Convergence properties of minimization algorithms for convex constraints using a structured trust region

by A.R. Conn[1], Nick Gould[2], and Ph.L. Toint[3]

RAL 92-069

November 16, 1992

Abstract. We present in this paper a class of trust region algorithms in which the structure of the problem is explicitly used in the very definition of the trust region itself. This development is intended to reflect the possibility that some parts of the problem may be more "trusted" than others, a commonly occurring situation in large-scale nonlinear applications. After describing the structured trust region mechanism, we prove global convergence for all algorithms in our class. We also prove that, when convex constraints are present, the correct set of such constraints active at the problem's solution is identified by these algorithms after a finite number of iterations.

[1] Mathematical Sciences Department,
IBM T.J. Watson Research Center,
PO Box 218, Yorktown Heights, NY 10598, USA
[2] Central Computing Department,
Rutherford Appleton Laboratory,
Chilton, Oxfordshire, OX11 0QX, England
[3] Department of Mathematics,
Facultés Universitaires ND de la Paix,
B-5000 Namur, Belgium

Keywords : Trust region methods, structured problems, large-scale optimization, partial separability, convex constraints.

Mathematics Subject Classifications : 65K05, 90C30

Convergence properties of minimization algorithms for convex constraints using a structured trust region

A. R. Conn, Nick Gould and Ph. L. Toint

November 16, 1992

Abstract

We present in this paper a class of trust region algorithms in which the structure of the problem is explicitly used in the very definition of the trust region itself. This development is intended to reflect the possibility that some parts of the problem may be more "trusted" than others, a commonly occurring situation in large-scale nonlinear applications. After describing the structured trust region mechanism, we prove global convergence for all algorithms in our class. We also prove that, when convex constraints are present, the correct set of such constraints active at the problem's solution is identified by these algorithms after a finite number of iterations.

1 Introduction

Trust region algorithms have enjoyed a long and successful history as tools for the solution of nonlinear, nonconvex, optimization problems. They have been studied and applied to unconstrained problems (see [6], [16], [24], [28], [29], [30], [31], [33], [34], [38]) and to problems involving various classes of constraints: simple bounds ([5], [10], [11], [27], [32]), convex constraints ([1], [2], [9], [41]), and also nonconvex ones ([4], [7], [15], [35], [42]). This long lasting interest is probably justified by the attractive combination of a solid convergence theory, a noted algorithmic robustness, the existence of numerically efficient implementations and an intuitively appealing justification. The main idea behind trust region algorithms is that, if a nonlinear function (objective and/or constraints) is expensive to compute or difficult to handle explicitly, one can replace it by a suitable *model.* This model may be "trusted" within a certain *trust region* around the current point, whose size (the trust region *radius*) is then expanded if the model and function sufficiently agree, or decreased if they differ too much. The minimization then proceeds by replacing the difficult nonlinear function(s) with the corresponding easier model(s).

It is remarkable that, up to now, all algorithms that we are aware of use a *single* trust region radius to measure the degree of trustworthiness of the models employed, even if several different functions are involved. This choice is somewhat surprising if one admits that some of the modelled functions could be substantially "better behaved" than others in the same problem, which means that the region in which their models can be trusted might also be substantially larger. In this context, the single trust region choice might be viewed as a conservative strategy ensuring that *all* models may be trusted in what amounts to a "safe minimal" region. This might be reasonable for small problems where each involved function depends on all the problem's variables, but the strategy becomes clearly questionable for large-scale applications, where each of the problem's function typically depends only on a small number of variables. For instance, one might consider the minimization of an unconstrained objective function consisting of many quadratic terms and just a few very nonlinear parts involving a small subset of the variables. If a classical single trust region algorithm with quadratic model is used, the quadratic terms are perfectly modelled, but the steps that one can make are (unnecessarily) limited by the very nonlinear behaviour of a small subset of the variables!

It is the purpose of this paper to present and analyze a class of algorithms that use the problem's *structure* in the very definition of the trust region, allowing large steps in directions in which the model has proved to be adequate while restricting the movement in directions where the model seems unreliable. To be more precise, we will consider the problem of minimizing a *partially separable* objective function subject to convex constraints; we will then use the decomposition of the objective function into element functions as the basis for our structured trust region definition. The choice of the partially separable structure, a concept introduced in [20], is motivated by the very general geometric nature of this structure and by the increasing recognition of its practical use (see [3], [8], [12], [13], [17], [18], [19], [21], [26], [39], amongst others). Furthermore, partial separability already provides a decomposition of the considered nonlinear function into a linear combination of smaller *element functions* which can then be modelled separately (see [40]). It is then quite natural to assign one trust region radius per element functions and to decide on their increase or decrease separately. Because different element functions typically involve different sets of variables, each *elemental trust region* only restricts the components of the step corresponding to its elemental variables.

A first approach to this idea could use the freedom left in the scaling matrices present in the available theory ([41], for instance), reflecting the difference in model adequacy between elements in the scaling (shape) of the trust region. This would be satisfactory if the theory did not require that the scaling matrices be of uniformly bounded condition number. In fact, this last condition prevents the trust region radius of well-modelled elements (linear, for instance) from increasing to infinity while other radii, corresponding to more nonlinear element functions, remain bounded. Furthermore, this strategy would probably cause numerical difficulties: we will indeed see below that additional algorithmic safeguards may be important for simultaneously handling trust regions of vastly different sizes. Hence, we do not pursue this first approach any further.

Section 2 of the paper presents the problem in more detail and the new class of algorithms using the principle of structured trust regions. Global convergence for all algorithms in the class is proved in Section 3. We discuss the identification of active constraints in Section 4. We finally give some comments and perspectives in Section 5.

2 Structured trust region for partially separable problems

2.1 A structured model of the objective and the corresponding structured trust region

2.1.1 The problem

The problem we consider is that of minimizing a smooth objective function subject to convex constraints, i.e. we wish to solve the problem

$$\min_{x \in X} f(x) \tag{2.1}$$

where X is a closed convex subset of $\mathbf{R}\ n$. We will denote by $\langle \cdot, \cdot \rangle$ the Euclidean inner product on $\mathbf{R}\ n$ and by $\| \cdot \|_2$ the associated ℓ_2-norm. Given Y a closed convex subset of $\mathbf{R}\ n$, we also define the operator $P_Y(\cdot)$ to be the orthogonal projection onto Y. We now list our further assumptions on (2.1).

AS.1 X has a non-empty interior.

AS.2 f is bounded below on X.

AS.3 f is partially separable, which means that

$$f(x) = \sum_{i=1}^{p} f_i(x) \tag{2.2}$$

and that, for each $i \in \{1, \ldots, p\}$, there exists a subspace $\mathcal{N}_i \neq \{0\}$ such that, for all $w \in \mathcal{N}_i$ and all $x \in X$,

$$f_i(x + w) = f_i(x). \tag{2.3}$$

AS.4 For each $i \in \{1, \ldots, p\}$, f_i is continuously differentiable in an open set containing X and its gradient is uniformly bounded on X.

Note that we admit the case where X is unbounded or even identical to $\mathbf{R}\ n$ itself, in which case we obtain an unconstrained problem. In relation to the partial separability of the objective function, we also consider the *range subspace* (see [22]) associated with each element function f_i, which is defined as

$$\mathcal{R}_i \stackrel{\text{def}}{=} \mathcal{N}_i^{\perp}. \tag{2.4}$$

We are mostly interested in the case where the dimension of each $\mathcal{R}_i$ is small compared to n. A commonly occurring case is when each element function f_i only depends on a small subset of the problem's variables: $\mathcal{R}_i$ is then the subspace spanned by the vectors of the canonical basis corresponding to the variables that occur in f_i (the *elemental variables*). For each $x \in \mathbf{R}\ n$, the range of the projection operator $P_{\mathcal{R}_i}(x)$ is therefore of low dimensionality. The reader is referred to [12] for a more detailed introduction to partially separable functions.

2.1.2 The element models

The algorithm we have in mind is iterative and we will associate, at iteration k, a model $m_{i,k}$ with each element function f_i. This model, defined on $\mathcal{R}_i$ in a neighbourhood of the projection of the k-th iterate x_k on this subspace, is meant to approximate f_i for all x in the *element trust region*

$$B_{i,k} \stackrel{\rm def}{=} \{x \in X \mid \|P_{\mathcal{R}_i}(x - x_k)\|_{(i,k)} \leq \nu_1 \Delta_{i,k}\}, \tag{2.5}$$

where ν_1 is a positive constant[1], $\Delta_{i,k}$ is the i-th trust region radius at iteration k and the norm $\|\cdot\|_{(i,k)}$ is a norm defined on the range subspace $\mathcal{R}_i$ and associated with iteration k. In what follows, we will slightly abuse notation by writing $m_{i,k}(x)$ for an $x \in \mathbf{R}\ n$, instead of the more complete $m_{i,k}(P_{\mathcal{R}_i}(x))$. We will furthermore assume that each model $m_{i,k}$ $(i \in \{1, \ldots, p\}, k = 0, 1, 2, \ldots)$ is differentiable and has Lipschitz continuous first derivatives on an open set containing $B_{i,k}$, and that

$$m_{i,k}(x_k) = f_i(x_k) \quad (i \in \{1, \ldots, p\}, k = 0, 1, 2, \ldots). \tag{2.6}$$

Moreover, we assume that $g_{i,k} \stackrel{\rm def}{=} \nabla m_{i,k}(x_k) \in \mathcal{R}_i$ approximates $\nabla f_i(x_k) \in \mathcal{R}_i$ in the sense that $g_{i,k} - \nabla f_i(x_k) \stackrel{\rm def}{=} e_{i,k}$, where, for all $i \in \{1, \ldots, p\}$ and all k,

$$\|e_{i,k}\|_{[i,k]} \leq \kappa_1 \Delta_{min,k} \tag{2.7}$$

for a non-negative constant $\kappa_1 > 0$ and $\Delta_{min,k}$ is defined by

$$\Delta_{min,k} \stackrel{\rm def}{=} \min_{i \in \{1,\ldots,p\}} \Delta_{i,k}. \tag{2.8}$$

The norm $\|\cdot\|_{[i,k]}$ is any norm that satisfies

$$|\langle x, y \rangle| \leq \|x\|_{(i,k)} \|y\|_{[i,k]} \tag{2.9}$$

for all $x, y \in \mathbf{R}\ n$. In particular, one can choose the *dual norm* of $\|\cdot\|_{(i,k)}$ defined by

$$\|y\|_{[i,k]} \stackrel{\rm def}{=} \sup_{x \neq 0} \frac{|\langle x, y \rangle|}{\|x\|_{(i,k)}}. \tag{2.10}$$

Condition (2.7) is quite weak, as it merely requires that the first order information on the considered element function be reasonably accurate whenever a short step must be taken. Indeed, one expects this first order behaviour to dominate for small steps. Further arguments supporting a choice similar to (2.7) for problems with convex constraints are presented in [9].

Amongst the most commonly used element models, linear or quadratic approximations are pre-eminent. One can, for instance, consider the quadratic model given by the first two terms of the element function Taylor series around the current iterate. Another popular choice is a quadratic model where the second derivative matrix is recurred using quasi-Newton formulae.

[1] We could use positive constants $\nu_i > 0$ depending on the element, but we will restrict ourselves to the case where they are all equal for the sake of simplicity.

2.1.3 Consistent norms

With iteration k, we also associate an overall norm $\|\cdot\|_{(k)}$ defined on the whole of $\mathbf{R}\ n$, whose purpose is to reflect the relative weighting of the different elemental norms $\|\cdot\|_{(i,k)}$ in a global measure.

Clearly, for the above conditions to be coherent from one iteration to the next, we need to assume some relationship between the various norms that we introduced. More precisely, we will assume that all the norms associated to the same subspace are *uniformly equivalent* in the following sense.

AS.5 There exist constants $\sigma_1, \sigma_2, \sigma_3 \geq 1$ such that, for all $k_1 \geq 0$ and $k_2 \geq 0$ and all $x \in \mathbf{R}\ n$,

$$\|x\|_{(i,k_1)} \leq \sigma_1 \|x\|_{(i,k_2)} \tag{2.11}$$

and

$$\|x\|_{[i,k_1]} \leq \sigma_2 \|x\|_{[i,k_2]} \tag{2.12}$$

for all $i \in \{1, \ldots, p\}$ and all $x \in \mathcal{R}_i$, and also that

$$\|x\|_{(k_1)} \leq \sigma_3 \|x\|_{(k_2)}. \tag{2.13}$$

We observe that the equivalence of all norms in finite dimensional spaces implies the existence of a $\sigma_4 \geq 1$ such that

$$\|P_{\mathcal{R}_i}(x)\|_{(i,k)} \leq \sigma_4 \|x\|_{(k)}. \tag{2.14}$$

for all $x \in \mathbf{R}\ n$. We also note that (2.11)–(2.13) necessarily hold if the norms $\|\cdot\|_{(i,k)}$, $\|\cdot\|_{[i,k]}$ and $\|\cdot\|_{(k)}$ are replaced by the ℓ_1,ℓ_2 or ℓ_∞ norms. We simplify the notation by defining

$$\sigma \stackrel{\text{def}}{=} \max\{\sigma_1, \sigma_2, \sigma_3, \sigma_4\} \tag{2.15}$$

which can then play the role of "universal" norm equivalence constant for all the norms so far considered.

2.1.4 The overall model and trust region

With all the elemental models at hand, we are now in position to define the overall model at iteration k, denoted m_k, whose purpose is to approximate the overall objective function f in a neighbourhood of the current iterate x_k. From (2.1), it is natural to use the overall model

$$m_k(x) \stackrel{\text{def}}{=} \sum_{i=1}^{p} m_{i,k}(x) \tag{2.16}$$

for all x in the overall trust region whose definition is now discussed.

First consider the set

$$D_k \stackrel{\text{def}}{=} \{i \in \{1, \ldots, p\} \mid P_{\mathcal{R}_i}(g_k) \neq 0\}, \tag{2.17}$$

where g_k is the vector

$$g_k \stackrel{\rm def}{=} \nabla m_k(x_k) = \sum_{i=1}^{p} g_{i,k}. \tag{2.18}$$

Next define

$$\Delta_{g,k} \stackrel{\rm def}{=} \max_{i \in D_k} \Delta_{i,k}, \tag{2.19}$$

the largest trust region radius associated with this set of elements, the associated "feasible ball" in $\mathbf{R}\ n$

$$B_k^{\diamond} = \{x \in X \mid \|x - x_k\|_{(k)} \leq \nu_2 \Delta_{g,k}\} \tag{2.20}$$

where $\nu_2 > 0$ is a constant. In addition, define

$$B_k^{\square} = \bigcap_{i \in \{1,\ldots,p\}} B_{i,k}, \tag{2.21}$$

the intersection of all elemental trust regions. We then define the overall trust region B_k by

$$B_k \stackrel{\rm def}{=} B_k^{\square} \cap B_k^{\diamond} = \left(\bigcap_{i \in D_k} B_{i,k} \right) \cap \left(B_k^{\diamond} \cap \bigcap_{i \notin D_k} B_{i,k} \right) \tag{2.22}$$

We now interpret this definition. First observe that D_k is the set of elements in the ranges of which a descent direction for the overall model can be found. The first term on the last right hand-side of (2.22) thus guarantees that all descent directions at x_k on m_k can be used up to the point where the involved models cease to be trusted. The second term does not impose any additional restriction on descent directions, but merely prevents too large steps that are orthogonal to the gradient. It should be noted that its effect is quite different from that of an angle test of the type

$$|\langle g_k, s_k \rangle| \geq \zeta \|g_k\|_{[k]} \|s_k\|_{(k)} \quad (\zeta \in (0,1)) \tag{2.23}$$

because it does not prevent the steps being orthogonal to the steepest descent direction, but only restricts the size of such steps. This is useful because these steps may occur when moving away from a saddle point of the objective function. A similar restriction is obviously present in the case where the objective has only one element, the role of $\Delta_{g,k}$ being played by the (unique) trust region radius in this case.

2.1.5 Curvature

We now follow [9] and [41] and define the *generalized Rayleigh quotient* of f at x along $s \neq 0$ by

$$\omega_n(f,x,s) \stackrel{\rm def}{=} \frac{2}{\|s\|_n^2}[f(x+s) - f(x) - \langle \nabla f(x), s \rangle], \tag{2.24}$$

where the subscript in ω_n indicates the norm used in the definition. Obviously, this definition is valid only if s is such that $x+s$ belongs to the domain of definition of f. Note that, by convention,

$$\omega_n(f,x,s) = 0 \ \text{ whenever } \ s = 0. \tag{2.25}$$

If we assume that f is twice continuously differentiable, the mean-value theorem (see [23]) implies that

$$\omega_n(f,x,s) = 2\int_0^1\int_0^1 t\frac{\langle s, \nabla^2 f(x+tvs)s\rangle}{\|s\|_n^2}\,dv\,dt. \tag{2.26}$$

Furthermore, if f is quadratic and the ℓ_2-norm is used, then one easily verifies that $\omega_2(f_i,x,s)$ is independent of x and is equal to the Rayleigh quotient of the matrix $\nabla^2 f$ in the direction s. We note that, because of AS.4, $\omega_2(f_i,x,s)$ is bounded by some constant $L_i \geq 0$ (see [23]). Hence we obtain that

$$\omega_{(i,k)}(f_i,x,s) \leq \max\left\{\sigma^2 \max_{i\in\{1,\ldots,p\}} L_i, 1\right\} \stackrel{\rm def}{=} L \tag{2.27}$$

for all $x, x+s \in X$, all $i \in \{1,\ldots,p\}$ and all k. The quantity that we need in our algorithm statement and analysis is an monotonically increasing upper bound on the generalized Rayleigh quotient $\omega_{(i,k)}(m_{i,k},x_k,s_{i,k})$ defined by

$$\beta_k \stackrel{\rm def}{=} 1 + \max_{\substack{q\in\{1,\ldots,k\}\\ i\in\{1,\ldots,p\}}} \{\max[0, \omega_{i,q}(m_{i,q},x_q,s_{i,q})]\}, \tag{2.28}$$

where s_k is defined below as the actual trial step computed by the algorithm and $s_{i,k} \stackrel{\rm def}{=} P_{\mathcal{R}_i}(s_k)$. This quantity measures the curvature of the model m_k in the direction of the trial step s_k. If a quadratic model m_k is considered, an upper bound on β_k is given by the largest positive eigenvalue of its Hessian matrix, plus one. We will assume that our choice of models is such that this curvature does not increase too fast, which could lead to premature convergence of the algorithm to a non-critical point (see [41]). More precisely, we make the following assumption, as in [9], [10], [34] and [41].

AS.6

$$\sum_{i=0}^{\infty}\frac{1}{\beta_k} = +\infty. \tag{2.29}$$

This condition is weaker that the common assumption that the model's second derivative matrices are uniformly bounded [32], which holds, for instance, for the classical Newton's method, where quadratic models using analytical second derivatives are used on a compact domain. It also weaker than the condition

$$\omega_{(i,k)}(m_{i,k},x_k,s) \leq c_0 k \tag{2.30}$$

for some constant $c_0 > 0$, which holds in the case where quadratic element models are used and updated using either the BFGS or the safeguarded Symmetric Rank One quasi-Newton formulae.

2.1.6 Criticality

Before we can describe our algorithm in detail, we also need a *criticality criterion* for our problem. A critical point of our problem is a feasible point x where the negative gradient of the objective function $-\nabla f(x)$ belongs to the *normal cone* of X at $x \in X$, which is defined by

$$\mathcal{N}(x) \overset{\text{def}}{=} \{y \in \mathbf{R}\ n \mid \langle y, u - x \rangle \leq 0, \forall u \in X\}. \tag{2.31}$$

The associated *tangent cone* of X at $x \in X$ is the polar of $\mathcal{N}(x)$, that is

$$\mathcal{T}(x) \overset{\text{def}}{=} \mathcal{N}(x)^0 = \text{closure}\{\lambda(u - x) \mid \lambda \geq 0 \text{ and } u \in X\}. \tag{2.32}$$

Thus every measure of criticality has to depend on the (differentiable) objective f and on the geometry of the feasible set at the current point. We will use the symbol $\alpha(x, f, X)$ to denote such a criticality measure.

AS.7 The criticality measure $\alpha(x, h, X)$ is non-negative for all $x \in X$ and all h differentiable in a neighbourhood of x. Moreover $\alpha(x, h, X) = 0$ if and only if x is critical for the problem

$$\min_{x \in X} h(x). \tag{2.33}$$

But, within the algorithm, only approximate gradient vectors might be available, namely the vectors g_k and $g_{i,k}$, the gradient of the models. It is therefore natural to use

$$\alpha_k \overset{\text{def}}{=} \alpha(x_k, m_k, X), \tag{2.34}$$

the criticality measure for the problem

$$\min_{x \in X} m_k(x). \tag{2.35}$$

as an "approximate" criticality measure for (2.1).

In unconstrained optimization, one typically chooses

$$\alpha_k = \|g_k\|, \tag{2.36}$$

the obvious criticality measure (see [31] or [33]). When bound constraints are present, the choice

$$\alpha_k = \|P_X(x_k - g_k) - x_k\| \tag{2.37}$$

is made in [10]. For the infinite dimensional case, the definition

$$\alpha_k = \|P_X(x_k - g_k) - x_k\|^2. \tag{2.38}$$

is used in [41]. For the case where convex constraints are considered,

$$\alpha_k = \frac{\|P(x_k - t_k^C g_k) - x_k\|}{t_k^C}, \tag{2.39}$$

is chosen in [32], where $t_k^C > 0$ is the line coordinate of the so-called "generalized Cauchy Point" to be discussed below. In a similar context,

$$\alpha_k = |\min_{\substack{x_k + d \in X \\ \|d\|_{(k)} \leq 1}} \langle g_k, d \rangle|. \tag{2.40}$$

is used in [9].

2.2 Ensuring sufficient model decrease

2.2.1 An overview of the classical sufficient decrease condition

A key to trust region algorithm is to choose a step s_k at iteration k that is guaranteed to provide a sufficient decrease on the overall objective function model m_k. In other words, a step such that

$$\delta m_k \overset{\text{def}}{=} m_k(x_k) - m_k(x_k + s_k) \tag{2.41}$$

is sufficiently positive, given the value of the criticality measure α_k. This concept of "sufficient decrease" is usually made more formal by introducing the notion of the *(generalized) Cauchy point*. This remarkable point, denoted x_k^C, is typically computed by the trust region algorithms as a point on (or close to) the projected gradient path $P_X(x_k - tg_k)$ $(t \geq 0)$ that is also within the trust region and sufficiently reduces the overall model in the sense that

$$m_k(x_k) - m_k(x_k^C) \geq \bar{\kappa}_2 \frac{\alpha_k^2}{\beta_k}, \tag{2.42}$$

where $\bar{\kappa}_2 > 0$ is a constant. However, such a point may not exist when the trust region radius Δ_k is small compared with α_k^2/β_k. In this case, the generalized Cauchy point is chosen as (or close to) the intersection of projected gradient path with the boundary of the trust region, yielding an inequality of the form

$$m(x_k) - m(x_k^C) \geq \bar{\kappa}_2 \alpha_k \Delta_k. \tag{2.43}$$

One then ensures the "sufficient decrease" by requiring that the chosen step s_k produces at least a fixed fraction of the overall model reduction achieved by the generalized Cauchy point, which is to say that

$$\delta m_k \geq \kappa_2 \alpha_k \min\left\{\frac{\alpha_k}{\beta_k}, \Delta_k\right\}, \tag{2.44}$$

where $\kappa_2 \in (0, \bar{\kappa}_2]$.

Many variants on the above scheme exist in the literature for the single trust region case. The best known is for unconstrained problems when the ℓ_2-norm is used to define the trust region shape. In that case, the projected gradient path is simply given by all negative multiples of the gradient g_k and the Cauchy point is simply the point that minimizes the model m_k in the intersection of the steepest descent direction and the trust region. When other norms are used, for example the ℓ_∞ norm, one can then choose either to minimize the model in the intersection of this steepest descent direction and the trust region, as before, or to "bend" the projected gradient path onto the boundary of the trust region and to choose the generalized Cauchy point as a point which satisfies classical Goldstein-type linesearch conditions along that path while staying within the trust region. This latter strategy is used in the LANCELOT software [13], for instance. When additional convex constraints are present, the projected gradient path is additionally "bent" to follow the boundary of the intersection of the feasible domain. Thus the philosophy is the same, in that (2.44) is guaranteed in any case. This last condition has indeed been obtained for all the choices for α_k given in (2.36), (2.37), (2.38), (2.39) and (2.40) in the paper where they were respectively introduced.

2.2.2 Sufficient decrease for structured model and trust region

We will use the same approach in our structured model and trust region framework. We first observe that

$$\nu_1 \Delta_{g,k} = \lim_{t \to \infty} \|P_{B_k^\square}(x_k - t g_k) - x_k\|_{(k)}, \tag{2.45}$$

where $\Delta_{g,k}$ is defined in (2.19). Hence $\Delta_{g,k}$ can be viewed as the distance from x_k to the farthest point from x_k that lies on the projected gradient path defined by $P_{B_k^\square}(x_k - t g_k)$ for $t \geq 0$. Recall that $B_k^\square$ is defined in (2.21) as the intersection of the feasible domain and the region where all relevant element models can be trusted. It is therefore natural to choose a generalized Cauchy point such that for some constant $\kappa_2 \in (0,1)$

$$m_k(x_k) - m_k(x_k^C) \geq \kappa_2 \alpha_k \Delta_{g,k} \tag{2.46}$$

(as in (2.43)) when a point satisfying (2.42) cannot be found in (or close to) the intersection of the gradient path projected on $B_k^\square$ with the structured trust region. If we again request that the final step s_k produces at least a fraction of the decrease at the generalized Cauchy point, we obtain the condition that

$$\delta m_k \geq \kappa_2 \alpha_k \min\left\{\frac{\alpha_k}{\beta_k}, \Delta_{g,k}\right\}. \tag{2.47}$$

This is the condition that will be used in our algorithm. Note that condition (2.47) reduces to (2.44) in the case where only one trust region radius is considered.

2.3 A class of structured trust region algorithms

We now describe the class of algorithms that we consider for solving (2.1). Besides κ_1 used in (2.7), κ_2 used (2.47), ν_1 used in (2.5) and ν_2 used in (2.20), it depends on the constants

$$0 < \gamma_1 \leq \gamma_2 < 1 \leq \gamma_3, \tag{2.48}$$

$$0 < \eta_1 \leq \eta_2 < \eta_3 < 1 \tag{2.49}$$

and

$$0 < \mu_1 < \mu_2 < 1. \tag{2.50}$$

In addition to the above conditions, we also require a compatibility condition between the η_i's and the μ_i's. Specifically, we request that

$$\eta_2 - \eta_1 \geq \frac{p-1}{p}(\mu_1 + \mu_2). \tag{2.51}$$

Typical values for these constants are $\kappa_1 = 0.1$, $\kappa_2 = 0.01$, $\nu_1 = \nu_2 = 1$, $\gamma_1 = 0.1$, $\gamma_2 = 0.5$, $\gamma_3 = 2$, $\eta_1 = 0.01$, $\eta_2 = 0.25$, $\eta_3 = 0.75$, $\mu_1 = 0.05$ and $\mu_2 = 0.1$.

Algorithm

step 0: initialization.

The starting point $x_0 \in X$ is given, together with the element function values $\{f_i(x_0)\}_{i=1}^p$ and the initial trust region radii $\{\Delta_{i,0}\}_{i=1}^p$. Set $k = 0$.

step 1: model choice.

For $i \in \{1, \ldots, p\}$, choose the model $m_{i,k}$ of the element function f_i in the trust region $B_{i,k}$ centered at x_k (as defined in (2.5)), satisfying (2.6) and (2.7).

step 2: determination of the step.

Choose a step s_k such that (2.47) holds and

$$x_k + s_k \in B_k. \tag{2.52}$$

step 3: measure overall model fit.

If

$$\delta f_k \overset{\text{def}}{=} f(x_k) - f(x_k + s_k) \geq \eta_1 \delta m_k \tag{2.53}$$

then

$$x_{k+1} = x_k + s_k, \tag{2.54}$$

else

$$x_{k+1} = x_k. \tag{2.55}$$

step 4: update the elemental trust region radii.

Denote the achieved changes in the element functions and their models by

$$\delta f_{i,k} \overset{\text{def}}{=} f_i(x_k) - f_i(x_k + s_k), \quad i \in \{1, \ldots, p\} \tag{2.56}$$

and

$$\delta m_{i,k} \overset{\text{def}}{=} m_{i,k}(x_k) - m_{i,k}(x_k + s_k), \quad i \in \{1, \ldots, p\}, \tag{2.57}$$

respectively. Then define the set of *negligible* elements at iteration k as

$$N_k \overset{\text{def}}{=} \{i \in \{1, \ldots, p\} \mid |\delta m_{i,k}| \leq \frac{\mu_1}{p} \delta m_k\} \tag{2.58}$$

and the set of *meaningful* elements as its complement, that is

$$M_k = \{1, \ldots, p\} \setminus N_k. \tag{2.59}$$

Then, for each $i \in \{1, \ldots, p\}$, perform the following.

Case 1: $i \in M_k$**.**

- If

$$\delta f_{i,k} \geq \delta m_{i,k} - \frac{1 - \eta_3}{p} \delta m_k \tag{2.60}$$

and (2.53) both hold, then choose

$$\Delta_{i,k+1} \in [\Delta_{i,k}, \gamma_3 \Delta_{i,k}]. \tag{2.61}$$

- If (2.60) holds but (2.53) fails then choose

$$\Delta_{i,k+1} = \Delta_{i,k}. \tag{2.62}$$

- If (2.60) fails, but

$$\delta f_{i,k} \geq \delta m_{i,k} - \frac{1-\eta_2}{p}\delta m_k \tag{2.63}$$

holds, then choose

$$\Delta_{i,k+1} \in [\gamma_2 \Delta_{i,k}, \Delta_{i,k}]. \tag{2.64}$$

- If (2.63) fails, then choose

$$\Delta_{i,k+1} \in [\gamma_1 \Delta_{i,k}, \gamma_2 \Delta_{i,k}]. \tag{2.65}$$

Case 2: $i \in N_k$

- If

$$|\delta f_{i,k}| \leq \frac{\mu_2}{p}\delta m_k, \tag{2.66}$$

and (2.53) both hold, then choose

$$\Delta_{i,k+1} \in [\Delta_{i,k}, \gamma_3 \Delta_{i,k}]. \tag{2.67}$$

- If (2.66) holds but (2.53) fails, then choose

$$\Delta_{i,k+1} = \Delta_{i,k}. \tag{2.68}$$

- If (2.66) fails, then choose

$$\Delta_{i,k+1} \in [\gamma_1 \Delta_{i,k}, \gamma_2 \Delta_{i,k}]. \tag{2.69}$$

Increment k by one and return to step 1.

End of Algorithm

As is traditional in trust region algorithm, we will call an iteration *successful* if the test (2.53) is satisfied, that is when the achieved objective reduction δf_k is large enough compared to the reduction δm_k predicted by the overall model. If (2.53) fails, the iteration is said to be *unsuccessful*. In what follows, we will denote by $\mathcal{S}$ the set of all successful iterations.

We now comment on various aspects of the algorithm.

1. The choice of the element models $m_{i,k}$ is left rather open in the above description. It clearly needs to be made precise for any practical implementation of the algorithm. One common choice would be to set

$$m_{i,k}(x_k+s) = f_i(x_k) + \langle g_{i,k}, s\rangle + \tfrac{1}{2}\langle s, H_{i,k}s\rangle, \tag{2.70}$$

where $H_{i,k}$ is a symmetric approximation to $\nabla^2 f_i(x_k)$ whose nullspace contains the subspace $\mathcal{N}_i$. In particular, Newton's method corresponds to the choice $g_{i,k} = \nabla f_i(x_k)$ and $H_{i,k} = \nabla^2 f_i(x_k)$, which is guaranteed to satisfy this latter condition. Another possible choice is $m_{i,k}(x_k+s) = f_i(x_k+s)$, which may be attractive for the simpler element functions. In this case, the model's fit to the true function is always good for the i-th element, and $\Delta_{i,k}$ is a non-decreasing sequence.

2. We note that (2.7) gives a practical rule for determining the required element gradient accuracy *before* it is actually needed in the computation to form the model $m_{i,k}$. A decreasing elemental trust region radius might impose a higher accuracy requirement on the corresponding model.

3. If the model change for an element is negligible, that is small compared to the overall predicted change, we do not need to restrict its element trust region size unless the true element change is relatively large compared with the same overall predicted change. We can therefore afford to ignore negligible items until they stop being relatively negligible, something which is inevitable when convergence occurs. Hence our distinction between "negligible" elements (in N_k) and "meaningful" ones (in M_k).

4. The apparent intricacy of (2.60) and (2.63) is caused by two complications which arise in the context of multiple elements. The first is that $\delta m_{i,k}$ cannot be assumed to be positive in general, even if δm_k always is (because of (2.47)). The second is that possible cancellation between elements makes it necessary to consider the "accuracy of model fit" for an element to be relative to the *overall* model fit. Indeed, requiring small relative errors for models with very large values may result in large absolute errors. If δm_k is small, these large errors will then cause δm_k to be a poor prediction of δf_k and the iteration might be unsuccessful. This explains why the perhaps more intuitive tests

   $$\delta f_{i,k} \geq \delta m_{i,k} - (1-\eta_j)|\delta m_{i,k}| \quad (j = 2, 3) \tag{2.71}$$

   cannot not be used instead of (2.63) ($j = 2$) and (2.60) ($j = 3$).

   Observe also that conditions (2.60) and (2.63) reduces to the familiar

   $$\delta f_k \geq \eta_j \delta m_k \quad (j = 2, 3), \tag{2.72}$$

   when $p = 1$.

5. Note again the consistency between the trust region radii updates in step 4 and the case where $p = 1$. In this latter case, the set N_k is always empty and (2.63) then implies (2.53), because of (2.49). Equation (2.62) is thus never invoked.

6. No stopping criterion has been explicitly included in our algorithm description. This is adequate for the theoretical analysis that we consider in the present paper, where we are interested in the asymptotic behaviour of the method, but is should be completed for any practical use. The choice of a particular stopping criterion will depend on the type of models being used.

7. The mechanism that we specified for updating the trust region radii does not exclude the additional requirement that the radii be uniformly bounded, if that is judged suitable for the type of models used. In practice, keeping the radii bounded is essential to prevent numerical overflow.

Before starting our global convergence analysis, we first state, for future reference, some properties that result from the mechanism of the algorithm.

Lemma 1 *Assume that AS.3 holds. At iteration k of the algorithm,*

1. *M_k contains at least one element. Furthermore*

$$\left(1-\frac{p-1}{p}\mu_1\right)\delta m_k \leq \sum_{i\in M_k}\delta m_{i,k} \leq \left(1+\frac{p-1}{p}\mu_1\right)\delta m_k; \tag{2.73}$$

2.

$$\|s_k\|_{(k)} \leq \nu\Delta_{g,k}, \tag{2.74}$$

where $\nu \stackrel{\rm def}{=} \max\{\nu_1,\nu_2\}$;

3.

$$\gamma_1\Delta_{i,k} \leq \Delta_{i,k+1} \leq \gamma_3\Delta_{i,k} \tag{2.75}$$

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

Proof. The first result immediately follows from the definition of N_k and the inequality $\mu_1 < 1$. One then deduces that N_k contains at most $p-1$ elements. Hence,

$$\delta m_k = \sum_{i\in M_k}\delta m_{i,k} + \sum_{i\in N_k}\delta m_{i,k} \leq \sum_{i\in M_k}\delta m_{i,k} + \mu_1\frac{|N_k|}{p}\delta m_k \tag{2.76}$$

from which the first part of (2.73) may be deduced. The second inequality in this result is obtained from

$$\sum_{i\in M_k}\delta m_{i,k} = \delta m_k - \sum_{i\in N_k}\delta m_{i,k} \leq \delta m_k + \sum_{i\in N_k}|\delta m_{i,k}|, \tag{2.77}$$

the relation (2.58) and $|N_k| \leq p-1$. The bound (2.74) immediately follows from (2.52) and (2.22). The bound (2.75) results from (2.61), (2.65), (2.67) and (2.69). □

We also investigate the coherency between the measure of fit for individual elements and that for the overall model.

Lemma 2 *Assume AS.3 holds and that, at iteration k of the algorithm, (2.63) holds for all $i \in M_k$ and that (2.66) holds for all $i \in N_k$. Then iteration k is successful, i.e. $k \in \mathcal{S}$.*

Proof. Because (2.63) holds for $i \in M_k$, one has that

$$\sum_{i\in M_k}\delta f_{i,k} \geq \sum_{i\in M_k}\delta m_{i,k} - (1-\eta_2)\frac{|M_k|}{p}\delta m_k \geq \left(\eta_2 - \frac{p-1}{p}\mu_1\right)\delta m_k \tag{2.78}$$

for all such i, where we used the inequality $|M_k| \leq p$ and Lemma 1 to deduce the second inequality. On the other hand, since (2.66) holds for $i \in N_k$, one obtains for these i that

$$\sum_{i\in N_k}|\delta f_{i,k}| \leq \frac{p-1}{p}\mu_2\delta m_k, \tag{2.79}$$

where we used Lemma 1 again to bound $|N_k|$. Now,

$$\delta f_k = \sum_{i \in M_k} \delta f_{i,k} + \sum_{i \in N_k} \delta f_{i,k} \geq \sum_{i \in M_k} \delta f_{i,k} - \sum_{i \in N_k} |\delta f_{i,k}|. \tag{2.80}$$

Combining this last inequality with (2.78) and (2.79) gives that

$$\delta f_k \geq \left(\eta_2 - \frac{p-1}{p}\mu_1 - \frac{p-1}{p}\mu_2 \right) \delta m_k \tag{2.81}$$

which then yields (2.53) because of (2.51). □

We therefore see that (2.53) is coherent with the measure of the fit between the element models and element functions.

3 Global convergence

We now study the convergence properties of the class of algorithms that we introduced in the preceding section. Our analysis follows the pattern of similar proofs with a single trust region (see [9] or [41]). The central idea in the proof is that the algorithm will continue to make progress as long as a critical point is not reached. We first start by bounding the error between the true element functions and their models. We next derive a lower bound on the size of the smallest trust region radius at a non-critical point. This lower bound ensures that the trust region constraint will not prevent further progress towards a critical point. Only with this bound can we then prove that limit points of the sequence of iterates produced by the Algorithm are indeed critical for the models used. We close the section by deriving some simple consequences of these results on the criticality of the limit points for the true objective function.

We first start by bounding the error made between the model of any element function and the element function itself at $x_k + s_k$.

Lemma 3 *Assume that AS.4 holds and consider a sequence $\{x_k\}$ of iterates generated by the algorithm. Then there exists a positive constant $c_1 > 1$ such that,*

$$|f_i(x_k + s_k) - m_{i,k}(x_k + s_k)| \leq c_1 \beta_k \Delta_{i,k}^2 \tag{3.1}$$

for all $i \in \{1, \ldots, p\}$ and all k.

Proof. We first observe that, for each $i \in \{1, \ldots, p\}$ and for all k, (2.6), the inequality (2.9) and the definition (2.24) imply that

$$\begin{aligned} |f_i(x_k + s_k) - m_{i,k}(x_k + s_k)| &\leq |\langle \nabla f_i(x_k) - g_{i,k}, s_{i,k} \rangle| \\ &\quad + \tfrac{1}{2} \|s_{i,k}\|_{(i,k)}^2 |\omega_{(i,k)}(f_i, x_k, s_{i,k}) - \omega_{(i,k)}(m_{i,k}, x_k, s_{i,k})| \\ &\leq \|e_{i,k}\|_{[i,k]} \, \|s_{i,k}\|_{(i,k)} \\ &\quad + \tfrac{1}{2} \|s_{i,k}\|_{(i,k)}^2 (|\omega_{(i,k)}(f_i, x_k, s_{i,k})| + |\omega_{(i,k)}(m_{i,k}, x_k, s_{i,k})|). \end{aligned} \tag{3.2}$$

But $\|s_{i,k}\|_{(i,k)} \leq \nu \Delta_{i,k}$, and hence we obtain from (2.7), (2.8), (2.27) and (2.28) that

$$|f_i(x_k + s_k) - m_{i,k}(x_k + s_k)| \leq \kappa_1 \nu \Delta_{min,k} \Delta_{i,k} + \tfrac{1}{2} \nu^2 (L + \beta_k) \Delta_{i,k}^2. \tag{3.3}$$

This then yields (3.1) with

$$c_1 \overset{\text{def}}{=} \max\{1, \kappa\nu + \tfrac{1}{2}\nu^2(L+1)\} \tag{3.4}$$

□

We now examine the relation between the change predicted by the overall model and that predicted for an element at a non critical point.

Lemma 4 *Assume that AS.1, AS.3–AS.5 hold. Consider iteration k of the algorithm and assume that*

$$\beta_k \Delta_{i,k} \le \min\{1, \frac{\alpha_k}{\nu}\} \tag{3.5}$$

and

$$\Delta_{min,k} \le 1. \tag{3.6}$$

Then one has that

$$|\delta m_{i,k}| \le c_2 \|s_{i,k}\|_{(i,k)} \tag{3.7}$$

for all $i \in \{1, \dots, p\}$ and for some constant $c_2 > 0$ independent of i and k.

Proof. Using (2.28) and (2.7), we obtain that

$$|\delta m_{i,k}| \le |\langle g_{i,k}, s_{i.k}\rangle| + \tfrac{1}{2}\beta_k \|s_{i,k}\|^2_{(i,k)} \le |\langle \nabla f_i(x_k), s_{i,k}\rangle| + |\langle e_{i,k}, s_{i,k}\rangle| + \tfrac{1}{2}\beta_k \|s_{i,k}\|^2_{(i,k)} \tag{3.8}$$

Remembering now AS.1, (2.7), (3.5) and (3.6), we can deduce that

$$\begin{aligned} |\delta m_{i,k}| &\le \sigma \max_{x\in X} (\|\nabla f_i(x)\|_2) \|s_{i,k}\|_{(i,k)} + \kappa_1 \Delta_{min,k} \|s_{i,k}\|_{(i,k)} + \tfrac{1}{2}\beta_k \|s_{i,k}\|^2_{(i,k)} \\ &\le [\sigma \max_{x\in X} (\|\nabla f_i(x)\|_2) + \kappa_1 + \tfrac{1}{2}] \|s_{i,k}\|_{(i,k)}. \end{aligned} \tag{3.9}$$

Inequality (3.9) then gives (3.7) with

$$c_2 \overset{\text{def}}{=} \sigma(\sigma \max_{x\in X} (\|\nabla f_i(x)\|_2) + \kappa_1 + \tfrac{1}{2}). \tag{3.10}$$

□

We next prove the important fact that the trust region radii stay bounded away from zero as long as a critical point is not reached, therefore allowing further progress to be made.

Theorem 5 *Assume that AS.1–AS.5 hold. Consider a sequence $\{x_k\}$ of iterates generated by the algorithm and assume that there exists a constant $\epsilon > 0$ such that*

$$\alpha_k \ge \epsilon \tag{3.11}$$

for all k. Then there is a constant $c_3 > 0$ such that

$$\Delta_{min,k} \ge \frac{c_3}{\beta_k} \tag{3.12}$$

for all k.

Proof. Assume, without loss of generality, that

$$\epsilon < \nu\beta_0\Delta_{min,0}. \tag{3.13}$$

In order to derive a contradiction, assume that there exists a k such that

$$\beta_k\Delta_{min,k} \leq \gamma_1 \min\left\{1, \frac{\epsilon}{\nu}, \frac{\mu_1 c_4^2(1-\eta_3)}{c_1c_2\nu p^2}, \frac{c_4(\mu_2-\mu_1)}{c_1 p}\right\} \overset{\text{def}}{=} c_3, \tag{3.14}$$

where $c_4 \overset{\text{def}}{=} \kappa_2\gamma_1\epsilon$. Now define r to be the smallest iteration number such that (3.14) holds. (Note that $r \geq 1$ because of (3.13) and the inequality $\gamma_1 < 1$.) Also fix i such that $\Delta_{min,r} = \Delta_{i,r}$. The bound (2.75) and the monotonic nature of the sequence $\{\beta_k\}$ then ensure that

$$\beta_{r-1}\Delta_{i,r-1} \leq \beta_r\frac{\Delta_{i,r}}{\gamma_1} \leq \frac{c_3}{\gamma_1} \leq \frac{\epsilon}{\nu} \tag{3.15}$$

where we used the bound $\beta_r\Delta_{i,r} \leq \gamma_1\epsilon/\nu$ from (3.14). We note that, because of (2.47), (3.11), (3.15) and the relation $\Delta_{min,r-1} \in (\Delta_{i,r}, \Delta_{i,r-1}]$,

$$\delta m_{r-1} \geq \kappa_2\epsilon\min\left\{\frac{\epsilon}{\beta_{r-1}}, \Delta_{min,r-1}\right\} \geq \kappa_2\epsilon\Delta_{min,r-1} \geq \kappa_2\epsilon\Delta_{i,r}, \tag{3.16}$$

which implies, because of (2.75), that

$$\delta m_{r-1} \geq c_4\Delta_{i,r-1}. \tag{3.17}$$

But (3.15) and (3.14) imply that

$$\Delta_{min,r-1} \leq \beta_{r-1}\Delta_{i,r-1} \leq 1. \tag{3.18}$$

Hence, this inequality together with (3.15) and (3.16) now allow us to apply Lemma 4 and to deduce that

$$|\delta m_{i,r-1}| \leq c_2\|s_{i,r-1}\|_{(i,r-1)} \leq c_2\nu\Delta_{i,r-1} \leq \frac{c_2\nu}{c_4}\delta m_{r-1}. \tag{3.19}$$

Assume first that $i \in M_{r-1}$. Then, using (2.58) and (3.17),

$$|\delta m_{i,r-1}| > \frac{\mu_1}{p}\delta m_{r-1} \geq \frac{\mu_1 c_4}{p}\Delta_{i,r-1}. \tag{3.20}$$

Because of (2.6), (3.1), (3.20), (2.8) and (2.5), we therefore obtain that

$$\left|\frac{\delta f_{i,r-1}}{\delta m_{i,r-1}} - 1\right| = \frac{|f_i(x_{r-1}+s_{r-1}) - m_{i,r-1}(x_{r-1}+s_{r-1})|}{|\delta m_{i,r-1}|} \leq \frac{c_1 p}{\mu_1 c_4}\beta_{r-1}\Delta_{i,r-1}. \tag{3.21}$$

But (3.14) and the first inequality of (3.14) together give that

$$\beta_{r-1}\Delta_{i,r-1} \leq (1-\eta_3)\frac{\mu_1 c_4^2}{c_1c_2\nu p^2}, \tag{3.22}$$

which, with (3.21), implies that

$$\left|\frac{\delta f_{i,r-1}}{\delta m_{i,r-1}} - 1\right| \leq \frac{(1-\eta_3)c_4}{c_2\nu p}. \tag{3.23}$$

Consider first the case where $\delta m_{i,r-1} > 0$. We may then apply (3.19) and deduce that

$$\delta m_{i,r-1} - \frac{1-\eta_3}{p}\delta m_{r-1} = \delta m_{i,r-1}\left(1 - \frac{(1-\eta_3)}{p}\frac{\delta m_{r-1}}{|\delta m_{i,r-1}|}\right) \leq \delta m_{i,r-1}\left(1 - \frac{(1-\eta_3)c_4}{c_2\nu p}\right). \tag{3.24}$$

Using (3.23), we now deduce that

$$\frac{\delta f_{i,r-1}}{\delta m_{i,r-1}} \geq 1 - \frac{(1-\eta_3)c_4}{c_2\nu p} \tag{3.25}$$

and therefore, because of (3.24), that

$$\delta f_{i,r-1} \geq \delta m_{i,r-1}\left(1 - \frac{(1-\eta_3)c_4}{c_2\nu p}\right) \geq \delta m_{i,r-1} - \frac{1-\eta_3}{p}\delta m_{r-1} \tag{3.26}$$

which implies that (2.60) holds for element i at iteration $r-1$. Now turn to the case where $\delta m_{i,r-1} < 0$. Because of (3.19), we deduce that

$$\delta m_{i,r-1} - \frac{1-\eta_3}{p}\delta m_{r-1} = \delta m_{i,r-1}\left(1 + \frac{(1-\eta_3)}{p}\frac{\delta m_{r-1}}{|\delta m_{i,r-1}|}\right) \leq \delta m_{i,r-1}\left(1 + \frac{(1-\eta_3)c_4}{c_2\nu p}\right). \tag{3.27}$$

As above, we use (3.23) to obtain that

$$\frac{\delta f_{i,r-1}}{\delta m_{i,r-1}} \leq 1 + \frac{(1-\eta_3)c_4}{c_2\nu p} \tag{3.28}$$

and therefore, because of (3.27), that

$$\delta f_{i,r-1} \geq \delta m_{i,r-1}\left(1 + \frac{(1-\eta_3)c_4}{c_2\nu p}\right) \geq \delta m_{i,r-1} - \frac{1-\eta_3}{p}\delta m_{r-1} \tag{3.29}$$

which again implies that (2.60) holds for element i at iteration $r-1$.

Assume now that $i \in N_{r-1}$. Then, because of (2.58) and (3.1), we have that

$$\begin{array}{rcl} |\delta f_{i,r-1}| & \leq & |\delta m_{i,r-1}| + |f_i(x_{r-1}+s_{r-1}) - m_{i,r-1}(x_{r-1}+s_{r-1})| \\ & \leq & \frac{\mu_1}{p}\delta m_{r-1} + c_1\beta_{r-1}\Delta_{i,r-1}^2. \end{array} \tag{3.30}$$

Now, multiplying (3.17) by $\Delta_{i,r-1}$, we obtain that

$$\Delta_{i,r-1}^2 \leq \frac{\Delta_{i,r-1}}{c_4}\delta m_{r-1}. \tag{3.31}$$

Gathering (3.30) and (3.31), we deduce that

$$|\delta f_{i,r-1}| \leq \left(\frac{\mu_1}{p} + \frac{c_1}{c_4}\beta_{r-1}\Delta_{i,r-1}\right)\delta m_{r-1}. \tag{3.32}$$

Observing now that (3.14) and the first inequality of (3.14) imply that

$$\beta_{r-1}\Delta_{i,r-1} \leq \frac{c_4}{c_1}.\frac{\mu_2-\mu_1}{p}, \tag{3.33}$$

we obtain from (3.32) that

$$|\delta f_{i,r-1}| \leq \frac{\mu_2}{p}\delta m_{r-1}. \tag{3.34}$$

But this inequality implies that (2.66) holds for element i at iteration $r-1$. Thus either (2.60) or (2.66) holds for element i at iteration $r-1$ and the mechanism of the algorithm then implies that $\Delta_{i,r} \geq \Delta_{i,r-1}$. But we may deduce from this inequality that

$$\beta_{r-1}\Delta_{min,r-1} \leq \beta_{r-1}\Delta_{i,r-1} \leq \beta_r\Delta_{i,r} \tag{3.35}$$

which contradicts the assumption that r is the smallest iteration number such that (3.14) holds. The inequality (3.14) therefore never holds and we obtain that (3.12) is satisfied for all k. □

We now turn to one of the main results in this section, which proves a weak form of global convergence. The technique is inspired by [34].

Theorem 6 *Assume that AS.1–AS.7 hold. Consider a sequence* $\{x_k\}$ *of iterates generated by the algorithm. Then*

$$\liminf_{k\rightarrow\infty} \alpha_k = 0. \tag{3.36}$$

Proof. Assume, for the purpose of obtaining a contradiction, that there exists an $\epsilon \in (0,1)$ such that (3.11) holds for all $k \geq 0$. Then

$$\begin{array}{rcl} \sum_{k\in\mathcal{S}} \delta f_k & \geq & \eta_1 \sum_{k\in\mathcal{S}} \delta m_k \\ & \geq & \eta_1\kappa_2\epsilon \sum_{k\in\mathcal{S}} \min\left\{\frac{\epsilon}{\beta_k}, \Delta_{g,k}\right\} \\ & \geq & \eta_1\kappa_2\epsilon \sum_{k\in\mathcal{S}} \min\left\{\frac{\epsilon}{\beta_k}, \Delta_{min,k}\right\} \\ & \geq & \eta_1\kappa_2\epsilon \min\{\epsilon, c_3\} \sum_{k\in\mathcal{S}} \frac{1}{\beta_k}, \end{array} \tag{3.37}$$

where we used successively (2.53), (2.47), (3.11), (2.19) and Theorem 5. We note that AS.2 then implies that

$$\sum_{k\in\mathcal{S}} \frac{1}{\beta_k} < +\infty. \tag{3.38}$$

Now let r be an integer such that

$$\gamma_3\gamma_2^{(r-1)/p} < 1 \tag{3.39}$$

and define

$$\mathcal{S}(k) \stackrel{\rm def}{=} |\mathcal{S} \cap \{1,\ldots,k\}|, \tag{3.40}$$

the number of successful iterations up to iteration k ($k \geq 1$). Then define

$$\mathcal{F}_1 \stackrel{\rm def}{=} \{k \mid k \leq r\mathcal{S}(k)\} \;\text{ and }\; \mathcal{F}_2 \stackrel{\rm def}{=} \{k \mid k > r\mathcal{S}(k)\}. \tag{3.41}$$

We now wish to show that both sums

$$\sum_{k\in\mathcal{F}_1} \frac{1}{\beta_k} \;\text{ and }\; \sum_{k\in\mathcal{F}_2} \frac{1}{\beta_k} \tag{3.42}$$

are finite. Consider the first. If it has only finitely many terms, its convergence is obvious. Otherwise, we may assume that $\mathcal{F}_1$ has an infinite number of elements, and we then construct two subsequences. The first consists of the indices of $\mathcal{F}_1$ in ascending order and the second, $\mathcal{F}_3$ say, of the set of indices in $\mathcal{S}$ (in ascending order) with each index repeated

r times. Hence the j-th element of $\mathcal{F}_3$ is no greater than the j-th element of $\mathcal{F}_1$. This gives that

$$\sum_{k\in\mathcal{F}_1}\frac{1}{\beta_k}\leq\sum_{k\in\mathcal{F}_3}\frac{1}{\beta_k}=r\sum_{k\in\mathcal{S}}\frac{1}{\beta_k}\leq+\infty \tag{3.43}$$

because of the nondecreasing nature of the sequence $\{\beta_k\}$ and (3.38). Now turn to the second sum in (3.42). Lemma 2 implies that, at each unsuccessful iteration, at least one element trust region radius satisfies (2.65) or (2.69) and none of them is allowed to increase. Hence

$$\prod_{i=1}^{p}\Delta_{i,k}\leq\gamma_3^{p\mathcal{S}(k)}\gamma_2^{k-\mathcal{S}(k)}\prod_{i=1}^{p}\Delta_{i,0}, \tag{3.44}$$

which immediately implies that

$$\Delta_{min,k}\leq\gamma_3^{\mathcal{S}(k)}\gamma_2^{(k-\mathcal{S}(k))/p}\Delta_{max,0}, \tag{3.45}$$

where $\Delta_{max,0}\stackrel{\rm def}{=}\max_{i\in\{1,\ldots,p\}}\Delta_{i,0}$. We deduce from this inequality that, for $k\in\mathcal{F}_2$,

$$\frac{c_3}{\beta_k}\leq\Delta_{min,k}\leq\gamma_3^{\mathcal{S}(k)}\gamma_2^{(k-\mathcal{S}(k))/p}\Delta_{max,0}\leq\gamma_3^{k/r}\gamma_2^{(k-k/r)/p}\Delta_{max,0}\leq\left[\gamma_3\gamma_2^{(r-1)/p}\right]^{k/r}\Delta_{max,0}, \tag{3.46}$$

where we have also used Theorem 5 and the definition of $\mathcal{F}_2$ in (3.41). This gives that

$$\sum_{k\in\mathcal{F}_2}\frac{1}{\beta_k}\leq\frac{\Delta_{max,0}}{c_3}\sum_{k\in\mathcal{F}_2}\left[\gamma_3\gamma_2^{(r-1)/p}\right]^{k/r}<+\infty \tag{3.47}$$

and the second sum is convergent. Therefore the sum

$$\sum_{k=0}^{\infty}\frac{1}{\beta_k}=\sum_{k\in\mathcal{F}_1}\frac{1}{\beta_k}+\sum_{k\in\mathcal{F}_2}\frac{1}{\beta_k} \tag{3.48}$$

is finite, which contradicts AS.6. Hence condition (3.11) is impossible and (3.36) follows. □

Notice that the relation between α_k, the criticality measure for problem (2.35), and $\alpha(x_k,f,X)$, the criticality measure for problem (2.1), has been left rather unspecified up to this point. It is indeed remarkable that we can prove Theorem 6 assuming so little on α. In order to derive convergence properties for the original problem from Theorem 6, we have to be slightly more specific and request that, if both function and model have the same first order information, then the criticality measures on the original problem and on the model problem agree.

AS.8 Let h_1 and h_2 be two continuously differentiable functions in the intersection of a neighbourhood of the feasible point x and X, such that $h_1(x)=h_2(x)$. Then, the difference $\alpha(x,h_1,X)-\alpha(x,h_2,X)$ tends to zero when $\nabla h_1(x)-\nabla h_2(x)$ tends to zero.

In other words, we require the criticality measure to be continuous (near zero) in the *gradient* of its second argument. Again, this is true for the choices (2.36)–(2.37) and (2.40).

With this additional assumption, we are now ready to examine the criticality of the limit points of the sequence of iterates generated by the algorithm for the original problem (2.1).

Corollary 7 *Assume that AS.1–AS.8 hold. Consider a sequence $\{x_k\}$ of iterates generated by the algorithm and assume that*

$$\lim_{k\to\infty} \|e_{i,k}\|_{[k,i]} = 0 \tag{3.49}$$

for all $i \in \{1,\ldots,p\}$. Then this sequence has at least one critical limit point x_*.

Proof. From AS.8 and (3.49), we obtain that

$$\lim_{k\to\infty} [\alpha(x_k, f, X) - \alpha_k] = 0, \tag{3.50}$$

which, with (3.36), guarantees

$$\liminf_{k\to\infty} \alpha(x_k, f, X) = 0. \tag{3.51}$$

The desired conclusion then follows by taking a subsequence of $\{x_k\}$ if necessary. □

Condition (3.49) is important, otherwise the situation might arise that an iterate is critical for the current overall model (because its gradient is inexact) while not being critical for the original problem. There are various ways in which (3.49) can be achieved in a practical algorithm, the simplest being to make the norm of $e_{i,k}$ also depend on α_k itself, ensuring that the first goes to zero if the latter does.

Corollary 8 *Assume that AS.1–AS.8 hold. If $\mathcal{S}$, the set of successful iterations generated by the algorithm is finite, then all iterates x_k are equal to some x_* for k large enough, and x_* is critical.*

Proof. Assume indeed that $\mathcal{S}$ is finite. It then clear from (2.55) that x_k is unchanged for k large enough, and therefore that $x_* = x_{j+1}$, where j is the largest index in $\mathcal{S}$. Note now that Lemma 2 implies that, if $k \notin \mathcal{S}$, then (2.63) or (2.66) must be violated for at least one element. Hence we obtain that $\Delta_{min,k}$ converges to zero. But (2.7) then implies that $e_{i,k}$ also converges to zero for all $i \in \{1,\ldots,p\}$ and g_k converges to $\nabla f(x_k)$. Thus AS.8 and Theorem 6 then guarantee the criticality of x_*. □

As in existing theories for the single trust region case, it is possible to replace the limit inferior in (3.36) by a true limit, therefore ensuring (if the gradients are asymptotically exact) that *all* limit points are critical. As in these theories, a slight strengthening of our assumptions is however necessary.

AS.9 We assume that

$$\lim_{k\to\infty} \beta_k \delta f_k = 0. \tag{3.52}$$

This assumption is identical to that used in [9] and [41], where it is motivated in detail. We only mention here that (3.52) holds for Newton's method on bounded domains.

With this additional assumption, we are now able to replace the limit inferior by a true limit.

Theorem 9 *Assume that AS.1–AS.9 hold. Consider the sequence $\{x_k\}$ of iterates generated by the algorithm and assume that there are infinitely many successful iterations. Then*

$$\lim_{k\in\mathcal{S}} \alpha_k = 0, \tag{3.53}$$

where $\mathcal{S}$ is, as above, the set of successful iterations.

Proof. We again proceed by contradiction. Assume therefore that there exists an $\epsilon_1 \in (0,1)$ and a subsequence $\{q_j\}$ of successful iterates such that, for all q_j in this subsequence

$$\alpha_{q_j} \geq \epsilon_1. \tag{3.54}$$

Theorem 6 guarantees the existence of another subsequence $\{l_j\}$ such that

$$\alpha_k \geq \epsilon_2 \text{ for } q_j \leq k < l_j \text{ and } \alpha_{l_j} < \epsilon_2, \tag{3.55}$$

where we have chosen $\epsilon_2 \in (0,\epsilon_1)$. We may now restrict our attention to the subsequence of successful iterations whose indices are in the set

$$\mathcal{K} \stackrel{\text{def}}{=} \{k \mid k \in \mathcal{S} \text{ and } q_j \leq k < l_j\}, \tag{3.56}$$

where q_j and l_j belong, respectively, to the two subsequences defined above. Applying now (2.47) for $k \in \mathcal{K}$, we obtain from (2.53) that

$$\delta f_k \geq \eta_1 \kappa_2 \epsilon_2 \min\left\{\frac{\epsilon_2}{\beta_k}, \Delta_{g,k}\right\}. \tag{3.57}$$

But AS.9, the inequality (2.74) and $\beta_k \geq 1$ imply that

$$\lim_{\substack{k\to\infty\\ k\in\mathcal{K}}} \beta_k \|s_k\|_{(k)} \leq \nu \lim_{\substack{k\to\infty\\ k\in\mathcal{K}}} \beta_k \Delta_{g,k} = 0. \tag{3.58}$$

and also that

$$\lim_{\substack{k\to\infty\\ k\in\mathcal{K}}} \Delta_{min,k} \leq \lim_{\substack{k\to\infty\\ k\in\mathcal{K}}} \beta_k \Delta_{g,k} = 0. \tag{3.59}$$

Therefore, we can deduce from (3.57) and (3.58), that, for j sufficiently large,

$$\begin{array}{rcl} \|x_{q_j} - x_{l_j}\|_2 & \leq & \sigma \sum_{k=q_j}^{l_j-1} \|x_{k+1} - x_k\|_{(k)} \\ & = & \sigma \sum_{k=q_j}^{l_j-1}{}^{(\mathcal{K})} \|s_k\|_{(k)} \\ & \leq & \sigma\nu \sum_{k=q_j}^{l_j-1}{}^{(\mathcal{K})} \Delta_{g,k} \\ & \leq & c_5 \sum_{k=q_j}^{l_j-1}{}^{(\mathcal{K})} [f(x_k) - f(x_{k+1})] \\ & \leq & c_5[f(x_{q_j}) - f(x_{l_j})], \end{array} \tag{3.60}$$

where the sums with superscript $(\mathcal{K})$ are restricted to the indices in $\mathcal{K}$, and where

$$c_5 \stackrel{\text{def}}{=} \frac{\sigma\nu}{\eta_1\kappa_2\epsilon_2}. \tag{3.61}$$

But AS.2 implies that the last right-hand side of (3.60) converges to zero as j tends to infinity. Hence the continuity of ∇f and AS.8 give that

$$|\alpha(x_{q_j}, f, X) - \alpha(x_{l_j}, f, X)| \leq \frac{1}{6}(\epsilon_1 - \epsilon_2) \tag{3.62}$$

for j sufficiently large. On the other hand, (3.59), the inequality $\beta_k \geq 1$, AS.8 and (2.7) imply that g_{q_j} is arbitrarily close to $\nabla f(x_{q_j})$ when j is large enough, and hence that

$$|\alpha_{q_j} - \alpha(x_{q_j}, f, X)| \leq \frac{1}{6}(\epsilon_1 - \epsilon_2) \tag{3.63}$$

for j sufficiently large. We note also that, because of (2.7), (2.12) and (2.75),

$$\|g_{l_j} - \nabla f(x_{l_j})\|_2 \leq \sum_{i=1}^{p} \|e_{i,l_j}\|_2 \leq \sigma\kappa_1 p \Delta_{min,l_j} \leq \sigma\kappa_1\gamma_3 p \Delta_{min,k_j}, \tag{3.64}$$

where k_j is the largest integer in $\mathcal{K}$ that is smaller than l_j. We now deduce from (3.59) that the left-hand side of (3.64) tends to zero when j tends to infinity, and therefore that, for j sufficiently large,

$$|\alpha_{l_j} - \alpha(x_{l_j}, f, X)| \leq \frac{1}{6}(\epsilon_1 - \epsilon_2) \tag{3.65}$$

because of AS.8. Combining (3.62), (3.63) and (3.65), we obtain that

$$\alpha_{q_j} \leq \alpha_{l_j} + \tfrac{1}{2}(\epsilon_1 - \epsilon_2) \leq \tfrac{1}{2}(\epsilon_1 + \epsilon_2) < \epsilon_1, \tag{3.66}$$

which is impossible because of (3.54). Hence our initial assumption cannot hold and the theorem is proved. □

As above, we now consider the case where we impose that the element gradient are asymptotically exact.

Corollary 10 *Assume that AS.1–AS.9 hold. Consider the sequence $\{x_k\}$ of iterates generated by the algorithm and assume furthermore that (3.49) holds for all $i \in \{1, \ldots, p\}$. Then all limit points of this sequence are critical.*

Proof. If the set $\mathcal{S}$ is finite, the conclusion immediately follows from Corollary 8. If, on the other hand, $\mathcal{S}$ has an infinite number of elements, (3.49) implies that g_k is arbitrarily close to $\nabla f(x_k)$ and the combination of AS.8 and Theorem 9 ensures the criticality of any limit point of the sequence of successful iterates. The desired conclusion then follows from (2.55). □

Of course, (3.49) might be impossible to achieve in practice, and one might consider the case where we can only assert that

$$\limsup_{k\to\infty} \left[\max_{i\in\{1,\ldots,p\}} \|e_{i,k}\|_2 \right] = \kappa_3, \tag{3.67}$$

for some small constant $\kappa_3 > 0$.

Corollary 11 *Assume that AS.1–AS.7 and AS.9 hold. Consider the sequence $\{x_k\}$ of iterates generated by the algorithm. Assume furthermore that (3.67) holds and that the criticality measure α satisfies*

$$|\alpha(x, h_1, X) - \alpha(x, h_2, X)| \leq L_\alpha \|\nabla h_1(x) - \nabla h_2(x)\|_2 \tag{3.68}$$

for all $x \in X$ and all functions h_1 and h_2 continuously differentiable in a neighbourhood of x such that $h_1(x) = h_2(x)$. Then, for each limit point x_* of the sequence,

$$\alpha(x_*, f, X) \leq \kappa_3 p L_\alpha. \tag{3.69}$$

Proof. As in Corollary 10, the desired conclusion immediately follows from Corollary 8 if $\mathcal{S}$ is finite. Assume therefore that $\mathcal{S}$ has infinitely many elements. We then deduce that, for all $k \in \mathcal{S}$,

$$\begin{aligned} \alpha(x_k, f, X) &\leq \alpha_k + |\alpha(x_k, m_k, X) - \alpha(x_k, f, X)| \\ &\leq \alpha_k + L_\alpha \|g_k - \nabla f(x_k)\|_2 \\ &\leq \alpha_k + L_\alpha p \max_{i \in \{1,\ldots,p\}} \|e_{i,k}\|_2. \end{aligned} \tag{3.70}$$

Taking the limit for k tending to infinity in $\mathcal{S}$ and using Theorem 9 and (3.67) then gives the desired conclusion. □

Finally observe that (3.68), although stronger than AS.8, is not a very strong condition. For instance, it is satisfied with $L_\alpha = 1$ for the choices (2.36), and also for (2.37) and (2.38) because of the non-expansive character of the projection operator P_X (see [41], for example). The same property also holds for the choice (2.40), as discussed in [9].

4 Finite identification of the correct active set

When applied to constrained problems, trust region algorithms typically use the notion of projected gradient or projected gradient path in order to identify a subset of inequality constraints that are satisfied as equalities. Ultimately, the aim thereby is to identify the constraints satisfied as equalities at the solution well before the solution is reached. The methods then reduce to an unconstrained calculation in the manifold defined by the currently "active" constraints. As a consequence, it is possible to guarantee fast asymptotic rates of convergence when using accurate models, as is the case when analytical second order information of the objective and constraint functions is available.

The main purpose of the present paper is to show that structured trust regions do not upset the theory developed in the unstructured case. Thus we will consider the active constraint identification problem from a quite general point of view. Our main observation is that a number of the existing theories for constraint identification are based on the definition of a special criticality measure that satisfies AS.7 while not satisfying AS.8. Let us denote this measure at iteration k by $\bar{\alpha}_k$. The steps leading to constraint identification are then as follows.

1. The first step is to prove that a sufficient decrease condition of the type (2.44) also holds with $\bar{\alpha}_k$ instead of α_k.

2. One then proceeds to prove that

$$\liminf_{k \to \infty} \bar{\alpha}_k = 0 \tag{4.1}$$

much in the same way as for (3.36).

3. The measure $\bar{\alpha}_k$ is also constructed to ensure that it is asymptotically bounded away from zero for all points such that their active set is not identical to that of a (close) critical point. (This, in particular, prevents AS.8 from holding.)

4. Some contradiction is then deduced from these last two properties.

We now make our assumptions on the problem structure, algorithm and criticality measure more precise.

4.1 Assumptions on the constraints

We first make our definition of the feasible set more precise. In what follows, we will assume that the convex set X is described by a finite collection of convex inequalities.

AS.10 We have that

$$X = \{x \in \mathbf{R}\ n \mid h_i(x) \geq 0 \quad (i \in \{1, \ldots, n_c\})\}, \tag{4.2}$$

where each function h_i is from $\mathbf{R}\ n$ into $\mathbf{R}$ and is continuously differentiable and convex.

We are interested by the *active set* at a given point $x \in X$, which we define as

$$A(x) \overset{\text{def}}{=} \{i \in \{1, \ldots, n_c\} \mid h_i(x) = 0\}. \tag{4.3}$$

If the sequence of iterates $\{x_k\}$ converges to x_*, the question we wish to analyze can then be phrased as "Is $A(x_k) = A(x_*)$ for k large enough?". We note that linear equality constraints could be added to our description of the set X without altering what follows (see [37] for details).

We recall here that we defined a point x_* to be critical for problem (2.1) if and only if $-\nabla f(x_*) \in \mathcal{N}(x_*)$, where $\mathcal{N}(x_*)$ is the normal cone to X at the point $x_* \in X$. If

$$-\nabla f(x_*) \in \text{ri}\,[\mathcal{N}(x_*)]\,, \tag{4.4}$$

where the notation $\text{ri}\,[\mathcal{N}(x)]$ denotes the *relative interior* of the normal cone $\mathcal{N}(x_*)$ (see [36, Section 6], then the critical point x_* is said to be *non-degenerate* (see [14]).

AS.11 We assume that all limit points of the sequence $\{x_k\}$ are finite and non-degenerate.

If we additionally assume the stronger constraint qualification where

AS.12

$$\{\nabla h_i(x_*)\}_{i \in A(x_*)} \text{ are linearly independent for any limit point } x_*, \tag{4.5}$$

AS.11 is then equivalent to the existence of a set of strictly positive Lagrange multipliers at x_*. That is

$$\nabla f(x_*) = \sum_{i \in A(x_*)} \lambda_i \nabla h_i(x_*) \tag{4.6}$$

for some uniquely defined $\lambda_i > 0$.

4.2 Geometric analysis

Given these assumptions, we recall here some results obtained in [9] for the unique connected component of critical points of the problem containing a given critical point x_*, which we denote by C_*.

Lemma 12 *Assume AS.3, AS.4 and AS.10–AS.12 hold. For each connected component of critical points C_*, there exists a set $A(C_*) \subseteq \{1, \ldots, n_c\}$ such that $A(x_*) = A(C_*)$ for all $x_* \in C_*$.*

Interpreting this result in the case where X is polyhedral, this indicates that a connected component of critical points cannot "spread" over more than a single face. We next note that different connected components of critical points are "well separated". This result uses the notion of distance between a vector x and a set Y defined by

$$\operatorname{dist}(x, Y) \stackrel{\text{def}}{=} \inf_{y \in Y} \|x - y\|_2. \tag{4.7}$$

Lemma 13 *Assume AS.3, AS.4 and AS.10–AS.12 hold. There exists a constant $\psi \in (0, 1)$ such that $\operatorname{dist}(x_*, C'_*) \geq \psi$ for every critical point x_* and each connected component of critical points C'_* such that $A(C_*) \neq A(C'_*)$.*

The reference [9] uses a slightly more restrictive definition of distance where the "inf" is replaced by a "min", and causes the version of Lemma 13 presented therein to be restricted to compact connected components of critical points. The extension stated here easily results from the definition of connected components ([25, p. 54]) and will not be discussed in detail.

The next result states that if we consider bounded sequences whose limit points are critical, then each member of such a sequence lies in the neighbourhood

$$\mathcal{V}(C_*, \phi) \stackrel{\text{def}}{=} \{x \in \mathbf{R}\ n \mid \operatorname{dist}(x, C_*) \leq \phi\} \tag{4.8}$$

of a well defined connected component of critical points.

Lemma 14 *Assume AS.3, AS.4 and AS.10–AS.12 hold. Assume that $\{y_k\}$ is any bounded sequence of feasible points whose limit points are critical. Then there exist a $\phi_1 \in (0, \frac{1}{4}\psi)$, where ψ is as in Lemma 13, and a $k_1 \geq 0$ such that, for all $k \geq k_1$, there exists a connected set of critical points $C_*(y_k)$ such that*

$$y_k \in \mathcal{V}(C_*(y_k), \phi_1) \tag{4.9}$$

and

$$A(x) \subseteq A(C_*(y_k)) \tag{4.10}$$

for all $x \in \mathcal{V}(C_*(y_k), \phi_1) \cap X$.

We complete our geometric analysis by the following result.

Lemma 15 *Assume AS.3, AS.4 and AS.10–AS.12 hold. Consider a sequence $\{z_k\}$ of points in X such that*

$$\lim_{k\to\infty} \text{dist}(z_k, C_*) = 0 \tag{4.11}$$

for some connected component of critical points C_* and a sequence $\{y_k\}$ of points in X such that

$$\lim_{k\to\infty} \|z_k - y_k\|_{(k)} = 0 \tag{4.12}$$

and

$$A(y_k) = A(C_*) \tag{4.13}$$

for all k. Then

$$\lim_{k\to\infty} P_{\mathcal{T}(y_k)}(\nabla f(z_k)) = 0, \tag{4.14}$$

where $P_{\mathcal{T}(y_k)}(\cdot)$ is the orthogonal projection on the tangent cone at y_k.

4.3 Yet another criticality measure

Instead of entering in the details of a definition of $\bar{\alpha}_k$ suitable for the different kinds of constraints we might consider (bounds, polyhedral sets, general convex sets), we will instead assume the generic properties of this measure and then proceed along the lines described above. We now describe in detail our assumptions.

AS.14 We assume that $\bar{\alpha}_k \geq 0$ for all k and that $\bar{\alpha}_k = 0$ if and only if x_k is critical for problem (2.1). Furthermore, we will assume that the step s_k in our class of algorithms is computed to ensure that

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_2 \bar{\alpha}_k \min\left\{\frac{\bar{\alpha}_k}{\beta_k}, \Delta_{g,k}\right\}, \tag{4.15}$$

where we have re-used the constant $\kappa_2 \in (0, 1)$.

Because of (4.15) and (2.47) are identical, one can use the theory presented in the preceding Section with $\bar{\alpha}$ replacing α and therefore deduce the analog of Theorems 5 and (6), including (4.1), as required.

We now assume that our new criticality measure is bounded away from zero in the neighbourhood of critical points, so long as the correct active set has not been identified.

AS.15 Given ψ and ϕ_1 as in Lemma 14, there exists a $\phi_2 \in (0, \phi_1]$ and an $\bar{\alpha}_* > 0$ such that, if x_k belongs to $\mathcal{V}(C_*, \phi_2)$ for some connected component of critical points and $A(x_k + s_k) \subset A(C_*)$, then $\bar{\alpha}_k \geq \bar{\alpha}_*$.

Observe that, because of Lemma 14 and $\phi_2 \leq \phi_1$, one has that $C_* = C_*(x_k)$ and that $A(x_k) \subseteq A(C_*)$.

AS.14 and AS.15 are not as strong as they might appear at first sight. Indeed, they are satisfied by existing criticality measures in the literature. They typically depends on the generalized Cauchy point whose definition varies with the considered algorithm and the problem solved. For instance, a framework similar to that presented above is considered in [9]. The criticality measure used is defined by

$$\bar{\alpha}_k = \Big| \min_{\substack{x_k+d \in X_k^C \\ \|d\|_{(k)} \leq 1}} \langle g_k, d \rangle \Big|, \tag{4.16}$$

where

$$X_k^C \overset{\text{def}}{=} \bigcap_{i \in A(x_k^C)} X_i, \tag{4.17}$$

with x_k^C being the generalized Cauchy point and $X_i = \{x \in \mathbf{R}\ n \mid h_i(x) \geq 0\}$. The reader is referred to [9, Lemma 31] or to [37, Lemmas 29 and 35] for more details, and in particular for the proof that this measure indeed satisfies both AS.14 and AS.15. Other choices of $\bar{\alpha}$ satisfying the same properties are possible. For example,

$$\bar{\alpha}_k \overset{\text{def}}{=} \|P_{\mathcal{T}(x_k^C)}(-g_k)\| \tag{4.18}$$

is considered in [2] and in [10, Theorem 14] to measure the criticality of the generalized Cauchy point x_k^C.

4.4 Constraint identification

Before starting our constraint identification analysis, we also slightly strengthen our assumption on model choice by assuming that the model's and objective's gradients coincide.

AS.13

$$\nabla f_i(x_k) = g_{i,k} \tag{4.19}$$

for all $i \in \{1, \ldots, p\}$ and all $k \geq 0$.

This is not the weakest assumption one can make on the gradient accuracy in order to obtain constraint identification results, but AS.13 considerably simplifies the exposition. A weaker alternative is discussed in Section 5.

We start our finite constraint identification theory by stating a simple variation on Lemma 3 in our new framework.

Lemma 16 *Assume AS.3, AS.4 and AS.13 hold. Consider a sequence $\{x_k\}$ of iterates generated by the algorithm. Then, there exists a $c_6 > 1$ such that*

$$|f(x_k + s_k) - m_k(x_k + s_k)| \leq c_6 \beta_k \Delta_{g,k}^2 \tag{4.20}$$

for all k.

Proof. We first observe that, for each $i \in \{1, \ldots, p\}$ and for all k, (2.6), AS.13 and the definition (2.24) imply that

$$\begin{aligned} |f_i(x_k + s_k) - m_{i,k}(x_k + s_k)| &\leq \tfrac{1}{2}\|s_{i,k}\|_{(i,k)}^2 |\omega_{(i,k)}(f_i, x_k, s_{i,k}) - \omega_{(i,k)}(m_{i,k}, x_k, s_{i,k})| \\ &\leq \tfrac{1}{2}\|s_{i,k}\|_{(i,k)}^2 (|\omega_{(i,k)}(f_i, x_k, s_{i,k})| + |\omega_{(i,k)}(m_{i,k}, x_k, s_{i,k})|). \end{aligned} \tag{4.21}$$

We now sum all the element contributions and obtain

$$|f(x_k+s_k)-m_k(x_k+s_k)| \le \sum_{i=1}^{p} |f_i(x_k+s_k)-m_{i,k}(x_k+s_k)| \le \tfrac{1}{2}\sigma\nu^2 p(L+1)\beta_k\Delta_{g,k}^2, \quad (4.22)$$

where we used (2.27), (2.28), (2.74) and the inequality $L+\beta_k \le (L+1)\beta_k$. This yields (4.20) with

$$c_6 \overset{\rm def}{=} \max\{1, \tfrac{1}{2}\sigma\nu^2 p(L+1)\}. \quad (4.23)$$

□

We also show that, for $\beta_k\Delta_{g,k}$ small, the iteration must be successful at a noncritical point.

Lemma 17 *Assume AS.3, AS.4, AS.13 and AS.14 hold, and that $\bar{\alpha}_k > 0$ and that*

$$\beta_k\Delta_{g,k} \le \frac{\kappa_2\bar{\alpha}_k\gamma_1(1-\eta_1)}{c_6} \quad (4.24)$$

for some $k \ge k_1$. Then iteration k is successful.

Proof. Observe first that

$$\frac{\kappa_2\gamma_1(1-\eta_1)}{c_6} \le 1 \quad (4.25)$$

where we used (2.48), (2.49), and the inequalities $\kappa_2 < 1$ and $c_6 > 1$. But this inequality and (4.15) then imply that

$$\delta m_k \ge \kappa_2\bar{\alpha}_k\Delta_{g,k}. \quad (4.26)$$

We can then verify that, because of (4.24),

$$\left|\frac{\delta f_k}{\delta m_k} - 1\right| = \frac{|f(x_k+s_k)-m_k(x_k+s_k)|}{\delta m_k} \le \frac{c_6}{\kappa_2\bar{\alpha}_k}\beta_k\Delta_{g,k} \le 1-\eta_1, \quad (4.27)$$

which implies (2.53) and hence proves the lemma. □

We are now ready to prove our first identification result, namely that the maximal active set is identified by a subsequence of iterates.

Theorem 18 *Assume AS.1–AS.7 and AS.9–AS.15 hold. Consider the sequence $\{x_k\}$ of iterates generated by the algorithm. Then there exists a subsequence $\{k_j\}$ of successful iterates such that*

$$A(x_{k_j}) = A_*, \quad (4.28)$$

where A_* is the maximal (largest) active set defined by any limit points of the sequence $\{x_k\}$.

Proof. We define the subsequence $\{k_j\}$ as the sequence of successful iterations whose iterates approach limit points with active set equal to A_*, that is

$$\{k_j\} \overset{\rm def}{=} \{k \in \mathcal{S} \mid A(C_*) = A_* \text{ and } \operatorname{dist}(x_k, C_*) \le \phi_2\}, \quad (4.29)$$

where C_* is the a connected component of critical points with maximal active set. We also assume, for the purpose of obtaining a contradiction, that

$$A(x_{k_j+1}) \neq A_* \tag{4.30}$$

for all j large enough. Because of Lemma 14, our definition of A_* and $k_j \in \mathcal{S}$, we deduce that

$$A(x_{k_j} + s_{k_j}) \subset A_* \tag{4.31}$$

for j sufficiently large. But AS.15 then implies that

$$\bar{\alpha}_k \geq \bar{\alpha}_* \tag{4.32}$$

for all $k \in \{k_j\}$ sufficiently large. Now, using (4.15) and (2.53), we obtain that

$$\beta_{k_j}[f(x_{k_j}) - f(x_{k_j+1})] = \beta_{k_j}\delta f_{k_j} \geq \eta_1\kappa_2\bar{\alpha}_* \min\{\bar{\alpha}_*, \beta_{k_j}\Delta_{g,k_j}\} \tag{4.33}$$

for j large enough. Because the left hand-side of this inequality must converge to zero as a result of AS.9, we have that

$$\|s_{k_j}\|_{(k_j)} \leq \nu\Delta_{g,k_j} < \phi_2 \leq \frac{1}{4}\psi \tag{4.34}$$

for j larger than $j_1 \geq 1$, say. But this last inequality, Lemma 13 and Lemma 14 imply that x_{k_k+1} cannot jump to the neighbourhood of any other connected component of critical points with a different active set, and hence that x_{k_j+1} belongs to $\mathcal{V}(C_*, \phi_2)$ again. The same property also holds for the next successful iterate, x_{k_j+q} say, and we have that $C_*(x_{k_j+q}) = C_*$ for all $q \geq 0$. Therefore, the subsequence $\{k_j\}$ is identical to the complete sequence of successful iterates with $k \geq k_{j_1}$. Hence we may deduce from (4.33) that

$$\lim_{k\to\infty} \beta_k\Delta_{g,k} = 0. \tag{4.35}$$

As a consequence of (2.74), we deduce that, for k large enough, x_k and $x_k + s_k$ both belong to $\mathcal{V}(C_*, \phi_2)$.

The next step in our proof is to show that ultimately all iterates must be successful. Suppose therefore that this is not the case. One can therefore find a subsequence K such that

$$k \notin \mathcal{S} \text{ and } k+1 \in \mathcal{S}. \tag{4.36}$$

for all $k \in K$. Note that, because of (4.35), (2.75) and the nondecreasing nature of the sequence $\{\beta_k\}$, one has that

$$\beta_k\Delta_{g,k} \leq \frac{1}{\gamma_1}\beta_{k+1}\Delta_{g,k+1} \leq \frac{\kappa_2\gamma_1\bar{\alpha}_*(1-\eta_1)}{2c_6} \tag{4.37}$$

for k sufficiently large. Now, if one has that

$$A(x_k + s_k) \subset A_*, \tag{4.38}$$

then AS.15 implies that $\bar{\alpha}_k \geq \bar{\alpha}_*$ and we may thus apply Lemma 17 to deduce from (4.37) that $k \in \mathcal{S}$, which contradicts (4.36). Hence (4.38) cannot hold and we must have that

$$A(x_k + s_k) = A_* \tag{4.39}$$

for all $k \in K$ sufficiently large. Observe now that, since $k \notin \mathcal{S}$, one has that $x_{k+1} = x_k$ and $g_{i,k+1} = g_{i,k}$ for all $i \in \{1, \ldots, p\}$. Hence, for all such i,

$$\begin{aligned} m_{i,k+1}(x_{k+1} + s_{k+1}) - m_{i,k}(x_k + s_k) \;&=\; \langle g_{i,k}, s_{k+1} - s_k \rangle - \tfrac{1}{2}\|s_{i,k}\|^2_{(k)} \omega_{(i,k)}(m_{i,k}, x_k, s_{i,k}) \\ &\quad + \tfrac{1}{2}\|s_{i,k+1}\|^2_{(k+1)} \omega_{(i,k+1)}(m_{i,k+1}, x_{k+1}, s_{i,k+1}). \end{aligned} \tag{4.40}$$

Summing over all elements and using (2.74), (2.75) and the definition of β_k, we obtain that

$$\begin{aligned} m_{k+1}(x_{k+1} + s_{k+1}) - m_k(x_k + s_k) \;&\geq\; \langle g_k, s_{k+1} - s_k \rangle - \tfrac{1}{2}\nu^2\sigma^2 p \left[\beta_{k+1}\Delta^2_{g,k+1} + \beta_k \Delta^2_{g,k}\right] \\ &\geq\; \langle -g_k, s_k - s_{k+1} \rangle - \tfrac{1}{2}\nu^2\sigma^2 p \left[1 + \tfrac{1}{\gamma_1^2}\right] \beta_{k+1}\Delta^2_{g,k+1}. \end{aligned} \tag{4.41}$$

We now note that, using the Moreau decomposition, $-g_k$ is given by

$$-g_k = P_{\mathcal{T}(x_k+s_k)}(-g_k) + P_{\mathcal{N}(x_k+s_k)}(-g_k), \tag{4.42}$$

and, since $k \notin \mathcal{S}$,

$$s_{k+1} - s_k = x_{k+1} + s_{k+1} - (x_k + s_k) \in \mathcal{T}(x_k + s_k), \tag{4.43}$$

which is the polar of $\mathcal{N}(x_k + s_k)$. Now from (2.74) and (2.75), we obtain that, for all $k \in K$ large enough,

$$\begin{aligned} \langle -g_k, s_k - s_{k+1} \rangle \;&=\; \langle P_{\mathcal{T}(x_k+s_k)}(-g_k), s_k - s_{k+1} \rangle + \langle P_{\mathcal{N}(x_k+s_k)}(-g_k), s_k - s_{k+1} \rangle \\ &\geq\; -\|P_{\mathcal{T}(x_k+s_k)}(-g_k)\|_{[k]} \|s_k - s_{k+1}\|_{(k)} \\ &\qquad -\langle P_{\mathcal{N}(x_k+s_k)}(-g_k), P_{\mathcal{T}(x_k+s_k)}(s_{k+1} - s_k) \rangle \\ &\geq\; -\|P_{\mathcal{T}(x_k+s_k)}(-g_k)\|_{[k]} \|s_k - s_{k+1}\|_{(k)} \\ &\geq\; -\nu \|P_{\mathcal{T}(x_k+s_k)}(-g_k)\|_{[k]} (\Delta_{g,k} + \Delta_{g,k+1}) \\ &\geq\; -\nu(1 + \tfrac{1}{\gamma_1}) \|P_{\mathcal{T}(x_k+s_k)}(-g_k)\|_{[k]} \Delta_{g,k+1}. \end{aligned} \tag{4.44}$$

Combining this last inequality with (4.41) gives that

$$\begin{aligned} & m_{k+1}(x_{k+1} + s_{k+1}) - m_k(x_k + s_k) \geq \\ & \quad -\Delta_{g,k+1} \left[(1 + \frac{1}{\gamma_1})\nu \|P_{\mathcal{T}(x_k+s_k)}(-g_k)\|_{[k]} + \frac{\nu^2\sigma^2 p}{2}(1 + \frac{1}{\gamma_1^2})\beta_{k+1}\Delta_{g,k+1} \right]. \end{aligned} \tag{4.45}$$

for all $k \in K$ sufficiently large. Now observe that, because of (2.74) and (2.75),

$$\|s_k\|_{(k)} \leq \Delta_{g,k} \leq \frac{1}{\gamma_1}\Delta_{g,k+1}. \tag{4.46}$$

Therefore, from (4.35), s_k tends to zero. Using (4.39), we may now apply Lemma 15 with $y_k = x_k + s_k$ to the subsequence K and deduce from (4.45) that

$$m_{k+1}(x_{k+1} + s_{k+1}) - m_k(x_k + s_k) \geq -\tfrac{1}{2}\kappa_2 \bar{\alpha}_* \Delta_{g,k+1} \tag{4.47}$$

for all $k \in K$ sufficiently large, where we also used (4.35). On the other hand, AS.15, (4.15), (4.36), the fact that (4.32) holds for all successful iterations and (4.35) imply that

$$\delta m_{k+1} \geq \kappa_2 \bar{\alpha}_* \Delta_{g,k+1}. \tag{4.48}$$

for $k \in K$ large enough. We therefore obtain, for such k, that

$$\delta m_k = \delta m_{k+1} + m_{k+1}(x_{k+1} + s_{k+1}) - m_k(x_k + s_k) \geq \tfrac{1}{2}\kappa_2 \bar{\alpha}_* \Delta_{g,k+1} \geq \tfrac{1}{2}\kappa_2 \gamma_1 \bar{\alpha}_* \Delta_{g,k}, \tag{4.49}$$

where we again used (2.75) to deduce the last inequality. This, together with Lemma 16 and (4.37), yields that

$$\left| \frac{\delta f_k}{\delta m_k} - 1 \right| = \frac{|f(x_k + s_k) - m_k(x_k + s_k)|}{\delta m_k} \leq \frac{2c_6}{\kappa_2 \gamma_1 \bar{\alpha}_*} \beta_k \Delta_{g,k} \leq 1 - \eta_1. \tag{4.50}$$

But this implies again that iteration k is successful, which is impossible because of (4.36). Hence (4.36) cannot hold for sufficiently large k. This in turn implies that all iterations are eventually successful and that the sequence $\{k_j\}$ defined at the beginning of the proof is the complete sequence of all iterates. The limit (4.35) then contradicts Theorem 5 (with $\bar{\alpha}_k$ playing the role of α_k). As a consequence our initial assumption is impossible and the theorem is proved. □

We now prove that, for large enough k, once found, the correct active set cannot be abandoned.

Theorem 19 *Assume AS.1–AS.7 and AS.9–AS.15 hold. Then there exists a unique active set A_* such that*

$$A(x_*) = A_* \tag{4.51}$$

for all limit points x_* of the sequence $\{x_k\}$. Furthermore

$$A(x_k) = A_* \tag{4.52}$$

for all sufficiently large k.

Proof. Let $\{k_i\}$ be the subsequence of successful iterates such that (4.28) holds, as given by Theorem 18. Assume furthermore that this subsequence is restricted to sufficiently large k, that is $k_i \geq k_1$ for all i. Assume finally that there exists a subsequence of $\{k_i\}$, $\{k_j\}$ say, such that for each j there is an ℓ_j with

$$\ell_j \in A(x_{k_j}) \text{ and } \ell_j \notin A(x_{k_j+1}). \tag{4.53}$$

Because $k_j \in \mathcal{S}$, we deduce that

$$\ell_j \notin A(x_{k_j} + s_{k_j}). \tag{4.54}$$

Now Lemma 14 and (4.28) together with the maximality of the connected component A_* imply that $A(C_*(x_{k_j})) = A_*$. Now AS.15 and (4.54) then ensure that

$$\bar{\alpha}_{k_j} \geq \bar{\alpha}_* \tag{4.55}$$

for all j. Combining this inequality with (4.15), one obtains again that

$$\beta_{k_j}\delta f_{k_j} \geq \eta_1\kappa_2\bar{\alpha}_*\min\{\bar{\alpha}_*, \beta_{k_j}\Delta_{g,k_j}\}. \tag{4.56}$$

Using AS.9, we then deduce that

$$\lim_{j\to\infty}\beta_{k_j}\Delta_{g,k_j} = 0. \tag{4.57}$$

As a consequence, we may deduce from (4.15) that

$$\delta m_{k_j} \geq \kappa_2\bar{\alpha}_*\Delta_{g,k_j} \tag{4.58}$$

for all j sufficiently large. On the other hand, we have that, for all $i \in \{1,\ldots,p\}$ and all k,

$$\delta m_{i,k} = \langle g_{i,k}, s_k\rangle + \tfrac{1}{2}\|s_{i,k}\|^2_{(i,k)}\omega_{(i,k)}(m_k, x_k, s_{i,k}) \tag{4.59}$$

because of (2.24) and the fact that $g_{i,k} \in \mathcal{R}_i$. Summing on all elements, we obtain that

$$\begin{aligned}\delta m_k &\leq |\langle g_k, s_k\rangle| + \tfrac{1}{2}\textstyle\sum_{i=1}^p \|s_{i,k}\|^2_{(i,k)}\omega_{(i,k)}(m_k, x_k, s_{i,k})\\ &\leq \|P_{\mathcal{T}(x_k)}(-g_k)\|_{[k]}\|s_k\|_{(k)} + \tfrac{1}{2}\sigma^2 p\beta_k\|s_k\|^2_{(k)}\\ &\leq \nu\|P_{\mathcal{T}(x_k)}(-g_k)\|_{[k]}\Delta_{g,k} + \tfrac{1}{2}\sigma^2\nu^2 p\beta_k\Delta^2_{g,k},\end{aligned} \tag{4.60}$$

where we used (2.14), (2.15), (2.74) and the fact that $s_k \in \mathcal{T}(x_k)$. Combining (4.58) and (4.59) (with $k = k_j$) and dividing by Δ_{g,k_j} yields that

$$\kappa_2\bar{\alpha}_* \leq \nu\|P_{\mathcal{T}(x_k)}(-g_k)\|_{[k]} + \tfrac{1}{2}\sigma^2\nu^2 p\beta_k\Delta_{g,k}. \tag{4.61}$$

Assuming that the sequence $\{x_{k_j}\}$ converges to some x_* in some C_* (or taking a further subsequence if necessary) and using Lemma 15 on the subsequence $\{k_j\}$ (with $y_k = x_{k_j}$) and (4.57), we deduce that (4.61) is impossible because its left-hand-side is a positive constant and its right-hand-side tends to zero. Hence, no such subsequence $\{k_j\}$ exists and the maximality of A_* then implies that

$$A_* = A(x_{k_i+1}) \tag{4.62}$$

for all i large enough. Therefore

$$A(x_{k_i+q}) = A_* \tag{4.63}$$

for i sufficiently large, where $k_i + q$ is the index of the next successful iteration after iteration k_i. Hence $k_i + q \in \{k_i\}$. Using this argument repeatedly, we thus deduce that $\{k_i\}$ is the sequence of all successful iteration with sufficiently large index. As a consequence, $A(x_k) = A_*$ for all such k, which proves (4.52). Moreover, A_* is then the only possible active set for all limit points, which gives (4.51). □

5 Extensions

We examine in this section some extensions and variants of the results presented above.

5.1 A hybrid technique

One of the possible drawbacks of our Algorithm is that steps might be constrained to be unnecessarily small in directions corresponding to very nonlinear element functions. Indeed, the negative effect of inaccurate models for these elements might be compensated by a successful step in directions corresponding to less nonlinear elements. This compromise between the different parts of the objective is, of course, inherent to the classical method using a single trust region.

We might try to obtain the best of both classical and structured approaches by using a hydrid technique. In this technique, a *global* trust region radius Δ_k is recurred for the objective function considered as a single element (using the Algorithm given above, which is then equivalent to the classical one), along with the individual radii $\Delta_{i,k}$. We then define the individual "hybrid" radii by

$$\Delta_{i,k}^h \stackrel{\rm def}{=} \max\{\Delta_k, \Delta_{i,k}\} \tag{5.1}$$

for each $i \in \{1, \ldots, p\}$ and redefine $B_{i,k}$ as

$$B_{i,k} \stackrel{\rm def}{=} \{x \in X \mid \|P_{\mathcal{R}_i}(x - x_k)\|_{(i,k)} \leq \nu_1 \Delta_{i,k}^h\}, \tag{5.2}$$

Similarly, $\Delta_{g,k}$ is then replaced by

$$\Delta_{g,k} \stackrel{\rm def}{=} \max_{i \in D_k} \Delta_{i,k}^h. \tag{5.3}$$

We can then apply our Algorithm with these new quantities, to the effect that gentle elements have their associated trust regions possibly extended without having to contract those corresponding to more nonlinear ones, as long as the global result is satisfactory.

It is not difficult to verify that the theory presented above still holds for his hybrid modification. The key points are to observe that the definition of $\Delta_{g,k}$ in (5.3) implies that

$$\delta m_k \geq \kappa_2 \alpha_k \min\left\{\frac{\alpha_k}{\beta_k}, \Delta_k\right\}, \tag{5.4}$$

which is the classical sufficient decrease condition (2.44), that the inequalities (2.75) are still valid with $\Delta_{i,k}$ replaced by $\Delta_{i,k}^h$, and also that an analogous to Theorem 5 also holds for the global trust region radius, as is already well known from the single trust region case (see [9], for instance).

Some extremely preliminary numerical tests indicate that this modification might be computationally advantageous compared to both the single trust region case and the original formulation of Section 2.3.

5.2 An alternative definition of success

An immediate consequence of inequality (2.73) in Lemma 1 is that it would be possible to replace the condition for an iteration to be successful (2.53) by

$$\delta f_k \geq \eta_1 \sum_{i \in M_k} \delta m_{i,k}(x_k), \tag{5.5}$$

without altering the developments presented above. Indeed, (2.73) shows the equivalence (2.53) and (5.5). We have chosen to use (2.53) above, because it seems natural to consider the same collection of elements on both sides of the inequality.

5.3 Weaker sufficient decrease conditions

It is remarkable to note that Theorems 5 and 6 can be proved in a weaker context. Indeed, we could define the overall trust region as

$$B_k = B_k^{\square}, \tag{5.6}$$

(a possibly larger region than that defined by (2.22)), require the weaker sufficient decrease condition

$$\delta m_k \geq \kappa_2 \alpha_k \min\left\{\frac{\alpha_k}{\beta_k}, \Delta_{min,k}\right\} \tag{5.7}$$

instead of (2.47), and still prove Theorems 5 and 6. However, we have not been able to prove Theorem 9 nor active constraint identification with these assumptions, because (5.7) only controls the length of the step in a possibly small subspace of $\mathbf{R}\ n$. If we replace (5.7) by the stronger condition

$$\delta m_k \geq \kappa_2 \alpha_k \min\left\{\frac{\alpha_k}{\beta_k}, \max\left[\Delta_{min,k}, \nu_3 \|s_k\|_{(k)}\right]\right\} \tag{5.8}$$

for some $\nu_3 > 0$, it is then possible to obtain the conclusion of Theorem 9 as well , namely that all limit points of the sequence of iterates are critical. But the strengthening introduced by (5.8) has not been sufficient for the authors to prove active constraint identification for general convex constraints.

5.4 Inexact gradients and constraint identification

As we have mentioned already, AS.13 is not the weakest possible assumption for proving finite active constraint identification. Weaker conditions are presented in [37]. As shown in this reference, it is sufficient to assume that all limit points of the sequence of iterates converges to a single limit point x_* and that the sequence of approximate gradients itself has a single limit point g_* such that

$$-g_* \in \mathrm{ri}\left[\mathcal{N}(x_*)\right]. \tag{5.9}$$

The technique of proof for this extension is very similar to that discussed above.

5.5 Constraint identification without linear independence of constraints normals

The linear independence assumption AS.12 can be somewhat restrictive in practice, especially for problems where X is a polyhedron defined by many linear inequality constraints. Fortunately, the "weak constraint identification" result of [9] can be applied in our context when AS.12 does not hold. This result implies the following useful consequence.

Theorem 20 *Assume AS.1–AS.7, AS.9–AS.11 and AS.13–AS.15 hold. Assume also that all functions h_i of (4.2) are linear and consider a sequence $\{x_k\}$ of iterates generated by the Algorithm. Then the active constraints are identified (i.e. (4.52) holds) for all k sufficiently large.*

We refer the reader to [9] or [37] for further details.

5.6 Convergence to a single limit point

As the results obtained above for structured trust regions are identical to those obtained in the unstructured case in [9] and by Sartenaer in [37], the theory developed in these papers for the convergence of the iterates to a minimizer (as opposed to a mere critical point) holds with only minor modifications. The main result is the following.

Theorem 21 *Assume AS.1–AS.7 and AS.9–AS.15 hold. Consider a sequence $\{x_k\}$ of iterates generated by the Algorithm. Assume that there are infinitely many successful iterations and that there exists an $\epsilon > 0$ such that*

$$\liminf_{\substack{k\in\mathcal{S}\\ k\to\infty}} \min_{i|s_{i,k}\neq 0} \omega_{(i,k)}(m_{i,k}, x_k, s_{i,k}) \geq \epsilon. \tag{5.10}$$

Assume furthermore that all limit points of the sequence $\{s_k\}_{k\in\mathcal{S}}$ belong to the subspace $\sum_{i=1}^p \mathcal{R}_i$. Assume finally that, for some limit point x_*, $\nabla^2 f(x_*)$ is positive definite on the tangent plane to X at x_*. Then

$$\lim_{k\to\infty} x_k = x_*. \tag{5.11}$$

Proof. The only modification needed to apply the theory developed in [9] and [37] is to deduce from our assumptions that $\omega_{(k)}(m_k, x_k, s_k)$ is asymptotically bounded away from zero. To obtain this result, we first note that our condition on the limit points of the sequence $\{s_k\}$ implies that, for $k \in \mathcal{S}$ sufficiently large,

$$\|s_k\|_2^2 \leq 2\sum_{i=1}^p \|s_{i,k}\|_2^2. \tag{5.12}$$

We can then deduce, for $k \in \mathcal{S}$ large enough, that

$$\begin{array}{rcl} \omega_{(k)}(m_k, x_k, s_k) & = & 2\left[\sum_{i=1}^p m_{i,k}(x_k+s_k) - \sum_{i=1}^p m_{i,k}(x_k) - \langle \sum_{i=1}^p g_i(x_k), s_k\rangle\right] / \|s\|_{(k)}^2 \\ & = & \sum_{i=1}^p \left[\omega_{(i,k)}(m_{i,k}, x_k, s_{i,k}) \|s_{i,k}\|_{i,k}^2 / \|s_k\|_{(k)}^2\right] \\ & \geq & \min_{i|s_{i,k}\neq 0} \omega_{(i,k)}(m_{i,k}, x_k, s_{i,k}) \left[\sum_{i=1}^p \|s_{i,k}\|_2^2\right] / \sigma^4 \|s_k\|_2^2 \\ & \geq & \min_{i|s_{i,k}\neq 0} \omega_{(i,k)}(m_{i,k}, x_k, s_{i,k}) / 2\sigma^4 \end{array} \tag{5.13}$$

where we successively used (2.24), (2.25), (2.14) and (5.12). Combining (5.13) and (5.10) yields that

$$\liminf_{\substack{k\in\mathcal{S}\\ k\to\infty}} \omega_{(k)}(m_k, x_k, s_k) \geq \frac{\epsilon}{2\sigma^4}, \tag{5.14}$$

which is the desired bound. □

Note that both $f(\cdot)$ and all models $m_{i,k}(\cdot)$ are invariant for any component of the step in the subspace $(\sum_{i=1}^{p} \mathcal{R}_i)^{\perp}$. Our assumption on the limit points of successful steps is thus very weak, because any nonzero component of the steps along this subspace is irrelevant for the minimization the objective function. Moreover, sensible implementations of the algorithm would typically yield that

$$s_k \in \sum_{i=1}^{p} \mathcal{R}_i. \tag{5.15}$$

This last condition would, for instance, be automatically fulfilled if s_k is the classical Newton's step $-[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$ or, more generally, any step chosen in a Krylov subspace derived from $\nabla f(x_k)$ and $\nabla^2 f(x_k)$, as would be the case with a truncated conjugate gradient technique. Our assumption is however necessary because, if the sequence of successful step has a limit point with a nonzero component along that subspace, then the sequence of iterates then remains in the subspace, preventing convergence of the algorithm to a single limit point.

As in [37], we can also deduce the convergence of the iterates to a single critical point whenever the feasible set is polyhedral.

Theorem 22 *Assume AS.1–AS.7 and AS.9–AS.15 hold. Consider a sequence $\{x_k\}$ of iterates generated by the Algorithm. Assume that there are infinitely many successful iterations, that all limit points of the sequence $\{s_k\}_{k \in \mathcal{S}}$ belong to $\sum_{i=1}^{p} \mathcal{R}_i$ and that there exists an $\epsilon > 0$ such that (5.10) holds. Assume furthermore that, for some limit point x_*, $\nabla^2 f(x_*)$ is nonsingular on the tangent plane to X at x_* and that X is polyhedral. Then*

$$\lim_{k \to \infty} x_k = x_*. \tag{5.16}$$

Again, we only need to deduce (5.14) from (5.10) and (5.15) to use the proof of [37]. This last results shows that convergence can occur to a critical point which is not a minimizer if the element models are asymptotically uniformly convex.

5.7 Noisy functions

In contrast to the description of [9], we have not extended in the present paper the application of trust region to noisy functions. However we believe this extension to be possible.

6 Conclusions

We have shown in this paper that the trust region concept, one of the most powerful tools for building efficient and robust algorithms for optimization, can be extended in a very natural way to reflect the structure of the underlying problem. The algorithm proposed above is indeed a direct generalization of the more usual case where only a single uniform trust region is considered. Similar global convergence properties can be proved for the new algorithm, including the case where dynamic scaling is performed on the variables and the situation where the gradients are only known approximately.

It remains to see if this modification of a trust region algorithm will prove efficient in practice and justify the slight additional complexity of the method. We will report on this issue later. However, we note that the results of preliminary numerical experiments have been extremely encouraging, especially those based on the hybrid method proposed in Section 5.1.

One of the nice features of the partially separable functions considered in the present theory is that the objective is a *linear* combination of its elements. While group partially separability, as used in [12] or [13], has computational advantages in terms of economy of derivative calculation, this structure involves a *nonlinear* relationship between the elements and the overall function. This makes exploiting the link between local and global models much harder. While we would be interested in deriving structured trust-region methods for group partially separable functions, the methods would undoubtedly be more complicated and less amenable to analysis. Thus, we are content, in the present paper, to consider the simpler, but nonetheless very general, partially separable structure.

Finally, there might be other ways to introduce structure in trust region methods than considering (group) partially separable objective functions. In particular, trust region methods for nonlinearly constrained problems seems attractive candidates for an alternative approach that would separate the trust region(s) on the objective from those on the constraints.

References

[1] J. V. Burke and J. J. Moré. On the identification of active constraints. *SIAM Journal on Numerical Analysis*, 25:1197–1211, 1988.

[2] J. V. Burke, J. J. Moré, and G. Toraldo. Convergence properties of trust region methods for linear and convex constraints. *Mathematical Programming, Series A*, 47(3):305–336, 1990.

[3] R. H. Byrd, E. Eskow, R. B. Schnabel, and S. L. Smith. Parallel global optimization: numerical methods, dynamic scheduling methods, and application to molecular configuration. Technical Report CU-CS-553-91, Department of Computer Science, University of Colorado at Boulder, Boulder, USA, 1991.

[4] R. H. Byrd, R. B. Schnabel, and G. A. Schultz. A trust region algorithm for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis*, 24:1152–1170, 1987.

[5] P. H. Calamai and J. J. Moré. Projected gradient methods for linearly constrained problems. *Mathematical Programming*, 39:93–116, 1987.

[6] R. G. Carter. On the global convergence of trust region methods using inexact gradient information. *SIAM Journal on Numerical Analysis*, 28(1):251–265, 1991.

[7] M. R. Celis, J. E. Dennis, and R. A. Tapia. A trust region strategy for nonlinear equality constrained optimization. In P. T. Boggs, R. H. Byrd, and R. B. Schnabel, editors, *Numerical Optimization 1984*, pages 71–82, 1985.

[8] A. R. Conn, N. I. M. Gould, M. Lescrenier, and Ph. L. Toint. Performance of a multifrontal scheme for partially separable optimization. In *Advances in numerical partial differential equations and optimization, Proceedings of the sixth Mexico-United States Workshop*, Philadelphia, USA, 1992. SIAM.

[9] A. R. Conn, N. I. M. Gould, A. Sartenaer, and Ph. L. Toint. Global convergence of a class of trust region algorithms for optimization using inexact projections on convex constraints. *SIAM Journal on Optimization*, (to appear), 1992.

[10] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM Journal on Numerical Analysis*, 25:433–460, 1988. See also same journal 26:764–767, 1989.

[11] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Testing a class of methods for solving minimization problems with simple bounds on the variables. *Mathematics of Computation*, 50:399–430, 1988.

[12] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. An introduction to the structure of large scale nonlinear optimization problems and the LANCELOT project. In R. Glowinski and A. Lichnewsky, editors, *Computing Methods in Applied Sciences and Engineering*, pages 42–54, Philadelphia, USA, 1990. SIAM.

[13] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. LANCELOT*: a Fortran package for large-scale nonlinear optimization (Release A)*. Number 17 in Springer Series in Computational Mathematics. Springer Verlag, Heidelberg, Berlin, New York, 1992.

[14] J. C. Dunn. On the convergence of projected gradient processes to singular critical points. *Journal of Optimization Theory and Applications*, 55:203–216, 1987.

[15] M. El-Alem. A global convergence theory for the Dennis-Celis-Tapia trust-region algorithm for constrained optimization. *SIAM Journal on Numerical Analysis*, 28(1):266–290, 1991.

[16] R. Fletcher. *Practical Methods of Optimization: Unconstrained Optimization.* J. Wiley and Sons, New-York, 1980.

[17] D. M. Gay. Is exploiting partial separability useful? Talk at the 6th Mexico United States Workshop, Oaxaca, Mexico, January 1992.

[18] D. Goldfarb and S. Wang. Partial-update Newton methods for unary factorable and partially separable optimization. *SIAM Journal on Optimization*, (to appear), 1989.

[19] A. Griewank. The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients. *Mathematical Programming, Series A*, 50(2), 1991.

[20] A. Griewank and Ph. L. Toint. On the unconstained optimization of partially separable functions. In M. J. D. Powell, editor, *Nonlinear Optimization 1981*, pages 301–312, London and New York, 1982. Academic Press.

[21] A. Griewank and Ph. L. Toint. Numerical experiments with partially separable optimization problems. In D. F. Griffiths, editor, *Numerical Analysis: Proceedings Dundee 1983*, pages 203–220, Berlin, 1984. Springer Verlag. Lecture Notes in Mathematics 1066.

[22] A. Griewank and Ph. L. Toint. On the existence of convex decomposition of partially separable functions. *Mathematical Programming*, 24:25–49, 1984.

[23] W. A. Gruver and E. Sachs. *Algorithmic methods in optimal control.* Pitman, Boston, USA, 1980.

[24] M. D. Hebden. An algorithm for minimization using exact second derivatives. Technical Report T. P. 515, AERE Harwell Laboratory, Harwell, UK, 1973.

[25] J. L. Kelley. *General Topology.* Springer Verlag, Heidelberg, Berlin, New York, 1955.

[26] M. Lescrenier. Partially separable optimization and parallel computing. In R. R. Meyer and S. Zenios, editors, *Parallel Optimization on Novel Computer Architectures*, Switzerland, 1989. A. C. Baltzer Scientific Publishing Co.

[27] M. Lescrenier. Convergence of trust region algorithms for optimization with bounds when strict complementarity does not hold. *SIAM Journal on Numerical Analysis*, 28(2):476–495, 1991.

[28] K. Levenberg. A method for the solution of certain problems in least squares. *Quarterly Journal on Applied Mathematics*, 2:164–168, 1944.

[29] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11:431–441, 1963.

[30] J. J. Moré. The Levenberg-Marquardt algorithm: implementation and theory. In G. A. Watson, editor, *Proceedings Dundee 1977*, Berlin, 1978. Springer Verlag. Lecture Notes in Mathematics.

[31] J. J. Moré. Recent developments in algorithms and software for trust region methods. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical Programming: The State of the Art*, pages 258–287, Berlin, 1983. Springer Verlag.

[32] J. J. Moré. Trust regions and projected gradients. In M. Iri and K. Yajima, editors, *System Modelling and Optimization*, volume 113, pages 1–13, Berlin, 1988. Springer Verlag. Lecture Notes in Control and Information Sciences.

[33] M. J. D. Powell. A new algorithm for unconstrained optimization. In J. B. Rose, O. L. Mangasarian, and K. Ritter, editors, *Nonlinear Programming*, New York, 1970. Academic Press.

[34] M. J. D. Powell. On the global convergence of trsut region algorithms for unconstrained optimization. *Mathematical Programming*, 29:297–303, 1984.

[35] M. J. D. Powell and Y. Yuan. A trust tregion algorithm for equality constrained optimization. *Mathematical Programming*, 49(2), 1990.

[36] R. T. Rockafellar. *Convex Analyis.* Princeton University Press, Princeton, USA, 1970.

[37] A. Sartenaer. *On some strategies for handling constraints in nonlinear optimization.* PhD thesis, Department of Mathematics, FUNDP, Namur, Belgium, 1991.

[38] Ph. L. Toint. Towards an efficient sparsity exploiting Newton method for minimization. In I. S. Duff, editor, *Sparse Matrices and Their Uses*, London, 1981. Academic Press.

[39] Ph. L. Toint. Global convergence of the partitioned BFGS algorithm for convex partially separable optimization. *Mathematical Programming*, 36(3):290–307, 1986.

[40] Ph. L. Toint. On large scale nonlinear least squares calculations. *SIAM Journal on Scientific and Statistical Computing*, 8(3):416–435, 1987.

[41] Ph. L. Toint. Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space. *IMA Journal of Numerical Analysis*, 8:231–252, 1988.

[42] A. Vardi. A trust region algorithm for equality constrained minimization: convergence properties and implementation. *SIAM Journal on Numerical Analysis*, 22(3):575–591, 1985.