

Accuracy and Relationships of Quadratic Models in Derivative-free Optimization

Yiwen Chen · Warren Hare · Lindon Roberts

May 12, 2026

Abstract We study three quadratic models in model-based derivative-free optimization: the minimum norm (MN), minimum Frobenius norm (MFN), and quadratic generalized simplex derivative (QS) models. Despite their widespread use, their approximation accuracy and relationships have not been systematically explored. We establish fully linear error bounds for all three models, removing the uniformly bounded model Hessian assumption required in existing MN analyses and deriving the first such results for the QS model. We further analyze Hessian approximation accuracy via directional error bounds, showing that all three models achieve fully quadratic accuracy along sample directions under a mild condition on the sample set. This reveals a form of directional fully quadratic accuracy not captured by existing theory. Finally, we characterize the relationships among these models, identifying conditions under which they coincide and clarifying their structural connections.

Keywords: derivative-free optimization; quadratic models; error bounds

MSC codes: primary 90C56; secondary 65G99, 65K05.

Yiwen Chen

Department of Mathematics, University of British Columbia, Kelowna, British Columbia, V1V 1V7, Canada. Chen's research is partially funded by the Natural Sciences and Engineering Research Council of Canada (les recherches de Chen sont partiellement financées par le Conseil de recherches en sciences naturelles et en génie du Canada), Discover Grant #2023-03555. Chen was also supported by the Australian Research Council Discovery Early Career Award DE240100006.
E-mail: yiwchen@student.ubc.ca

Warren Hare

Department of Mathematics, University of British Columbia, Kelowna, British Columbia, V1V 1V7, Canada. Hare's research is partially funded by the Natural Sciences and Engineering Research Council of Canada (les recherches de Hare sont partiellement financées par le Conseil de recherches en sciences naturelles et en génie du Canada), Discover Grant #2023-03555.
E-mail: warren.hare@ubc.ca

Lindon Roberts

School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria 3010, Australia. Roberts was supported by the Australian Research Council Discovery Early Career Award DE240100006.
E-mail: lindon.roberts@unimelb.edu.au

1 Introduction

Derivative-free optimization (DFO) methods are an essential tool for solving optimization problems where the derivatives of the objective and/or constraint functions are unavailable, unreliable, or prohibitively expensive to compute [1, 11]. One major class of DFO methods, known as model-based DFO methods, constructs surrogate models using only function evaluations. The approximation accuracy of these models depends critically on the geometry of the set of sampled points, which we refer to as a *sample set*.

Convergence analyses of model-based DFO methods typically rely on model accuracy assumptions, most notably that the models are *fully linear* or *fully quadratic* [2, 19]. A common approach for constructing such models is quadratic interpolation. In this paper, we study three quadratic models: the minimum norm (MN) model [12], the minimum Frobenius norm (MFN) model [9], and the quadratic generalized simplex derivative (QS) model [5]. The QS model is constructed using generalized simplex gradients [10, 13, 14, 18] and generalized simplex Hessians [16]. Although these models have appeared in various forms in the literature, a unified theoretical comparison of their approximation accuracy and structural relationships remains lacking.

Our first contribution is the derivation of fully linear error bounds for all three models. Existing results for the MN model rely on a uniformly bounded model Hessian, while those for the MFN model are established under different settings or sample set geometry measures. We remove the bounded Hessian assumption for the MN model and establish the *first* fully linear error bounds for the QS model, thereby providing a unified perspective on their approximation accuracy.

We next study the accuracy of the Hessian approximations arising from these models through directional error bounds. Under a mild structural assumption on the sample set, we show that all three models achieve fully quadratic accuracy along sample directions. More precisely, the directional error ranges from $\mathcal{O}(1)$ to $\mathcal{O}(\Delta_{D^Y})$, where Δ_{D^Y} is the maximum length of the sample directions. This reveals that these models can significantly exceed the accuracy of standard fully linear models in a directional sense. Similar phenomena have been observed in [3], where the sharp error bound of linear interpolation/extrapolation depends on the location of the point of interest. These findings provide new insights into how second-order information can be captured without requiring fully determined quadratic interpolation.

Finally, we analyze the relationships among the three models. We identify conditions under which they coincide and clarify the structural connections between their constructions. These results not only enhance the theoretical understanding of these models but also suggest simpler and potentially more efficient construction procedures.

In summary, this paper advances the theoretical foundations of model-based DFO by establishing new fully linear error bounds, deriving directional Hessian error bounds, and clarifying the relationships among several widely used quadratic models. These results have implications for both the analysis and practical implementation of DFO methods.

The remainder of the paper is organized as follows. Section 2 introduces notation and background definitions. Section 3 establishes fully linear error bounds. Section 4 presents directional Hessian error bounds. Section 5 analyzes the relationships among the three models. Section 6 concludes this paper and discusses future directions.

2 Notation and Background

In this section, we introduce the notation and background definitions used throughout the paper. We begin by presenting the notation, followed by the formal definitions, and defer all examples to Subsection 2.4.

2.1 Notation

We use $\mathbf{0}$ and $\mathbf{1}$ to denote the all-zero and all-one vectors with appropriate dimensions. We use e^i to denote the i -th coordinate vector with appropriate dimensions. We use I_n to denote the identity matrix of dimension $n \times n$. We use $\mathbf{0}_{m \times n}$ to denote the all-zero matrix of dimension $m \times n$. We use $S_n(\mathbb{R})$ to denote the set of all $n \times n$ real symmetric matrices. For any matrix A , we use A_{ij} to denote its (i, j) -th element and $\text{col}(A)$ to denote its column space. We use $a \in A$ to mean that a is a column of A and define $\Delta_A := \max_{a \in A} \|a\|$. For a non-zero matrix A , we define $\bar{A} := A/\Delta_A$. For a square matrix A , we use $\text{tr}(A)$ to denote its trace and $\text{diag}(A)$ to denote the vector consisting of its diagonal elements. For a vector $v \in \mathbb{R}^n$, we use $\text{Diag}(v)$ to denote the $n \times n$ diagonal matrix whose diagonal elements are given by the components of v . We use $\|\cdot\|$ to denote the Euclidean norm of a vector and the spectral norm of a matrix. We use $\|\cdot\|_F$ to denote the Frobenius norm of a matrix. We use $B(x; \Delta)$ to denote the closed Euclidean ball centered at x with radius Δ . We use $[n]$ to denote the set $\{1, \dots, n\}$. For any set Y written in the form $Y = \{x^0, x^0 + d^i : i \in [m]\}$, we define $D^Y := [d^1 \dots d^m]$.

2.2 Generalized simplex gradient, generalized simplex Hessian, and quadratic generalized simplex derivative models

Definition 1 Let $x^0 \in \mathbb{R}^n$ and $S = [s^1 \dots s^p] \in \mathbb{R}^{n \times p}$. The generalized simplex gradient (GSG) of f at x^0 over S is

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

where the i -th element of $\delta_f(x^0; S)$ is $f(x^0 + s^i) - f(x^0)$, $i \in [p]$.

The sample set of the GSG of f at x^0 over S is $Y = \{x^0, x^0 + s^i : i \in [p]\}$.

Definition 2 Let $x^0 \in \mathbb{R}^n$ and $S = [s^1 \dots s^p] \in \mathbb{R}^{n \times p}$. Let matrices $T_i = [t_i^1 \dots t_i^{q_i}] \in \mathbb{R}^{n \times q_i}$, $i \in [p]$. Denote $T_{1:p} = \{T_1, \dots, T_p\}$. The generalized simplex Hessian (GSH) of f at x^0 over S and $T_{1:p}$ is

$$\nabla_s^2 f(x^0; S; T_{1:p}) := (S^\top)^\dagger \delta_f^2(x^0; S; T_{1:p}) \in \mathbb{R}^{n \times n}$$

where the i -th row of $\delta_f^2(x^0; S; T_{1:p})$ is $(\nabla_s f(x^0 + s^i; T_i) - \nabla_s f(x^0; T_i))^\top$, $i \in [p]$.

Remark 3 Following [16], if $T_1 = \dots = T_p = T \in \mathbb{R}^{n \times q}$, then, with a little misuse of notation, we denote $T_{1:p} = T$. In this case, we also denote by $\delta_{\delta_f}(x^0; S; T)$ the matrix with the i -th row given by $(\delta_f(x^0 + s^i; T) - \delta_f(x^0; T))^\top$. Equivalently, for $i \in [p]$ and $j \in [q]$,

$$(\delta_{\delta_f}(x^0; S; T))_{ij} := f(x^0 + s^i + t^j) - f(x^0 + s^i) - f(x^0 + t^j) + f(x^0).$$

Notice that $\nabla_s^2 f(x^0; S; T) = (S^\top)^\dagger \delta_{\delta_f}(x^0; S; T) T^\top$.

Remark 4 As we will see in Example 12, the GSH is not necessarily symmetric. In general, we have $\nabla_s^2 f(x^0; S; T)^\top = \nabla_s^2 f(x^0; T; S)$ [16, Prop. 4.4].

The sample set of the GSH of f at x^0 over S and $T_{1:p}$ is

$$Y = \left\{ x^0, x^0 + s^i, x^0 + t_i^j, x^0 + s^i + t_i^j : i \in [p], j \in [q_i] \right\}. \quad (1)$$

There is no unique way to construct quadratic models using the GSG and GSH. In this paper, we use the definition given in [5, Def. 7], which defines a general class of such models called the class of quadratic generalized simplex derivative models.

Definition 5 Let $x^0 \in \mathbb{R}^n$, Y be given by (1), and $d = x - x^0$. The class of quadratic generalized simplex derivative models of f at x^0 over Y is

$$\left\{ m_{\text{QS}}^{x^0, Y}(x) = f(x^0) + \nabla_{\text{QS}} f(x^0; Y)^\top d + \frac{1}{2} d^\top \nabla_{\text{QS}}^2 f(x^0; Y) d \right\}$$

where $\nabla_{\text{QS}} f(x^0; Y)$ and $\nabla_{\text{QS}}^2 f(x^0; Y)$ are linear combinations of GSG and GSH constructed using points in Y , chosen such that the union of the points used in both constructions equals Y .

Remark 6 As mentioned in Remark 4, the GSH need not be symmetric, so the QS model is not necessarily an interpolation model. One case where it is an interpolation model is given in Example 44 and analyzed in [16, Sec. 5]. Further details and additional cases are discussed in Section 5.

2.3 Minimum norm models and minimum Frobenius norm models

Two common choices of quadratic interpolation models are the minimum norm models [12] and the minimum Frobenius norm models [9]. Here, we adopt the definition given in [5, Defs. 2 and 3].

Definition 7 Let $x^0 \in \mathbb{R}^n$, $Y = \{x^0, x^0 + d^i : i \in [m]\}$, and $d = x - x^0$. The class of minimum norm models of f at x^0 over Y is

$$\left\{ m_{\text{MN}}^{x^0, Y}(x) = f(x^0) + \nabla_{\text{MN}} f(x^0; Y)^\top d + \frac{1}{2} d^\top \nabla_{\text{MN}}^2 f(x^0; Y) d \right\}$$

where $(\nabla_{\text{MN}} f(x^0; Y), \nabla_{\text{MN}}^2 f(x^0; Y)) \in \mathcal{M}_{\text{MN}}^{x^0, Y}$, the set of minimizers of

$$\begin{aligned} \min_{\substack{\alpha \in \mathbb{R}^n \\ H \in \mathcal{S}_n(\mathbb{R})}} \quad & \frac{1}{2} \|\alpha\|^2 + \frac{1}{2} \|H\|_F^2 \\ \text{s.t.} \quad & (d^i)^\top \alpha + \frac{1}{2} (d^i)^\top H (d^i) = f(x^0 + d^i) - f(x^0), \quad i \in [m]. \end{aligned} \tag{MN}$$

We say Y is poised for MN interpolation if the solution to Problem (MN) exists and is unique. Notice that the objective function is strongly convex and all constraints are linear with respect to α and H , the solution to Problem (MN) exists and is unique if and only if Problem (MN) is feasible, i.e., there exists a quadratic function that interpolates f on Y .

Definition 8 Let $x^0 \in \mathbb{R}^n$, $Y = \{x^0, x^0 + d^i : i \in [m]\}$, and $d = x - x^0$. The class of minimum Frobenius norm models of f at x^0 over Y is

$$\left\{ m_{\text{MFN}}^{x^0, Y}(x) = f(x^0) + \nabla_{\text{MFN}} f(x^0; Y)^\top d + \frac{1}{2} d^\top \nabla_{\text{MFN}}^2 f(x^0; Y) d \right\}$$

where $(\nabla_{\text{MFN}} f(x^0; Y), \nabla_{\text{MFN}}^2 f(x^0; Y)) \in \mathcal{M}_{\text{MFN}}^{x^0, Y}$, the set of minimizers of

$$\begin{aligned} \min_{\substack{\alpha \in \mathbb{R}^n \\ H \in \mathcal{S}_n(\mathbb{R})}} \quad & \frac{1}{2} \|H\|_F^2 \\ \text{s.t.} \quad & (d^i)^\top \alpha + \frac{1}{2} (d^i)^\top H (d^i) = f(x^0 + d^i) - f(x^0), \quad i \in [m]. \end{aligned} \tag{MFN}$$

We say Y is poised for MFN interpolation if the solution to Problem (MFN) exists and is unique. An equivalent condition for the existence and uniqueness of Problem (MFN) can be found in [11, Sec. 5.3]. As mentioned in [11, Sec. 5.3], if Y is poised for MFN interpolation, then it is poised for MN interpolation.

2.4 Examples

In this subsection, we present examples of the MN, MFN, and QS models. We begin with examples of the MN and MFN models. The following example shows that not all sets poised for MN interpolation are poised for MFN interpolation.

Example 9 Consider $n = 3$, $x^0 = \mathbf{0}$, $Y = \{\mathbf{0}, e^1, e^2, -e^1, -e^2\} \subseteq \mathbb{R}^3$, and $f(x) = \|x\|^2$. We can compute that

$$\mathcal{M}_{\text{MN}}^{x^0, Y} = \left\{ \left(\mathbf{0}, \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \right\}, \quad \mathcal{M}_{\text{MFN}}^{x^0, Y} = \left\{ \left(\begin{bmatrix} 0 \\ 0 \\ r \end{bmatrix}, \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) : r \in \mathbb{R} \right\}.$$

This implies that Y is poised for MN interpolation but not poised for MFN interpolation, as the solution to Problem (MFN) is not unique.

We also notice that even if Y is poised for both MN and MFN interpolation, the MN and MFN models may not be the same.

Example 10 Consider $n = 2$, $x^0 = \mathbf{0}$, $Y = \{\mathbf{0}, e^1, e^2, 2e^1, e^1 + e^2\} \subseteq \mathbb{R}^2$, and $f(x) = \|x\|^2$. We can compute that

$$\mathcal{M}_{\text{MN}}^{x^0, Y} = \left\{ \left(\begin{bmatrix} 0 \\ \frac{4}{5} \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & \frac{2}{5} \end{bmatrix} \right) \right\} \quad \text{and} \quad \mathcal{M}_{\text{MFN}}^{x^0, Y} = \left\{ \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \right) \right\}.$$

Next, we provide an example of the QS model illustrating that it may or may not coincide with the MN and MFN models.

Example 11 We use the same setting as Example 9. One possible QS model can be constructed by letting

$$\begin{aligned} \nabla_{\text{QS}} f(x^0; Y) &= \frac{1}{2} (\nabla_s f(x^0; [e^1 \ e^2]) + \nabla_s f(x^0; [-e^1 \ -e^2])) = \mathbf{0} \\ \nabla_{\text{QS}}^2 f(x^0; Y) &= \nabla_s^2 f(x^0; [e^1 \ e^2]; \{[-e^1], [-e^2]\}) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

As seen in Example 9, this QS model coincides with the MN model and belongs to the class of MFN models of f at x^0 over Y . However, if we change $\nabla_{\text{QS}} f(x^0; Y)$ to $\nabla_s f(x^0; [e^1]) = e^1$, then the corresponding QS model no longer coincides with MN or MFN models.

Finally, we note that, unlike the MN and MFN models, the Hessian of the QS model need not be symmetric. This asymmetry arises because the GSH itself may be non-symmetric, as illustrated in the following example.

Example 12 Consider $n = 2$, $x^0 = \mathbf{0}$, $S = [e^1 \ e^2]$, $T = [e^1]$, and $f(x) = (\mathbf{1}^\top x)^2$. Then, the GSH of f at x^0 over S and T is

$$\nabla_s^2 f(x^0; S; T) = \begin{bmatrix} f(2e^1) - f(e^1) - f(e^1) + f(\mathbf{0}) \\ f(e^1 + e^2) - f(e^2) - f(e^1) + f(\mathbf{0}) \end{bmatrix} (e^1)^\top = \begin{bmatrix} 2 & 0 \\ 2 & 0 \end{bmatrix} \notin S_2(\mathbb{R}).$$

The examples in this subsection illustrate both distinctions and potential connections among the MN, MFN, and QS models. In this paper, we establish error bounds and further investigate the relationships among these models.

3 Fully linear error bounds

In this section, we present and derive fully linear error bounds for the MN, MFN, and QS models. We note that fully linear bounds of the MN and MFN models have previously been studied in [11, Chap. 5], and the errors of the MN model projected onto an appropriate linear subspace are given in [10, Thm. 5.12]. However, the fully linear bounds for the MN model provided in [11, Chap. 5] rely on the assumption that the model Hessians are uniformly bounded. Our analysis demonstrates that this assumption is unnecessary. Moreover, we establish the first fully linear error bounds for the QS model.

Following [1, Def. 9.1] and [11, Def. 6.1], we first define a class of fully linear models.

Definition 13 Suppose that $f \in \mathcal{C}^1$, $x^0 \in \mathbb{R}^n$, and $\bar{\Delta} > 0$. We say that $\mathcal{M}_{\bar{\Delta}} = \{m_{\Delta} : \mathbb{R}^n \rightarrow \mathbb{R}\}_{\Delta \in (0, \bar{\Delta}]}$ is a class of fully linear models of f at x^0 parameterized by Δ if there exists constants $\kappa^{ef} > 0$ and $\kappa^{eg} > 0$ independent of x^0 such that for all $\Delta \in (0, \bar{\Delta}]$ and $x \in B(x^0; \Delta)$,

$$|f(x) - m_{\Delta}(x)| \leq \kappa^{ef} \Delta^2 \quad \text{and} \quad \|\nabla f(x) - \nabla m_{\Delta}(x)\| \leq \kappa^{eg} \Delta.$$

In the following results, we may use the following assumptions.

Assumption 14 *The sample set $Y = \{x^0, x^0 + d^i : i \in [m]\} \subseteq \mathbb{R}^n$ satisfies that D^Y has full row rank, i.e., $\text{col}(D^Y) = \mathbb{R}^n$.*

Assumption 15 *The objective $f \in \mathcal{C}^{1+}$ with $L_{\nabla f}$ -Lipschitz gradient.*

We note that Assumption 14 is imposed primarily to simplify the notation and presentation of the analysis. By applying the conversion framework developed in [5], all results stated under Assumption 14 extend to the case where the columns of D^Y span a linear subspace. In that setting, the analysis can be carried out in intrinsic coordinates on the subspace, after which the resulting bounds for the subspace model can be translated back to the full-space model on Y via the conversion formula in [5].

Assumption 15 is standard in error analysis for model-based DFO, as it provides a bound on the error of the first-order Taylor approximation.

Lemma 16 (*[1, Lem. 9.4]*) *Suppose that Assumption 15 holds. Then, for all $x, d \in \mathbb{R}^n$, we have $|f(x+d) - f(x) - \nabla f(x)^\top d| \leq \frac{L_{\nabla f}}{2} \|d\|^2$.*

The next theorem, analogous to [11, Thm. 5.4], provides a fully linear bound for general quadratic interpolation models.

Theorem 17 *Suppose that Assumptions 14 and 15 hold. Let $m^{x^0, Y}(x)$ be any quadratic interpolation model of f at x^0 over Y . If $\|\nabla^2 m^{x^0, Y}(x)\| \leq \kappa^{mH}$ for some $\kappa^{mH} > 0$ on a set $K \subseteq \mathbb{R}^n$, then for all $x^0 \in K$, $m^{x^0, Y}(x)$ belongs to a class of fully linear models of f at x^0 parameterized by Δ_{D^Y} , with*

$$\kappa^{ef} := \frac{L_{\nabla f} + \kappa^{mH}}{2} \sqrt{n} \left\| \overline{D^Y}^\dagger \right\|_1 + \frac{L_{\nabla f} + \kappa^{mH}}{2}, \quad \kappa^{eg} := 2\kappa^{ef} + 2\kappa^{mH}.$$

Proof. Fix $x \in B(x^0; \Delta_{D^Y})$ and denote $d = x - x^0$. From Assumption 14, there exist $v \in \mathbb{R}^m$ such that $d = \sum_{i=1}^m v_i d^i$. This implies that $\Delta_{D^Y} \overline{D^Y} v = d$ is a consistent linear system. Let v be the minimum norm solution to this system, i.e., $v = \overline{D^Y}^\dagger d / \Delta_{D^Y}$. Suppose that $m^{x^0, Y}(x) = f(x^0) + \alpha^\top d + \frac{1}{2} d^\top H d$. Since $m^{x^0, Y}(x)$ interpolates f on Y , $\alpha^\top d^i = f(x^0 + d^i) - f(x^0) - \frac{1}{2} (d^i)^\top H d^i$ and so

$$\left| m^{x^0, Y}(x) - f(x^0) - \nabla f(x^0)^\top d \right|$$

$$\begin{aligned}
&\leq \left| (\alpha - \nabla f(x^0))^\top d \right| + \frac{1}{2} |d^\top H d| \leq \left| \sum_{i=1}^m v_i (\alpha - \nabla f(x^0))^\top d^i \right| + \frac{\kappa^{mH}}{2} \Delta_{D^Y}^2 \\
&\leq \left| \sum_{i=1}^m v_i (f(x^0 + d^i) - f(x^0) - \nabla f(x^0)^\top d^i) \right| + \frac{\kappa^{mH}}{2} (\|v\|_1 + 1) \Delta_{D^Y}^2 \\
&\leq \left(\frac{L_{\nabla f} + \kappa^{mH}}{2} \|v\|_1 + \frac{\kappa^{mH}}{2} \right) \Delta_{D^Y}^2.
\end{aligned}$$

Since $\|d\|_1 \leq \sqrt{n}\|d\|$, we have $\|v\|_1 \leq \|\overline{D^Y}^\dagger\|_1 \|d\|_1 / \Delta_{D^Y} \leq \sqrt{n} \|\overline{D^Y}^\dagger\|_1$. The result follows from [19, Lem. 5.1(a)]. \square

Remark 18 Theorem 17 implies that any quadratic interpolation model with a bounded Hessian belongs to a class of fully linear models. Therefore, to prove that a class of models is fully linear, it suffices to prove, under reasonable assumptions, that the corresponding Hessians are bounded.

3.1 Error bounds of MFN models

For the MFN model, fully linear error bounds can be found in, e.g., [19, Thm. 5.15] and [11, Thms. 5.4 and 5.7]. However, their definitions of MFN models differ slightly from Definition 8. In particular, Definition 8 requires the model to interpolate f at x^0 , which is not imposed in [19, Thm. 5.15]. Moreover, [19, Thm. 5.15] replaces the objective in Problem (MFN) with $\frac{1}{4}\|H\|_F$, and [11, Thms. 5.4 and 5.7] express the bounds in terms of a specific geometric measure of Y known as the Λ -poisedness constant. Following the approach of [19, Thm. 5.15], we establish a fully linear error bound for the MFN model in Definition 8.

Theorem 19 *Suppose that Assumption 15 holds and Y is poised for MFN interpolation. Define $P \in \mathbb{R}^{m \times m}$ with $P_{ij} := \frac{1}{4}((d^i / \Delta_{D^Y})^\top (d^j / \Delta_{D^Y}))^2$, and $F := \begin{bmatrix} P & \overline{D^Y}^\top \\ \overline{D^Y} & \mathbf{0}_{n \times n} \end{bmatrix}$. Then, $\nabla_{\text{MFN}}^2 f(x^0; Y)$ is unique and satisfies*

$$\|\nabla_{\text{MFN}}^2 f(x^0; Y)\| \leq \|\nabla_{\text{MFN}}^2 f(x^0; Y)\|_F \leq \kappa_{\text{MFN}}^{mH} := \frac{L_{\nabla f}}{4} m \|F^{-1}\|_\infty.$$

Proof. Similar to [19, Lem. 5.11], we can show that the solution to Problem (MFN), denoted by (α^*, H^*) , can be obtained by solving

$$\underbrace{\begin{bmatrix} \Delta_{D^Y}^4 P & (\overline{D^Y})^\top \\ \overline{D^Y} & \mathbf{0}_{n \times n} \end{bmatrix}}_{\tilde{F}} \begin{bmatrix} \lambda \\ \alpha^* \end{bmatrix} = \begin{bmatrix} \delta_f(x^0; D^Y) \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad H^* = \frac{1}{2} \sum_{i=1}^m \lambda_i d^i (d^i)^\top,$$

where $\lambda \in \mathbb{R}^m$ is the Lagrange multiplier. Since Y is poised for MFN interpolation, \tilde{F} is invertible and so F is invertible. The proof of [19, Lem. 5.14] gives $|\lambda_i| \leq \frac{L_{\nabla f}}{2} \|F^{-1}\|_\infty \Delta_{D^Y}^{-2}$, so

$$\|H^*\| \leq \|H^*\|_F \leq \frac{1}{2} \sum_{i=1}^m |\lambda_i| \|d^i\|^2 \leq \frac{1}{2} \sum_{i=1}^m |\lambda_i| \Delta_{D^Y}^2 \leq \frac{L_{\nabla f}}{4} m \|F^{-1}\|_\infty.$$

\square

Theorem 20 Under the assumptions of Theorems 17 and 19, $m_{\text{MFN}}^{x^0; Y}(x)$ belongs to a class of fully linear models of f at x^0 parameterized by Δ_{D^Y} , with

$$\kappa_{\text{MFN}}^{ef} := \frac{L_{\nabla f} + \kappa_{\text{MFN}}^{mH}}{2} \sqrt{n} \left\| \overline{D^Y}^\dagger \right\|_1 + \frac{L_{\nabla f} + \kappa_{\text{MFN}}^{mH}}{2}, \quad \kappa_{\text{MFN}}^{eg} := 2\kappa_{\text{MFN}}^{ef} + 2\kappa_{\text{MFN}}^{mH}.$$

Proof. Directly apply Theorem 17 with $\kappa^{mH} = \kappa_{\text{MFN}}^{mH}$. \square

3.2 Error bounds of MN models

For the MN model, a fully linear bound follows from [11, Thm. 5.4] if the model Hessians are uniformly bounded. In this section, we show that this assumption is unnecessary, provided that the sample set is poised for MFN interpolation.

Remark 21 In practice, we often assume that $\{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ is compact. This is a common choice for the set K in the following theorems.

Theorem 22 Suppose that Assumption 15 holds and Y is poised for MFN interpolation. Let $K \subseteq \mathbb{R}^n$ be a compact set and so there exists $\kappa^g > 0$ such that $\|\nabla f(x)\| \leq \kappa^g$ on K . Then, $\nabla_{\text{MN}}^2 f(x^0; Y)$ is unique and satisfies

$$\|\nabla_{\text{MN}}^2 f(x^0; Y)\| \leq \kappa_{\text{MN}}^{mH} := \sqrt{(\kappa^g + \kappa_{\text{MFN}}^{eg} \overline{\Delta})^2 + (\kappa_{\text{MFN}}^{mH})^2} \quad \text{for all } x^0 \in K,$$

for any $\overline{\Delta} \geq \Delta_{D^Y}$.

Proof. Since $(\nabla_{\text{MFN}} f(x^0; Y), \nabla_{\text{MFN}}^2 f(x^0; Y))$ is feasible for Problem (MN),

$$\begin{aligned} \|\nabla_{\text{MN}}^2 f(x^0; Y)\|^2 &\leq \|\nabla_{\text{MN}}^2 f(x^0; Y)\|_F^2 \leq \|\nabla_{\text{MN}} f(x^0; Y)\|^2 + \|\nabla_{\text{MFN}}^2 f(x^0; Y)\|_F^2 \\ &\leq \|\nabla_{\text{MFN}} f(x^0; Y)\|^2 + \|\nabla_{\text{MFN}}^2 f(x^0; Y)\|_F^2. \end{aligned}$$

Using the triangle inequality and the fact that $m_{\text{MFN}}^{x^0; Y}(x)$ is fully linear, we get $\|\nabla_{\text{MFN}} f(x^0; Y)\| \leq \|\nabla f(x^0)\| + \kappa_{\text{MFN}}^{eg} \Delta_{D^Y}$. The proof is complete by combining the inequalities above and noticing that $\|\nabla f(x)\| \leq \kappa^g$ on K . \square

Theorem 23 Under the assumptions of Theorems 17 and 22, for all $x^0 \in K$, $m_{\text{MN}}^{x^0; Y}(x)$ belongs to a class of fully linear models of f at x^0 parameterized by Δ_{D^Y} , with

$$\kappa_{\text{MN}}^{ef} := \frac{L_{\nabla f} + \kappa_{\text{MN}}^{mH}}{2} \sqrt{n} \left\| \overline{D^Y}^\dagger \right\|_1 + \frac{L_{\nabla f} + \kappa_{\text{MN}}^{mH}}{2}, \quad \kappa_{\text{MN}}^{eg} := 2\kappa_{\text{MN}}^{ef} + 2\kappa_{\text{MN}}^{mH}.$$

Proof. Directly apply Theorem 17 with $\kappa^{mH} = \kappa_{\text{MN}}^{mH}$. \square

3.3 Error bounds of QS models

In this subsection, we derive a fully linear error bound for the QS model in the case where it is an interpolation model. To our knowledge, this is the first fully linear bound for the QS model.

Theorem 24 Suppose that Assumption 15 holds and $\nabla_{\text{QS}}^2 f(x^0; Y)$ has the form $\nabla_{\text{QS}}^2 f(x^0; Y) = \sum_{k=1}^l \beta_k \nabla_s^2 f(x^0; S^k; T_{1:p^k}^k)$, where $l \geq 1$, $S^k \in \mathbb{R}^{n \times p^k}$, and $T_{1:p^k}^k = \{T_1^k, \dots, T_{p^k}^k\}$ with each $T_i^k \in \mathbb{R}^{n \times q_i^k}$. Then,

$$\|\nabla_{\text{QS}}^2 f(x^0; Y)\| \leq \kappa_{\text{QS}}^{mH} := L_{\nabla f} \sum_{k=1}^l |\beta_k| \|S^{k\dagger}\| \sqrt{\sum_{i=1}^{p^k} q_i^k \|\overline{T_i^k}^\dagger\|^2}.$$

Proof. It suffices to show $\|\nabla_s^2 f(x^0; S; T_{1:p})\| \leq L_{\nabla f} \|S^\dagger\| \sqrt{\sum_{i=1}^p q_i \|\overline{T_i}^\dagger\|^2}$ for any $S \in \mathbb{R}^{n \times p}$ and $T_i \in \mathbb{R}^{n \times q_i}$, $i \in [p]$, since the result then follows from the triangle inequality. The j -th element of $\delta_f(x^0 + s^i; T_i) - \delta_f(x^0; T_i)$ satisfies

$$\begin{aligned} & \left| f(x^0 + s^i + t_i^j) - f(x^0 + s^i) - f(x^0 + t_i^j) + f(x^0) \right| \\ &= \left| \int_0^1 \left(\nabla f(x^0 + s^i + \tau t_i^j) - \nabla f(x^0 + \tau t_i^j) \right)^\top t_i^j d\tau \right| \leq L_{\nabla f} \Delta_S \Delta_{T_i}. \end{aligned}$$

We have the i -th row of $\delta_f^2(x^0; S; T_{1:p})$ satisfies

$$\begin{aligned} & \left\| \nabla_s f(x^0 + s^i; T_i) - \nabla_s f(x^0; T_i) \right\| = \left\| (T_i^\top)^\dagger (\delta_f(x^0 + s^i; T_i) - \delta_f(x^0; T_i)) \right\| \\ & \leq \sqrt{q_i} L_{\nabla f} \Delta_S \Delta_{T_i} \|T_i^\dagger\| = \sqrt{q_i} L_{\nabla f} \|\overline{T_i}^\dagger\| \Delta_S. \end{aligned}$$

The result follows from

$$\|\nabla_s^2 f(x^0; S; T_{1:p})\| \leq \|S^\dagger\| \|\delta_f^2(x^0; S; T_{1:p})\|_F \leq \|S^\dagger\| L_{\nabla f} \Delta_S \sqrt{\sum_{i=1}^p q_i \|\overline{T_i}^\dagger\|^2}.$$

□

Theorem 25 Suppose that the assumptions of Theorems 17 and 24 hold. If $m_{\text{QS}}^{x^0, Y}(x)$ interpolates f on Y , then for all $x^0 \in \mathbb{R}^n$, $m_{\text{QS}}^{x^0, Y}(x)$ belongs to a class of fully linear models of f at x^0 parameterized by Δ_{D^Y} , with

$$\kappa_{\text{QS}}^{ef} := \frac{L_{\nabla f} + \kappa_{\text{QS}}^{mH}}{2} \sqrt{n} \|\overline{D^Y}^\dagger\|_1 + \frac{L_{\nabla f} + \kappa_{\text{QS}}^{mH}}{2}, \quad \kappa_{\text{QS}}^{eg} := 2\kappa_{\text{QS}}^{ef} + 2\kappa_{\text{QS}}^{mH}.$$

Proof. Directly apply Theorem 17 with $\kappa^{mH} = \kappa_{\text{QS}}^{mH}$. □

4 Directional Hessian error bounds with structured Y

In this section, we derive directional Hessian error bounds for the Hessian approximations arising from the MN, MFN, and QS models, under a mild structural assumption on Y . Our results show that the directional errors range from $\mathcal{O}(1)$ to $\mathcal{O}(\Delta_{D^Y})$, depending on the direction of interest. Notably, they achieve fully quadratic accuracy along sample directions. This highlights that all three models achieve significantly higher accuracy than general fully linear models in a directional sense. To our knowledge, these are the first directional error bounds established for these Hessian approximations.

We may use the following assumptions.

Assumption 26 The set $Y = \{x^0, x^0 + d^i, x^0 - d^i : i \in [p]\}$.

Assumption 27 The objective $f \in \mathcal{C}^{2+}$ with $L_{\nabla^2 f}$ -Lipschitz Hessian.

Assumption 28 The Problem (MN) is feasible.

Notice that Assumption 28 guarantees the existence of a solution to Problem (MFN). However, the solution need not be unique. Indeed, as shown in [5, Rem. 2 and Exp. 1], in this case, $\nabla_{\text{MFN}}^2 f(x^0; Y)$ is always unique, whereas $\nabla_{\text{MFN}} f(x^0; Y)$ may not be unique.

The following two lemmas present basic properties of \mathcal{C}^{2+} functions and will be used multiple times in the proofs in this subsection.

Lemma 29 Suppose that Assumption 27 holds. For $x^0, u, v, w^1, w^2 \in \mathbb{R}^n$, define $F(u, v, w^1, w^2) := \int_0^1 (1-t) u^\top (\nabla^2 f(x^0 + tw^1) - \nabla^2 f(x^0 + tw^2)) v dt$. Then, $|F(u, v, w^1, w^2)| \leq \frac{L_{\nabla^2 f}}{6} \|u\| \|v\| \|w^1 - w^2\|$.

Proof. By the triangle inequality and Assumption 27, we obtain

$$|F(u, v, w^1, w^2)| \leq L_{\nabla^2 f} \|u\| \|v\| \|w^1 - w^2\| \int_0^1 (1-t) t dt,$$

and the result follows. \square

Lemma 30 Suppose that Assumption 27 holds. Let $\Delta > 0$. Then, for any $x^0 \in \mathbb{R}^n$, $u, v \in B(\mathbf{0}; \Delta) \subseteq \mathbb{R}^n$,

$$\begin{aligned} E(u, v) := & \left| 2f(x^0 + \frac{1}{2}u + \frac{1}{2}v) + 2f(x^0 - \frac{1}{2}u - \frac{1}{2}v) - \frac{1}{2}f(x^0 + u) - \frac{1}{2}f(x^0 + v) \right. \\ & \left. - \frac{1}{2}f(x^0 - u) - \frac{1}{2}f(x^0 - v) - 2f(x^0) - u^\top \nabla^2 f(x^0)v \right| \leq \frac{2L_{\nabla^2 f}}{3} \Delta^3. \end{aligned}$$

Moreover,

$$E(u, u) = |f(x^0 + u) + f(x^0 - u) - 2f(x^0) - u^\top \nabla^2 f(x^0)u| \leq \frac{L_{\nabla^2 f}}{3} \Delta^3.$$

Proof. Notice that $u^\top \nabla^2 f(x^0)v = 2 \int_0^1 (1-t) u^\top \nabla^2 f(x^0)v dt$ and, for all $d \in \mathbb{R}^n$, $f(x^0 + d) = f(x^0) + \nabla f(x^0)^\top d + \int_0^1 (1-t) d^\top \nabla^2 f(x^0 + td)d dt$. We can compute that

$$\begin{aligned} E(u, v) &= \left| \frac{1}{2}F(u, u, \frac{1}{2}(u+v), u) + \frac{1}{2}F(v, v, \frac{1}{2}(u+v), v) + \frac{1}{2}F(u, u, -\frac{1}{2}(u+v), -u) \right. \\ & \quad \left. + \frac{1}{2}F(v, v, -\frac{1}{2}(u+v), -v) + F(u, v, \frac{1}{2}(u+v), \mathbf{0}) + F(u, v, -\frac{1}{2}(u+v), \mathbf{0}) \right| \\ &\leq \frac{L_{\nabla^2 f}}{6} \left(\frac{1}{2}\|u\|^2 \|u-v\| + \frac{1}{2}\|v\|^2 \|u-v\| + \|u\| \|v\| \|u+v\| \right) \leq \frac{2L_{\nabla^2 f}}{3} \Delta^3. \end{aligned}$$

Finally, if we take $u = v$ in the above inequality, then $\|u-v\| = 0$ and so $E(u, u) \leq \frac{L_{\nabla^2 f}}{6} \|u\| \|v\| \|u+v\| \leq \frac{L_{\nabla^2 f}}{3} \Delta^3$. \square

We begin with a directional Hessian error bound in the directions $d^i/\|d^i\|$. Lemma 31 shows that this bound is $\mathcal{O}(\Delta_{D^Y})$ for the MN and MFN models.

Lemma 31 *Suppose that Assumptions 26 to 28 hold. Let $\mathcal{X} \in \{\text{MN}, \text{MFN}\}$. Then, for any non-zero $d^i \in D^Y$,*

$$\left| \left(\frac{d^i}{\|d^i\|} \right)^\top (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) \frac{d^i}{\|d^i\|} \right| \leq \frac{L_{\nabla^2 f}}{3} \Delta_{D^Y}. \quad (2)$$

Proof. Under Assumption 26, direct row operations show that the equality constraints of Problems (MN) and (MFN) are

$$\begin{cases} (d^i)^\top \alpha = \frac{1}{2} (f(x^0 + d^i) - f(x^0)) - \frac{1}{2} (f(x^0 - d^i) - f(x^0)), & i \in [p], \\ (d^i)^\top H(d^i) = f(x^0 + d^i) + f(x^0 - d^i) - 2f(x^0), & i \in [p]. \end{cases}$$

Using Lemma 30 with $u = d^i$ and $\Delta = \|d^i\|$, we obtain

$$\left| \left(\frac{d^i}{\|d^i\|} \right)^\top (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) \frac{d^i}{\|d^i\|} \right| = \frac{1}{\|d^i\|^2} E(d^i, d^i) \leq \frac{L_{\nabla^2 f}}{3} \Delta_{D^Y}.$$

□

If we replace one of the $d^i/\|d^i\|$ on the left-hand side of Inequality (2) by $d^j/\|d^j\|$ where $i \neq j$, then the error bound is slightly worse than that in Inequality (2). As shown in Lemma 32, the bound in this case is $\mathcal{O}(1)$.

Lemma 32 *Suppose that Assumptions 14 to 28 hold and Y is poised for MFN interpolation. Let $\mathcal{X} \in \{\text{MN}, \text{MFN}\}$. Then, for any non-zero $d^i, d^j \in D^Y$ with $i \neq j$,*

$$\left| \left(\frac{d^i}{\|d^i\|} \right)^\top (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) \frac{d^j}{\|d^j\|} \right| \leq 4\kappa_{\mathcal{X}}^{ef} \frac{\Delta_{D^Y}^2}{\|d^i\|\|d^j\|} + \frac{2L_{\nabla^2 f}}{3} \frac{\Delta_{D^Y}^3}{\|d^i\|\|d^j\|}.$$

Proof. Under the assumptions, we have that $m_{\mathcal{X}}^{x^0, Y}(x)$ is fully linear by Theorems 20 and 23. Since $(\nabla_{\mathcal{X}} f(x^0; Y), \nabla_{\mathcal{X}}^2 f(x^0; Y))$ satisfies the constraints of Problems (MN) and (MFN), we can compute that

$$\begin{aligned} & (d^i)^\top \nabla_{\mathcal{X}}^2 f(x^0; Y) d^j \\ &= 2m_{\mathcal{X}}^{x^0, Y}(x^0 + \frac{1}{2}d^i + \frac{1}{2}d^j) + 2m_{\mathcal{X}}^{x^0, Y}(x^0 - \frac{1}{2}d^i - \frac{1}{2}d^j) \\ & \quad - \frac{1}{2}f(x^0 + d^i) - \frac{1}{2}f(x^0 + d^j) - \frac{1}{2}f(x^0 - d^i) - \frac{1}{2}f(x^0 - d^j) - 2f(x^0). \end{aligned}$$

Using the triangle inequality and the fact that $m_{\mathcal{X}}^{x^0, Y}(x)$ is fully linear, we get

$$\left| (d^i)^\top (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) d^j \right| \leq 4\kappa_{\mathcal{X}}^{ef} \Delta_{D^Y}^2 + E(d^i, d^j).$$

The result follows from Lemma 30 with $u = d^i$, $v = d^j$, and $\Delta = \Delta_{D^Y}$. □

Using Lemmas 31 and 32, we establish the following bounds on the directional Hessian error for any direction $d/\|d\| \in \mathbb{R}^n$. These results further yield a $\mathcal{O}(1)$ bound on the Hessian approximation error.

Theorem 33 Suppose that the assumptions of Lemma 32 hold. Denote $D := [d^1 \cdots d^p]$. Then, $\text{col}(D) = \mathbb{R}^n$ and for any non-zero $d \in \mathbb{R}^n$, we can write $d = Dv$ with $v = D^\dagger d$. Moreover,

$$\begin{aligned} & \left| \left(\frac{d}{\|d\|} \right)^\top (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) \frac{d}{\|d\|} \right| \\ & \leq 4 \frac{\|v\|_1^2 - \|v\|_\infty^2}{\|v\|^2} \kappa_{\mathcal{X}}^{ef} \|\bar{D}^\dagger\|^2 + \frac{L_{\nabla^2 f}}{3} \|\bar{D}^\dagger\|^2 \left(2 \frac{\|v\|_1^2 - \|v\|_\infty^2}{\|v\|^2} + 1 \right) \Delta_D. \end{aligned}$$

In particular, we have that

$$\begin{aligned} & \|\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)\| = \max_{\|d\|=1} |d^\top (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) d| \\ & \leq 4 \left(p - \frac{1}{p} \right) \kappa_{\mathcal{X}}^{ef} \|\bar{D}^\dagger\|^2 + \frac{L_{\nabla^2 f}}{3} \|\bar{D}^\dagger\|^2 \left(2p - \frac{2}{p} + 1 \right) \Delta_D. \end{aligned}$$

Proof. Under Assumption 14, we have that $\text{col}(D) = \text{col}(D^Y) = \mathbb{R}^n$ and $d = DD^\dagger d = Dv = \sum_{i=1}^p v_i d^i$. Moreover, $\Delta_D = \Delta_{D^Y}$. Combining the triangle inequality and previous results in this section, we get

$$\begin{aligned} & \left| \left(\frac{d}{\|d\|} \right)^\top (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) \frac{d}{\|d\|} \right| \\ & = \frac{1}{\|d\|^2} \left| \left(\sum_{i=1}^p v_i (d^i)^\top \right) (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) \left(\sum_{i=1}^p v_i d^i \right) \right| \\ & \leq \frac{L_{\nabla^2 f}}{3} \frac{\|v\|^2}{\|d\|^2} \Delta_D^3 + \frac{1}{\|d\|^2} \left(4\kappa_{\mathcal{X}}^{ef} \Delta_D^2 + \frac{2L_{\nabla^2 f}}{3} \Delta_D^3 \right) \sum_{i,j=1, i \neq j}^p |v_i v_j|. \end{aligned}$$

Without loss of generality, we suppose that v_k has the largest absolute value among all v_i , where $i \in [p]$. Then,

$$\begin{aligned} \sum_{i,j=1, i \neq j}^p |v_i v_j| & = 2|v_k| \sum_{i=1, i \neq k}^p |v_i| + \left(\sum_{i=1, i \neq k}^p |v_i| \right)^2 - \sum_{i=1, i \neq k}^p |v_i|^2 \\ & \leq 2|v_k| \sum_{i=1, i \neq k}^p |v_i| + \left(\sum_{i=1, i \neq k}^p |v_i| \right)^2 \\ & = 2\|v\|_\infty (\|v\|_1 - \|v\|_\infty) + (\|v\|_1 - \|v\|_\infty)^2 = \|v\|_1^2 - \|v\|_\infty^2. \end{aligned}$$

Notice that $\|\bar{D}^\dagger\| = \|D^\dagger\| \Delta_D$, and $\|v\| \leq \|D^\dagger\| \|d\|$ implies $\|d\| \geq \|v\| / \|D^\dagger\|$. We obtain

$$\begin{aligned} & \left| \left(\frac{d}{\|d\|} \right)^\top (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) \frac{d}{\|d\|} \right| \\ & \leq \frac{L_{\nabla^2 f}}{3} \frac{\|v\|^2}{\|d\|^2} \Delta_D^3 + \left(4\kappa_{\mathcal{X}}^{ef} \Delta_D^2 + \frac{2L_{\nabla^2 f}}{3} \Delta_D^3 \right) \frac{\|v\|_1^2 - \|v\|_\infty^2}{\|d\|^2} \\ & \leq 4 \frac{\|v\|_1^2 - \|v\|_\infty^2}{\|v\|^2} \kappa_{\mathcal{X}}^{ef} \|\bar{D}^\dagger\|^2 + \frac{L_{\nabla^2 f}}{3} \|\bar{D}^\dagger\|^2 \left(2 \frac{\|v\|_1^2 - \|v\|_\infty^2}{\|v\|^2} + 1 \right) \Delta_D. \end{aligned}$$

Finally, we take the maximum over all d with $\|d\| = 1$, which is equivalent to maximizing over all $v \in \text{col}(D^\top)$ with $\|Dv\| = 1$. Since we are only seeking an upper bound, we ignore the constraint $v \in \text{col}(D^\top)$. Let $w = v/\|v\|$. Then,

$$\max_{\|Dv\|=1} \frac{\|v\|_1^2 - \|v\|_\infty^2}{\|v\|^2} = \max_{\|Dw\|\|v\|=1} (\|w\|_1^2 - \|w\|_\infty^2) = \max_{\substack{\|w\|=1 \\ \|Dw\|>0}} (\|w\|_1^2 - \|w\|_\infty^2).$$

Notice that the set $\text{argmax}_{\|w\|=1} (\|w\|_1^2 - \|w\|_\infty^2) = \{\frac{1}{\sqrt{p}}[\pm 1, \dots, \pm 1]^\top\} \subseteq \mathbb{R}^p$ spans \mathbb{R}^p . Since $\text{rank}(D) = n > 0$, there must exist an element in the set such that $\|Dw\| > 0$. This gives

$$\max_{\|w\|=1, \|Dw\|>0} (\|w\|_1^2 - \|w\|_\infty^2) = \max_{\|w\|=1} (\|w\|_1^2 - \|w\|_\infty^2) = p - \frac{1}{p}$$

and so

$$\begin{aligned} \|\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)\| &= \max_{\|d\|=1} |d^\top (\nabla_{\mathcal{X}}^2 f(x^0; Y) - \nabla^2 f(x^0)) d| \\ &\leq 4 \left(p - \frac{1}{p} \right) \kappa_{\mathcal{X}}^{ef} \|\bar{D}^\dagger\|^2 + \frac{L_{\nabla^2 f}}{3} \|\bar{D}^\dagger\|^2 \left(2p - \frac{2}{p} + 1 \right) \Delta_D. \end{aligned}$$

□

Remark 34 Comparing Lemmas 31, 32, and Theorem 33, we see that the Hessian approximations arising from the MN and MFN models have directional errors ranging from $\mathcal{O}(1)$ to $\mathcal{O}(\Delta_{D^Y})$. In particular, along the directions $d^i/\|d^i\|$, the error is $\mathcal{O}(\Delta_{D^Y})$, indicating that the MN and MFN models attain the same level of accuracy as fully quadratic models in these directions.

Remark 35 In Lemma 32 and Theorem 33, Assumptions 14 and 15, together with the requirement that Y be poised for MFN interpolation, are imposed solely to ensure that $m_{\mathcal{X}}^{x^0, Y}(x)$ is fully linear. However, an inspection of the proofs shows that the conclusions remain valid if these assumptions are replaced by the weaker requirement that $m_{\mathcal{X}}^{x^0, Y}(x)$ be fully linear only on the affine subspace $x^0 + \text{col}(D^Y)$ where $\text{col}(D^Y)$ is not necessarily \mathbb{R}^n .

A detailed study of fully linear models defined on affine subspaces can be found in [4, 7], while a thorough analysis of the relationship between full-space and subspace models is provided in [5].

Now, we develop similar results for the GSH. The results can be applied to derive similar directional error bounds for the Hessian of QS models. We will use the following, more specific version of Assumption 26.

Assumption 36 *The matrix $S = [d^1 \dots d^p]$ has full column rank and for $i \in [p]$, $T_i = [-d^i]$.*

We have $\Delta_{D^Y} = \Delta_S$ under Assumption 36. Similar to Lemma 31, the directional Hessian error bound of GSH in the directions $d^i/\|d^i\|$ is $\mathcal{O}(\Delta_S)$.

Lemma 37 *Suppose that Assumptions 27, 36 hold. Then, for any non-zero $d^i \in D^Y$,*

$$\left| \left(\frac{d^i}{\|d^i\|} \right)^\top (\nabla_s^2 f(x^0; S; T_{1:p}) - \nabla^2 f(x^0)) \frac{d^i}{\|d^i\|} \right| \leq \frac{L_{\nabla^2 f}}{3} \Delta_S. \quad (3)$$

Proof. Notice that $\nabla_s f(x^0 + d^i; T_i) = (f(x^0 + d^i) - f(x^0)) d^i / \|d^i\|^2$ and $\nabla_s f(x^0; T_i) = (f(x^0) - f(x^0 - d^i)) d^i / \|d^i\|^2$, so

$$\begin{aligned} & \nabla_s^2 f(x^0; S; T_{1:p}) \\ &= (S^\top)^\dagger \text{Diag} \left(\left[\frac{f(x^0 + d^1) + f(x^0 - d^1) - 2f(x^0)}{\|d^1\|^2} \quad \dots \quad \frac{f(x^0 + d^p) + f(x^0 - d^p) - 2f(x^0)}{\|d^p\|^2} \right] \right) S^\top. \end{aligned}$$

By Assumption 36, S has full column rank and so $S^\dagger d^i = e^i$. Therefore,

$$\begin{aligned} & \left| \left(\frac{d^i}{\|d^i\|} \right)^\top (\nabla_s^2 f(x^0; S; T_{1:p}) - \nabla^2 f(x^0)) \frac{d^i}{\|d^i\|} \right| \\ &= \frac{1}{\|d^i\|^2} \left| \left(\frac{f(x^0 + d^i) + f(x^0 - d^i) - 2f(x^0)}{\|d^i\|^2} e^i \right)^\top S^\top d^i - (d^i)^\top \nabla^2 f(x^0) d^i \right| \\ &= \frac{1}{\|d^i\|^2} E(d^i, d^i) \leq \frac{L_{\nabla^2 f}}{3} \|d^i\| \leq \frac{L_{\nabla^2 f}}{3} \Delta_S, \end{aligned}$$

where the last line follows from Lemma 30 with $u = d^i$ and $\Delta = \|d^i\|$. \square

Similar to Lemma 32, if we replace one of the $d^i / \|d^i\|$ on the left-hand side of Inequality (3) with $d^j / \|d^j\|$ where $i \neq j$, then the error bound deteriorates to $\mathcal{O}(1)$. A key difference between Lemmas 32 and 38 is that the latter does not require the corresponding model to be fully linear. Consequently, no full-rank assumption on D^Y , nor the requirement that Y be poised for MFN interpolation, is needed.

Lemma 38 *Suppose that Assumptions 27, 36 hold. Then, for any non-zero $d^i, d^j \in D^Y$ with $i \neq j$,*

$$\left| \left(\frac{d^i}{\|d^i\|} \right)^\top (\nabla_s^2 f(x^0; S; T_{1:p}) - \nabla^2 f(x^0)) \frac{d^j}{\|d^j\|} \right| \leq \|\nabla^2 f(x^0)\| + \frac{L_{\nabla^2 f}}{3} \Delta_S.$$

Proof. Using similar proof techniques as in Lemma 37, we obtain

$$\begin{aligned} & \left| \left(\frac{d^i}{\|d^i\|} \right)^\top (\nabla_s^2 f(x^0; S; T_{1:p}) - \nabla^2 f(x^0)) \frac{d^j}{\|d^j\|} \right| \\ &= \frac{1}{\|d^i\| \|d^j\|} \left| \frac{(d^i)^\top d^j}{\|d^i\|^2} (f(x^0 + d^i) + f(x^0 - d^i) - 2f(x^0)) - (d^i)^\top \nabla^2 f(x^0) d^j \right| \\ &\leq \frac{1}{\|d^i\|^2} E(d^i, d^i) + \frac{1}{\|d^i\| \|d^j\|} \left| (d^i)^\top \nabla^2 f(x^0) \left(\frac{(d^i)^\top d^j}{\|d^i\|^2} d^i - d^j \right) \right| \\ &\leq \frac{1}{\|d^j\|} \|\nabla^2 f(x^0)\| \left\| \frac{(d^i)^\top d^j}{\|d^i\|^2} d^i - d^j \right\| + \frac{L_{\nabla^2 f}}{3} \Delta_S. \end{aligned}$$

The proof is complete by noticing that $\frac{(d^i)^\top d^j}{\|d^i\|^2} d^i$ is the projection of d^j onto the linear span of d^i , so $\left\| \frac{(d^i)^\top d^j}{\|d^i\|^2} d^i - d^j \right\| \leq \|d^j\|$. \square

Using Lemmas 37 and 38, we can establish results similar to those in Theorem 33. However, as we noted before Lemma 38, we do not require any full-rank assumptions on D^Y . Consequently, the directional Hessian error bounds in Theorem 39 are established for any direction $d / \|d\|$ belonging to the linear span of d^1, \dots, d^p . These results further yield an $\mathcal{O}(1)$ bound on the Hessian approximation error restricted to the linear span of d^1, \dots, d^p .

Theorem 39 *Suppose that Assumptions 27, 36 hold. Then, for any non-zero $d \in \text{col}(S)$, we can write $d = Sv$ with $v = S^\dagger d$. Moreover,*

$$\begin{aligned} & \left| \left(\frac{d}{\|d\|} \right)^\top (\nabla_s^2 f(x^0; S; T_{1:p}) - \nabla^2 f(x^0)) \frac{d}{\|d\|} \right| \\ & \leq \frac{\|v\|_1^2 - \|v\|_\infty^2}{\|v\|^2} \|\nabla^2 f(x^0)\| \|\bar{S}^\dagger\|^2 + \frac{L_{\nabla^2 f}}{3} \|\bar{S}^\dagger\|^2 \left(\frac{\|v\|_1^2 - \|v\|_\infty^2}{\|v\|^2} + 1 \right) \Delta_S. \end{aligned}$$

In particular, we have that

$$\begin{aligned} & \max_{d \in \text{col}(S), \|d\|=1} |d^\top (\nabla_s^2 f(x^0; S; T_{1:p}) - \nabla^2 f(x^0)) d| \\ & \leq \left(p - \frac{1}{p} \right) \|\nabla^2 f(x^0)\| \|\bar{S}^\dagger\|^2 + \frac{L_{\nabla^2 f}}{3} \|\bar{S}^\dagger\|^2 \left(p - \frac{1}{p} + 1 \right) \Delta_S. \end{aligned}$$

Proof. Using similar proof techniques as in Theorem 33, we obtain

$$\begin{aligned} & \left| \left(\frac{d}{\|d\|} \right)^\top (\nabla_s^2 f(x^0; S; T_{1:p}) - \nabla^2 f(x^0)) \frac{d}{\|d\|} \right| \\ & \leq \frac{\|v\|_1^2 - \|v\|_\infty^2}{\|d\|^2} \|\nabla^2 f(x^0)\| \Delta_S^2 + \frac{L_{\nabla^2 f}}{3} \frac{\|v\|^2}{\|d\|^2} \Delta_S^3 + \frac{L_{\nabla^2 f}}{3} \frac{\|v\|_1^2 - \|v\|_\infty^2}{\|d\|^2} \Delta_S^3. \end{aligned}$$

The remaining proof is the same as that of Theorem 33. \square

Remark 40 Comparing Lemmas 37, 38, and Theorem 39, we see that the GSH have directional errors ranging from $\mathcal{O}(1)$ to $\mathcal{O}(\Delta_S)$. Similar to the Hessian approximations arising from the MN and MFN models, it achieves the same level of accuracy as the Hessian approximation of fully quadratic models along the directions $d^i/\|d^i\|$.

5 Relationships among MN, MFN, and QS models

In this section, we explore the relationships among the MN, MFN, and QS models. In particular, we examine the conditions under which they coincide. These results not only deepen our understanding of these models but also provide an alternative, potentially simpler approach to constructing them.

5.1 Relationships among MN, MFN, and QS models in the case $T_{1:p} = T$

In this subsection, we suppose that all $T_i = T = [t^1 \dots t^q]$ and

$$Y = \{x^0, x^0 + s^i, x^0 + t^j, x^0 + s^i + t^j : i \in [p], j \in [q]\}.$$

As mentioned in Remark 4, we have $\nabla_s^2 f(x^0; S; T)^\top = \nabla_s^2 f(x^0; T; S)$. Therefore, in this subsection we present results only for $\nabla_s^2 f(x^0; S; T)$. Similar results can be established for $\nabla_s^2 f(x^0; T; S)$ by interchanging the roles of S and T .

Direct row operations can show that the equality constraints of Problems (MN) and (MFN) are equivalent to

$$\begin{cases} (s^i)^\top \alpha + \frac{1}{2} (s^i)^\top H(s^i) = f(x^0 + s^i) - f(x^0), & i \in [p], \\ (t^j)^\top \alpha + \frac{1}{2} (t^j)^\top H(t^j) = f(x^0 + t^j) - f(x^0), & j \in [q], \\ S^\top HT = \delta_{\delta_f}(x^0; S; T). \end{cases} \quad (4)$$

Consider the following problem that only involves H as a variable:

$$\min_{H \in \mathbb{R}^{n \times n}} \frac{1}{2} \|H\|_F^2 \quad s.t. \quad S^\top H T = \delta_{\delta_f}(x^0; S; T), \quad H \in S_n(\mathbb{R}). \quad (5)$$

Clearly, for all (α, H) feasible for System (4), H is feasible for Problem (5). Therefore, the solution to Problem (5) should be closely related to the solution to Problems (MN) and (MFN).

We begin by introducing several lemmas that examine the properties of $\nabla_s^2 f(x^0; S; T)$. In particular, we discuss when it is the solution to Problem (5). As will be shown later, these results are crucial to linking the MN, MFN, and QS models. The next lemma shows that $\nabla_s^2 f(x^0; S; T)$ is always the unique solution to Problem (5) when the symmetry requirement is removed.

Lemma 41 *Suppose that Problem (6) is feasible. Then, $\nabla_s^2 f(x^0; S; T)$ is the unique solution to*

$$\min_{H \in \mathbb{R}^{n \times n}} \frac{1}{2} \|H\|_F^2 \quad s.t. \quad S^\top H T = \delta_{\delta_f}(x^0; S; T). \quad (6)$$

In particular, $\nabla_s^2 f(x^0; S; T)$ is the unique solution to Problem (5) if and only if Problem (5) is feasible and $\nabla_s^2 f(x^0; S; T) \in S_n(\mathbb{R})$ (or equivalently, $\nabla_s^2 f(x^0; S; T)$ is feasible for Problem (5)).

Proof. Notice that the objective functions of Problems (5) and (6) are strongly convex, and all constraints are linear with respect to H . The solution exists and is unique if and only if the problem is feasible. Moreover, the KKT conditions are necessary and sufficient for optimality.

Let $\Lambda \in \mathbb{R}^{q \times p}$ be the Lagrange multiplier. The Lagrangian function of Problem (6) is $L(H, \Lambda) = \frac{1}{2} \|H\|_F^2 + \text{tr}((S^\top H T - \delta_{\delta_f}(x^0; S; T))\Lambda)$. The KKT conditions are $H + S\Lambda^\top T^\top = \mathbf{0}_{n \times n}$ and $S^\top H T - \delta_{\delta_f}(x^0; S; T) = \mathbf{0}_{p \times q}$.

We claim that $(H^*, \Lambda^*) = (\nabla_s^2 f(x^0; S; T), -(S^\dagger H^* (T^\top)^\dagger)^\top)$ satisfies the KKT conditions. Indeed, since $S^\top H T = \delta_{\delta_f}(x^0; S; T)$ is feasible, every column of $\delta_{\delta_f}(x^0; S; T)$ lies in $\text{col}(S^\top)$ and every row of $\delta_{\delta_f}(x^0; S; T)$ lies in $\text{col}(T^\top)$, i.e., $S^\top (S^\top)^\dagger \delta_{\delta_f}(x^0; S; T) = \delta_{\delta_f}(x^0; S; T)$ and $\delta_{\delta_f}(x^0; S; T) T^\dagger T = \delta_{\delta_f}(x^0; S; T)$. Moreover, every column of H^* lies in $\text{col}((S^\top)^\dagger) = \text{col}(S)$ and every row of H^* lies in $\text{col}((T^\top)^\dagger) = \text{col}(T)$, i.e., $S S^\dagger H^* = H^*$ and $H^* (T^\top)^\dagger T^\top = H^*$. Therefore, $S^\top H^* T - \delta_{\delta_f}(x^0; S; T) = S^\top (S^\top)^\dagger \delta_{\delta_f}(x^0; S; T) T^\dagger T - \delta_{\delta_f}(x^0; S; T) = \mathbf{0}_{p \times q}$ and $H^* + S(\Lambda^*)^\top T^\top = H^* - S S^\dagger H^* (T^\top)^\dagger T^\top = \mathbf{0}_{n \times n}$, i.e., $H^* = \nabla_s^2 f(x^0; S; T)$ is a KKT point and therefore the unique solution. It immediately follows that $\nabla_s^2 f(x^0; S; T)$ is the unique solution to Problem (5) if and only if Problem (5) is feasible and $\nabla_s^2 f(x^0; S; T) \in S_n(\mathbb{R})$. \square

Lemma 41 provides a necessary and sufficient condition for $\nabla_s^2 f(x^0; S; T)$ to be the unique solution to Problem (5). Now, we provide a detailed characterization of the solution to Problem (5), under an extra assumption on $\text{col}(T)$ and $\text{col}(S)$.

Lemma 42 *Suppose that Problem (5) is feasible and $\text{col}(T) \subseteq \text{col}(S)$. The unique solution to Problem (5) can be expressed, and in fact must take the form $\nabla_s^2 f(x^0; S; T) + (S^\top)^\dagger Z (I_n - T T^\dagger) \in S_n(\mathbb{R})$, for some $Z \in \mathbb{R}^{p \times n}$. In particular, if T has full row rank, then $\nabla_s^2 f(x^0; S; T)$ is the unique solution to Problem (5).*

Proof. As mentioned in the proof of Lemma 41, the solution to Problem (5) exists and is unique if and only if the problem is feasible. Moreover, the KKT conditions are necessary and sufficient conditions for optimality.

Let $\Lambda_1 \in \mathbb{R}^{q \times p}$ and $\Lambda_2 \in \mathbb{R}^{n \times n}$. The Lagrangian function is

$$L(H, \Lambda_1, \Lambda_2) = \frac{1}{2} \|H\|_F^2 + \text{tr}((S^\top H T - \delta_{\delta_f}(x^0; S; T)) \Lambda_1) + \text{tr}((H - H^\top) \Lambda_2).$$

The KKT conditions are $H + SA_1^\top T^\top + A_2^\top - A_2 = \mathbf{0}_{n \times n}$, $S^\top HT = \delta_{\delta_f}(x^0; S; T)$, and $H = H^\top$. Since H is symmetric, we separate the symmetric and skew parts from the first KKT condition and get

$$H = -\frac{1}{2}(SA_1^\top T^\top + TA_1 S^\top) \quad (7)$$

$$\frac{1}{2}(SA_1^\top T^\top - TA_1 S^\top) = A_2 - A_2^\top. \quad (8)$$

The second KKT condition gives $-\frac{1}{2}S^\top(SA_1^\top T^\top + TA_1 S^\top)T = \delta_{\delta_f}(x^0; S; T)$, i.e., A_1 is the solution to

$$-\frac{1}{2}(S^\top SA_1^\top T^\top T + S^\top TA_1 S^\top T) = \delta_{\delta_f}(x^0; S; T). \quad (9)$$

The KKT conditions of Problem (5) are equivalent to Equations (7), (8), and (9). Equation (8) is always feasible by letting $A_2 = \frac{1}{2}SA_1^\top T^\top$.

Since $\text{col}(T) \subseteq \text{col}(S)$, we have $T = SM$ for some matrix $M \in \mathbb{R}^{p \times q}$. Then, Equation (7) implies that all columns of H lie in $\text{col}(S) = \text{col}((S^\top)^\dagger)$, i.e., $(S^\top)^\dagger S^\top H = H$. Moreover, Equation (9) becomes

$$-\frac{1}{2}S^\top S(A_1^\top M^\top + MA_1)S^\top T = \delta_{\delta_f}(x^0; S; T),$$

with the general solution in the form

$$-\frac{1}{2}S^\top S(A_1^\top M^\top + MA_1)S^\top = \delta_{\delta_f}(x^0; S; T)T^\dagger + Z(I_n - TT^\dagger),$$

where $Z \in \mathbb{R}^{p \times n}$. Since the solution to Problem (5) exists, there must exist Z and A_1 such that the equation above is feasible and so Equation (7) gives

$$\begin{aligned} H &= (S^\top)^\dagger S^\top H = -\frac{1}{2}(S^\top)^\dagger S^\top S(A_1^\top M^\top + MA_1)S^\top \\ &= \nabla_s^2 f(x^0; S; T) + (S^\top)^\dagger Z(I_n - TT^\dagger), \end{aligned}$$

i.e., $H = \nabla_s^2 f(x^0; S; T) + (S^\top)^\dagger Z(I_n - TT^\dagger)$ is a KKT point and thus is the unique solution. The proof is complete by noticing that if T has full row rank, then $I_n - TT^\dagger = \mathbf{0}_{n \times n}$. \square

The special case in which $\nabla_s^2 f(x^0; S; T)$ is the unique solution to Problem (5), as given in Lemma 42, in fact imply that $\text{col}(S) = \text{col}(T)$. More generally, we can show that if Problem (5) is feasible and $\text{col}(S) = \text{col}(T)$, then $\nabla_s^2 f(x^0; S; T)$ is the unique solution to Problem (5), i.e., the result holds without any full rank assumption on T .

Corollary 43 *Suppose Problem (5) is feasible and $\text{col}(T) = \text{col}(S)$. Then, $\nabla_s^2 f(x^0; S; T)$ must be the unique solution to Problem (5).*

Proof. Notice that we have $\text{col}(T) \subseteq \text{col}(S)$. By Lemma 42, the unique solution to Problem (5) has the form $H = \nabla_s^2 f(x^0; S; T) + (S^\top)^\dagger Z(I_n - TT^\dagger)$ for some $Z \in \mathbb{R}^{p \times n}$. Now, since $\text{col}(T) = \text{col}(S)$, Equation (7) implies that $HTT^\dagger = H$. Using the identity $T^\dagger TT^\dagger = T^\dagger$, we obtain

$$H = \nabla_s^2 f(x^0; S; T)TT^\dagger + (S^\top)^\dagger Z(I_n - TT^\dagger)TT^\dagger = \nabla_s^2 f(x^0; S; T).$$

\square

The results of Lemmas 41, 42, and Corollary 43 are shown in Table 1.

Now, we discuss some special structures of T that make $\nabla_s^2 f(x^0; S; T)$ feasible for Problem (5), which implies that $\nabla_s^2 f(x^0; S; T)$ is the unique solution to Problem (5). In fact, if T has these structures, then the feasibility is independent of f .

Table 1: Summary of the unique solution to Problems (5) and (6)

Assumption	Unique Solution
(5) is feasible and $\nabla_s^2 f(x^0; S; T) \in S_n(\mathbb{R})$	$\nabla_s^2 f(x^0; S; T)$
(5) is feasible and $\text{col}(T) \subseteq \text{col}(S)$	$\nabla_s^2 f(x^0; S; T) + (S^\top)^\dagger Z (I_n - TT^\dagger)$
(5) is feasible and $\text{col}(T) = \text{col}(S)$	$\nabla_s^2 f(x^0; S; T)$
(6) is feasible	$\nabla_s^2 f(x^0; S; T)$

Example 44 Suppose that S has full column rank. Let $\ell \in \{0, 1, \dots, p\}$. Define $U_S^0 := S$ and $U_S^\ell := [s^1 - s^\ell \dots s^{\ell-1} - s^\ell \quad -s^\ell \quad s^{\ell+1} - s^\ell \dots s^p - s^\ell]$ for $\ell > 0$. If $T = U_S^\ell$, then for all functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\nabla_s^2 f(x^0; S; T)$ is the unique solution to Problem (5).

Proof. Combining [16, Prop. 5.7] with [5, Thm. 5], we obtain that there exists $\alpha \in \mathbb{R}^n$ such that $(\alpha, \nabla_s^2 f(x^0; S; T))$ is feasible for System (4). Therefore, $\nabla_s^2 f(x^0; S; T)$ is feasible for Problem (5), and so is the unique solution. \square

The next lemma, inspired by [16, Prop. 5.5], discusses how certain transformations on S and T influence the symmetry of $\nabla_s^2 f(x^0; S; T)$ and its feasibility for Problem (5).

Lemma 45 *Let $N \in \mathbb{R}^{n \times n}$ be any orthogonal matrix. Let $P_1 \in \mathbb{R}^{p \times p}$ and $P_2 \in \mathbb{R}^{q \times q}$ be any permutation matrices. Define $\tilde{S} := NSP_1$, $\tilde{T} := NTP_2$, and $\tilde{f}(x) := f(x^0 + N^{-1}(x - x^0))$. If $\nabla_s^2 f(x^0; S; T)$ is symmetric, then $\nabla_s^2 \tilde{f}(x^0; \tilde{S}; \tilde{T})$ is symmetric. Moreover, if $\nabla_s^2 f(x^0; S; T)$ is feasible for Problem (5), then $\nabla_s^2 \tilde{f}(x^0; \tilde{S}; \tilde{T})$ is feasible for Problem (5) defined on \tilde{S} , \tilde{T} , and \tilde{f} .*

Proof. Suppose that $\nabla_s^2 f(x^0; S; T) = (S^\top)^\dagger \delta_{\delta_f}(x^0; S; T) T^\dagger$ is symmetric. By definition, we have that $\delta_{\delta_{\tilde{f}}}(x^0; \tilde{S}; \tilde{T}) = \delta_{\delta_f}(x^0; SP_1; TP_2)$. Suppose that $SP_1 = [s^{k_1} \dots s^{k_p}]$ and $TP_2 = [t^{l_1} \dots t^{l_q}]$. Then, direction computation can show that $(\delta_{\delta_{\tilde{f}}}(x^0; SP_1; TP_2))_{ij} = (\delta_{\delta_f}(x^0; S; T))_{k_i l_j}$. This means we have $\delta_{\delta_{\tilde{f}}}(x^0; SP_1; TP_2) = P_1^\top \delta_{\delta_f}(x^0; S; T) P_2$ and so

$$\nabla_s^2 \tilde{f}(x^0; \tilde{S}; \tilde{T}) = (\tilde{S}^\top)^\dagger \delta_{\delta_{\tilde{f}}}(x^0; SP_1; TP_2) \tilde{T}^\dagger = N (S^\top)^\dagger \delta_{\delta_f}(x^0; S; T) T^\dagger N^\top,$$

is symmetric. Now, suppose that $\nabla_s^2 f(x^0; S; T)$ is feasible for Problem (5). Notice that $\tilde{S}^\top (\tilde{S}^\top)^\dagger = P_1^\top S^\top N^{-1} N (S^\top)^\dagger P_1 = P_1^\top S^\top (S^\top)^\dagger P_1$ and, similarly, $\tilde{T}^\dagger \tilde{T} = P_2^\top T^\dagger TP_2$. We obtain

$$\begin{aligned} & \tilde{S}^\top \nabla_s^2 \tilde{f}(x^0; \tilde{S}; \tilde{T}) \tilde{T} \\ &= P_1^\top S^\top (S^\top)^\dagger P_1 \delta_{\delta_f}(x^0; SP_1; TP_2) P_2^\top T^\dagger TP_2 \\ &= P_1^\top S^\top (S^\top)^\dagger \delta_{\delta_f}(x^0; S; T) T^\dagger TP_2 = P_1^\top S^\top \nabla_s^2 f(x^0; S; T) TP_2 \\ &= P_1^\top \delta_{\delta_f}(x^0; S; T) P_2 = \delta_{\delta_f}(x^0; SP_1; TP_2) = \delta_{\delta_{\tilde{f}}}(x^0; \tilde{S}; \tilde{T}), \end{aligned}$$

and so $\nabla_s^2 \tilde{f}(x^0; \tilde{S}; \tilde{T})$ is feasible for Problem (5) defined on \tilde{S} , \tilde{T} , and \tilde{f} . \square

5.1.1 Constructing MN and MFN models from GSG and GSH in the case $\text{col}(T) = \text{col}(S)$

In this subsection, we specifically consider the case where $\text{col}(T) = \text{col}(S)$. We have proved that in this case, if Problem (5) is feasible, then $\nabla^2 f(x^0; S; T)$ is the unique solution to Problem (5). Based on this, we provide exact formulas to construct MN and MFN models using the GSG and GSH. We also discuss when the MN, MFN, and QS models coincide.

The following theorem characterizes the relationship among the GSG, GSH, and MN models. In particular, $\nabla_{\text{MN}}^2 f(x^0; Y)$ coincides exactly with the GSH of f at x^0 over S and T , while $\nabla_{\text{MN}} f(x^0; Y)$ corresponds to the GSG of f at x^0 over S , augmented by an additional correction term.

Theorem 46 *Suppose Problem (MN) is feasible and $\text{col}(T) = \text{col}(S)$. Then, we have $\nabla_{\text{MN}} f(x^0; Y) = \nabla_s f(x^0; S) - \frac{1}{2}(S^\top)^\dagger \text{diag}(\delta_{\delta_f}(x^0; S; T)T^\dagger S)$ and $\nabla_{\text{MN}}^2 f(x^0; Y) = \nabla_s^2 f(x^0; S; T)$.*

Proof. Since $\text{col}(S) = \text{col}(T)$, there exist $M_1 \in \mathbb{R}^{p \times q}$ and $M_2 \in \mathbb{R}^{q \times p}$ such that $T = SM_1$ and $S = TM_2$. We take $M_1 = S^\dagger T$ and $M_2 = T^\dagger S$. The last equality in System (4) is $S^\top HT = S^\top HSM_1 = M_2^\top T^\top HT = \delta_{\delta_f}(x^0; S; T)$, which gives

$$\begin{aligned} S^\top HS &= M_2^\top T^\top HTM_2 = \delta_{\delta_f}(x^0; S; T)M_2 = \delta_{\delta_f}(x^0; S; T)T^\dagger S \\ T^\top HT &= M_1^\top S^\top HSM_1 = M_1^\top \delta_{\delta_f}(x^0; S; T) = T^\top (S^\top)^\dagger \delta_{\delta_f}(x^0; S; T). \end{aligned}$$

Using these equations to cancel the term containing H in the remaining equalities in System (4), we obtain that System (4) is equivalent to

$$\begin{cases} S^\top \alpha = \delta_f(x^0; S) - \frac{1}{2} \text{diag}(\delta_{\delta_f}(x^0; S; T)T^\dagger S), \\ T^\top \alpha = \delta_f(x^0; T) - \frac{1}{2} \text{diag}(T^\top (S^\top)^\dagger \delta_{\delta_f}(x^0; S; T)), \\ S^\top HT = \delta_{\delta_f}(x^0; S; T). \end{cases}$$

Since $\text{col}(S) = \text{col}(T)$ and System (4) is feasible, the first two constraints in the system above must be equivalent, and Problem (5) is feasible. Thus, $\nabla_{\text{MN}}^2 f(x^0; Y)$ is the solution to Problem (5) and $\nabla_{\text{MN}} f(x^0; Y)$ is the minimum norm solution to $S^\top \alpha = \delta_f(x^0; S) - \frac{1}{2} \text{diag}(\delta_{\delta_f}(x^0; S; T)T^\dagger S)$. Therefore, we have $\nabla_{\text{MN}}^2 f(x^0; Y) = \nabla_s^2 f(x^0; S; T)$ by Corollary 43, and

$$\begin{aligned} \nabla_{\text{MN}} f(x^0; Y) &= (S^\top)^\dagger \delta_f(x^0; S) - \frac{1}{2} (S^\top)^\dagger \text{diag}(\delta_{\delta_f}(x^0; S; T)T^\dagger S) \\ &= \nabla_s f(x^0; S) - \frac{1}{2} (S^\top)^\dagger \text{diag}(\delta_{\delta_f}(x^0; S; T)T^\dagger S). \end{aligned}$$

□

Following the same proof as that of Theorem 46, we obtain similar results for the MFN model. A key distinction, however, is that Corollary 43 guarantees only the uniqueness of $\nabla_{\text{MFN}}^2 f(x^0; Y)$. In contrast, the uniqueness of $\nabla_{\text{MFN}} f(x^0; Y)$ requires an additional assumption, namely that Y is poised for MFN interpolation.

Theorem 47 *Suppose Problem (MFN) is feasible and $\text{col}(T) = \text{col}(S)$. Then, $\nabla_{\text{MFN}}^2 f(x^0; Y) = \nabla_s^2 f(x^0; S; T)$ and $\nabla_{\text{MFN}} f(x^0; Y)$ is any vector such that $(\nabla_{\text{MFN}} f(x^0; Y), \nabla_s^2 f(x^0; S; T))$ is feasible. If Y is poised for MFN interpolation, then $\nabla_{\text{MFN}} f(x^0; Y) = \nabla_s f(x^0; S) - \frac{1}{2}(S^\top)^\dagger \text{diag}(\delta_{\delta_f}(x^0; S; T)T^\dagger S)$.*

Remark 48 Comparing Theorems 46 and 47, we observe that the MN and MFN models coincide when $\text{col}(T) = \text{col}(S)$ and Y is poised for MFN interpolation. As we can see from the proof of Theorem 46, this ultimately stems from the fact that this particular structure of Y allows Problem (MN) to be separable in (α, H) .

An interesting special case is where S has full column rank and T has the structure discussed in Example 44. The proof of Example 44 shows that System (4) is feasible. Moreover, $\nabla_{\text{MN}}f(x^0; Y)$ and $\nabla_{\text{MN}}^2f(x^0; Y)$ are linear combinations of GSG and GSH. In particular, $m_{\text{MN}}^{x^0, Y}(x)$ belongs to the class of QS models of f at x^0 over Y .

Corollary 49 *Suppose S has full column rank and T is as in Example 44. Then, $\nabla_{\text{MN}}f(x^0; Y) = \nabla_s f(x^0; S) + \nabla_s f(x^0 - s^\ell; S) - \nabla_s f(x^0 - s^\ell; 2S)$ and $\nabla_{\text{MN}}^2f(x^0; Y) = \nabla_s^2 f(x^0; S; T)$. In particular, $m_{\text{MN}}^{x^0, Y}(x)$ belongs to the class of QS models of f at x^0 over Y .*

Proof. Let $T = U_S^\ell$. Notice that $\text{col}(S) = \text{col}(T)$. For any vector v , we define $v^0 := \mathbf{0}$. Define $M^\ell := I_p - e^\ell \mathbf{1}^\top - e^\ell (e^\ell)^\top$. Then, $T = SM^\ell$. Since $(M^\ell)^{-1} = M^\ell$ and S has full column rank, $T^\dagger S = M^\ell S^\dagger S = M^\ell$. Thus, the i -th element of $\text{diag}(\delta_{\delta_f}(x^0; S; T)T^\dagger S)$ is $f(x^0 + 2s^i - s^\ell) - 2f(x^0 + s^i - s^\ell) + f(x^0 - s^\ell)$ and so $\text{diag}(\delta_{\delta_f}(x^0; S; T)T^\dagger S) = \delta_f(x^0 - s^\ell; 2S) - 2\delta_f(x^0 - s^\ell; S)$. Therefore,

$$\begin{aligned} \nabla_{\text{MN}}f(x^0; Y) &= \nabla_s f(x^0; S) - \frac{1}{2} (S^\top)^\dagger (\delta_f(x^0 - s^\ell; 2S) - 2\delta_f(x^0 - s^\ell; S)) \\ &= \nabla_s f(x^0; S) + (S^\top)^\dagger \delta_f(x^0 - s^\ell; S) - (2S^\top)^\dagger \delta_f(x^0 - s^\ell; 2S) \\ &= \nabla_s f(x^0; S) + \nabla_s f(x^0 - s^\ell; S) - \nabla_s f(x^0 - s^\ell; 2S). \end{aligned}$$

□

Remark 50 When $\ell = 0$, $\nabla_{\text{MN}}f(x^0; Y) = 2\nabla_s f(x^0; S) - \nabla_s f(x^0; 2S)$ and $\nabla_{\text{MN}}^2f(x^0; Y) = \nabla_s^2 f(x^0; S; S)$. This extends the result in [7, Thm. 1], where it was only proved that the corresponding model satisfies interpolation conditions. Moreover, $2\nabla_s f(x^0; S) - \nabla_s f(x^0; 2S)$ is the *generalized adapted centred simplex gradient* over $\mathbb{Y} = \mathbb{Y}^+ \cup \tilde{\mathbb{Y}}$, where $\mathbb{Y}^+ = \{x^0, x^0 + s^i : i \in [p]\}$ and $\tilde{\mathbb{Y}} = \{x^0, x^0 + 2s^i : i \in [p]\}$. Detailed explanations and more properties of the generalized adapted centred simplex gradient can be found in [6, 8].

Similarly, we obtain corresponding results for the MFN model. Once again, an additional assumption that Y is poised for MFN interpolation is required to guarantee the uniqueness of $\nabla_{\text{MFN}}f(x^0; Y)$.

Corollary 51 *Suppose S has full column rank and T is as in Example 44. Then, $\nabla_{\text{MFN}}^2f(x^0; Y) = \nabla_s^2 f(x^0; S; T)$. If Y is poised for MFN interpolation, then $\nabla_{\text{MFN}}f(x^0; Y) = \nabla_s f(x^0; S) + \nabla_s f(x^0 - s^\ell; S) - \nabla_s f(x^0 - s^\ell; 2S)$ and $m_{\text{MFN}}^{x^0, Y}(x)$ belongs to the class of QS models of f at x^0 over Y .*

Remark 52 Comparing Corollaries 49 and 51, we observe that the MN, MFN, and QS models coincide when S has full column rank, T has the structure given in Example 44, and Y is poised for MFN interpolation.

5.2 Relationships among MN, MFN, and QS models in the case of different T_i

In this subsection, we suppose that all T_i are different. We show that, when S and T_i have a specific structure, results similar to these in Subsection 5.1.1 exist.

The following theorem is analogous to Theorem 46. Given the structures of S and T_i , the MN model can be constructed using GSG and GSH. In particular, it belongs to the class of QS models of f at x^0 over Y . Notably, these structures alone are sufficient to establish the results. No assumptions on the column spaces are required.

Theorem 53 Suppose that $S = [e^1 \cdots e^p]$ and $T_i = [-e^i]$, $i \in [p]$, where $p \leq n$. Then, we have $\nabla_{\text{MN}} f(x^0; Y) = \frac{1}{2} (\nabla_s f(x^0; S) + \nabla_s f(x^0; -S))$ and $\nabla_{\text{MN}}^2 f(x^0; Y) = \nabla_s^2 f(x^0; S; T_{1:p})$. In particular, $m_{\text{MN}}^{x^0, Y}(x)$ belongs to the class of QS models of f at x^0 over Y .

Proof. Note that $Y = \{x^0, x^0 + e^i, x^0 - e^i : i \in [p]\}$. Direct row operations show that the equality constraints of Problem (MN) are equivalent to

$$\begin{cases} (e^i)^\top \alpha = \frac{1}{2} (f(x^0 + e^i) - f(x^0)) - \frac{1}{2} (f(x^0 - e^i) - f(x^0)), & i \in [p], \\ (e^i)^\top H(e^i) = f(x^0 + e^i) + f(x^0 - e^i) - 2f(x^0), & i \in [p]. \end{cases}$$

Writing the first set of equalities as $S^\top \alpha = \frac{1}{2} (\delta_f(x^0; S) - \delta_f(x^0; -S))$, we get

$$\begin{aligned} \nabla_{\text{MN}} f(x^0; Y) &= \frac{1}{2} (\nabla_s f(x^0; S) + \nabla_s f(x^0; -S)) \\ \nabla_{\text{MN}}^2 f(x^0; Y) &= \text{Diag} \left(\begin{bmatrix} f(x^0 + e^1) + f(x^0 - e^1) - 2f(x^0) \\ \vdots \\ f(x^0 + e^p) + f(x^0 - e^p) - 2f(x^0) \\ \mathbf{0} \end{bmatrix} \right) \in \mathbb{R}^{n \times n}. \end{aligned}$$

It is straightforward to check that $\nabla_{\text{MN}}^2 f(x^0; Y) = \nabla_s^2 f(x^0; S; T_{1:p})$. \square

Remark 54 The $\frac{1}{2} (\nabla_s f(x^0; S) + \nabla_s f(x^0; -S))$ is the *generalized centred simplex gradient* over $\mathbb{Y} = \mathbb{Y}^+ \cup \mathbb{Y}^-$, where $\mathbb{Y}^+ = \{x^0, x^0 + s^i : i \in [p]\}$ and $\mathbb{Y}^- = \{x^0, x^0 - s^i : i \in [p]\}$. Detailed explanations and more properties of the generalized centred simplex gradient can be found in [8, 15]. In particular, we note that this is a special case of the generalized adapted centred simplex gradient [6, 8] mentioned in Remark 50.

Moreover, with the structure of S and $T_{1:p}$ defined in Theorem 53, the GSH coincides with the *generalized centered simplex Hessian*, and its diagonal elements form the *centered simplex Hessian diagonal*. Detailed discussions can be found in [17].

Once again, similar results exist for the MFN model, with the uniqueness of $\nabla_{\text{MFN}} f(x^0; Y)$ guaranteed by the poisedness of Y for MFN interpolation.

Theorem 55 Suppose that $S = [e^1 \cdots e^p]$ and $T_i = [-e^i]$, $i \in [p]$, where $p \leq n$. Then, we have $\nabla_{\text{MFN}}^2 f(x^0; Y) = \nabla_s^2 f(x^0; S; T_{1:p})$. If Y is poised for MFN interpolation, then $\nabla_{\text{MFN}} f(x^0; Y) = \frac{1}{2} (\nabla_s f(x^0; S) + \nabla_s f(x^0; -S))$ and $m_{\text{MFN}}^{x^0, Y}(x)$ belongs to the class of QS models of f at x^0 over Y .

Remark 56 Comparing Theorems 53 and 55, we observe that the MFN, MN, and QS models coincide when $S = [e^1 \cdots e^p]$, $T_i = [-e^i]$, and Y is poised for MFN interpolation. Once again, this is because this particular structure of Y allows Problem (MN) to be separable in (α, H) .

6 Conclusion

In this paper, we investigated the approximation accuracy and relationships among the MN, MFN, and QS models. We established fully linear error bounds and derived directional Hessian error bounds for all three models. In particular, we showed that their Hessian approximations exhibit directional errors ranging from $\mathcal{O}(1)$ to $\mathcal{O}(\Delta_{DY})$ and achieve fully quadratic accuracy along sample directions. These results demonstrate that meaningful second-order information can be obtained without fully determined quadratic interpolation. We also clarified the structural

relationships among the three models, identifying conditions under which they coincide. These insights not only deepen theoretical understanding but also point toward simpler and potentially more efficient strategies for constructing quadratic models in DFO.

A natural direction for future work is to study how different weightings of $\|\alpha\|^2$ and $\|H\|_F^2$ in the objective of Problem (MN) affect model accuracy and construction. The analysis developed here for the MN and MFN models extends directly to this more general setting.

References

1. C. AUDET AND W. HARE, *Derivative-free and blackbox optimization*, Springer, 2017.
2. ———, *Model-based methods in derivative-free nonsmooth optimization*, in Numerical Nonsmooth Optimization, Springer, 2020, pp. 655–691.
3. L. CAO, Z. WEN, AND Y.-X. YUAN, *The error in multivariate linear extrapolation with applications to derivative-free optimization*, 2023. <https://arxiv.org/abs/2307.00358>.
4. C. CARTIS AND L. ROBERTS, *Scalable subspace methods for derivative-free nonlinear least-squares optimization*, Math. Program., 199 (2023), pp. 461–524.
5. Y. CHEN, *Relationships between full-space and subspace quadratic interpolation models and simplex derivatives*, 2026. <https://arxiv.org/abs/2602.10374>.
6. Y. CHEN AND W. HARE, *Adapting the centred simplex gradient to compensate for misaligned sample points*, IMA J. Numer. Anal., (2023).
7. Y. CHEN, W. HARE, AND A. WIEBE, *Q-fully quadratic modeling and its application in a random subspace derivative-free method*, Comput. Optim. Appl., 89 (2024), pp. 317–360.
8. ———, *A general framework for floating point error analysis of first-order simplex derivatives*, Optim. Methods Softw., (2025).
9. A. R. CONN, K. SCHEINBERG, AND P. L. TOINT, *A derivative free optimization algorithm in practice*, in 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, 1998.
10. A. R. CONN, K. SCHEINBERG, AND L. N. VICENTE, *Geometry of sample sets in derivative-free optimization: Polynomial regression and underdetermined interpolation*, IMA J. Numer. Anal., 28 (2008), pp. 721–748.
11. ———, *Introduction to derivative-free optimization*, SIAM, 2009.
12. A. R. CONN AND P. L. TOINT, *An algorithm using quadratic interpolation for unconstrained derivative free optimization*, in Nonlinear Optimization and Applications, Springer, 1996, pp. 27–47.
13. A. L. CUSTÓDIO, J. E. DENNIS, AND L. N. VICENTE, *Using simplex gradients of nonsmooth functions in direct search methods*, IMA J. Numer. Anal., 28 (2008), pp. 770–784.
14. A. L. CUSTÓDIO AND L. N. VICENTE, *Using sampling and simplex derivatives in pattern search methods*, SIAM J. Optim., 18 (2007), pp. 537–555.
15. W. HARE, G. JARRY-BOLDUC, AND C. PLANIDEN, *Error bounds for overdetermined and underdetermined generalized centred simplex gradients*, IMA J. Numer. Anal., 42 (2022), pp. 744–770.
16. ———, *A matrix algebra approach to approximate Hessians*, IMA J. Numer. Anal., 44 (2024), pp. 2220–2250.
17. G. JARRY-BOLDUC AND C. PLANIDEN, *Using generalized simplex methods to approximate derivatives*, IMA J. Numer. Anal., (2025).
18. R. G. REGIS, *The calculus of simplex gradients*, Optim. Lett., 9 (2015), pp. 845–865.
19. L. ROBERTS, *Introduction to interpolation-based optimization*, 2025. <https://arxiv.org/abs/2510.04473>.