Using generalized simplex methods to approximate derivatives

Gabriel Jarry-Bolduc · Chayne Planiden

January 18, 2024

Abstract This paper presents two methods for approximating a proper subset of the entries of a Hessian using only function evaluations. These approximations are obtained using the techniques called *generalized simplex Hessian* and *generalized centered simplex Hessian*. We show how to choose the matrices of directions involved in the computation of these two techniques depending on the entries of the Hessian of interest. We discuss the number of function evaluations required in each case and develop a general formula to approximate all order-P partial derivatives. Since only function evaluations are required to compute the methods discussed in this paper, they are suitable for use in derivative-free optimization methods.

Keywords Partial Hessian · Generalized simplex Hessian · Generalized centered simplex Hessian · Approximating order-P partial derivatives · derivative-free optimization methods

1 Introduction

Approximating Hessians is a popular topic in numerical analysis and optimization. The Hessian captures the curvature of a function, thus providing additional information that the gradient does not have and aiding in the optimization process. There exist many approaches to approximate Hessians, including automatic differentiation [11], graph coloring approach [4,10], Lagrange polynomials [6], Newton fundamental polynomials [6], regression nonlinear models or underdetermined interpolating models [5]. Arguably, two of the most well-known methods to approximate Hessian are *forward-finite-difference approximation* and *centered-finite-difference* approximation [1, Section 4.6].

True Hessians and approximate Hessians are often used in optimization methods. One of the most known optimization method using (true) Hessians is *Newton's Method* [2, Section 4.3], which boasts a quadratic rate of convergence [2, Theorem 4.3]. In derivative-free optimization (DFO)

Gabriel Jarry-Bolduc
Mathematics and Statistics Department, Saint Francis Xavier University,
Antigonish, Nova Scotia, Canada.
ORCID 0000-0002-1827-8508
E-mail: gabjarry@alumni.ubc.ca

Chayne Planiden
School of Mathematics and Applied Statistics, University of Wollongong,
Wollongong, NSW, 2500, Australia. Research supported by University of Wollongong. chayne@uow.edu.au,
ORCID 0000-0002-0412-8445

methods, it is assumed that derivative information is not directly available. For this reason, true Hessians are not employed. Approximate Hessians have been used in DFO methods since at least 1970 [30]. Researchers from the DFO community have previously explored methods to approximate full Hessians or some of the entries of the Hessian. In [8], the authors outline an idea for a *simplex Hessian* that is constructed via quadratic interpolation through $(n+1)(n+2)/2$ well-poised sample points. They further posit that if only the diagonal entries are desired, then $2n+1$ sample points are sufficient. These ideas are formalized in [5] through quadratic interpolation and analyzed through the use of Lagrange polynomials. Obtaining an approximation of the diagonal component of a Hessian is also discussed in [7,16], and in [14] it is shown that the diagonal entries can be obtained for free (in terms of function evaluations) if the gradient has been previously approximated via the *(generalized) centered simplex gradient* technique.

There are now many DFO algorithms that employ approximate Hessians to solve optimization problems (see, for example, [9,18,19,20,21,22,23,24,25,29]). To develop strong convergence results, a DFO algorithm uses numerical analysis techniques to approximate Hessians (and gradients) in a manner that has controllable error bounds. In [13,17], two techniques based on simple matrix algebra to approximate a full Hessian called the *generalized simplex Hessian (GSH)* and the *generalized centered simplex Hessian (GCSH)* are introduced. It is shown that the GSH is an order-1 accurate approximation of the full Hessian and that the GCSH is an order-2 accurate approximation of the full Hessian. The GSH can be viewed as a generalization of the *simplex Hessian* discussed in [6,8]. The simplex Hessian requires $(n+1)(n+2)/2$ sample points poised for quadratic interpolation. On the other hand, the GSH and the GCSH are well-defined as long as the matrices of directions utilized are nonempty. Hence, they offer enough flexibility to approximate only a proper subset of the entries of a Hessian.

The main goal of this paper is to investigate how to approximate a proper subset of the entries of the Hessian with the GSH and the GCSH. Error bounds are provided, showing that the GSH can provide order-1 accuracy of the appropriate subset of the entries of the Hessian. Using the GCSG, error bounds show that we can obtain order-2 accuracy on the appropriate subset of the entries of the Hessian. Secondary goals include to show how to obtain an order-2 accurate approximation of the full Hessian with a low number of function evaluations. Lastly, a general recursive formula to approximate all order-P partial derivatives is provided.

This paper is organized as follows. Section 2 contains a description of the notation and some needed definitions, including those of the generalized simplex gradient (GSG), the GSH and the GCSH. In Section 3, we discuss how to obtain an order-2 accurate approximation of the full Hessian with a low number of function evaluations. In Section 4, we investigate how to approximate a proper subset of the entries of a Hessian. Diagonal entries and the relation between the GSH and the centered simplex Hessian diagonal are discussed in Section 4 as well, with details provided on how to approximate the off-diagonal entries of a Hessian and a column of a Hessian. Error bounds are provided in each section. The properties of the matrices of directions, number of function evaluations required, and error bounds are described. In Section 5, a formula to approximate all order-P partial derivatives is introduced. Finally, Section 6 contains concluding remarks and recommends areas of future research in this vein.

2 Preliminaries

Throughout this work, we use the standard notation found in [27]. The domain of a function f is denoted by $\operatorname{dom} f$. The transpose of matrix A is denoted by $A^\top$. We work in finite-dimensional space $\mathbb{R}^n$ with inner product $x^\top y = \sum_{i=1}^n x_i y_i$ and induced norm $\|x\| = \sqrt{x^\top x}$. The identity matrix in $\mathbb{R}^{n\times n}$ is denoted by Id_n. We use $e_n^i \in \mathbb{R}^n$ where $i \in \{1,2,\ldots,n\}$, to denote the standard unit basis vectors in $\mathbb{R}^n$, i.e. the i^{th} column of Id_n. When there is no ambiguity about the dimension, we may omit the subscript and simply write e^i or Id. The zero vector in $\mathbb{R}^n$ is denoted by $\mathbf{0}_n$ and the zero matrix in $\mathbb{R}^{n\times m}$ is denoted by $\mathbf{0}_{n\times m}$. The entry in the i^{th} row and j^{th} column of a matrix $A \in \mathbb{R}^{n\times m}$ is denoted by $A_{i,j}$. If the matrix already involves a subscript, say k, then we use the notation $[A_k]_{i,j}$. The matrix $D = \mathrm{Diag}(v) = \mathrm{Diag}[v_1, \dots, v_n] \in \mathbb{R}^{n\times n}$, where $v \in \mathbb{R}^n$, represents a diagonal matrix with diagonal entries $D_{j,j} = v_j$ for all $j \in \{1, \dots, n\}$. The linear span of a matrix $A \in \mathbb{R}^{n\times m}$, denoted by $\operatorname{span} A$, represents the set generated by all linear combinations of the columns in A. The *Minkowski sum* of two sets of vectors A and B is denoted by $A \oplus B$ and defined as follows:

$$A \oplus B = \{a + b : a \in A, b \in B\}.$$

Given a matrix $A \in \mathbb{R}^{n\times m}$, we use the induced matrix norm

$$\|A\| = \|A\|_2 = \max\{\|Ax\|_2 \, : \, \|x\|_2 = 1\},$$

and the Frobenius norm

$$\|A\|_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} A_{i,j}^2 \right)^{\frac{1}{2}}.$$

We denote by $B_n(x^0; \Delta)$ and $\overline{B}_n(x^0; \Delta)$ the open and closed balls, respectively, centered at $x^0 \in \mathbb{R}^n$ with radius $\Delta > 0$. We define a quadratic function $Q : \mathbb{R}^n \to \mathbb{R}$ to be a function of the form $Q(x) = \alpha_0 + \alpha^\top x + \frac{1}{2} x^\top \mathbf{H} x$ where $\alpha_0 \in \mathbb{R}, \alpha \in \mathbb{R}^n$ and $\mathbf{H} = \mathbf{H}^\top \in \mathbb{R}^{n\times n}$. An affine function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}$ is defined to be any function that can be written in the form $\mathcal{L}(x) = \alpha_0 + \alpha^\top x$. Note that affine functions and constant functions $C(x) = \alpha_0$ are also considered quadratic functions, with $\mathbf{H} = \mathbf{0}_{n\times n}$.

In order to introduce the definitions of the GSG, the GSH and the GCSH, we require the Moore–Penrose pseudo-inverse of a matrix.

Definition 1 (Moore-Penrose pseudo-inverse) [28, Chapter 17]
Let $A \in \mathbb{R}^{n\times m}$. The unique matrix $A^\dagger \in \mathbb{R}^{m\times n}$ that satisfies the following four equations is called the Moore-Penrose pseudo-inverse of A:

(i) $AA^\dagger A = A$
(ii) $A^\dagger A A^\dagger = A^\dagger$
(iii) $(AA^\dagger)^\top = AA^\dagger$
(iv) $(A^\dagger A)^\top = A^\dagger A$.

The Moore–Penrose pseudoinverse $A^\dagger$ is not always an inverse of A, but the following two properties hold.

- If A has full column rank m, then $A^\dagger$ is a left inverse of A. That is, $A^\dagger A = \mathrm{Id}_m$ and

$$A^\dagger = (A^\top A)^{-1} A^\top. \tag{1}$$

- If A has full row rank n, then $A^\dagger$ is a right inverse of A. That is, $AA^\dagger = \mathrm{Id}_n$ and

$$A^\dagger = A^\top (AA^\top)^{-1}. \tag{2}$$

Definition 2 (GSG) [12, Definition 2] Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $x^0 \in \operatorname{dom} f$ be the point of interest. Let $S = \begin{bmatrix} s^1 \ s^2 \ \cdots \ s^m \end{bmatrix} \in \mathbb{R}^{n \times m}$ with $x^0 \oplus S \subseteq \operatorname{dom} f$. The *generalized simplex gradient* of f at x^0 over S is denoted by $\nabla_s f(x^0; S)$ and defined by

$$\nabla_s f(x^0; S) = (S^\top)^\dagger \delta_s f(x^0; S) \in \mathbb{R}^n,$$

where

$$\delta_s f(x^0; S) = \begin{bmatrix} f(x^0 + s^1) - f(x^0) \\ \vdots \\ f(x^0 + s^m) - f(x^0) \end{bmatrix} \in \mathbb{R}^m.$$

In the next definition, we recall key notation used in the construction of the GSH and the GCSH. Within, we write a set of vectors in matrix form, by which we mean that each column of the matrix is a vector in the set.

Definition 3 (GSH and GCSH notation) Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $x^0 \in \operatorname{dom} f$ be the point of interest. Let

$$\begin{aligned} S &= \begin{bmatrix} s^1 \ s^2 \ \cdots \ s^m \end{bmatrix} \in \mathbb{R}^{n \times m} \quad \text{and} \\ T_j &= \begin{bmatrix} t_j^1 \ t_j^2 \ \cdots \ t_j^k \end{bmatrix} \in \mathbb{R}^{n \times k_j}, j \in \{1, \dots, m\}, \end{aligned}$$

be sets of directions contained in $\mathbb{R}^n$, written in matrix form. Define

$$T_{1:m} = \{T_1, \dots, T_m\},$$

and

$$\Delta_S = \max_{j \in \{1, \dots, m\}} \|s^j\|, \quad \Delta_{T_j} = \max_{\ell \in \{1, \dots, k_j\}} \|t_j^\ell\|, \quad \Delta_T = \max_{j \in \{1, \dots, m\}} \Delta_{T_j}.$$

The *normalized matrices* $\widehat{S}$ and $\widehat{T_j}$ are respectively defined by

$$\widehat{S} = \frac{1}{\Delta_S} S, \quad \widehat{T_j} = \frac{T_j}{\Delta_{T_j}}, \quad j \in \{1, \dots, n\}. \tag{3}$$

In this paper, it is always assumed that the matrix S and all matrices T_j are non-empty and have non-null rank. This ensures that the matrices in (3) are well-defined.

Definition 4 (Generalized simplex Hessian) [13, Definition 2.6] Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $x^0 \in \operatorname{dom} f$ be the point of interest. Let $S = \begin{bmatrix} s^1 \ s^2 \ \cdots \ s^m \end{bmatrix} \in \mathbb{R}^{n \times m}$ and $T_j \in \mathbb{R}^{n \times k_j}$ with $x^0 \oplus T_j, x^0 \oplus S, x^0 + s^j \oplus T_j$ contained in $\operatorname{dom} f$ for all $j \in \{1, \dots, m\}$. The *generalized simplex Hessian of f at x^0 over S and $T_{1:m}$* is denoted by $\nabla_s^2 f(x^0; S, T_{1:m})$ and defined by

$$\nabla_s^2 f(x^0; S, T_{1:m}) = (S^\top)^\dagger \delta_s^2 f(x^0; S, T_{1:m}),$$

where

$$\delta_s^2 f(x^0; S, T_{1:m}) = \begin{bmatrix} (\nabla_s f(x^0 + s^1; T_1) - \nabla_s f(x^0; T_1))^\top \\ (\nabla_s f(x^0 + s^2; T_2) - \nabla_s f(x^0; T_2))^\top \\ \vdots \\ (\nabla_s f(x^0 + s^m; T_m) - \nabla_s f(x^0; T_m))^\top \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

Note that the number of columns m in S can be any positive integer and each number of columns k_j in T_j can be any positive integer for all $j \in \{1, \dots, m\}$.

In the case where all matrices T_j are equal, we define $\overline{T} = T_1 = \cdots = T_m \in \mathbb{R}^{n \times k}$ and write $\nabla_s^2 f(x^0; S, \overline{T})$ to emphasize this special case.

Definition 5 (Generalized centered simplex Hessian) [13, Definition 2.7] Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $x^0 \in \operatorname{dom} f$ be the point of interest. Let $S = \begin{bmatrix} s^1 & \cdots & s^m \end{bmatrix} \in \mathbb{R}^{n \times m}$ and $T_j \in \mathbb{R}^{n \times k_j}$ with $x^0 + s^j \oplus T_j, x^0 - s^j \oplus (-T_j), x^0 \oplus (\pm S)$, and $x^0 \oplus (\pm T_j)$ contained in $\operatorname{dom} f$ for all $j \in \{1, \dots, m\}$. The *generalized centered simplex Hessian of f at x^0 over S and $T_{1:m}$* is denoted by $\nabla_c^2 f(x^0; S, T_{1:m})$ and defined by

$$\nabla_c^2 f(x^0; S, T_{1:m}) = \frac{1}{2} \left(\nabla_s^2 f(x^0; S, T_{1:m}) + \nabla_s^2 f(x^0; -S, -T_{1:m}) \right). \tag{4}$$

The relation between the GSH and the GCSH is investigated in [13]. It is shown that the GCSH is a particular case of the GSH.

Proposition 6 *Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $x^0 \in \operatorname{dom} f$ be the point of interest. Let $A = \begin{bmatrix} a^1 & a^2 & \cdots & a^{2m} \end{bmatrix} = \begin{bmatrix} S & -S \end{bmatrix} \in \mathbb{R}^{n \times 2m}$ for some $S = \begin{bmatrix} s^1 & \cdots & s^m \end{bmatrix} \in \mathbb{R}^{n \times m}$ and let $B_j = T_j \in \mathbb{R}^{n \times k_j}$, and $B_{m+j} = -T_j$ for all $j \in \{1, \dots, m\}$. Suppose that $x^0 + s^j \oplus T_j, x^0 + (-s^j) \oplus (-T_j), x^0 \oplus (\pm S)$, and $x^0 \oplus (\pm T_j)$ are contained in $\operatorname{dom} f$ for all $j \in \{1, \dots, m\}$. Then*

$$\nabla_s^2 f(x^0; A, B_{1:2m}) = \nabla_c^2 f(x^0; S, T_{1:m}).$$

In [26], it is shown that the order of the sample points does not affect the value of the generalized simplex gradient. We now provide a simpler proof of this statement. It shows that the order of the columns in the matrices of directions used does not affect the value of the generalized simplex gradient. Consequently, the order of the columns does not affect the value of the GSH nor the GCSH.

Proposition 7 *Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $x^0 \in \operatorname{dom} f$ be the point of interest. Let $S = \begin{bmatrix} s^1 & s^2 & \cdots & s^m \end{bmatrix} \in \mathbb{R}^{n \times m}$ with $x^0 \oplus (\pm S) \subseteq \operatorname{dom} f$. Let $P \in \mathbb{R}^{m \times m}$ be a permutation matrix. Then*

$$\nabla_s f(x^0; SP) = \nabla_s f(x^0; S).$$

Proof We have

$$\begin{aligned}
\nabla_s f(x^0; SP) &= \left((SP)^\top \right)^\dagger \delta_s f(x^0; SP) \\
&= (P^\top S^\top)^\dagger P^\top \delta_s f(x^0; S) \\
&= (S^\top)^\dagger P P^\top \delta_s f(x^0; S) && \text{(since } P \text{ is an orthonormal matrix)} \\
&= (S^\top)^\dagger \operatorname{Id} \delta_s f(x^0; S) = \nabla_s f(x^0; S).
\end{aligned}$$

□

Next, we recall the different cases to classify a GSH or a GCSH.

Definition 8 [13, Definition 2.9] Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $x^0 \in \operatorname{dom} f$ be the point of interest. Let $S = \begin{bmatrix} s^1 & s^2 & \cdots & s^m \end{bmatrix} \in \mathbb{R}^{n \times m}$ and $T_j \in \mathbb{R}^{n \times k_j}$ with $x^0 \oplus s^j \oplus T_j, x^0 - s^j \oplus (-T_j), x^0 \oplus (\pm S)$, and $x^0 \oplus (\pm T_j)$ contained in $\operatorname{dom} f$ for all $j \in \{1, \dots, m\}$. Assume that all matrices are non-null rank. We define the following four cases to characterize the matrix $S \in \mathbb{R}^{n \times m}$ and the set of matrices $T_{1:m} = \{T_1, \dots, T_m\}$.

- Underdetermined: the GSH (GCSH) is said to be *S-underdetermined* if S is non-square and full column rank. We say that it is *$T_{1:m}$-underdetermined* if all matrices in the set $T_{1:m}$ are full column rank and at least one matrix is non-square.
- Determined: the GSH (GCSH) is said to be *S-determined* if S is square and full rank. It is *$T_{1:m}$-determined* if all matrices in the set $T_{1:m}$ are square and full rank.

- Overdetermined: the GSH (GCSH) is said to be S-*overdetermined* if S is non-square and full row rank. It is $T_{1:m}$-*overdetermined* if all matrices in the set $T_{1:m}$ are full row rank and at least one is non-square.
- Nondetermined: the GSH (GCSH) is said to be S-*nondetermined* if it is not in any of the previous three cases. It is $T_{1:m}$-*nondetermined* if the set $T_{1:m}$ is not in any of the previous three cases.

In the special case where all matrices T_j are equal, we may write $\overline{T}$- instead of $T_{1:m}$-. Note that the definition of an S-underdetermined GSH (GCSH) implies that $\operatorname{span} S \neq \mathbb{R}^n$, which is true if and only if $SS^\dagger \neq \mathrm{Id}_n$. Similarly, the definition of a $T_{1:m}$-underdetermined GSH (GCSH) implies that $\operatorname{span} T_j \neq \mathbb{R}^n$ for some $j \in \{1, \dots, m\}$, which is true if and only if $T_j T_j^\dagger \neq \mathrm{Id}_n$ for some j.

It turns out that error bounds can be defined between the GSH (GCSH) and some of the entries of the true Hessian. The appropriate entries of the true Hessian are obtained via a projection operator. The projection operator involves all matrices of directions utilized to compute the GSH (GCSH).

Given matrices $S \in \mathbb{R}^{n\times m}$ and $T_j \in \mathbb{R}^{n \times k_j}$, the projection of the matrix $H \in \mathbb{R}^{n\times n}$ onto S and $T_{1:m}$ is denoted by $\operatorname{Proj}_{S,T_{1:m}} H$ and defined by

$$\operatorname{Proj}_{S,T_{1:m}} H = \sum_{j=1}^{m} (S^\top)^\dagger e_m^j (e_m^j)^\top S^\top H T_j T_j^\dagger.$$

In the case where $T_1 = T_2 = \cdots = T_m = \overline{T}$, the projection of H onto S and $\overline{T}$ is denoted by $\operatorname{Proj}_{S,\overline{T}} H$, and reduces to

$$\begin{aligned}
\operatorname{Proj}_{S,\overline{T}} H &= \sum_{j=1}^{m} (S^\top)^\dagger e_m^j (e_m^j)^\top S^\top H \overline{T}\,\overline{T}^\dagger \\
&= (S^\top)^\dagger \left(\sum_{j=1}^{m} e_m^j (e_m^j)^\top \right) S^\top H \overline{T}\,\overline{T}^\dagger \\
&= (S^\top)^\dagger \mathrm{Id}_m S^\top H \overline{T}\,\overline{T}^\dagger \\
&= (S^\top)^\dagger S^\top H \overline{T}\,\overline{T}^\dagger.
\end{aligned}$$

Note that $\operatorname{Proj}_{S,T_{1:m}}$ is a linear operator. The following proposition demonstrates that in certain situations, the projection of the GSH (GCSH) onto S and $T_{1:m}$ is equal to the GSH (GCSH)

Proposition 9 [13, Proposition 4.1] *Let* $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ *and* $x^0 \in \operatorname{dom} f$. *Let* $S = \begin{bmatrix} s^1 \cdots s^m \end{bmatrix} \in \mathbb{R}^{n\times m}$, *and* $T_j \in \mathbb{R}^{n\times k_j}$. *Assume that* $x^0 \oplus (\pm S), x^0 \oplus (\pm T_j)$, *and* $x^0 \oplus (\pm(S \oplus T_j))$ *are contained in* $\operatorname{dom} f$ *for all* j. *Then the following hold.*

(i) If S *is full column rank* **or** T_j *is full row rank for all* $j \in \{1, \dots, m\}$, *then*

$$\operatorname{Proj}_{S,T_{1:m}} \nabla_s^2 f(x^0; S, T_{1:m}) = \nabla_s^2 f(x^0; S, T_{1:m})$$

and

$$\operatorname{Proj}_{S,T_{1:m}} \nabla_c^2 f(x^0; S, T_{1:m}) = \nabla_c^2 f(x^0; S, T_{1:m}).$$

The following error bounds for the GSH and the GCSH were introduced in [13]. In the following theorem and the remainder of this paper, we use the notation

$$\begin{aligned}
\Delta_u &= \max\{\Delta_S, \Delta_{T_1}, \dots, \Delta_{T_m}\},\\
\Delta_l &= \min\{\Delta_S, \Delta_{T_1}, \dots, \Delta_{T_m}\},\\
\widehat{T} &= \widehat{T}_j \quad \text{such that} \quad \left\|\widehat{T}_j^\dagger\right\| \quad \text{is maximal}, \quad j \in \{1, \dots, m\},\\
k &= \max\{k_1, \dots, k_m\}.
\end{aligned}$$

Theorem 10 (Error bounds for the GSH) *Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^3$ on $B_n(x^0; \overline{\Delta})$ where $x^0 \in \operatorname{dom} f$ is the point of interest and $\overline{\Delta} > 0$. Denote by $L_{\nabla^2 f} \geq 0$ the Lipschitz constant of $\nabla^2 f$ on $\overline{B}_n(x^0; \overline{\Delta})$. Let $S = \begin{bmatrix} s^1 \ s^2 \ \cdots \ s^m \end{bmatrix} \in \mathbb{R}^{n\times m}$ and $T_j = \begin{bmatrix} t_j^1 \ t_j^2 \ \cdots \ t_j^{k_j} \end{bmatrix} \in \mathbb{R}^{n\times k_j}$ for all $j \in \{1, \dots, m\}$. Assume that $B_n(x^0; \Delta_{T_j}) \subset B_n(x^0; \overline{\Delta})$ and $B_n(x^0 + s^j; \Delta_{T_j}) \subset B_n(x^0; \overline{\Delta})$ for all $j \in \{1, \dots m\}$. Then the following hold.*

(i) If S is full column rank **or** *T_j is full row rank for all $j \in \{1, \dots, m\}$,, then*

$$\begin{aligned}
&\| \operatorname{Proj}_{S,T_{1:m}} \nabla_s^2 f(x^0; S, T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} \nabla^2 f(x^0)\|\\
&= \|\nabla_s^2 f(x^0; S, T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} \nabla^2 f(x^0)\|\\
&\leq 4m\sqrt{k} L_{\nabla^2 f} \|(\widehat{S}^\top)^\dagger\| \|\widehat{T}^\dagger\| \left(\frac{\Delta_u}{\Delta_l}\right)^2 \Delta_u. \qquad (5)
\end{aligned}$$

(ii) If $T_1 = T_2 = \cdots = T_m = \overline{T}$, then

$$\begin{aligned}
&\| \operatorname{Proj}_{S,\overline{T}} \nabla_s^2(x^0, S, \overline{T}) - \operatorname{Proj}_{S,\overline{T}} \nabla^2 f(x^0)\|\\
&= \|\nabla_s^2(x^0, S, \overline{T}) - \operatorname{Proj}_{S,\overline{T}} \nabla^2 f(x^0)\|\\
&\leq 4\sqrt{mk} L_{\nabla^2 f} \frac{\Delta_u}{\Delta_l} \left\|(\widehat{S}^\top)^\dagger\right\| \left\|\widehat{\overline{T}}^\dagger\right\| \Delta_u. \qquad (6)
\end{aligned}$$

Theorem 11 (Error bounds for the GCSH) *Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^4$ on $B_n(x^0; \overline{\Delta})$ where $x^0 \in \operatorname{dom} f$ is the point of interest and $\overline{\Delta} > 0$. Denote by $L_{\nabla^3 f}$ the Lipschitz constant of $\nabla^3 f$ on $\overline{B}_n(x^0; \overline{\Delta})$. Let $S = \begin{bmatrix} s^1 \ s^2 \ \cdots \ s^m \end{bmatrix} \in \mathbb{R}^{n\times m}$, $T_j = \begin{bmatrix} t_j^1 \ t_j^2 \ \cdots \ t_j^{k_j} \end{bmatrix} \in \mathbb{R}^{n\times k_j}$ with the ball $B_n(x^0 + s^j; \Delta_{T_j}) \subset B_n(x^0; \overline{\Delta})$ for all $j \in \{1, \dots, m\}$. Then the following hold.*

(i) If S is full column rank **or** *T_j is full row rank for all $j \in \{1, \dots, m\}$, then*

$$\begin{aligned}
&\| \operatorname{Proj}_{S,T_{1:m}} \nabla_c^2 f(x^0; S, T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} \nabla^2 f(x^0)\|\\
&= \|\nabla_c^2 f(x^0; S, T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} \nabla^2 f(x^0)\|\\
&\leq 2m\sqrt{k} L_{\nabla^3 f} \left(\frac{\Delta_u}{\Delta_l}\right)^2 \left\|(\widehat{S}^\top)^\dagger\right\| \left\|\left(\widehat{T}\right)^\dagger\right\| \Delta_u^2. \qquad (7)
\end{aligned}$$

(ii) If $T_1 = T_2 = \cdots = T_m = \overline{T} \in \mathbb{R}^{n\times k}$, then

$$\begin{aligned}
&\left\|\operatorname{Proj}_{S,\overline{T}} \nabla_c^2(x^0; S, \overline{T}) - \operatorname{Proj}_{S,\overline{T}} \nabla^2 f(x^0)\right\|\\
&= \left\|\nabla_c^2(x^0; S, \overline{T}) - \operatorname{Proj}_{S,\overline{T}} \nabla^2 f(x^0)\right\|\\
&\leq 2\sqrt{mk} L_{\nabla^3 f} \frac{\Delta_u}{\Delta_l} \left\|(\widehat{S}^\top)^\dagger\right\| \left\|\left(\widehat{\overline{T}}\right)^\dagger\right\| \Delta_u^2. \qquad (8)
\end{aligned}$$

We use the following definition to describe the order at which an approximation technique converges.

Definition 12 [3, Definition 1.19] Let $f : \mathbb{R}_+ \to \mathbb{R}^{n\times m}$ and $g : \mathbb{R}_+ \to \mathbb{R}$. Suppose that $\lim_{\Delta\to 0} g(\Delta) = 0$ and $\lim_{\Delta\to 0} f(\Delta) = L \in \mathbb{R}^{n\times m}$. If there exists a scalar $\kappa \geq 0$ with

$$\|f(\Delta) - L\| \leq \kappa\, g(\Delta) \quad \text{for sufficiently small } \Delta,$$

then we say $f(\Delta)$ is $O(g(\Delta))$, or $f(\Delta)$ is a $O(g(\Delta))$ accurate approximation of L.

In this paper, $g(\Delta)$ takes the form $g(\Delta) = \Delta^N$, where $N \in \mathbb{N}$. If Definition 12 is satisfied, we will say that $f(\Delta)$ is an *order-N accurate approximation of* L where N is the greatest positive integer satisfying Definition 12.

3 Minimal poised set for the GCSH

In this section, we investigate how to obtain an order-2 accurate approximation of the full Hessian with a minimal number of function evaluations. We define a *minimal poised set* for the GCSH; the following notation is used.

$$\begin{aligned}
\mathcal{S}_s(x^0; S, \overline{T}) &: \text{set of all distinct points utilized to compute } \nabla^2_s f(x^0; S, \overline{T}),\\
\mathcal{S}_c(x^0; S, \overline{T}) &: \text{set of all distinct points utilized to compute } \nabla^2_c f(x^0; S, \overline{T}).
\end{aligned}$$

First, recall the definition of a minimal poised set for the GSH introduced in [13, Definition 5.2].

Definition 13 (Minimal poised set for the GSH) Let $x^0 \in \mathbb{R}^n$ be the point of interest. We say that $\mathcal{S}_s(x^0; S, \overline{T})$ is a minimal poised set for the GSH at x^0 if and only if it is S-determined, $\overline{T}$-determined and $\mathcal{S}_s(x^0; S, \overline{T})$ contains exactly $(n+1)(n+2)/2$ distinct points.

Similar to the previous definition, we introduce the definition of a minimal poised set for the GCSH.

Definition 14 (Minimal poised set for the GCSH) Let $x^0 \in \mathbb{R}^n$ be the point of interest. We say that $\mathcal{S}_c(x^0; S, \overline{T})$ is a minimal poised set for the GCSH at x^0 if and only if it is S-determined, $\overline{T}$-determined and $\mathcal{S}_c(x^0; S, \overline{T})$ contains exactly $n^2 + n + 1$ distinct points.

Next, we provide a choice of matrices of directions that creates a minimal poised set for the GCSH.

Proposition 15 *Let $S \in \mathbb{R}^{n\times n}$ be full rank and let $T = -S$. Then $\mathcal{S}_c(x^0; S, \overline{T})$ is a minimal poised set for the GCSH at x^0.*

Proof The matrices S and T are clearly full row rank. The set $\mathcal{S}_c(x^0; S, \overline{T})$ contains the following sample points:

$$x^0 \oplus S \oplus -S, x^0 \oplus -S \oplus S, x^0 \oplus \pm S, x^0.$$

Since the set $x^0 \oplus S \oplus -S$ is equal to to the set $x^0 \oplus -S \oplus S$, we drop $x^0 \oplus S \oplus -S$. The set $x^0 \oplus \pm S$ contains $2n$ distinct sample points.

The set $x^0 \oplus -S \oplus S$ contains $n^2 - n + 1$ distinct sample points and it contains the point x^0. Since S contains n linearly independent directions, it follows that a direction in $x^0 \oplus -S \oplus S$ cannot be contained in $x^0 \oplus S$ nor $x^0 \oplus -S$. Hence, the number of distinct sample points is

$$2n + (n^2 - n + 1) = n^2 + n + 1.$$

Therefore, $\mathcal{S}_c(x^0; S, \overline{T})$ is a minimal poised set for the GCSH.

□

It seems that n^2+n+1 sample points is the minimal amount of sample points to obtain an order-2 accuracy of the full Hessian. However, a rigorous proof of this statement is a future direction to investigate.

4 Approximating a proper subset of the entries of the Hessian

In this section, we provide details on how to choose the matrices of directions S and T_j when we are interested in a proper subset of the entries of the Hessian. In particular, we investigate how to approximate the diagonal entries, the off-diagonal entries and a column of the Hessian. The number of function evaluations required is discussed and an error bound is provided in each case. The relation between the *centered simplex Hessian diagonal* (CSHD) introduced in [16] and the GCSH is discussed. We begin by presenting results on how to approximate some or all diagonal entries of the Hessian.

4.1 Approximating the diagonal entries of the Hessian

An explicit formula to compute all the diagonal entries of the Hessian, which is well-defined regardless of the number of sample points utilized, is discussed in [16]. The CSHD is an approximation technique that provides an order-2 accurate approximation of the diagonal entries of the Hessian. We begin by showing that the CSHD is a specific case of the GCSH when the appropriate matrices of directions S and T_j are employed. First, recall the definitions of the *Hadamard product* and the CSHD.

Definition 16 [15] Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times m}$. The Hadamard product of A and B, denoted $A \odot B$ is the component-wise product. That is $[A \odot B]_{i,j} = A_{i,j}B_{i,j}$ for all $i \in \{1,\dots n\}$ and $j \in \{1,\dots,m\}$.

Definition 17 (Centered simplex Hessian diagonal) [16] Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$, $x^0 \in \operatorname{dom} f$ be the point of interest, $S = \begin{bmatrix} s^1 & s^2 & \cdots & s^m\end{bmatrix} \in \mathbb{R}^{n\times m}$ and $W = \begin{bmatrix} s^1 \odot s^1 & \cdots & s^m \odot s^m\end{bmatrix} \in \mathbb{R}^{n\times m}$. Assume that $x^0 \oplus (\pm S) \subset \operatorname{dom} f$. The *centered simplex Hessian diagonal of f at x^0 over S*, denoted by $d\nabla^2 f(x^0;S)$ is a vector in $\mathbb{R}^n$ given by

$$d\nabla^2 f(x^0;S) = (W^\top)^\dagger \varepsilon f(x^0;S),$$

where

$$\varepsilon f(x^0;S) = \begin{bmatrix} f(x^0+s^1)+f(x^0-s^1)-2f(x^0) \\ \vdots \\ f(x^0+s^m)+f(x^0-s^m)-2f(x^0)\end{bmatrix} \in \mathbb{R}^m.$$

Definition 18 (Partial diagonal matrix) Let $M \in \mathbb{R}^{n\times m}, m \leq n$. We say that M is a partial diagonal matrix if there exists a diagonal matrix $D \in \mathbb{R}^{n\times n}$ such that for each column $Me^j, j \in \{1,\dots m\}$, there exists an index $i \in \{1,\dots,n\}$ that yields $Me^j = De^i$.

In other words, a partial diagonal matrix is a subset of the columns of a single diagonal matrix. For example, the matrix

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 2\end{bmatrix}$$

is a partial diagonal matrix, but

$$\widetilde{M} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix}$$

is not a partial diagonal matrix.

Note that a partial diagonal matrix is full column rank if and only if it does not contain a column equal to the zero vector in $\mathbb{R}^n$.

The following lemma provides details about the Moore–Penrose pseudo-inverse of a partial diagonal matrix with full column rank.

Lemma 19 *Let $S = \begin{bmatrix} s^1 \ s^2 \ \cdots \ s^m \end{bmatrix} \in \mathbb{R}^{n \times m}$ where $m \leq n$ be a partial diagonal matrix with full column rank. Then*

$$S^\dagger = \begin{bmatrix} (s^1)^\dagger (s^2)^\dagger \cdots (s^m)^\dagger \end{bmatrix}^\top .$$

Proof Let u_j be the index in $\{1, \ldots, m\}$ of the only non-zero entry in column s^j. Since S is full column rank, using (1), we have

$$\begin{aligned} S^\dagger &= (S^\top S)^{-1} S^\top \\ &= \left(\operatorname{Diag}\left[(s^1_{u_1})^2 \cdots (s^m_{u_m})^2\right]\right)^{-1} S^\top \\ &= \operatorname{Diag}\left[\frac{1}{(s^1_{u_1})^2} \cdots \frac{1}{(s^m_{u_m})^2}\right] S^\top \\ &= \left[\frac{1}{s^1_{u_1}} e^{u_1} \cdots \frac{1}{s^m_{u_m}} e^{u_m}\right]^\top . \end{aligned}$$

Since S is full column rank, using (1) we find $(s^j)^\dagger = \frac{1}{s^j_{u_j}} (e^{u_j})^\top$ for all $j \in \{1, \ldots, m\}$. Therefore, $S^\dagger = \begin{bmatrix} (s^1)^\dagger \ (s^2)^\dagger \ \cdots \ (s^m)^\dagger \end{bmatrix}^\top$. □

The following theorem provides a sufficient condition for the GCSH to return the same approximation of the diagonal entries of the Hessian as the CSHD.

Theorem 20 *Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}, x^0 \in \operatorname{dom} f$ be the point of interest, $S = \begin{bmatrix} s^1 \ s^2 \ \cdots \ s^m \end{bmatrix} \in \mathbb{R}^{n \times m}$ and $T_j = -s^j \in \mathbb{R}^n$ for all $j \in \{1, \ldots, m\}$. Let $z \in \mathbb{R}^n$ be a vector containing the n diagonal entries of $\nabla^2_c f(x^0; S, T_{1:m})$. That is $z_i = \left[\nabla^2_c f(x^0; S, T_{1:m})\right]_{i,i}$ for all $i \in \{1, \ldots, n\}$. If S is a partial diagonal matrix with full column rank, then $z = d\nabla^2 f(x^0; S)$.*

Proof Let $A = \begin{bmatrix} S & -S \end{bmatrix} \in \mathbb{R}^{n \times 2m}$ and $T_{m+j} = s^j$ for $j \in \{1, \ldots, m\}$. We have

$$\begin{aligned} \nabla^2_c f(x^0; S, T_{1:m}) &= \nabla^2_s f(x^0; A, T_{1:2m}) \qquad \text{(by Proposition 6)} \\ &= (A^\top)^\dagger \begin{bmatrix} \left(\nabla_s f(x^0 + s^1; -s^1) - \nabla_s f(x^0; -s^1)\right)^\top \\ \vdots \\ \left(\nabla_s f(x^0 + s^m; -s^m) - \nabla_s f(x^0; -s^m)\right)^\top \\ \left(\nabla_s f(x^0 - s^1; s^1) - \nabla_s f(x^0; s^1)\right)^\top \\ \vdots \\ \left(\nabla_s f(x^0 - s^m; s^m) - \nabla_s f(x^0; s^m)\right)^\top \end{bmatrix} . \end{aligned}$$

Since $(A^\top)^\dagger = \frac{1}{2}\left[(S^\top)^\dagger \; -(S^\top)^\dagger\right]$, and expanding each row of the form

$$\left(\nabla_s f(x^0 \pm s^j; \mp s^j) - \nabla_s f(x^0; \mp s^j)\right)^\top,$$

we obtain

$$\begin{aligned}
\nabla_c^2 f(x^0; S, T_{1:m}) &= \frac{1}{2}\left[(S^\top)^\dagger \; -(S^\top)^\dagger\right] \begin{bmatrix} (-s^1)^\dagger \left(-f(x^0+s^1) - f(x^0-s^1) + 2f(x^0)\right) \\ \vdots \\ (-s^m)^\dagger \left(-f(x^0+s^m) - f(x^0-s^m) + 2f(x^0)\right) \\ (s^1)^\dagger \left(2f(x^0) - f(x^0-s^1) - f(x^0+s^1)\right) \\ \vdots \\ (s^m)^\dagger \left(2f(x^0) - f(x^0-s^m) - f(x^0+s^m)\right) \end{bmatrix} \\
&= (S^\dagger)^\top \begin{bmatrix} (s^1)^\dagger \left(f(x^0-s^1) + f(x^0+s^1) - 2f(x^0)\right) \\ \vdots \\ (s^m)^\dagger \left(f(x^0-s^m) + f(x^0+s^m) - 2f(x^0)\right) \end{bmatrix} \\
&= \left[((s^1)^\dagger)^\top \cdots ((s^m)^\dagger)^\top\right] \begin{bmatrix} (s^1)^\dagger \left(f(x^0-s^1) + f(x^0+s^1) - 2f(x^0)\right) \\ \vdots \\ (s^m)^\dagger \left(f(x^0-s^m) + f(x^0+s^m) - 2f(x^0)\right) \end{bmatrix}
\end{aligned}$$

by Lemma 19. Let $z \in \mathbb{R}^n$ be the vector containing the n diagonal entries of the previous equation. Then

$$\begin{aligned}
z &= \left[((s^1)^\top)^\dagger \odot ((s^1)^\top)^\dagger \cdots ((s^m)^\top)^\dagger \odot ((s^m)^\top)^\dagger\right] \varepsilon f(x^0; S) \\
&= (W^\top)^\dagger \varepsilon f(x^0; S) = d\nabla^2 f(x^0; S).
\end{aligned}$$

By defining the sets T_j as in Theorem 20, the CSHD and the GCSH use the same set of sample points. However, if S is not a partial diagonal matrix with full column rank, then the vector z containing the diagonal entries of the GCSH is not necessarily equal to the CSHD. Moreover, the GCSH is not necessarily a diagonal matrix. The following two examples illustrate these claims.

Example 21 Let

$$S = \begin{bmatrix} s^1 & s^2 & s^3 \end{bmatrix} = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0.2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let $T_j = -s^j$ for all $j \in \{1, 2, 3\}$. Let $f(x) = -2x_1^4 + x_2^4 + 10x_3^4$ and $x^0 = \begin{bmatrix} 2 & -2 & 5 \end{bmatrix}^\top$. Note that

$$\nabla^2 f(x^0) = \mathrm{Diag}[-96 \; 48 \; 3000].$$

The GCSH is

$$\nabla_c^2 f(x^0; S, T_{1:3}) = \mathrm{Diag}[-96.04 \; 48.068 \; 0],$$

and the CSHD is

$$d\nabla^2 f(x^0; S) = \begin{bmatrix} -96.04 & 48.0765 & 0 \end{bmatrix}^\top.$$

The next example shows that the GCSH is not necessarily a diagonal matrix, even when we use the same set of sample points.

Example 22 Let

$$S = \begin{bmatrix} s^1 & s^2 \end{bmatrix} = \begin{bmatrix} 0.1 & 0.1 \\ 0 & 0.1 \\ 0 & 0 \end{bmatrix}.$$

Let $T_j = -s^j$ for all $j \in \{1, 2\}$ Consider the same function and point of interest as in the previous example. That is $f(x) = -2x_1^4 + x_2^4 + 10x_3^4$ and $x^0 = \begin{bmatrix} 2 & -2 & 5 \end{bmatrix}^\top$. Then the GCSH is

$$\nabla_c^2 f(x^0; S, T_{1:2}) = \begin{bmatrix} -96.04 & 0 & 0 \\ 72.03 & -24.01 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and the CSHD is

$$d\nabla^2 f(x^0; S) = \begin{bmatrix} -96.04 & 48.02 & 0 \end{bmatrix}^\top.$$

The next theorem presents a general error bound when the matrices of directions T_j used in the computation of the GCSH have the form $T_j = -s^j$ for all $j \in \{1, \dots, m\}$. We begin by introducing a result concerning the projection of a matrix over S and $T_{1:m}$.

Proposition 23 *Let $M \in \mathbb{R}^{n\times n}$. Let $S = \begin{bmatrix} s^1 & \cdots & s^m \end{bmatrix} \in \mathbb{R}^{n\times m}$ and let $T_j = -s^j$ for all $j \in \{1, \dots, m\}$. If S is a partial diagonal matrix with full column rank, then*

$$\operatorname{Proj}_{S,T_{1:m}} M = \operatorname{Proj}_{S,T_{1:m}} \operatorname{Diag} \begin{bmatrix} M_{1,1} & \cdots & M_{n,n} \end{bmatrix}.$$

Moreover, if $(e^i)^\top S \neq \mathbf{0}_m^\top$ for some $i \in \{1, \dots, n\}$, then

$$\left[\operatorname{Proj}_{S,T_{1:m}} M\right]_{i,i} = M_{i,i}.$$

If $(e^i)^\top S = \mathbf{0}_m^\top$ for some $i \in \{1, \dots, n\}$, then

$$\left[\operatorname{Proj}_{S,T_{1:m}} M\right]_{i,i} = 0.$$

Proof We have

$$\begin{aligned}
\sum_{j=1}^m (S^\top)^\dagger e^j (e^j)^\top S^\top M T_j T_j^\dagger &= \sum_{j=1}^m ((s^j)^\top)^\dagger (s^j)^\top M(-s^j)(-s^j)^\dagger \\
&= \sum_{j=1}^m e^{u_j} (e^{u_j})^\top M e^{u_j} (e^{u_j})^\top
\end{aligned}$$

where u_j represents the index of the only nonzero entry in s^j, $u_j \in \{1, \dots n\}$, and $j \in \{1, \dots, m\}$. From the definition of a partial diagonal matrix, we know that $u_j \neq u_{\bar{j}}$ whenever $j \neq \bar{j}$, j and $\bar{j}$ in $\{1, \dots, m\}$. Noticing that $e^{u_j}(e^{u_j})^\top = \operatorname{Diag}(e^{u_j})$, we get

$$\begin{aligned}
\sum_{j=1}^m (S^\top)^\dagger e^j (e^j)^\top S^\top M T_j T_j^\dagger &= \sum_{j=1}^m \operatorname{Diag}(e^{u_j}) M \operatorname{Diag}(e^{u_j}) \\
&= \sum_{j=1}^m \operatorname{Diag}(e^{u_j}) \cdot M_{u_j,u_j} = \operatorname{Proj}_{S,T_{1:m}} \operatorname{Diag}[M_{1,1} \ \cdots \ M_{n,n}].
\end{aligned}$$

The rest of the proof follows immediately from the fact that $m \leq n$, and $u_j \neq u_{\bar{j}}$ whenever $j \neq \bar{j}$, j and $\bar{j}$ in $\{1, \dots, m\}$. $\square$

The notation D is now used to represent the diagonal matrix in $\mathbb{R}^{n\times n}$ containing the diagonal entries of the Hessian $\nabla^2 f(x^0)$. That is $D_{i,i} = [\nabla^2 f(x^0)]_{i,i}$ for all $i \in \{1,\dots,n\}$. If S is a diagonal matrix with full column rank and $T_j = -s^j$ for all $j \in \{1,\dots,n\}$, it follows from the previous proposition that

$$\operatorname{Proj}_{S,T_{1:m}} \nabla^2 f(x^0) = \operatorname{Proj}_{S,T_{1:m}} D.$$

In other words, the projection of the full true Hessian is a diagonal matrix that keeps intact all diagonal entries of the true Hessian. In the case where S is a non-square partial diagonal matrix, then it makes the (i,i) diagonal entry of the true Hessian equal to zero if S does not contain a multiple of the identity column e^i. Also, since S is full column rank, it follows from Propositions 23 and 9(ii) that $\nabla^2_s f(x^0;S,T_{1:m})$ and $\nabla^2_c f(x^0;S,T_{1:m})$ are diagonal matrices.

The next theorem presents an error bound when the GCSH is used to approximate some or all diagonal entries of the true Hessian.

Theorem 24 (Error bound for the diagonal entries of the Hessian) *Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^4$ on an open domain containing $\overline{B}_n(x^0;\Delta_S)$ where $x^0 \in \operatorname{dom} f$ is the point of interest and $\Delta_S > 0$ is the radius of $S = \begin{bmatrix} s^1 \cdots s^m \end{bmatrix} \in \mathbb{R}^{n\times m}$. Let $T_j = -s^j$ for all $j \in \{1,\dots,m\}$. Denote by $L_{\nabla^3 f} \geq 0$ the Lipschitz constant of $\nabla^3 f$ on $\overline{B}_n(x^0;\Delta_S)$. If S is a partial diagonal matrix with full column rank, then*

$$\begin{aligned}
&\left\|\operatorname{Proj}_{S,T_{1:m}} \nabla^2_c f(x^0;S,T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} \nabla^2 f(x^0)\right\| \\
&\quad = \left\|\nabla^2_c f(x^0;S,T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} D\right\| \leq \frac{1}{12} L_{\nabla^3 f}\Delta_S^2. \qquad (9)
\end{aligned}$$

Proof By Proposition 9(i) and Proposition 23, we get the equality. To make notation more compact, let $\varepsilon = \varepsilon f(x^0;S) \in \mathbb{R}^m$. We have

$$\begin{aligned}
&\left\|\nabla^2_c f(x^0;S,T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} D\right\| \\
&= \left\|\sum_{i=1}^m ((s^i)^\top)^\dagger (s^i)^\top (S^\top)^\dagger \operatorname{Diag}(\varepsilon) S^\dagger(-s^i)(-s^i)^\dagger - \sum_{i=1}^m ((s^i)^\top)^\dagger (s^i)^\top D s^i (s^i)^\dagger\right\| \\
&\leq \max_{i=1,\dots,m} \left(\|((s^i)^\top)^\dagger\|\|(s^i)^\dagger\| \left|(s^i)^\top (S^\top)^\dagger \operatorname{Diag}(\epsilon) S^\dagger s^i - (s^i)^\top D s^i\right|\right) \\
&= \max_{j=1,\dots,m} \left(\frac{1}{\|s^j\|^2}\left|\varepsilon_j - (s^j)^\top D s^j\right|\right).
\end{aligned}$$

By Taylor's Theorem, using a similar process as in the proof in [16, Theorem 3.3], we obtain

$$\begin{aligned}
\left\|\nabla^2_c f(x^0;S,T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} D\right\| &\leq \max_{j=1,\dots,m}\left(\frac{1}{\|s^j\|^2}\frac{1}{12}L_{\nabla^3 f}\|s^j\|^4\right) \\
&= \max_{j=1,\dots,m}\left(\frac{1}{12}L_{\nabla^3 f}\|s^j\|^2\right) \\
&\leq \frac{1}{12}L_{\nabla^3 f}\Delta_S^2.
\end{aligned}$$

By defining S and T_j as in the previous theorem, the GCSH is S-underdetermined\determined and $T_{1:m}$-underdetermined. Hence, the general error bound proposed for the GCSG in Theorem 11(ii) is also valid. The previous proof utilized properties of partial diagonal matrices to obtain a tighter error bound than the one proposed in Theorem 11(ii).

The previous theorem shows how to obtain an order-2 accurate approximation of some or all diagonal entries of the Hessian. This requires $2n+1$ function evaluations when S is square. If we are interested in approximating only one diagonal entry of a Hessian $\nabla^2 f(x^0)$, say $\left[\nabla^2 f(x^0)\right]_{i,i}$, then the computational cost is three function evaluations. In this case, we can choose $S = he^i$ and $T_1 = -he^i$. Each additional diagonal entry can be obtained for two more function evaluations.

Other matrices of directions S and T_j may be used to obtain an approximation of all diagonal entries of a Hessian. For instance, the following matrices can be used:

$$S = \left[s^1 \cdots s^n\right] = h\,\mathrm{Id}, \quad T_j = s^j, \quad \text{for all}\, j \in \{1, \dots, n\}, h \neq 0.$$

In this case, S is diagonal with full column rank and it follows from Proposition 23 that $\mathrm{Proj}_{S,T_{1:m}} \nabla^2 f(x^0) = \mathrm{Proj}_{S,T_{1:m}} D = D$ where $D \in \mathbb{R}^{n\times n}$ is the diagonal matrix such that $D_{i,i} = [\nabla^2 f(x^0)]_{i,i}$ for all $i \in \{1, \dots, n\}$. By Theorem 10(ii), this choice of matrices provides an order-1 accurate approximation of all diagonal entries of the Hessian. The computation of $\nabla_s^2 f(x^0; S, T_{1:n})$ requires $2n+1$ function evaluations. Hence, it is preferable to choose the matrices of directions S and T_j as in Theorem 24, since it provides a greater order of accuracy for the same number of function evaluations.

In the next section, we investigate the approximation of some or all off-diagonal entries of the Hessian.

4.2 Approximating the off-diagonal entries of the Hessian

In this section, how to approximate some or all off-diagonal entries of the Hessian is examined. First, recall that the Hessian $\nabla^2 f(x^0)$ is symmetric whenever $f \in \mathcal{C}^2$. Therefore, it is sufficient to consider the off-diagonal entries $[\nabla^2 f(x^0)]_{i,j}$ such that $i < j$. It is possible to approximate some or all off-diagonal entries of the Hessian by setting the matrices of directions S and T_j in the following way. Define

$$\begin{aligned}
&\widetilde{S} \in \mathbb{R}^{n\times n-1} : \text{a partial diagonal matrix with full column rank}\\
&\qquad\qquad\qquad \text{such that the } n^{\text{th}} \text{ row is equal to } \mathbf{0}_{n-1}^\top,\\
&S = \left[s^1 \cdots s^m\right] \in \mathbb{R}^{n\times m} : \text{a non-empty subset of the columns of } \widetilde{S}, \qquad (10)\\
&T = \left[t^1 \cdots t^n\right] \in \mathbb{R}^{n\times n} : \text{a diagonal matrix with full column rank},\\
&\widetilde{T}_j = \left[t^{u_j+1} \cdots t^{n-1}\; t^n\right] \in \mathbb{R}^{n\times n-u_j} \text{where } u_j \text{ represents the index}\\
&\text{of the non-zero entry in } s^j, j \in \{1, \dots, m\},\\
&T_j \in \mathbb{R}^{n\times k_j} : \text{a subset of directions contained in } \widetilde{T}_j \text{ for all } j \in \{1, \dots, m\}. \qquad (11)
\end{aligned}$$

In the next theorem, the matrix $U \in \mathbb{R}^{n\times n}$ denotes a strictly upper triangular matrix such that

$$U_{i,j} = \begin{cases} \left[\nabla^2 f(x^0)\right]_{i,j}, & \text{if} \quad 1 \le i < j \le n,\\ 0, & \text{otherwise.}\end{cases}$$

Using a similar process to the one in Proposition 23, it can be shown that

$$\mathrm{Proj}_{S,T_{1:m}} \nabla^2 f(x^0) = \mathrm{Proj}_{S,T_{1:m}} U$$

and that the GSH (GCSH) is a strictly upper triangular matrix whenever the matrices of directions S and T_j are defined as in (10) and (11).

The following two error bounds follow from Theorem 10(ii) and Theorem 11(ii), respectively.

Corollary 25 (Error bound for the off-diagonal entries of the Hessian) *Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^4$ on $B_n(x^0;\overline{\Delta})$ where $x^0 \in \operatorname{dom} f$ is the point of interest and $\overline{\Delta} > 0$. Denote by $L_{\nabla^2 f} \geq 0$ and $L_{\nabla^3 f} \geq 0$ the Lipschitz constant of $\nabla^2 f$ and $\nabla^3 f$ on $\overline{B}_n(x^0;\overline{\Delta})$ respectively. Let $S = \begin{bmatrix} s^1 & s^2 & \cdots & s^m \end{bmatrix} \in \mathbb{R}^{n\times m}$ and $T_j \in \mathbb{R}^{n\times k_j}$ be defined as in (10) and (11) respectively. Assume that $B_n(x^0+s^j;\Delta_{T_j}) \subset B_n(x^0;\overline{\Delta})$ for all $j \in \{1,\ldots,m\}$. Then*

(i)

$$\begin{aligned}&\left\|\operatorname{Proj}_{S,T_{1:m}} \nabla^2_s f(x^0;S,T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} \nabla^2 f(x^0)\right\| \\ &\quad = \|\nabla^2_s f(x^0;S,T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} U\| \\ &\quad \leq 4m\sqrt{k} L_{\nabla^2 f}\left(\frac{\Delta_u}{\Delta_l}\right)^2 \|(\widehat{S}^\top)^\dagger\|\|\widehat{T}^\dagger\|\Delta_u,\end{aligned}$$

and

(ii)

$$\begin{aligned}&\left\|\operatorname{Proj}_{S,T_{1:m}} \nabla^2_c f(x^0;S,T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} \nabla^2 f(x^0)\right\| \\ &\quad = \left\|\nabla^2_c f(x^0;S,T_{1:m}) - \operatorname{Proj}_{S,T_{1:m}} U\right\| \\ &\quad \leq 2m\sqrt{k} L_{\nabla^3 f}\left(\frac{\Delta_u}{\Delta_l}\right)^2 \|(\widehat{S}^\top)^\dagger\|\|\widehat{T}^\dagger\|\Delta_u^2.\end{aligned}$$

A simple choice for S and T_j if **all** off-diagonal entries are of interest is to set

$$S = h\begin{bmatrix} e^1 & \cdots & e^{n-1}\end{bmatrix}, \tag{12}$$
$$T_j = h\begin{bmatrix} e^{j+1} & \cdots & e^{n}\end{bmatrix}, \quad \text{for all} \quad j \in \{1,\ldots,n-1\} \tag{13}$$

where $h \neq 0$. In this case, the GSH is an order-1 accurate approximation of all off-diagonal entries of the Hessian. To compute this GSH, the function must be evaluated at the points $x^0, x^0 \oplus S, x^0 \oplus T_j$ and $x^0 + s^j \oplus T_j$ for all $j \in \{1,\ldots n-1\}$. Hence, the number of distinct function evaluations is

$$1+(n-1)+(n-1)+\frac{(n-1)n}{2}-(n-2) = n+\frac{(n-1)n}{2} = \frac{n(n+1)+2}{2}.$$

In the previous equation, we subtracted $(n-2)$ since $x^0 \oplus h\begin{bmatrix} e^2 \cdots e^{n-1}\end{bmatrix}$ appears in $x^0 \oplus T_j$ and $x^0 \oplus S$.

Note that this number of function evaluations is smaller than $(n+1)(n+2)/2$ whenever $n \geq 1$, which is the number of function evaluations require to compute a GSH with a minimal poised set for the GSH. Therefore, if we are only interested in the off-diagonal entries of a Hessian, it is preferable to set the matrices S and T_j as described in this section, rather than using a minimal poised set for GSH.

In the previous corollary, Item (ii) shows that the GCSH is an order-2 accurate approximation of all off-diagonal entries of the Hessian. In this case, the sample points used are $x^0, x^0 \oplus (\pm S), x^0 \oplus (\pm T_j), x^0 \oplus S \oplus T_j$, and $x^0 \oplus (-S) \oplus (-T_j)$ for all $j \in \{1,\ldots n-1\}$. The number of distinct function evaluations is

$$1+2\left(\frac{n(n+1)}{2}\right) = n^2+n+1.$$

Notice that that this is the same amount of function evaluations utilized when using a minimal poised set for the GCSH (Definition 14). Therefore, there is no advantage in terms of function evaluations to choose S and T_j as described in (10) and (11) over a minimal poised set for the GCSH.

It does not seem to be possible to obtain an order-1 accurate approximation of all off-diagonal entries of a Hessian with fewer than $\frac{n(n+1)}{2}+1$ function evaluations, nor an order-2 accurate approximation with fewer than n^2+n+1 function evaluations. An obvious future research direction is to investigate this conjecture and mathematically prove or disprove it.

In the next section, we discuss how to approximate one row of the Hessian.

4.3 Approximating a row of the Hessian

In this section, we discuss how to approximate some or all entries of a row in the Hessian. Since the Hessian is symmetric, approximating a row also provides an approximation of the corresponding column.

Let $M \in \mathbb{R}^{n\times n}$. We denote by $R_i \in \mathbb{R}^{n\times n}$ the square matrix such that $R_i = \operatorname{Diag}(e^i)M$ for all $i \in \{1,\dots,n\}$. We begin by introducing the following lemma.

Lemma 26 *Let $M \in \mathbb{R}^{n\times n}$, $S = he^i \in \mathbb{R}^n$ where $h \neq 0$, and $\overline{T} \in \mathbb{R}^{n\times k}$. Define $R_i = \operatorname{Diag}(e^i)M$ for all $i \in \{1,\dots,n\}$. Then for all $i \in \{1,\dots,n\}$,*

$$\operatorname{Proj}_{S,\overline{T}} M = \operatorname{Proj}_{S,\overline{T}} R_i.$$

Proof We have

$$\begin{aligned}
\operatorname{Proj}_{S,\overline{T}} M &= ((he^i)^\top)^\dagger (he^i)^\top M\overline{T}\,\overline{T}^\dagger \\
&= e^i(e^i)^\top M\overline{T}\,\overline{T}^\dagger \\
&= R_i\overline{T}\,\overline{T}^\dagger \\
&= (e^i)(e^i)^\top R_i\overline{T}\,\overline{T}^\dagger \\
&= ((he^i)^\top)^\dagger (he^i)^\top R_i\overline{T}\,\overline{T}^\dagger = \operatorname{Proj}_{S,\overline{T}} R_i.
\end{aligned}$$

□

In words, the previous result says that the projection onto S and $\overline{T}$ of a matrix M is equal to the projection onto S and $\overline{T}$ of row i of this matrix whenever $S = he^i$.

When S and $\overline{T}$ are defined as in the previous proposition, S is full column rank and it follows from Proposition 9 (ii) that $\nabla_s^2(x^0, S, \overline{T}) = \operatorname{Diag}(e^i)\nabla_s^2(x^0, S, \overline{T})$ and $\nabla_c^2(x^0; S, \overline{T}) = \operatorname{Diag}(e^i)\nabla_c^2(x^0; S, \overline{T})$. Moreover, the projection of the Hessian is

$$\operatorname{Proj}_{S,\overline{T}} \nabla^2 f(x^0) = \operatorname{Proj}_{S,\overline{T}} \operatorname{Diag}(e^i)\nabla^2 f(x^0)$$

for all $i \in \{1,\dots,n\}$.

Next, we present two error bounds; one for the GSH and one for the GCSH. These error bounds follow immediately from Theorems 10 (iii) and 11 (iii) respectively.

Corollary 27 (General error bounds for one row of a Hessian) *Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^4$ on $B_n(x^0; \overline{\Delta})$ where $x^0 \in \operatorname{dom} f$ is the point of interest and $\overline{\Delta} > 0$. Denote by $L_{\nabla^2 f} \geq 0$ and $L_{\nabla^3 f} \geq 0$ the Lipschitz constant of $\nabla^2 f$ and $\nabla^3 f$ on $\overline{B}_n(x^0; \overline{\Delta})$ respectively. Let $S = he^i \in \mathbb{R}^n$ where $h \neq 0$, and $\overline{T} \in \mathbb{R}^{n\times k}$. Assume that $B_n(x^0 + he^i; \Delta_T) \subset B_n(x^0; \overline{\Delta})$. Then*

(i)

$$\left\|\operatorname{Proj}_{S,\overline{T}} \nabla_s^2 f(x^0; S, \overline{T}) - \operatorname{Proj}_{S,\overline{T}} \nabla^2 f(x^0)\right\| = \left\|\nabla_s^2 f(x^0; S, \overline{T}) - \operatorname{Proj}_{S,\overline{T}} \operatorname{Diag}(e^j)\nabla^2 f(x^0)\right\|$$
$$\leq 4\sqrt{k} L_{\nabla^2 f} \left(\frac{\Delta_u}{\Delta_l}\right) \|\widehat{\overline{T}}^\dagger\| \Delta_u,$$

and

(ii)

$$\left\|\operatorname{Proj}_{S,\overline{T}} \nabla_c^2(x^0; S, \overline{T}) - \operatorname{Proj}_{S,\overline{T}} \nabla^2 f(x^0)\right\| = \left\|\nabla_c^2(x^0; S, \overline{T}) - \operatorname{Proj}_{S,\overline{T}} \operatorname{Diag}(e^i)\nabla^2 f(x^0)\right\|$$
$$\leq 2\sqrt{k} L_{\nabla^3 f} \left(\frac{\Delta_u}{\Delta_l}\right) \|\widehat{\overline{T}}^\dagger\| \Delta_u^2.$$

Note that $\|(\widehat{S}^\top)^\dagger\|$ does not appear in the previous error bounds since $\|(\widehat{S}^\top)^\dagger\| = 1$.

One simple choice to approximate all entries of the i^{th} row is to choose

$$S = he^i, \ \overline{T} = h\mathrm{Id}_n$$

where $h \neq 0$. In this case, $\nabla_s^2 f(x^0; S, \overline{T})$ is an order-1 accurate approximation of the whole i^{th} row of the Hessian. This choice uses the set of sample points $x^0, x^0 + he^i, x^0 \oplus h\mathrm{Id}_n$ and $x^0 + he^i \oplus h\mathrm{Id}_n$. In this case, the number of function evaluations is

$$1 + 1 + n + n - 1 = 2n + 1.$$

We subtract one in the previous equation since one point is reused: $x^0 + he^i$. Note that $2n + 1 \leq (n + 1)(n + 2)/2$ for all $n \in \{1, 2, \dots\}$. Therefore, if we are only interested by the entries of row i, setting S and $\overline{T}$ in this fashion saves function evaluations compared to using a minimal poised set for the GSH.

To obtain an order-2 accurate approximation of the whole i^{th} row of the Hessian, we may choose once again

$$S = he^i, \overline{T} = h\mathrm{Id}_n,$$

where $h \neq 0$. In this case, the set of sample points is $x^0, x^0 \pm he^i, x^0 \oplus (\pm h\mathrm{Id}_n), x^0 + he^i \oplus h\mathrm{Id}$, and $x^0 - he^i \oplus -\mathrm{Id}_n$. Two sample points are reused: $x^0 \pm he^i$. The number of function evaluations is

$$1 + 2(2n) = 4n + 1.$$

Note that $4n + 1 < n^2 + n + 1$ when $n \geq 4$. Therefore, if $n \in \{1, 2, 3\}$, then using a minimal poised set for GCSH is preferable since it uses fewer function evaluations.

It seems that the minimum number of function evaluations to obtain an order-1 accurate approximation of a full row (column) in a Hessian is $2n + 1$. To obtain an order-2 accurate approximation of a full row, the minimum number seems to be $n^2 + n + 1$ when $n \in \{1, 2, 3\}$ and $4n + 1$ when $n \geq 4$. Future research could focus on mathematically proving or disproving this claim.

5 Approximating order-P derivatives

Now that we have a general method to approximate first-order derivatives called the generalized simplex gradient and a general method to approximate second-order derivatives called the generalized simplex Hessian, we may develop a general method to approximate P-order derivatives. The object containing all P-order derivatives can be viewed as a P-dimensional matrix. We begin by providing a formula to approximate all third-order derivatives and then we propose a formula to compute P-order derivatives.

We refer to $\nabla^3 f(x^0)$ as the *Tressian* of f at x^0. A Tressian can be viewed as a three-dimensional matrix $\mathbf{M}$ in $\mathbb{R}^{n\times n\times n}$ where the third dimension represents the depth of $\mathbf{M}$. In this section, a three-dimensional matrix $\mathbf{M} \in \mathbb{R}^{r\times c\times p}$ will be either thought as an object containing r floors where each floor is a matrix in $\mathbb{R}^{c\times p}$ or as an object containing p layers where each layer is a matrix in $\mathbb{R}^{r\times c}$. A three-dimensional matrix $\mathbf{M} \in \mathbb{R}^{r\times c\times p}$ is written by "floor" in the following way:

$$\mathbf{M} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_r \end{bmatrix}_{(r,\cdot,\cdot)} \tag{14}$$

where $F_i \in \mathbb{R}^{c\times p}$ for all $i \in \{1, 2, \ldots, r\}$ and $[M_i]_{j,k} = \mathbf{M}_{i,j,k}$ for all i, j, k. The subscript in (14) is used to make it clear that $\mathbf{M}$ is written by floor. The matrix $\mathbf{M}$ can also be written in terms of layers:

$$\mathbf{M} = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_p \end{bmatrix}_{(\cdot,\cdot,p)}$$

where $[L_k]_{i,j} = \mathbf{M}_{i,j,k}$ for all i, j, k.

We are now ready to introduce the formula to approximate $\nabla^3 f(x^0)$. The technique requires one more set of matrices of directions than the generalized simplex Hessian. The letter U is used to denote this new of set of matrices. This set of matrices associated could contain $k_1+k_2+\cdots+k_m$ different matrices. To keep things relatively simple, we provide the formula for the case where all matrices T_j, are equal and all matrices $U_k \in \mathbb{R}^{n\times \ell_k}$ are equal. To emphasize this special case where all matrices U_k are equal, we use the notation $\overline{U} \in \mathbb{R}^{n\times \ell}$. Hence, the three matrices of directions involved in the computation of the approximation technique are $S \in \mathbb{R}^{n\times m}, \overline{T} \in \mathbb{R}^{n\times k}$, and $\overline{U} \in \mathbb{R}^{n\times \ell}$.

Before introducing the approximation technique, we define the multiplication of a two-dimensional matrix with a three-dimensional matrix.

Let $A \in \mathbb{R}^{n\times m}$ and let $\mathbf{M} \in \mathbb{R}^{m\times n\times p}$ and let $\mathbf{M}$ be written as layers:

$$\mathbf{M} = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_p \end{bmatrix}_{(\cdot,\cdot,p)} \in \mathbb{R}^{m\times n\times p},$$

where $[L_k]_{i,j} = \mathbf{M}_{i,j,k}$. Then

$$A \otimes \mathbf{M} = \begin{bmatrix} AL_1 \\ AL_2 \\ \vdots \\ AL_p \end{bmatrix}_{(\cdot,\cdot,p)} \in \mathbb{R}^{n\times n\times p}$$

Definition 28 (Generalized simplex Tressian) Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $x^0 \in \operatorname{dom} f$ be the point of interest. Let $S = \begin{bmatrix} s^1 & s^2 & \cdots & s^m \end{bmatrix} \in \mathbb{R}^{n\times m}$ and $\overline{T} \in \mathbb{R}^{n\times k}, \overline{U} \in \mathbb{R}^{n\times \ell}$ with the set of sample points $\mathcal{S}(x^0; S, \overline{T}, \overline{U})$ contained in $\operatorname{dom} f$. The *generalized simplex Tressian* of f at x^0 over $S, \overline{T}$ and $\overline{\mathrm{U}}$ is denoted by $\nabla^3_s f(x^0; S, \overline{T}, \overline{\mathrm{U}})$ and defined by

$$\nabla^3_s f(x^0; S, \overline{T}, \overline{\mathrm{U}}) = (S^\top)^\dagger \otimes \delta^{\mathbf{3}}_{\mathbf{s}} f(x^0; S, \overline{T}, \overline{\mathrm{U}}) \in \mathbb{R}^{n\times n\times n},$$

where

$$\delta^{\mathbf{3}}_{\mathbf{s}} f(x^0; S, \overline{T}, \overline{\mathrm{U}}) = \begin{bmatrix} (\nabla^2_s f(x^0 + s^1; \overline{T}, \overline{\mathrm{U}}) - \nabla^2_s f(x^0; \overline{T}, \overline{\mathrm{U}}))^\top \\ (\nabla^2_s f(x^0 + s^2; \overline{T}, \overline{\mathrm{U}}) - \nabla^2_s f(x^0; \overline{T}, \overline{\mathrm{U}}))^\top \\ \vdots \\ (\nabla^2_s f(x^0 + s^m; \overline{T}, \overline{\mathrm{U}}) - \nabla_s f(x^0; \overline{T}, \overline{\mathrm{U}}))^\top \end{bmatrix}_{(m,\cdot,\cdot)} \in \mathbb{R}^{m\times n\times n}.$$

Recursively, we may now define a simple formula to approximate order-P derivatives of a function at a point of interest $x^0 \in \mathbb{R}^n$. Before introducing the formula, notation needs to be slightly modified to make it easier to discuss general order-P derivatives. To approximate order-P derivatives, we use a matrix $S_1 \in \mathbb{R}^{n\times m_1}$, and set of matrices $S_2, S_3, \ldots, S_P$. To keep notation relatively simple, we consider the case where all matrices of directions are the same in the sets $S_2, \ldots, S_P$. As before, we write $\overline{S_i}$ to emphasize that all matrices of directions are identical in each set $\overline{S_i}, i \in \{2, \ldots, P\}$. A matrix in the set $\overline{S_i}$ has dimensions $n \times m_i$, for $i \in \{2, 3, \ldots, P\}$.

The transpose of a P-dimensional matrix $\mathbf{M} \in \mathbb{R}^{n\times m_1\times\cdots\times m_{P-1}}$ is denoted by $\mathbf{M}^\top$ where the entries of $\mathbf{M}^\top$ are equal to

$$[\mathbf{M}^\top]_{i,j_1,\ldots,j_{P-1}} = \mathbf{M}_{j_{P-1},j_{P-2},\ldots,j_1,i}, \quad i \in \{1, \ldots, n\}, j_k \in \{1, 2, \ldots, m_k\}, k \in \{1, 2, \ldots, P-1\}.$$

Definition 29 (Order-P simplex derivative matrix) Let $f : \operatorname{dom} f \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $x^0 \in \operatorname{dom} f$ be the point of interest. Let $S_1 = \begin{bmatrix} s^1 & s^2 & \cdots & s^m \end{bmatrix} \in \mathbb{R}^{n\times m_1}$ and $\overline{S_i} \in \mathbb{R}^{n\times m_i}$ for all $i \in \{2, 3, \ldots, P\}$ with the set of sample points $\mathcal{S}(x^0; S_1, \overline{S_2}, \ldots, \overline{S_P})$ contained in $\operatorname{dom} f$. The *order-P simplex derivative tensor of f at x^0 over* $S_1, \overline{S_2}, \ldots, \overline{S_P}$ is denoted by $\nabla^P_s f(x^0; S_1, \overline{S_2}, \ldots, \overline{S_P})$ and defined by

$$\nabla^P_s f(x^0; S_1, \overline{S_2}, \ldots, \overline{S_P}) = (S_1^\top)^\dagger \otimes \delta^P_s f(x^0; S_1, \overline{S_2}, \ldots, \overline{S_P}) \in \mathbb{R}^{n\times n\times\cdots\times n},$$

where

$$\delta^P_s f(x^0; S_1, \overline{S_2}, \ldots, \overline{S_P}) = \begin{bmatrix} (\nabla^{P-1}_s f(x^0 + s^1; \overline{S_2}, \ldots, \overline{S_P}) - \nabla^{P-1}_s f(x^0; \overline{S_2}, \ldots, \overline{S_P}))^\top \\ (\nabla^{P-1}_s f(x^0 + s^2; \overline{S_2}, \ldots, \overline{S_P}) - \nabla^{P-1}_s f(x^0; \overline{S_2}, \ldots, \overline{S_P}))^\top \\ \vdots \\ (\nabla^{P-1}_s f(x^0 + s^m; \overline{S_2}, \ldots, \overline{S_P}) - \nabla^{P-1}_s f(x^0; \overline{S_2}, \ldots, \overline{S_P}))^\top \end{bmatrix}_{(m,\cdot,\ldots,\cdot)}.$$

6 Conclusion

We have defined minimal poised set for the GCSH, which demonstrates how to obtain an order-2 accurate approximation of the full Hessian with $n^2 + n + 1$ distinct sample points. A future research direction is to investigate if there exist choices other than $\overline{T} = -S$ that make the set of sample points a minimal poised set for the GCSH. Furthermore, proving that $n^2 + n + 1$ is the minimal number of function evaluations to obtain an order-2 accurate approximation of the full Hessian is an obvious future research direction. It remains to verify if an order-1 accurate approximation of the main diagonal of a Tressian can be obtained for free in terms of function evaluations, if an order-2 accurate approximation of the Hessian has been previously computed with the GCSH.

In Section 4, we provided details on how to choose the matrices S and T_j when we are only interested in a proper subset of the entries of the Hessian. In particular, we investigated how to approximate the diagonal entries of a Hessian, the off-diagonal entries of a Hessian, and a row of a Hessian. The number of function evaluations to obtain an order-1 accurate approximation, or an order-2 accurate approximation of the entries of the Hessian of interest has been discussed. The relation between the CSHD introduced in [16] and the GCSH is clarified. It is shown that the CSHD is equal to the GCSH whenever S is a partial diagonal matrix with full column rank and $T_j = -s^j$ for all j (Theorem 20).

In Section 5, the approximation technique is generalized to higher-order derivatives than two. First, it is discussed how to obtain an approximation of the third-order derivatives. Then a simple recursive formula is introduced to computer order-P derivatives of a function at a point of interest. On a final note, an implementation in MATLAB of each approximation technique discussed in this paper is available upon request.

References

1. N. Andrei, *A Derivative-Free Two-Level Random Search Method for Unconstrained Optimization*, Springer, 2021.
2. ———, *Modern Numerical Nonlinear Optimization*, vol. 195, Springer Nature, 2022.
3. R. Burden, J. Faires, and A. Burden, *Numerical Analysis 10/e IE*, Brooks/Cole Cengage Learning, 2016.
4. T. Coleman and J. Moré, *Estimation of sparse Hessian matrices and graph coloring problems*, Mathematical programming, 28 (1984), pp. 243–270.
5. A. Conn, K. Scheinberg, and L. Vicente, *Geometry of sample sets in derivative-free optimization: polynomial regression and underdetermined interpolation*, IMA Journal of Numerical Analysis, 28 (2008), pp. 721–748.
6. ———, *Introduction to Derivative-Free Optimization*, SIAM, 2009.
7. I. Coope and R. Tappenden, *Gradient and diagonal Hessian approximations using quadratic interpolation models and aligned regular bases*, Numerical Algorithms, 88 (2021), pp. 767–791.
8. A. Custódio and L. Vicente, *Using sampling and simplex derivatives in pattern search methods*, SIAM Journal on Optimization, 18 (2007), pp. 537–555.
9. A. Custódio, H. Rocha, and L. Vicente, *Incorporating minimum Frobenius norm models in direct search*, Computational Optimization and Applications, 46 (2010), pp. 265–278.
10. A. Gebremedhin, F. Manne, and A. Pothen, *What color is your Jacobian? graph coloring for computing derivatives*, SIAM review, 47 (2005), pp. 629–705.
11. A. Griewank and G. Corliss, *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, Society for Industrial and Applied Mathematics, Philadelphia, 1991.
12. W. Hare and G. Jarry-Bolduc, *Calculus identities for generalized simplex gradients: Rules and applications*, SIAM Journal on Optimization, 30 (2020), pp. 853–884.
13. W. Hare, G. Jarry-Bolduc, and C. Planiden, *A matrix algebra approach to approximate Hessians*, IMA Journal of Numerical Analysis, (2023).
14. W. Hare, G. Jarry–Bolduc, and C. Planiden, *Error bounds for overdetermined and underdetermined generalized centred simplex gradients*, IMA Journal of Numerical Analysis, 42 (2020), pp. 744–770.
15. R. Horn and C. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.
16. G. Jarry-Bolduc, *Approximating the diagonal of a Hessian: which sample set of points should be used*, Numerical Algorithms, 91 (2022), pp. 1349–1361.
17. G. Jarry-Bolduc, *Numerical analysis for derivative-free optimization*, PhD thesis, University of British Columbia, 2023.
18. C. Kelley, *Implicit Filtering*, vol. 23, SIAM, 2011.
19. M. Powell, *The use of band matrices for second derivative approximations in trust region algorithms*, in Advances in Nonlinear Programming: Proceedings of the 96 International Conference on Nonlinear Programming, Springer, 1998, pp. 3–28.
20. ———, *Least Frobenius norm updating of quadratic models that satisfy interpolation conditions*, Mathematical Programming, 100 (2004), pp. 183–215.
21. ———, *On the use of quadratic models in unconstrained minimization without derivatives*, Optimization Methods and Software, 19 (2004), pp. 399–411.
22. ———, *On updating the inverse of a KKT matrix*, Numerical Linear Algebra and Optimization, (2004), pp. 56–78.
23. ———, *The NEWUOA software for unconstrained optimization without derivatives*, in Large-scale nonlinear optimization, Springer, 2006, pp. 255–297.
24. ———, *A view of algorithms for optimization without derivatives*, Mathematics Today-Bulletin of the Institute of Mathematics and its Applications, 43 (2007), pp. 170–174.
25. ———, *Developments of NEWUOA for minimization without derivatives*, IMA Journal of Numerical Analysis, 28 (2008), pp. 649–664.
26. R. Regis, *The calculus of simplex gradients*, Optimization Letters, 9 (2015), pp. 845–865.
27. R. Rockafellar and R. Wets, *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1998.
28. S. Roman, *Advanced Linear Algebra*, vol. 135, Springer, New York, 3rd ed., 2007.
29. S. Wild, R. Regis, and C. Shoemaker, *ORBIT: optimization by radial basis function interpolation in trust-regions*, SIAM Journal on Scientific Computing, 30 (2008), pp. 3197–3219.
30. D. Winfield, *Function and Functional Optimization by Interpolation in Data Tables*, PhD thesis, Harvard University, 1970.