A bundle filter method for nonsmooth nonlinear optimization *

R. Fletcher and S. Leyffer[†]

22 December 1999

Abstract

We consider minimizing a nonsmooth objective subject to nonsmooth constraints. The nonsmooth functions are approximated by a bundle of subgradients. The novel idea of a filter is used to promote global convergence.

Keywords: Nonsmooth optimization, bundle method, filter method.

1 Introduction

This paper is concerned with nonsmooth optimization problems where a nonsmooth objective is minimized subject to a nonsmooth constraint. This type of problem can be stated as

$$(P) \left\{ \begin{array}{ll} \underset{x}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & c(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{x} \in X. \end{array} \right.$$

Throughout the paper, the following assumptions are made:

A1 $X \subset \mathbb{R}$ is a bounded polyhedral set.

A2 f, c are convex, possibly nonsmooth, locally Lipschitz continuous functions from $\mathbb{R}^n$ to $\mathbb{R}$.

A3 For every $\mathbf{x}^{(k)} \in X$ we can evaluate $f^{(k)} = f(\mathbf{x}^{(k)})$, $c^{(k)} = c(\mathbf{x}^{(k)})$ and one arbitrary element of their respective generalized gradients $\mathbf{g}^{(k)} \in \partial f^{(k)} = \partial f(\mathbf{x}^{(k)})$ and $\mathbf{a}^{(k)} \in \partial c^{(k)} = \partial c(\mathbf{x}^{(k)})$, where the generalized gradient (or subdifferential) is defined as

$$\partial f(\mathbf{x}) := \text{conv} \left\{ \mathbf{g} \mid \mathbf{g} = \lim_{i \to \infty} \nabla f(\mathbf{x}_i),\ \mathbf{x}_i \to \mathbf{x},\ \nabla f(\mathbf{x}_i) \text{ exists \& converges} \right\}$$

Problems of type (P) arise as master problems in decomposition methods such as Benders Decomposition (e.g. [2], [10] and [9]). An important evolving area of application

*Numerical Analysis Report NA/195, Department of Mathematics, University of Dundee

†Department of Mathematics, University of Dundee, DD1 4HN, U.K. `fletcher@maths.dundee.ac.uk`, `sleyffer@maths.dundee.ac.uk`

is in stochastic Nonlinear Programming (NLP). Here, the decomposition of the deterministic equivalent of a stochastic NLP gives rise to a master problem in the first stage variables only of the form (P).

Several methods have been proposed for solving (P), see Kiwiel [12] and references therein. Most methods are of bundle type, where the nonsmooth functions f and c are approximated by a bundle of supporting hyperplanes (e.g. Hiriart-Urruty and Lémarechal [11]).

Clearly, it is possible to re-write (P) as a linearly constraint optimization problem by introducing an exact penalty function. However, Kiwiel [12] points out that this introduces numerical difficulties as the bundle method has to cope with subgradients of widely varying magnitudes.

The method proposed in this paper is closest in concept to Kiwiel's method [12] but it avoids the use a penalty function. Instead the novel concept of a filter [6] is used to force global convergence. A filter accepts a trial point whenever the objective or the constraint violation is improved compared to all previous iterates.

Filter methods have several advantages over penalty function methods. They do not require a penalty parameter estimate which could be problematic to obtain. Practical experience with filter methods shows that they exhibit a certain degree of nonmonotonicity which can be beneficial.

Our method can be extended easily to the case of more than a single nonlinear constraint either by setting $c(\mathbf{x}) = \max\{\mathbf{c}(\mathbf{x})\}$ or directly at the expense of introducing a bundle for *each* nonsmooth constraint.

In Section 2 nonsmooth Fritz-John type optimality conditions are derived. These conditions motivate a Sequential Linear Programming bundle-method for solving (P). The method is stabilized by a trust-region and global convergence is induced by a filter. The method is described in Section 3 and its convergence properties are analysed in Section 4.

2 Strictly feasible strict descent

First, nonsmooth Fritz-John (NSFJ) conditions are derived for (P) in Lemma 2.1. Next these conditions are generalized to allow ϵ-subgradients in Corollary 2.2. This motivates a bundle approach to solving problem (P).

Lemma 2.1 *Let $\mathbf{x}^*$ be a solution to (P). Then it follows that the set*

$$S := \{\mathbf{s} \in I\!R^n \mid \|\mathbf{s}\| = 1 \quad , \quad \mathbf{s}^T\mathbf{g} < 0 \ , \ \forall \mathbf{g} \in \partial f(\mathbf{x}^*) \qquad (2.1)$$

$$\mathbf{s}^T\mathbf{a} < 0 \ , \ \forall \mathbf{a} \in \partial c(\mathbf{x}^*) \ \textit{if} \ c(\mathbf{x}^*) = 0 \ \} \qquad (2.2)$$

is empty.

A point that satisfies $\mathcal{S} = \emptyset$ will be referred to as a *nonsmooth Fritz-John (NSFJ) point*. Lemma 2.1 states that if $\mathbf{x}^*$ is optimal, then there exist no strictly feasible strict descent directions. In passing we note that condition (2.2) can be more conveniently written as

$$\mathbf{s}^T\mathbf{a} < 0 \ , \ \forall \mathbf{a} \in \partial_C c(\mathbf{x}^*) \qquad (2.3)$$

by introducing Clarke's *relative generalized gradient* [4, Section 6.2]

$$\partial_C c(\mathbf{x}) := \text{conv}\left\{\mathbf{a} \in \mathbb{R}^n \mid \mathbf{a} = \lim_{i\to\infty} \mathbf{a}^{(i)},\ \text{where } \mathbf{a}^{(i)} \in \partial c(\mathbf{x}^{(i)}),\ \mathbf{x}^{(i)} \in C \text{ and } \mathbf{x} = \lim_{i\to\infty} \mathbf{x}^{(i)}\right\}$$

where $C := \{\mathbf{x} \in \mathbb{R}^n \mid c(\mathbf{x}) > 0\}$ is the set of all infeasible points. Clearly, $\partial_C c(\mathbf{x}^*) = \emptyset$ if the solution is strictly feasible, which simplifies (2.2)

Proof of Lemma 2.1.

The proof is given for the case $c(\mathbf{x}^*) = 0$. The proof for $c(\mathbf{x}^*) < 0$ is similar.

Assume that the set in (2.1) and (2.2) is *not* empty and seek a contradiction. If the set $\mathcal{S} \neq \emptyset$, then there exists a direction $\mathbf{s}$ with $\|\mathbf{s}\| = 1$ and $\epsilon_1 > 0$ such that

$$\mathbf{s}^T\mathbf{g} \leq \epsilon_1 \text{ and } \mathbf{s}^T\mathbf{a} \leq \epsilon_1 \ ,\ \forall \mathbf{g} \in \partial f^* \text{ and } \forall \mathbf{a} \in \partial c^*.$$

This follows from the fact that the subgradients are closed and convex sets (e.g. [4, Prop. 2.1.2]). The upper semi-continuity ([4, Prop. 2.1.5]) of the subgradients in turn implies that there exists $\epsilon_2 > 0$ such that

$$\begin{array}{l} \partial f(\hat{\mathbf{x}}) \subset \mathcal{N}(\partial f^*, \epsilon_2)\ ,\ \forall \hat{\mathbf{x}} \in \mathcal{N}(\mathbf{x}^*, \delta) \text{ and crucially } \mathbf{s}^T\mathbf{g} < 0,\ \forall \mathbf{g} \in \partial f(\hat{\mathbf{x}}) \\ \partial c(\tilde{\mathbf{x}}) \subset \mathcal{N}(\partial c^*, \epsilon_2)\ ,\ \forall \tilde{\mathbf{x}} \in \mathcal{N}(\mathbf{x}^*, \delta) \text{ and crucially } \mathbf{s}^T\mathbf{a} < 0,\ \forall \mathbf{a} \in \partial c(\tilde{\mathbf{x}}), \end{array}$$

where $\mathcal{N}(\mathbf{x}^*, \delta)$ is a neighbourhood of $\mathbf{x}^*$ of radius δ and $\mathcal{N}(\partial f^*, \epsilon_2)$ is an ϵ_2 neighbourhood of the set ∂f^*. Now consider the effect of a step of length α in the direction $\mathbf{s}$. From the mean-value Theorem it follows that there exists $\hat{\mathbf{x}},\ \tilde{\mathbf{x}} \in \mathcal{N}(\mathbf{x}^*, \delta)$ such that

$$\begin{array}{lrcl} \exists \hat{\mathbf{g}} \in \partial f(\hat{\mathbf{x}}): & f(\mathbf{x}^* + \alpha\mathbf{s}) - f(\mathbf{x}^*) &=& \alpha\mathbf{s}^T\hat{\mathbf{g}} \ < \ 0 \\ \exists \tilde{\mathbf{a}} \in \partial c(\tilde{\mathbf{x}}): & c(\mathbf{x}^* + \alpha\mathbf{s}) - c(\mathbf{x}^*) &=& \alpha\mathbf{s}^T\tilde{\mathbf{a}} \ < \ 0. \end{array}$$

Thus, for α sufficiently small, $\mathbf{x}^* + \alpha\mathbf{s}$ is feasible and has a lower function value than $f(\mathbf{x}^*)$. This contradicts the optimality of $\mathbf{x}^*$, therefore the assumption must be wrong and the set $\mathcal{S}$ must be empty. □

Lemma 2.1 is related to the notion of complete descent of Clarke [4]. However, it is more useful in the context of bundle methods. If $\mathbf{x}$ is not optimal, then the solution of the following (semi-infinite) Linear Program (LP) provides a descent direction.

$$\left\{\begin{array}{ll} \underset{\eta,s}{\text{minimize}} & \eta \\ \text{subject to} & \eta \geq f(\mathbf{x}) + \mathbf{s}^T\mathbf{g}\ ,\ \forall \mathbf{g} \in \partial f(\mathbf{x}) \\ & 0 \geq c(\mathbf{x}) + \mathbf{s}^T\mathbf{a}\ ,\ \forall \mathbf{a} \in \partial c(\mathbf{x}) \\ & \mathbf{x} + \mathbf{s} \in X\ ,\ \|\mathbf{s}\| = 1. \end{array}\right.$$

Note that this problem is generally intractable as it may have an infinite number of constraints. If $\partial f(\mathbf{x})$ and $\partial c(\mathbf{x})$ are finitely generated, as is the case for some composite functions [5, Chapter 14], then this LP is finite dimensional and can become the basis of an algorithm to solve (P).

This paper is concerned with the more general situation where the complete subdifferential is not available, but a single element of it can be computed at any point. The proposed algorithm solves a sequence of LPs in which the generalized gradients $\partial f(\mathbf{x})$ and $\partial c(\mathbf{x})$ are approximated by bundles of ϵ-subgradients.

The following Corollary extends Lemma 2.1 to the situation where ϵ-subgradients are used. For $\epsilon \geq 0$ the ϵ-subgradient is defined (e.g. Hiriart-Urruty and Lemaréchal [11, XI.1]) as

$$\partial_\epsilon f(\mathbf{x}) := \{\mathbf{g} \in \mathbb{R}^n \mid f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{g}^T(\mathbf{y} - \mathbf{x}) - \epsilon\ ,\ \forall \mathbf{y} \in \mathbb{R}^n\}.$$

Corollary 2.2 *Let $\mathbf{x}^*$ be a solution to (P). Then it follows that the set*

$$S_\epsilon := \{\mathbf{s} \in I\!R^n \mid \|\mathbf{s}\| = 1 \quad , \quad \mathbf{s}^T\mathbf{g} < 0 \; , \; \forall \mathbf{g} \in \partial_\epsilon f(\mathbf{x}^*) \tag{2.4}$$

$$\mathbf{s}^T\mathbf{a} < 0 \; , \; \forall \mathbf{a} \in \partial_\epsilon c(\mathbf{x}^*) \textit{ if } c(\mathbf{x}^*) = 0 \; \} \tag{2.5}$$

is empty.

Proof. In [11, Chapter XI.1] it is shown that $\partial f(\mathbf{x}) = \partial_0 f(\mathbf{x})$. Therefore it follows that $\partial_\epsilon f^* \supset \partial f^*$ and $\partial_\epsilon c^* \supset \partial c^*$ $\forall \epsilon \geq 0$. Thus, the conditions defining S_ϵ are stronger and $S_\epsilon \subset S = \emptyset$. □

3 A bundle-filter method for nonsmooth NLP

This section presents a bundle method for nonsmooth NLP. The LP solved at every iteration is motivated by the total descent LP of the previous section. The treatment in this section is similar to that of Kiwiel [12]. The key idea is to build up a cutting plane model of the nonsmooth functions about the current point $\mathbf{x}^{(k)}$ by collecting subgradients from "nearby" points. Unlike other methods, the present algorithm does not make use of a penalty function. Instead, a filter is introduced which ensures global convergence.

3.1 The bundle

This section introduces the concept of a bundle (e.g. [12]). The idea behind the bundle is to approximate the nonsmooth NLP (P) by a cutting plane model. This gives rise to an LP which can be solved to either make progress towards the solution or to enhance the cutting plane model.

The algorithm which is formally described in Section 3.4. It generates a sequence of points $\{\mathbf{z}^{(i)}\}$ which consists of trial points and serious points. The sequence of serious points is denoted by $\{\mathbf{x}^{(k)}\}$. At every serious point $\mathbf{x}^{(k)}$, a cutting plane model of (P) is formed using cutting planes from the current point $\mathbf{x}^{(k)}$ and some auxiliary points $\mathbf{z}^{(i)}$. This cutting plane model gives rise to the following LP

$$LP(\mathbf{x}^{(k)}, j, \rho) \left\{ \begin{array}{lll} \underset{\eta,d}{\text{minimize}} & \eta & \\ \text{subject to} & \eta \geq f^{k,i} + \mathbf{d}^T\mathbf{g}^{(i)} & \forall i \in \mathcal{B}_f^{(j)} \\ & 0 \geq c^{k,i} + \mathbf{d}^T\mathbf{a}^{(i)} & \forall i \in \mathcal{B}_c^{(j)} \\ & \mathbf{x}^{(k)} + \mathbf{d} \in X \; , \; \|\mathbf{d}\| \leq \rho, & \end{array} \right.$$

where we use k to index the serious steps and j to index *all* trial points, including the serious steps. The sets $\mathcal{B}_f^{(j)}$ and $\mathcal{B}_c^{(j)}$ are sets of indices of two bundles of subgradients $\mathbf{g}^{(i)} \in \partial f(\mathbf{z}^{(i)})$ and $\mathbf{a}^{(i)} \in \partial c(\mathbf{z}^{(i)})$ from auxiliary points $\mathbf{z}^{(i)}$. We assume that the subgradients about $\mathbf{x}^{(k)}$ are contained in the bundles of $LP(\mathbf{x}^{(k)}, j, \rho)$. In order to avoid storing the $\mathbf{z}^{(i)}$, the shifted function values

$$f^{k,i} = f^{(i)} + \mathbf{g}^{(i)^T}(\mathbf{x}^{(k)} - \mathbf{z}^{(i)}) \leq f^{(k)} \tag{3.1}$$

and

$$c^{k,i} = c^{(i)} + \mathbf{a}^{(i)^T}(\mathbf{x}^{(k)} - \mathbf{z}^{(i)}) \leq c^{(k)} \tag{3.2}$$

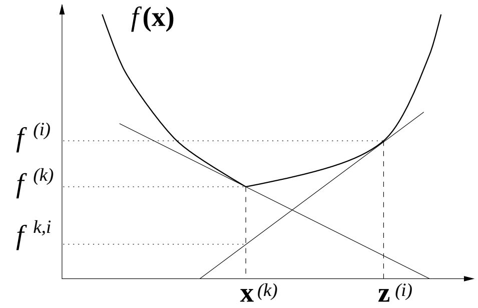

Figure 1: Shifted value $f^{k,i} = f(\mathbf{z}^{(i)}) + g(\mathbf{z}^{(i)})^T(\mathbf{x}^{(k)} - \mathbf{z}^{(i)}))$

are introduced, where the inequalities follow from the convexity of f and c. The $f^{k,i}, c^{k,i}$ can be recurred whenever a serious step is taken from $\mathbf{x}^{(k)}$ to $\mathbf{x}^{(k+1)}$ by using $f^{k+1,i} = f^{k,i} + \mathbf{g}^{(i)^T}(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})$. Figure 1 shows a bundle and the shifted value $f^{k,i}$.

Solving $LP(\mathbf{x}^{(k)}, j, \rho)$ generates a new trial point $\mathbf{x}^+ = \mathbf{x}^{(k)} + \mathbf{d}$. The algorithm now has three possible scenarios.

1. The step is judged to make sufficient progress towards a solution of (P). This will be referred to as a *serious step*. We set $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{d}$ and $\mathbf{z}^{(j+1)} = \mathbf{x}^{(k)} + \mathbf{d}$, choose new bundles $\mathcal{B}_f^{(j+1)} \subset \mathcal{B}_f^{(j)} \cup \{j+1\}$ and $\mathcal{B}_c^{(j+1)} \subset \mathcal{B}_c^{(j)} \cup \{j+1\}$, increase k and j and solve a new LP about $\mathbf{x}^{(k+1)}$.

2. The step fails to make sufficient progress towards a solution, but the new cutting planes at $\mathbf{x}^+$ change $LP(\mathbf{x}^{(k)}, j, \rho)$ significantly. This will be referred to as a *null step*. The new cuts are added to the bundle by setting $\mathbf{z}^{(j+1)} = \mathbf{x}^{(k)} + \mathbf{d}$. Then j is increased and the modified $LP(\mathbf{x}^{(k)}, j, \rho)$ is solved.

3. The step fails to make sufficient progress towards a solution and the new cutting planes at $\mathbf{x}^+$ fail to change $LP(\mathbf{x}^{(k)}, j, \rho)$ significantly. In this case, $\mathbf{x}^+$ is rejected, the trust region radius ρ is reduced and $LP(\mathbf{x}^{(k)}, j, \rho)$ is re-solved.

This description leaves open a number of important issues which are addressed in the next two sections.

The traditional way to handle constraints is by introducing a penalty function. This can either be done by reformulating (P) as an unconstrained minimization of the penalty function or by solving a problem like $LP(\mathbf{x}^{(k)}, \rho)$ and performing a line-search on the penalty function (as in [12]). Here we prefer to use a filter which is introduced in the next section.

3.2 The filter

This section introduces a filter for nonsmooth optimization. The filter will be used as a criterion for accepting or rejecting a step generated by $LP(\mathbf{x}^{(k)}, j, \rho)$. Filter methods have recently been introduced for smooth optimization [6], [8], and [7]. Audet and Dennis [1] have used a filter to derive a constraint pattern search algorithm.

A *filter* is a list of pairs $(h^{(j)}, f^{(j)})$, $j \in \mathcal{F}$ such that no pair dominates any other pair, i.e.

$$\text{either } h^{(j)} < h^{(i)} \ \text{ or } \ f^{(j)} < f^{(i)} \ , \ \forall i \neq j \ , \ i, j \in \mathcal{F}$$

where $h^{(j)} = h(c(\mathbf{x}^{(j)})) = \max\{0, c(\mathbf{x}^{(j)})\}$ is the constraint violation. A typical filter is given by the solid lines in Figure 2.

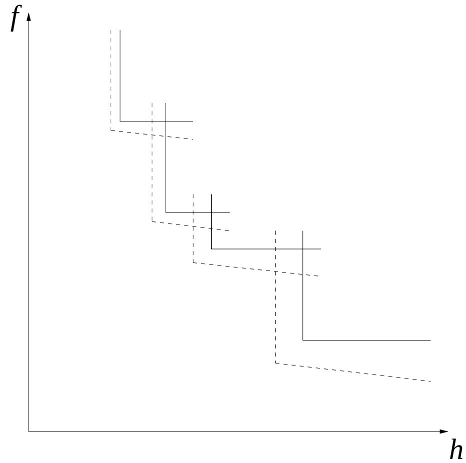

Figure 2: A filter with envelope ($\beta = 0.9,\ \gamma = 0.1$)

The key idea is then to accept only points whose (h, f) pair lies below and to the left of the step function defined by the filter. This turns out to be not sufficient as it allows points to accumulate arbitrarily close to filter entries with $h > 0$. To avoid this a small envelope around the filter is introduced. It is also useful to "tilt" this envelope in the h direction as this enables stronger results about the feasibility of limit points to be established, see Chin and Fletcher [3]. A point $\mathbf{x}^+$ is now said to be *acceptable to the filter* if its (h^+, f^+) pair satisfies

$$h^+ \leq \beta h^{(j)} \ \text{ or } \ f^+ \leq f^{(j)} - \gamma h^+ \ , \ \forall j \in \mathcal{F} \tag{3.3}$$

where $0 < \gamma < \beta < 1$ are constants. This envelope is illustrated by the dashed lines in Figure 2.

Note that this still allows points to accumulate at feasible but non-optimal limit points. To avoid this it is necessary to extend ideas from unconstrained optimization. Define the *actual reduction* as

$$\Delta f = f(\mathbf{x}) - f(\mathbf{x} + \mathbf{d}) \tag{3.4}$$

and the *predicted reduction* as

$$\Delta l = f(\mathbf{x}) - \eta. \tag{3.5}$$

Steps near the feasible region are then required to satisfy a *sufficient reduction* condition

$$\Delta f \geq \sigma_1 \Delta l \ \text{ if } \ \Delta l \geq \delta h. \tag{3.6}$$

If (3.6) is satisfied, then the step is referred to as an *f-type step*. Otherwise, if $\Delta l^{(k)} < \delta(h^{(k)})$, it is labelled an *h-type step*. These definitions are in fact the same as the ones required by filter methods for smooth optimization problems.

3.3 Null steps

The nonsmoothness of (P) requires the introduction of a bundle of subgradients as in Section 3.1. The bundle is accumulated by taking a number of *null-steps* at which the current point remains unchanged but new subgradient information is added to the bundle. A null step is taken, if the trial point $\mathbf{x}+\mathbf{d}$ is rejected and satisfies

$$f(\mathbf{x}+\mathbf{d}) \geq \eta + \sigma_2 \Delta l \tag{3.7}$$

or

$$c(\mathbf{x}+\mathbf{d}) \geq \beta \tau^{(k)}, \tag{3.8}$$

where $0 < \sigma_2 < 1$ and $\sigma_1 + \sigma_2 \leq 1$ and $\tau^{(k)} = \min\{h^{(j)}, j \in \mathcal{F}^{(k)}\} > 0$.

3.4 The algorithm

The algorithm is stated by way of Figure 3. It consists of an inner iteration in which the trust-region radius is reduced until either a serious step (moving to a new point) or a null step (enhancing the bundle) is found. After either a null step or a serious step have been accepted, the trust-region radius is reset to any value of $\rho \geq \rho_0$.

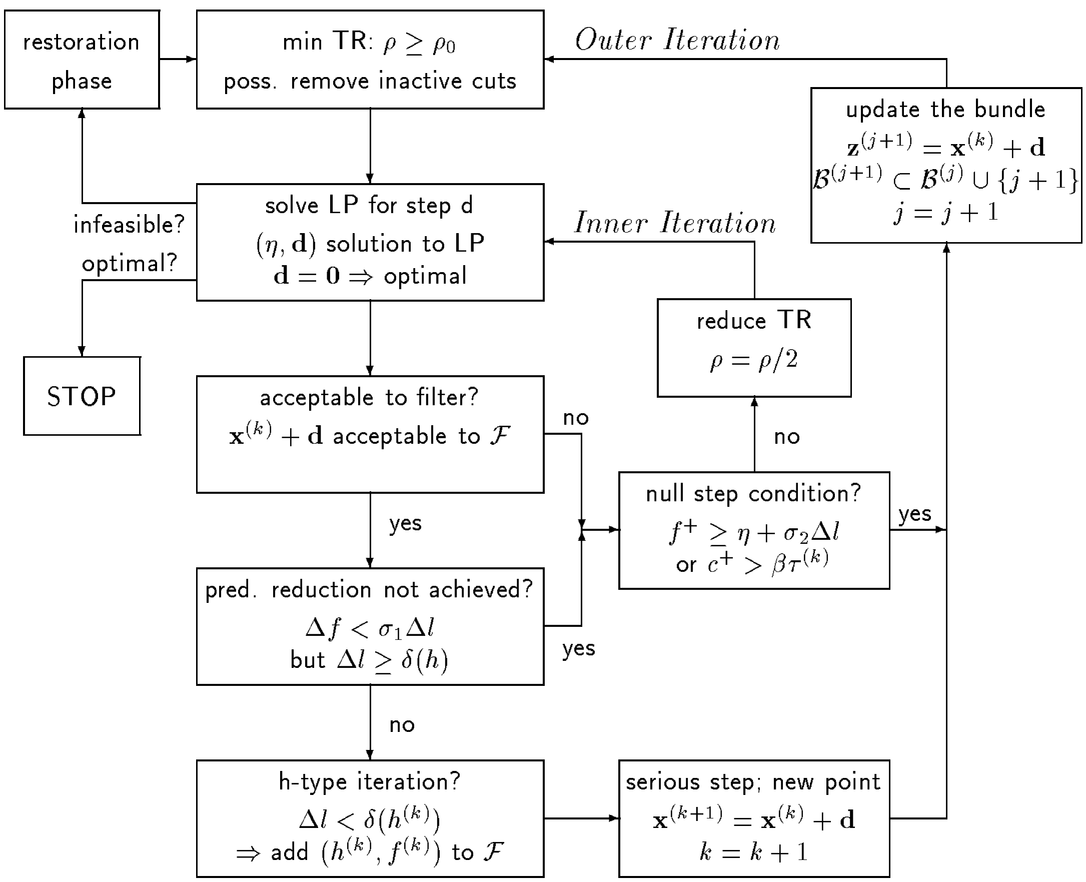

Figure 3: Bundle-filter method for nonsmooth NLP

The algorithm contains a restoration phase which is entered whenever the current $LP(\mathbf{x}^{(k)}, j, \rho)$ is inconsistent. The aim of this phase is to move closer to the feasible region and restore consistency of the LP. This may be achieved by using a trust-region strategy to minimize the norm of the infeasibility. This problem is then a linearly constraint, convex nonsmooth optimization problem which can be solved with a number of traditional techniques such as [13].

In the restoration phase, a number of serious steps are taken and k is increased. The restoration phase has two possible outcomes: *either* for some k, a consistent $LP(\mathbf{x}^{(k)}, j, \rho)$ approximation is encountered *or* all $LP(\mathbf{x}^{(k)}, j, \rho)$ are inconsistent. In the first case, the restoration phase terminates and the algorithms returns to solving (P). In the latter case, the restoration phase converges to a minimum of the constraint violation of (P). If this minimum is positive, then the original problem (P) is inconsistent.

After accepting a serious step or a null step, it is possible to remove some inactive cuts from $LP(\mathbf{x}^{(k)}, j, \rho)$. Thus

$$\mathcal{B}_f^{(j+1)} \subset \mathcal{B}_f^{(j)} \cup \{j+1\} \text{ and } \mathcal{B}_c^{(j+1)} \subset \mathcal{B}_c^{(j)} \cup \{j+1\},$$

where it is assumed that the new cutting planes in f and c are both added to the new bundle, but older (inactive) cuts may be removed from the bundle.

4 Convergence analysis

This section gives a global convergence proof for the bundle-filter algorithm presented in the previous section. First, in Lemma 4.1, it is shown that the inner iteration is finite. This implies that the algorithm is well defined.

Lemma 4.1 *The inner iteration is finite.*

Proof. Assume it is not finite. Then it follows that $\rho \to 0$. Distinguish two cases depending on whether $h^{(k)} > 0$ or $h^{(k)} = 0$.

Case 1: $h^{(k)} > 0$ implies that $LP(\mathbf{x}^{(k)}, j, \rho)$ becomes incompatible for $\rho < \hat{\rho}$ sufficiently small. Thus the inner iteration terminates by entering the restoration phase.

Case 2: If $h^{(k)} = 0$, only cuts in the LP that are active at $\mathbf{x}^{(k)}$ play a role as $\rho \to 0$. These are the cuts with $f^{k,i} = f^{(k)}$ or $c^{k,i} = c^{(k)} = h^{(k)} = 0$. Denoting these active cuts by $\mathcal{A}_f^{(k)}$ and $\mathcal{A}_c^{(k)}$ and replacing $\eta = \nu + f^{(k)}$, the $LP(\mathbf{x}^{(k)}, j, \rho)$ becomes

$$\left\{ \begin{array}{ll} \underset{\eta,d}{\text{minimize}} & \nu + f^{(k)} \\ \text{subject to} & \nu \geq \mathbf{d}^T \mathbf{g}^{(i)}, \ i \in \mathcal{A}_f^{(k)} \\ & 0 \geq \mathbf{d}^T \mathbf{a}^{(i)}, \ i \in \mathcal{A}_c^{(k)} \\ & \mathbf{x}^{(k)} + \mathbf{d} \in X \ , \ \|\mathbf{d}\| \leq \rho. \end{array} \right.$$

Dividing each constraint by ρ, it follows that $\mathbf{d}$ solves $LP(\mathbf{x}^{(k)}, j, \rho)$ if and only if $\mathbf{s} = \mathbf{d}/\rho$ and $\theta = \nu/\rho$ solve

$$\left\{ \begin{array}{ll} \underset{\theta,s}{\text{minimize}} & \theta\rho + f^{(k)} \\ \text{subject to} & \theta \geq \mathbf{s}^T \mathbf{g}^{(i)}, i \in \mathcal{A}_f^{(k)} \\ & 0 \geq \mathbf{s}^T \mathbf{a}^{(i)}, i \in \mathcal{A}_c^{(k)} \\ & \mathbf{s} \in \mathcal{T}(\mathbf{x}^{(k)}; X) \ , \ \|\mathbf{s}\| \leq 1, \end{array} \right.$$

where

$$\mathcal{T}(\mathbf{x}^{(k)};X) := \left\{ \mathbf{s} \in \mathrm{I\!R}^n \mid \mathbf{s} = \lim_{i\to\infty} \frac{\mathbf{x}^{(i)} - \mathbf{x}^{(k)}}{\tau_i} \ , \ \mathbf{x}^{(i)} \in X \ , \ \tau_i \searrow 0 \right\}$$

is the tangent cone to X at $\mathbf{x}^{(k)}$. The minimizer $(\theta, \mathbf{s})$ of this last LP is independent of ρ. Moreover, $\theta \leq 0$, since $(\theta, \mathbf{s}) = (0, \mathbf{0})$ is feasible, which implies that the predicted reduction is $\Delta l = -\theta\rho \geq 0$ for ρ sufficiently small. Assume that $c^+ \leq \beta\tau^{(k)}$, since otherwise a null step must be taken and the inner iteration terminates. Thus, the steps must have been rejected due to not attaining sufficient reduction. This implies

$$\begin{aligned} f(\mathbf{x}^{(k)} + \mathbf{d}) \ &> f^{(k)} - \sigma_1 \Delta l \ = \ f^{(k)} + \sigma_1 \theta \rho \\ &\geq f^{(k)} + (1 - \sigma_2)\theta\rho \ = \ f^{(k)} + \nu + \sigma_2 \Delta l \ = \ \eta + \sigma_2 \Delta l, \end{aligned}$$

where the first inequality follows from the fact that sufficient reduction does not hold and the second inequality follows from $\sigma_1 + \sigma_2 \leq 1$. Thus a null step is taken for ρ sufficiently small and the inner iteration terminates. □

Remark 4.2 *After a null step, the solution of $LP(\mathbf{x}^{(k)}, j, \rho)$ differs significantly from the previous LP solution in the following sense.*

Let $(\eta^{(j)}, \mathbf{d}^{(j)})$ be the solution of $LP(\mathbf{x}^{(k)}, j, \rho)$ before the null step and let $(\eta^{(j+1)}, \mathbf{d}^{(j+1)})$ be any feasible point of $LP(\mathbf{x}^{(k)}, j+1, \rho)$. Now consider the effect of both null step conditions in turn.

1. If the null step condition on f holds, then $f^{(j+1)} = f(\mathbf{x}^{(k)} + \mathbf{d}^{(j)}) \geq \eta^{(j)} + \sigma_2 \Delta l$, and $\Delta l > 0$ and the cut $\eta \geq f^{k,j+1} + \mathbf{g}^{(j+1)^T}\mathbf{d}$ is added. Substituting for $f^{k,j+1}$ shows that any feasible point $(\eta^{(j+1)}, \mathbf{d}^{(j+1)})$ in $LP(\mathbf{x}^{(k)}, j+1, \rho)$ must satisfy

   $$\eta^{(j+1)} \ \geq \ f^{(j+1)} + \mathbf{g}^{(j+1)^T}(-\mathbf{d}^{(j)}) + \mathbf{g}^{(j+1)^T}\mathbf{d}^{(j+1)} \ \geq \ \eta^{(j)} + \sigma_2 \Delta l + \mathbf{g}^{(j+1)^T}(\mathbf{d}^{(j+1)} - \mathbf{d}^{(j)}).$$

   Thus

   $$(\eta^{(j+1)} - \eta^{(j)}) \ + \ \mathbf{g}^{(j+1)^T}(\mathbf{d}^{(j)} - \mathbf{d}^{(j+1)}) \ \geq \ \sigma_2 \Delta l > 0$$

   which shows that $(\eta^{(j+1)}, \mathbf{d}^{(j+1)})$ differ significantly from $(\eta^{(j)}, \mathbf{d}^{(j)})$ after the null step.

2. If the null step condition on c holds, then the cut $0 \geq c^{k,j+1} + \mathbf{a}^{(j+1)^T}\mathbf{d}$ is added. Substituting for $c^{k,j+1}$, and evaluating at any feasible point $(\eta^{(j+1)}, \mathbf{d}^{(j+1)})$ gives

   $$0 \ \geq \ c^{(j+1)} + \mathbf{a}^{(j+1)^T}(-\mathbf{d}^{(j)}) + \mathbf{a}^{(j+1)^T}\mathbf{d}^{(j+1)} \ \geq \ \beta\tau^{(k)} + \mathbf{a}^{(j+1)^T}(\mathbf{d}^{(j+1)} - \mathbf{d}^{(j)}).$$

   Thus

   $$\mathbf{a}^{(j+1)^T}(\mathbf{d}^{(j)} - \mathbf{d}^{(j+1)}) \ \geq \ \beta\tau^{(k)} > 0$$

   which shows that $\mathbf{d}^{(j+1)}$ differs significantly from $\mathbf{d}^{(j)}$ after the null step.

In the subsequent analysis, it is useful to distinguish four different cases:

$\mathbb{A}$ The Restoration phase fails. (P) is infeasible.

$\mathbb{B}$ An optimal solution is found.

$\mathbb{C}$ There exists an infinite subsequence of serious steps.

$\mathbb{D}$ There exists an infinite sequence of null steps after a finite number of serious steps.

The fact that all iterates lie in the compact set X and Lemma 4.1 imply that if the algorithm does not terminate finitely, then it generates at least one converging subsequence. In case $\mathbb{C}$, denote this subsequence of serious steps by $\{\mathbf{x}^{(k)},\ k \in \mathcal{S}\}$ and its limit point by $\mathbf{x}^\infty = \lim_{k\in\mathcal{S}} \mathbf{x}^{(k)}$. In case $\mathbb{D}$, denote this subsequence of null steps by $\{\mathbf{z}^{(i)},\ i \in \mathcal{S}\}$ and its limit point by $\mathbf{z}^\infty = \lim_{i\in\mathcal{S}} \mathbf{z}^{(i)}$.

In case $\mathbb{C}$ it is shown in Lemma 4.3 that the filter ensures feasibility of all limit points. Moreover, if this limit point is not an NSFJ point, then we show in Lemma 4.4 that an f-type step that is acceptable to the filter is generated after a finite number of null steps. In case $\mathbb{D}$, Lemma 4.5 shows that any limit point of the final sequence of null steps is a feasible NSFJ point. Finally, Theorem 4.6 summarizes these results.

Lemma 4.3 *Every limit point* $\mathbf{x}^\infty$ *of a subsequence of type* $\mathbb{C}$ *is feasible.*

Proof. This result is established in [3]. A weaker result which ensures only the existence of a feasible limit point is established in Lemma 3, [8]. □

Lemma 4.4 *In case* $\mathbb{C}$, *assume that* $\mathbf{x}^\infty$ *is* not *an NSFJ point and that* $k \in \mathcal{S}$, *k sufficiently large is fixed. Then a serious step must be accepted by the algorithm after a finite number of null steps. Moreover, this serious step is an f-type step.*

Proof. If $\mathbf{x}^\infty$ is not an NSFJ point, then there exist $\epsilon > 0$ and $\epsilon_0 > 0$ and a neighbourhood $\mathcal{N}^\infty$ of $\mathbf{x}^\infty$ such that for every $\mathbf{x} \in \mathcal{N}^\infty$ there exists a strictly feasible strict descent direction $\mathbf{s}$, $\|\mathbf{s}\| = 1$ with

$$\mathbf{s}^T\mathbf{g} \le -\epsilon \text{ and } \mathbf{s}^T\mathbf{a} \le -\epsilon\ ,\ \forall \mathbf{g} \in \partial_{\epsilon_0} f(\mathbf{x}) \text{ and } \forall \mathbf{a} \in \partial_{\epsilon_0} c(\mathbf{x}).$$

In what follows we may assume $\mathbf{x}^{(k)} \in \mathcal{N}^\infty$ for all k sufficiently large. Now we assume that an infinite number of null-steps is taken and seek a contradiction.
First we show that for k sufficiently large, $LP(\mathbf{x}^{(k)}, j, \rho)$ is consistent. From Lemma 4.3 it follows that $h^{(k)} \to 0$.
The constraints $i \in \mathcal{B}_f^{(j)}$ of $LP(\mathbf{x}^{(k)}, j, \rho)$ are clearly all feasible for any $\rho \ge 0$.
Now consider the cutting planes $i \in \mathcal{B}_c^{(j)}$. Since some cutting planes may come from auxiliary points $\mathbf{z}^{(i)} \ne \mathbf{x}^{(k)}$ it is necessary to distinguish two cases, depending on whether the cutting plane is "nearly" active at $\mathbf{x}^{(k)}$ or not. This can be made precise by distinguishing whether $\mathbf{a}^{(i)} \in \partial_{\epsilon_0} c(\mathbf{x})$ for some $\mathbf{x} \in \mathcal{N}^\infty$ or not.
Case 1: If $\mathbf{a}^{(i)} \in \partial_{\epsilon_0} c(\mathbf{x})$ for some $\mathbf{x} \in \mathcal{N}^\infty$, then $\mathbf{s}^T\mathbf{a}^{(i)} \le -\epsilon$. The convexity of $c(\mathbf{x})$ implies that $c^{k,i} \le c^{(k)} = h^{(k)}$. Thus the cut

$$c^{k,i} + \rho\mathbf{s}^T\mathbf{a}^{(i)} \le h^{(k)} - \rho\epsilon$$

is satisfied whenever $\rho \ge h^{(k)}/\epsilon$.
Case 2: If $\mathbf{a}^{(i)} \notin \partial_{\epsilon_0} c(\mathbf{x})$ for some $\mathbf{x} \in \mathcal{N}^\infty$, then the definition of the ϵ_0 subdifferential implies that $c^{\infty,i} < -\epsilon_0$. Now observe

$$\begin{aligned} c^{k,i} + \rho\mathbf{s}^T\mathbf{a}^{(i)} &= c^{(i)} + \mathbf{a}^{(i)^T}(\mathbf{x}^{(k)} - \mathbf{z}^{(i)}) + \rho\mathbf{s}^T\mathbf{a}^{(i)} \\ &= c^{\infty,i} + \mathbf{a}^{(i)^T}(\mathbf{x}^{(k)} - \mathbf{x}^\infty) + \rho\mathbf{s}^T\mathbf{a}^{(i)} \\ &\le -\epsilon_0 + L\|\mathbf{x}^{(k)} - \mathbf{x}^\infty\| + \rho\mathbf{s}^T\mathbf{a}^{(i)} \end{aligned}$$

where L is the local Lipschitz constant of $c(\mathbf{x})$. Now, if

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{\infty}\| \leq \frac{\epsilon_0}{2L}$$

then it follows that

$$c^{k,i} + \rho \mathbf{s}^T \mathbf{a}^{(i)} \leq -\frac{\epsilon_0}{2} + \rho \mathbf{s}^T \mathbf{a}^{(i)}$$

Let $\overline{a} > 0$ be a constant such that $\overline{a} \geq \mathbf{s}^T \mathbf{a}^{(i)}$ ($\overline{a}$ exists by the compactness of $\partial c(\mathbf{x})$) and define

$$\kappa := \frac{\epsilon_0}{2\overline{a}}. \tag{4.1}$$

The above argument shows that the cutting plane i is consistent if $\rho \leq \kappa$ independent of i. Combining both cases, it follows that the LP is consistent, if

$$\frac{h^{(k)}}{\epsilon} \leq \rho \leq \frac{\epsilon_1}{\overline{a}}.$$

Since the lower bound tends to zero, it follows that for k sufficiently large, the LP is consistent for $\rho = \frac{1}{2}\min(\rho_0, \kappa)$.

Now assume that an infinite number of null steps is taken and seek a contradiction. If such an infinite sequence exists, then the null step conditions (3.7) or (3.8) must hold infinitely often. Now consider both conditions in turn.

If an infinite sequence of null steps with (3.8) is taken, then there exists an infinite sequence of null steps $\mathbf{d}^{(j)} = \mathbf{z}^{(j)} - \mathbf{x}^{(k)}$ with $c(\mathbf{x}^{(k)} + \mathbf{d}^{(j)}) \geq \beta\tau^{(k)}$. The compactness of the trust region and X ensures that there exists a converging subsequence with limit $\mathbf{z}^{\infty}$. At every such iteration the following cut is added to the bundle.

$$0 \;\geq\; c^{k,j} + \mathbf{a}^{(j)^T}(\mathbf{x} - \mathbf{x}^{(k)}) \;=\; c^{k,j} - \mathbf{a}^{(j)^T}(\mathbf{x}^{(k)} - \mathbf{z}^{(j)}) + \mathbf{a}^{(j)^T}(\mathbf{x} - \mathbf{z}^{(j)}) \;=\; c^{(j)} + \mathbf{a}^{(j)^T}(\mathbf{x} - \mathbf{z}^{(j)}),$$

where we have used (3.2). The limit $\mathbf{z}^{\infty}$ must satisfy these cuts for all j. Substituting $\mathbf{x} = \mathbf{z}^{\infty}$ and taking limits on both sides gives

$$0 \;\geq\; c(\mathbf{z}^{\infty}) + \mathbf{a}^{\infty^T}(\mathbf{z}^{\infty} - \mathbf{z}^{\infty}) \;=\; c(\mathbf{z}^{\infty}),$$

where the equality follows from the boundedness of ∂c. Thus any limit point $\mathbf{z}^{\infty}$ is feasible.

Since $\mathbf{z}^{\infty}$ is feasible it follows in particular, that there exists an index j_0 such that $c(\mathbf{z}^{(j)}) \leq \beta\tau^{(k)}$, $\forall j \geq j_0$ which means that $\mathbf{z}^{(j)}$ for all j sufficiently large is acceptable to the filter.

Therefore, the sequence of null steps can only be infinite, if the sufficient reduction condition is *not* satisfied. The cuts that are added at every null step include

$$\eta \geq f^{k,j} + \mathbf{g}^{(j)^T}\mathbf{d}.$$

Substituting for $f^{k,j}$, see (3.1) and evaluating these cuts at $\mathbf{d}^{(j)}$ shows that the sequence of LP solutions $(\eta^{(j)}, \mathbf{d}^{(j)})$ must satisfy

$$\begin{aligned} & \eta^{(j+1)} - \eta^{(j)} + \mathbf{g}^{(j)^T}(\mathbf{d}^{(j)} - \mathbf{d}^{(j+1)}) \;\geq\; \sigma_2\epsilon\rho_{\min} \\ \Rightarrow\; & \eta^{(j+1)} - \eta^{(j)} + \mathbf{g}^{(j)^T}(\mathbf{z}^{(j)} - \mathbf{z}^{(j+1)}) \;\geq\; \sigma_2\epsilon\rho_{\min} > 0, \end{aligned}$$

where $\rho_{\min} = \frac{1}{2}\min(\rho_0, \kappa)$, κ is defined in (4.1) and we have used Remark 4.2 and the fact that $\Delta l \geq \rho\epsilon$. This contradicts the fact that the sequence $(\eta^{(j)}, \mathbf{z}^{(j)})$ converges (the

LP values $\eta^{(j)}$ converge since they are bounded below by the supports on $f(\mathbf{x})$ and above by f^∞ due to convexity).
Therefore, the assumption must be wrong and a serious step is accepted after a finite number of null steps. Moreover, this serious step satisfies the conditions for an f-type step. This can be seen from the fact that

$$\Delta l^{(k)} \geq \rho\epsilon \geq \delta(h^{(k)})$$

for k sufficiently large (since otherwise, $h^{(k)}$ would be bounded away from zero). □

Lemma 4.5 *In case $\mathbb{D}$, suppose that the algorithm generates a finite number of serious steps and that k is the last serious step. Then it follows that $\mathbf{x}^{(k)}$ is a feasible NSFJ point.*

Proof. First it is shown that $\mathbf{x}^{(k)}$ is feasible. Then we can apply Lemma 4.4 to show that $\mathbf{x}^{(k)}$ is also an NSFJ point. To show that $\mathbf{x}^{(k)}$ is feasible we distinguish two cases, depending on whether $\rho \to 0$ or not.

If $\rho \to 0$, then the sequence of null steps $\mathbf{z}^{(j)} \to \mathbf{z}^\infty = \mathbf{x}^{(k)}$. Now assume that this limit is *not* feasible, i.e. $c^{(k)} \geq \epsilon > 0$. Since $\mathbf{z}^{(j)} \to \mathbf{x}^{(k)}$, it follows that

$$0 \;\geq\; c^{(k)} + \mathbf{a}^{(k)^T}(\mathbf{z}^{(j)} - \mathbf{x}^{(k)}) \;=\; c^{(k)} \;\geq\; \epsilon \;>\; 0$$

which is a contradiction. Thus $\mathbf{x}^{(k)}$ must be feasible if $\rho \to 0$.

Next consider the case where $\rho \geq \overline{\rho} > 0 \;\forall j$. In this case, the algorithm behaves like a cutting plane method on the ball of radius $\overline{\rho}$ about the point $\mathbf{x}^{(k)}$. The cutting plane method converges (e.g. [11]) to a feasible minimum, say $\mathbf{z}^\infty$.
Again assume that $\mathbf{x}^{(k)}$ is not feasible ($h^{(k)} > 0$). Since $\mathbf{z}^\infty$ is feasible, it follows that $h^{(j)} \to 0$ so that for j sufficiently large the null step is acceptable to the filter, i.e. $h^{(j)} \leq \beta\tau^{(k)}$. Now, if there exists a j such that the predicted reduction $\Delta l^{(j)} \leq \delta h^{(k)}$, then a serious h-type step is accepted which contradicts the fact that only null steps are taken.
Therefore there exists a constant $\gamma > 0$ such that

$$\Delta l^{(j)} \geq \gamma := \delta h^{(k)} > 0 \;\forall j.$$

From the convergence of the cutting plane method it follows that

$$\eta^{(j)} \nearrow f^\infty \quad \text{and} \quad f^{(j)} \to f^\infty$$

the LP function values, $\eta^{(j)}$, and the function values, $f^{(j)}$, converge. Therefore, there exists j_0 such that

$$\begin{aligned} f^{(j)} - \eta^{(j)} \;&\leq\; (1-\sigma_1)\gamma \\ \Rightarrow\; f^{(j)} - \eta^{(j)} \;&\leq\; (1-\sigma_1)\Delta l^{(j)} \;=\; (1-\sigma_1)(f^{(k)} - \eta^{(j)}) \end{aligned}$$

for all $j \geq j_0$. Rearranging this inequality and substituting for $f^{(j)}$ in the actual reduction condition (3.4) implies that

$$\Delta f \;=\; f^{(k)} - f^{(j)} \;\geq\; f^{(k)} - \eta^{(j)} - (1-\sigma_1)(f^{(k)} - \eta^{(j)}) \;=\; \sigma_1 \Delta l^{(j)},$$

for all $j \geq j_0$. This shows that a serious f-type step is accepted, again contradicting the fact that no more serious steps are taken. Thus it follows that $\mathbf{x}^{(k)}$ is feasible.

Since $h^{(k)} = 0$, we can now apply Lemma 4.4 which shows that if $\mathbf{x}^{(k)}$ is not an NSFJ point, then a serious f-type step is accepted after a finite number of null steps. This contradicts the fact that no more serious steps are taken. Therefore, $\mathbf{x}^{(k)}$ must be an NSFJ point. □

Theorem 4.6 *If the assumptions* **A1** *to* **A3** *hold, then for the algorithm in Figure 3, either* $\mathbb{A}$ *or* $\mathbb{B}$ *holds or there exists an infinite subsequence of type* $\mathbb{C}$ *or* $\mathbb{D}$*. Any accumulation point of this subsequence is an NSFJ point.*

Proof. It suffices to consider the case in which neither ($\mathbb{A}$) nor ($\mathbb{B}$) occurs. The existence of an infinite subsequence of type ($\mathbb{C}$) or ($\mathbb{D}$) follows from Lemma 4.1.
In case ($\mathbb{C}$), let $\mathbf{x}^{\infty}$ be any accumulation point of this subsequence and consider the (thinner) subsequence of serious steps $\{\mathbf{x}^{(k)},\ k \in \mathcal{S}\}$ converging to $\mathbf{x}^{\infty}$. We will show that $\mathbf{x}^{\infty}$ is an NSFJ point.
The feasibility of $\mathbf{x}^{\infty}$ follows from Lemma 4.3. Now assume that $\mathbf{x}^{\infty}$ is not an NSFJ point. Then it follows from Lemma 4.4 that an f-type step is taken after a finite number of null steps. This f-type step is taken for $\rho_{\min} = \frac{1}{2}\min\{\kappa, \rho_0\}$ and therefore $\Delta l \geq \epsilon\rho_{\min}$. The fact that an f-type step is taken implies that $\Delta f^{(k)} > \sigma\epsilon\rho_{\min}$ which contradicts the boundedness of $\sum \Delta f^{(k)}$. Therefore, $\mathbf{x}^{\infty}$ must be an NSFJ point.
In case ($\mathbb{D}$), it follows that after a final serious step, only null steps are taken. In this case, Lemma 4.5 shows that $\mathbf{x}^{(k)}$ is an NSFJ point and must be feasible. □

References

[1] Audet, C. and Dennis, J.E.Jr. Combining pattern search and filter algorithms for derivative free optimization,. Technical report in progress, Department of Computational & Applied Mathematics, Rice University, Houston, TX, USA, November 1999. (Presented at Nonlinear Programming Conference, Coimbra, Oct. 99).

[2] J.F. Benders. Partitioning procedures for solving mixed-variable programming problems. *Numerische Mathematik*, 4:238–252, 1962.

[3] Choong Ming Chin and R. Fletcher. A really cool paper on filter methods. Numerical Analysis Report NA/XXX, Department of Mathematics, University of Dundee, Dundee, DD1 4HN, UK, December 1999.

[4] F.H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley, reprinted by SIAM, New York, 1983 (1990).

[5] Fletcher, R. *Practical Methods of Optimization, 2nd edition.* John Wiley, Chichester, 1987.

[6] Fletcher, R. and Leyffer, S. Nonlinear programming without a penalty function. Numerical Analysis Report NA/171, Department of Mathematics, University of Dundee, September 1997. (Submitted to Mathematical Programming).

[7] Fletcher, R., Gould, N.I.M., Leyffer, S. and Toint, Ph.L. Global convergence of trust-region SQP-filter algorithms for general nonlinear programming. Report 99/03, Department of Mathematics, University of Namur, 61, rue de Bruxelles, B-5000 Belgium, November 1999.

[8] Fletcher, R., Leyffer, S. and Toint, Ph.L. On the global convergence of an SLP-filter algorithm. Numerical Analysis Report NA/183, University of Dundee, UK, August 1998.

[9] Flippo, O.E. and Rinnoy Kan, A.H.G. Decomposition in general mathematical programming. *Mathematical Programming*, 60:361–382, 1993.

[10] Geoffrion, A.M. Generalized Benders decomposition. *Journal of Optimization Theory and Applications*, 10:237–260, 1972.

[11] Hirriart-Urruty, J.-B. and Lemaréchal, C. *Convex Analysis and Minimization Algorithms II.* Grundlehren der mathematischen Wissenschaften. Springer Verlag, Berlin, 1996.

[12] K.C. Kiwiel. A constraint linearization method for nondifferentiable convex minimization. *Numerische Mathematik*, 51:395–414, 1987.

[13] Schramm, H. and Zowe, J. A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. *SIAM Journal of Optimization*, 2(1):121–152, 1992.