UNIVERSITÉ DE NAMUR

naXys
Namur Center for Complex Systems

A trust-funnel method for nonlinear optimization problems with general nonlinear constraints and its application to derivative-free optimization

by Ph. R. Sampaio and Ph. L. Toint

Report naXys-03-2015 January 26, 2015

University of Namur, 61, rue de Bruxelles, B5000 Namur (Belgium)

http://www.unamur.be/sciences/naxys

A trust-funnel method for nonlinear optimization problems with general nonlinear constraints and its application to derivative-free optimization

Ph. R. Sampaio* and Ph. L. Toint†

January 26, 2015

Abstract

A trust-funnel method is proposed for solving nonlinear optimization problems with general nonlinear constraints. It extends the one presented by Gould and Toint (*Math. Prog.*, 122(1):155-196, 2010), originally proposed for equality-constrained optimization problems only, to problems with both equality and inequality constraints and where simple bounds are also considered. As the original one, our method makes use of neither filter nor penalty functions and consider the objective function and the constraints as independently as possible. To handle the bounds, an active-set approach is employed. At last, we exploit and incorporate techniques developed for derivative-free optimization to obtain a final method that can also be used to solve problems where the derivatives are unavailable or are available at a prohibitive cost.

Keywords: Constrained nonlinear optimization, trust-region method, trust funnel, derivative-free optimization.

1 Introduction

We consider a trust-funnel method for the solution of the nonlinear optimization problem

$$\begin{cases} \min\limits_{x} & f(x) \\ \text{s.t.:} & l^s \leq c(x) \leq u^s, \\ & l^x \leq x \leq u^x, \end{cases} \tag{1.1}$$

where we assume that $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable, and that f is bounded below on the feasible domain. The vectors l^s and u^s are lower and upper bounds, respectively, on the constraints' values $c(x)$, while l^x and u^x are bounds on the x variables, with $l^s \in (\mathbb{R} \cup -\infty)^m$, $u^s \in (\mathbb{R} \cup \infty)^m$, $l^x \in (\mathbb{R} \cup -\infty)^n$ and $u^x \in (\mathbb{R} \cup \infty)^n$.

By defining $f(x, s) \stackrel{\text{def}}{=} f(x)$ and $c(x, s) \stackrel{\text{def}}{=} c(x) - s$, the problem above may be rewritten as the following equality-constrained optimization problem with simple bounds

$$\begin{cases} \min\limits_{(x,s)} & f(x, s) \\ \text{s.t.:} & c(x, s) = 0, \\ & l^s \leq s \leq u^s, \\ & l^x \leq x \leq u^x, \end{cases} \tag{1.2}$$

*Namur Center for Complex Systems (naXys) and Department of Mathematics, University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium. Email: phillipe.sampaio@unamur.be

†Namur Center for Complex Systems (naXys) and Department of Mathematics, University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium. Email: philippe.toint@unamur.be

which is the one we address throughout this paper.

The trust-funnel method was firstly introduced by Gould and Toint [12] (see correction in [11]) as a SQP (*Sequential Quadratic Programming*) algorithm for equality-constrained optimization problems whose convergence is driven by an adaptive bound imposed on the allowed infeasibility at each iteration. Such bound is monotonically decreased as the algorithm progresses, assuring its global convergence whilst seeking optimality and hence originating the name "trust funnel". It belongs to the class of trust-region methods and makes use of a composite-step approach to calculate a new direction at each iteration: a normal step is firstly computed in the hope of reducing the infeasibility measure ensuing from the constraint functions' values, and a tangent step is subsequently calculated with the aim of improving optimality of the iterates with regard to the objective function. These computations are carried out with the use of two different trust regions, one for each step component. The main idea is to consider the objective function and the constraints as independently as possible. The method is noticeable among others for constrained problems as a parameter-free alternative, for neither filter nor penalties are needed, freeing the user from common difficulties encountered when choosing the initial penalty parameter, for instance. An extension to problems with both equalities and inequalities was developed of late by Curtis *et al.* [8], who presented an interior-point trust-funnel algorithm for solving large-scale problems that may be characterized as a barrier-SQP method.

The fast-growing need for optimization methods that do not consider derivatives in fields like medicine, chemistry, engineering and many others, aroused the optimization community's interest to come up with new algorithms, unleashing the blossom of a new research field on optimization and hatching new ideas for optimizing functions. Although many methods have been designed for unconstrained DFO (*Derivative-Free Optimization*), the range of possibilities for the constrained case remains quite open. A derivative-free trust-funnel method called DEFT-FUNNEL (DErivative-Free Trust FUNNEL), proposed by Sampaio and Toint [22] for equality-constrained problems, has given good results in comparison with a filter-SQP method introduced by Colson [4] and a trust-region method by Powell named COBYLA (Constrained Optimization BY Linear Approximations) [19, 20], all centred on local polynomial interpolation models. Recently, another trust-region method by Powell [21] entitled LINCOA has been developed for linearly-constrained optimization without derivatives and makes use of quadratic interpolation-based models for the objective function. Many other methods that use direct search instead of trust-region techniques combined to interpolation-based models also have been developed for the constrained case (e.g., Lewis and Torczon [15, 16, 17], Audet and Dennis [1], Lucidi *et al.* [18], Yu and Li [24]).

The final method described here may be considered as an extension of the DEFT-FUNNEL method presented in [22], for it employs many of its techniques for the maintenace of the quality of the interpolation models. It features four main steps to solve problems with general nonlinear constraints, namely: a subspace minimization approach to handle the bounds on the x variables, which makes the DEFT-FUNNEL an active-set method, a bounded linear least-squares solver to calculate the normal step, a projected gradient method to calculate the tangent step and the control of the permitted infeasibility of the iterates through the funnel bound. The reason behind the choice of exploring subspaces defined by the active bounds is dual. Besides the fact that we aim to avoid treating the bounds on the x variables as general inequality constraints, the reduction of the dimension of the problem after having identified the active bounds helps to thwart a possible degeneration of the interpolation set when the sample points become close to each other and thus affinely dependent, which happens often as the optimal solution is approached. The fact that the s variables play no role on the choice of the interpolation set vindicates the construction of the subspaces based upon the x variables only.

This paper is organized as follows. Section 2 introduces the proposed trust-funnel algorithm step by step. The subsection 2.1 explains the subspace minimization approach, while the subsections 2.2 and 2.3 give details on how the computation of the composite step components is conducted. The

subsections 2.4 and 2.5 then address the whole mechanism of the algorithm itself. Section 3 concerns the application of the method to derivative-free optimization problems and Section 4 gives the description of the final method, which assembles the techniques elaborated in Sections 2 and 3. In Section 5, we discuss some initial numerical experiments and compare its performance to the COBYLA's in a selected set of test problems. Ultimately, we draw some final conclusions in Section 6.

Notation. Unless otherwise specified, our norm $\|\cdot\|$ is the standard Euclidean norm. Given any set $\mathcal{A}$, $|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$. We let $\mathcal{B}(z;\Delta)$ denote the closed Euclidian ball centered at z, with radius $\Delta > 0$. By $\mathcal{P}_n^d$, we mean the space of all polynomials of degree at most d in $\mathbb{R}^n$. Finally, given any subspace $\mathcal{S}$, we denote its dimension by $dim(\mathcal{S})$.

2 An active-set trust-funnel method

Our method generates a sequence of points $\{(x_k, s_k)\}$ such that, at each iteration k, the bound constraints below are satisfied

$$l^s \leq s_k \leq u^s, \tag{2.3}$$
$$l^x \leq x_k \leq u^x. \tag{2.4}$$

By using a composite-step approach, each trial step $d_k \stackrel{\text{def}}{=} (d_k^x, d_k^s)^T$ is decomposed as

$$d_k = \begin{pmatrix} d_k^x \\ d_k^s \end{pmatrix} = \begin{pmatrix} n_k^x \\ n_k^s \end{pmatrix} + \begin{pmatrix} t_k^x \\ t_k^s \end{pmatrix} = n_k + t_k,$$

where the *normal step* component $n_k \stackrel{\text{def}}{=} (n_k^x, n_k^s)^T$ aims to improve feasibility, and the *tangent step* component $t_k \stackrel{\text{def}}{=} (t_k^x, t_k^s)^T$ reduces the objective function model's value while preserving any gains in feasibility obtained through n_k.

In the following subsections, we will describe how each component is computed. Before that, we briefly explain how the subspace minimization is employed in our algorithm.

2.1 Subspace minimization

As in the method proposed by Gratton *et al.* [13] for bound-constrained optimization problems, our algorithm makes use of an active-set approach where the minimization is restricted to subspaces defined by the active x variables.

At each iteration k, we define the subspace $\mathcal{S}_k$ as follows

$$\mathcal{S}_k \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid x_i = l_k^x \ \text{ for } \ i \in \mathcal{L}_k \ \text{ and } \ x_i = u_k^x \ \text{ for } \ i \in \mathcal{U}_k\},$$

where $\mathcal{L}_k \stackrel{\text{def}}{=} \{i \mid x_{k_i} - l_i^x \leq \epsilon_b\}$ and $\mathcal{U}_k \stackrel{\text{def}}{=} \{i \mid u_i^x - x_{k_i} \leq \epsilon_b\}$ define the index sets of (nearly) active variables at their bounds, for some small constant $\epsilon_b > 0$ defined *a priori*. After that $\mathcal{S}_k$ has been defined, the minimization at iteration k is then restricted to the new subspace $\mathcal{S}_k$. Once a direction d_k for (x_k, s_k) has been computed, we set $(x_{k+1}, s_{k+1}) = (x_k, s_k) + d_k$ if k is a successful iteration; otherwise, we set $(x_{k+1}, s_{k+1}) = (x_k, s_k)$.

If a solution $(\tilde{x}_k, \tilde{s}_k)$ for the subproblem defined by $\mathcal{S}_k$ satisfies the optimality conditions for the subproblem, we check whether it is also optimal for the original problem (2.5). If it is, the solution encountered is returned to the user and the algorithm halts; otherwise, it proceeds in the full space by computing a new direction for $(\tilde{x}_k, \tilde{s}_k)$ and repeats the above process at iteration $k+1$ by defining $\mathcal{S}_{k+1}$.

2.2 The normal step

For any point (x, s), we measure the constraint violation by

$$v(x,s) \stackrel{\text{def}}{=} \tfrac{1}{2}\|c(x,s)\|^2. \tag{2.5}$$

Analogous to the trust-funnel method for equality-constrained problems proposed in [12] and in [22], we also have a funnel bound $v_k^{\max}$ for v such that, for each iteration k,

$$v_k \leq v_k^{\max},$$

where $v_k \stackrel{\text{def}}{=} v(x_k, s_k)$. As this bound is monotonically decreased, the algorithm is driven towards feasibility, guaranteeing the convergence of the algorithm.

Before computing the normal step at any iteration k, we verify that the measure of constraint violation is substantial enough at the iterate. This is done by comparing it to the measure of optimality to be defined in the next subsection. If the constraint violation is significant, a normal step n_k can be computed by reducing the Gauss-Newton model of v at (x_k, s_k) within a trust region while care is taken to ensure that the conditions (2.3) and (2.4) are fulfilled.

To ensure that the normal step is indeed "normal", we add the following condition that asks that it mostly lies in the space spanned by the columns of the matrix $J(x_k, s_k)^T$:

$$\|n_k\|_\infty \leq \kappa_n \|c(x_k, s_k)\|, \tag{2.6}$$

for some $\kappa_n \geq 1$. We then perform the calculation of n_k by solving the trust-region bound-constrained linear least-squares problem

$$\left\{\begin{array}{rc} \min\limits_{n=(n^x,n^s)} & \tfrac{1}{2}\|c(x_k,s_k)+J(x_k,s_k)n\|^2 \\ \text{s.t.:} & l^s \leq s_k + n^s \leq u^s, \\ & l^x \leq x_k + n^x \leq u^x, \\ & x_k + n^x \in \mathcal{S}_k, \\ & n \in \mathcal{N}_k, \end{array}\right. \tag{2.7}$$

where $J(x,s) \stackrel{\text{def}}{=} (J(x) \ -I)$ represents the Jacobian of $c(x,s)$ with respect to (x,s) and

$$\mathcal{N}_k \stackrel{\text{def}}{=} \{z \in \mathbb{R}^{n+m} \mid \|z\|_\infty \leq \min[\,\Delta_k^c,\, \kappa_n \|c(x_k,s_k)\|\,]\}, \tag{2.8}$$

for some trust-region radius $\Delta_k^c > 0$.

Rather than solving (2.7) exactly, we simply require that the computed step n_k produces a reduction in the linear part of the Gauss-Newton model of v at (x_k, s_k) which is at least a fraction of that achieved by the projected Cauchy direction. We know from Section 12.2.1 of [5] that the following modified Cauchy condition is then ensured:

$$\begin{aligned} \delta_k^{c,n} &\stackrel{\text{def}}{=} \tfrac{1}{2}\|c(x_k,s_k)\|^2 - \tfrac{1}{2}\|c(x_k,s_k)+J(x_k,s_k)n_k\|^2 \\ &\geq \kappa_{nC}\pi_k^v \min\left[\frac{\pi_k^v}{1+\|W_k\|}, \Delta_k^c, 1\right] \geq 0, \end{aligned} \tag{2.9}$$

where $W_k = J(x_k,s_k)^T J(x_k,s_k)$ is the symmetric Gauss-Newton approximation of the Hessian of v at (x_k, s_k), $\kappa_{nC} \in (0, \frac{1}{2}]$ and $\pi_k^v \stackrel{\text{def}}{=} -\langle J(x_k,s_k)^T c(x_k,s_k), b_k\rangle$ is a v-criticality measure that indicates how much decrease could be obtained locally along the projection of the negative gradient of the

Gauss-Newton model of v at (x_k, s_k) onto the region delimited by the bounds. The projected Cauchy direction b_k is, in turn, given by the solution of

$$\left\{\begin{array}{rc} \min\limits_{b=(b^x,b^s)} & \langle J(x_k,s_k)^T c(x_k,s_k), b\rangle \\ \text{s.t.:} & l^s \leq s_k + b^s \leq u^s, \\ & l^x \leq x_k + b^x \leq u^x, \\ & x_k + b^x \in \mathcal{S}_k, \\ & \|b\|_\infty \leq 1. \end{array}\right. \tag{2.10}$$

2.3 The tangent step

After attempting to reduce infeasibility through the normal step, a tangent step is calculated with the aim of improving optimality. The computation of the latter is conducted carefully so that the gains in feasibility obtained by the former are not abandoned without good reasons.

The SQP model for the function f is defined as

$$\psi_k((x_k,s_k)+d) \stackrel{\text{def}}{=} f_k + \langle g_k, d\rangle + \tfrac{1}{2}\langle d, B_k d\rangle, \tag{2.11}$$

where $f_k \stackrel{\text{def}}{=} f(x_k,s_k)$, $g_k \stackrel{\text{def}}{=} \nabla_{(x,s)} f(x_k,s_k)$, and B_k is the approximate Hessian of the Lagrangian function

$$\mathcal{L}(x,s,\mu,z^s,w^s,z^x,w^x) = f(x) + \langle \mu, c(x,s))\rangle + \langle w^s, s-u^s\rangle + \langle z^s, l^s - s\rangle + \langle w^x, x-u^x\rangle + \langle z^x, l^x - x\rangle$$

with respect to (x,s), given by

$$B_k = \begin{pmatrix} G_k & 0 \\ 0 & 0 \end{pmatrix},$$

where G_k is a symmetric approximation of the Hessian of the Lagrangian with respect to x defined as

$$G_k \stackrel{\text{def}}{=} H_k + \sum_{i=1}^{m} [\hat{\mu}_k]_i C_{ik}, \tag{2.12}$$

z^s and w^s are the Lagrange multipliers associated to the lower and upper bounds, respectively, on the slack variables s, and z^x and w^x, the Lagrange multipliers associated to the lower and upper bounds on the x variables. In (2.12), H_k is a bounded symmetric approximation of $\nabla^2_{xx} f(x_k,s_k) = \nabla^2 f(x_k)$, the matrices C_{ik} are bounded symmetric approximations of the constraints' Hessians $\nabla^2_{xx} c_{ik}(x_k,s_k) = \nabla^2 c_{ik}(x_k)$ and the vector $\hat{\mu}_k$ may be viewed as a bounded local approximation of the Lagrange multipliers with respect to the equality constraints $c(x,s)$, in the sense that we require that

$$\|\hat{\mu}_k\| \leq \kappa_\mu, \tag{2.13}$$

for some $\kappa_\mu > 0$.

By using the decomposition $d_k = n_k + t_k$, we then have that

$$\psi_k((x_k,s_k)+n_k) = f_k + \langle g_k, n_k\rangle + \tfrac{1}{2}\langle n_k, B_k n_k\rangle \tag{2.14}$$

and

$$\begin{array}{rcl} \psi_k((x_k,s_k)+n_k+t) & = & f_k + \langle g_k, n_k+t\rangle + \tfrac{1}{2}\langle n_k+t, B_k\,(n_k+t)\rangle \\ & = & \psi_k((x_k,s_k)+n_k) + \langle g_k^N, t\rangle + \tfrac{1}{2}\langle t, B_k t\rangle, \end{array} \tag{2.15}$$

where

$$g_k^N \stackrel{\text{def}}{=} g_k + B_k\, n_k. \tag{2.16}$$

In the interest of assuring that (2.15) it is a proper local approximation for the function $f((x_k, s_k) + n_k + t)$, the complete step $d = n_k + t$ must belong to

$$\mathcal{T}_k \stackrel{\text{def}}{=} \{d \in \mathbb{R}^{n+m} \mid \|d\|_\infty \leq \Delta_k^f\}, \tag{2.17}$$

for some radius Δ_k^f. The minimization of (2.15) should then be restricted to the intersection of $\mathcal{N}_k$ and $\mathcal{T}_k$, which imposes that the *tangent step* t_k results in a complete step $d_k = n_k + t_k$ that satisfies the inclusion

$$d_k \in \mathcal{R}_k \stackrel{\text{def}}{=} \mathcal{N}_k \cap \mathcal{T}_k \stackrel{\text{def}}{=} \{d \in \mathbb{R}^{n+m} \mid \|d\|_\infty \leq \Delta_k\}, \tag{2.18}$$

where the radius Δ_k of $\mathcal{R}_k$ is thus given by

$$\Delta_k = \min[\Delta_k^c, \Delta_k^f]. \tag{2.19}$$

Similarly to the computation of the normal step, we intend to remain in the subspace $\mathcal{S}_k$ after walking along the tangent direction. We accomplish that by imposing the following condition

$$x_k + n_k^x + t^x \in \mathcal{S}_k.$$

Additionaly, the conditions (2.3) and (2.4) must be satisfied at the final point $(x_k, s_k) + d_k$, i.e., we must have

$$\begin{aligned} l^s \leq s_k + n_k^s + t^s \leq u^s, \\ l^x \leq x_k + n_k^x + t^x \leq u^x. \end{aligned}$$

Due to (2.18), we ask n_k to belong to $\mathcal{R}_k$ before attempting the computation of t_k by requiring that

$$\|n_k\|_\infty \leq \kappa_{\mathcal{R}} \Delta_k, \tag{2.20}$$

for some $\kappa_{\mathcal{R}} \in (0, 1)$. If (2.20) happens, which means that there is "enough space left" to make another step without crossing the trust region border, the tangent step is finally computed by (approximately) solving the following problem

$$\left\{ \begin{array}{rc} \min\limits_{t=(t^x,t^s)} & \langle g_k^N, t\rangle + \frac{1}{2}\langle t, B_k t\rangle \\ \text{s.t.:} & J(x_k, s_k)t = 0, \\ & l^s \leq s_k + n_k^s + t^s \leq u^s, \\ & l^x \leq x_k + n_k^x + t^x \leq u^x, \\ & x_k + n_k^x + t^x \in \mathcal{S}_k. \\ & n_k + t \in \mathcal{R}_k. \end{array} \right. \tag{2.21}$$

In practice, we do not solve the problem (2.21) exactly either, rather we only require a "sufficient" reduction of (2.15) in the sense that the tangent step must produce a reduction in the model ψ_k which is at least a fraction of that achieved by solving the modified Cauchy point subproblem

$$\min_{\substack{\tau>0 \\ (x_k,s_k)+n_k+\tau r_k \in \mathcal{R}_k}} \psi_k((x_k, s_k) + n_k + \tau r_k), \tag{2.22}$$

where r_k is the projected Cauchy direction obtained by solving the linear optimization problem

$$\left\{ \begin{array}{rc} \min\limits_{r=(r^x,r^s)} & \langle g_k^N, r\rangle \\ \text{s.t.:} & J(x_k, s_k)r = 0, \\ & l^s \leq s_k + n_k^s + r^s \leq u^s, \\ & l^x \leq x_k + n_k^x + r^x \leq u^x, \\ & x_k + n_k^x + r^x \in \mathcal{S}_k. \\ & \|r\|_\infty \leq 1. \end{array} \right. \tag{2.23}$$

We then define our f-criticality measure as

$$\pi_k^f \stackrel{\text{def}}{=} -\langle g_k^N, r_k \rangle. \tag{2.24}$$

By definition, π_k^f measures how much decrease could be obtained locally along the projection of the negative of the approximate gradient g_k^N onto the nullspace of $J(x_k, s_k)$ intersected to the region delimited by the bounds. This procedure ensures, for some $\kappa_{tC} \in (0, 1]$, the modified Cauchy condition

$$\delta_k^{f,t} \stackrel{\text{def}}{=} \psi_k((x_k, s_k) + n_k) - \psi_k((x_k, s_k) + n_k + t_k) \geq \kappa_{tC} \pi_k^f \min\left[\frac{\pi_k^f}{1 + \|B_k\|}, \Delta_k, 1\right] > 0. \tag{2.25}$$

A new local estimate of the Lagrange multipliers $(\mu_k, z_k^s, w_k^s, z_k^x, w_k^x)$ are computed by solving the following bound-constrained linear least-squares problem

$$\left\{ \begin{array}{rl} \min\limits_{(\mu, \hat{z}^s, \hat{w}^s, \hat{z}^x, \hat{w}^x)} & \frac{1}{2}\|\mathcal{M}_k(\mu, \hat{z}^s, \hat{w}^s, \hat{z}^x, \hat{w}^x)\|^2. \\ \text{s.t.:} & \hat{z}^s, \hat{w}^s, \hat{z}^x, \hat{w}^x \geq 0, \end{array} \right. \tag{2.26}$$

where

$$\begin{aligned} \mathcal{M}_k(\mu, \hat{z}^s, \hat{w}^s, \hat{z}^x, \hat{w}^x) \stackrel{\text{def}}{=} & \begin{pmatrix} g_k^N \\ 0 \end{pmatrix} + \begin{pmatrix} J(x_k)^T \\ -I \end{pmatrix} \mu + \begin{pmatrix} 0 \\ I_w^s \end{pmatrix} \hat{w}^s + \begin{pmatrix} 0 \\ -I_z^s \end{pmatrix} \hat{z}^s \\ & + \begin{pmatrix} I_w^x \\ 0 \end{pmatrix} \hat{w}^x + \begin{pmatrix} -I_z^x \\ 0 \end{pmatrix} \hat{z}^x, \end{aligned}$$

the matrix I is the $m \times m$ identity matrix, the matrices I_z^s and I_w^s are obtained from I by removing the columns whose indices are not associated to any active (lower and upper, respectively) bound at $s_k + n_k^s$, the matrices I_z^x and I_w^x are obtained from the $n \times n$ identity matrix by removing the columns whose indices are not associated to any active (lower and upper, respectively) bound at $x_k + n_k^x$, and the Lagrange multipliers $(\hat{z}^s, \hat{w}^s, \hat{z}^x, \hat{w}^x)$ are those in (z^s, w^s, z^x, w^x) associated to active bounds at $s_k + n_k^s$ and $x_k + n_k^x$. All the other Lagrange multipliers are set to zero.

2.4 Which steps to compute and retain

We now explain when the normal step n_k and the tangent step t_k should be computed at iteration k given the constraint violation and the measure of optimality described in the previous subsection.

The normal step is computed when $k = 0$ or the current constraint violation is "significant", which is now formally defined by the conditions

$$\|c(x_k, s_k)\| > \omega_n(\pi_{k-1}^f) \text{ or } v_k > \kappa_{vv} v_k^{\max}, \tag{2.27}$$

where ω_n is some bounding function, $\kappa_{vv} \in (0, 1)$ is a constant. If (2.27) fails, the computation of the normal step is not required and so we set $n_k = 0$.

If the solution of (2.23) is $r_k = 0$, then by (2.24) we have $\pi_k^f = 0$. In this case, the computation of the tangent step is skipped, and we simply set $t_k = 0$. If π_k^f is unsubstantial compared to the current infeasibility, i.e., for a given a monotonic bounding function ω_t, the condition

$$\pi_k^f > \omega_t(\|c(x_k, s_k)\|) \tag{2.28}$$

fails, then the current iterate is still too far from feasibility to worry about optimality, and we again skip the tangent step computation by setting $t_k = 0$.

For technical reasons, we require that

$$\omega_n(t) = 0 \iff t = 0 \quad \text{and} \quad \omega_t(\omega_n(t)) \leq \kappa_\omega t, \tag{2.29}$$

for all $t \geq 0$ and for some $\kappa_\omega \in (0, 1)$. Examples for the functios $\omega_n(t)$ and $\omega_t(t)$ might be

$$\omega_n(t) \stackrel{\text{def}}{=} 0.01 \min[1,\ t] \quad \text{and} \quad \omega_t(t) \stackrel{\text{def}}{=} 0.01 \min[1,\ t^2]. \tag{2.30}$$

We say that (x_k, s_k) is an infeasible stationary point if $c(x_k, s_k) \neq 0$ and $\pi_k^v = 0$. While (2.28) and (2.27) together provide considerable flexibility in our algorithm in that a normal or tangent step is only computed when relevant, our setting also produces the possibility that both these conditions fail. In this case, we have that $d_k = n_k + t_k$ is identically zero, and the sole computation in the iteration is that of the new Lagrange multipliers (2.26).

The conditions (2.29) are important to prove a version of the following lemma for equality-constrained problems, which can be found in [12].

Lemma 2.1. For all k such that $d_k = 0$ and (x_k, s_k) is not an infeasible stationary point,

$$\pi_k^f \leq \kappa_\omega \pi_{k-1}^f.$$

It is argued in that reference that the initial assumption that is f bounded below on the feasible domain and the fact the the algorithm terminates when an infeasible stationary point is found imply that such behavior ($d_k = 0$) cannot persist unless x_k is optimal.

Finally, we may evaluate the usefulness of the tangent step t_k after (or during) its computation, in the sense that we would like a relatively large tangent step to cause a clear decrease in the model (2.15) of the objective function. We therefore check whether the conditions

$$\|t_k\| > \kappa_{\mathcal{CS}} \|n_k\| \tag{2.31}$$

and

$$\delta_k^f \stackrel{\text{def}}{=} \delta_k^{f,t} + \delta_k^{f,n} \geq \kappa_\delta \delta_k^{f,t}, \tag{2.32}$$

where

$$\delta_k^{f,n} \stackrel{\text{def}}{=} \psi_k(x_k, s_k) - \psi_k((x_k, s_k) + n_k), \tag{2.33}$$

are satisfied for some $\kappa_{\mathcal{CS}} > 1$ and for $\kappa_\delta \in (0, 1)$. The inequality (2.32) indicates that the *predicted* improvement in the objective function obtained in the tangent step is not negligible compared to the *predicted* change in f resulting from the normal step. If (2.31) holds but (2.32) fails, the tangent step is not useful in the sense just discussed, and we choose to ignore it by resetting $t_k = 0$.

2.5 Iterations types

Once we have computed the step d_k and the trial point

$$(x_k^+, s_k^+) \stackrel{\text{def}}{=} (x_k, s_k) + d_k, \tag{2.34}$$

we are left with the task of accepting or rejecting it. If $n_k = t_k = 0$, iteration k is said to be a μ-iteration because the only computation potentially performed is that of a new vector of Lagrange multiplier estimates. We will say that iteration k is an f-iteration if $t_k \neq 0$, (2.32) holds, and

$$v(x_k^+, s_k^+) \leq v_k^{\max}. \tag{2.35}$$

Condition (2.35) ensures that the step keeps feasibility within reasonable bounds. Thus the iteration's expected major achievement is, in this case, a decrease in the value of the objective function f, hence its name. If iteration k is neither a μ-iteration nor a f-iteration, then it is said to be a c-iteration. If (2.32) fails, then the expected major achievement (or failure) of iteration k is, *a contrario*, to improve feasibility, which is also the case when the step only contains its normal component.

The main idea behind the technique for accepting the trial point is to measure whether the major expected achievement of the iteration has been realized.

- If iteration k is a μ-iteration, we do not have any other choice than to restart with $(x_{k+1}, s_{k+1}) = (x_k, s_k)$ using the new multipliers. We then define

$$\Delta_{k+1}^f = \Delta_k^f \text{ and } \Delta_{k+1}^c = \Delta_k^c \tag{2.36}$$

  and keep the current value of the maximal infeasibility $v_{k+1}^{\max} = v_k^{\max}$.

- If iteration k is an f-iteration, we accept the trial point (i.e., $(x_{k+1}, s_{k+1}) = (x_k^+, s_k^+)$) if

$$\rho_k^f \stackrel{\text{def}}{=} \frac{f(x_k, s_k) - f(x_k^+, s_k^+)}{\delta_k^f} \geq \eta_1, \tag{2.37}$$

  and reject it (i.e., $(x_{k+1}, s_{k+1}) = (x_k, s_k)$), otherwise. The value of the maximal infeasibility measure is left unchanged, that is $v_{k+1}^{\max} = v_k^{\max}$. Note that $\delta_k^f > 0$ (because of (2.25) and (2.32)) unless (x_k, s_k) is first-order critical, and hence that condition (2.37) is well-defined.

- If iteration k is a c-iteration, we accept the trial point if the improvement in feasibility is comparable to its predicted value

$$\delta_k^c \stackrel{\text{def}}{=} \tfrac{1}{2}\|c(x_k, s_k)\|^2 - \tfrac{1}{2}\|c(x_k, s_k) + J(x_k, s_k)d_k\|^2,$$

  and the latter is itself comparable to its predicted decrease along the normal step, that is

$$n_k \neq 0, \quad \delta_k^c \geq \kappa_{cn}\delta_k^{c,n} \text{ and } \rho_k^c \stackrel{\text{def}}{=} \frac{v(x_k, s_k) - v(x_k^+, s_k^+)}{\delta_k^c} \geq \eta_1, \tag{2.38}$$

  for some $\kappa_{cn} \in (0, 1 - \kappa_{tg}]$. If (2.38) fails, the trial point is rejected. We update the value of the maximal infeasibility by

$$v_{k+1}^{\max} = \begin{cases} \max\left[\kappa_{tx1} v_k^{\max}, v(x_k^+, s_k^+) + \kappa_{tx2}(v(x_k, s_k) - v(x_k^+, s_k^+))\right] & \text{if (2.38) hold,} \\ v_k^{\max} & \text{otherwise,} \end{cases} \tag{2.39}$$

  for some $\kappa_{tx1} \in (0, 1)$ and $\kappa_{tx2} \in (0, 1)$.

We now describe why the last condition in (2.38) is well-defined. Firstly, we only check the third condition *after* the first two conditions have been verified. Assuming that $n_k \neq 0$, the Cauchy condition (2.9) and $c(x_k, s_k) \neq 0$ ensure that $\delta_k^{c,n} > 0$ provided $\pi_k^v \neq 0$. Thus the third condition is well defined, unless (x_k, s_k) is an infeasible stationary point, in which case the algorithm is terminated.

3 Application to derivative-free optimization

Our algorithm makes use of surrogate models $m^f(x)$ and $m^c(x) = (m_1^c(x), m_2^c(x), \ldots, m_m^c(x))$ built from polynomial interpolation over a set of sample points $\mathcal{Y} = \{y^0, y^1, \ldots, y^p\}$ that replace the functions $f(x)$ and $c(x) = (c_1(x), c_2(x), \ldots, c_m(x))$. Therefore, at each iteration k, the following interpolation conditions are satisfied

$$\begin{aligned} m^f(y^i) &= f(y^i), \\ m_j^c(y^i) &= c_j(y^i), \end{aligned} \tag{3.40}$$

for all $y^i \in \mathcal{Y}_k$.

Whenever $\mathcal{Y}_k$ changes, the models m_k^c and m_k^f are updated to satisfy the interpolation conditions (3.40) for the new set $\mathcal{Y}_{k+1}$, thereby implying that new function evaluations of f and c are carried out for the additional points obtained at iteration k.

The algorithm developed in this work employs the commonly used idea of starting with incomplete interpolation models with linear accuracy and then enhancing them with curvature information, which provides an actual accuracy at least as good as that for linear models and, hopefully, better. We thus consider underdetermined quadratic interpolation — i.e., $n+1 \leq |\mathcal{Y}| \leq (n+1)(n+2)/2$ — with initial sample sets that are poised for linear interpolation.

As there are many possibilities for building underdetermined quadratic interpolation-based models, we consider the following distinct ways in our algorithm for building them: subbasis selection approach, minimum ℓ_2-norm models and minimum Frobenius norm models. In addition, we also consider least-squares regression for the construction of quadratic models, in which case one has $n+1 \leq |\mathcal{Y}| \leq (n+1)(n+2)$, giving the user a total of four possibilities for model building. A complete description of these approaches is available in Conn, Scheinberg and Vicente [6].

In order to present the error bounds for the interpolating models, we firstly introduce a measure of poisedness of the interpolation set $\mathcal{Y}$. Such measure indicates how well $\mathcal{Y}$ "spans" the region where interpolation is of interest.

Definition 3.1. *Let $\Lambda > 0$ and $\mathcal{P}_n^d$ be a space of polynomials of degree less than or equal to d on $\mathbb{R}^n$, with a basis $\varphi = \{\varphi_0(x), \varphi_1(x), \ldots, \varphi_p(x)\}$. Then, a set $\mathcal{Y} = \{y^0, y^1, \ldots, y^p\}$ is said to be Λ-poised in $\mathcal{B}$ for $\mathcal{P}_n^d$ (in the interpolation sense) with respect to the basis φ if and only if for any $x \in \mathcal{B} \subset \mathbb{R}^n$ there exists $\lambda(x) \in \mathbb{R}^{p+1}$ such that:*

$$\sum_{i=0}^{p} \lambda_i(x)\varphi(y^i) = \varphi(x) \quad \textit{with} \quad \|\lambda(x)\|_\infty \leq \Lambda.$$

We also require the following assumption, which is readily achieved, for instance, by applying the procedures described in [6] once the interpolanting model is built with the Lagrange polynomials.

Assumption 3.2. *Assume we are given any set $\mathcal{Y} \subset \mathcal{B}(z, \Delta)$ with $n+1 \leq |\mathcal{Y}| \leq (n+1)(n+2)/2$ and $z \in \mathbb{R}^n$. Then we can apply a finite number of substitutions of the points in $\mathcal{Y}$, in fact, at most $|\mathcal{Y}|-1$, such that the new resultant set is Λ-poised in $B(z, \Delta)$ for a polynomial space $\mathcal{P}$, with dimension $|\mathcal{Y}|$ and $\mathcal{P}_n^1 \subseteq \mathcal{P} \subseteq \mathcal{P}_n^2$.*

The next lemma states the error bounds for at most fully quadratic models. As one might expect, the accuracy inherent in undetermined quadratic interpolation models is similar to the linear interpolation ones, having the error bounds being linear in Δ for the first derivatives and quadratic for the function values. The proof of the lemma can be found in [7]. Similar error bounds for the regression case are shown in [6].

Lemma 3.3. Given any $\Delta > 0$, $z \in \mathbb{R}^n$ and $\mathcal{Y} = \{y^0, y^1, \ldots, y^p\} \subset \mathcal{B}(y^0, \Delta)$ a Λ-poised set in $\mathcal{B}(y^0, \Delta)$ with $n+1 \leq |\mathcal{Y}| \leq (n+1)(n+2)/2$, let $p(\cdot) \in \mathcal{P}_n^2$ be a interpolating polynomial of f on $\mathcal{Y}$, i.e.

$$p(y^i) = f(y^i), \quad i = 1, \ldots, |\mathcal{Y}|.$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable in an open domain containing $\mathcal{B}(y^0, \Delta)$ and ∇f is Lipschitz continuous in Ω with constant L, then for any point $y \in \mathcal{B}(y^0, \Delta)$ we have

$$\begin{aligned} \|\nabla f(y) - \nabla p(y)\| &\leq \hat{\kappa}_{eg}(n, \Lambda, L)(\|\nabla^2 p\| + 1)\Delta, \\ |f(y) - p(y)| &\leq \hat{\kappa}_{ef}(n, \Lambda, L)(\|\nabla^2 p\| + 1)\Delta^2, \end{aligned}$$

where $\hat{\kappa}_{eg}$ and $\hat{\kappa}_{ef}$ are positive constants depending only on n, Λ and L.

3.1 Maintenance of the interpolation set

It is well known that derivative-based trust-regions methods usually decrease the trust region radius at unsuccessful iterations to converge. However, in derivative-free optimization, unsuccessful iterations in trust-region methods might be due to a bad quality of the interpolating model rather than a large trust region size. In order to maintain the quality of the surrogate models without resorting to costly model improvement steps at unsuccessful iterations, which might require another optimization problem to be globally solved, we apply a self-correcting geometry scheme proposed by Scheinberg and Toint [23] for the management of the geometry of the interpolation set. Such scheme is based upon the idea that unsuccessful trial points can still be useful as they may improve the geometry of the interpolation set. At such iterations, the trust region radii are reduced only if the algorithm fails to improve the geometry of the interpolation set by replacing one of its points by the trial point, thereby implying that there is no lack of poisedness in this case. Firstly, an interpolation point that is far from the current point x_k is sought to be replaced by the trial point x_k^+. If there is no such a far point, one tries to replace an interpolation point $y^{k,j}$ whose associated Lagrange polynomial value at x_k^+, $\ell_{k,j}(x_k^+)$, is bigger than a predefined threshold Λ. By doing that, one attempts to obtain a Λ-poised set $\mathcal{Y}_{k+1}$, or, since at most one interpolation point is replaced by iteration, improve its poisedness, at least. If no interpolation point is replaced, the trust regions are then shrunk.

A slight modification on the scheme is used on DEFT-FUNNEL. Following the idea proposed in [13], the trust region radii can also be reduced in cases where it is possible to improve poisedness. The number of reductions of Δ_k^f and Δ_k^c allowed in such cases, however, is limited by constants $\nu_f^{\max} > 0$ and $\nu_c^{\max} > 0$ predefined by the user as a means to prevent the trust regions to become too small.

3.2 Recursive call in subspaces

As pointed out before, by reducing the dimension of the problem we attempt to diminish the chances of a possible degeneration of the interpolation set when the sample points become too close to each other and thus affinely dependent. Such case might happen, for instance, as the optimal solution is approached.

In our algorithm, we apply the recursive approach found in the derivative-free method proposed by Gratton *et al.* [13] for bound-constrained optimization problems. Once the subspace $\mathcal{S}_k$ has been defined at iteration k, the algorithm calls itself recursively and the dimension of the problem is

then reduced to $\hat{n} = n - |\mathcal{L}_k \cup \mathcal{U}_k|$, where n denotes here the dimension of $\mathbb{R}^n$. A new well-poised interpolation set $\mathcal{Z}_k$ is then constructed from a suitable choice of points in $\mathcal{X}_k \cap \mathcal{S}_k$, where $\mathcal{X}_k$ is the set of all points obtained up to iteration k. In order to save function evaluations in the building process of the new interpolation set, all the points in $\mathcal{X}_k$ that are nearly but not in $\mathcal{S}_k$ are projected onto $\mathcal{S}_k$ and used to build $\mathcal{Z}_k$ with their model values instead of their real function values. In a more formal description, we define the set of the points that are close to (but not on) the active bounds at x_k as

$$\mathcal{A}_k \stackrel{\rm def}{=} \{y \in \mathcal{X}_k \mid 0 \leq |y_i - l_i^x| \leq \epsilon_b \ \text{ for } \ i \in \mathcal{L}_k \ \text{ and } \ 0 \leq |u_i^x - y_i| \leq \epsilon_b \ \text{ for } \ i \in \mathcal{U}_k\},$$

where, for at least one i, the strict inequality $0 < |y_i - l_i^x|$, $i \in \mathcal{L}_k$, or $0 < |u_i^x - y_i|$, $i \in \mathcal{U}_k$, must hold. We then project all the points $y \in \mathcal{A}_k$ onto $\mathcal{S}_k$, obtaining new "dummy" points y_s that are added to $\mathcal{X}_k$ with associated values $m_k^f(y_s)$ and $m_k^c(y_s)$ rather than the values of the original functions. These dummy points are progressively replaced by other points with true function values with high priority during the minimization in $\mathcal{S}_k$.

Convergence in a subspace is only declared if the interpolation set contains no dummy points. If a solution has been found for a subspace and there are still dummy points in the interpolation set, evaluations of the original functions $f(x)$ and $c(x)$ at such points are carried out and the interpolating models are recomputed from the original function values. Once convergence has been declared in a subspace $\mathcal{S}_k$, the $|\mathcal{L}_k \cup \mathcal{U}_k|$ fixed components x_i associated with the active bounds and the component x of the approximate solution found in $\mathcal{S}_k$ of dimension $\hat{n} = n - |\mathcal{L}_k \cup \mathcal{U}_k|$ are assembled to compose a full-dimensional vector $x_{\mathcal{S}}^*$ in $\mathbb{R}^n$. The algorithm then checks whether $(x_{\mathcal{S}}^*, s_{\mathcal{S}}^*)$ is optimal for the full-dimensional problem or not. Firstly, a full-space interpolation set of degree $n+1$ is built in an ϵ-neighborhood around the point $x_{\mathcal{S}}^*$. Subsequently, the corresponding interpolating models m_k^f and m_k^c are recomputed and the f-criticality measure π_{k-1}^f is calculated anew using information of the updated models. Finally, the criticality step in the full space is then entered.

4 The algorithm

We now provide a formal description of our complete algorithm for solving nonlinear optimization problems with general nonlinear constraints without using derivatives.

Algorithm 4.1: DEFT-FUNNEL($\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, (x,s), Δ^f, Δ^c, $v^{\max}$)

Step 0: Initialization. An initial accuracy threshold ϵ_0, an initial vector of multipliers μ_{-1} and positive initial trust-region radii Δ_0^f and Δ_0^c are given, as well as the constants

$$\alpha \in (0,1), \quad 0 < \gamma_1 < 1 < \gamma_2, \quad \zeta \geq 1, \quad 0 < \eta_1 < \eta_2 < 1 \quad \text{and} \quad \beta, \eta_3 > 0.$$

An initial set of interpolation points is also given, $\mathcal{Y}_0$, with $x_0 \in \mathcal{Y}_0 \subset \mathcal{B}(x_0, \Delta_0)$ and $|\mathcal{Y}_0| \geq n+1$, as well as the maximum number of interpolation points $p_{\max} \geq |\mathcal{Y}_0|$ in $\mathcal{Y}_k$ at the end. Let p_k denote the cardinality of $\mathcal{Y}_k$. This interpolation set defines interpolation models m_0^f and m_0^c around x_0 and associated Lagrange polynomials $\{l_{0,j}\}_{j=0}^p$. Define $\Delta_0 = \min[\Delta_0^f, \Delta_0^c] \leq \Delta^{\max}$, and $v_0^{\max} = \max[\kappa_{ca}, \kappa_{cr} v(x_0, s_0)]$ for some constants $\kappa_{ca} > 0$ and $\kappa_{cr} > 1$. Compute r_{-1} by solving (2.23) with normal step $n_{-1} = 0$ and define π_{-1}^f as in (2.24). Define $\nu_f^{\max} > 0$ and $\nu_c^{\max} > 0$, the maximum number of times that the tangential and normal trust regions sizes can be reduced when an interpolation point is replaced at unsuccessful iterations. Initialize the corresponding counters $\nu_f = \nu_c = 0$. Define $k = 0$ and $i = 0$.

Step 1: Subspace minimization. Check for (nearly) active bounds at x_k and define $\mathcal{S}_k$.

Step 1.1: If $\mathcal{A}(x_k) = \emptyset$ or if $\mathcal{S}_k$ has already been explored, go to Step 2.

Step 1.2: Project points in $\mathcal{X}_k$ which lie close to the (nearly) active bounds on $\mathcal{S}_k$ and associate with them suitable function values estimates.

Step 1.3: Build a new interpolation set $\mathcal{Z}_k$ in $\mathcal{S}_k$ including the projected points, if any.

Step 1.4: Call recursively DEFT-FUNNEL($\mathcal{S}_k$, $\mathcal{X}_k$, $\mathcal{Z}_k$, (x_k, s_k), Δ_k^f, Δ_k^c, $v_k^{\max}$) and let $(x_{\mathcal{S}}^*, s_{\mathcal{S}}^*)$ be the solution of the subspace problem after adding the fixed components.

Step 1.5: If $dim(\mathcal{S}_k) < n$ (where n denotes here the dimension of $\mathbb{R}^n$), return $(x_{\mathcal{S}}^*, s_{\mathcal{S}}^*)$. Otherwise, reset $(x_k, s_k) = (x_{\mathcal{S}}^*, s_{\mathcal{S}}^*)$, construct a new interpolation set $\mathcal{Y}_k$ around x_k, build the corresponding models m_k^f and m_k^c and recompute π_{k-1}^f using information of the new models.

Step 2: Criticality step. Define $\hat{m}_i^f = m_k^f$, $\hat{m}_i^c = m_k^c$ and $\hat{\pi}_i^f = \pi_{k-1}^f$.

Step 2.1: If $\|c(x_k, s_k)\| \leq \epsilon_i$ and $\hat{\pi}_i^f \leq \epsilon_i$, set $\epsilon_{i+1} = \max\left[\alpha\|c(x_k, s_k)\|, \alpha\hat{\pi}_i^f, \epsilon\right]$ and modify $\mathcal{Y}_k$ as needed to ensure it is Λ-poised in $\mathcal{B}(x_k, \epsilon_{i+1})$. If $\mathcal{Y}_k$ was modified, compute new models $\hat{m}_i^f$ and $\hat{m}_i^c$, calculate $\hat{r}_i$ and $\hat{\pi}_i^f$ associated to these models and increment i by one. If $\|c(x_k, s_k)\| \leq \epsilon$ and $\hat{\pi}_i^f \leq \epsilon$, return (x_k, s_k); otherwise, start Step 2.1 again;

Step 2.2: Set $m_k^f = \hat{m}_i^f$, $m_k^c = \hat{m}_i^c$, $\pi_{k-1}^f = \hat{\pi}_i^f$, $\Delta_k = \beta \max\left[\|c(x_k, s_k)\|, \pi_{k-1}^f\right]$ and define $v_i = x_k$ if a new model has been computed.

Step 3: Normal step. If $c(x_k, s_k) \neq 0$ and $\pi_k^v = 0$, STOP (infeasible stationary point). Otherwise, compute a normal step n_k by solving the problem (2.7). This computation must be performed if $k = 0$ or if (2.27) holds when $k > 0$. If n_k has not been computed, set $n_k = 0$.

Step 4: Tangent step. If (2.20) holds, then

Step 4.1: select a vector $\hat{\mu}_k$ satisfying (2.13) and define G_k as in (2.12) to obtain B_k;

Step 4.2: compute μ_k by solving (2.26);

Step 4.3: compute the modified Cauchy direction r_k by solving (2.23) and define π_k^f as (2.24);

Step 4.4: if (2.28) holds, compute a tangent step t_k by solving (2.21).

If (2.20) fails, set $\mu_k = \mu_{k-1}$. In this case, or if (2.28) fails, or if (2.31) holds but (2.32) fails, set $t_k = 0$ and $d_k = n_k$. In all cases, define $(x_k^+, s_k^+) = (x_k, s_k) + d_k$.

Step 5: Conclude a μ-iteration. If $n_k = t_k = 0$, then

Step 5.1: set $(x_{k+1}, s_{k+1}) = (x_k, s_k)$;

Step 5.2: define $\Delta_{k+1}^f = \Delta_k^f$ and $\Delta_{k+1}^c = \Delta_k^c$;

Step 5.3: set $v_{k+1}^{\max} = v_k^{\max}$, $\Delta_{k+1} = \min[\Delta_{k+1}^f, \Delta_{k+1}^c]$ and $\mathcal{Y}_{k+1} = \mathcal{Y}_k$.

Step 6: Conclude an f-iteration. If $t_k \neq 0$ and (2.32) and (2.35) hold,

Step 6.1: Augment the interpolation set. If $p_k < p_{\max}$, then define $\mathcal{Y}_{k+1} = \mathcal{Y}_k \cup \{x_k^+\}$.

- If $\rho_k^f \geq \eta_1$, set $(x_{k+1}, s_{k+1}) = (x_k^+, s_k^+)$ and $\nu_f = 0$.
  If $\rho_k^f \geq \eta_2$, set $\Delta_{k+1}^f = \min[\max[\gamma_2\|d_k\|, \Delta_k^f], \Delta^{\max}]$; otherwise, set $\Delta_{k+1}^f = \Delta_k^f$.
  If $v(x_k^+, s_k^+) < \eta_3\, v_k^{\max}$, set $\Delta_{k+1}^c = \min[\max[\gamma_2\|n_k\|, \Delta_k^c], \Delta^{\max}]$; otherwise, set $\Delta_{k+1}^c = \Delta_k^c$.

- If $\rho_k^f < \eta_1$, set $(x_{k+1}, s_{k+1}) = (x_k, s_k)$ and $\Delta_{k+1}^c = \Delta_k^c$.
  If $\nu_f \leq \nu_f^{\max}$, set $\Delta_{k+1}^f = \gamma_1 \Delta_k^f$ and $\nu_f = \nu_f + 1$; otherwise, set $\Delta_{k+1}^f = \Delta_k^f$.

Step 6.2: Successful iteration. If $p_k = p_{\max}$, $\rho_k^f \geq \eta_1$, set $(x_{k+1}, s_{k+1}) = (x_k^+, s_k^+)$ and define $\mathcal{Y}_{k+1} = \mathcal{Y}_k \setminus \{y^{k,r}\} \cup \{x_k^+\}$ for

$$y^{k,r} = \arg \max_{y^{k,j} \in \mathcal{Y}_k} \|y^{k,j} - x_k^+\|^2 |\ell_{k,j}(x_k^+)|. \tag{4.41}$$

Set $\nu_f = 0$. If $\rho_k^f \geq \eta_2$, set $\Delta_{k+1}^f = \min[\max[\gamma_2 \|d_k\|, \Delta_k^f], \Delta^{\max}]$; otherwise, set $\Delta_{k+1}^f = \Delta_k^f$.
If $v(x_k^+, s_k^+) < \eta_3 \, v_k^{\max}$, set $\Delta_{k+1}^c = \min[\max[\gamma_2 \|n_k\|, \Delta_k^c], \Delta^{\max}]$; otherwise, set $\Delta_{k+1}^c = \Delta_k^c$.

Step 6.3: Replace a far interpolation point. If $p_k = p_{\max}$, $\rho_k^f < \eta_1$, either $x_k \neq v_i$ or $\Delta_k \leq \epsilon_i$, and the set

$$\mathcal{F}_k \stackrel{\text{def}}{=} \{y^{k,j} \in \mathcal{Y}_k \text{ such that } \|y^{k,j} - x_k\| > \zeta \Delta \text{ and } \ell_{k,j}(x_k^+) \neq 0\} \tag{4.42}$$

is non-empty, then define $(x_{k+1}, s_{k+1}) = (x_k, s_k)$, and set $\Delta_{k+1}^f = \gamma_1 \|d_k\|$ if $\nu_f \leq \nu_f^{\max}$ or $\Delta_{k+1}^f = \Delta_k^f$ otherwise. Define $\mathcal{Y}_{k+1} = \mathcal{Y}_k \setminus \{y^{k,r}\} \cup \{x_k^+\}$, where

$$y^{k,r} = \arg \max_{y^{k,j} \in \mathcal{F}_k} \|y^{k,j} - x_k^+\|^2 |\ell_{k,j}(x_k^+)|. \tag{4.43}$$

If $\nu_f \leq \nu_f^{\max}$, update $\nu_f = \nu_f + 1$.

Step 6.4: Replace a close interpolation point. If $p_k = p_{\max}$, $\rho_k^f < \eta_1$, either $x_k \neq v_i$ or $\Delta_k \leq \epsilon_i$, the set $\mathcal{F}_k$ is empty, and the set

$$\mathcal{C}_k \stackrel{\text{def}}{=} \{y^{k,j} \in \mathcal{Y}_k \text{ such that } \|y^{k,j} - x_k\| \leq \zeta \Delta \text{ and } \ell_{k,j}(x_k^+) > \lambda\} \tag{4.44}$$

is non-empty, then define $(x_{k+1}, s_{k+1}) = (x_k, s_k)$ and set $\Delta_{k+1}^f = \gamma_1 \|d_k\|$ if $\nu_f \leq \nu_f^{\max}$ or $\Delta_{k+1}^f = \Delta_k^f$ otherwise. Define $\mathcal{Y}_{k+1} = \mathcal{Y}_k \setminus \{y^{k,r}\} \cup \{x_k^+\}$, where

$$y^{k,r} = \arg \max_{y^{k,j} \in \mathcal{C}_k} \|y^{k,j} - x_k^+\|^2 |\ell_{k,j}(x_k^+)|. \tag{4.45}$$

If $\nu_f \leq \nu_f^{\max}$, update $\nu_f = \nu_f + 1$.

Step 6.5: Reduce the trust-region radius. If $p_k = p_{\max}$, $\rho_k^f < \eta_1$ and either $x_k = v_i$ and $\Delta_k^f > \epsilon_i$ or $\mathcal{F}_k \cup \mathcal{C}_k = \emptyset$, then set $(x_{k+1}, s_{k+1}) = (x_k, s_k)$, $\Delta_{k+1}^c = \Delta_k^c$, $\Delta_{k+1}^f = \gamma_1 \|d_k\|$ and $\mathcal{Y}_{k+1} = \mathcal{Y}_k$.

Step 6.6: Update the combined radius. Set $\Delta_{k+1} = \min[\Delta_{k+1}^f, \Delta_{k+1}^c]$ and $v_{k+1}^{\max} = v_k^{\max}$.

Step 7: Conclude a c-iteration. If either $n_k \neq 0$ and $t_k = 0$, or either one of (2.32) or (2.35) fails,

Step 7.1: Augment the interpolation set. If $p_k < p_{\max}$, then define $\mathcal{Y}_{k+1} = \mathcal{Y}_k \cup \{x_k^+\}$.

- If $\rho_k^c \geq \eta_1$, set $(x_{k+1}, s_{k+1}) = (x_k^+, s_k^+)$, $\Delta_{k+1}^f = \Delta_k^f$ and $\nu_c = 0$.
  If $\rho_k^c \geq \eta_2$, set $\Delta_{k+1}^c = \min[\max[\gamma_2 \|n_k\|, \Delta_k^c], \Delta^{\max}]$; otherwise, set $\Delta_{k+1}^c = \Delta_k^c$.
- If $\rho_k^c < \eta_1$, set $(x_{k+1}, s_{k+1}) = (x_k, s_k)$ and $\Delta_{k+1}^f = \Delta_k^f$.
  If $\nu_c \leq \nu_c^{\max}$, then, set $\Delta_{k+1}^c = \gamma_1 \|n_k\|$ if $\|n_k\| \neq 0$, and $\Delta_{k+1}^c = \gamma_1 \Delta_k^c$, otherwise ($\|n_k\| = 0$). Update $\nu_c = \nu_c + 1$. If $\nu_c > \nu_c^{\max}$, set $\Delta_{k+1}^c = \Delta_k^c$.

Step 7.2: Successful iteration. If $p_k = p_{\max}$, (2.38) holds, set $(x_{k+1}, s_{k+1}) = (x_k^+, s_k^+)$ and define $\mathcal{Y}_{k+1} = \mathcal{Y}_k \setminus \{y^{k,r}\} \cup \{x_k^+\}$ for

$$y^{k,r} = \arg \max_{y^{k,j} \in \mathcal{Y}_k} \|y^{k,j} - x_k^+\|^2 |\ell_{k,j}(x_k^+)|. \tag{4.46}$$

Set $\Delta_{k+1}^f = \Delta_k^f$ and $\nu_c = 0$. Set $\Delta_{k+1}^c = \min[\max[\gamma_2 \|n_k\|, \Delta_k^c], \Delta^{\max}]$ if $\rho_k^c \geq \eta_2$ or $\Delta_{k+1}^c = \Delta_k^c$ otherwise.

Step 7.3: Replace a far interpolation point. If $p_k = p_{\max}$, (2.38) fails, either $x_k \neq v_i$ or $\Delta_k \leq \epsilon_i$, and the set

$$\mathcal{F}_k \stackrel{\text{def}}{=} \{y^{k,j} \in \mathcal{Y}_k \text{ such that } \|y^{k,j} - x_k \mid > \zeta\Delta \text{ and } \ell_{k,j}(x_k^+) \neq 0\} \tag{4.47}$$

is non-empty, then define $(x_{k+1}, s_{k+1}) = (x_k, s_k)$ and set $\Delta_{k+1}^f = \Delta_k^f$. If $\nu_c \leq \nu_c^{\max}$, then set $\Delta_{k+1}^c = \gamma_1 \|n_k\|$ if $\|n_k\| \neq 0$, or $\Delta_{k+1}^c = \gamma_1 \Delta_k^c$ otherwise ($\|n_k\| = 0$). If $\nu_c > \nu_c^{\max}$, set $\Delta_{k+1}^c = \Delta_k^c$. Define $\mathcal{Y}_{k+1} = \mathcal{Y}_k \setminus \{y^{k,r}\} \cup \{x_k^+\}$, where

$$y^{k,r} = \arg \max_{y^{k,j} \in \mathcal{F}_k} \|y^{k,j} - x_k^+\|^2 |\ell_{k,j}(x_k^+)|. \tag{4.48}$$

If $\nu_c \leq \nu_c^{\max}$, update $\nu_c = \nu_c + 1$.

Step 7.4: Replace a close interpolation point. If $p_k = p_{\max}$, (2.38) fails, either $x_k \neq v_i$ or $\Delta_k \leq \epsilon_i$, the set $\mathcal{F}_k$ is empty, and the set

$$\mathcal{C}_k \stackrel{\text{def}}{=} \{y^{k,j} \in \mathcal{Y}_k \text{ such that } \|y^{k,j} - x_k\| \leq \zeta\Delta \text{ and } \ell_{k,j}(x_k^+) > \lambda\} \tag{4.49}$$

is non-empty, then set $(x_{k+1}, s_{k+1}) = (x_k, s_k)$ and $\Delta_{k+1}^f = \Delta_k^f$. If $\nu_c \leq \nu_c^{\max}$, then set $\Delta_{k+1}^c = \gamma_1 \|n_k\|$ if $\|n_k\| \neq 0$, or $\Delta_{k+1}^c = \gamma_1 \Delta_k^c$ otherwise ($\|n_k\| = 0$). If $\nu_c > \nu_c^{\max}$, set $\Delta_{k+1}^c = \Delta_k^c$. Define $\mathcal{Y}_{k+1} = \mathcal{Y}_k \setminus \{y^{k,r}\} \cup \{x_k^+\}$, where

$$y^{k,r} = \arg \max_{y^{k,j} \in \mathcal{C}_k} \|y^{k,j} - x_k^+\|^2 |\ell_{k,j}(x_k^+)|. \tag{4.50}$$

If $\nu_c \leq \nu_c^{\max}$, update $\nu_c = \nu_c + 1$.

Step 7.5: Reduce the trust-region radius. If $p_k = p_{\max}$, (2.38) fails and either $x_k = v_i$ and $\Delta_k^c > \epsilon_i$ or $\mathcal{F}_k \cup \mathcal{C}_k = \emptyset$, then set $(x_{k+1}, s_{k+1}) = (x_k, s_k)$ and $\Delta_{k+1}^f = \Delta_k^f$. If $\|n_k\| \neq 0$, set $\Delta_{k+1}^c = \gamma_1 \|n_k\|$, otherwise set $\Delta_{k+1}^c = \gamma_1 \Delta_k^c$. Define $\mathcal{Y}_{k+1} = \mathcal{Y}_k$.

Step 7.6: Update the combined radius and the maximal infeasibility. Set $\Delta_{k+1} = \min[\Delta_{k+1}^f, \Delta_{k+1}^c]$ and update $v_k^{\max}$ using (2.39).

Step 8: Update the models and the Lagrange polynomials. If $\mathcal{Y}_{k+1} \neq \mathcal{Y}_k$, compute the interpolation models m_{k+1}^f and m_{k+1}^c around x_{k+1} using $\mathcal{Y}_{k+1}$ and the associated Lagrange polynomials $\{l_{k+1,j}\}_{j=0}^p$. Set $\Delta_{k+1} = \min[\Delta_{k+1}^f, \Delta_{k+1}^c]$, increment k by one and go to Step 1.

5 Numerical experiments

We implemented DEFT-FUNNEL in Matlab and tested it firstly on a set of 80 small-scale constrained problems from the CUTEst [10] collection. The problems contain at least one equality or inequality constraint, many of them containing both types and some containing simple bounds as well. 63 of the problems in our experiments are in the range of the first 113 test examples from the Hock-Schittowski collection [14], while the remaining are nonlinearly constrained optimization problems found in [3].

For the calculation of the normal step and the approximate Lagrange multipliers, we used a Matlab code named BLLS developed in collaboration with Anke Tröltzsch for solving bound-constrained linear least-squares problems. This method is intended for small-dimensional problems and is an active-set algorithm where the unconstrained problem is solved at each iteration in the subspace defined by the currently active bounds, which are determined by a projected Cauchy step. As for the computation of the tangent step, we used a nonmonotone spectral projected gradient method [2] to solve the (possibly) indefinite quadratic subproblems (2.21).

We compare the results of the four variants of our method with regard to the way of building the interpolating models — subbasis selection, mininum Frobenius norm, minimum ℓ_2-norm and regression — to those obtained with the popular Fortran code COBYLA [19], a trust-region method for constrained problems that models the objective and constraint functions by linear interpolation. The only criterion for comparison is the number of calls on the routine that evaluates the objective function and the constraints at the same time at the required points, which is motivated by the fact that these costs very often dominate those of all other algorithmic computations. Therefore, we do not count evaluations of the objective function and the constraints separately, but rather the number of calls to the single routine that calculates both simultaneously. Comparisons of the computational results obtained by DEFT-FUNNEL for equality-constrained problems only also have been made with another method named CDFO, a derivative-free filter-SQP algorithm proposed in [4]. The results are reported in [22] and show that DEFT-FUNNEL was superior to CDFO in our tests.

The threshold ϵ for declaring convergence in the criticality step in DEFT-FUNNEL is set to $\epsilon = 10^{-4}$. The stopping criterion used in COBYLA is based on the trust region radius ρ from the interval $[\rho_{end}, \rho_{beg}]$, where ρ_{beg} and ρ_{end} are constants predefined by the user. The parameter ρ is decreased by a constant factor during the execution of the algorithm and is never increased. The algorithm stops when $\rho = \rho_{end}$. Therefore, ρ_{end} should have the magnitude of the required accuracy in the final values of the variables. In our experiments, we set $\rho_{end} = 10^{-4}$.

In the DEFT-FUNNEL method, we fixed the trust-region parameters to $\Delta_0 = 1$, $\eta_1 = 0.0001$, $\eta_2 = 0.9$, $\eta_3 = 0.5$, $\gamma_1 = 0.5$, $\gamma_2 = 2.5$ and $\Delta^{\max} = 10^{10}$. The parameter ζ used in the definition of the sets $\mathcal{F}_k$ and $\mathcal{C}_k$ of far points and close points, respectively, is set to $\zeta = 1$. For the limit number of times to reduce the trust regions sizes when a far or close interpolation point is replaced at unsuccessful iterations, we choose $\nu_f^{\max} = \nu_c^{\max} = 10$. We set $p_{\max} = (n+1)(n+2)/2$ for the subbasis, minimum Frobenius norm and minimum ℓ_2-norm approaches, and $p_{\max} = (n+1)(n+2)$ for the regression case. Finally, we set $\epsilon_0 = 0.01$ as the initial value for the loop in the criticality step, $\alpha = 0.1$ and $\beta = 1$.

The performance of the methods are compared by means of their *performance profiles* [9]. Let n_s denote the number of solvers, and n_p, the number of problems. Denote the set of solvers by $\mathcal{S}$, and the set of problems by $\mathcal{P}$. We compare the performance on problem $p \in \mathcal{P}$ by solver $s \in \mathcal{S}$ with the best performance by any solver on this problem; that is, we use the *performance ratio*

$$r_{p,s} \stackrel{\text{def}}{=} \frac{t_{p,s}}{\min\{t_{p,s} : s \in \mathcal{S}\}},$$

where $t_{p,s}$ is the number of function evaluations required to solve problem p by solver s. If we define

$$\rho_s(\tau) \stackrel{\text{def}}{=} \frac{1}{n_p}\text{size}\{p \in \mathcal{P} : r_{p,s} \leq \tau\},$$

then $\rho_s(\tau)$ is an approximation to the probability for solver $s \in \mathcal{S}$ that a performance ratio $r_{p,s}$ is within a factor $\tau \in \mathbb{R}$ of the best possible ratio.

In Figure 5.1, we provide the performance profile of the four variants of DEFT-FUNNEL, whereas, in Figure 5.2, we compare the performance of each variant of our method to COBYLA's individually. As it can be seen, DEFT-FUNNEL has shown superior results on the set of test problems when the interpolating models are built from subbasis selection, minimum Frobenius norm and minimum ℓ_2-norm approaches. For the regression variant, Figure 5.2d reveals that COBYLA was faster than DEFT-FUNNEL in most of the problems, although the latter was able to solve more problems than the former for large values of τ.

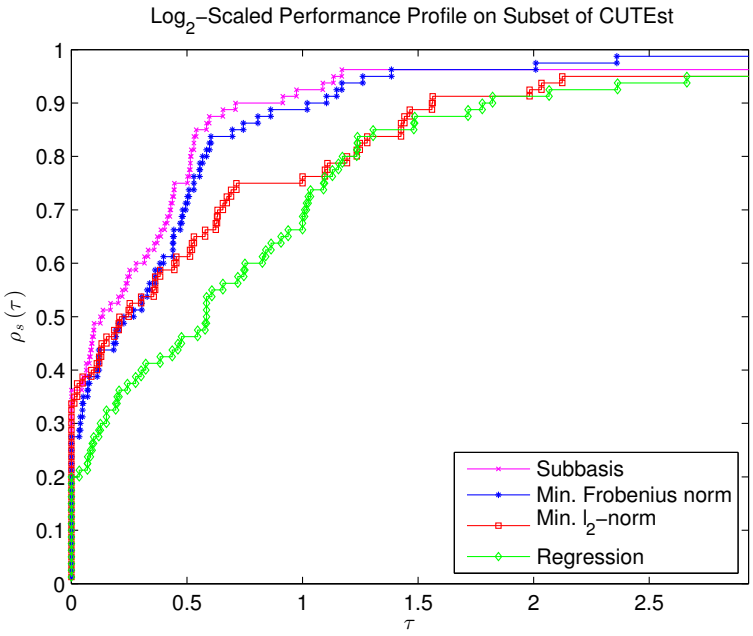
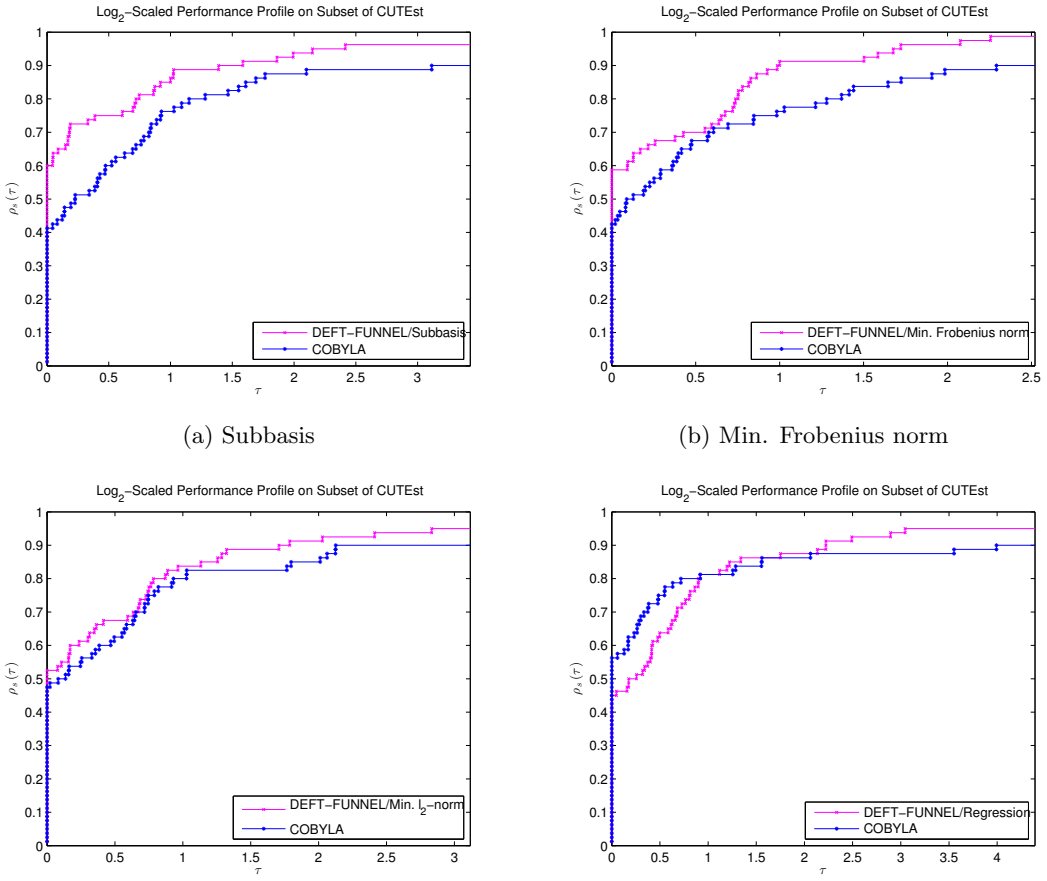


Figure 5.1: Log_2-scaled performance profile of DEFT-FUNNEL with different approaches to build the models on a set of 80 problems from CUTEst.

We also tested DEFT-FUNNEL on a set of 20 small-scale linearly constrained optimization problems from CUTEst and compare its results with those of the software LINCOA, a newly developed trust-region method by Powell [21] for linearly constrained optimization without derivatives. His software combines active-set methods with truncated conjugate gradients and use quadratic interpolation-based surrogate models for the objective function. The user is required to provide the Jacobian matrix and the vector of constants in the right side of the constraints as inputs, as well as a feasible starting point. Since many of CUTEst problems provide an infeasible starting point, we chose a feasible one instead for the sake of comparison. The stopping criterion used in LINCOA is the same found in COBYLA, i.e., the algorithm stops when the trust region radius $\rho \in [\rho_{end}, \rho_{beg}]$ reaches the smallest value allowed ρ_{end}. In our experiments, we set $\rho_{end} = 10^{-4}$. Regarding the parameter setting in DEFT-FUNNEL, the same values were kept for comparison with LINCOA.

In Figure 5.3, we provide the performance profile of the four variants of DEFT-FUNNEL for the 20 linearly constrained problems, whereas, in Figure 5.4, we compare the performance of each variant of



(a) Subbasis

(b) Min. Frobenius norm

(c) Min. ℓ_2-norm

(d) Regression

Figure 5.2: *Log*$_2$-scaled performance profile of the methods DEFT-FUNNEL (with different approaches to build the models) and COBYLA on a set of 80 problems from CUTEst.

our method to LINCOA's individually. The results reported in Figures 5.4a and 5.4b reveal a superior performance of LINCOA over DEFT-FUNNEL/Subbasis and DEFT-FUNNEL/Frobenius. On the other hand, Figure 5.4c shows that DEFT-FUNNEL/Min. ℓ_2-norm was faster than LINCOA and, in Figure 5.4d, it can be seen that DEFT-FUNNEL/Regression also was slightly faster, although the latter presented superior performance for large values of τ, which indicates more robustness.

We conclude this section by noticing that the performance profiles in the Figures 5.2 and 5.4 give clear indication that DEFT-FUNNEL provides encouraging results for small-scale nonlinear optimization problems with general nonlinear constraints. For the case where the user knows that the constraint functions are linear and he is able to provide their gradients, it might be interesting to the handle those constraints separately rather than lumping them and other general nonlinear constraints together in $g(x)$.

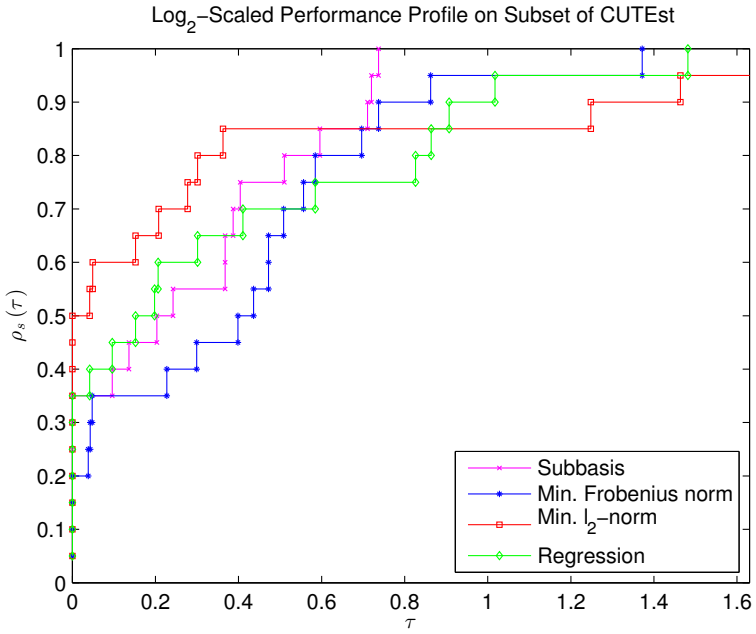
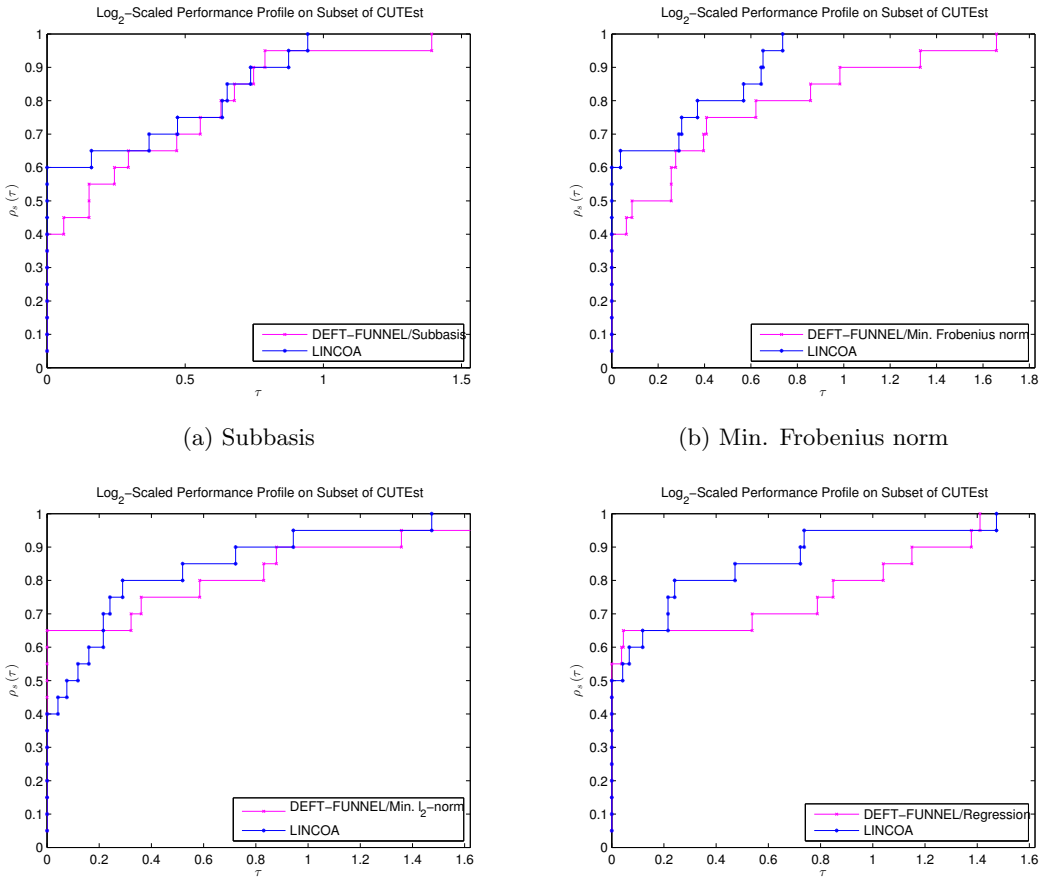


Figure 5.3: *Log*$_2$-scaled performance profile of DEFT-FUNNEL with different approaches to build the models on a set of 20 linearly constrained problems from CUTEst.



(a) Subbasis

(b) Min. Frobenius norm

(c) Min. ℓ_2-norm

(d) Regression

Figure 5.4: Log_2-scaled performance profile of the methods DEFT-FUNNEL (with different approaches to build the models) and COBYLA on a set of 20 linearly constrained problems from CUTEst.

6 Conclusions

We have presented a SQP algorithm for solving nonlinear optimization problems with general nonlinear constraints. It extends the one introduced by Gould and Toint [12] for problems with equality constraints only. When simple bounds are given, the algorithm makes use of an active-set approach to perform minimization on the subspaces defined by the active bounds.

We also applied our algorithm to the field of derivative-free optimization by employing an ensemble of underlying techniques such as multivariate polynomial interpolation for the construction of surrogate models, a self-correcting geometry mechanism for the maintenance of interpolation set and the reduction of the dimension of the problem when entering the subspaces by using a recursive subspace minimization approach. The final method, called DEFT-FUNNEL, makes use of neither the derivatives of the objective function nor the derivatives of the constraints. It also considers both the objective and constraints as black-box functions. Numerical experiments on a set of 80 small-scale nonlinear optimization problems with general nonlinear constraints and on a set of 20 small-scale linearly constrained optimization problems were performed and revealed good results of the new algorithm in practice.

For future research, we plan to consider the possibility of having two different interpolations sets for the objective function and the constraints and analyze its performance for cases where the constraints are linear, while the objective function is of higher degree. Convergence results for both derivative-full and -free versions of the algorithm are also of interest and are left as a development line for future work.

Acknowledgements

The first author gratefully acknowledges a CERUNA-UNamur scholarship.

References

[1] C. Audet and J. E. Dennis. A pattern search filter method for nonlinear programming without derivatives. *SIAM Journal on Optimization*, 14(4):980–1010, 2004.

[2] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211, 2000.

[3] P. T. Boggs and J. W. Tolle. A strategy for global convergence in a sequential quadratic programming algorithm. *SIAM Journal on Numerical Analysis*, 26(3):600–623, 1989.

[4] B. Colson. *Trust-Region Algorithms for Derivative-Free Optimization and Nonlinear Bilevel Programming.* PhD thesis, Department of Mathematics, FUNDP - University of Namur, Namur, Belgium, 2004.

[5] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods.* MOS-SIAM Series on Optimization, Philadelphia, 2000.

[6] A. R. Conn, K. Scheinberg, and L. N. Vincente. *Introduction to derivative-free optimization.* MPS-SIAM Book Series on Optimization, Philadelphia, 2009.

[7] A. R. Conn, K. Scheinberg, and H. Zang. A derivative-free algorithm for least-squares minimization. *SIAM Journal on Optimization*, 20(6):3555–3576, 2010.

[8] F. Curtis, N. I. M. Gould, D. Robinson, and Ph. L. Toint. An interior-point trust-funnel algorithm for nonlinear optimization. Technical Report naXys-02-2014, Department of Mathematics, University of Namur, 2014.

[9] E. Dolan and J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–13, 2002.

[10] N. I. M. Gould, D. Orban, and Ph. L. Toint. Cutest: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Computational Optimization and Applications*, pages 1–13, 2014.

[11] N. I. M. Gould, D. P. Robinson, and Ph. L. Toint. Corrigendum: Nonlinear programming without a penalty function or a filter. Technical report, Namur Centre for Complex Systems (naXys), University of Namur, Namur, Belgium, 2011.

[12] N. I. M. Gould and Ph. L. Toint. Nonlinear programming without a penalty function or a filter. *Mathematical Programming*, 122(1):155–196, 2010.

[13] S. Gratton, Ph. L. Toint, and A. Tröltzsch. An active-set trust-region method for bound-constrained nonlinear optimization without derivatives. *Optimization Methods and Software*, 26(4-5):875–896, 2011.

[14] W. Hock and K. Schittkowski. Test example for nonlinear programming codes. *Journal of Optimization Theory and Applications*, 30(1):127–129, 1980.

[15] R. M. Lewis and V. Torczon. Pattern search algorithms for linearly constrained minimization. *SIAM Journal on Optimization*, 10:917–941, 2000.

[16] R. M. Lewis and V. Torczon. A globally convergent augmented langragian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Optimization*, 12(4):1075–1089, 2002.

[17] R. M. Lewis and V. Torczon. Active set identification for linearly constrained minimization without explicit derivatives. *SIAM Journal on Optimization*, 20(3):1378–1405, 2009.

[18] S. Lucidi, M. Sciandrone, and P. Tseng. Objective-derivative-free methods for constrained optimization. *Mathematical Programming, Series A*, 92:37–59, 2002.

[19] M. J. D. Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in Optimization and Numerical Analysis, Proceedings of the Sixth Workshop on Optimization and Numerical Analysis, Oaxaca, Mexico*, volume 275, pages 51–67, Dordrecht, The Netherlands, 1994. Kluwer Academic Publishers.

[20] M. J. D. Powell. Direct search algorithms for optimization calculations. *Acta Numerica*, 7:287–336, 1998.

[21] M. J. D. Powell. On fast trust region methods for quadratic models with linear constraints. Technical Report DAMTP 2014/NA02, Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, England, 2014.

[22] Ph.R. Sampaio and Ph.L. Toint. A derivative-free trust-funnel method for equality-constrained nonlinear optimization. *Computational Optimization and Applications*, pages 1–25, 2014.

[23] K. Scheinberg and Ph. L. Toint. Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization. *SIAM Journal on Optimization*, 20(6):3512–3532, 2010.

[24] W.-C. Yu and Y.-X. Li. A direct search method by the local positive basis for the nonlinearly constrained optimization. *Chinese Annals of Mathematics*, 2:269–280, 1981.