sets do not depend on the particular hyperrectangle \mathcal{B} as pointed out in Exmp. 2, but they are uniquely determined by the function φ itself. Consequently, all index sets need to be determined only once. This step can be carried out at the time of construction of the extended codelist. In particular, it need not be repeated at the time of evaluating the codelist for a particular \mathcal{B} . Once \mathcal{I}_k and \mathcal{L}_k have been determined, each line of the extended codelist is specified by the rules in Tab. A.1 (and Tab. 3.1). It is easy to show that the evaluation of every expression in the second column of Tab. A.1 requires at most $\mathcal{O}(n)$ basic operations (like additions, multiplications, or comparisons of two real numbers; see [9, Sect. 4.1] for further details). Thus, under the assumption that $[\varphi]$ and $[\nabla\varphi]$ are known, we need $\mathcal{O}(n)N(\varphi)$ basic operations for the computation of $[\lambda_\varphi]$. Since the calculation of $[\varphi]$ and $[\nabla\varphi]$ require $\mathcal{O}(1)$ and $\mathcal{O}(n)$ basic operations according to standard results in AD and IA (see, e.g., [4, 7]), we obtain $N([\varphi], [\nabla\varphi], [\lambda_\varphi]) = \mathcal{O}(1)N(\varphi) + \mathcal{O}(n)N(\varphi) + \mathcal{O}(n)N(\varphi) = \mathcal{O}(n)N(\varphi)$.

5. Numerical experiments for a large number of examples. In this section, we analyze 1522 numerical examples taken from the COCONUT collection of optimization problems [14]. We consider all COCONUT problems with $1 < n \leq 10$ variables and extract those cost and constraint functions that can be decomposed into

the operations listed in Tabs. 3.1 and A.1. For each function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, we consider 100 (randomly generated) hyperrectangles $\mathcal{B} \subset \mathcal{D}$ in the domain \mathcal{D} of φ specified in the respective COCONUT problem. For ease of comparison, the set of examples as well as the associated hyperrectangles are identical to the examples considered in [13].

For each of the resulting $1522 \cdot 100$ sample problems, we solve problem (1.1) using the improved algorithm (A^\dagger for short) in Prop. 4.17. We compare the resulting eigenvalue bounds with those obtained from two established methods using interval Hessians (see problem (1.3)) and either Gershgorin’s circle criterion (G for short) or Hertz and Rohn’s method (H for short) for the computation of spectral bounds of interval matrices (see problem (1.4)). We choose G and H as reference procedures due to the favorable computational complexity of G and since H provides tight eigenvalue bounds for problem (1.4) (cf. Sect. 1). We refer to the original papers [2, 5, 6, 12] or the summaries in [9, 13] for a detailed description of methods G and H.

TABLE 5.1

Classes used to aggregate results in Tab. 5.2. Symbols $[\lambda_{A^\dagger}] = [\underline{\lambda}_{A^\dagger}, \overline{\lambda}_{A^\dagger}]$, $[\lambda_G] = [\underline{\lambda}_G, \overline{\lambda}_G]$, and $[\lambda_H] = [\underline{\lambda}_H, \overline{\lambda}_H]$ denote eigenvalue bounds calculated by the improved algorithm A^\dagger (see Prop. 4.17), Gershgorin’s circle criterion and Hertz and Rohn’s method, respectively.

class	verbal definition	formal definition	
		lower bound $\underline{\lambda}$	upper bound $\overline{\lambda}$
1	A^\dagger worse than G (and H)	$\underline{\lambda}_{A^\dagger} < \underline{\lambda}_G \leq \underline{\lambda}_H$	$\overline{\lambda}_H \leq \overline{\lambda}_G < \overline{\lambda}_{A^\dagger}$
2	A^\dagger equal to G but worse than H	$\underline{\lambda}_G \approx \underline{\lambda}_{A^\dagger} < \underline{\lambda}_H$	$\overline{\lambda}_H < \overline{\lambda}_{A^\dagger} \approx \overline{\lambda}_G$
3	A^\dagger better than G but worse than H	$\underline{\lambda}_G < \underline{\lambda}_{A^\dagger} < \underline{\lambda}_H$	$\overline{\lambda}_H < \overline{\lambda}_{A^\dagger} < \overline{\lambda}_G$
4	A^\dagger equal to H (and equal to or better than G)	$\underline{\lambda}_G \leq \underline{\lambda}_H \approx \underline{\lambda}_{A^\dagger}$	$\overline{\lambda}_{A^\dagger} \approx \overline{\lambda}_H \leq \overline{\lambda}_G$
5	A^\dagger better than H (and G)	$\underline{\lambda}_G \leq \underline{\lambda}_H < \underline{\lambda}_{A^\dagger}$	$\overline{\lambda}_{A^\dagger} < \overline{\lambda}_H \leq \overline{\lambda}_G$

For each sample problem, we analyze whether A^\dagger performs better than, equally good as, or worse than G and H. We independently compare the lower and upper eigenvalue bounds of the particular methods and categorize the results according to the five classes in Tab. 5.1. Note that G never performs better than H (since H provides tight bounds for (1.4)). Consequently, the relations $\underline{\lambda}_G \leq \underline{\lambda}_H$ and $\overline{\lambda}_H \leq \overline{\lambda}_G$ always hold. Hence, the list of classes in Tab. 5.1 is complete in the sense that every example can be uniquely classified into one of the five classes. It remains to comment on the precise meaning of $a > b$ and $a \approx b$ as used for the classification in Tab. 5.1. To this end, we introduce the function

$$\text{dev}(a, b) = \frac{a - b}{1 + 0.5|a + b|}$$

which evaluates a weighted difference of $a, b \in \mathbb{R}$. Based on $\text{dev}(a, b)$, we define

$$a > b \iff \text{dev}(a, b) > \epsilon \quad \text{and} \quad a \approx b \iff |\text{dev}(a, b)| \leq \epsilon, \quad (5.1)$$

where $\epsilon \in \mathbb{R}_+$ represents an error bound. Note that $|\text{dev}(a, b)|$ is approximately equal to the relative error for two large but almost equal numbers $a, b \in \mathbb{R}$ and almost equal to the absolute error for two small but almost equal numbers $a, b \in \mathbb{R}$. This behavior is useful since the absolute values of the computed eigenvalue bounds range across multiple magnitudes.

We summarize numerical results for the analyzed examples in Tab. 5.2 (with $\epsilon = 10^{-6}$). We list the percentage of samples that fall into the classes 1 to 5 from Tab. 5.1 separated by dimension n of the underlying example. In order to compare

TABLE 5.2

Numerical results for 1522 · 100 sample problems. The classification of the results is carried out according to Tab. 5.1 and Eq. (5.1) with $\epsilon = 10^{-6}$. For each class, the left respectively right column refers to eigenvalue bounds obtained by the original algorithm A from [9] and the improved variant A^\dagger taking sparsity into account.

examples		1		2		3		4		5	
n	#	A	A^\dagger	A	A^\dagger	A	A^\dagger	A	A^\dagger	A	A^\dagger
2	62	56.37	29.06	1.07	1.08	12.24	16.06	18.89	41.27	11.43	12.54
3	1078	2.98	0.95	77.53	0.31	0.85	1.01	17.70	96.72	0.95	1.01
4	67	60.75	35.28	4.84	4.63	8.45	16.76	15.04	32.14	10.92	11.19
5	88	56.80	35.34	3.32	0.21	10.16	14.90	14.91	34.47	14.81	15.09
6	95	35.05	25.65	5.31	4.13	31.91	34.25	13.87	21.38	13.86	14.58
7	27	65.80	27.07	11.93	8.50	0.02	19.44	22.26	44.98	0.00	0.00
8	15	94.23	63.83	3.20	4.23	1.27	29.80	1.30	1.53	0.00	0.60
9	24	57.27	25.02	8.29	0.02	18.71	34.94	4.25	22.58	11.48	17.44
10	66	51.17	12.24	1.00	0.22	22.12	34.11	15.86	41.54	9.86	11.89
all	1522	17.77	9.09	56.11	0.94	5.32	7.78	16.86	77.90	3.95	4.30

the improved algorithm in Prop. 4.17 to the original method from [9] (see Thm. 3.1), we also list the classification results using the original algorithm (A for short). The numerical results confirm that the consideration of sparsity significantly improves the tightness of the computed eigenvalue bounds. To see this, note that for each dimension n , the percentages in class 1 (where the established approaches outperform the direct computation of eigenvalue bounds) decrease while the percentages in classes 4 and 5 (where the direct computation of eigenvalue bounds performs as good as or better than Hertz and Rohn’s method) increases using the improved algorithm A^\dagger instead of the original A. In particular, it is remarkable that the improved algorithm A^\dagger results in worse eigenvalue bounds than G in only 9.09% of all cases in contrast to 17.77% for the original method A. Moreover, A^\dagger provides equally good or better eigenvalue bounds than H in 82.20% = 77.90% + 4.30% of all cases while the corresponding percentage only reads 20.83% = 16.86% + 3.95% for A.

Another observation is that the ratios in the particular classes seem to be independent of the dimension n (i.e., there is no trend). This is important since the numerical complexities of the established approaches G and H vary between $\mathcal{O}(n)N(\varphi) + \mathcal{O}(n^2)$ and $\mathcal{O}(n^2)N(\varphi) + \mathcal{O}(2^n n^3)$ operations (see Sect. 1 and the benchmark in [13]), while the direct eigenvalue bound computation requires $\mathcal{O}(n)N(\varphi)$. Thus, methods A and A^\dagger become numerically very attractive for high dimensions n .

TABLE 5.3

The last line of Tab. 5.2 for different choices of the error bound ϵ in (5.1).

ϵ	1		2		3		4		5	
	A	A^\dagger	A	A^\dagger	A	A^\dagger	A	A^\dagger	A	A^\dagger
10^{-5}	17.16	8.78	54.11	0.91	4.29	6.57	20.78	79.74	3.66	4.01
10^{-6}	17.77	9.09	56.11	0.94	5.32	7.78	16.86	77.90	3.95	4.30
10^{-7}	18.04	9.28	57.62	0.66	5.61	8.24	14.67	77.41	4.06	4.41

According to (5.1), the classification in Tab. 5.2 depends on the choice of the error bound ϵ . We repeated all calculations for various choices of ϵ and present the results reported in the last line of Tab. 5.2 for $\epsilon = 10^{-5}$ and $\epsilon = 10^{-7}$ in Tab. 5.3.

As expected, the ratios in classes 1 and 5 increase for decreasing ϵ , since we detect $\underline{\lambda}_{A^\dagger} < \underline{\lambda}_G$ (as well as $\bar{\lambda}_G < \bar{\lambda}_{A^\dagger}$, $\underline{\lambda}_H < \underline{\lambda}_{A^\dagger}$, and $\bar{\lambda}_{A^\dagger} < \bar{\lambda}_H$) for a larger number of examples (cf. (5.1)). However, beside this observation, the results are robust w.r.t. the value of ϵ .

6. Conclusion. We significantly improved a method recently introduced in [9] for the efficient computation of spectral bounds for Hessian matrices of twice continuously differentiable functions on hyperrectangles. The improvements build on the identification and utilization of sparsity that naturally arises in the first lines of every codelist for a function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$.

The improved method was applied to a set of 1522 examples previously analyzed in [13]. The numerical results show that the consideration of sparsity results in significantly tighter eigenvalue bounds. In fact, the improved method provided equally good or better eigenvalue bounds than Hertz and Rohn’s method in 82.20 % of the examples while the corresponding percentage only reads 20.83 % for the original procedure.

In addition to illustrating the practical usefulness of the proposed improvements, we provided an important theoretic result. In fact, it is well-known that the original method from [9] results in spectral bounds with $0 \in [\lambda_\varphi]$ for any function that involves the multiplications of two or more variables (see [9, Rem. 4.3]). Consequently, convex functions that involve such a multiplication cannot be detected to be convex with the original method. We showed that this restrictions does not apply for the improved method.

Acknowledgements. Funding by Deutsche Forschungsgemeinschaft grant MO-1086/9 is gratefully acknowledged.

REFERENCES

- [1] C. S. Adjiman, I. P. Androulakis, and C. A. Floudas. A global optimization method, α BB, for general twice-differentiable constrained NLPs – II. Implementation and computational results. *Computers and Chemical Engineering*, 22(9):1159–1179, 1998.
- [2] C. S. Adjiman, S. Dallwig, C. A. Floudas, and A. Neumaier. A global optimization method, α BB, for general twice-differentiable constrained NLPs – I. Theoretical advances. *Computers and Chemical Engineering*, 22(9):1137–1158, 1998.
- [3] I. P. Androulakis, C. D. Maranas, and C. A. Floudas. α BB: A global optimization method for general constrained nonconvex problems. *Journal of Global Optimization*, 7(4):337–363, 1995.
- [4] H. Fischer. Automatisches Differenzieren. In J. Herzberger, editor, *Wissenschaftliches Rechnen: Eine Einführung in das Scientific Computing*, pages 53–103. Akademie Verlag Berlin, 1995.
- [5] S. Gerschgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk SSSR, Ser. fizmat.*, 6:749–754, 1931.
- [6] D. Hertz. The extreme eigenvalues and stability of real symmetric interval matrices. *IEEE Transactions on Automatic Control*, 37(4):532–535, 1992.
- [7] R. B. Kearfott. *Rigorous Global Search: Continuous Problems*. Kluwer Academic Publishers, 1996.
- [8] M. Mönnigmann. Positive invariance tests with efficient Hessian matrix eigenvalue bounds. In *Proceedings of the 17th IFAC World Congress*, 2008.
- [9] M. Mönnigmann. Fast calculation of spectral bounds for Hessian matrices on hyperrectangles. *SIAM Journal on Scientific Computing*, 2011.
- [10] A. Neumaier. *Interval Methods for Systems of Equations*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1st edition, 2008.
- [11] L. B. Rall. *Automatic Differentiation: Techniques and Applications*, volume 120 of *Lecture Notes in Computer Science*. Springer Berlin/Heidelberg, 1981.
- [12] J. Rohn. Positive definiteness and stability of interval matrices. *SIAM Journal on Matrix Analysis and Applications*, 15(1):175–184, 1994.
- [13] M. Schulze Darup, M. Kastnian, S. Mross, and M. Mönnigmann. Efficient computation of

spectral bounds for Hessian matrices on hyperrectangles for global optimization. *Journal of Global Optimization*, 58:631–652, 2014.

- [14] O. Shcherbina, A. Neumaier, D. Sam-Haroud, X.-H. Vu, and T.-V. Nguyen. Benchmarking global optimization and constraint satisfaction codes. In C. Blik, C. Jermann, and A. Neumaier, editors, *Global Optimization and Constraint Satisfaction*, pages 211–222. Springer Verlag, 2003.

Appendix A. Supplementary results.

Lemma A.1. *Let $[a], [b], [c] \subset \mathbb{R}$ be real intervals, let $[\Lambda_\star([a], [b], [c])]$ be defined as in (4.21), and consider the matrix set*

$$\mathcal{H} = \{H \in \mathbb{R}^{2 \times 2} \mid H_{11} \in [a], H_{22} \in [b], H_{12} \in [c], H = H^T\}.$$

Then,

$$\underline{\Lambda}_\star([a], [b], [c]) = \min_{H \in \mathcal{H}} \lambda_{\min}(H) \quad \text{and} \quad \max_{H \in \mathcal{H}} \lambda_{\max}(H) = \overline{\Lambda}_\star([a], [b], [c]). \quad (\text{A.1})$$

Proof. The eigenvalue bounds of a symmetric matrix $H = \begin{pmatrix} a & c \\ c & b \end{pmatrix} \in \mathcal{H}$ read

$$\lambda_{\min}(H) = \frac{1}{2} \left(a + b - \sqrt{(a - b)^2 + 4c^2} \right), \quad \lambda_{\max}(H) = \frac{1}{2} \left(a + b + \sqrt{(a - b)^2 + 4c^2} \right).$$

We therefore have to show that

$$\underline{a} + \underline{b} - \sqrt{(\underline{a} - \underline{b})^2 + d} = \min_{a \in [a], b \in [b], c \in [c]} a + b - \sqrt{(a - b)^2 + 4c^2} \quad \text{and} \quad (\text{A.2})$$

$$\overline{a} + \overline{b} + \sqrt{(\overline{a} - \overline{b})^2 + d} = \max_{a \in [a], b \in [b], c \in [c]} a + b + \sqrt{(a - b)^2 + 4c^2}, \quad (\text{A.3})$$

where the l.h.s. results from (4.21) and where $d = 4 \max\{\underline{c}^2, \overline{c}^2\}$. We show that (A.2) holds and claim (A.3) can be proven analogously. First note that the r.h.s. in (A.2) can be simplified to

$$\min_{a \in [a], b \in [b], c \in [c]} a + b - \sqrt{(a - b)^2 + 4c^2} = \min_{a \in [a], b \in [b]} a + b - \sqrt{(a - b)^2 + d} \quad (\text{A.4})$$

since c only occurs in the radicand. Consider the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(a, b) = a + b - \sqrt{(a - b)^2 + d}$, which occurs on the r.h.s. of (A.4) and note that $f(a, b)$ is concave (since $g(a, b) = \sqrt{(a - b)^2 + d}$ is convex). Since the hyperrectangle $\mathcal{B} = [a] \times [b]$ is convex, the minimum on the r.h.s. of (A.4) is attained at one of the vertices of \mathcal{B} . Among the candidate tuples $(\underline{a}, \underline{b})$, $(\underline{a}, \overline{b})$, $(\overline{a}, \underline{b})$, and $(\overline{a}, \overline{b})$, it is easy to show that $(\underline{a}, \underline{b})$ results in the smallest function value, i.e., $f(\underline{a}, \underline{b}) \leq \min\{f(\underline{a}, \overline{b}), f(\overline{a}, \underline{b}), f(\overline{a}, \overline{b})\}$. Thus, (A.2) holds. \square

TABLE A.1

Rules for the calculation of the eigenvalue bounds $[\lambda_k^\dagger]$ in the k -th line of the codelist (3.1) according to Prop. 4.17. The intervals $[\lambda_t]$, $[\lambda_{i,0}]$ and $[\lambda_{j,0}]$ are shorthand notations for $[\lambda_t] = [\Lambda_t([\nabla_{\mathcal{L}_k^c} y_i], [\nabla_{\mathcal{L}_k^c} y_j])]$, $[\lambda_{i,0}] = [\min\{\underline{\lambda}_i^\dagger, 0\}, \max\{\overline{\lambda}_i^\dagger, 0\}]$, and $[\lambda_{j,0}] = [\min\{\underline{\lambda}_j^\dagger, 0\}, \max\{\overline{\lambda}_j^\dagger, 0\}]$. The index sets \mathcal{L}_\cup and \mathcal{L}_\cap are shorthand notations for $\mathcal{L}_\cup = \mathcal{L}_i \cup \mathcal{L}_j$ and $\mathcal{L}_\cap = \mathcal{L}_i \cap \mathcal{L}_j$. Condition C_\star reads $(\mathcal{I}_i \cup \mathcal{I}_j = \mathcal{N}) \wedge (|\mathcal{I}_i| = n-1) \wedge (|\mathcal{I}_j| = n-1)$.

op Φ_k	$[\lambda_k^\dagger]$	condition
add	$[0, 0]$ $[\lambda_i^\dagger]$ $[\lambda_j^\dagger]$ $[\Lambda_r([\lambda_i^\dagger], [\lambda_j^\dagger])]$ $[\lambda_i^\dagger] + [\lambda_j^\dagger]$ $[\lambda_i^\dagger] + [\lambda_{j,0}]$ $[\lambda_{i,0}] + [\lambda_j^\dagger]$ $[\lambda_{i,0}] + [\lambda_{j,0}]$	$\mathcal{L}_i = \mathcal{N} \wedge \mathcal{L}_j = \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_j = \mathcal{N}$ $\mathcal{L}_i = \mathcal{N} \wedge \mathcal{L}_j \subset \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_j \subset \mathcal{N} \wedge \mathcal{L}_\cup = \mathcal{N}$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_i = \mathcal{L}_j$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_i \subset \mathcal{L}_j$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_j \subset \mathcal{L}_i$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_i \not\subseteq \mathcal{L}_j \wedge \mathcal{L}_j \not\subseteq \mathcal{L}_i$
mul	$[\lambda_t]$ $[\lambda_t] + [y_j] [\lambda_i^\dagger]$ $[\lambda_t] + [y_j] [\lambda_{i,0}]$ $[\Lambda_\star([y_j] [\lambda_i^\dagger], [0, 0], [\nabla_{\mathcal{I}_i^c} y_i][\nabla_{\mathcal{I}_j^c} y_j])]$ $[\lambda_t] + [y_i] [\lambda_j^\dagger]$ $[\lambda_t] + [y_i] [\lambda_{j,0}]$ $[\Lambda_\star([0, 0], [y_i] [\lambda_j^\dagger], [\nabla_{\mathcal{I}_i^c} y_i][\nabla_{\mathcal{I}_j^c} y_j])]$ $[\lambda_t] + [\Lambda_r([\Lambda_r([y_j] [\lambda_i^\dagger], [y_i] [\lambda_j^\dagger]), [0, 0])]$ $[\lambda_t] + [\Lambda_r([y_j] [\lambda_i^\dagger], [y_i] [\lambda_j^\dagger])]$ $[\Lambda_\star([y_j] [\lambda_i^\dagger], [y_i] [\lambda_j^\dagger], [\nabla_{\mathcal{I}_i^c} y_i][\nabla_{\mathcal{I}_j^c} y_j])]$ $[\lambda_t] + [y_j] [\lambda_i^\dagger] + [y_i] [\lambda_j^\dagger]$ $[\lambda_t] + [y_j] [\lambda_i^\dagger] + [y_i] [\lambda_{j,0}]$ $[\lambda_t] + [y_j] [\lambda_{i,0}] + [y_i] [\lambda_j^\dagger]$ $[\lambda_t] + [\Lambda_r([y_j] [\lambda_i^\dagger] + [y_i] [\lambda_j^\dagger], [0, 0])]$ $[\lambda_t] + [\Lambda_r([y_j] [\lambda_i^\dagger] + [y_i] [\lambda_{j,0}], [0, 0])]$ $[\lambda_t] + [\Lambda_r([y_j] [\lambda_{i,0}] + [y_i] [\lambda_j^\dagger], [0, 0])]$ $[\lambda_t] + [y_j] [\lambda_{i,0}] + [y_i] [\lambda_{j,0}]$	$\mathcal{L}_i = \mathcal{N} \wedge \mathcal{L}_j = \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_j = \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_i$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_j = \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_i \wedge \neg C_\star$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_j = \mathcal{N} \wedge C_\star$ $\mathcal{L}_i = \mathcal{N} \wedge \mathcal{L}_j \subset \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_j$ $\mathcal{L}_i = \mathcal{N} \wedge \mathcal{L}_j \subset \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_j \wedge \neg C_\star$ $\mathcal{L}_i = \mathcal{N} \wedge \mathcal{L}_j \subset \mathcal{N} \wedge C_\star$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_j \subset \mathcal{N} \wedge \mathcal{L}_\cup = \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_\cap$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_j \subset \mathcal{N} \wedge \mathcal{L}_\cup = \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_\cap \wedge \neg C_\star$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_j \subset \mathcal{N} \wedge C_\star$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_i = \mathcal{L}_j$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_i \subset \mathcal{L}_j$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_j \subset \mathcal{L}_i$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_i = \mathcal{L}_j$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_i \subset \mathcal{L}_j$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_j \subset \mathcal{L}_i$ $\mathcal{L}_\cup \subset \mathcal{N} \wedge \mathcal{L}_i \not\subseteq \mathcal{L}_j \wedge \mathcal{L}_j \not\subseteq \mathcal{L}_i$
powNat	$m(m-1)[y_i]^{m-2} [\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i])]$ $m[y_i]^{m-2} ((m-1)[\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i]) + [y_i] [\lambda_i^\dagger])]$ $m[y_i]^{m-2} ((m-1)[\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i]) + [y_i] [\lambda_{i,0}])]$	$\mathcal{L}_i = \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_i$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_i$
oneOver	$2[y_k]^3 [\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i])]$ $[y_k]^2 (2[y_k] [\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i)]) - [\lambda_i^\dagger])$ $[y_k]^2 (2[y_k] [\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i)]) - [\lambda_{i,0}])$	$\mathcal{L}_i = \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_i$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_i$
sqrt	$1/(-4[y_k]^3) [\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i])]$ $1/(2[y_k])(1/(-2[y_i])[\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i)]) + [\lambda_i^\dagger])$ $1/(2[y_k])(1/(-2[y_i])[\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i)]) + [\lambda_{i,0}])$	$\mathcal{L}_i = \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_i$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_i$
exp	$[y_k] [\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i])]$ $[y_k] ([\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i)]) + [\lambda_i^\dagger])$ $[y_k] ([\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i)]) + [\lambda_{i,0}])$	$\mathcal{L}_i = \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_i$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_i$
ln	$-1/[y_i]^2 [\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i])]$ $1/[y_i] ([\lambda_i^\dagger] - 1/[y_i] [\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i)])]$ $1/[y_i] ([\lambda_{i,0}] - 1/[y_i] [\Lambda_s([\nabla_{\mathcal{L}_k^c} y_i)])]$	$\mathcal{L}_i = \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k = \mathcal{L}_i$ $\mathcal{L}_i \subset \mathcal{N} \wedge \mathcal{L}_k \subset \mathcal{L}_i$
addC	$[0, 0]$ $[\lambda_i^\dagger]$	$\mathcal{L}_i = \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N}$
mulByC	$[0, 0]$ $c[\lambda_i^\dagger]$	$\mathcal{L}_i = \mathcal{N}$ $\mathcal{L}_i \subset \mathcal{N}$