Limiting behaviour of the generalized simplex gradient as the number of points tends to infinity on a fixed shape in $\mathbb{R}^n$
Throughout this work, we use the standard notation found in [RW98]. The domain of a function \( f \) considers a ball as the sample region and provides an error bound ad infinitum. Illustrative examples are included setting when the sample region is a hyperrectangle, and in Section 4, we present its error bound ad infinitum. Section adapted to other shapes. We discuss the option of other shapes further in the concluding section.

One of the main concerns regarding any DFO gradient approximation method is the establishment of an upper bound on the error between the gradient approximation and the true gradient of the function. If the error can be controlled, then efficient, convergent optimization algorithms can be designed around the method. Recent work [HJBP20] has been done in this regard on several expansions of a popular approximation technique called the simplex gradient. This paper focuses on the expansion known as the \textit{generalized simplex gradient}. The definition of this method is given in the next section.

Given a \( C^2 \) function \( f : \mathbb{R}^n \rightarrow \mathbb{R}^p \), the simplex gradient approximates the gradient \( \nabla f \) at a point \( x \) by building an affine function \( \varphi \) that is a good approximation to \( f \) near \( x \) and then calculating \( \nabla \varphi \). The construction is accomplished by selecting \( n \) suitably spaced points in the domain of \( f \) that together with the reference point \( x \) form a simplex and serve to define \( \varphi \). Then \( \nabla \varphi \) is calculated using \( n + 1 \) function evaluations of \( f \). Some early examples of the simplex gradient are found in [Zie72, Zie73], in which a particular simplex based on coordinate directions is used to approximate gradients of quadratic and cubic functions. The \textit{Simplex Gradient} is introduced in [Kel99], where the ideas are generalized to allow for an arbitrary simplex. The error bound on the simplex gradient is a function of the dimension \( n \) and the radius of the simplex \( \Delta_S \) [Ke99].

The \textit{Generalized Simplex Gradient} (GSG) relaxes the requirement of evaluating exactly \( n+1 \) points in \( \mathbb{R}^n \), making it possible to obtain an approximate gradient with controlled error bound by using any finite number of points \( m + 1 \) (the positive integer \( m \) can be greater than, equal to, or less than \( n \) [Reg15]. This method has an error bound that depends on the number of points used \( m \) and the sampling radius \( \Delta_S \), which means that if one were to increase the number of sample points used and consider the limit as \( m \rightarrow \infty \), then the error bound does not necessarily remain bounded. Indeed, it is possible to construct examples in which the classical error bound tends to infinity as \( m \) increases [BHJB21]. This is a counter-intuitive event, as it is reasonable to conjecture that more sample points would provide better accuracy of the model function. This problem is investigated in [BHJB21] for functions \( f : \mathbb{R} \rightarrow \mathbb{R} \), in which new error bounds are developed and shown to have a more desirable behaviour at the limit.

In [BHJB21], it is shown that the norm of the difference between the approximate derivative and the true derivative of a one-dimensional function has a limit that is a factor of the Lipschitz constant of \( \nabla f \) and the sampling radius \( \Delta_S \). However, [BHJB21] considers only single-variable functions. The purpose of the present work is to extend those results to the multivariable setting.

We first consider sampling over a hyperrectangle with the reference point at a corner point. We explore the limiting behaviour of the gradient approximation on \( \mathbb{R}^n \) and calculate the limit of the corresponding error bound as the number of sample points tends to infinity. We then repeat the process considering the case where the sampling set is a ball with the reference point at the center. From these results, it becomes clear how the method can be adapted to other shapes. We discuss the option of other shapes further in the concluding section.

The remainder of this paper is organized as follows. Section 2 defines the notation used throughout the paper and provides some definitions for later use. In Section 3, we investigate the limiting behaviour of the GSG in the Cartesian setting when the sample region is a hyperrectangle, and in Section 4, we present its error bound ad infinitum. Section 5 considers a ball as the sample region and provides an error bound ad infinitum. Illustrative examples are included throughout. We make some concluding remarks and discuss avenues of future research in Section 6.

2 Preliminaries

2.1 Notation

Throughout this work, we use the standard notation found in [RW98]. The domain of a function \( f \) is denoted by \( \text{dom} \ f \). The transpose of a 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 \( \text{Id}_n \), or by \( \text{Id} \) if the dimension is clear. The vector \( e_i \in \mathbb{R}^n \) denotes the \( i \)-th column of \( \text{Id}_n \). The vector of all ones in \( \mathbb{R}^n \) is denoted by \( \mathbf{1}_n \), or \( \mathbf{1} \) if the dimension is clear. The \( i \)-th component of a vector \( v \) is denoted by \( v_i \), and of a vector \( v_n \) by \( [v_n]_i \). Similarly, the entry in the \( i \)-th row and \( j \)-th column of a matrix \( A \) is denoted by \( A_{i,j} \), and of a matrix \( A_n \) by \( [A_n]_{i,j} \). The diagonal matrix \( D \in \mathbb{R}^{n \times n} \) is written \( \text{diag}\{D_{1,1}, D_{2,2}, \ldots, D_{n,n}\} \) when convenient. Given a matrix \( A \in \mathbb{R}^{n \times m} \), we use the induced matrix norm

\[
\|A\| = \|A\|_2 = \max\{\|Ax\|_2 : \|x\|_2 = 1\}.
\]
The sphere and the ball of radius \( r > 0 \) centered at \( x^0 \in \mathbb{R}^n \) are denoted by \( S_n(x^0; r) \), and \( B_n(x^0; r) \) respectively. That is, 
\[
S_n(x^0; r) = \{ x \in \mathbb{R}^n : \| x - x^0 \| = r \}, \quad B_n(x^0; r) = \{ x \in \mathbb{R}^n : \| x - x^0 \| \leq r \}.
\]

### 2.2 Definitions and minor results

In this section, we list some concepts and properties that will be useful in developing the main results, as well as the formal definition of the GSG. Central to the GSG is the Moore–Penrose matrix pseudoinverse.

**Definition 2.1** (Moore–Penrose pseudoinverse). Let \( A \in \mathbb{R}^{n \times m} \). The Moore–Penrose pseudoinverse of \( A \), denoted by \( A^\dagger \), is the unique matrix in \( \mathbb{R}^{m \times n} \) that satisfies the following four equations:

\[
\begin{align*}
AA^\dagger A &= A, \\
A^\dagger AA^\dagger &= A^\dagger, \\
(AA^\dagger)^\top &= AA^\dagger, \\
(A^\dagger A)^\top &= A^\dagger A.
\end{align*}
\]

Note that for any \( A \in \mathbb{R}^{n \times m} \), there exists a unique Moore-Penrose pseudoinverse \( A^\dagger \in \mathbb{R}^{m \times n} \). The following property will be used frequently in the sequel.

- If \( A \) has full row rank \( n \), then \( A^\dagger \) is a right-inverse of \( A \), so that \( AA^\dagger = \text{Id}_n \). In this case, \( A^\dagger = A^\top (AA^\top)^{-1} \).

Next, we introduce the definition of the GSG and provide its error bound.

**Definition 2.2** (Generalized simplex gradient). Let \( f : \text{dom } f \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \). Let \( x^0 \in \text{dom } f \) be the reference point. Let \( S \in \mathbb{R}^{n \times N} \) with \( x^0 + S e_j \in \text{dom } f \) for all \( j \in \{1, 2, \ldots, N\} \). The generalized simplex gradient of \( f \) at \( x^0 \) over \( S \) is denoted by \( \nabla \_S f(x^0; S) \) and defined by

\[
\nabla \_S f(x^0; S) = (S^\top)^\top \delta_j(x^0; S),
\]

where \( \delta_j(x^0; S) = [f(x^0 + S e_1) - f(x^0) \quad \cdots \quad f(x^0 + S e_N) - f(x^0)]^\top \in \mathbb{R}^N \).

Occasionally, (1) is written in terms of \( (S^\top)^\top \). These two forms are equivalent, as \( (S^\top)^\top = (S^\top)^\top \). The following theorem establishes an error bound for the GSG. The error bound depends on the radius \( \Delta_S \) of the matrix \( S \), that is,

\[
\Delta_S = \max \{ \| S e_j \| : j \in \{1, 2, \ldots, N\} \}.
\]

**Theorem 2.3** (Classical error bound for the GSG). [Reg15, Cor.1] & [HJBP20, Thm.3.3] Let \( S \in \mathbb{R}^{n \times N} \) have full row rank and radius \( \Delta_S \). Let \( f \) be \( C^2 \) on an open domain containing \( B_n(x^0; \Delta_S) \) where \( x^0 \) is the reference point and \( \Delta_S > 0 \). Then

\[
\| \nabla \_S f(x^0; S) - \nabla f(x^0) \| \leq \frac{\sqrt{N}}{2} L_{\nabla f} \| (\hat{S}^\top)^\top \| \Delta_S,
\]

where \( \hat{S} = S / \Delta_S \) and \( L_{\nabla f} \) denotes the Lipschitz constant of \( \nabla f \) on \( B_n(x^0; \Delta_S) \).

Moreover, if \( f \in C^3 \) on an open domain containing \( B_n(x^0; \Delta_S) \) and \( S \) (or some permutation of \( S \)) has the form \( S = [A \quad -A] \) for some \( A \in \mathbb{R}^{n \times \frac{N}{2}} \), then

\[
\| \nabla \_S f(x^0; S) - \nabla f(x^0) \| \leq \frac{\sqrt{N}}{6} L_H \| (\hat{A}^\top)^\top \| \Delta_S^2,
\]

where \( \hat{A} = A / \Delta_S \) and \( L_H \) denotes the Lipschitz constant of \( \nabla^2 f \) on \( B(x^0; \Delta_S) \).

Note that (4) is the error bound defined for the generalized centered simplex gradient (GCSG) in [HJBP20, Thm.3.3]. It was shown in [HJBP20, Prop.2.10] that the GSG over \( S = [A \quad -A] \) and the GCSG over \( A \) are equivalent. So, we consider the GCSG a specific case of the GSG.

Finally, we recall the Sherman–Morrison–Woodbury formula for calculating matrix inverses.

---

3
where \( \Delta \) is partitioned into sub-hyperrectangles with lengths two equal parts, so \( N \). Hence, \( x \). Let \( \Delta \) be the reference point. Consider a hyperrectangular sample region with side lengths \( \Delta \). Suppose the longer side is divided into three equal parts, so \( n \). Then \( S_R \) is defined to be a matrix in \( \mathbb{R}^{n \times N} \) that contains all the directions used to form the sample points when the rightmost endpoint of each sub-hyperrectangle is chosen. 1 Hence, \( R(x^0; d) \) contains \( N \) sample points and one reference point, \( x^0 \).

Let \( D = \text{diag} \{ \Delta_1, \Delta_2, \ldots, \Delta_n \} \in \mathbb{R}^{n \times n} \). Then \( S_R \) can be written as a block matrix in the following way:

\[
S_R = \begin{bmatrix}
B_R^{1,1,1} & B_R^{1,1,2} & \ldots & B_R^{N_2,N_3,\ldots,N_{n-1},N_n}
\end{bmatrix} \in \mathbb{R}^{n \times N},
\]

where

\[
B_R^x = B_R^{z_2, z_3, \ldots, z_n} = D B_R^x = \begin{bmatrix}
1 & 2 & \ldots & N_1 \\
z_2 & z_2 & \ldots & z_2 \\
z_3 & z_3 & \ldots & z_3 \\
\vdots & \vdots & \ddots & \vdots \\
z_{n-1} & z_{n-1} & \ldots & z_{n-1} \\
z_n & z_n & \ldots & z_n
\end{bmatrix} \in \mathbb{R}^{n \times N_1}, 
\]

The matrix \( S_R \) contains \( N_2 N_3 \ldots N_n \) blocks \( B_R^x \), each of which contains \( N_1 \) directions. Thus, \( S_R \) contains \( N_1 N_2 \ldots N_n = N \) columns in total. Note that a block is identified using a vector \( x^0 \) containing \( n \) components labeled \( z_2, z_3, \ldots, z_n \). Next, we provide an example in \( \mathbb{R}^2 \) to get the reader accustomed to the notation.

**Example 3.1.** Select the reference point \( x^0 = [0, 0]^\top \) and the sample region \([0, 12] \times [0, 6] \). Then \( d = [12, 6]^\top = [\Delta_1, \Delta_2]^\top \). Suppose the longer side is divided into three equal parts, so \( N_1 = 3 \), and the shorter side is divided into two equal parts, so \( N_2 = 2 \). Thus, the side lengths of one partition are \( \Delta_1 = 12/3 = 4 \) and \( \Delta_2 = 6/2 = 3 \). The matrix \( S_R \) is given by

\[
S_R = \begin{bmatrix}
B_R^1 & B_R^2
\end{bmatrix},
\]

where

\[
B_R^1 = \begin{bmatrix}
12 & 0 & 0 \\
0 & 1 & 1 \\
0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
4 & 8 & 12 \\
3 & 3 & 3 \\
6 & 6 & 6
\end{bmatrix},
\]

\[
B_R^2 = \begin{bmatrix}
12 & 0 & 0 \\
0 & 1 & 1 \\
0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
4 & 8 & 12 \\
6 & 6 & 6
\end{bmatrix}.
\]

1 The process will be generalized later in this section to allow choosing any arbitrary point in each partition of \( R(x^0; d) \).
The sample points $x^j$ are obtained by setting $x^j = x^0 + S_R e_j$ for all $j \in \{1, 2, \ldots, 6\}$. We see that $x^1, x^2, x^3$ are associated with $B_R^1$ and $x_4, x_5, x_6$ are associated with $B_R^2$. Figure 1 illustrates the sample points built from $S_R$.

Figure 1: An example of a sample set built from a matrix $S_R$ in $\mathbb{R}^2$

The first goal of this paper is to find an expression for the GSG over $R(x^0; d)$ when the set of sample points forms a dense subset of $\mathbb{R}^n$, that is, as $N_i \to \infty$ for each $i$. As long as each $N_i$ tends to infinity (not necessarily at the same speed), we will show that an expression for the GSG over $R(x^0; d)$ can be found. Recall that the formula (1) to compute the GSG of $f$ at $x^0$ over $S$ is

$$\nabla_s f(x^0; S) = (S^\dagger)^\top \delta f(x^0; S).$$

When $S$ is full row rank, then the above formula can be written as

$$\nabla_s f(x^0; S) = (S^\top (SS^\top)^{-1})^\top \delta f(x^0; S) = (SS^\top)^{-\top} S \delta f(x^0; S). \quad (8)$$

Note that $S_R \in \mathbb{R}^{n \times N}$ as defined in (6) is full row rank whenever $N_i \geq 2$ for all $i \in \{1, 2, \ldots, n\}$. For the remainder of this section and in Section 4, we will assume $N_i \geq 2$ for all $i$. In the following lemma, we begin our investigation of (8) by finding an expression for the matrix $S_R S_R^\top$.

**Lemma 3.2.** Let $S_R \in \mathbb{R}^{n \times N}$ be defined as in (6). Then

$$S_R S_R^\top = \sum_{z_2=1}^{N_2} \sum_{z_3=1}^{N_3} \cdots \sum_{z_n=1}^{N_n} B_R^z (B_R^z)^\top = U_n \in \mathbb{R}^{n \times n},$$

where

$$[U_n]_{i,i} = N (N_i + 1)(2N_i + 1) \frac{\Delta^2}{6}, \quad i \in \{1, 2, \ldots, n\},$$

$$[U_n]_{i,j} = N (N_i + 1)(N_j + 1) \frac{\Delta_i \Delta_j}{4}, \quad i, j \in \{1, 2, \ldots, n\}, \quad i \neq j.$$
Proof. The proof is by induction on $n$. First, we prove the case $n = 2$. We have

$$\sum_{z_2=1}^{N_2} B_{R}^{z_2 \top} (B_{R}^{z_2})^\top = \left[ \begin{array}{c} \overline{A}_1 \\ 0 \\ \overline{A}_2 \end{array} \right] \left( \sum_{z_2=1}^{N_2} \left[ \begin{array}{ccc} 1 & 2 & \cdots \\ z_2 & z_2 & \cdots \\ \vdots & \vdots & \vdots \\ 0 & 0 & \cdots \\ N_1 & z_2 & 2 \end{array} \right] \right) \left[ \begin{array}{c} \overline{A}_1 \\ 0 \\ \overline{A}_2 \end{array} \right]$$

$$= \left[ \begin{array}{c} \overline{A}_1 \\ 0 \\ \overline{A}_2 \end{array} \right] \left( \sum_{z_2=1}^{N_2} \left[ \frac{N_1(N_1+1)(2N_1+1)}{2} \overline{A} \right] \left[ \begin{array}{c} z_2 \\ \overline{A} \end{array} \right] \right) \left[ \begin{array}{c} \overline{A}_1 \\ 0 \\ \overline{A}_2 \end{array} \right]$$

$$= \left[ \begin{array}{c} \overline{A}_1 \\ 0 \\ \overline{A}_2 \end{array} \right] \left( \sum_{z_2=1}^{N_2} \left[ \frac{N_1N_2(N_1+1)(N_2+1)}{6} \right] \left[ \begin{array}{c} z_2 \\ \overline{A} \end{array} \right] \right) \left[ \begin{array}{c} \overline{A}_1 \\ 0 \\ \overline{A}_2 \end{array} \right]$$

$$= N_1N_2 \left[ \frac{(N_1+1)(N_1+2)}{6} \overline{A} \right] \left[ \begin{array}{c} \overline{A} \end{array} \right] = U_2.$$  

Now, let $n = k$ for some $k \in \mathbb{N} \setminus \{1\}$. For clarity, we will write $B_{R}^{z_k \top}$ instead of $B_{R}^{z_k}$ to make it clear that the last component in the vector of indices $\overline{z}$ is $z_k$. Suppose that the induction hypothesis is true for $n = k$. We want to show that it is true for $n = k + 1$. We have

$$\sum_{z_2=1}^{N_2} \cdots \sum_{z_{k+1}=1}^{N_{k+1}} B_{R}^{z_2 \top} \cdots (B_{R}^{z_{k+1} \top}) = \sum_{z_2=1}^{N_2} \cdots \sum_{z_{k+1}=1}^{N_{k+1}} \left[ \overline{A} \right] B_{R}^{z_2 \top} \cdots \left[ \overline{A} \right] (B_{R}^{z_{k+1} \top})$$

$$= \sum_{z_2=1}^{N_2} \cdots \sum_{z_{k+1}=1}^{N_{k+1}} \left[ \overline{A} \right] B_{R}^{z_2 \top} \cdots \left[ \overline{A} \right] (B_{R}^{z_{k+1} \top})$$

Now, we compute the last sum of the k-tuple sum. We have

$$\sum_{z_{k+1}=1}^{N_{k+1}} \left[ \begin{array}{c} B_{R}^{z_{k+1} \top} \end{array} \right] \left[ \begin{array}{c} \overline{A} \end{array} \right] \left( B_{R}^{z_{k+1}} \right) = \sum_{z_{k+1}=1}^{N_{k+1}} \left[ \begin{array}{c} B_{R}^{z_{k+1} \top} \end{array} \right] \left[ \begin{array}{c} \overline{A} \end{array} \right] \left( B_{R}^{z_{k+1}} \right)$$

$$= \left[ \begin{array}{c} \overline{A} \end{array} \right] \left[ \begin{array}{c} \overline{A} \end{array} \right] \left( B_{R}^{z_{k+1}} \right)$$

Substituting (10) into (9), we obtain

$$\sum_{z_2=1}^{N_2} \cdots \sum_{z_{k+1}=1}^{N_{k+1}} B_{R}^{z_2 \top} \cdots (B_{R}^{z_{k+1} \top}) = \sum_{z_2=1}^{N_2} \cdots \sum_{z_{k+1}=1}^{N_{k+1}} \left[ \overline{A} \right] B_{R}^{z_2 \top} \cdots \left[ \overline{A} \right] (B_{R}^{z_{k+1} \top})$$

(11) Let us find an expression for each of the four blocks in (11). Using the induction assumption, we have

$$\sum_{z_2=1}^{N_2} \cdots \sum_{z_{k+1}=1}^{N_{k+1}} B_{R}^{z_2 \top} \cdots (B_{R}^{z_{k+1} \top}) = N_{k+1}U_k.$$
Therefore, using (5), we obtain
\[
\sum_{z_2=1}^{N_2} \sum_{z_3=1}^{N_3} \cdots \sum_{z_k=1}^{N_k} \Delta_{k+1} \frac{N_{k+1}(N_{k+1} + 1)}{2} B_R^{zz} 1_{N_1} = \Delta_{k+1} \frac{N_{k+1}(N_{k+1} + 1)}{2} \begin{bmatrix}
\Delta_1 N_1^2 \frac{(N_1+1)}{2} N_3 \cdots N_k \\
\Delta_2 N_2 \frac{(N_2+1)}{2} N_3 N_4 \cdots N_k \\
\vdots \\
\Delta_k N_1 N_2 \cdots N_{k-1} N_k \frac{(N_k+1)}{2} 
\end{bmatrix}
\]

The last block is
\[
\sum_{z_2=1}^{N_2} \sum_{z_3=1}^{N_3} \cdots \sum_{z_k=1}^{N_k} \sum_{i=1}^{N_i} \Delta_{k+1} N_i N_{k+1} + 1 \frac{(2N_{k+1} + 1)}{6} = N \Delta_{k+1} \frac{(N_{k+1} + 1)(2N_{k+1} + 1)}{6}.
\]

Hence, (11) is equal to \( U_{k+1} \). By the principle of induction, the claim is true for all \( n \in \mathbb{N} \setminus \{1\} \).

The GSG uses the inverse of \( (S_R S_R^T)^T \). The following lemma finds an expression for this inverse.

Lemma 3.3. Let \( S_R \in \mathbb{R}^{n \times N} \) be defined as in (6). Then
\[
(S_R S_R^T)^{-T} = \frac{12}{N} \left( E - \frac{3}{1 + 3s} yy^T \right),
\]

where
\[
E = \text{diag} \left[ \frac{1}{(N_1^2-1) \Delta_1^2} \cdots \frac{1}{(N_n^2-1) \Delta_n^2} \right] \in \mathbb{R}^{n \times n}, \ s = \sum_{i=1}^{n} \frac{N_i + 1}{N_i - 1}, \text{ and } y = \left[ \frac{1}{(N_1-1) \Delta_1} \cdots \frac{1}{(N_n+1) \Delta_n} \right]^T \in \mathbb{R}^n.
\]

Proof. Note that for all \( n \in \mathbb{N} \setminus \{1\} \) the symmetric matrix \( U_n \) can be written as
\[
U_n = N \left( \text{diag} \left[ \frac{(N_1-1)(N_1+1)}{12} \Delta_1^2 \cdots \frac{(N_n-1)(N_n+1)}{12} \Delta_n^2 \right] + \left[ \frac{(N_1+1)}{2} \Delta_1 \cdots \frac{(N_n+1)}{2} \Delta_n \right]^T \right).
\]

Let \( \bar{d} = \left[ \frac{(N_1+1)}{2} \Delta_1 \cdots \frac{(N_n+1)}{2} \Delta_n \right]^T \in \mathbb{R}^n \) and \( \bar{D} = \text{diag} \left[ \frac{N_1^2-1}{12} \Delta_1^2 \cdots \frac{N_n^2-1}{12} \Delta_n^2 \right] \in \mathbb{R}^{n \times n} \). Then
\[
U_n = N(\bar{D} + \bar{d} \bar{d}^T).
\]

Therefore, using (5), we obtain
\[
(U_n)^{-T} = (U_n^T)^{-1} = U_n^{-1}
\]
\[
= \frac{1}{N} \left( \bar{D}^{-1} - \bar{D}^{-1} \bar{d} \bar{d}^T \bar{D}^{-1} \right).
\]
The denominator of the second term in (12) is

\begin{align*}
1 + d^\top \tilde{D}^{-1} d &= 1 + \left[ \frac{(N_1+1)}{2} \Delta_1, \ldots, \frac{(N_n+1)}{2} \Delta_n \right] \text{diag} \left[ \frac{12}{(N_1^2-1)\Delta_1^2}, \ldots, \frac{12}{(N_n^2-1)\Delta_n^2} \right] \\
&= 1 + \left[ \frac{6}{(N_1-1)\Delta_1}, \ldots, \frac{6}{(N_n-1)\Delta_n} \right] \\
&= 1 + 3 \sum_{i=1}^{n} \frac{N_i+1}{N_i-1} = 1 + 3s.
\end{align*}

The numerator of the second term in (12) is

\[ \tilde{D}^{-1} d \tilde{D}^{-1} = \left[ \frac{6}{(N_1-1)\Delta_1}, \ldots, \frac{6}{(N_n-1)\Delta_n} \right] \text{diag} \left[ \frac{6}{(N_1-1)\Delta_1}, \ldots, \frac{6}{(N_n-1)\Delta_n} \right] = 36yy^\top. \]

Since \( \tilde{D}^{-1} = 12E \), we get

\[ (U_n)^\top = \frac{1}{N} \left( \tilde{D}^{-1} - \frac{36}{1+3s} yy^\top \right) = \frac{12}{N} \left( E - \frac{3}{1+3s} yy^\top \right). \]

Using the previous lemma, we can now provide an expression for the transpose of the Moore–Penrose inverse \( S_R^\top \).

**Corollary 3.4** (The matrix \( (S_R^\top)^\top \)). Let \( S_R \in \mathbb{R}^{n \times N} \) be defined as in (6). Then

\[ (S_R^\top)^\top = \frac{12}{N} \left( E - \frac{3}{1+3s} yy^\top \right) S_R, \]

where \( E = \text{diag} \left[ \frac{1}{(N_1^2-1)\Delta_1^2}, \ldots, \frac{1}{(N_n^2-1)\Delta_n^2} \right] \in \mathbb{R}^{n \times n} \), \( y = \left[ \frac{1}{(N_1-1)\Delta_1}, \ldots, \frac{1}{(N_n-1)\Delta_n} \right]^\top \in \mathbb{R}^n \), and the scalar \( s = \sum_{i=1}^{n} \frac{N_i+1}{N_i-1} \).

**Proof.** Since \( S_R \) has full row rank, we have

\[ (S_R^\top)^\top = (S_R^\top (S_R S_R^\top)^{-1})^\top = (S_R S_R^\top)^{-1} S_R = \frac{12}{N} \left( E - \frac{3}{1+3s} yy^\top \right) S_R, \]

by Lemma 3.3.

In the following proposition, we investigate the limit of \( N(S_R S_R^\top)^{-1} \) when all \( N_i \) go to infinity. Applying these limits, the sample points contained in \( R(x_0, d) \) form a dense grid of \( \mathbb{R}^n \). To make the notation compact, we write \( \lim_{N \to \infty} \) to represent the limit as \( N_1 \to \infty, N_2 \to \infty, \ldots, N_n \to \infty \). The reason we are interested in the limit of \( N(S_R S_R^\top)^{-1} \) is that, assuming the limits exist, we may write

\[ \lim_{N \to \infty} \nabla f(x^0; S_R) = \lim_{N \to \infty} (S_R^\top)^\top \delta_f(x^0; S_R) \]

\[ = \lim_{N \to \infty} N \frac{1}{N} (S_R S_R^\top)^{-1} \Delta \frac{1}{N} N S_R \delta_f(x^0, S_R) \]

\[ = \frac{1}{\Delta} \left[ \lim_{N \to \infty} N (S_R S_R^\top)^{-1} \right] \left[ \lim_{N \to \infty} \frac{1}{N} N S_R \delta_f(x^0; S_R) \right], \]

where \( \Delta = \Delta_1 \Delta_2 \cdots \Delta_n \). So if we can show that the two limits in (13) exist, then we have found the limit of the GSG over a dense hyperrectangle. Note that the term \( \Delta \) remains inside the second limit, as we will show that the expression in the second limit is a \( n \)-tuple Riemann sum.
**Proposition 3.5.** Let \( S_R \) be defined as in (6). Then
\[
\lim_{N \to \infty} N(S_R S_R^\top)^{-\top} = L_n \in \mathbb{R}^{n \times n}
\]
where
\[
[L_n]_{i,i} = \frac{12(3n - 2)}{\Delta_i^2 (3n + 1)}, \quad i \in \{1, 2, \ldots, n\},
\]
\[
[L_n]_{i,j} = \frac{-36}{\Delta_i \Delta_j (3n + 1)}, \quad i, j \in \{1, 2, \ldots, n\}, i \neq j.
\]

**Proof.** By Corollary 3.4 we have
\[
\lim_{N \to \infty} N(S_R S_R^\top)^{-\top} = \lim_{N \to \infty} N \frac{12}{N} \left( E - \frac{3}{1 + 3s} yy^\top \right) = \lim_{N \to \infty} 12 \left( E - \frac{3}{1 + 3s} yy^\top \right),
\]
where \( E = \text{diag} \left[ \frac{1}{(N_1 - 1)\Delta_1}, \ldots, \frac{1}{(N_n - 1)\Delta_n} \right] \in \mathbb{R}^{n \times n} \), \( y = \left[ \frac{1}{(N_1 - 1)\Delta_1}, \ldots, \frac{1}{(N_n - 1)\Delta_n} \right]^\top \in \mathbb{R}^n \) and the scalar \( s = \sum_{i=1}^n \frac{N_i t_i}{N_i - 1} \). We show that this converges componentwise to \( L \).

We begin with the diagonal entries. Applying \( \Delta_i^2 = N_i^2 / \Delta_i^2 \), note that
\[
[yy^\top]_{i,i} = \frac{1}{(N_i - 1)^2 \Delta_i} = \frac{N_i^2}{(N_i - 1)^2 \Delta_i^2}.
\]
Substituting this and \( \Delta_i^2 = N_i^2 / \Delta_i^2 \) into the definition of \( E \) yields
\[
\lim_{N \to \infty} 12 \left( E - \frac{3}{1 + 3s} yy^\top \right)_{i,i} = 12 \lim_{N \to \infty} \frac{N_i^2}{(N_i - 1)(N_i + 1) \Delta_i^2} = \frac{3}{1 + 3 \sum_{i=1}^n \frac{N_i t_i}{N_i - 1}} \left( N_i - 1 \right)^2 \Delta_i^2
\]
\[
= 12 \left( \frac{1}{\Delta_i^2} - \frac{3}{(1 + 3n) \Delta_i^2} \right)
\]
\[
= \frac{12(3n - 2)}{\Delta_i^2 (3n + 1)}, \quad i \in \{1, 2, \ldots, n\}.
\]
Similarly, the off-diagonal entries of the matrix in (14) are given by
\[
\lim_{N \to \infty} \left[ \frac{-36}{1 + 3 \sum_{i=1}^n \frac{N_i t_i}{N_i - 1} (N_i - 1)(N_j - 1) \Delta_i \Delta_j} \right]_{i,j} = \frac{-36}{(1 + 3n) \Delta_i \Delta_j}, \quad i, j \in \{1, 2, \ldots, n\}, i \neq j.
\]

### 3.2 Generalization: using an arbitrary point in each partition

We now generalize \( S_R \) to a matrix that allows choosing an arbitrary point in each partition of the sample region, not necessarily the right endpoint of each partition. The matrix containing all directions used to obtain an arbitrary sample point in each partition will be denoted by \( S \). Let \( N = N_1 N_2 \cdots N_n \) and \( \mathcal{D} = \text{diag} [\Delta_1 \Delta_2 \cdots \Delta_n] \in \mathbb{R}^{n \times n} \) where \( \Delta_i = \Delta_i / N_i \) for all \( i \). The matrix \( S \in \mathbb{R}^{n \times N} \) can be written as a block matrix in the following way:
\[
S = \begin{bmatrix} B_{1,1}^{1,1} & B_{1,1}^{1,2} & \cdots & B_{1,n}^{1,N_n} \\ \vdots & \ddots & \vdots & \vdots \\ B_{n,1}^{N_1,1} & \cdots & B_{n,n}^{N_1,N_n} \end{bmatrix}
\]
where
\[
B_{R}^R = B_{R}^{R} - B_{M}^{R} \in \mathbb{R}^{n \times N_1}.
\]
The block \( B_{R}^{R} \in \mathbb{R}^{n \times N_1} \) is defined in (7) and the block \( B_{M}^{R} \in \mathbb{R}^{n \times N_1} \) is
\[
B_{M}^{R} = \mathcal{D} B_{M}^{R}.
\]
where all entries of \( B_M^\# \) are in \([0, 1]\). Let \( S_M \in \mathbb{R}^{n \times N} \) be defined as
\[
S_M = \begin{bmatrix}
B_M^{1,1,\ldots,1,1} & B_M^{1,1,\ldots,1,2} & \cdots & B_M^{N_2,N_3,\ldots,N_{n-1},N_n}
\end{bmatrix}.
\]

Then \( S \) can be written as \( S = S_R - S_M \). Let us provide an example of \( S \) in \( \mathbb{R}^2 \).

**Example 3.6.** Consider the same sample region as Example 3.1. Select the ‘arbitrary’ points \( [2, 2]^T, [5, 1]^T, [8, 3]^T, [1, 1]^T, [6, 4.5]^T, \) and \( [12, 6]^T \). These are visualized in Figure 2.

**Figure 2:** An example of sample set built from a matrix \( S \) in \( \mathbb{R}^2 \)

The **matrix** \( S \in \mathbb{R}^{2 \times 6} \) is given by
\[
S = \begin{bmatrix} B^1 & B^2 \end{bmatrix},
\]
where
\[
B^1 = B^1_R - B^1_M = \begin{bmatrix} 12 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 12 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 8 & 12 \\ 3 & 3 & 3 \end{bmatrix} - \begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 5 & 8 \\ 2 & 1 & 3 \end{bmatrix}.
\]

and
\[
B^2 = B^2_R - B^2_M = \begin{bmatrix} 12 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 3 \end{bmatrix} - \begin{bmatrix} 12 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 8 & 12 \\ 6 & 6 & 6 \end{bmatrix} - \begin{bmatrix} 4 & 2 & 0 \\ 3 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 6 & 12 \\ 3 & 3 & 2 \end{bmatrix}.
\]

The sample points \( x^j \) are built by setting \( x^j = x^0 + S e_j \) for all \( j \in \{1, 2, \ldots, 6\} \). We see that \( x^1, x^2, x^3 \) are associated with \( B^1 \) and \( x^4, x^5, x^6 \) are associated with \( B^2 \).

The following proposition generalizes Proposition 3.5 by considering \( S \) in place of \( S_R \).

**Proposition 3.7.** Let \( S \in \mathbb{R}^{n \times N} \) be defined as in (15). Then
\[
\lim_{N \to \infty} N (SS^T)^{-T} = L_n \in \mathbb{R}^{n \times n},
\]
where \( L_n \) is defined as in Proposition 3.5.
Proof. We have

\[
\lim_{N \to \infty} N(SS^\top)^{-T} = \lim_{N \to \infty} \left( \frac{1}{N} SS^\top \right)^{-T}.
\]

Note that the inverse of \( SS^\top \) is well-defined, since \( S \) is full row rank whenever \( N_i \in \mathbb{N} \setminus \{1\} \) for all \( i \). It follows that \( SS'^\top \) is full rank, so the inverse of \((SS'^\top)^\top\) exists. Since the inverse of \((SS'^\top)^\top\) is a continuous function with respect to \( N \), we may take the limit inside the inverse. We obtain

\[
\lim_{N \to \infty} \left( \frac{1}{N} SS^\top \right)^{-T} = \left( \lim_{N \to \infty} \frac{1}{N} SS^\top \right)^{-T} = \left( \lim_{N \to \infty} \frac{1}{N} S R S_R^\top - \lim_{N \to \infty} \frac{1}{N} S M S_M^\top - \lim_{N \to \infty} \frac{1}{N} S R S_M^\top + \lim_{N \to \infty} \frac{1}{N} S M S_M^\top \right)^{-T}.
\tag{16}
\]

Now, we show that \(\lim_{N \to \infty} \frac{1}{N} S M S_M^\top \), \(\lim_{N \to \infty} \frac{1}{N} S R S_R^\top \), and \(\lim_{N \to \infty} \frac{1}{N} S R S_M^\top \) are equal to the \( n \times n \) zero matrix.

We begin with showing \(\lim_{N \to \infty} \frac{1}{N} S M S_M^\top = 0_{n \times n}\). Let \( D = \mathbb{R}^{n \times n} \) be the diagonal matrix with entries \( \Delta_1/N_1, \ldots, \Delta_n/N_n \). We have

\[
\frac{1}{N} S M S_M^\top = \frac{1}{N} \sum_{i=2}^{N_2} \cdots \sum_{n=1}^{N_n} B_M^x B_M^x = \frac{1}{N} D \left( \sum_{i=2}^{N_2} \cdots \sum_{n=1}^{N_n} \bar{B}_M^x (\bar{B}_M^x)^\top \right) D.
\tag{17}
\]

Since all entries in the matrix \( \bar{B}_M^x \) are contained in \([0, 1]\), the \((n - 1)\)-tuple sum in \(17\) is bounded componentwise below by the matrix \( 0_{n \times n} \) and above by

\[
\sum_{i=2}^{N_2} \cdots \sum_{n=1}^{N_n} \bar{B}_M^x (\bar{B}_M^x)^\top \leq \sum_{i=2}^{N_2} \cdots \sum_{n=1}^{N_n} 1_n \mathbf{1}_n^\top = N_1 \mathbf{1}_n \mathbf{1}_n^\top = N_1 \mathbf{1}_n \mathbf{1}_n^\top.
\]

It follows that componentwise

\[
\lim_{N \to \infty} \frac{1}{N} D \left( \sum_{i=2}^{N_2} \cdots \sum_{n=1}^{N_n} \bar{B}_M^x (\bar{B}_M^x)^\top \right) D \leq \lim_{N \to \infty} \frac{1}{N} D N_1 \mathbf{1}_n \mathbf{1}_n^\top \leq 0_{n \times n}.
\]

By the Squeeze Theorem, \(\lim_{N \to \infty} \frac{1}{N} S M S_M^\top = 0_{n \times n}\).

Now, we show that \(\lim_{N \to \infty} \frac{1}{N} S R S_R^\top = 0_{n \times n}\). We have

\[
\frac{1}{N} S R S_R^\top = \frac{1}{N} D \left( \sum_{i=2}^{N_2} \cdots \sum_{n=1}^{N_n} \bar{B}_R^x (\bar{B}_M^x)^\top \right) D.
\tag{18}
\]

\[
\frac{1}{N} S R S_R^\top = \frac{1}{N} D \left( \sum_{i=2}^{N_2} \cdots \sum_{n=1}^{N_n} \bar{B}_R^x (\bar{B}_M^x)^\top \right) D.
\tag{18}
\]
The \((n - 1)\)-tuple sum in (18) is bounded componentwise below by \(0 \times \ldots \times 0\). A componentwise upper bound for (18) is
\[
\frac{1}{N} \mathcal{D} \left( \sum_{z_2=1}^{N_2} \cdots \sum_{z_n=1}^{N_n} \hat{B}_{R}^{z} (\hat{M}_{R}^{z})^{\top} \right) \mathcal{D} \leq \frac{1}{N} \mathcal{D} \left( \sum_{z_2=1}^{N_2} \cdots \sum_{z_n=1}^{N_n} \hat{B}_{R}^{z} 1_{N-1}^{\top} \right) \mathcal{D}
\]
\[
= \frac{1}{N} \mathcal{D} \left( \frac{N}{2} \begin{bmatrix} N_1 + 1 & N_1 + 1 & \cdots & N_1 + 1 \\ N_2 + 1 & N_2 + 1 & \cdots & N_2 + 1 \\ \vdots & \vdots & \ddots & \vdots \\ N_n + 1 & N_n + 1 & \cdots & N_n + 1 \end{bmatrix} \right) \mathcal{D}
\]
\[
= \frac{1}{2} \mathcal{D} \left( \begin{bmatrix} \Delta_1(N_1+1) & \Delta_1 \Delta_2(N_1+1) & \cdots & \Delta_1 \Delta_n(N_1+1) \\ \Delta_1 \Delta_2(N_2+1) & \Delta_2^2(N_2+1) & \cdots & \Delta_2 \Delta_n(N_2+1) \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_1 \Delta_n(N_n+1) & \Delta_2 \Delta_n(N_n+1) & \cdots & \Delta_n^2(N_n+1) \end{bmatrix} \right) \mathcal{D}.
\]
It follows that \(\lim_{N \to \infty} S_R S_M^\top \leq 0 \times \ldots \times 0\), and by the Squeeze Theorem, \(\lim_{N \to \infty} S_R S_M^\top = 0 \times \ldots \times 0\). As \(S_M S_R^\top = (S_R S_M^\top)^\top\), we also have \(\lim_{N \to \infty} S_M S_R^\top = 0 \times \ldots \times 0\). Thus, (16) reduces to
\[
\left( \lim_{N \to \infty} \frac{1}{N} S_R S_R^\top - \lim_{N \to \infty} \frac{1}{N} S_M S_R^\top - \lim_{N \to \infty} \frac{1}{N} S_R S_M^\top + \lim_{N \to \infty} \frac{1}{N} S_M S_M^\top \right)^{-\top} = \left( \lim_{N \to \infty} \frac{1}{N} S_R S_R^\top \right)^{-\top} = \lim_{N \to \infty} N (S_R S_R^\top)^{-\top} = L_n
\]
by Proposition 3.5.

The previous proposition gives an expression for the first limit in (13). The following theorem gives the limit of the product \(\frac{\tan}{N} \Delta S \delta f(x^0; S)\), the second limit in (13), as a multiple integral over \(R(0;d)\).

**Proposition 3.8.** Let \(f \in C^0\) on an open domain containing \(R(x^0;d) \subseteq \text{dom} f\). Let \(d = [\Delta_1 \Delta_2 \cdots \Delta_n]^\top > 0\), \(x = [x_1 \cdots x_n]^\top\) and let \(S \in \mathbb{R}^{n \times N}\) be defined as in (15). Then
\[
\lim_{N \to \infty} \frac{\Delta}{N} S \delta f(x^0; S) = T_n \in \mathbb{R}^n,
\]
where
\[
[T_n]_i = \int_{R(0;d)} x_i (f(x^0 + x) - f(x^0)) \, dx, \quad i \in \{1, 2, \ldots, n\}.
\]

**Proof.** We have
\[
\lim_{N \to \infty} \frac{\Delta}{N} S \delta f(x^0; S) = \lim_{N \to \infty} \frac{\Delta}{N} \left[ B^{1,1, \ldots, 1} \cdots B^{N_2, N_3, \ldots, N_n} \right] \delta f(x^0; B^{1,1, \ldots, 1})
\]
\[
= \lim_{N \to \infty} \sum_{z_2=1}^{N_2} \sum_{z_3=1}^{N_3} \cdots \sum_{z_n=1}^{N_n} \frac{\Delta}{N} B^{z} \delta f(x^0; B^{z}).
\]
(19)

Note that \(\frac{\Delta}{N}\) is the volume of one partition of \(R(x^0;d)\). Recall that
\[
\delta f(x^0; B^{z}) = [f(x^0 + B^{z} e_1) - f(x^0) \cdots f(x^0 + B^{z} e_N) - f(x^0)]^\top \in \mathbb{R}^{N_1}.
\]
The matrix $B^T$ has dimension $n \times N_1$ so (19) is the limit of a vector in $\mathbb{R}^n$. Since $f \in C^0$ on an open domain containing $R(x^0; d)$, (19) is a vector of $n$ definite integrals:

$$
\lim_{N \to \infty} \frac{N_2}{N_1} \sum_{x_2=1}^{N_2} \sum_{x_3=1}^{N_2} \cdots \sum_{x_n=1}^{N_2} \sum_{x_2=1}^{N_2} \sum_{x_3=1}^{N_2} \cdots \sum_{x_n=1}^{N_2} \frac{\Delta}{N} B^T \delta f(x^0; B^T) = \left[ \int_{R(0,d)} x_1 (f(x^0 + x) - f(x^0)) \, dx \\
\int_{R(0,d)} x_2 (f(x^0 + x) - f(x^0)) \, dx \\
\vdots \\
\int_{R(0,d)} x_n (f(x^0 + x) - f(x^0)) \, dx \right] = T_n.
$$

Now we are ready for our first main result: the limiting behaviour of the GSG at $x^0$ over the dense grid $R(x^0; d)$. The result of Theorem 3.9 below is expressed as a multiple definite integral.

**Theorem 3.9** (Limiting behaviour of the GSG of $f$ at $x^0$ over $S$). Let $f \in C^0$ on an open domain containing $R(x^0; d) \subseteq \text{dom } f$, $d = [\Delta_1 \, \Delta_2 \, \cdots \, \Delta_n]^{\top}$ > 0. Let $S \in \mathbb{R}^{n \times N}$ be defined as in (15). Then

$$
\lim_{N \to \infty} \nabla_S f(x^0; S) = \Delta^{-1} L_n T_n,
$$

where the entries of $L_n \in \mathbb{R}^{n \times n}$ are given by

$$
[L_n]_{i,i} = \frac{12(3n - 2)}{\Delta_i^2 (3n + 1)}, \quad i \in \{1, 2, \ldots, n\},
$$

$$
[L_n]_{i,j} = \frac{-36}{\Delta_i \Delta_j (3n + 1)}, \quad i, j \in \{1, 2, \ldots, n\}, i \neq j,
$$

and the entries of $T_n \in \mathbb{R}^n$ are given by

$$
[T_n]_i = \int_{R(0,d)} x_i (f(x^0 + x) - f(x^0)) \, dx, \quad i \in \{1, 2, \ldots, n\}.
$$

**Proof.** We have

$$
\lim_{N \to \infty} \nabla_S f(x^0; S) = \lim_{N \to \infty} (S^T)^{-1} \delta f(x^0; S)
$$

$$
= \lim_{N \to \infty} (SS^T)^{-1} S \delta f(x^0; S)
$$

$$
= \lim_{N \to \infty} \frac{N}{\Delta} (SS^T)^{-1} S \delta f(x^0; S).
$$

By Proposition 3.7,

$$
\lim_{N \to \infty} N(SS^T)^{-1} = L_n \in \mathbb{R}^{n \times n}.
$$

By Proposition 3.8,

$$
\lim_{N \to \infty} \frac{\Delta}{N} S \delta f(x^0; S) = T_n \in \mathbb{R}^n.
$$

Therefore,

$$
\lim_{N \to \infty} \nabla_S f(x^0; S) = \frac{1}{\Delta} \lim_{N \to \infty} N(SS^T)^{-1} \lim_{N \to \infty} \frac{\Delta}{N} S \delta f(x^0; S)
$$

$$
= \Delta^{-1} L_n T_n.
$$

**Remark 3.10.** The previous result also agrees with what was found in [BHJB21] for $n = 1$. Indeed, in $\mathbb{R}$ [BHJB21, Theorem 4.1] found that

$$
\lim_{N \to \infty} \nabla_S f(x^0; S) = \frac{3}{\Delta^2} \int_0^\Delta x(f(x^0 + x) - f(x^0)) \, dx,
$$

which is what (20) becomes when $n = 1$. 

13
In this section, we use the results obtained thus far to formulate an error bound ad infinitum that does not depend on the number of points used in the hyperrectangle \( R(x^0; d) \). Let \( f : \mathbb{R}^2 \rightarrow \mathbb{R} : x \mapsto x_1^2 + x_2^2 \) and the reference point \( x^0 = \begin{bmatrix} 3 & 1 \end{bmatrix}^\top \). Consider the sample region \( R(x^0; [1 \ 1]^\top) \), a square of side length 1. By Theorem 3.9, we know that

\[
\lim_{N \to \infty} \nabla_s f(x^0; S) = \frac{1}{\Delta} L_2 T_2
\]

\[
= \frac{1}{(1)(1)} \begin{bmatrix} \frac{48}{7} & \frac{36}{7} \\ \frac{36}{7} & \frac{48}{7} \end{bmatrix} \begin{bmatrix} \int_0^1 \int_0^1 x_1 \left( (3 + x_1)^2 + (1 + x_2)^2 - 10 \right) dx_1 dx_2 \\ \int_0^1 \int_0^1 x_2 \left( (3 + x_1)^2 + (1 + x_2)^2 - 10 \right) dx_1 dx_2 \end{bmatrix}
\]

\[
= \begin{bmatrix} \frac{48}{7} & \frac{36}{7} \\ \frac{36}{7} & \frac{48}{7} \end{bmatrix} \begin{bmatrix} 11 \frac{7}{7} \\ 1231 \frac{7}{7} \end{bmatrix} \approx \begin{bmatrix} 6.71 \\ 2.71 \end{bmatrix}
\]

The absolute error is approximately \( \left\| \begin{bmatrix} 6.71 \\ 2.71 \end{bmatrix} - \begin{bmatrix} 6 \\ 2 \end{bmatrix} \right\| \approx 1.01 \).

Now that we have an expression for the GSG as the number of points tends to infinity, the next step is to define an error bound ad infinitum. This is the focus of the next section.

4 Error bound ad infinitum of GSG over a hyperrectangle

In this section, we use the results obtained thus far to formulate an error bound ad infinitum that does not depend on the number of points used in the hyperrectangle \( R(x^0; d) \). To obtain an error bound ad infinitum for the GSG at \( x^0 \) over \( R(x^0; d) \), we require the function \( f \) to be \( C^2 \) on an open domain containing \( R(x^0; d) \). We will write \( f(x^0 + x) \) as the first-order Taylor expansion. That is

\[
f(x^0 + x) = f(x^0) + \nabla f(x^0)^\top x + R_1(x^0 + x), \quad \text{where} \quad R_1(x^0 + x) = \frac{1}{2} x^\top \nabla^2 f(\xi)x,
\]

for \( \xi \in \mathbb{R}^n \) between \( x^0 \) and \( x^0 + x \).

By rewriting \( f(x^0 + x) \) in the form of (22), the components of the vector \( T_n \) defined in (21) can be written as

\[
[T_n]_i = \int_{R(0; d)} x_i \left( f(x^0 + x) - f(x^0) \right) dx
\]

\[
= \int_{R(0; d)} x_i \left( \nabla f(x^0)^\top x + R_1(x^0 + x) \right) dx
\]

\[
= \int_{R(0; d)} x_i \nabla f(x^0)^\top x dx + \int_{R(0; d)} x_i R_1(x^0 + x) dx.
\]

for \( i \in \{1, 2, \ldots, n\} \). Let \( v \in \mathbb{R}^n \) be the vector defined by

\[
v_i = \int_{R(0; d)} x_i \nabla f(x^0)^\top x dx, \quad i \in \{1, 2, \ldots, n\},
\]

and \( w \in \mathbb{R}^n \) be the vector defined by

\[
w_i = \int_{R(0; d)} x_i R_1(x^0 + x) dx, \quad i \in \{1, 2, \ldots, n\}.
\]

Then the expression for \( \lim_{N \to \infty} \nabla_s f(x^0; S) \) given in (20) can be expressed as

\[
\lim_{N \to \infty} \nabla_s f(x^0; S) = \frac{1}{\Delta} L_n v + \frac{1}{\Delta} L_n w.
\]

We begin by showing that the first term in (24) is equal to \( \nabla f(x^0) \).
Lemma 4.1. Let $L_n \in \mathbb{R}^{n \times n}$ be defined by
\[
[L_n]_{i,i} = \frac{12(3n-2)}{\Delta_i^2(3n+1)}, \quad i \in \{1, 2, \ldots, n\},
\]
\[
[L_n]_{i,j} = \frac{-36}{\Delta_i \Delta_j (3n+1)}, \quad i, j \in \{1, 2, \ldots, n\}, i \neq j.
\]

Let $v \in \mathbb{R}^n$ be defined by
\[
v_i = \int_{R(0,d)} x_i \nabla f(x^0)^\top x \, dx, \quad i \in \{1, 2, \ldots, n\}.
\]

Then $\frac{1}{\Delta} L_n v = \nabla f(x^0)$.

Proof. First, we find an expression for $v_i, i \in \{1, 2, \ldots, n\}$. To make notation tighter, let $g = \nabla f(x^0)$. We have
\[
v_i = \int_{R(0,d)} x_i g_i^\top x \, dx
\]
\[
= \int_{R(0,d)} x_i \left( \sum_{j=1}^{n} g_j x_j \right) \, dx
\]
\[
= \int_{R(0,d)} x_i^2 g_i \, dx + \sum_{j 
eq i} \int_{R(0,d)} x_i x_j g_j \, dx
\]
\[
= \frac{\Delta_i^2}{3} \Delta_i g_i + \sum_{j 
eq i} \frac{\Delta_i^2}{2} \frac{\Delta_j^2}{2} \frac{\Delta_i \Delta_j}{\Delta} g_j
\]
\[
= \frac{\Delta_i^2}{3} g_i + \sum_{j 
eq i} \frac{\Delta_i \Delta_j}{4} g_j = \frac{\Delta_i}{12} \left( 4\Delta_i g_i + 3 \sum_{j 
eq i} \Delta_j g_j \right) .
\]

Let $s = \sum_{j=1}^{n} \Delta_j g_j$. Then
\[
v_i = \frac{\Delta \Delta_i}{12} \left( 4\Delta_i g_i + 3 \sum_{j 
eq i} \Delta_j g_j \right) = \frac{\Delta \Delta_i}{12} (\Delta_i g_i + 3s).
\]

Now, let us compute $\frac{1}{\Delta} L_n v$. Let $D \in \mathbb{R}^{n \times n}$ be the diagonal matrix with entries $\Delta_1, \Delta_2, \ldots, \Delta_n$. Note that
\[
L_n = \frac{12}{3n+1} D^{-1} L'_n D^{-1},
\]
where $[L'_n]_{i,i} = 3n - 2$ for all $i \in \{1, 2, \ldots, n\}$ and $[L'_n]_{i,j} = -3$ for all $i, j \in \{1, \ldots, n\}, i \neq j$. Let the vector $d = [\Delta_1 \Delta_2 \cdots \Delta_n]^\top \in \mathbb{R}^n$. The vector $v$ can be written as
\[
v = \frac{\Delta}{12} (D^2 g + 3sd).
\]

We obtain
\[
\frac{1}{\Delta} L_n v = \frac{12}{\Delta(3n+1)} D^{-1} L'_n D^{-1} \frac{\Delta}{12} (D^2 g + 3sd)
\]
\[
= \frac{1}{3n+1} D^{-1} L'_n D^{-1} (D^2 g + 3sd)
\]
\[
= \frac{1}{3n+1} D^{-1} L'_n D g + \frac{3s}{3n+1} D^{-1} L'_n D^{-1} d.
\]
The first term in (25) is equal to
\[
\frac{1}{3n+1} D^{-1} L_n' D g = \frac{1}{3n+1} D^{-1} \left[ \begin{array}{c} (3n-2) \Delta_1 g_1 - 3 \sum_{j \neq 1} \Delta_j g_j \\ \vdots \\ (3n-2) \Delta_n g_n - 3 \sum_{j \neq n} \Delta_j g_j \end{array} \right] 
\]
\[
= \frac{1}{3n+1} D^{-1} \left[ \begin{array}{c} (3n-2) \Delta_1 g_1 - 3 \sum_{j \neq 1} \Delta_j g_j \\ \vdots \\ (3n-2) \Delta_n g_n - 3 \sum_{j \neq n} \Delta_j g_j - 3 \Delta_1 g_1 + 3 \Delta_1 g_1 \end{array} \right] 
\]
\[
= \frac{1}{3n+1} D^{-1} \left[ \begin{array}{c} (3n-2) \Delta_1 g_1 - 3s \\ \vdots \\ (3n-1) \Delta_1 g_1 - 3s \end{array} \right] 
\]
\[
= g - \frac{3s}{3n+1} D^{-1} 1_n.
\]
Proof. We find the norm by finding the largest singular value of \( \tilde{L}_n^\top \tilde{L}_n \). We have

\[
[\tilde{L}_n^\top \tilde{L}_n]_{i,j} = \begin{cases} 
144 \frac{(3n+1)^2}{2} (9n^2 - 3n - 5), & \text{if } i = j, \\
144 \frac{(3n+1)^2}{2} (-9n - 6), & \text{if } i \neq j, i, j \in \{1, 2, \ldots, n\}.
\end{cases}
\]

Let \( t = \frac{144}{(3n+1)^2} \). It follows that

\[ |\tilde{L}_n^\top \tilde{L}_n - \lambda \text{Id} | = t^n \begin{pmatrix}
1 - \frac{\lambda}{t} & -9n - 6 & \cdots & -9n - 6 \\
9n^2 - 3n - 5 - \frac{\lambda}{t} & -9n - 6 & \cdots & -9n - 6 \\
\vdots & \vdots & \ddots & \vdots \\
-9n - 6 & -9n - 6 & \cdots & 9n^2 - 3n - 5 - \frac{\lambda}{t}
\end{pmatrix}.
\]

Now, we apply elementary row and column operations on the matrix in (29) to make it an upper-diagonal matrix. First, let Row \( i = \text{Row } i - \text{Row } 1 \) for \( i \in \{2, 3, \ldots, n\} \). Second, using the new matrix, let Column 1 = Column 1 + \( \sum_{i=2}^n \) Column \( i \). This generates the matrix

\[
\begin{pmatrix}
1 - \frac{\lambda}{t} & -9n - 6 & \cdots & -9n - 6 \\
0 & 9n^2 + 6n + 1 - \frac{\lambda}{t} & \cdots & 0 \\
0 & 0 & 9n^2 + 6n + 1 - \frac{\lambda}{t} & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 9n^2 + 6n + 1 - \frac{\lambda}{t} \\
0 & 0 & 0 & \cdots & 0 & 9n^2 + 6n + 1 - \frac{\lambda}{t}
\end{pmatrix}
\]

and therefore we have

\[ |\tilde{L}_n^\top \tilde{L}_n - \lambda \text{Id} | = t^n \left(1 - \frac{\lambda}{t}\right) \left(9n^2 + 6n + 1 - \frac{\lambda}{t}\right)^{n-1}.\]

Hence, the eigenvalues of \( \tilde{L}_n^\top \tilde{L}_n \) are

\[ \lambda_1 = t = \frac{144}{(3n+1)^2} \quad \text{and} \quad \lambda_{2,3,\ldots,n} = \frac{t}{(9n^2 + 6n + 1)} = 144.\]

We see that the maximum eigenvalue, denoted \( \lambda_{\text{max}}(\tilde{L}_n^\top \tilde{L}_n) \), is 144. Therefore,

\[ \| \tilde{L}_n \| = \sqrt{\lambda_{\text{max}}(\tilde{L}_n^\top \tilde{L}_n)} = 12.\]

\[ \square \]

Lemma 4.3. Let \( f : \text{dom } f \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \) be \( C^2 \) on an open domain containing \( R(x_0; d) \). Let \( L_{\nabla f} \) denote the Lipschitz constant of \( \nabla f \) on \( R(x_0; d) \). Let \( D = \text{diag} \left[ \Delta_1, \cdots, \Delta_n \right] \in \mathbb{R}^{n \times n} \) and let \( w \in \mathbb{R}^n \) be defined by

\[ w_i = \int_{R(0,d)} x_i R_i(x_0 + x) \, dx, \quad i \in \{1, 2, \ldots, n\}, \]

where \( R_i(x_0 + x) \) is the remainder term of the first-order Taylor expansion of \( f(x_0 + x) \) about \( x_0 \). Then

\[ \| D^{-1} w \| \leq \frac{\sqrt{n}}{8} L_{\nabla f} \Delta^2 \Delta_s. \]

Moreover, if all \( \Delta_i \) are equal (i.e. the sample region is a hypercube), then

\[ \| D^{-1} w \| \leq \frac{\sqrt{24}}{\sqrt{n}} \frac{(2n+1)}{n} L_{\nabla f} \Delta^2 \Delta_s. \]
Proof. First, let us find an upper bound for \( \frac{w_i}{\Delta_i} \), \( i \in \{1, 2, \ldots, n\} \). We have

\[
\left| \frac{w_i}{\Delta_i} \right| = \left| \int_{R(0,d)} \frac{x_i}{\Delta_i} R_i(x^0 + x) \, dx \right| = \frac{1}{2} \left| \int_{R(0,d)} \frac{x_i}{\Delta_i} x^T \nabla^2 f(x) \, dx \right| \leq \frac{1}{2} \int_{R(0,d)} \frac{x_i}{\Delta_i} \|x\|^2 \|\nabla^2 f(x)\| \, dx.
\]

Using \( \|\nabla^2 f\| \leq L_{\nabla f} \) and \( R(0;d) \subseteq \mathbb{R}^n_+ \), we obtain

\[
\frac{1}{2} \left| \int_{R(0,d)} \frac{x_i}{\Delta_i} \|x\|^2 \|\nabla^2 f(x)\| \, dx \right| \leq \frac{L_{\nabla f}}{2} \int_{R(0,d)} \left( \frac{x_i}{\Delta_i} \right) \left( \sum_{j=1}^{n} x_j^2 \right) \, dx = \frac{L_{\nabla f}}{2} \left( \int_{R(0,d)} \frac{x_i^3}{\Delta_i} \, dx + \sum_{j \neq i} \int_{R(0,d)} \frac{x_i x_j^2}{\Delta_i} \, dx \right) = \frac{L_{\nabla f}}{2} \left( \frac{\Delta \Delta_i^2}{4} + \sum_{j \neq i} \frac{\Delta \Delta_j^2}{6} \right).
\]

Therefore,

\[
\|D^{-1} w\| \leq \sqrt{\sum_{i=1}^{n} \left( \frac{L_{\nabla f}}{24} \Delta \left( \Delta_i^2 + 2 \Delta_S^2 \right) \right)^2} \leq \frac{L_{\nabla f}}{24} \Delta \left( \sum_{i=1}^{n} \frac{3 \Delta \Delta_i^2}{2} \right) = \frac{L_{\nabla f}}{24} \Delta \left( \Delta_i^2 + 2 \Delta_S^2 \right).
\]

When all \( \Delta_i \) are equal, then (30) becomes

\[
\frac{L_{\nabla f}}{24} \Delta \left( \Delta_i^2 + 2 \Delta_S^2 \right) = \frac{L_{\nabla f}}{24} \Delta \left( \frac{\Delta_S^2}{n} + 2 \Delta_S^2 \right)
\]

and it follows that

\[
\|D^{-1} w\| \leq \frac{\sqrt{n} (2n + 1)}{12} L_{\nabla f} \Delta_S^2.
\]

We are now ready to introduce an error bound ad infinitum for the GSG.

**Theorem 4.4** (Error bound ad infinitum for the GSG). Let \( f : \text{dom} f \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \) be \( C^2 \) on an open domain containing \( R(x^0;d) \) where \( d = \left[ \Delta_1 \ \Delta_2 \ \cdots \ \Delta_n \right] > 0 \) and \( x^0 \) is the reference point. Let \( \Delta_S \) be the radius of \( S \subseteq \mathbb{R}^{n \times N} \) as defined in (2). Let \( \Delta_{\min} = \min_{i \in \{1, \ldots, n\}} \Delta_i \). Let \( L_{\nabla f} \) denote the Lipschitz constant of \( \nabla f \) on \( R(x^0;d) \). Then

\[
\lim_{N \rightarrow \infty} \nabla_s f(x^0; S) - \nabla f(x^0) \leq \frac{3}{2} \sqrt{n} L_{\nabla f} \frac{\Delta_S^2}{\Delta_{\min}}.
\]

Moreover, if all \( \Delta_i \) are equal, then

\[
\lim_{N \rightarrow \infty} \nabla_s f(x^0; S) - \nabla f(x^0) \leq \frac{1}{2} (2n + 1) L_{\nabla f} \Delta_S.
\]
Proof. We have

\[ \left\| \lim_{N \to \infty} \nabla_x f(x^0; S) - \nabla f(x^0) \right\| = \left\| \Delta^{-1} L_n (v + w) - \nabla f(x^0) \right\| \\
= \left\| \Delta^{-1} L_n v - \nabla f(x^0) + \Delta^{-1} L_n w \right\| \\
\leq \left\| \Delta^{-1} L_n v - \nabla f(x^0) \right\| + \Delta^{-1} \left\| D^{-1} \right\| L_n \left\| D^{-1} w \right\|. \]

By Lemma 4.1, we know \( \left\| \Delta^{-1} L_n v - \nabla f(x^0) \right\| = 0 \). By Lemma 4.2, Lemma 4.3 and since \( D^{-1} = \Delta_{\text{min}} \), we obtain

\[ \left\| \lim_{N \to \infty} \nabla f(x^0; S) - \nabla f(x^0) \right\| = \Delta^{-1} \left\| D^{-1} \right\| L_n \left\| D^{-1} w \right\| \leq \frac{3}{2} \sqrt{n} L_{\varphi f} \eta \Delta_S. \]

When all \( \Delta_i \) are equal, \( \Delta_{\text{min}} = \Delta = \Delta_S / \sqrt{n} \) for any \( i \in \{1, \ldots, n\} \). We obtain

\[ \left\| \lim_{N \to \infty} \nabla f(x^0; S) - \nabla f(x^0) \right\| \leq \Delta^{-1} \left\| D^{-1} \right\| L_n \left\| D^{-1} w \right\| \leq \Delta^{-1} \left( \frac{\sqrt{n}}{\Delta_S} (12) \frac{\sqrt{n}}{24} (2n + 1) \right) L_{\varphi f} \Delta_S^2 \]

\[ = \frac{1}{2} (2n + 1) L_{\varphi f} \Delta_S. \]

In the previous theorem, we see that the error bound is \( O\left( \frac{\Delta_S^2}{\Delta_{\text{min}}} \right) \). As \( \Delta_S \) is the radius of the sample region and \( \Delta_{\text{min}} \) is the length of the shortest side of the sample region, the theorem suggests that the more uniform the sample region the smaller the error. In other words, we want the simplex with vertices \( x^0, x^0 + \Delta_1, \ldots, x^0 + \Delta_n \) to be ‘as uniform as possible’. Analyzing the ratio \( \Delta_S / \Delta_{\text{min}} \), we see that the minimum value of this ratio is \( \sqrt{n} \), which occurs when the hyperrectangle is a hypercube.

5 The GSG over a ball

In this section, we find an error bound ad infinitum for the GSG of \( f \) at \( x^0 \) over a ball. First, we present some results on integration over a ball. In the next theorem, \( P(x) \) denotes a monomial. That is,

\[ P(x) = x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, \] (32)

where \( \alpha_i \in \mathbb{N} \cup \{0\} \) for all \( i \).

**Theorem 5.1** (Integrating over a ball). [Fol01] Let \( P \) be a monomial defined as in (32). Let \( \sigma \) be the \( (n-1) \)-dimensional surface measure on \( S_n(0; r) \). Let \( \beta_i = \frac{1}{2}(\alpha_i + 1) \) for all \( i \). Then

\[ \int_{B_n(0; r)} P(x) dx = \frac{r^{\alpha_1 + \cdots + \alpha_n + n}}{\alpha_1 + \cdots + \alpha_n + n} \int_{S_n(0; r)} P d\sigma, \]

where

\[ \int_{S_n(0; r)} P d\sigma = \begin{cases} 0, & \text{if any } \alpha_i \text{ is odd}, \\
\frac{2\Gamma(\beta_1)\Gamma(\beta_2)\cdots\Gamma(\beta_n)}{\Gamma(\beta_1 + \beta_2 + \cdots + \beta_n)}, & \text{if all } \alpha_i \text{ are even.} \end{cases} \]

We will also need an expression for the integral of \( Q(x) = |x_1|^{\alpha_1} |x_2|^{\alpha_2} \cdots |x_n|^{\alpha_n} \) over the ball \( B_n(0; r) \).

**Proposition 5.2**. Let \( Q(x) = |x_1|^{\alpha_1} |x_2|^{\alpha_2} \cdots |x_n|^{\alpha_n} \), and let \( \beta_i = \frac{1}{2}(\alpha_i + 1) \) for all \( i \). Then

\[ \int_{B_n(0; r)} Q(x) dx = \frac{r^{\alpha_1 + \cdots + \alpha_n + n}}{(\alpha_1 + \cdots + \alpha_n + n)} \frac{\Gamma(\beta_1)\cdots\Gamma(\beta_n)}{\Gamma(\beta_1 + \cdots + \beta_n)}. \]
Proof. Note that \(|x_i|^{\alpha_i}\) is an even function for any \(\alpha_i \in \mathbb{N} \cup \{0\}\). Using this fact and following the same scheme of the proof for Theorem 5.1 in [Fol01] yields the result.

Now, let us define the matrix of directions \(S_R\) that is used to form the sample points. Recall that in \(\mathbb{R}^n\), the conversion from Cartesian coordinates \(x = [x_1, x_2, \ldots, x_n]^\top\) to \(n\)-spherical coordinates is
\[
x_1 = \rho \cos \phi_1, \\
x_2 = \rho \sin \phi_1 \cos \phi_2, \\
x_3 = \rho \sin \phi_1 \sin \phi_2 \cos \phi_3, \\
\vdots \\
x_{n-2} = \rho \sin \phi_1 \cdots \sin \phi_{n-3} \cos \phi_{n-2}, \\
x_{n-1} = \rho \sin \phi_1 \cdots \sin \phi_{n-2} \cos \theta, \\
x_n = \rho \sin \phi_1 \cdots \sin \phi_{n-2} \sin \theta,
\]
where \(\rho = \|x\|\) and \(\phi_1, \ldots, \phi_{n-2}\) are the angles that identify the direction of \(x\). The angles have domains \(\theta \in [0, 2\pi)\) and \(\phi_i \in [0, \pi)\) for all \(i \in \{1, 2, \ldots, n-2\}\). To keep the same notation as the hyperrectangle, we define
\[
\Delta_1 = r, \quad \Delta_2 = 2\pi, \quad \Delta_3 = \Delta_4 = \cdots = \Delta_n = \pi.
\]
As before, \(N_i\) represents the number of subdivisions used to build the partitions in the ball. Once again, we define \(\Delta_i = \Delta_i/N_i\) for all \(i \in \{1, 2, \ldots, n\}\), \(\Delta = \Delta_1 \Delta_2 \cdots \Delta_n\), and \(N = N_1 N_2 \cdots N_n\).

Now we build the matrix \(S_R \in \mathbb{R}^{n \times N}\). The matrix \(S_R\) contains all directions to add to the reference point \(x^0\) to obtain a sample point in each partition of the ball \(B_n(x^0; r)\). A polar grid is built, in which each partition is a "bent" hyperrectangle. When using \(S_R\), the sample point chosen in each partition is the rightmost endpoint (the point with the largest values of \(\rho, \theta, \phi_1, \ldots, \phi_{n-2}\)). Let \(\Psi = [y_1, y_2, \ldots, y_n]^\top\) be a vector of indices in \(\mathbb{R}^n\) (not \(\mathbb{R}^{n-1}\) as it is the case for \(\mathbb{Z}\)). Define
\[
S_R = \frac{\partial \Psi}{N_1} \left[ \begin{array}{cccc}
\cos \frac{\pi y_1}{N_1} & \sin \frac{\pi y_1}{N_1} & \cdots & \sin \frac{\pi y_n}{N_1} \\
\cos \frac{\pi y_2}{N_2} & \sin \frac{\pi y_2}{N_2} & \cdots & \sin \frac{\pi y_n}{N_2} \\
\vdots & \vdots & \ddots & \vdots \\
\cos \frac{\pi y_n}{N_n} & \sin \frac{\pi y_n}{N_n} & \cdots & \sin \frac{\pi y_n}{N_n}
\end{array} \right]^\top \in \mathbb{R}^n.
\]
Then \(S_R\) can be written as
\[
S_R = \left[ s_{1,1}^{1,1,1,\ldots,1,1} s_{1,1}^{1,1,1,\ldots,1,2} \cdots s_{1,1}^{1,1,1,\ldots,N_1} s_{1,1}^{1,1,1,\ldots,N_2} \cdots s_{1,1}^{1,1,1,\ldots,N_n} \right].
\]

Let us provide an example of a sample set built by using \(S_R\) in \(\mathbb{R}^2\).

Example 5.3. In this example, the reference point is \(x^0 = [0, 0]^\top\). The sample region is a ball with radius \(\Delta_1 = r = 30\). Set \(N_1 = 3\) and \(N_2 = 4\). Hence, \(\Delta_1 = \frac{30}{3} = 10\) and \(\Delta_2 = \frac{30}{4} = \frac{15}{2}\). The matrix \(S_R \in \mathbb{R}^{2 \times 12}\) is given by
\[
S_R = \begin{bmatrix}
1 & 3 & 4 & 2 & 1 & 3 & 4 & 2 & 1 & 3 & 4 & 2 \\
10 & 0 & -10 & 0 & 10 & 0 & -20 & 0 & 20 & 0 & -30 & 0
\end{bmatrix}
\]
\[
S_R = \begin{bmatrix}
0 & -10 & 0 & 10 & 0 & -20 & 0 & 20 & 0 & -30 & 0 & -30
\end{bmatrix}
\]

The sample points \(x^i\) are obtained by setting \(x^i = x^0 + S_R e_j\) for all \(i \in \{1, 2, \ldots, 12\}\). Figure 3 illustrates the sample points built from the matrix \(S_R\).
Figure 3: An example of a sample set built from $S_R$ in $\mathbb{R}^2$.

Note that $S_R$ is full row rank whenever all $N_i > 2$. For the remainder of this section, assume $N_i > 2$ for all $i$. Hence, we want to find the limit of the following expression:

$$\lim_{N \to \infty} \nabla_x f(x^0; S_R) = \lim_{N \to \infty} (S_R S_R^\top)^{-\top} S_R \delta f(x^0; S_R). \quad (33)$$

Define the determinant of the Jacobian as a function $J : \mathbb{R}^n \to \mathbb{R}$:

$$J(y_1, \ldots, y_n) = \left( \frac{\rho y_1}{N_1} \right)^{n-1} \sin^{n-2} \frac{\pi y_3}{N_3} \sin^{n-3} \frac{\pi y_4}{N_4} \cdots \sin^2 \frac{\pi y_{n-1}}{N_{n-1}} \sin \frac{\pi y_n}{N_n}. \quad (34)$$

Let $J \in \mathbb{R}^{N \times N}$ be the matrix defined by

$$J = \text{diag} \left[ j^{1,1}, \ldots, j^{1,1}, \ldots, j^{N_1, N_2, \ldots, N_n} \right]$$

where $j^\pi = J(y_1, y_2, \ldots, y_n)$.

Define

$$K = S_R J S_R^\top = S_R J S_R^\top (S_R S_R^\top)^{-1}. \quad (35)$$

Notice that $K \in \mathbb{R}^{n \times n}$ is an invertible diagonal matrix such that

$$S_R J = K S_R. \quad (35)$$

Considering (33), notice that

$$(S_R S_R^\top)^{-\top} S_R \delta f(x^0; S_R) = (S_R S_R^\top)^{-\top} \frac{1}{\Delta} K^{-1} K S_R \delta f(x^0; S_R) \Delta$$

$$(K S_R S_R^\top \Delta)^{-\top} K S_R \delta f(x^0; S_R) \Delta$$

$$(S_R J S_R^\top \Delta)^{-\top} S_R J \delta f(x^0, S_R) \Delta.$$
Proposition 5.4. Let \( M \) be defined as in (36). Then

\[
[M]_{i,j} = \begin{cases} 
\frac{V_{n+2}}{2\pi} & \text{if } i = j, \\
0 & \text{if } i \neq j.
\end{cases}
\]

Consequently

\[
\lim_{N \to \infty} \left( \frac{N_1}{y_1} \cdots \frac{N_n}{y_n} \right)^T \left( \sum_{y_1=1}^{N_1} \cdots \sum_{y_n=1}^{N_n} \frac{V(x_0; y)}{s} \right)^T J(y_1, \ldots, y_n) \Delta = \frac{2\pi}{V_{n+2}} \text{Id}_n.
\]
Proof. Each entry of $M_n$ is a $n$-tuple Riemann sum over $B_n(0; r)$. Taking the limit as $N \to \infty$, each entry of $M_n$ can be written as the following integral (in Cartesian coordinates):

$$[M_n]_{i,j} = \int_{B_n(0; r)} x_i x_j \, dx, \quad i, j \in \{1, 2, \ldots, n\}.$$  

The off-diagonal entries of $M_n$ are zero by Proposition 5.1. The diagonal entries are given by

$$[M_n]_{i,i} = \frac{r^{n+2}}{(n+2)} \frac{2\Gamma \left( \frac{n+1}{2} \right)}{\Gamma \left( \frac{n+1}{2} + \frac{3}{2} \right)} = \frac{r^{n+2}}{(n+2)} \frac{2\pi^{\frac{n+1}{2}}}{\Gamma \left( \frac{n+1}{2} + \frac{3}{2} \right)} = \frac{r^{n+2}}{(n+2)} \frac{2\pi^{\frac{n+1}{2}}}{\Gamma \left( \frac{n+1}{2} + \frac{3}{2} \right)}.$$

From (37), we see that the diagonal entries are simply

$$[M_n]_{i,i} = \frac{V_{n+2}}{2\pi}.$$  

The second equation follows trivially. \hfill \Box

In the next proposition, we give an expression for the second limit in (36).

**Proposition 5.5.** Let $f : \text{dom } f \subseteq \mathbb{R}^n \to \mathbb{R}$ with $B_n(x^0; r) \subseteq \text{dom } f$. Then the following limit can be written as a vector of integrals (in Cartesian coordinates):

$$\lim_{N \to \infty} \left[ \sum_{y_1=1}^{N_1} \cdots \sum_{y_n=1}^{N_n} s^\top \delta_f(x^0; s^\top) J(y_1, \ldots, y_n) \Delta \right] = \left[ \int_{B_n(0; r)} x_1 (f(x^0 + x) - f(x^0)) \, dx \right] \left[ \int_{B_n(0; r)} x_2 (f(x^0 + x) - f(x^0)) \, dx \right] \cdots \left[ \int_{B_n(0; r)} x_n (f(x^0 + x) - f(x^0)) \, dx \right] = T_n \in \mathbb{R}^n. \quad (39)$$

Proof. The $n$-tuple sum of the left-hand side of (39) is a Riemann sum with a finite-sized sample region $B_n(x^0; r)$. The result follows by taking the limit as $N \to \infty$. \hfill \Box

Now we generalize the matrix $S_R$. Let

$$S = \begin{bmatrix} s^{1,1,1,\ldots,1,1} & s^{1,1,1,\ldots,1,2} & \cdots & s^{N_1, N_2, \ldots, N_{n-1}, N_n} \end{bmatrix}$$

be a matrix in $\mathbb{R}^{n \times N}$ in which each column $s$ is a direction to add to $x^0$ to form exactly one arbitrary sample point in each partition of $B_n(x^0; r)$. Note that Propositions 5.4 and 5.5 still hold by considering $S$ instead of $S_R$. Indeed, since $f$ is a continuous function, as long as exactly one sample point is used in each partition of the ball, the results of the previous two propositions hold. We are now ready to provide an expression for the GSG ad infinitum of $f$ at $x^0$ over $B_n(x^0; r)$.

**Theorem 5.6** (The GSG over a ball). Let $f : \text{dom } f \subseteq \mathbb{R}^n \to \mathbb{R}$ with $B_n(x^0; r) \subseteq \text{dom } f$. Let $S \in \mathbb{R}^{n \times N}$ be a matrix such that each sample point $x^0 + s e_j, j \in \{1, 2, \ldots, N \}$, is in exactly one partition of the ball $B_n(x^0; r)$. Let $V_{n+2}$ be the volume of a ball with radius $r$ in $\mathbb{R}^{n+2}$ and $T_n \in \mathbb{R}^n$ be defined as in (39). Then

$$\lim_{N \to \infty} \nabla_s f(x^0; S) = \frac{2\pi}{V_{n+2}} T_n. \quad (40)$$

Proof. we have

$$\lim_{N \to \infty} \nabla_s f(x^0; S)$$

$$= \lim_{N \to \infty} \left( SS^\top \right)^{-\top} S \delta_f(x^0; S)$$

$$= \lim_{N \to \infty} \left[ \left( \sum_{y_1=1}^{N_1} \cdots \sum_{y_n=1}^{N_n} s^\top \delta_f(x^0; s^\top) J(y_1, \ldots, y_n) \Delta \right) \right] \lim_{N \to \infty} \left[ \sum_{y_1=1}^{N_1} \cdots \sum_{y_n=1}^{N_n} s^\top \delta_f(x^0; s^\top) J(y_1, \ldots, y_n) \Delta \right]$$

$$= \frac{2\pi}{V_{n+2}} T_n,$$

by Propositions 5.4 and 5.5. \hfill \Box

23
Let us provide an example of the calculations necessary to obtain the limit of the the GSG.

**Example 5.7.** Let \( f : \mathbb{R}^2 \to \mathbb{R} : x \mapsto x_1^2 + x_2^2 \). Let the reference point be \( x^0 = \begin{bmatrix} 3 & 1 \end{bmatrix}^\top \). Let the sample region be \( B_2(x^0; 1) \). By Theorem 5.6, we know that

\[
\lim_{N \to \infty} \nabla_s f(x^0; S) = \frac{2\pi}{V_4} \left[ \int_{B_2(x^0; 1)} x_1 (f(x^0 + x) - f(x^0)) \, dx \right] + \frac{2\pi}{V_4} \left[ \int_{B_2(x^0; 1)} x_2 (f(x^0 + x) - f(x^0)) \, dx \right].
\]

Writing the vector of integrals in polar coordinates and since \( V_4 = \frac{\pi^2}{4} \), we obtain

\[
\lim_{N \to \infty} \nabla_s f(x^0; S) = 4\pi \left[ \int_0^{2\pi} \int_0^1 r \cos \theta \left( (3 + r \cos \theta)^2 + (1 + r \sin \theta)^2 - 10 \right) r \, dr \, d\theta \right] = 4\pi \left[ \frac{3\pi}{2} \right] = \frac{6\pi}{2} = 3.
\]

**Note that for this problem** \( \lim_{N \to \infty} \nabla_s f(x^0; S) = \nabla f(x^0) \).

The reason why the GSG is perfectly accurate in the previous example will be discussed at the end of this section, but first we develop an error bound ad infinitum for the GSG over a ball. To obtain such an error bound, we require \( f \) to be \( C^3 \) on an open domain containing \( B_n(x^0; r) \). The function \( f \) at \( x^0 + x \) is written as the second-order Taylor expansion. That is,

\[
f(x^0 + x) = f(x^0) + \nabla f(x^0)^\top x + \frac{1}{2} x^\top \nabla^2 f(x^0) x + R_2(x^0 + x),
\]

where the remainder term satisfies

\[
|R_2(x^0 + x)| \leq \frac{1}{6} L_H \Delta_S^3
\]

and \( L_H \) denotes the Lipschitz constant of the Hessian on \( B_n(x^0; r) \). By rewriting \( f(x^0 + x) \) in the form of (41), each entry of the vector \( T_n \) defined in Proposition 5.5 can be written as

\[
[T_n]_i = \int_{B_n(0; r)} x_i (f(x^0 + x) - f(x^0)) \, dx
\]

\[
= \int_{B_n(0; r)} x_i \left( \nabla f(x^0)^\top x + \frac{1}{2} x^\top \nabla^2 f(x^0) x + R_2(x^0 + x) \right) \, dx
\]

\[
= \int_{B_n(0; r)} x_i \nabla f(x^0)^\top x \, dx + \int_{B_n(0; r)} x_i x^\top \nabla^2 f(x^0) x \, dx + \int_{B_n(0; r)} x_i R_2(x^0 + x) \, dx,
\]

for \( i = \{1, 2, \ldots, n\} \). Let \( v \in \mathbb{R}^n \) be the vector defined by

\[
v_i = \int_{B_n(0; r)} x_i \nabla f(x^0)^\top x \, dx, \quad i = \{1, 2, \ldots, n\}.
\]

\( w \in \mathbb{R}^n \) be the vector defined by

\[
w_i = \int_{B_n(0; r)} x_i x^\top \nabla^2 f(x^0) x \, dx, \quad i = \{1, 2, \ldots, n\},
\]

and \( z \in \mathbb{R}^n \) be the vector defined by

\[
z_i = \int_{B_n(0; r)} x_i R_2(x^0 + x) \, dx, \quad i = \{1, 2, \ldots, n\}.
\]

Then the expression for \( \lim_{N \to \infty} \nabla_s f(x^0; S) \) given in (40) can be written as

\[
\lim_{N \to \infty} \nabla_s f(x^0; S) = \frac{2\pi}{V_{n+2}} (v + w + z) = \frac{2\pi}{V_{n+2}} v + \frac{2\pi}{V_{n+2}} w + \frac{2\pi}{V_{n+2}} z.
\]

We now find an expression for the first two terms in (45).
Lemma 5.8. Let $f : \text{dom } f \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be $C^3$ on an open domain containing $B_n(x^0; r)$. Let $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^n$ be defined as in (42) and (43). Then

$$
\frac{2\pi}{V_{n+2}}v = \nabla f(x^0) \text{ and } \frac{2\pi}{V_{n+2}}w = 0.
$$

Proof. Let $g_i = [\nabla f(x^0)]_i$ for all $i \in \{1, 2, \ldots, n\}$. We have

$$
\frac{2\pi}{V_{n+2}} \int_{B_n(0, r)} x_i \nabla f(x^0) dx = \frac{2\pi}{V_{n+2}} \int_{B_n(0, r)} x_i^2 g_i dx + \sum_{j \not= i} \frac{2\pi}{V_{n+2}} x_i x_j g_j dx = \frac{2\pi}{V_{n+2}} \frac{V_n}{2\pi} g_i + 0 = g_i,
$$

by Proposition 5.4. Therefore $\frac{2\pi}{V_{n+2}}v = \nabla f(x^0)$. Let $\nabla^2 f(x^0) = H \in \mathbb{R}^{n \times n}$. We have

$$
\frac{\pi}{V_{n+2}} \int_{B_n(0, r)} x_i x^\top \nabla^2 f(x^0) dx = \frac{\pi}{V_{n+2}} \left( \sum_{j=1}^n \int_{B_n(0, r)} x_i x_j^2 H_{j,j} dx + 2 \sum_{j=1}^n \sum_{k \not= j} \int_{B_n(0, r)} x_i x_j x_k H_{j,k} dx \right) = 0,
$$

by Theorem 5.1. Therefore, $\frac{2\pi}{V_{n+2}}w = 0$. \qed

Next, we find an upper bound for the third term in (45). In Lemma 5.9 we create the redundant variable $\Delta_S = r$. This allows for easy and immediate comparison to the results in Section 3.

Lemma 5.9. Let $f : \text{dom } f \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be $C^3$ on an open domain containing $B_n(x^0; r)$. Let $z \in \mathbb{R}^n$ be defined as in (44). Let

$$
\eta = \frac{\Gamma\left(\frac{n+4}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{n+2}{2}\right)}.
$$

Also, denote by $L_H$ the Lipschitz constant of the Hessian on $B_n(x^0; r)$. Let $\Delta_S = r$. Then

$$
\frac{2\pi}{V_n + 2} \|z\| \leq \frac{\sqrt{n}}{3\sqrt{\pi}} L_H \eta \Delta_S^2.
$$

Proof. We have

$$
|z_i| = \left| \frac{2\pi}{V_{n+2}} \int_{B_n(0, r)} x_i R_2(x^0 + x) dx \right|
\leq \frac{2\pi}{V_{n+2}} \int_{B_n(0, r)} |x_i||R_2(x^0 + x)| dx
\leq \frac{2\pi}{V_{n+2}} \frac{1}{6} L_H \Delta_S^2 \int_{B_n(0, r)} |x_i| dx.
$$

(46)

By Proposition 5.2, we know that

$$
\int_{B_n(0, r)} |x_i| dx = 2r^{n+1} \frac{\Gamma(1)\Gamma\left(\frac{n}{2}\right)}{(n + 1)\Gamma\left(1 + \frac{n+1}{2}\right)} = 2r^{n+1} \pi \frac{n+1}{(n+1)\Gamma\left(\frac{n+3}{2}\right)}.
$$

(47)

Note that

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

Hence, (47) can be written as

$$
\int_{B_n(0, r)} |x_i| dx = \pi \frac{n+1}{\Gamma\left(\frac{n+3}{2}\right)} = \frac{V_{n+1}}{\pi}.
$$

(48)
Substituting (48) in (46) gives
\[
\frac{2\pi}{V_{n+2}} \int_{B_n(0,r)} x_i R_2(x^0 + x) dx \leq \frac{1}{3} \frac{V_{n+1}}{V_{n+2}} L_H \Delta_S^3.
\] (49)

The term \( V_{n+1}/V_{n+2} \) in (49) is
\[
\frac{V_{n+1}}{V_{n+2}} = \frac{\pi^{n+1} n_{n+1}}{\Gamma(n+2)} \frac{\Gamma(n+4/2)}{\pi^{n+2} r^{n+2}} = \frac{1}{\sqrt{\pi}} \eta.
\]

Thus,
\[
|z_i| \leq \frac{1}{3\sqrt{\pi}} L_H \eta \Delta_S^2, \quad \forall i \in \{1, 2, \ldots, n\}.
\]

Therefore,
\[
\|z\| \leq \frac{\sqrt{n}}{3\sqrt{\pi}} L_H \eta \Delta_S^2.
\]

\[\square\]

**Theorem 5.10** (Error bound ad infinitum for the GSG over a ball). Let \( f : \text{dom } f \subset \mathbb{R}^n \to \mathbb{R} \) be \( C^3 \) on an open domain containing \( B_n(x^0; r) \). Let the vectors \( v, w, z \in \mathbb{R}^n \) be defined as in (42), (43), (44), respectively. Let
\[
\eta = \frac{\Gamma(n+4/2)}{\sqrt{\pi} \Gamma(n+2)}.
\]

Denote by \( L_H \) the Lipschitz constant of \( \nabla^2 f \) on \( B_n(x^0; r) \). Let \( \Delta_S \) be defined as in (2). Then
\[
\left\| \lim_{N \to \infty} \nabla_x f(x^0; S) - \nabla f(x^0) \right\| \leq \frac{\sqrt{n}}{3\sqrt{\pi}} L_H \eta \Delta_S^2.
\] (50)

**Proof.** We have
\[
\left\| \lim_{N \to \infty} \nabla_x f(x^0; S) - \nabla f(x^0) \right\| = \left\| \frac{2\pi}{V_{n+2}} v + \frac{2\pi}{V_{n+2}} w + \frac{2\pi}{V_{n+2}} z - \nabla f(x^0) \right\|
\]
\[
= \left\| \nabla f(x^0) + 0 + \frac{2\pi}{V_{n+2}} z - \nabla f(x^0) \right\|
\]
\[
= \frac{2\pi}{V_{n+2}} \|z\| \leq \frac{\sqrt{n}}{3\sqrt{\pi}} L_H \eta \Delta_S^2.
\]

\[\square\]

Note that the error bound ad infinitum over a ball is \( O(\Delta_S^2) \), which is not the case for the error bound ad infinitum defined in Section 4. This is due to the fact that, for each column \( s \in S \), its opposite \(-s\) is also in \( S \) as \( \bar{N} \to \infty \). Therefore, the limit of the GSG over \( B_n(x^0; r) \) is equivalent to the limit of the generalized centered simplex gradient over a half-ball centered at \( x^0 \) of radius \( r \). Notice that the shape of the sample region is not the key point to obtain an error bound \( O(\Delta_S^2) \). The position of the reference point \( x^0 \) is what matters. Indeed, we could get an error bound ad infinitum of accuracy \( O(\Delta_S^2) \) by considering a hyperrectangle, but instead of letting \( x^0 \) be the left endpoint of the sample region as we did in the previous sections, let \( x^0 \) be located at the intersection of all diagonals of the hyperrectangular sample region. Finally, note that the error bound in (50) involves the Lipschitz constant of the Hessian of \( f \), \( L_H \). Therefore, the error bound reduces to zero whenever \( f \) is a polynomial of degree at most 2. This explains why the GSG is a perfect approximation in Example 5.7.
6 Conclusion and future research directions

In this paper, we have provided an expression for the GSG ad infinitum and an error bound ad infinitum for the GSG over both a hyperrectangle and a ball. In both cases, we note that the error bound is independent of the number of sample points, which is critical in allowing the analysis of the limits. Examining the techniques used in each case, it seems likely that an error bound ad infinitum (independent of \( N \)) for the GSG of \( f \) at \( x^0 \) over any reasonable sample region can be defined. However, repeating the process for every possible region is clearly an unreasonable proposition. A more practical open question is the following: Given a set of sample points \( \Omega \subseteq \mathbb{R}^n \) and a bijection \( T : \mathbb{R}^n \mapsto \mathbb{R}^n \) such that \( T(\Omega) = R(x^0; d) \), can the bijection be used to determine the GSG ad infinitum and an error bound ad infinitum for the GSG over \( \Omega \)?

Comparing Theorems 4.4 and 5.10, we see that the position of the reference point \( x^0 \) has an impact on the error bound. Indeed, when \( x^0 \) is the center of the sample region, then we obtained an error bound ad infinitum of \( O(\Delta s^2) \). But, when \( x^0 \) is on the boundary of the sample region, then we obtained an error bound ad infinitum of \( O(\Delta s) \). It is unclear how these conclusions change if the reference point is at another location within the sample region.

In [BHJB21] it was found that under certain conditions, the limit of the classical error bound for the GSG in \( \mathbb{R}^n \) (as herein) could be taken directly. It is possible that this is true in \( \mathbb{R}^n \) as well. However, the techniques in [BHJB21] do not adapt directly.

We conclude this paper with a comparison of classical error bounds (as given by (3) and (4)) as \( N \) gets large to the error bounds ad infinitum derived in Theorems 4.4 and 5.10.

**Example 6.1.** In this example, we consider the function \( f : \mathbb{R}^2 \rightarrow \mathbb{R} : x = [x_1 \ x_2]^T \mapsto x_1^2 + x_2^2 \). The reference point is set to \( x^0 = [1 \ 1]^T \). Two sample regions are considered: the square \( [0,1] \times [0,1] \) and the ball \( B_2(x^0; 1) \).

We set \( N_1 = N_2 \), so \( N = (N_1)^2 \). The classical error bounds are computed using (3) and (4). The error bounds ad infinitum are computed using Theorems 4.4 and 5.10. Finally, the GSG is constructed and the true absolute error is computed for both sample regions. Figure 4 visualizes the results for \( N_1 \in \{2^2, 2^3, \ldots, 2^{10}\} \) (so, \( N \in \{2^4, 2^6, \ldots, 2^{20}\} \)). Note the bounds ad infinitum are independent of \( N_1 \), so constants.

![Figure 4: The error bound ad infinitum in \( \mathbb{R}^2 \) for two different sample regions](image)

*Based on this example, the error bounds ad infinitum provides an accurate upper bound as low as \( N = 16 \). It also appears that the error bound ad infinitum over the ball \( B_2(x^0; 1) \) provides a tighter error bound than the classical error bound for \( N \geq 16 \). Finally, in this example, it appears that the classical error bounds may converge as \( N \) tends to infinity as conjectured.*
Acknowledgements

Hare’s research is partially funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada, Discover Grant #2018-03865. Jarry-Bolduc’s research is partially funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada, Discover Grant #2018-03865. Jarry-Bolduc would like to acknowledge UBC for the funding received through the University Graduate Fellowship award.

References

[AH17] C. Audet and W. Hare. *Derivative-free and Blackbox Optimization*. Springer International Publishing AG, Switzerland, 2017.

[BBN19] A. Berahas, R. Byrd, and J. Nocedal. Derivative-free optimization of noisy functions via quasi-Newton methods. *SIAM Journal on Optimization*, 2019. To appear.

[BHJB21] Phillip Braun, Warren Hare, and Gabriel Jarry-Bolduc. Limiting behavior of derivative approximation techniques as the number of points tends to infinity on a fixed interval in $\mathbb{R}$. *J. Comput. Appl. Math.*, 386:113218, 22, 2021.

[CSV08] A. Conn, K. Scheinberg, and L. Vicente. Geometry of interpolation sets in derivative free optimization. *Math. Program.*, 111(1-2):141–172, 2008.

[CSV09] A.R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to Derivative-Free Optimization*, volume 8 of MPS/SIAM Book Series on Optimization. SIAM, 2009.

[Fol01] G. Folland. How to integrate a polynomial over a sphere. *The American Mathematical Monthly*, 108(5):446–448, 2001.

[HJBP20] W. Hare, G. Jarry-Bolduc, and C. Planiden. Error bounds for overdetermined and underdetermined generalized centred simplex gradients. *arXiv preprint arXiv:2006.00742*, 2020.

[Kel99] C. T. Kelley. *Iterative methods for optimization*, volume 18 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

[MW19] M. Menickelly and S. Wild. Derivative-free robust optimization by outer approximations. *Mathematical Programming*, 2019. To appear.

[MWDS18] A. Maggiar, A. Wächter, I. Dolinskaya, and J. Staum. A derivative-free trust-region algorithm for the optimization of functions smoothed via Gaussian convolution using adaptive multiple importance sampling. *SIAM Journal on Optimization*, 28(2):1478–1507, 2018.

[Pow09] M. Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA0909/06, University of Cambridge, Cambridge*, pages 26–46, 2009.

[Reg15] R.G. Regis. The calculus of simplex gradients. *Optim. Lett.*, 9(5):845–865, 2015.

[RW98] R. Rockafellar and R. Wets. *Variational analysis*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1998.

[VKPS17] A. Verdério, E. Karas, L. Pedrosa, and K. Scheinberg. On the construction of quadratic models for derivative-free trust-region algorithms. *EURO Journal on Computational Optimization*, 5(4):501–527, 2017.

[Zie72] R. Ziemniński. A simplex design for gradient estimation in quadratic regression. *Zastos. Mat.*, 13:493–496, 1972.

[Zie73] Ryszard Ziemniński. A randomized finite-differential estimator of the gradient. *Algorytmy*, 10(18):21–30, 1973.