Model-Based Trust Region Methods for Derivative-Free Optimization with Unrelaxable Linear Constraints

Stephen C. Billups and Trever Hallock

January 3, 2022

Abstract

We propose a model-based trust-region algorithm for constrained optimization problems with linear constraints in which derivatives of the objective function are not available and the objective function values outside the feasible region are not available. In each iteration, the objective function is approximated by an interpolation model, which is then minimized over a trust region. To ensure feasibility of all sample points and iterates, we consider two trust region strategies in which the trust regions are contained in the feasible region. Computational results are presented on a suite of test problems.

1 TO DO

1. Shorten literature review
2. Consider removing material on finding the maximum volume ellipsoid
3. Remove ellipsoid searches. Instead, just use the one center defined in the analysis.
4. Remove polyhedral trust region method.

2 Introduction

Derivative-free optimization (DFO) refers to mathematical programs involving functions for which derivative information is not explicitly available. Such problems arise, for example, when the functions are evaluated by simulations or by laboratory experiments. Applications of DFO appear in many fields, including photo-injector optimization [1], circuitry arrangements [2], machine learning [3], volume optimization [4], and reliability-based optimization [5]. In such applications, function evaluations are typically expensive, so it is sensible to invest significant computational resources to minimize the number of function evaluations required.

This work is aimed at developing algorithms to solve constrained optimization problems of the form

$$\begin{array}{ll} \min_{x\in\mathbb{R}^n} & f(x) \\ \text{subject to} & c_i(x) \leq 0, \qquad \forall 1 \leq i \leq m, \end{array} \tag{1}$$

where f and c_i for $1 \leq i \leq m$ are real-valued functions on $\mathbb{R}^n$ with at least one of these functions being a *black-box* function, meaning that derivatives cannot be evaluated directly. We let the feasible set be represented as

$$\mathcal{F} = \{x \in \mathbb{R}^n | c_i(x) \leq 0,\ \forall 1 \leq i \leq m\}. \tag{2}$$

We are interested in developing *model-based*, *trust-region* algorithms for solving these problems. Model-based methods work by constructing model functions to approximate the black box functions at each iteration. The model functions are determined by fitting previously evaluated function values on a set of sample points. In trust-region methods, the model functions are used to define a trust-region subproblem whose solution determines the next iterate. For example, the trust-region subproblem might have the form

$$\begin{array}{ll} \min_{\|s\|\leq\Delta_k} & {m_f}^{(k)}\left(x^{(k)}+s\right) \\ \text{subject to} & m_{c_i}^{(k)}\left(x^{(k)}+s\right) \leq 0 \qquad 1 \leq i \leq m, \\ & \|s\| \leq \Delta_k \end{array}$$

where $x^{(k)}$ is the current iterate, ${m_f}^{(k)}$ is the model function approximating f, and $m_{c_i}^{(k)}$ are the model functions approximating the constraint functions $c_i, \forall 1 \leq i \leq m$, and Δ_k is the radius of the trust-region. The key differences between the trust-region subproblem problem and the original are that all functions are replaced with their model functions, and a trust region constraint ($\|s\| \leq \Delta_k$) has been added.

Conceptually, the model functions are "trusted" only within a distance Δ_k of the current iterate $x^{(k)}$; so the trust-region subproblem restricts the length of step s to be no larger than Δ_k. To ensure that the model functions are good approximations of the true functions over the trust region, the sample points are typically chosen to lie within the trust-region.

We are specifically interested in applications where some of the black box functions cannot be evaluated outside the feasible region. Constraints of this type can arise in several ways, such as when a simulation cannot be carried out on particular inputs. For example, the authors of [6] optimize rapid-cycling synchroton lattice design, but some overlapping synchrotron elements cannot be simulated. Following the taxonomy in [7], such constraints are called *unrelaxable*.

For problems involving unrelaxable constraints, we require all iterates and sample points to be feasible. ~~Notice that this feasibility requirement means that if the algorithm is stopped at any time, it will still produce a feasible output.~~ As a first step toward developing algorithms to solve such problems, this paper considers a simplified problem where all of the constraints are linear and unrelaxable; namely:

$$\begin{array}{ll} \min_{x\in\mathbb{R}^n} & f(x) \\ & Ax \leq b, \end{array} \tag{3}$$

where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$.

An important consideration in fitting the model functions is the "geometry" of the sample set. This will be discussed in more detail in Section 5.3, but the key point is that the relative positions of the sample points within the trust region have a significant effect on the accuracy of the model functions over the trust region. When the geometry of the sample set is poor, it is sometimes necessary to evaluate the functions at new points within the trust region to improve the geometry of the sample set. It is well understood how to do this for unconstrained problems, but for constrained problems our requirement that the sample points must be feasible poses some interesting challenges to maintaining good geometry.

SCB note: The following new paragraph discusses geometry

In unconstrained optimization, it is typical and natural that the sample trust region be centered set at the current iterate. Thus, much of the standard theory (see for example, Chapter 3 of [8]) about the geometry of the sample set assumes that the sample points are chosen within a ball centered at the current iterate. However, for constrained problems, this is no longer a useful assumption. In particular, when iterates are close to the boundary of the feasible region, it is advantageous to shift the center of the sample trust region away from the current iterate. Because of this, we rework some of the standard results so that they don't require that one of the sample points be at the center of the sample trust region.

We propose an algorithm for solving (3) whose main idea is to construct a feasible ellipsoid at each iteration that lies within the intersection of the feasible region and the current trust region. We call this ellipsoid the *sample trust region*. To choose well-poised sample points for this ellipsoidal region, we adapt a model improvement algorithm presented in [8] for spherical trust regions and establish error bounds on the accuracy of the model functions over this region.

Our convergence analysis is based on an algorithmic framework presented [9], which describes a class of trust-region algorithms for convex constrained minimization without derivatives. This framework assumes that a model of the objective function can be constructed at each iteration that satisfies certain accuracy assumptions, but it does not specify how to construct such a model function. Here, we show that the model functions constructed by our algorithm satisfy these assumptions. Hence, using the results from [9], we show that the criticality measure for the iterates generated by our algorithm converges to zero. As a consequence, any accumulation point of the iterates satisfies the Karush-Kuhn-Tucker first order optimality conditions.

SCB note: Consider deleting the following paragraph

We then discuss several variants to improve on the algorithm, as well as provide numerical results. We present several variants of how to construct the feasible ellipsoid. We first show how to find the maximum volume ellipsoid contained within a polytope given a fixed center. We then explore several strategies for shifting the center of the ellipsoid.

3 Notation

SCB note: Make sure we actually use all of these conventions in the paper. That is, delete anything used in the thesis that isn't used in the linear constraint paper We use the following notational conventions throughout the paper. Any variables that depend on the iteration will be super-scripted by k. For example, the k-th iterate is given by $x^{(k)}$, and the model of the objective function for iteration k is given by ${m_f}^{(k)}$. Subscripts on vectors are used as an index into the vector, while vectors in a sequence of vectors use superscripts: that is, $x_i^{(k)}$ is the i-th component of the k-th vector in the sequence $\{x^{(k)}\}_k$. We use $\text{Proj}_X(x) = \arg\min_{x' \in X} \|x' - x\|$ to denote the projection of x onto a convex set X. For any $m \in \mathbb{N}$, we define $[m] = \{i \in \mathbb{N} | 1 \leq i \leq m\}$.

Matrices. Matrices are denoted with capital letters. The i-th row of the matrix A is denoted A_i. For any index set $S \subseteq \{i \in \mathbb{N} | 1 \leq i \leq m\}$, the $|S| \times n$ matrix formed by only using rows of the $m \times n$ matrix A from the set S is A_S. Let the condition number of a matrix Q be denoted $\kappa(Q)$. We use e_i to denote the i-th unit vector and e to denote the vector of all ones. $B_k(c; \Delta)$ is the ball of radius Δ in the k norm, centered at point c. That is,

$$B_k(c; \Delta) = \left\{ x \in \mathbb{R}^n \,\middle|\, \|x - c\|_k \leq \Delta \right\}. \tag{4}$$

Note that some of the common norms are:

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|, \quad \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}, \quad \text{and} \quad \|x\|_1 = \sum_{i=1}^n |x_i|.$$

When a norm is presented without a subscript, it is assumed to be the 2 norm: $\|x\| = \|x\|_2$ for any $x \in \mathbb{R}^n$. For matrices, we will also use the Frobenius norm $\|\cdot\|_F$ which is $\|A\|_F = \sqrt{\sum_i \sum_j A_{i,j}^2}$.

Sets. SCB note: I am worried about overloading the bar notation. We use $\bar{\Phi}$ and $\bar{Y}$ (and maybe other examples) to mean something other than the complement of Φ and Y. Maybe we should use S^c to denote the complement of a set S. The complement of a set S is denoted as $\bar{S}$. Define a point x subtracted from a set S as $S - x = \{s \in \mathbb{R}^n | s - x \in S\}$. Set addition is $X + Y = \{x + y | x \in X, y \in Y\}$.

$\delta_{i,j}$ is the Kronecker delta, $\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$. We define the positive part of a real number to be

$$a^+ = \begin{cases} a & \text{if} \quad a \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

4 Background

SCB note: I commented out the literature review for now. We may not need to have a literature review section.

5 Model-Based Trust-Region Methods for DFO

The main idea of model-based, trust-region methods is that trial points are selected at each iteration by solving a trust-region subproblem. Each subproblem has the form

$$\begin{array}{ll} \min_{s\in\mathbb{R}^n} & m_f^{(k)}\left(x^{(k)}+s\right) \\ \text{such that} & x^{(k)}+s\in\mathcal{F}_m^{(k)} \\ & \|s\|\leq\Delta_k \end{array}$$

where $m_f^{(k)}$ is a model function approximating the objective f, Δ_k is the trust region radius, and $\mathcal{F}_m^{(k)}$ is an approximation of the feasible region. The solution $s^{(k)}$ of the trust-region subproblem determines a *trial point* $x^{(k)}+s^{(k)}$. The objective function and constraints are then evaluated at the trial point to determine whether to accept the trial point. If the trial point is rejected, the trust region radius is decreased and a new trial point is computed by solving a revised trust-region subproblem. If the trial point is accepted, then the trust region radius may be increased or decreased for the next iteration depending on how well the sample point improved upon the previous iterate. This will be discussed in Section 5.1.

5.1 Assessing Model Accuracy and Trust Region Radius Management

In trust region methods, each iteration that evaluates a trial point must also test the accuracy of the model functions. To test the accuracy, we calculate a quantity

$$\rho_k=\frac{f(x^{(k)})-f(x^{(k)}+s^{(k)})}{{m_f}^{(k)}(x^{(k)})-{m_f}^{(k)}(x^{(k)}+s^{(k)})}, \tag{5}$$

which measures the actual improvement of the trial point $x^{(k)}+s^{(k)}$ divided by the predicted improvement. If ρ_k is negative, the trial point is rejected and the trust region radius is decreased. On the other hand, if the trial point is accepted, the trust region radius for the next iteration may still be decreased since a small value of ρ_k implies that the model functions are not sufficiently accurate. For larger values of ρ_k the trial point is accepted and the trust region radius may be increased. This strategy has been widely used within trust region frameworks such as [10] and within a derivative-free context [8].

To implement the above strategy, we define parameters $0<\gamma_{\text{min}}<\gamma_{\text{suff}}\leq 1$ as thresholds on ρ_k and $0<\omega_{\text{dec}}<1\leq\omega_{\text{inc}}$ as decrement and increment factors to determine the trust region update policy. That is, when $\rho_k<\gamma_{\text{min}}$, the trust region is decreased by a factor of ω_{dec}, and the trust region is increased by a factor of ω_{inc} during some iterations in which $\rho_k>\gamma_{\text{suff}}$.

5.2 Interpolation-Based Models

Model-based methods for derivative-free optimization construct models of a function $f(x)$ from a family of functions spanned by a set of $p+1\in\mathbb{N}$ basis functions $\Phi=\{\phi_0,\phi_1,\ldots,\phi_p\}$. Each member of this family has the form $m_f(x)=\sum_{i=0}^{p}\alpha_i\phi_i(x)$ for some scalar coefficients $\alpha_i, i\in\{0,\ldots,p\}$.

We use interpolation to choose the coefficients $\alpha=[\alpha_0,\ldots,\alpha_p]^T$ so that ${m_f}^{(k)}$ agrees with f on a set of $p+1$ sample points $Y=\{y^{(0)},y^{(1)},\ldots,y^{(p)}\}$ at which the function f has been evaluated. Thus, the coefficients α must satisfy the *interpolation condition*

$$m_f(y^{(i)})=\sum_{j=0}^{p}\alpha_j\phi_j(y^{(i)})=f(y^{(i)})\quad\forall\quad 0\leq i\leq p. \tag{6}$$

This equation can be written more compactly in the form

$$M\alpha = \bar{f}, \tag{7}$$

where $\bar{f} = [f(y^{(0)}), f(y^{(1)}), \ldots, f(y^{(p)})]^T$ and the Vandermonde matrix M is defined by

$$M = M(\Phi, Y) := \begin{bmatrix} \phi_0(y^{(0)}) & \phi_1(y^{(0)}) & \ldots & \phi_p(y^{(0)}) \\ \phi_0(y^{(1)}) & \phi_1(y^{(1)}) & \ldots & \phi_p(y^{(1)}) \\ & & \vdots & \\ \phi_0(y^{(p)}) & \phi_1(y^{(p)}) & \ldots & \phi_p(y^{(p)}) \end{bmatrix}. \tag{8}$$

SCB note: Changed notation for the Vandermonde matrix from V to M for consistency. V is used elsewhere to denote the volume of the ellipsoid and also in the eigendecomposition of matrix. The interpolation equation (7) has a unique solution if and only if M is non-singular. In this case, we say that the sample set Y is *poised* for interpolation with respect to Φ. Note also that if Y is poised with respect to one basis for a family of functions, then it is poised with respect to any basis for that family. However, even when M is non-singular, the interpolation model may be inaccurate if the Vandermonde matrix is poorly conditioned. Informally, we say that a sample set is *well-poised* if it produces accurate model functions. The following section formalizes this idea.

5.3 Sample Set Geometry

The term *geometry* refers to how the distribution of points in the sample set Y affects the accuracy of the model function. In the case of polynomial model functions, a rigorous analysis of model accuracy can be performed using *Lagrange polynomials*. Let the space of polynomials on $\mathbb{R}^n$ with degree less than or equal to d be denoted by $\mathcal{P}_n^d$ and have dimension $p+1$. The Lagrange polynomials $\ell_0, \ell_1, \ldots, \ell_p$ for the sample set Y are a basis of $\mathcal{P}_n^d$ such that

$$\ell_i(y^{(j)}) = \delta_{i,j} \tag{9}$$

where $\delta_{i,j} = \{0 \text{ if } i \neq j, \ 1 \text{ if } i = j\}$ is the Kronecker-delta function. Thus, as shown in [8], the interpolating model function $m_f(x)$ satisfies

$$m_f(x) = \sum_{j=0}^{p} f(y^{(i)})\ell_i(x). \tag{10}$$

The following definition allows us to quantify how well-poised a sample set is with respect to interpolation on a polynomial subspace.

SCB note: Corrected the definition of Λ-poised. We don't need a basis for the polynomial space in the definition since the Lagrange polynomials depend only on the space, not the basis. The definition also needs to include a set B

Definition 5.1 ~~*We say that a set Y is* Λ-poised *for a fixed constant Λ with respect to a basis Φ on the set $B \subset \mathbb{R}^n$*~~

Given a fixed constant $\Lambda > 0$ and a set $B \subset \mathbb{R}^n$, we say that a sample set Y is Λ-poised *in B for interpolation on $\mathcal{P}_n^d$ if and only if the Lagrange polynomials ℓ_i associated with Y satisfy*

$$\Lambda \geq \max_{0 \leq i \leq p} \max_{x \in B} |\ell_i(x)|. \tag{11}$$

To examine how Λ-poisedness relates to the accuracy of the interpolation model functions, we shall consider the Vandermonde matrix of a shifted and scaled sample set $\hat{Y} = \frac{1}{\Delta}(Y - \bar{y})$. The analysis is similar to that presented in [8]; however their analysis defines a particular $\hat{Y}$ by choosing $\bar{y} = y^{(0)}$ and $\Delta = \max_{1 \le i \le p} \|y^{(i)} - y^{(0)}\|$. The resulting set $\hat{Y}$ is contained in the unit ball centered at $\hat{y}^{(0)} = 0$ and has at least one point on the boundary of the ball. This approach is sensible for unconstrained optimization, where $y^{(0)}$ is typically chosen to be the current iterate and we are interested in constructing sample sets that are well-poised over the trust region $B_2(y^{(0)}; \Delta)$. However, for constrained optimization, it is disadvantageous to require the current iterate to be at the center of the sample trust region or to require that there be a sample point on the boundary of the ball. We therefore consider a more general transformation by specifying an arbitrary center $\bar{y}$ and radius Δ.

The following result from [8, Lemmas 3.8 & 3.9] shows that the Lagrange polynomials for a sample set are invariant under a shift or scaling of coordinates. Hence, the poisedness constant Λ in Theorem 5.1 is also invariant with respect to a shift or scaling of coordinates.

Lemma 5.2 *Let $Y = \{y^{(0)}, \ldots, y^{(p)}\}$ be a set that is poised for interpolation on $\mathcal{P}_n^d$. Let $\ell_i(x)$, $i = 0, \ldots, p$, be the set of Lagrange polynomials satisfying* (9) *for the set Y. Let $\Delta > 0$ and $a \in \mathbb{R}^n$ be given and define the shifted and scaled sample set $\hat{Y} = \left\{\frac{y^{(0)} - a}{\Delta}, \ldots, \frac{y^{(p)} - a}{\Delta}\right\}$. Let $\hat{\ell}_i(x)$, $i = 0, \ldots, p$ be the Lagrange polynomials satisfying* (9) *for the set $\hat{Y}$. Then for any $x \in \mathbb{R}^n$,*

$$\ell_i(x) = \hat{\ell}_i\left(\frac{x - a}{\Delta}\right), i = 0, \ldots, p.$$

We now state two technical lemmas that will be used in the proof of our next theorem.

SCB note: Maybe we don't need to restate these lemmas, and can instead just reference the lemmas from the book in our proofs .

Lemma 5.3 *([8, Lemma 3.10]) Let $\bar{\Phi}$ be the monomial basis* (12)*. There exists a number $\sigma_\infty > 0$ such that, for any choice of v satisfying $\|v\|_\infty = 1$, there exists a $y \in B_2(0; 1)$ such that $\left|v^T \bar{\Phi}(y)\right| \ge \sigma_\infty$.*

Lemma 5.4 *([8, Lemma 3.13]) Let w be a normalized right-singular vector of a nonsingular $n \times n$ matrix A corresponding to its largest singular value. Then, for any vector $r \in \mathbb{R}^n$,*

$$\|Ar\|_2 \ge |w^T r| \|A\|_2.$$

The following result, which is a straightforward generalization of [8, Theorem 3.14], establishes a relationship between the Λ-poisedness of Y on $B_2(y^{(0)}; \Delta))$ and the condition number of the Vandermonde matrix $\hat{M} = M(\bar{\Phi}, \hat{Y})$, where $\bar{\Phi}$ denotes the monomial basis for $\mathcal{P}_n^2$. Specifically,

$$\bar{\Phi} = \{\bar{\phi}_0, \ldots, \bar{\phi}_p\} = \{1, x_1, \ldots, x_n, x_1^2/2, \ldots x_n^2/2, x_1 x_2, \ldots, x_{n-1} x_n\}. \tag{12}$$

Theorem 5.5 SCB note: I reworked this theorem to allow for an arbitrary center, and to remove the requirement that the sample set is contained in the ball *Let $\Delta > 0$ and $\bar{y} \in \mathbb{R}^n$ be given and let Y be a sample set consisting of $p + 1 = (n + 1)(n + 2)/2$ points in $\mathbb{R}^n$. Define the shifted and scaled set $\hat{Y} = \frac{1}{\Delta}(Y - \bar{y})$ and let $\hat{M} = M(\bar{\Phi}, \hat{Y})$ where $\bar{\Phi}$ is the monomial basis* (12)*. If $\hat{M}$ is non-singular, and $\left\|\hat{M}^{-1}\right\|_2 \le \Lambda$, then Y is $\hat{\Lambda}$-poised in $B_2(\bar{y}; \Delta)$, where $\hat{\Lambda} = \Lambda (p + 1)^{-\frac{1}{2}}$. Conversely, there exists a constant κ_Λ depending only on Λ and p such that if Y is Λ-poised on $B_2(\bar{y}; \Delta)$, then $\left\|\hat{M}^{-1}\right\| \le \kappa_\Lambda \Lambda$.*

Proof:

(of Theorem 5.5) Suppose that $\left\|\hat{M}^{-1}\right\| \leq \Lambda$. Let $x \in B_2(0;1)$ be arbitrary and let $\ell(x) = (\ell_0(x), \ldots, \ell_p(x))^T$, where $\ell_i, i = 0, \ldots, p$ denote the Lagrange polynomials for $\hat{Y}$. As shown in [8], we can write $\ell(x) = \hat{M}^{-T}\bar{\phi}(x)$, where $\bar{\phi}(x) = (\bar{\phi}_0(x), \ldots, \bar{\phi}_p(x))$. Thus,

$$\begin{aligned}
\|\ell(x)\|_\infty &\leq \left\|\hat{M}^{-T}\right\|_\infty \left\|\bar{\phi(x)}\right\|_\infty \\
&\leq (p+1)^{-\frac{1}{2}} \left\|\hat{M}^{-1}\right\|_2 \left\|\bar{\phi}(x)\right\|_\infty \\
&\leq \Lambda\,(p+1)^{-\frac{1}{2}} \max\left\{1, |x_1|, \ldots, |x_n|, \frac{1}{2}x_1^2, \ldots, \frac{1}{2}x_n^2, x_1x_2, \ldots, x_{n-1}x_n\right\} \\
&\leq \Lambda\,(p+1)^{-\frac{1}{2}} = \hat{\Lambda}.
\end{aligned}$$

Since x was chosen arbitrarily from $B_2(0;1)$, it follows that $\hat{Y}$ is Λ-poised over $B_2(0;1)$. Since Λ-poisedness is invariant with respect to shifting and scaling, we conclude that Y is Λ-poised over $B_2(\bar{y};\Delta)$.

Conversely, suppose Y is Λ-poised over $B_2(\bar{y};\Delta)$. Then $\hat{Y}$ is Λ-poised over $B_2(0;1)$. Let v be a right singular vector of $M^{-1}\left(\bar{\Phi}, Y\right)$ corresponding to its largest singular value and normalized so that $\|v\|_\infty = 1$. By Lemma 5.3, there exists $y \in B_2(0;1)$ such that $\left|v^T\bar{\Phi}(y)\right| \geq \sigma_\infty$, where $\sigma_\infty > 0$ is the constant guaranteed by Lemma 5.3. Then, because $\hat{Y}$ is Λ-poised on $B_2(0;1)$, we have that

$$\Lambda \geq \|l(y)\|_\infty = \left\|M\left(\bar{\Phi}, Y\right)^{-T}\bar{\Phi}(y)\right\|_\infty \geq (p+1)^{-\frac{1}{2}}\left\|M\left(\bar{\Phi}, \hat{Y}\right)^{-T}\bar{\Phi}(y)\right\|.$$

Applying Lemma 5.4 with $A = \bar{M}\left(\bar{\Phi}, \hat{Y}\right)^{-T}$, $w = v$, and $r = \bar{\Phi}(y)$, we have

$$\begin{aligned}
\Lambda \geq (p+1)^{-\frac{1}{2}}\left\|M\left(\bar{\Phi}, \hat{Y}\right)^{-T}\bar{\Phi}(y)\right\| &\geq (p+1)^{-\frac{1}{2}}\left|v^T\bar{\Phi}(y)\right|\left\|M\left(\bar{\Phi}, Y\right)^{-T}\right\| \\
&\geq (p+1)^{-\frac{1}{2}}\,\sigma_\infty\left\|M\left(\bar{\Phi}, Y\right)^{-T}\right\|.
\end{aligned}$$

The conclusion follows with $\kappa_\Lambda = \frac{\sqrt{p+1}}{\sigma_\infty}$. □

Having proven this relationship between the Λ-poisedness of Y and the Vandermonde matrix for $\hat{Y}$, we now establish error bounds for the model function and its derivatives based on the Λ-poisedness of Y. We first state the following technical lemma from [11]:

Lemma 5.6 *([11, Lemma 4.1.14]) Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz twice continuously differentiable on an open convex set $\Omega \subset \mathbb{R}^n$, with Lipschitz constant L_{∇^2}. Then, for any $x + p \in \Omega$,*

$$\left|f(x+p) - \left(f(x) + \nabla f(x)^T p + \frac{1}{2}p^T\nabla^2 f(x)p\right)\right| \leq \frac{L_{\nabla^2}}{6}\|p\|^3.$$

Theorem 5.7 *Let $\Delta > 0$ be given. Let $Y = \left\{y^{(0)}, y^{(1)}, \ldots y^{(p)}\right\}$ be a set of $p + 1 = \frac{(n+1)(n+2)}{2}$ points in $B_2(0;\Delta)$. Let $\hat{Y} = \frac{1}{\Delta}Y$ be a scaled sample set and let $\hat{M} = M(\bar{\Phi}, \hat{Y})$, where $\bar{\Phi}$ is the monomial basis for the quadratic polynomials. Let Ω be an open convex set containing $B_2(0;\Delta)$. Suppose that f is Lipschitz*

twice continuously differentiable on Ω with Lipschitz constant L_{∇^2}. Let $m : \mathbb{R}^n \to \mathbb{R}$ be a quadratic function interpolating f on Y; that is

$$m\left(y^{(i)}\right) = f\left(y^{(i)}\right) \quad i = 0, \ldots, p. \tag{13}$$

If $\hat{M}$ is nonsingular, then there exists a constant κ depending only on p and L_{∇^2} such that the following error bounds holds for all $y \in B_2(0; \Delta)$:

$$|m(y) - f(y)| \leq \kappa \left\|\hat{M}^{-1}\right\|_2 \Delta^3, \tag{14}$$

$$\|\nabla m(y) - \nabla f(y)\| \leq \kappa \left\|\hat{M}^{-1}\right\|_2 \Delta^2, \tag{15}$$

$$\left\|\nabla^2 m(y) - \nabla^2 f(y)\right\| \leq \kappa \left\|\hat{M}^{-1}\right\|_2 \Delta. \tag{16}$$

Proof:

Let m have the form $m(x) = c + g^T + \frac{1}{2}x^T H x$ for some $c \in \mathbb{R}, g \in \mathbb{R}^n$, and symmetric matrix $H \in \mathbb{R}^{n \times n}$. Define error functions e_f, e_g and e_h by

$$\begin{aligned} e_f(x) &= m(x) - f(x) = c + g^T x + \frac{1}{2}x^T H x - f(x), \\ e_g(x) &= \nabla m(x) - \nabla f(x) = g + Hx - \nabla f(x), \\ e_h(x) &= \nabla^2 m(x) - \nabla^2 f(x) = H - \nabla^2 f(x). \end{aligned} \tag{17}$$

Let $y \in B_2(0; \Delta)$ be arbitrary. For each $0 \leq i \leq p$, we can subtract (17) from (13) to find that

$$\begin{aligned} f\left(y^{(i)}\right) - f(y) - e_f(y) &= m\left(y^{(i)}\right) - m(y) \\ &= g^T\left(y^{(i)} - y\right) + \frac{1}{2}\left(y^{(i)}\right)^T H y^{(i)} - \frac{1}{2}y^T H y \qquad (18) \\ &= g^T\left(y^{(i)} - y\right) + \frac{1}{2}\left(y^{(i)} - y\right)^T H\left(y^{(i)} - y\right) + \left(y^{(i)} - y\right)^T H y. \qquad (19) \end{aligned}$$

By defining

$$e_O^{(i)}(y) = f\left(y^{(i)}\right) - \left[f(y) + \nabla f(y)^T\left(y^{(i)} - y\right) + \frac{1}{2}\left(y^{(i)} - y\right)^T \nabla^2 f(y)\left(y^{(i)} - y\right)\right].$$

we can use Theorem 5.6 with $p = y^{(i)} - y$ to conclude that

$$\left|e_O^{(i)}(y)\right| \leq \frac{1}{6}L_{\nabla^2}\left\|y^{(i)} - y\right\|^3 \leq \frac{4}{3}L_{\nabla^2}\Delta^3. \tag{20}$$

Furthermore, notice that

$$\begin{aligned} f\left(y^{(i)}\right) - f(y) &= \nabla f(y)^T\left(y^{(i)} - y\right) + \frac{1}{2}\left(y^{(i)} - y\right)^T \nabla^2 f(y)\left(y^{(i)} - y\right) + e_O^{(i)}(y) \\ &= \left(Hy + g - e_g(y)\right)^T\left(y^{(i)} - y\right) + \frac{1}{2}\left(y^{(i)} - y\right)^T \nabla^2 f(y)\left(y^{(i)} - y\right) + e_O^{(i)}(y). \end{aligned}$$

Subtracting this from (19), we find that

$$e_O^{(i)}(y) - e_f(y) = e_g(y)^T \left(y^{(i)} - y\right) + \frac{1}{2}\left(y^{(i)} - y\right)^T \left(H - \nabla^2 f(y)\right)\left(y^{(i)} - y\right)$$

or

$$\begin{aligned} e_O^{(i)}(y) = &\left[e_f(y) - e_g(y)^T y + \frac{1}{2} y^T \left(H - \nabla^2 f(y)\right) y\right] \\ &+ \left[e_g(y) + \left(H - \nabla^2 f(y)\right) y\right]^T y^{(i)} \\ &+ \frac{1}{2}\left(y^{(i)}\right)^T \left(H - \nabla^2 f(y)\right) y^{(i)}. \end{aligned} \tag{21}$$

To put this equation in matrix form, we introduce some notation. Given some $n \times n$ matrix A, let $\mathrm{vec} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{\frac{n(n+1)}{2}}$ vectorize the upper triangular portion of A:

$$\mathrm{vec}(A) = \left(\frac{1}{2}A_{1,1}\ ,\ \frac{1}{2}A_{2,2}\ ,\ \ldots\ ,\ \frac{1}{2}A_{n,n}\ ,\ A_{1,2}\ ,\ A_{1,3}\ ,\ \ldots\ ,\ A_{n-1,n}\right)^T$$

in a similar order to that used in $\bar{\Phi}$ (see (12)). Then, with $e_O(y) = \left(e_O^{(0)}(y), e_O^{(1)}(y), \ldots, e_O^{(p)}(y)\right)^T$, we can write (21) in matrix form as

$$M\left(\bar{\Phi}, Y\right)\begin{pmatrix} e_f(y) + y^T e_g(y) + y^T e_h(y) y \\ e_g(y) + e_h(y) y \\ \mathrm{vec}\left(e_h(y)\right) \end{pmatrix} = e_O(y).$$

Thus,

$$\begin{aligned} \left\| M(\bar{\Phi}, Y)\begin{pmatrix} e_f(y) + y^T e_g(y) + y^T e_h(y) y \\ e_g(y) + e_h(y) y \\ \mathrm{vec}\left(e_h(y)\right) \end{pmatrix}\right\|_2 &\leq \|e_O(y)\|_2 \\ &\leq \sqrt{p+1}\,\|e_O(y)\|_\infty \\ &\leq \frac{4}{3}\sqrt{p+1} L_{\nabla^2} \Delta^3 \quad \text{(by (20))} \\ &= \kappa_0 \Delta^3, \end{aligned} \tag{22}$$

where $\kappa_0 = \frac{4}{3}\sqrt{p+1} L_{\nabla^2}$.

To remove the dependence on Δ of $M(\bar{\Phi}, Y)$, observe that

$$\hat{M} = M\left(\bar{\Phi}, \hat{Y}\right) = M\left(\bar{\Phi}, Y\right)\begin{pmatrix} 1 & 0 & 0 \\ 0 & \Delta^{-1} I_n & 0 \\ 0 & 0 & \Delta^{-2} I_{\frac{n(n+1)}{2}} \end{pmatrix}. \tag{23}$$

Thus, (22) can be rewritten as

$$\left\| \hat{M}\begin{pmatrix} e_f(y) + y^T e_g(y) + y^T e_H(y) y \\ \Delta\left(e_g(y) + e_H(y) y\right) \\ \Delta^2 \mathrm{vec}\left(e_H(y)\right) \end{pmatrix}\right\|_2 \leq \kappa_0 \Delta^3,$$

which implies that

$$\left\|e_f(y) + y^T e_g(y) + y^T e_H(y) y\right\|_2 \leq \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta^3, \tag{24}$$

$$\Delta \left\|e_g(y) + e_H(y) y\right\|_2 \leq \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta^3, \tag{25}$$

$$\Delta^2 \left\|\mathrm{vec}\left(e_H(y)\right)\right\|_2 \leq \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta^3. \tag{26}$$

We see that (26) immediately provides

$$\left\|e_H(y)\right\|_2 \leq \left\|e_H(y)\right\|_F \leq \sqrt{2} \left\|\mathrm{vec}\left(e_H(y)\right)\right\|_2 \leq \sqrt{2}\kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta$$

where $\|\cdot\|_F$ is the Frobenius norm $\|A\|_F = \sqrt{\sum_i \sum_j A_{i,j}^2}$. Similarily, the triangle inequality and (25) imply

$$\begin{aligned}
\left\|e_g\left(y\right)\right\|_2 &\leq \left\|e_g(y) + e_H(y) y\right\|_2 + \left\|e_H\left(y\right)\right\|_2 \left\|y\right\|_2 \\
&\leq \kappa_0 \left\|\hat{M}^{-1}(\bar{\Phi}, Y)\right\|_2 \Delta^2 + \left[\sqrt{2}\kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta\right] \Delta \\
&\leq \left(1 + \sqrt{2}\right) \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta^2.
\end{aligned}$$

Finally, using (24), we have

$$\begin{aligned}
\left|e^f\left(y\right)\right| &\leq \left\|e_g(y)\right\| \Delta + \left\|e_H(y)\right\| \Delta^2 + \kappa_0 \left\|\hat{M}^{-1}(\bar{\Phi}, Y)\right\|_2 \Delta^3 \\
&\leq \left(2 + 2\sqrt{2}\right) \kappa_0 \left\|\hat{M}^{-1}\right\| \Delta^3.
\end{aligned}$$

Thus, the conclusion of the theorem holds with $\kappa = \frac{4}{3}(2 + 2\sqrt{2})\sqrt{p+1} L_{\nabla^2}$.
□

Corollary 5.8 *Let $\Delta > 0$ and $a \in \mathbb{R}^n$ be given. Let $Y = \left\{y^{(0)}, y^{(1)}, \ldots y^{(p)}\right\}$ be a set of $p+1 = \frac{(n+1)(n+2)}{2}$ points in $B_2\left(a; \Delta\right)$. Let $\hat{Y} = \frac{1}{\Delta}\left(Y - a\right)$ be a shifted and scaled sample set and let $\hat{M} = M(\bar{\Phi}, \hat{Y})$, where $\bar{\Phi}$ is the monomial basis for the quadratic polynomials. Let Ω be an open convex set containing $B_2(a; \Delta)$. Suppose that f is Lipschitz twice continuously differentiable on Ω with Lipschitz constant L_{∇^2}. Let $m : \mathbb{R}^n \to \mathbb{R}$ be a quadratic function interpolating f on Y.*

If $\hat{M}$ is nonsingular, then there exists a constant κ_0 depending only on p and L_{∇^2} such that the following error bounds holds for all $y \in B_2\left(a; \Delta\right)$:

$$\left|m(y) - f(y)\right| \leq \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta^3, \tag{27}$$

$$\left\|\nabla m(y) - \nabla f(y)\right\| \leq \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta^2, \tag{28}$$

$$\left\|\nabla^2 m(y) - \nabla^2 f(y)\right\| \leq \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta. \tag{29}$$

Proof:

Define $\bar{Y} = Y - a$ and let $\bar{\ell}_i(x), i = 0, \ldots, p$ be the Lagrange polynomials for the shifted set $\bar{Y}$. By Lemma 5.2 $\ell_i(x) = \bar{\ell}_i\left(\frac{x-a}{\Delta}\right)$, $i = 0, \ldots, p$. Define $\bar{f}(x) = f(\Delta x + a)$. Let $\bar{m}(x)$ be the quadratic polynomial interpolating $\bar{f}$ on the set $\bar{Y}$.

Let $y \in B_2(a;\Delta)$ be arbitrary and define $\bar{y} = y - a$. Observe that $\bar{y} \in B_2(0;\Delta)$. By (10),

$$m(y) = \sum_{i=0}^{p} f(y^{(i)})\ell_i(y) = \sum_{i=0}^{p} \bar{f}(y^{(i)} - a)\bar{\ell}_i(\bar{y}) = \bar{m}(\bar{y}).$$

Note also that $\nabla m(y) = \nabla \bar{m}(\bar{y})$, $\nabla^2 m(y) = \nabla^2 \bar{m}(\bar{y})$, $\nabla f(y) = \nabla \bar{f}(\bar{y})$ and $\nabla^2 f(y) = \nabla^2 \bar{f}(\bar{y})$. Applying Theorem 5.7, for all $y \in B_2(a;\Delta)$, we have

$$|m(y) - f(y)| = |\bar{m}(\bar{y}) - \bar{f}(\bar{y})| \leq \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta^3, \tag{30}$$

$$|\nabla m(y) - \nabla f(y)| = |\nabla \bar{m}(\bar{y}) - \nabla \bar{f}(\bar{y})| \leq \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta^3, \tag{31}$$

$$|\nabla^2 m(y) - \nabla^2 f(y)| = |\nabla^2 \bar{m}(\bar{y}) - \nabla^2 \bar{f}(\bar{y})| \leq \kappa_0 \left\|\hat{M}^{-1}\right\|_2 \Delta^3, \tag{32}$$

where κ is the constant guaranteed by Theorem 5.7. □

Corollary 5.9 *Let $\Delta > 0$ and $a \in \mathbb{R}^n$ be given. Let $Y = \left\{y^{(0)}, y^{(1)}, \ldots y^{(p)}\right\}$ be a set of $p + 1 = \frac{(n+1)(n+2)}{2}$ points in $\mathbb{R}^n$. Let Ω be an open convex set containing $B_2(a;\Delta) \cup Y$. Suppose that f is Lipschitz twice continuously differentiable on Ω with Lipschitz constant L_{∇^2}. Then there exists a constant $\kappa > 0$, independent of Y adn f such that if Y is Λ-poised in $B_2((a;\Delta)$ for interpolation on $\mathcal{P}_n^d$, then the following error bounds hold for all $y \in B_2(a;\Delta)$, where $m(x)$ is the quadratic function interpolating f on Y.*

$$|m(y) - f(y)| \leq \kappa \Lambda \Delta^3, \tag{33}$$

$$\|\nabla m(y) - \nabla f(y)\| \leq \kappa \Lambda \Delta^2, \tag{34}$$

$$\left\|\nabla^2 m(y) - \nabla^2 f(y)\right\| \leq \kappa \Lambda \Delta. \tag{35}$$

Proof:

Let $\hat{Y} = \frac{1}{\Delta}(Y - a)$ and $\hat{M} = M(\bar{\Phi}, \hat{Y})$. By Theorem 5.5, there exists a constant κ_Λ such that $\left\|\hat{M}^{-1}\right\| \leq \kappa_\Lambda \Lambda$. Let κ_0 be the constant guaranteed by Theorem 5.8. Then by Theorem 5.8, the conclusion holds with $\kappa = \kappa_0 \kappa_\Lambda$. □

The following result, which is a special case of [12, Corollary 4.6]. SCB note: Is the following Lemma true for any norm?

Lemma 5.10 *Let f be twice continuously differentiable on an open convex set Ω and suppose that the Hessian of f is Lipschitz continuous on Ω with constant L. Let $\ell_0, \ell_1, \ldots, \ell_p$ denote the Lagrange polynomials for a poised sample set $Y = \{y^{(0)}, \ldots, y^{(p)}\} \subset \Omega$. Let $m(x)$ be the model function determined by* (10)*. Then, for any $x \in \Omega$ the following error bounds hold:*

- $|f(x) - m(x)| \leq \sum_{i=0}^{p} \frac{L}{2}\left\|y^{(i)} - x\right\|^3 |\ell_i(x)|$
- $\|\nabla f(x) - \nabla m(x)\| \leq \sum_{i=0}^{p} \frac{L}{2}\left\|y^{(i)} - x\right\|^3 \|\nabla \ell_i(x)\|$
- $\left\|\nabla^2 f(x) - \nabla^2 m(x)\right\| \leq \sum_{i=0}^{p} \frac{L}{2}\left\|y^{(i)} - x\right\|^3 \left\|\nabla^2 \ell_i(x)\right\|$.

5.4 Model Improvement Algorithms

Efficient implementations of model-based methods re-use sample points from previous iterations that fall within (or at least near) the current trust region. New points are then added to the sample set using a model improvement algorithm as described in [8] and stated here within Algorithm 1.

The model improvement algorithm starts with a set of $p+1$ sample points that have been shifted and scaled so that they lie within a ball $B_2(0;\Delta)$, with $\Delta \geq 1$. The algorithm then uses LU factorization with partial pivoting of the associated Vandermonde matrix to construct a set of pivot polynomials $\{u_0, \ldots, u_p\}$ that are closely related to the Lagrange polynomials.

Each iteration of the algorithm identifies a point to include in the final sample set. In particular, on the ith iteration, the points $y^{(0)}, \ldots, y^{(i-1)}$ have already been included. If a point $y^{(j)}$, $j \geq i$ from the original set can be found such that $u_i(y^{(j)})$ has a sufficiently large magnitude (i.e., $|u_i(y^{(j)})| \geq \xi_{\min}$), then that point is added to the final sample set (by swapping it with $y^{(i)}$). However, if no such point can be found, it indicates that including any of the remaining points in the final sample set would result in a poorly poised set. Therefore, the point $y^{(i)}$ is replaced by a new point that is obtained by maximizing $|u_i(x)|$ over the unit ball $B_2(0;1)$. The pivot polynomials are then updated so that $u_j(y^{(i)}) = 0$ for $j > i$. At the successful completion of the algorithm, the final set of points Y is Λ-poised over $B_2(0;\Delta)$, where Λ depends on Δ, n and $\xi_{\min}$, and is inversely proportional to $\xi_{\min}$ (see Theorem 5.13 below). SCB note: Discuss how to choose initial sample set.

Note. Typically, we have fewer than $p+1$ previously evaluated sample points within the trust region at the beginning of each iteration. Since the Model Improvement Algorithm requires a starting set of $p+1$ points, we add copies of $y^{(0)}$ to create a set with $p+1$ points.

Algorithm 1: Model Improvement Algorithm

Step 0 (Initialization)
Given $\xi_{\text{min}} \in (0, 1/4]$, $\Delta \geq 1$, a set $Y = \left\{y^{(0)}, y^{(1)}, \ldots, y^{(p)}\right\} \subset B_2(0;\Delta)$ of $p+1$ points, initialize $i = 0$, and let $u_i(x) = \bar{\phi}_i(x)$ be the monomial basis for $\mathcal{P}_n^d$.

Step 1 (Pivot)
If $i = 0$, go to Step 3.
Compute the next pivot index $j_i^{\max} = \arg\max_{i \leq j \leq |Y|-1} \left|u_i\left(y^{(j)}\right)\right|$, and swap points $y^{(i)}$ and $y^{(j_i^{\max})}$ within Y.

Step 2 (Check threshold)
If $\left|u_i\left(y^{(i)}\right)\right| \geq \xi_{\text{min}}$ then go to Step 3.
If $\left|u_i\left(y^{(i)}\right)\right| < \xi_{\text{min}}$, then replace $y^{(i)} \leftarrow \arg\max_{x \in B_2(0;1)} |u_i(x)|$.
If $\left|u_i\left(y^{(i)}\right)\right| < \xi_{\text{min}}$ after this replacement, **Stop**: ξ_{min} was chosen too small.

Step 3 (Gaussian elimination)
For $j = i+1, \ldots, p$
Set $u_j(x) \leftarrow u_j(x) - \frac{u_j\left(y^{(i)}\right)}{u_i\left(y^{(i)}\right)} u_i(x)$
If $i = p$ then **Stop**, otherwise. Set $i \leftarrow i+1$ and go to Step 1

Notice that Algorithm 1 constructs a set of pivot polynomials $u = \{u_0, u_1, \ldots, u_p\}$, and that $M(u, Y)$ is the upper triangular factor of the LU factorization of $M(\bar{\Phi}, Y)$. The first instruction of Step 1 of Algorithm 1 ensures that the first point is not replaced. The first basis polynomial is always 1, so we can be sure that the first pivot will not need to be replaced.

Observe that Algorithm 1 uses a threshold parameter $\xi_{\min}$ to select the next sample point. If $\xi_{\min}$ is too large, it is possible that there might not exist a point $x \in B_2(0;1)$ in the unit ball such that $\|u_i(x)\| \geq \xi_{\min}$. In this case, the algorithm stops prematurely in Step 2 with a certification that $\xi_{\min}$ is too small. However, in Lemma 5.12 below, we show that this cannot happen as long as $0 < \xi_{\min} \leq \frac{1}{4}$. To establish the result, we first need the following Lemma from [8].

Lemma 5.11 *([8, Lemma 6.7])*
Let $v^T\bar{\Phi}(x)$ be a quadratic polynomial, where $\|v\|_\infty = 1$ and $\bar{\Phi}$ is the monomial basis. Then,

$$\max_{x \in B_2(0;1)} \left\| v^T \bar{\Phi}(x) \right\| \geq \frac{1}{4}.$$

Lemma 5.12 *For any given $\xi_{\min} \in (0, 1/4]$, Algorithm 1 computes a set Y of $p+1$ points in the ball $B_2(0;\Delta)$ for which the pivots of the LU factorization of $M = M(\bar{\Phi}, Y)$ satisfy*

$$\left\| u_i(y^{(i)}) \right\| \geq \xi_{\min}, \quad i = 0, \ldots, p.$$

Proof:

The first pivot polynomial is $u_0(x) = 1 = (1, 0, \ldots, 0)\cdot\bar{\Phi}(x)$, so that $\left\|u_0\left(y^{(0)}\right)\right\| = 1 \geq \xi_{\min}$. For $i > 1$, the $(i+1)$st pivot polynomial $u_i(x)$ can be expressed as $v^T\bar{\Phi}(x)$, where $v = (v_0, \ldots, v_{i-1}, 1, 0, \ldots, 0)^T$. Observe that $\|v\|_\infty \geq 1$. Let $\bar{v} = v/\|v\|_\infty$. Then by Lemma 5.11,

$$\max_{x \in B_2(0;1)} u_i(x) = \max_{x \in B_2(0;1)} \left| v^T \bar{\phi}(x) \right| = \|v\|_\infty \max_{x \in B_2(0;1)} \left| \bar{v}^T \bar{\phi}(x) \right| \geq \frac{1}{4}\|v\|_\infty \geq \frac{1}{4}.$$

It follows that the algorithm does not stop in Step 2 for $\xi_{\min} \leq 1/4$. Hence, at the end of the i-th iteration, we have that $\left|u_i\left(y^{(i)}\right)\right| \geq \xi_{\min}$. Moreover, after the i iteration has been completed, u_i, and $y^{(i)}$ are never altered. Thus, the algorithm terminates with $\left|u_i\left(y^{(i)}\right)\right| \geq \xi_{\min}$ for $i = 0, \ldots, p$.

Finally, observe that all points in the final sample set Y are selected either from the original sample set, which is contained in $B_2(0;\Delta)$, or are obtained in Step 2 from within $B_2(0;1)$. Since $\Delta \geq 1$, it follows that $Y \subset B_2(0;\Delta)$.
□

Theorem 5.13 *Let $\Delta \geq 1$, and suppose that Algorithm 1 is called on a sample set $Y = \left\{y^{(0)}, y^{(1)}, \ldots, y^{(p)}\right\} \subset B_2(0;\Delta)$ resulting in a new sample set $\hat{Y}$, and let $\bar{\Phi}$ be the monomial basis* (12)*. Then,*

- $\left\| M(\bar{\Phi}, \hat{Y})^{-1} \right\|_2$ *as defined by* (8)*, is bounded by a constant depending only on p and ξ_{min}.*
- *The resulting sample set $\hat{Y}$ is Λ-poised over $B_2(0;\Delta)$ for quadratic interpolation, where $\Lambda > 0$ is a constant that depends only on Δ, ξ_{min} and n and is inversely proportional to ξ_{min}.*
- *The initial point $y^{(0)}$ is retained in the resulting sample set: $y^{(0)} \in \hat{Y}$.*

Proof:

Note that Algorithm 1 never swaps or replaces $y^{(0)}$. By Lemma 5.12, $\hat{Y} \subset B_2(0;\Delta)$ and the pivots in the LU-factorization of $M(\bar{\phi}, Y)$ are bounded below by $\xi_{\min}$; that is, $\left|u_i\left(y^{(i)}\right)\right| \geq \xi_{\min},\ \forall 0 \leq i \leq p$. Thus, by [8, Section 6.7, Exercise 3], we have that

$$\left\|M\left(\bar{\Phi}, Y\right)^{-1}\right\| \leq \frac{\sqrt{p+1}\epsilon_{growth}}{\xi_{\min}}$$

where ϵ_{growth} is a potentially large, but finite estimate of the growth factor in the LU factorizations. By Theorem 5.5, using the fact that $\Delta \geq 1$, we have that Y is Λ-poised in $B_2(0;\Delta)$, where

$$\Lambda = \left(\frac{\sqrt{p+1}\epsilon_{growth}}{\xi_{\min}}\right)\sqrt{p+1}\frac{\Delta^2}{2} = \frac{(p+1)\epsilon_{growth}}{2\xi_{\min}}\Delta^2.$$

□

5.5 Criticality Measure

SCB note: Simplify this section, since we are dealing only with linear constraints.

A key component of any optimization algorithm is the *criticality measure*, which is used to define stopping criteria for the algorithm. The criticality measure $\chi(x^{(k)})$ at an iterate $x^{(k)}$ is a nonnegative quantity that is zero if and only if $x^{(k)}$ satisfies necessary optimality conditions. For unconstrained problems, a typical choice of criticality measure for classical (gradient-based) algorithms is given by $\chi(x) = \|\nabla f(x)\|$ since the first order optimality condition is $\nabla f(x) = 0$. The algorithm then terminates when $\chi(x) < \tau_\xi$, where $\tau_\xi > 0$ is a stopping tolerance.

For derivative-free optimization, $\nabla f(x)$ is not available, so the true criticality measure is replaced by the model criticality measure $\chi_m^{(k)}(x) = \left\|\nabla {m_f}^{(k)}(x)\right\|$. Note however that $\chi_m^{(k)}$ is only an accurate approximation of $\chi(x)$ if the gradient of the model function $\nabla {m_f}^{(k)}(x)$ is a close approximation to $\nabla f(x)$. For this reason, derivative-free algorithms require not only that $\chi_m^{(k)}(x)$ is small, but also that the model functions are accurate. To accomplish this, we require poised sample sets as discussed in Section 5.3 with the trust-region radius converging to zero. Once the criticality measure has reached a small enough threshold $\tau_\xi > 0$ and the trust region is small enough ($\Delta_k < \tau_\Delta$), we can terminate the algorithm.

If the feasible region $\mathcal{F}$ is convex, a classic criticality measure (see, for example [9] [10]) is given by

$$\chi(x) = \left\|x - \text{Proj}_{\mathcal{F}}\left(x - \nabla f\left(x\right)\right)\right\|.$$

If the feasible region is not convex, the projection operation is not well-defined, so this criticality measure is not helpful. Moreover, even with a convex feasible region, we need a way to determine the projection, so we prefer a criticality measure based on the functions defining the constraints. In particular, for a first order criticality measure at $x^{(k)}$, we define

$$F_c^{(k)} = \left\{x \in \mathbb{R}^n \middle| c_i\left(x^{(k)}\right) + \nabla c_i\left(x^{(k)}\right)^T\left(x - x^{(k)}\right) \leq 0\ \forall i \in [m]\right\} \tag{36}$$

and use

$$\chi_c^{(k)}(x) = \left\|x - \text{Proj}_{F_c^{(k)}}\left(x - \nabla f\left(x\right)\right)\right\|. \tag{37}$$

Note that $\chi_c^{(k)}(x^{(k)}) = 0$ if and only if $x^{(k)}$ satisfies the first order Karush-Kuhn-Tucker conditions. Unfortunately, without derivative information, we cannot calculate $\chi_c^{(k)}$ directly, so in our algorithms we approximate it with

$$\chi_m^{(k)} = \left\| x^{(k)} - \text{Proj}_{\mathcal{F}_m^{(k)}} \left(x^{(k)} - \nabla {m_f}^{(k)} \left(x^{(k)} \right) \right) \right\|, \tag{38}$$

where $\mathcal{F}_m^{(k)}$ is the model feasible region defined by

$$\mathcal{F}_m^{(k)} = \left\{ x \in \mathbb{R}^n \middle| m_{c_i}^{(k)}(x) \leq 0 \quad \forall 1 \leq i \leq m \right\}. \tag{39}$$

This quantity measures how far the current iterate $x^{(k)}$ is from satisfying the first order optimality conditions for the model function ${m_f}^{(k)}$. In turn, as $\Delta_k \to 0$, the model ${m_f}^{(k)}$ better approximates f, $\mathcal{F}_m^{(k)}$ better approximates $\mathcal{F}$ and $x^{(k)}$ approaches a critical point for f.

6 Algorithm Description

We now describe our algorithm for solving the linearly constrained problem

$$\begin{array}{ll} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & Ax \leq b \end{array} \tag{40}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a black-box function, A is an $m \times n$ matrix and $b \in \mathbb{R}^m$. For convenience, we assume that each row of the constraint matrix has been normalized so that $\|A_i\| = 1$ for each $i \in [m]$.

SCB note: It would sure be nice to get rid of $\frac{1}{2}\delta_k^2$ in the definition of $T_{\text{sample}}^{(k)}$).

A key idea of our method is to define an ellipsoidal sample trust region $T_{\text{sample}}^{(k)}$ at each iteration from which new sample points will be chosen. The sample trust region has the form

$$T_{\text{sample}}^{(k)} = \left\{ x \in \mathbb{R}^n \middle| \left(x - c^{(k)} \right)^T Q^{(k)} \left(x - c^{(k)} \right) \leq \frac{1}{2}\delta_k^2 \right\}, \tag{41}$$

where we choose the symmetric positive definite matrix $Q^{(k)}$, vector $c^{(k)} \in \mathbb{R}^n$, and constant $\delta_k > 0$ so that the sample region lies within the intersection of the feasible region and an outer trust region defined by

$$T_{\text{out}}^{(k)} = B_\infty \left(x^{(k)}; \Delta_k \right), \tag{42}$$

where $x^{(k)}$ is the current iterate and Δ_k is the current trust region radius. After $T_{\text{sample}}^{(k)}$ is constructed, sample points are selected by applying the model improvement algorithm (Algorithm 1) to a transformed problem in which the ellipsoid $T_{\text{sample}}^{(k)}$ is mapped onto the unit ball.

To make this concrete, let $Q^{(k)} = VD^2V^T$ be the eigendecomposition of $Q^{(k)}$, where $V^TV = I$ and D is a diagonal matrix with positive entries. Define $T : \mathbb{R}^n \times \mathbb{R}_+^n \to \mathbb{R}^n$ by $T(x; \delta) = \frac{\delta}{\delta_k} DL^T \left(x - c^{(k)} \right)$. Note that $T(\cdot, \sqrt{2})$ is invertible and maps $T_{\text{sample}}^{(k)}$ onto the unit ball $B_2(0; 1)$: SCB note: maybe we can delete the following

$$\delta^2 \geq \| T\left(x; \sqrt{2}\delta \right) \|^2 = \left\| \frac{\sqrt{2}\delta}{\delta_k} DV^T (x - c^{(k)}) \right\|^2 = \frac{2\delta^2}{\delta_k^2} \left(x - c^{(k)} \right)^T Q \left(x - c^{(k)} \right)$$

$$\iff \left(x - c^{(k)}\right)^T Q \left(x - c^{(k)}\right) \leq \frac{1}{2}\delta_k^2.$$

To apply Algorithm 1, we calculate a transformed sample set $\hat{Y}$ by computing $T(y^{(i)}; \sqrt{2})$ for each $y^{(i)} \in Y$, where Y is a starting set of $p+1$ previously evaluated sample points.

SCB note: We need to specify somewhere how we choose Y for each iteration. For example, we might say the following: We construct Y by choosing at most p sample points from the previous iteration that lie within $T^{(k)}_{\text{sample}}$ and then add enough copies of the current iterate $x^{(k)}$ to create a set with $p+1$ points.

The complete process is formalized in Algorithm 2.

Algorithm 2: Model Construction Algorithm

Step 0 (Initialization)
Given a function $f : \mathbb{R}^n \to \mathbb{R}$; ellipsoid parameters $Q^{(k)} \succ 0$, $c^{(k)}$, and δ_k; and sample set Y.

Step 1 (Construct the Transformation)
Give $Q^{(k)}$ its eigen-decomposition $Q^{(k)} = VD^2V^T$, and define $T(x; \sqrt{2}) = \frac{\sqrt{2}}{\delta_k} DV^T \left(x - c^{(k)}\right)$.

Step 2 (Construct the Sample Set)
Construct the transformed sample set $\hat{Y}$ by computing $T\left(y^{(i)}; \sqrt{2}\right)$ for each $y^{(i)} \in Y$. Call Algorithm 1 with $\Delta = 2$ SCB note: why $\Delta = 2$? Should this be $\Delta = \sqrt{2}$? to produce an improved sample set $\hat{Y}'$. Construct the unshifted sample set Y' by computing $T^{-1}\left(\hat{y}^{(i)}; \sqrt{2}\right)$ for each $\hat{y}^{(i)} \in \hat{Y}'$.

Step 3 (Construct Model Functions)
Evaluate f at each $y^{(i)} \in Y'$, and construct m_f with (10)

Our main method follows an algorithmic template described in [9], and is given in Algorithm 3.

SCB note: Modify this algorithm description to use a starting sample set $Y^{(0)}$ instead of a starting point $x^{(0)}$. Also keep track of $Y^{(i)}$ in this outer algorithm so that we can refer to it when we call Algorithm 2. For example, in Step 1, " Construct ${m_f}^{(k)}$ by calling Algorithm 2 on $T^{(k)}_{\text{sample}}$ with initial sample set $Y = Y^{(k)} \cup x^{(k)}$ (or whatever is correct)"

Algorithm 3: Linear Always-feasible Constrained Derivative-free Algorithm

Step 0 (Initialization)
Choose $\tau_\xi \geq 0$, $\tau_\Delta \geq 0$, $\omega_{\text{dec}} \in (0,1)$, $\omega_{\text{inc}} \geq 1$, $\gamma_{\text{suff}} \in (0,1]$, $\gamma_{\text{min}} \in (0, \gamma_{\text{suff}})$, and $\alpha > 0$.
Initialize $k \leftarrow 0$, $x^{(0)} \in \mathcal{F}$, and $0 < \Delta_0 \leq \Delta_{\text{max}}$.

Step 1 (Construct the Model)
Construct the trust region $T^{(k)}_{\text{sample}}$.
Construct ${m_f}^{(k)}$ by calling Algorithm 2 on $T^{(k)}_{\text{sample}}$.

Step 2 (Check Stopping Criteria)
Compute the criticality measure $\chi^{(k)}_m$ as in (38).
If $\chi^{(k)}_m < \tau_\xi$ and $\Delta_k < \tau_\Delta$, **stop**: return $x^{(k)}$ as the solution.
Otherwise, if $\Delta_k > \alpha \chi^{(k)}_m$, set $\Delta_{k+1} \leftarrow \omega_{\text{dec}} \Delta_k$, $x^{(k+1)} \leftarrow x^{(k)}$, $k \leftarrow k+1$; **go to Step 1**.

Step 3 (Solve the Trust Region Subproblem)
Compute $s^{(k)} = \arg\min_{s \in T^{(k)}_{\text{search}} - x^{(k)}} {m_f}^{(k)} \left(x^{(k)} + s\right)$.

Step 4 (Test for Improvement)
Evaluate $f(x^{(k)} + s^{(k)})$ and ρ_k as in (5).
If $\rho_k < \gamma_{\text{min}}$, then $x^{(k+1)} \leftarrow x^{(k)}$, otherwise $x^{(k+1)} \leftarrow x^{(k)} + s^{(k)}$.
If $\rho_k < \gamma_{\text{suff}}$, then $\Delta_{k+1} \leftarrow \min\{\Delta_{\text{max}}, \omega_{\text{dec}} \Delta_k\}$.
If $\rho_k \geq \gamma_{\text{suff}}$, and $\|s^{(k)}\|_\infty = \Delta_k$, then $\Delta_{k+1} \leftarrow \omega_{\text{inc}} \Delta_k$.
If $\rho_k \geq \gamma_{\text{suff}}$, and $\|s^{(k)}\|_\infty < \Delta_k$, then $\Delta_{k+1} \leftarrow \Delta_k$.
Set $k \leftarrow k+1$ and **go to Step 1**.

7 Convergence Analysis

For our convergence analysis, we make the following assumptions about the problem. Let Ω be an open convex set containing $\mathcal{F}$.

Assumption 7.1 *The function f is twice continuously differentiable with Lipschitz continuous Hessian in Ω. That is, there exists constant $L_{\nabla^2} > 0$ such that*

$$\left\|\nabla^2 f(x) - \nabla^2 f(y)\right\| \leq L_{\nabla^2} \|x - y\| \quad \forall x, y \in \Omega. \tag{43}$$

Assumption 7.2 *The function f is bounded below over Ω. That is,*

$$f(x) \geq f_{min} \quad \forall x \in \Omega. \tag{44}$$

Assumption 7.3 *There exists a point $\bar{x}$ within the interior of the feasible region. Namely, using A and b from* (40)*:*

$$A\bar{x} < b. \tag{45}$$

Assumption 7.4 SCB note: I changed $\beta - 1$ to β in this assumption and in H3 from conejo et al. for simplicity . *The Hessians of f are uniformly bounded at each iterate. That is, there exists a constant $\beta \geq 0$ such that*

$$\left\|\nabla^2 f\left(x^{(k)}\right)\right\| \leq \beta \quad \forall k \in \mathbb{N}.$$

In addition to the above assumptions about the problem, we also require that the sequence of sample trust regions $\{T^{(k)}_{\text{sample}}\}$ are *suitable* as defined in Definition 7.1 below. Later, in Section 8.1.1, we describe one method for constructing such a sequence.

To simplify notation, define $G^{(k)} = \begin{pmatrix} A \\ I \\ -I \end{pmatrix}$ and $g^{(k)} = \begin{pmatrix} b \\ x_i^{(k)} + \Delta_k \\ -x_i^{(k)} + \Delta_k \end{pmatrix}$. We then define

$$\begin{aligned} \mathcal{P}^{(k)} = \mathcal{F} \cap T^{(k)}_{\text{out}} &= \{x \in \mathbb{R}^n |\ Ax \leq b, \|x - x^{(k)}\|_\infty \leq \Delta_k\} \\ &= \{x \in \mathbb{R}^n | G^{(k)} x \leq g^{(k)}\}. \end{aligned} \tag{46}$$

Note that we will still have $\|G_i^{(k)}\| = 1$ for each $i \in [m]$.

Definition 7.1 *Let $\left\{Q^{(k)}\right\}$ be a sequence of positive definite symmetric matrices, $\left\{c^{(k)}\right\} \subset \mathbb{R}^n$ be a sequence of points, and $\{\delta_k\}$ be a sequence of positive scalars. For each $k \in \mathbb{N}$, define the ellipsoids*

$$E^{(k)} = \left\{ x \in \mathbb{R}^n \middle| \left(x - c^{(k)}\right)^T Q^{(k)} \left(x - c^{(k)}\right) \leq \frac{1}{2}\delta_k^2 \right\}, \text{and}$$
$$\hat{E}^{(k)} = \left\{ x \in \mathbb{R}^n \middle| \left(x - c^{(k)}\right)^T Q^{(k)} \left(x - c^{(k)}\right) \leq \delta_k^2 \right\}.$$

The sequence $\left\{E^{(k)}\right\}$ is said to be suitable *if all of the following are satisfied:*

1. *$E^{(k)} \subseteq \mathcal{P}^{(k)}$, where $\mathcal{P}^{(k)}$ is defined by* (46)*,*
2. *$x^{(k)} \in \hat{E}^{(k)}$,*
3. *The condition number $\kappa\left(Q^{(k)}\right)$ is bounded independently of k.*

We begin our analysis by examining how the shapes of the ellipsoidal sample trust regions $T^{(k)}_{\text{sample}}$ affect the accuracy of the interpolating model functions. Recall that the standard model improvement algorithm (**??** produces a well-poised sample set over a ball $B_2(0;\Delta)$ resulting in accurate model functions. The following Lemma shows that applying that algorithm to the transformed sample set, as done in Algorithm 2, also produces accurate model functions for the ellipsoidal sample trust regions $T^{(k)}_{\text{sample}}$, provided that we have a uniform bound on the condition numbers of the matrices $Q^{(k)}$ in (41).

SCB note: I think κ is too overloaded. Perhaps we should use a different symbol for the condition number?

Lemma 7.2 *Let $\delta_k > 0$, $c^{(k)} \in \mathbb{R}^n$ and a symmetric matrix $Q^{(k)} \succ 0$ be given. Let the eigen decomposition of $Q^{(k)}$ be $Q^{(k)} = VD^2V^T$, $V^TV = I$, where D is diagonal with positive entries. Let $\hat{E}^{(k)}$ be defined by*

$$\hat{E}^{(k)} = \left\{ x \in \mathbb{R}^n \middle| \left(x - c^{(k)}\right)^T Q^{(k)} \left(x - c^{(k)}\right) \leq \delta_k^2 \right\}$$

For $\delta > 0$, define the transformation $T(x;\delta) = \frac{\delta}{\delta_k} DV^T\left(x - c^{(k)}\right)$, and let $\delta_r = \max_{x \in \hat{E}^{(k)}} \|x - c^{(k)}\|$. Let $\hat{m}_f(u)$ be a model of the shifted objective $\hat{f}(u) = f(T^{-1}(u;\delta_r))$ such that, for constants $\kappa_{ef}, \kappa_{eg}, \kappa_{eh} > 0$, the following error bounds hold for all $u \in B_2(0;\delta_r)$:

$$|\hat{m}_f(u) - \hat{f}(u)| \leq \kappa_{ef}\delta_r^3, \tag{47}$$
$$\|\nabla\hat{m}_f(u) - \nabla\hat{f}(u)\| \leq \kappa_{eg}\delta_r^2, \tag{48}$$
$$\|\nabla^2\hat{m}_f(u) - \nabla^2\hat{f}(u)\| \leq \kappa_{eh}\delta_r. \tag{49}$$

Then, with

$$\kappa'_{ef} = \kappa_{ef},$$
$$\kappa'_{eg} = \kappa_{eg}\sqrt{\kappa(Q^{(k)})},$$
$$\kappa'_{eh} = \kappa_{eh}\kappa(Q^{(k)}),$$

the model function $m_f(x) = \hat{m}_f(T(x;\delta_r))$ satisfies the following error bounds for all $x \in \hat{E}^{(k)}$:

$$|m(x) - f(x)| \leq \kappa'_{ef}\delta_r^3, \tag{50}$$

$$\|\nabla m(x) - \nabla f(x)\| \leq \kappa'_{eg}\delta_r^2, \tag{51}$$

$$\|\nabla^2 m(x) - \nabla^2 f(x)\| \leq \kappa'_{eh}\delta_r. \tag{52}$$

Proof:

Noting that $D_{i,i} > 0$, observe that

$$\delta_r^2 = \frac{\delta_k^2}{\lambda_{\min}\left(Q^{(k)}\right)} = \frac{\delta_k^2}{\min_{i\in[n]} D_{i,i}^2},$$

so $\min_i D_{i,i} = \frac{\delta_k^2}{\delta_r^2}$.

Thus,

$$\kappa\left(Q^{(k)}\right) = \kappa\left(D^2\right) = \frac{\max_i D_{i,i}^2}{\min_i D_{i,i}^2} = \frac{\delta_r^2}{\delta_k^2}\max_i D_{i,i}^2 = \frac{\delta_r^2}{\delta_k^2}\|D\|^2$$
$$\implies \frac{\delta_r}{\delta_k}\|D\| = \sqrt{\kappa\left(Q^{(k)}\right)}.$$

Let $x \in E^{(k)}$ be arbitrary and let $u = T(x) \in B_2(0;\delta_r)$. Then

$$|m_f(x) - f(x)| = |\hat{m}(u) - \hat{f}(u)| \leq \kappa'_{ef}\delta_r^3.$$

Similarly, for the gradient we find:

$$\begin{aligned}\|\nabla m_f(x) - \nabla f(x)\| &= \left\|\frac{\delta_r}{\delta_k}DL^T\left(\nabla\hat{m}_f(u) - \nabla\hat{f}(u)\right)\right\| \\ &\leq \frac{\delta_r}{\delta_k}\|DL^T\|\|\nabla\hat{m}_f(u) - \nabla\hat{f}(u)\| \\ &\leq \sqrt{\kappa(Q^{(k)})}\kappa_{eg}\delta_r^2 = \kappa'_{eg}\delta_r^2.\end{aligned}$$

Finally, for the Hessian, we have

$$\begin{aligned}\|\nabla^2 m_f(x) - \nabla^2 f(x)\| &= \left\|\frac{\delta_r^2}{\delta_k^2}DL^T\left(\nabla\hat{m}_f(u) - \nabla\hat{f}(u)\right)LD^T\right\| \\ &\leq \frac{\delta_r^2}{\delta_k^2}\|D\|^2\|\nabla\hat{m}_f(u) - \nabla\hat{f}(u)\| \\ &\leq \kappa(Q^{(k)})\kappa_{eh}\delta_r = \kappa'_{eh}\delta_r.\end{aligned}$$

□

The remainder of our analysis closely follows that of [9], which presents a class of trust region algorithms for derivative-free optimization with convex constraints. In fact, if the tolerances τ_χ and τ_Δ are set to zero, then Algorithm 3 is a particular implementation of the algorithm studied within this reference. The only

modification is in the trust region subproblem, in which we allow for more trial points with an L_∞ ball instead of an L_2 ball. We have also included a modification that $\Delta_k \leq \Delta_{\max}$; however, this does not alter any of the convergence results. In fact, the authors allow $\omega_{\text{inc}} = 1$, so that the trust region radius can never be increased. It is therefore sufficient to show that the model functions generated by our method satisfy the hypotheses required for their analysis.

The analysis in [9] assumes that the model of the objective is quadratic and satisfies an efficiency condition. Namely, there exists a constant κ_f such that for all $k \in \mathbb{N}$, SCB note: yet another κ! Make sure there is no conflict regarding the use of κ_f (versus, say κ_{ef}.

$$m_f{}^{(k)}\left(x^{(k)}\right) - m_f{}^{(k)}\left(x^{(k)} + s^{(k)}\right) \geq \kappa_f \chi_m^{(k)} \min\left\{\frac{\chi_m^{(k)}}{1 + \left\|\nabla^2 m_f{}^{(k)}(x^{(k)})\right\|}, \Delta_k, 1\right\}, \tag{53}$$

where $\chi_m^{(k)}$ is defined by (38). The Generalized Cauchy Point is shown to satisfy (53) within [10]. This is a reasonable choice for $s^{(k)}$, but the exact solution as chosen within Algorithm 3 necessarily also satisfies (53) as it can only improve upon the Generalized Cauchy Point. Also, note that although [9] presents a criticality measure based on the true projection onto the constraints, for linear constraints, this precisely coincides with the models used in (38).

The authors of [9] also make four explicit assumptions, which are satisfied by our algorithm under our assumptions. Their first assumption H_1 requires the function f to be differentiable and its gradient ∇f to be Lipschitz continuous with constant $L > 0$ in Ω. This is satisfied by our Assumption 7.1 which is a strictly stronger assumption. Their second assumption H_2 is equivalent to our Assumption 7.2. Their third assumption H_3 is that the Hessians of the model functions $m_f{}^{(k)}$ are uniformly bounded. In particular they assume that there exists a constant $\beta_2 \geq 0$ such that

$$\left\|\nabla^2 m_f{}^{(k)}(x^{(k)})\right\| \leq \beta_2 \quad \forall k \geq 0. \tag{54}$$

Our Assumption 7.4 is similar, except that it imposes a bound on the Hessian of the objective function f. We show in Lemma 7.5 below that under this assumption, the model functions $m_f{}^{(k)}$ constructed by our algorithm will satisfy (54).

The final assumption, H_4, from [9] is an accuracy condition. Namely, it assumes that there exists a constant $\delta_g > 0$ such that for all $k \in \mathbb{N}$,

$$\left\|\nabla m_f{}^{(k)}(x^{(k)}) - \nabla f(x^{(k)}\right\| \leq \delta_g \Delta_k. \tag{55}$$

In Theorem 7.4 we show that this accuracy condition is satisfied.

SCB note: The following theorem only states a small portion of what you actually prove. I would like to reword as in the the revised theorem below.

Theorem 7.3 *Suppose that Assumption 7.1, Assumption 7.4, and Assumption 7.3 hold. Suppose further that for each iterate k, the ellipsoids $E^{(k)}$ and $\hat{E}^{(k)}$ provided by the* ConstructTrustRegion *subroutine are suitable according to Definition 7.1, and the sample set is constructed by calling Algorithm 2.*

Then, $m_f\left(x^{(k)}\right) = f\left(x^{(k)}\right)$, and the accuracy condition (55) *is satisfied for each iterate k. Namely, there exists $\kappa_g > 0$ such that*

$$\left\|\nabla m_f\left(x^{(k)}\right) - \nabla f\left(x^{(k)}\right)\right\| \leq \kappa_g \Delta_k \quad \forall k \in \mathbb{N}.$$

Theorem 7.4 *Suppose that Assumption 7.1, Assumption 7.4, and Assumption 7.3 hold. Suppose further that ~~for each iterate k, the ellipsoids $E^{(k)}$ and $\hat{E}^{(k)}$ provided by the ConstructTrustRegion subroutine~~ the sample trust region ellipsoids $E^{(k)} = T^{(k)}_{sample}$, $k = 0, \ldots$, are suitable according to Definition 7.1, and let $\hat{E}^{(k)}$ be defined as in Definition 7.1. For each iterate k, let ${m_f}^{(k)}(x)$ be constructed by Algorithm 2. Then, there exist constants $\kappa_{ef}, \kappa_{eg}, \kappa_{eh}$, independent of k, such that for all $x \in \hat{E}^{(k)}$ the following error bounds hold:*

$$\begin{aligned} |{m_f}^{(k)}(x) - f(x)| &\leq \kappa_{ef}\delta_r^3, \\ \|\nabla {m_f}^{(k)}(x) - \nabla f(x)\| &\leq \kappa_{eg}\delta_r^2, \\ \|\nabla^2 {m_f}^{(k)}(x) - \nabla^2 f(x)\| &\leq \kappa_{eh}\delta_r. \end{aligned}$$

Proof:

By the definition of suitability, we have that $E^{(k)} \subset \mathcal{P}^{(k)}$ and $x^{(k)} \in \hat{E}^{(k)}$ and that there is a bound κ_{max} such that $\kappa(Q^{(k)}) \leq \kappa_{max}$ for all k. As in Lemma 7.2, give $Q^{(k)} = LD^2L^T$ its eigen-decomposition. Define the transformation $T(x;\delta) = \frac{\delta}{\delta_k^2} DL^T\left(x - c^{(k)}\right)$. Notice that the transformation $T\left(x;\sqrt{2}\right)$ maps $T^{(k)}_{\text{sample}} = E^{(k)}$ to the unit ball. After using Algorithm 1 to choose sample points, we know by Theorem 5.13 that there is a bound $\Lambda > 0$ with $\left\|M\left(\bar{\Phi}, \hat{Y}\right)^{-1}\right\| \leq \Lambda$ depending only on p and ξ_{min}. Next, consider the shifted sample set $\bar{Y} = \sqrt{2}\frac{\delta_r}{\delta_2}\hat{Y}$, where $\delta_r = \max_{x \in \hat{E}^{(k)}} \|x - c^{(k)}\|$. By Theorem 5.7, there are fixed constants $\kappa_f, \kappa_g, \kappa_h > 0$ such that

$$\begin{aligned} \left|\hat{f}(u) - \hat{m}_f(u)\right| &\leq \kappa_f \Lambda \delta_r^3, \\ \left\|\nabla \hat{f}(u) - \nabla \hat{m}_f(u)\right\| &\leq \kappa_g \Lambda \delta_r^2, \\ \left\|\nabla^2 \hat{f}(u) - \nabla^2 \hat{m}_f(u)\right\| &\leq \kappa_h \Lambda \delta_r. \end{aligned}$$

for all $u \in B_2(0, \delta_r)$. This means that the shifted functions $\hat{m}_f^{(k)}(u) = {m_f}^{(k)}(T^{-1}(u);\delta)$ and $\hat{f}(u) = f(T^{-1}(u;\delta))$ as described in Lemma 7.2, satisfy (47)—(49), and therefore (50)—(52).

Because $\kappa\left(Q^{(k)}\right)$ is bounded by some $\epsilon_\alpha > 0$ independently of k, and defining $\kappa_g' = \kappa_g\sqrt{\epsilon_\alpha}\Lambda\Delta_{\text{max}}$, we see that we can use Lemma 7.2 to conclude that for all $x_0 \in \hat{E}^{(k)}$:

$$\|\nabla f(x_0) - \nabla m_f(x_0)\| \leq \kappa_g \Lambda \Delta_k^2 \sqrt{\kappa\left(\frac{2}{\delta_k} Q^{(k)}\right)} = \kappa_g\sqrt{\epsilon_\alpha}\Lambda\Delta_{\text{max}}\Delta_k = \kappa_g'\Delta_k.$$

In particular, $x^{(k)} \in \hat{E}^{(k)}$. □

SCB note: The following lemma assumes that all of the sample sets are Λ-poised. So you need to prove that this is the case (based on suitability)

Lemma 7.5 *Suppose that Assumption 7.1 and Assumption 7.4 hold. Let ${m_f}^{(k)}$ be a quadratic model interpolating f on a Λ-poised sample set Y over $B_2\left(x^{(k)}, \Delta_k\right)$ for each $k \in \mathbb{N}$. Then* (54) *is also satisfied.*

Proof:

By Assumption 7.4, we can choose $\beta_1 \geq 1$ to be such that for all $k \in \mathbb{N}$:

$$\left\|\nabla^2 f\left(x^{(k)}\right)\right\| \leq \beta_1 - 1.$$

Because Y is Λ-poised, we can use Theorem 5.5 to bound $\left\|\hat{M}\left(\bar{\Phi}, Y\right)^{-1}\right\|$. Thus, by Theorem 5.7, we have

$$\left\|\nabla^2 f\left(x^{(k)}\right) - \nabla^2 m_f\left(x^{(k)}\right)\right\| \leq \kappa_h \Delta_k \leq \kappa_h \Delta_{\max}$$

Defining $\beta_2 = \kappa_h \Delta_{\max} + \beta_1 \geq 1$, we see that

$$\left\|\nabla^2 m_f\left(x^{(k)}\right)\right\| \leq \left\|\nabla^2 m_f\left(x^{(k)}\right) - \nabla^2 f\left(x^{(k)}\right)\right\| + \left\|\nabla^2 f\left(x^{(k)}\right)\right\| \leq \beta_2 - 1.$$

□

Lemma 7.6 *Suppose that Assumptions 7.1 and Assumption 7.4 hold. Let the sequence of sample trust regions $\{T^{(k)}_{sample}\}$ be suitable as defined in Definition 7.1. Let ${m_f}^{(k)}$ be the quadratic model functions constructed in Algorithm 2. Then* (54) *is satisfied.*

Proof:

Let $T^{(k)}_{\text{sample}}$ be defined by $Q^{(k)}, c^{(k)}, \delta_k$ as in (41). Let $E^{(k)} = T^{(k)}_{\text{sample}}$ and $\hat{E}^{(k)}$ be as defined in Definition 7.1. By the definition of suitability, $x^{(k)} \in \hat{E}^{(k)}$, so by Theorem 7.4, we have

$$\|\nabla^2 {m_f}^{(k)}(x) - \nabla^2 f(x)\| \leq \kappa'_{eh}\delta_r \leq \kappa_{eh}\Delta_k \leq \kappa_{eh}\Delta_{\max}.$$

By Assumption 7.4, we have

$$\left\|\nabla^2 m_f\left(x^{(k)}\right)\right\| \leq \left\|\nabla^2 m_f\left(x^{(k)}\right) - \nabla^2 f\left(x^{(k)}\right)\right\| + \left\|\nabla^2 f\left(x^{(k)}\right)\right\| \leq \kappa_{eh}\Delta_{\max} + \beta.$$

It follows that (54) is satisfied for all k with $\beta_2 = \beta + \kappa_{eh}\Delta_{\max}$. □

We can now state our main convergence result:

SCB note: Chapter 3 of the thesis never gave a "main convergence result" similar to Corollary IV.44 in the theis. This was an oversight that we need to correct in the journal article.

Theorem 7.7 *Suppose that Assumptions* **??**–**??** *are satisfied. Blah blah blah.*

8 Ellipsoidal Trust Region Approach

The main idea of the ellipsoidal method is to define the sample trust region to be a feasible ellipsoid as in (41):

$$T^{(k)}_{\text{sample}} = E^{(k)} = \left\{ x \in \mathbb{R}^n \middle| (x - c^{(k)})^T Q^{(k)} (x - c^{(k)}) \leq \frac{1}{2}\delta_k^2 \right\}$$

where $Q^{(k)}$ is a positive definite matrix, $c^{(k)}$ is the center of the ellipsoid, and δ_k is a constant determining the ellipsoid's radius. $Q^{(k)}$ and $c^{(k)}$ are chosen so that the ellipsoid conforms roughly to the shape of the feasible region near the current iterate, while ensuring that it lies entirely within the intersection of the feasible region and the outer trust region.

8.1 A safe ellipsoid

8.1.1 Construction

This section shows how to construct one possible ellipsoid that satisfies Definition 7.1. We define the active constraints at a point $x \in \mathcal{F}$ by

$$\mathcal{A}(x) = \{i \in [m] | A_i x = b_i\}. \tag{56}$$

Recall that A and b are defined by (40).

For any iterate k, if there are no active constraints, $\mathcal{A}\left(x^{(k)}\right) = \emptyset$, then we are free to choose

$$Q^{(k)} = I, \quad c^{(k)} = x^{(k)}, \quad \text{and} \quad \delta_k = \min\left\{\Delta_k, \min_{i\in[m]} b_i - A_i x\right\}. \tag{57}$$

We show within Lemma 8.1 that this is a suitable ellipsoid. Otherwise, we begin by constructing a direction $\hat{u}^{(k)}$ that is maximally feasible with respect to the active constraints. To make this precise, let $\mathcal{S} \subseteq \{1, \ldots, m\}$ be arbitrary. We define the set of "most" feasible direction from x, and quantify how feasible they are with the definitions:

$$u_{\text{feasible}}(\mathcal{S}) = \begin{cases} \arg\max_{\|u\|=1} \min_{i\in\mathcal{S}} -u^T A_i & \text{if } \mathcal{S} \neq \emptyset \\ \emptyset & \text{if } \mathcal{S} = \emptyset \end{cases} \tag{58}$$

$$\pi_1(\mathcal{S}) = \begin{cases} \max_{\|u\|=1} \min_{i\in\mathcal{S}} -u^T A_i & \text{if } \mathcal{S} \neq \emptyset \\ 1 & \text{if } \mathcal{S} = \emptyset \end{cases} \tag{59}$$

$$\pi_2(x) = \pi_1\left(\mathcal{A}(x)\right). \tag{60}$$

Here, u_{feasible} is a set of directions that head away from each of the active constraints at a point. For each iterate, this produces the quantities

$$\pi^{(k)} = \pi_2\left(x^{(k)}\right), \tag{61}$$

$$\hat{u}^{(k)} \in u_{\text{feasible}}\left(\mathcal{A}\left(x^{(k)}\right)\right). \tag{62}$$

We then compute

$$\pi_3^{(k)} = \left(1 + \frac{1}{\sqrt{2}}\right)\sqrt{1 + \left(\pi^{(k)}\right)^2}, \tag{63}$$

$$\delta_f = \frac{1}{\pi_3^{(k)}} \min\left\{\Delta_k, \min_{i\in[m]\setminus\mathcal{A}\left(x^{(k)}\right)} \left[b_i - \left(A_i\right)^T x^{(k)}\right]\right\}. \tag{64}$$

Notice that the minimization in the above expression computes the minimum distance to a non-active constraint. It will be convenient to define the rotation matrix and affine mapping

$$R^{(k+1)} = 2\frac{(e_1 + \hat{u}^{(k)})(e_1 + \hat{u}^{(k)})^T}{(e_1 + \hat{u}^{(k)})^T(e_1 + \hat{u}^{(k)})} - \boldsymbol{I}, \tag{65}$$

$$T_k(x) = R^{(k+1)}\left(x - x^{(k)}\right). \tag{66}$$

Recall that $e_1 = (1, 0, \ldots, 0)^T$, and observe that $R^{(k+1)}e_1 = \hat{u}^{(k)}$, $R^{(k+1)}\hat{u}^{(k)} = e_1$, $\det(R^{(k+1)}) = 1$, and $R^{(k+1)}R^{(k+1)^T} = R^{(k+1)^T}R^{(k+1)} = I$.

We can then define the ellipsoid as

$$Q^{(k)} = R^{(k+1)^T} \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \left(\pi^{(k)}\right)^{-2} \boldsymbol{I} \end{pmatrix} R^{(k+1)}, \quad c^{(k)} = x^{(k)} + \delta_f \hat{u}^{(k)}, \quad \text{and} \quad \delta_k = \delta_f \tag{67}$$

whenever $\mathcal{A}\left(x^{(k)}\right) \neq \emptyset$.

8.1.2 Suitability

We show that this construction provides a suitable ellipsoid according to Definition 7.1.

Lemma 8.1 *Let $\mathcal{A}$ be defined by* (56)*.*
If $\mathcal{A}\left(x^{(k)}\right) = \emptyset$ during iteration k, then (57) *defines a suitable ellipsoid for iteration k according to Definition 7.1.*

Proof:

If $\mathcal{A}\left(x^{(k)}\right) = \emptyset$, then we are free to use (57). This simplifies $\hat{E}^{(k)}$ to

$$\left(x - c^{(k)}\right)^2 Q^{(k)} \left(x - c^{(k)}\right)^2 \leq \delta_k^2 \Longrightarrow \left\|x - x^{(k)}\right\|^2 \leq \Delta_k^2.$$

Because $E^{(k)}$ is then a sphere within the outer trust region with radius less than the distance to the nearest constraint, $E^{(k)} \subseteq \mathcal{P}^{(k)}$. Because the sphere $\hat{E}^{(k)}$ is centered at $x^{(k)}$, $x^{(k)} \in \hat{E}^{(k)}$. Also, $\kappa\left(Q^{(k)}\right) = 1$.
□

Lemma 8.2 *Let π_2 be defined by* (60)*.*
Suppose that Assumption 7.3 holds.
Then, $1 \geq \pi_2(y) > 0$ for any $y \in \mathcal{F}$.

Proof:

Let $y \in \mathcal{F}$, and let $\mathcal{A}(x)$ be defined by (56). If $\mathcal{A}(y) = \emptyset$, then $\pi_2(y) = 1 > 0$. Otherwise, let $i \in \mathcal{A}(y)$, so that by (40), $c_i(y)(Ay - b)_i = 0$. We know that if $\bar{x}$ is defined as in Assumption 7.3, then

$$c_i(\bar{x}) = c_i(y) + A_i(\bar{x} - y) \Longrightarrow A_i(\bar{x} - y) \leq c_i(\bar{x}) - c_i(y) = c_i(\bar{x}) < 0.$$

Using (46), we can write this as

$$-A_i \frac{\bar{x} - y}{\|\bar{x} - y\|} > 0 \Longrightarrow \min_{i \in \mathcal{A}(y)} -A_i \frac{\bar{x} - y}{\|\bar{x} - y\|} > 0.$$

Using this along with the definitions of u_{feasible}, π_1, and π_2 in (58), (59), and (60); we see

$$\pi_2(y) = \pi_1\left(\mathcal{A}(y)\right) = \max_{\|u\|=1} \min_{i \in \mathcal{A}(y)} -A_i^T u \geq \min_{i \in \mathcal{A}(y)} -A_i^T \frac{\bar{x} - y}{\|\bar{x} - y\|} > 0.$$

We know that $\pi_2(y) \leq 1$ because it is the dot product of two vectors of length one: if $\|u\| = 1$, then $\left|u^T A_i\right|^2 \leq \|u\|\|A_i\| = 1$ by Cauchy–Schwartz.
□

Lemma 8.3 *Let $\pi^{(k)}$ be defined by* (61).
Suppose that Assumption 7.3 holds.
There exists an $\epsilon_\alpha > 0$ such that $\pi^{(k)} \geq \epsilon_\alpha \; \forall \; k \in \mathbb{N}$.

Proof:

Let $\mathcal{A}$ be defined by (56). Because there are m constraints, each $\mathcal{A}(x)$ is one of the 2^m subsets of $\{1, 2, 3, \ldots, m\}$. This means that u_{feasible}, π_1, and $\pi^{(k)}$ as defined by (58), (59), and (61) can only take on at most $1 + 2^m$ values. By Lemma 8.2, we know that each of these values must be positive. Thus, we are free to choose ϵ_α to be the smallest of these values. □

It will be useful to define some intermediate cones. Namely, for any scalar π, direction u, and point c, we define the cone

$$\mathcal{C}(\pi, u, c) = \left\{ c + tu + s \in \mathbb{R}^n \middle| \; s^T u = 0, t \geq 0, \|s\| \leq \pi t \right\}. \tag{68}$$

We can use this to define shifted and unshifted cones about the point $x^{(k)}$ along a direction $\hat{u}^{(k)}$:

$$C_{\text{sh}}^{(k)} = \mathcal{C}\left(\pi^{(k)}, e_1, 0\right), \tag{69}$$

$$C_{\text{unsh}}^{(k)} = \mathcal{C}\left(\pi^{(k)}, \hat{u}^{(k)}, x^{(k)}\right). \tag{70}$$

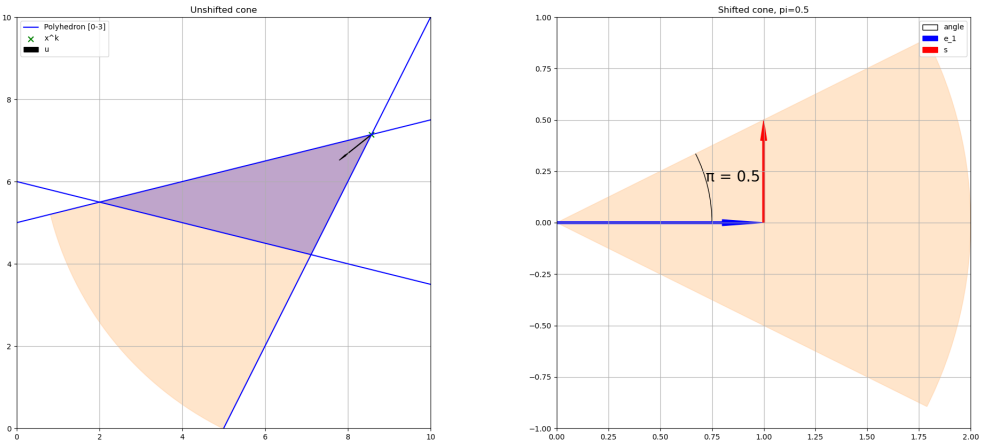

Figure 1: On the left is an unshifted cone $C_{\text{unsh}}^{(k)}$. The zeros of the polyhedron constraint functions are in blue, and the current iterate is in green. A feasible direction u from the current iterate is calculated, and the width of the cone is determined to lie within the active constraints of the polyhedra. On the right is a shifted cone $C_{\text{unsh}}^{(k)}$. It is centered at the origin, and points along e_1, opening at a rate of $\pi^{(k)}$.

Observe that, by construction, $C_{\text{unsh}}^{(k)}$ is feasible with respect to the active constraints. That is, $A_i x \leq b_i$ for all $x \in C_{\text{unsh}}^{(k)}$ and $i \in \mathcal{A}\left(x^{(k)}\right)$. Figure 1 contains a depiction of these cones.

Lemma 8.4 *Let $C_{unsh}^{(k)}$ and $\mathcal{P}^{(k)}$ be defined by* (70) *and* (46).

The set $C_{unsh}^{(k)}$ is feasible with respect to the active constraints of $\mathcal{P}^{(k)}$ at $x^{(k)}$.

Proof:

Note that the trust region boundary cannot be active at $x^{(k)}$ as $\Delta_k > 0$. Let $\mathcal{A}$, $\pi^{(k)}$, and $\hat{u}^{(k)}$ be defined by (56), (61), (62) and A and b be defined by (40). Let $y = x^{(k)} + t\hat{u}^{(k)} + s \in C_{\text{unsh}}^{(k)}$ and $i \in \mathcal{A}\left(x^{(k)}\right)$ be arbitrary. Then,

$$A_i^T y - b_i = A_i^T(t\hat{u}^{(k)} + s) = A_i^T s + tA_i^T\hat{u}^{(k)} \leq \|s\| - \pi^{(k)}t \leq 0.$$

□

The following function is useful for showing results about our ellipsoids:

$$f_e(\pi, \delta, r; x) = (x - \delta e_1)^T \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \pi^{-2}\boldsymbol{I} \end{pmatrix} (x - \delta e_1) - r. \tag{71}$$

Lemma 8.5 *Let $C_{sh}^{(k)}, f_e, \pi^{(k)}$ be defined as in* (69), (71), (61). *Then, for all $\delta > 0$, the ellipsoid*

$$\left\{ x \in \mathbb{R}^n \middle| f_e\left(\pi^{(k)}, \delta, \frac{1}{2}\delta^2; x\right) \leq 0 \right\} \subseteq C_{sh}^{(k)}. \tag{72}$$

Proof:

Suppose that $x \in \left\{ x \in \mathbb{R}^n \middle| f_e\left(\pi^{(k)}, \delta, \frac{1}{2}\delta^2; x\right) \leq 0 \right\}$, and let $t = x^T e_1$, $s = x - te_1$. Then

$$\begin{aligned} f_e(x) \leq 0 &\Longrightarrow (x - \delta e_1)^T \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \left(\pi^{(k)}\right)^{-2}\boldsymbol{I} \end{pmatrix} (x - \delta e_1) \leq \frac{1}{2}\delta^2 \\ &\Longrightarrow (t - \delta)^2 + \frac{1}{\left(\pi^{(k)}\right)^2}\|s\|^2 \leq \frac{1}{2}\delta^2 \\ &\Longrightarrow \|s\|^2 \leq \left(\pi^{(k)}\right)^2 \left[\frac{1}{2}\delta^2 - (t - \delta)^2\right] \\ &= \left(\pi^{(k)}\right)^2 \left[t^2 - 2(t - \frac{1}{2}\delta)^2\right] \leq \left(\pi^{(k)}\right)^2 t^2 \\ &\Longrightarrow \|s\| \leq \pi^{(k)}t. \end{aligned}$$

Thus, $x \in C_{\text{sh}}^{(k)}$. □

Lemma 8.6 *Let $R^{(k+1)}$, T_k, $C_{unsh}^{(k)}$, $C_{sh}^{(k)}$ be defined as in* (65), (66), (70), (69). *Then $T_k(C_{unsh}^{(k)}) = C_{sh}^{(k)}$.*

Proof:

Observe that $R^{(k+1)}e_1 = \hat{u}^{(k)}$, $R^{(k+1)}\hat{u}^{(k)} = e_1$, $\det(R^{(k+1)}) = 1$, and $R^{(k+1)}{R^{(k+1)}}^T = {R^{(k+1)}}^T R^{(k+1)} = I$.

Suppose that $x \in C_{\text{unsh}}^{(k)}$. Then there exists $t \geq 0$ and $s \in \mathbb{R}^n$ such that $x = x^{(k)} + t\hat{u}^{(k)} + s$ where $s^T \hat{u}^{(k)} = 0$ and $\|s\| \leq \pi^{(k)} t$. Then $T_k(x) = tR^{(k+1)}\hat{u}^{(k)} + R^{(k+1)}s = te_1 + R^{(k+1)}s$. Observe that $(Rs)_1 = (R^{(k+1)}s)^T e_1 = s^T {R^{(k+1)}}^T (R^{(k+1)}\hat{u}^{(k)}) = s^T \hat{u}^{(k)} = 0$. Hence, $T_k(x) = \begin{pmatrix} t \\ \sigma \end{pmatrix}$ where $\sigma \in \mathbb{R}^{n-1}$ satisfies $\|\sigma\| = \|s\| \leq \pi^{(k)} t$. Thus, $T_k(x) \in C_{\text{sh}}^{(k)}$. Conversely, if $\begin{pmatrix} t \\ \sigma \end{pmatrix} \in C_{\text{sh}}^{(k)}$, then let $s = {R^{(k+1)}}^T \begin{pmatrix} 0 \\ \sigma \end{pmatrix}$ to see that $x = T_k^{-1}\left(\begin{pmatrix} t \\ \sigma \end{pmatrix}\right) = {R^{(k+1)}}^T \left(te_1 + \begin{pmatrix} 0 \\ \sigma \end{pmatrix}\right) = t\hat{u}^{(k)} + s$, where $\|s\| = \|\sigma\| \leq \pi^{(k)} t$. Hence $T_k^{-1}\left(\begin{pmatrix} t \\ \sigma \end{pmatrix}\right) \in C_{\text{unsh}}^{(k)}$. □

Lemma 8.7 *Let $\mathcal{A}$, f_e, δ_f $\pi^{(k)}$, $R^{(k+1)}$, T_k, and $\mathcal{P}^{(k)}$ be defined by* (56), (71), (64), (61), (65), (69), *and* (46), *respectively.*

For each iteration k, if $\mathcal{A}\left(x^{(k)}\right) \neq \emptyset$, the ellipsoid

$$E^{(k)} = \left\{ x \in \mathbb{R}^n \middle| f_e\left(\pi^{(k)}, \delta_f, \frac{1}{2}\delta_f^2; T_k(x)\right) \leq 0 \right\} \tag{73}$$

satisfies $E^{(k)} \subseteq \mathcal{P}^{(k)}$.

Proof:

Let $\pi_3^{(k)}$, $R^{(k+1)}$, $C_{\text{unsh}}^{(k)}$, and $C_{\text{sh}}^{(k)}$ be defined by (63), (65), (70), and (69), respectively. We see that if $x \in E^{(k)}$, then by Lemma 8.5 we have that $T_k(x) = \begin{pmatrix} t \\ \sigma \end{pmatrix} \in C_{\text{sh}}^{(k)}$ for some $\sigma \in \mathbb{R}^{n-1}$ with $\|\sigma\| \leq \pi^{(k)}$, and

$$\begin{aligned} (x - \delta_f e_1)^T \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \left(\pi^{(k)}\right)^{-2} \boldsymbol{I} \end{pmatrix} (x - \delta_f e_1) &\leq \frac{1}{2}\delta_f^2 \\ \implies (t - \delta_f)^2 \leq (t - \delta_f)^2 + \frac{1}{\left(\pi^{(k)}\right)^2}\|\sigma\|^2 &\leq \frac{1}{2}\delta_f^2 \\ \implies t \leq \left(1 + \frac{1}{\sqrt{2}}\right)\delta_f & \end{aligned}$$

so that

$$\begin{aligned} \|x\|^2 = t^2 + \|\sigma\|^2 \leq \left(1 + \left(\pi^{(k)}\right)^2\right) t^2 \leq \left(1 + \left(\pi^{(k)}\right)^2\right)\left(1 + \frac{1}{\sqrt{2}}\right)^2 \delta_f^2 = \left(\pi_3^{(k)}\right)^2 \delta_f^2 \\ \implies \|x\| \leq \pi_3^{(k)} \delta_f \leq \min_{i \in [m] \setminus \mathcal{A}\left(x^{(k)}\right)} \left[b_i - (A_i)^T x^{(k)}\right]. \end{aligned}$$

Thus, all points within $E^{(k)}$ are closer than the nearest point of a non-active constraint. Combine this with Lemma 8.4 to see that $E^{(k)} \subseteq \mathcal{P}^{(k)}$. □

Lemma 8.8 *Let δ_f, $R^{(k+1)}$, T_k, $C_{unsh}^{(k)}$, $C_{sh}^{(k)}$, $\mathcal{P}^{(k)}$ be defined as in* (64), (65), (66), (70), (69), (46).

For any iteration k, the ellipsoid

$$\hat{E}^{(k)} = \left\{ x \in \mathbb{R}^n | f_e \left(\pi^{(k)}, \delta_k, \delta_k^2; T_k(x) \right) \leq 0 \right\} \tag{74}$$

satisfies $x^{(k)} \in \hat{E}^{(k)}$.

Proof:

We have that

$$f_e \left(\pi^{(k)}, \delta_f, \delta_f^2; T_k \left(x^{(k)} \right) \right) = f_e \left(\pi^{(k)}, \delta_f, \delta_f^2; 0 \right) =$$
$$(0 - \delta_f e_1)^T \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \left(\pi^{(k)}\right)^{-2} \boldsymbol{I} \end{pmatrix} (0 - \delta_f e_1) = \delta_f^2 \leq \delta_f^2.$$

□

Lemma 8.9 *Let $\mathcal{A}$ and $\pi^{(k)}$ be defined by* (56) *and* (61)*, respectively.*
Suppose that Assumption 7.3 holds.
For some iteration k, if $\mathcal{A} \neq \emptyset$, then the ellipsoid defined by (67) *is suitable according to Definition 7.1.*

Proof:

Let f_e, T_k, $R^{(k+1)}$, δ_f, and $\hat{\mathcal{E}}^k_{\text{feasible}}$ be defined as in (71), (66), (65), (73), and (74), respectively. Observe that Definition 7.1 with (67) defines $E^{(k)}$ and $\hat{E}^{(k)}$ to be

$$\begin{aligned} E^{(k)} &= \left\{ x \in \mathbb{R}^n \middle| f_e \left(\pi^{(k)}, \delta_k, \tfrac{1}{2}\delta_k^2; T_k(x) \right) \leq 0 \right\}, \qquad \text{and} \\ \hat{E}^{(k)} &= \left\{ x \in \mathbb{R}^n \middle| f_e \left(\pi^{(k)}, \delta_k, \delta_k^2; T_k(x) \right) \leq 0 \right\}. \end{aligned}$$

By Lemma 8.7, we know that $E^{(k)} \subseteq \mathcal{P}^{(k)}$. By Lemma 8.8, we know that $x^{(k)} \in \hat{E}^{(k)}$. By Lemma 8.3, there exists $\epsilon_\alpha > 0$, such that the condition number $\kappa \left(Q^{(k)} \right) = \frac{\max\{1, {\pi^{(k)}}^{-2}\}}{\min\{1, {\pi^{(k)}}^{-2}\}} = {\pi^{(k)}}^{-2} > 0$. This is because $\det \left(R^{(k+1)} \right) = 1$ means the condition number of $Q^{(k)}$ is not affected $R^{(k+1)}$. □

8.2 Ellipsoid searches

Within Section 8.1.1, we showed how to construct one possible ellipsoid that satisfies Definition 7.1. In practice, this ellipsoid could be less than desirable. Within this section, we discuss the requirements of other variants of the *ConstructTrustRegion* subroutine that we explored for practicality. The key requirement is these ellipsoids must still satisfy Definition 7.1 to share the same convergence results.

8.2.1 Ensuring suitability

To retain the convergence results, these algorithms maintain $E^{(k)} \subseteq \mathcal{P}^{(k)}$, $x^{(k)} \in \hat{E}^{(k)}$, and a bound on $\kappa \left(Q^{(k)} \right)$.

To ensure the condition number is bounded, the algorithm can compute the condition number of the safe ellipsoid, and only consider ellipsoids better conditioned. Alternatively, it could introduce its own bound.

We found it simplest to parameterize ellipsoids by their Cholesky factorization $Q^{(k)} = LL^T$. Namely, we parameterized the search space in terms of a lower triangular matrix L, and required the diagonal entries to be positive. To ensure a bounded condition number, we chose a $\epsilon_\alpha > 0$, and constrain $\max_{i\in[n]} \leq \epsilon_\alpha \min_{i\in[n]}$.

One potential difficulty created by moving the ellipsoid center $c^{(k)}$ is that the current iterate $x^{(k)}$ may not lie within near the resulting ellipsoid. The pitfall is that the model function may lose accuracy near the current iterate. Thus, we have implemented a few ways of ensuring the current iterate is within the search trust region. This can be done by either of the following two options:

- Adding a constraint to the ellipsoid problem to include the original point.
- Expanding the size of the ellipsoid.

Adding a constraint. In order to include the original point as a constraint, we add a constraint of the following form to the definition of the ellipsoid.

$$\frac{1}{2}\left(x^{(k)} - c^{(k)}\right)^T Q^{(k)} \left(x^{(k)} - c^{(k)}\right) \leq \frac{1}{2}\delta_k^2.$$

Constraints of this nature make finding the ellipsoid much more expensive: the optimization problem we construct uses ${Q^{(k)}}^{-1}$ as decision variables, so that constraints in terms of $Q^{(k)}$ must model matrix inversion.

Increasing the radius. An alternative is to scale $Q^{(k)}$ by a constant. We use a scaling factor $\delta_k > 0$ defined by

$$\delta_k = \max\left\{1, \sqrt{\left(x^{(k)} - c^{(k)}\right)^T Q^{(k)} \left(x^{(k)} - c^{(k)}\right)}\right\}$$

and let the ellipsoid be:

$$E^{(k)} = \left\{x \in \mathbb{R}^n \middle| \frac{1}{2}\left(x - c^{(k)}\right)^T Q^{(k)} \left(x - c^{(k)}\right) \leq \frac{1}{2}\delta_k^2\right\}.$$

However, this means that in general $E^{(k)} \not\subset \mathcal{F}$. so that the sample points must be contained to also lie within the feasible region: $E^{(k)} \cap \mathcal{F}_m^{(k)}$. For details on how to choose sample points with additional constraints, see Algorithm 1.

8.2.2 Maximal volume ellipsoids

The error bounds given in Lemma 7.2 suggest that we can obtain more accurate model functions by minimizing the condition number of the matrix $Q^{(k)}$. However, we also desire a large ellipsoid so that our model will satisfy the error bounds over more of $T_{\text{search}}^{(k)}$. Thus, one choice for $T_{\text{sample}}^{(k)}$ is to choose the maximum volume ellipsoid that is both feasible and lies within the outer trust region. We can accomplish this for various ellipsoid centers, by finding the maximal volume ellipsoid that is constrained to lie within the polytope

$$\mathcal{P}^{(k)} := \left\{x \in \mathbb{R}^n | Ax \leq b, x_i^{(k)} - \Delta_k \leq x_i \leq x_i^{(k)} + \Delta_k\right\}.$$

This is not the only reasonable approach: maximizing ensures a larger region for which the models will be accurate, but we are most interested in descent directions. Namely, we may wish to use previously

evaluated points to provide a hint of where the next ellipsoid should be. Another consideration is where points have been evaluated: it may be more economical to choose a smaller ellipsoid that can reuse existing points. However, in this section, we consider the problem of choosing $Q^{(k)}$ and δ_k to maximize the volume of $E^{(k)} \subseteq \mathcal{P}^{(k)}$ given a fixed center $c^{(k)}$. Later, in Section 9.1.2, we will explore strategies for moving the center of the ellipsoid to improve the performance of our trust region algorithm.

We adopt a method similar to that described in [13], which presents an algorithm for finding the maximum volume ellipsoid inscribed in a given polytope.

Let $\bar{g} = g^{(k)} - G^{(k)}c^{(k)}$ and $d = x - c^{(k)}$ so that the polytope becomes

$$\mathcal{P}^{(k)} = \left\{ c^{(k)} + d \in \mathbb{R}^n \mid G^{(k)}d \leq \bar{g} \right\}.$$

Using this transformation, the ellipsoid can then be centered at zero, and defined by a symmetric positive definite matrix $Q \succ 0$:

$$E = \left\{ d \in \mathbb{R}^n \;\middle|\; \frac{1}{2}d^T Q d \leq 1 \right\}.$$

Our goal is to determine the matrix Q that maximizes the volume of E such that $\mu + E \subset \mathcal{P}$. This is accomplished by defining $\bar{b} = b - Ac$ and solving the following problem for Q for a given center:

$$\begin{array}{cc} \sup_{Q \succeq 0} & \det(Q^{-1}) \\ \text{s.t.} & A_i^T Q^{-1} A_i \leq \frac{1}{2}\bar{b_i}^2. \end{array} \tag{75}$$

Theorem 8.10 *Let $\mathcal{P} = \{x \in \mathbb{R}^n | Ax \leq b\}$, where A is an $m \times n$ matrix, and $b \in \mathbb{R}^m$. Let $c \in \operatorname{int} \mathcal{P}$.*

Suppose that some positive definite, symmetric matrix Q solves (75)*, where $\bar{b} = b - Ac$. Then the ellipsoid $E = \{x \in \mathbb{R}^n | (x-c)^T Q (x-c) \leq 1\}$ has the maximum volume over all ellipsoids centered at c and contained in $\mathcal{P}$.*

Proof:

Define the auxiliary function $f(d) = \frac{1}{2}d^T Q d$ so that $E = \{d \in \mathbb{R}^n \mid f(d) \leq 1\}$. Because Q is positive definite, f has a unique minimum on each hyperplane $\{d \in \mathbb{R}^n | \ A_i d = b_i\}$. Let this minimum be $d^{(i)} = \arg\min_{A_i d = \bar{b}_i} f(d)$ for $i \in [m]$. By the first-order optimality conditions, there exists a $\lambda \in \mathbb{R}^m$ such that for each $i \in [m]$,

$$\nabla f(d^{(i)}) = Q d^{(i)} = \lambda_i A_i.$$

Since Q is invertible, we have $d^{(i)} = \lambda_i Q^{-1} A_i$. We also know that $A_i^T d^{(i)} = \bar{b_i}$, so

$$A_i^T \lambda_i Q^{-1} A_i = \bar{b_i} \Longrightarrow \lambda_i = \frac{\bar{b_i}}{A_i^T Q^{-1} A_i},$$

so that

$$d^{(i)} = \lambda_i Q^{-1} A_i = \frac{\bar{b_i}}{A_i^T Q^{-1} A_i} Q^{-1} A_i \quad \forall 1 \leq i \leq m.$$

Because $E \subset \mathcal{P}$, we also know that $f(d^{(i)}) \geq 1$ for each i. Thus,

$$\begin{aligned}
&\frac{1}{2}(d^{(i)})^T Q d^{(i)} \geq 1 \\
&\qquad \Longrightarrow \frac{1}{2}\left(\frac{\bar{b_i}}{A_i^T Q^{-1} A_i} Q^{-1} A_i\right)^T Q \frac{\bar{b_i}}{A_i^T Q^{-1} A_i} Q^{-1} A_i \geq 1 \\
&\qquad \Longrightarrow \frac{1}{2}\frac{1}{A_i^T Q^{-1} A_i} \bar{b_i} A_i^T Q^{-1} Q \frac{\bar{b_i}}{A_i^T Q^{-1} A_i} Q^{-1} A_i \geq 1 \\
&\qquad \Longrightarrow \frac{1}{2}\frac{1}{A_i^T Q^{-1} A_i} \frac{\bar{b_i}^2}{A_i^T Q^{-1} A_i} A_i^T Q^{-1} A_i \geq 1 \\
&\qquad \Longrightarrow \frac{1}{2}\frac{\bar{b_i}^2}{A_i^T Q^{-1} A_i} \geq 1 \\
&\qquad \Longrightarrow \frac{1}{2}\bar{b_i}^2 \geq A_i^T Q^{-1} A_i \\
&\qquad \Longrightarrow A_i^T Q^{-1} A_i \leq \frac{1}{2}\bar{b_i}^2.
\end{aligned}$$

Because the volume of the ellipsoid is proportional to the determinant of Q^{-1}, the maximal ellipsoid is defined by (75). □

Notice that the constraints for (75) can be simplified for the outer trust region's constraints. These take the form:

$$e_i^T \left(\frac{Q^{(k)}}{\frac{1}{2}\delta_k^2}\right)^{-1} e_i \leq \left[e_i^T \left(x^{(k)} - c^{(k)}\right) \pm \Delta_k\right]^2 \quad \forall i \in [n]$$

or

$$\left(\frac{Q^{(k)}}{\frac{1}{2}\delta_k^2}\right)^{-1}_{i,i} \leq \left[x_i^{(k)} - c_i^{(k)} \pm \Delta_k\right]^2 \quad \forall i \in [n]. \tag{76}$$

9 Results

9.1 Algorithm variants

Here, we present numerical results for our implementations of Algorithm 3. We begin by describing the different ways we implemented *ConstructTrustRegion*.

9.1.1 Circular trust region

The simplest approach to maintaining a feasible trust region is to set the inner trust region radius sufficiently small. Within the *ConstructTrustRegion* subroutine, this method sets the trust region radius to the distance to the closest constraint: $T_{\text{out}}^{(k)} = B_2\left(x^{(k)}, \min\left\{\Delta_k, \min_i \frac{|A_i x^{(k)} - b_i|}{\|A_i\|}\right\}\right)$. In practice, this does not work well as the radius can become too small to allow adequate progress.

Two general strategies were considered for addressing this issue as illustrated in Figure 2. One option is to shift the center of the inner trust region as long as it remains within the outer trust region. The second option is to elongate the trust region along the nearest constraint as discussed in the next section. Of course, both of these can be done at the same time.

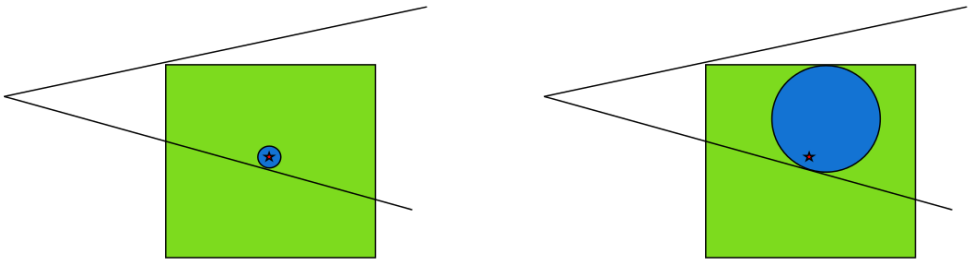

Figure 2: When the current iterate is too close to a constraint, the circular trust region becomes too small. Shifting the trust region center can help remedy this. The star is the current iterate, the outer trust region is in green, and the inner trust region is in blue.

In order to address this issue, we considered using ellipsoidal trust regions. Whereas the circle does not allow improvement when the current iterate lies along a constraint, an ellipsoid elongates along this constraint. In figure Figure 3, we have this type of iterate, but by using an ellipsoid we are still able to search towards the vertex of the feasible region.

More specifically, at iteration k, we choose a scaling factor δ_k and solve for an ellipsoid center $c^{(k)} \in \mathbb{R}^n$ and positive definite matrix $Q^{(k)}$ to define an ellipsoid

$$E^{(k)} = \left\{ x \in \mathbb{R}^n \,\middle|\, \frac{1}{2}\left(x - c^{(k)}\right)^T Q^{(k)} \left(x - c^{(k)}\right) \leq \frac{1}{2}\delta_k^2 \right\}.$$

The simplest approach is to simply let the center of the ellipsoid be the current iterate: $c^{(k)} = x^{(k)}$.

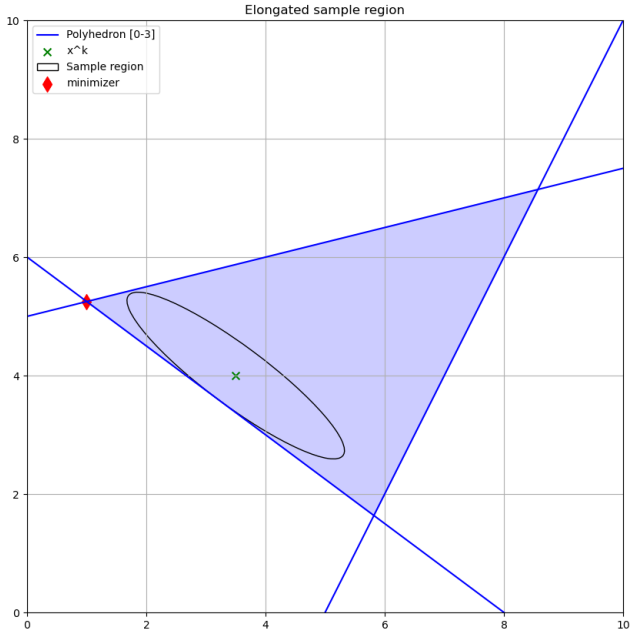

Figure 3: Although the center of the ellipsoid in green is close to the boundary of the blue constraints, it can elongate to allow progress.

9.1.2 Choosing the ellipsoid center

The most obvious choice for the center of the ellipsoid is to choose $c^{(k)} = x^{(k)}$ (i.e., the current iterate). However, if $x^{(k)}$ is too close to a boundary of the feasible region, this can result in a badly shaped or very small ellipsoid. We, therefore, explore strategies where the *ConstructTrustRegion* subroutine moves the center of the ellipsoid away from the boundary. This is depicted in Figure 3.

Outer trust region search. One approach is to search all possible centers within $\mathcal{F} \cap T_{\text{out}}^{(k)}$.

This has the advantage that it allows for the largest volume. However, one problem with this search is that it can force the trust region away from the descent direction. Notice that in Figure 4, although the ellipsoid found has larger volume than before being shifted, this ellipsoid contains points farther from the corner that happens to contain the trust region's minimizer. Within the numerical results, these algorithms are described as "ellipse everywhere".

The following section addresses this problem by proposing a path search method for choosing the ellipsoid center.

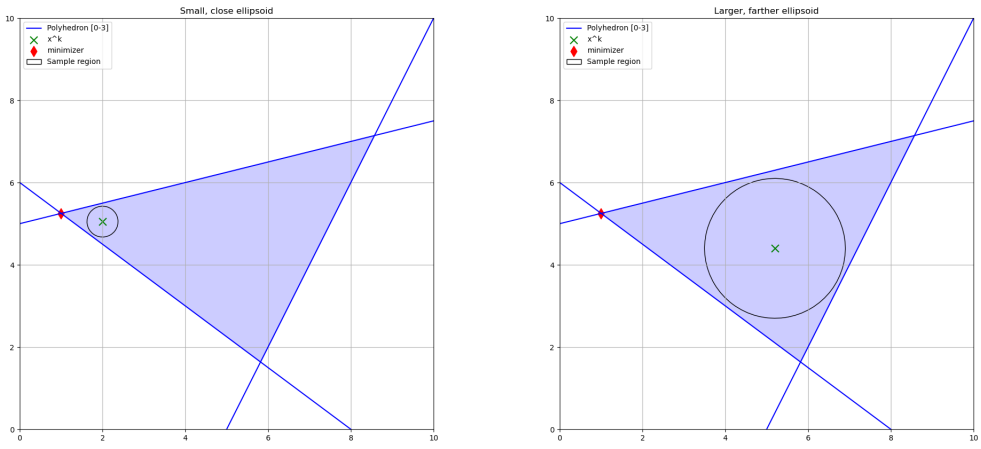

Figure 4: An example of how the search for the feasible region center can go wrong. On the left, a very small sample region is selected; however its proximity to the minimizer makes the model more accurate at the minimizer. On the right, a sample region with larger volume is chosen, but it is further from the trust region minimizer.

Path searches. Rather than searching over all possible centers, it may be more efficient to only move away from the boundary. This can be done using one-dimensional search along an appropriate direction. For example, our first attempt was to simply search a line directed orthogonally away from the closest constraint. This has obvious problems as shown in Figure 5: we should avoid letting the new center get closer to another constraint.

To fix this, we search along a piecewise linear path leading away from the closest constraints. The algorithm works by choosing a set of breakpoints $s_0, s_1, s_2, \ldots, s_{n_{\text{points}}}$ that are each equidistant to a subset of the constraint's faces. Intuitively, the search moves away from the nearest constraint until it reaches a point equidistant to the second nearest constraint, and so on. The center search then considers points along the line segments between these points.

More precisely, the first point is chosen to be the current iterate: $s_0 = x^{(k)}$. The algorithm then repeats the following process for i from 0 to $n_{\text{points}} - 1$. First, compute the set of nearest constraints, where the distance from the current point s_i to each constraint A_j for $j \in [m]$: is given by $d_j = b - A_j x$. Recall that the rows of A are normalized: $\|A_j\| = 1 \quad \forall j \in [m]$. While finding the next point s_{i+1}, let E be the indices of A whose constraints are equidistant and nearest to s_i: $\{j \in [m] | d_j = \min_{l \in [m]} d_l\}$. Let the remaining indices be $R = [m] \setminus E$. The algorithm chooses a search direction $p = {A_E}^T r$ as a linear combination of the normal vectors of the nearest constraints. This search ray r can be found by computing a point s_i whose distance to each equidistant constraint is twice its current value:

$$b_E - A_E(s_i + {A_E}^T r) = 2\left[b_E - A_E s_i\right] \Longrightarrow r = \left[A_E {A_E}^T\right]^{-1}\left(b_E + A_E s_i\right).$$

We can travel along this ray until we reach a point that is the same distance to a remaining constraint. By travelling a distance t, we see that the j-th constraint becomes active when $A_j(s_i + tp) = b_j \Longrightarrow t =$

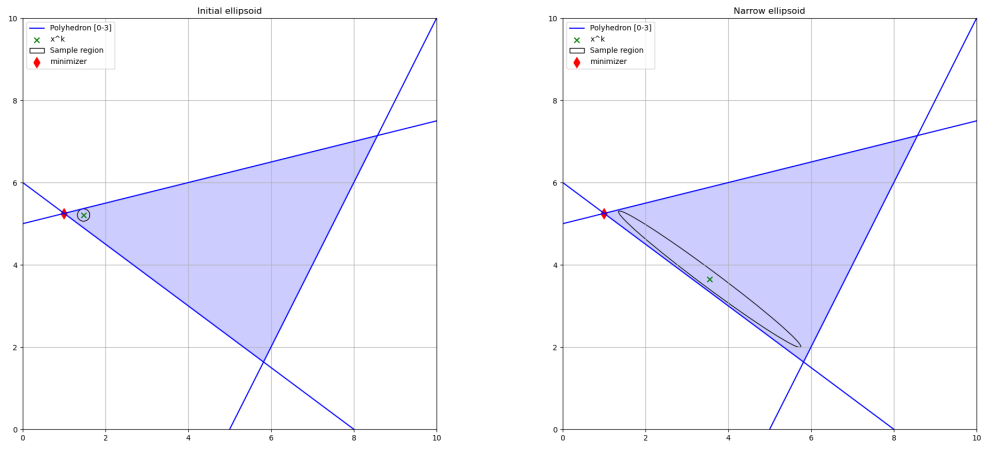

Figure 5: Why only considering the nearest constraint is not sufficient. On the left is the starting ellipsoid. By choosing centers further away from only the nearest constraint, the ellipsoid becomes narrow as another constraint is limiting the length of the second axis.

$\frac{b_j - A_j s_i}{A_j p}$ so that we can travel by $t = \min_{j \in R} \left\{ \frac{b_j - A_j s_i}{A_j p} \right\}$. We set $s_{i+1} = s_i + tp$. This process is described in Algorithm 4. Of course, n_{points} must be less than or equal to $n + 1$ in order for this to be defined.

Algorithm 4: Path segment construction

Step 0 **(Initialization)**
Choose a number of segments $n_{\text{points}} \leq n$, and set $s_0 = x^{(k)}$, $i = 1$.

Step 1 **(Compute nearest constraints)**
Compute $d_j = b - A_j s_{i-1}$ for each $j \in [m]$, and partition

$$E = \left\{ j \in [m] \middle| d_j = \min_{l \in [m]} d_l \right\}, \text{ and } R = [m] \setminus E.$$

If $|E| \geq n$, then **Stop**.

Step 2 **(Compute search segment)**
Compute the search direction $p = {A_E}^T \left[A_E {A_E}^T \right]^{-1} (b_E + A_E s_{i-1})$ and distance $t = \min_{j \in R} \left\{ \frac{b_j - A_j s_{i-1}}{A_j p} \right\}$.
Set $s_i \leftarrow s_{i-1} + tp$.

Step 4 **(Repeat)**
if $i = n_{\text{points}}$ **stop**, otherwise set $i \leftarrow i + 1$ and go to Step 1.

Once each end point s_i is computed, the algorithm searches along each line segment $[s_{i-1}, s_i]$ for $i \in [n_{\text{points}}]$. This means that we can define a class of searches that each limit the number of line segments to search n_{points}. Within the results, these algorithms are described as "ellipse segment n_{points}".

9.2 Buffered segments

Even with these modifications, the iterates may approach the boundary of the feasible region. Another way of dealing with this is to shift the constraints closer to the current iterate. Namely, we introduce a parameter υ to determine how far to scale the constraints. Then, within the trust region subproblem, we add constraints of $Ax \leq b\upsilon + (1-\upsilon)Ax^{(k)}$. This produces the buffered segment searches within the results.

Circumscribed ellipse. We also implemented a variant of the algorithm in which the sample region is a smallest volume ellipsoid to contain the feasible region. Notice the ellipsoid necessarily includes points that are outside the current trust region and may include infeasible points. Thus, both the trust region and linear constraints have to be added to the optimization problem defining the replacement point while computing the Lagrange polynomials. For more details about this modified model improvement algorithm, see **??**.

9.3 Sample problem

The first test was on a problem with simple constraints and a pathological objective. We let $f(x) = \epsilon x + (1-\epsilon)(y - \alpha x \sin(\gamma x))^2$ for a fixed constant ϵ, and set the constraints to be $x_2 \leq ax_1$, $x_2 \geq -ax_1$ for a fixed constant a. We summarize the number of function evaluations and iterations taken within **??**.

In general, we notice the linear models use fewer evaluations than quadratic models. We see that the method with the fewest iterations and function evaluations is the linear polyhedral shape. This is likely because the polyhedral shape is allowed to search the entire outer trust region. This also explains why the circumscribed ellipse and maximum volume simplex also perform well. Also, the scaled ellipsoid performs comparably to the unscaled version.

9.4 Hock-Schittkowksi test problems

We tested these algorithms on several problems from the Hock-Schittkowski problem set [14] and [15]. We selected the problems that have linear constraints: 21, 24, 25, 35, 36, 44, 45, 76, 224, 231, 232, 250, 251. We summarize the results within **??**.

Performance profile. In order to better evaluate the algorithms on the problems across in this test set, we use a performance profile developed in [16]. Given a set of Solvers $\mathcal{S}$ that solved a set of problems $\mathcal{P}$ with the number of evaluations of solver s on problem p being $N(s,p)$, the performance ratio is defined to be $r(s,p) = \frac{N(s,p)}{\min_{s\in\mathcal{S}} N(s,p)}$. If the algorithm does not complete, then the number of evaluations is set to ∞. The performance profile of a solver s and parameter $\alpha \in [0,\infty)$ is then the number of problems for which the performance ratio is less than or equal to α:

$$\rho(s,\alpha) = \frac{1}{\|\mathcal{P}\|} \|p \in \mathcal{P} | r(s,p) \leq \alpha\| . \tag{77}$$

The y axis of a performance plot is the performance profile, and the x axis is the parameter α. Note that algorithms with high performance profiles for small values of α solved a large number of problems the most with the fewest evaluations, while algorithms that eventually reach high performance profiles with larger values of α solve a large set of problems. The performance profile for the Hock-Schittkowski problem set is given in figure Figure 6.

The line segment search with 5 segments does not solve many problems, this is because several of the problems have dimension less than 5, so that it was not even run on these. Notice that the polyhedral search does very well. We conjecture that this may not hold with modeled, nonlinear constraints.

References

[1] N. Neveu, J. Larson, J. G. Power, and L. Spentzouris, "Photoinjector optimization using a derivative-free, model-based trust-region algorithm for the Argonne Wakefield Accelerator," *Journal of Physics: Conference Series*, vol. 874, no. 1, p. 012062, 2017.

[2] N. Ploskas, C. Laughman, A. U. Raghunathan, and N. V. Sahinidis, "Optimization of circuitry arrangements for heat exchangers using derivative-free optimization," *Chemical Engineering Research and Design*, vol. 131, pp. 16–28, 2018. Energy Systems Engineering.

[3] N. B. Kovachki and A. M. Stuart, "Ensemble Kalman inversion: a derivative-free technique for machine learning tasks," *Inverse Problems*, vol. 35, p. 095005, Aug 2019.

[4] Z. Cheng, E. Shaffer, R. Yeh, G. Zagaris, and L. Olson, "Efficient parallel optimization of volume meshes on heterogeneous computing systems," *Engineering with Computers*, vol. 33, no. 4, pp. 717–726, 2017.

[5] T. Gao and J. Li, "A derivative-free trust-region algorithm for reliability-based optimization," *Structural and Multidisciplinary Optimization*, vol. 55, no. 4, pp. 1535–1539, 2017.

[6] J. S. Eldred, J. Larson, M. Padidar, E. Stern, and S. M. Wild, "Derivative-free optimization of a rapid-cycling synchrotron," Tech. Rep. 2108.04774, ArXiv, 2021.

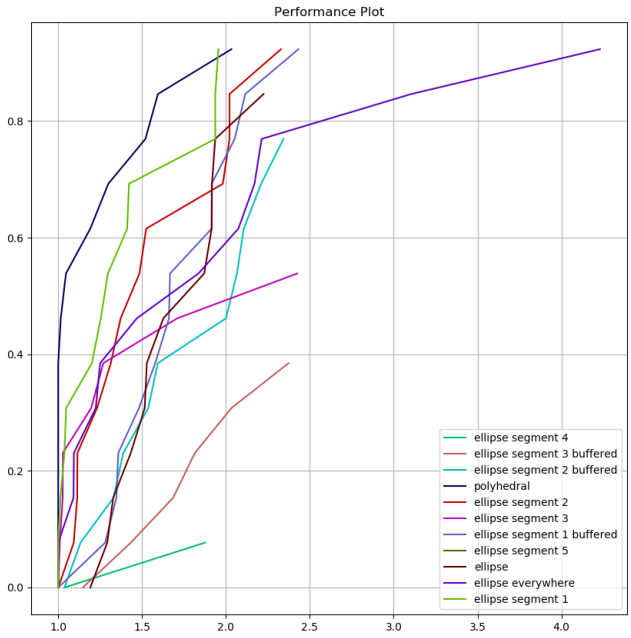

Figure 6: A performance profile comparing different variants of the algorithm for linear constraints. We see that the polyhedral algorithm is both efficient and robust.

[7] S. Le Digabel and S. M. Wild, "A taxonomy of constraints in simulation-based optimization," Tech. Rep. 1505.07881, ArXiv, 2015.

[8] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to Derivative-Free Optimization*. Society for Industrial and Applied Mathematics, 2009.

[9] P. D. Conejo, E. W. Karas, L. G. Pedroso, A. A. Ribeiro, and M. Sachine, "Global convergence of trust-region algorithms for convex constrained minimization without derivatives," *Appl. Math. Comput.*, vol. 220, pp. 324–330, 2013.

[10] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-region Methods*. Society for Industrial and Applied Mathematics, 2000.

[11] J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*. Englewood Cliffs, N.J: Prentice-Hall, 1983.

[12] S. C. Billups, J. Larson, and P. Graf, "Derivative-free optimization of expensive functions with computational error using weighted regressions," *SIAM Journal on Optimization*, vol. 23, no. 1, pp. 27–53, 2013.

[13] L. G. Khachiyan and M. J. Todd, "On the complexity of approximating the maximal inscribed ellipsoid for a polytope," *Mathematical Programming*, vol. 61, no. 1, pp. 137–159, 1993.

[14] K. Schittkowski, *More Test Examples for Nonlinear Programming Codes*. No. 282 in Lecture Notes in Economics and Mathematical Systems, Berlin, Heidelberg: Springer-Verlag, 1987.

[15] W. Hock and K. Schittkowski, "Test examples for nonlinear programming codes," *Journal of Optimization Theory and Applications*, vol. 30, no. 1, pp. 127–129, 1980.

[16] J. J. Moré and S. M. Wild, "Benchmarking derivative-free optimization algorithms," *SIAM Journal on Optimization*, vol. 20, no. 1, pp. 172–191, 2009.