A note on feasibility in Benders Decomposition

A. Grothey, S. Leyffer and K.I.M. McKinnon

Draft January 15, 1999

Abstract

A convex nonlinear programming problem is presented which causes a simplistic application of Benders Decomposition to converge to a non-stationary point. The failure of Benders Decomposition to converge is due to the way in which infeasible subproblems are handled. A new feasibility restoration is proposed and shown to resolve this problem. The new feasibility restoration is also relevant to bundle-based decomposition schemes.

Keywords: Nonlinear Programming, Benders Decomposition, Variable Decomposition, Bundle-based Decomposition.

1 Introduction

This note considers feasibility issues arising in Benders Decomposition (e.g. Geoffrion [3] or Flippo and Rinnoy Kan [2]). We are interested in Benders Decomposition as a mechanism for decomposing and solving large scale Nonlinear Programming (NLP) problems.

An important and often difficult part in solving large NLP problems is to attain feasible points. This task is harder in Benders Decomposition as points must be obtained which are feasible in the generated subproblems, whereas SQP methods for instance only require feasible QP approximations.

In the remainder of this section Benders Decomposition is briefly reviewed (see [3] and [2] for a more detailed description). Consider the NLP problem

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

where the variables $\mathbf{y}$ are regarded as the complicating variables. In the context of decomposition, $\mathbf{y}$ are the linking variables and (P) decomposes into a number of smaller independent NLP problems for fixed $\mathbf{y}$.

Throughout the paper the following assumptions are made.

A1 f and $\mathbf{c}$ are twice continuously differentiable convex functions.

A2 X and Y are compact polyhedral sets.

A3 The multipliers of (P) are bounded for every fixed $\mathbf{y}$.

These assumptions are standard assumptions in nonlinear programming except for the convexity assumption in **A1**. In practice heuristics can be used which do not require f and $\mathbf{c}$ to be convex. These heuristics usually involve the shifting down of cuts in the master problem. The following analysis can be readily extended to this case.

Fixing the complicating variables $\mathbf{y} = \hat{\mathbf{y}}$ a (simpler) NLP subproblem is obtained.

$$(S(\hat{\mathbf{y}}))\left\{\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(\mathbf{x},\mathbf{y}) \\ \text{subject to} & \mathbf{c}(\mathbf{x},\mathbf{y}) \le \mathbf{0} \\ & \mathbf{y} = \hat{\mathbf{y}} \\ & (\mathbf{x},\mathbf{y}) \in X \times Y. \end{array}\right.$$

Related to $(S(\hat{\mathbf{y}}))$ is the value function $v(\hat{\mathbf{y}}) := v(S(\hat{\mathbf{y}}))$. It can be shown that the multipliers of the constraint $\mathbf{y} = \hat{\mathbf{y}}$ are subgradients of $v(\hat{\mathbf{y}})$.

In general, not all choices of $\hat{\mathbf{y}} \in Y$ give rise to feasible NLP subproblems. Therefore, in [3] and [2] a feasibility subproblem is solved, if the subproblem is infeasible. A general form of this feasibility problem is

$$(F(\hat{\mathbf{y}}))\left\{\begin{array}{ll} \underset{x,y}{\text{minimize}} & \sum_{j \in J} c_j(\mathbf{x},\mathbf{y}) \\ \text{subject to} & c_j(\mathbf{x},\mathbf{y}) \le 0 \ , \ \forall j \in J^\perp \\ & \mathbf{y} = \hat{\mathbf{y}} \\ & (\mathbf{x},\mathbf{y}) \in X \times Y, \end{array}\right.$$

where the constraints have been divided into two disjoint sets. J collects the constraints with positive residual and $J^\perp$ the ones that can be satisfied. This form of feasibility problem is used in SQP methods (e.g. [1]) and is general enough to include more common forms such as the least l_1 or l_∞ norm feasibility problem.

Denote the value function corresponding to $(F(\hat{\mathbf{y}}))$ by $v^0(\hat{\mathbf{y}}) := v(F(\hat{\mathbf{y}}))$. In practice, it is not necessary to distinguish between $(S(\hat{\mathbf{y}}))$ and $(F(\hat{\mathbf{y}}))$ as most solvers will return a solution to some $(F(\hat{\mathbf{y}}))$ if $(S(\hat{\mathbf{y}}))$ is not feasible.

Next note that (P) is equivalent to

$$(M)\left\{\begin{array}{ll} \underset{\eta,y}{\text{minimize}} & \eta \\ \text{subject to} & \eta \ge v(\mathbf{y}) \\ & 0 \ge v^0(\mathbf{y}) \\ & \mathbf{y} \in Y. \end{array}\right.$$

And then define the Benders master program which is solved at iteration k as the following relaxation of this problem.

$$(M^k)\left\{\begin{array}{ll} \underset{\eta,y}{\text{minimize}} & \eta \\ \text{subject to} & \eta \ge v(\mathbf{y}^{(j)}) + \mathbf{w}^{(j)^T}(\mathbf{y} - \mathbf{y}^{(j)}) \ , \ \forall j \in \mathcal{S}^k \\ & 0 \ge v^0(\mathbf{y}^{(i)}) + \mathbf{w}^{(i)^T}(\mathbf{y} - \mathbf{y}^{(i)}) \ , \ \forall i \in \mathcal{F}^k \\ & \mathbf{y} \in Y. \end{array}\right.$$

where

$$\mathcal{S}^k := \left\{ j \le k : S(\mathbf{y}^{(j)}) \text{ feasible} \right\} \quad \text{and} \quad \mathcal{F}^k := \left\{ i \le k : S(\mathbf{y}^{(i)}) \text{ infeasible} \right\}$$

and $\mathbf{w}^{(j)} \in \partial v(\mathbf{y}^{(j)})$ and $\mathbf{w}^{(i)} \in \partial v^0(\mathbf{y}^{(j)})$.

Benders Decomposition solves an alternating sequence of NLP subproblems ($S(\mathbf{y}^{(k)})$ or $F(\mathbf{y}^{(k)})$ respectively) and (LP) master problems (M^k). The optimal values of (M^k) form a non-decreasing sequence of lower bounds on the optimum of (P) and the values $v(\mathbf{y}^{(j)})$ are upper bounds. Hence the algorithm terminates once both bounds are close.

Convergence of Benders Decomposition is established in [3] under the assumption that either Y is finite (which corresponds to the integer programming case) or that ($S(\hat{\mathbf{y}})$) is feasible for all $\hat{\mathbf{y}} \in Y$. In [2] convergence is established under the assumption that either Y is finite or that the set of dual multipliers is finite. Neither of these assumptions apply if (P) is a general NLP problem and this appears to rule Benders Decomposition out for decomposing large NLP problems.

In the next section we present a convex NLP example for which Benders Decomposition fails to converge. The failure to converge is due to the fact that the feasibility cuts derived from $F(\mathbf{y})$ do not provide subgradient information for $v(\mathbf{y})$. In Section 3 a new feasibility restoration is proposed which is shown to avoid these difficulties.

2 Convergence to non-stationary points

Consider the convex NLP

$$(P_1)\left\{\begin{array}{ll} \underset{x_1,x_2,y}{\text{minimize}} & y^2 - x_2 \\ \text{subject to} & (x_1-1)^2 + x_2^2 \le \ln(y) \\ & (x_1+1)^2 + x_2^2 \le \ln(y) \\ & y \ge 1. \end{array}\right.$$

Here y is the complicating variable and x_1, x_2 are the subproblem variables. By fixing $y = \hat{y}$ the following subproblem is obtained

$$(S_1(\hat{y}))\left\{\begin{array}{ll} \underset{x_1,x_2,y}{\text{minimize}} & y^2 - x_2 \\ \text{subject to} & (x_1-1)^2 + x_2^2 \le \ln(y) \\ & (x_1+1)^2 + x_2^2 \le \ln(y) \\ & y = \hat{y}. \end{array}\right.$$

The feasible region of ($S_1(\hat{y})$) is the intersection of two circles in the (x_1, x_2) plane, illustrated in Figure 1. The feasible region is controlled by $\sqrt{\ln(\hat{y})}$ the radius of the two circles, illustrated by the shaded area. For $\hat{y} < e$ the feasible region is empty.

Starting with $y = e^2$ the subproblem is feasible and the optimal solution is $x_1 = 0, x_2 = 1$. A Benders cut is derived from dual information of the subproblem and the master problem is given by

$$(M_1)\left\{\begin{array}{ll} \underset{\eta,y}{\text{minimize}} & \eta \\ \text{subject to} & \eta \ge (e^2-1) + (2e^2 - \frac{1}{2e^2})(y - e^2) \\ & y \ge 1. \end{array}\right.$$

The solution of (M_1) is $y = 1$. Note that the Benders cut "pushes" y to the left end of its range, as the slope $2e^2 - \frac{1}{2e^2}$ is positive.

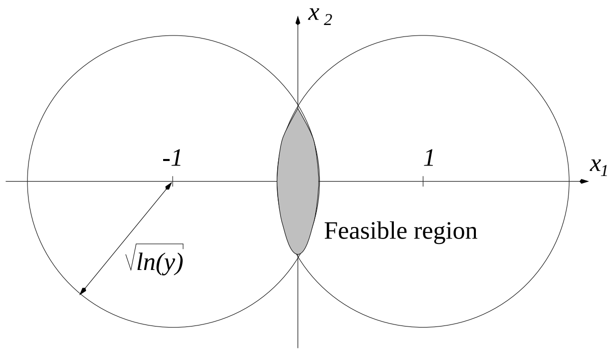

Figure 1: Feasible region of example NLP

The next subproblem is infeasible (as $y = 1 < e$) and an optimal solution to the corresponding feasibility problem is given by $x_1 = x_2 = 0$ ($J^\perp = \emptyset$). In general, if $y^{(k)} < e$, the following feasibility cut is generated and added to the master problem

$$0 \geq (2 - 2\ln(y^{(k)})) + (-\frac{2}{y^{(k)}})(y - y^{(k)}) \;\Leftrightarrow\; y \geq (2 - \ln(y^{(k)}))y^{(k)}.$$

The first Benders cut will always ensure that the new y is chosen as small as possible. Thus the solution of the next master problem will be $(2 - \ln(y^{(k)}))y^{(k)}$. This is formalized in the following lemma.

Lemma 2.1 If $1 \leq y^{(k)} < e$, then the next iterate of Benders Decomposition is $y^{(k+1)} = (2 - \ln(y^{(k)}))y^{(k)} < e$.

Proof: It is clear that the solution to the successive master problem relaxations is $y^{(k+1)} = (2 - \ln(y^{(k)}))y^{(k)}$. It only remains to show that $y < e$ implies $(2 - \ln(y))y < e$. Now write $y = e\gamma$ with $0 < \gamma < 1$. Then it follows that $(2 - \ln(y))y = (1 - \ln(\gamma))\gamma e$. Thus it suffices to show that $(1 - \ln(\gamma))\gamma < 1$, $\forall 0 < \gamma < 1$. Next examine the behaviour of $(1 - \ln(\gamma))\gamma$ over the open interval $(0, 1)$. It is easy to show that

$$\lim_{\gamma \to 1}(1 - \ln(\gamma))\gamma = 1 \quad \text{and} \quad \lim_{\gamma \to 0}(1 - \ln(\gamma))\gamma = 0$$

Differentiating with respect to γ gives

$$\frac{d}{d\gamma}(1 - \ln(\gamma))\gamma = -\ln(\gamma) > 0 \quad \forall 0 < \gamma < 1.$$

Thus $(1 - \ln(\gamma))\gamma$ is monotonic over $0 < \gamma < 1$ and attains its supremum at $\gamma = 1$. This supremum is 1, which completes the proof. □

A consequence of this lemma is that Benders Decomposition will never generate a feasible y but converge to

$$\lim_{k \to \infty} y^{(k)} = e$$

a feasible but *nonstationary* point.

2.1 Significance to convergence of Benders Decomposition

At a first glance, the example seems to indicate that there are errors in the various convergence results in the literature. However, this is not the case as the assumptions usually exclude our example. In the case of Geoffrion's analysis [3], the example does not satisfy the assumption in Theorem 2.5 that all points $y \in Y$ generate feasible subproblems. So the important issue of feasibility is really avoided. Flippo and Rinnoy Kan [2] make the assumption that either the primal or the dual feasible set are finite in their convergence Theorem 3.3. In the example both are violated.

As a consequence, Benders Decomposition of [3] and [2] is not applicable to the decomposition of general NLP problems. In the next section a general method for ensuring convergence to optimal solutions that avoids the pitfalls of earlier work is presented.

3 Feasibility Restoration for Benders Decomposition

The failure of Benders Decomposition to converge is due to the fact that the Benders cuts only approach feasibility in the limit and never collect subgradient information from the objective of (P).

In this section a new version of Benders Decomposition is presented that resolves the problem highlighted in the previous section. The key idea is to solve an NLP problem with enlarged feasible region whenever an infeasible subproblem is encountered. The subgradient information from this second NLP is then used in Benders Decomposition. In this way it is possible to glean objective information from infeasible subproblems.

Let $\mathbf{y}^{(k)}$ be a set of complicating variables for which the subproblem $(S(\mathbf{y}^{(k)}))$ is infeasible. The NLP solver will detect infeasibility by generating a point $\hat{\mathbf{x}}^{(k)}$ which solves the feasibility problem $(F(\mathbf{y}^{(k)}))$. Let $\hat{\mathbf{c}}^{(k)} = \mathbf{c}(\hat{\mathbf{x}}^{(k)}, \mathbf{y}^{(k)})$ denote the corresponding constraint values and let $\hat{\mathbf{c}}^{(k)+} = \max(\mathbf{0}, \hat{\mathbf{c}}^{(k)})$ be the vector of infeasibilities (here the max is taken componentwise). Let $m_1 > 1$ be a constant and define a relaxed NLP subproblem as

$$(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)+}))\left\{\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(\mathbf{x},\mathbf{y}) \\ \text{subject to} & \mathbf{c}(\mathbf{x},\mathbf{y}) \le m_1\hat{\mathbf{c}}^{(k)+} \\ & \mathbf{y} = \mathbf{y}^{(k)} \\ & (\mathbf{x},\mathbf{y}) \in X \times Y. \end{array}\right.$$

Clearly, $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)+}))$ is feasible by construction and its solution provides objective information that can be used in Benders Decomposition. The Benders cut derived from this auxiliary problem has the form

$$\eta \ge f^{(k)} + \boldsymbol{\lambda}^{(k)^T}\mathbf{c}^{(k)} + \hat{\mathbf{w}}^T(\mathbf{y} - \mathbf{y}^{(k)}),$$

where $\hat{\mathbf{w}} \in \partial v(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)+}))$ and $\lambda^{(k)}$ are the multipliers of $\mathbf{c}$ in $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)+}))$. The choice $m_1 > 1$ implies that $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)+}))$ satisfies a Slater type constraint qualification.

This gives rise to the following algorithm which only differs from the algorithms presented in [3] and [2] in the way in which infeasible subproblems are treated.

Benders Decomposition with Feasibility Restoration

Initialization: Given y^0, set $k = 0$, upper bound $U^k = \infty$ lower bound $L^k = -\infty$.

REPEAT

Attempt to solve $(S(\mathbf{y}^{(k)}))$.
IF $((S(\mathbf{y}^{(k)}))$ feasible) **THEN**
Add an optimality cut to the master problem, set $\mathcal{S}^{k+1} = \mathcal{S}^k \cup \{k\}$.
Update the upper bound $U^k = \min(U^k, f^{(k)})$ and the current best point.
ELSE
Solve $(F(\mathbf{y}^{(k)}))$ and determine $\mathbf{c}^{(k)^+}$. and solve $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)^+}))$.
Derive a feasibility cut from $(F(\mathbf{y}^{(k)}))$ and an optimality cut from $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)^+}))$.
Set $\mathcal{F}^{k+1} = \mathcal{F}^k \cup \{k\}$.
ENDIF
Solve the master program relaxation (M^k) and obtain a new $\mathbf{y}^{(k+1)}$ and a new lower bound L^k.
Increase $k = k + 1$. **UNTIL** $(U^{k-1} - L^{k-1} < \epsilon)$

The algorithm appears cumbersome, if infeasible subproblems are encountered, solving apparently 3 different NLP problems. Note, however, that most NLP codes provide a solution to $(F(\mathbf{y}^{(k)}))$, if infeasibility is detected. Moreover, this point is usually a good starting point for the solution of $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)^+}))$ so that the overhead is not too severe.

Next some Lemmas are presented which are useful in establishing convergence for the new feasibility restoration. Since the algorithm only differs in the feasibility restoration, it suffices to show that the cuts generated in that part are underestimators and that they ensure that the algorithm does not get trapped at a nonstationary point. Throughout it is assumed that assumptions **A1** - **A3** hold.

Lemma 3.1 There exists a converging subsequence $\mathbf{y}^{(k)} \to \mathbf{y}^*$ and $\mathbf{y}^*$ is feasible.

Proof: The existence of a converging subsequence follows from the compactness of Y (**A2**). Now assume that the limit $\mathbf{y}^*$ is *not* feasible and seek a contradiction.
Since $\mathbf{y}^*$ is not feasible, there exists a (possibly different) subsequence $\mathbf{y}^{(k_i)} \to \mathbf{y}^*$ such that $\mathbf{y}^{(k_i)}$ are not feasible. Benders Decomposition adds a feasibility cut at each iteration of this subsequence which has the form

$$0 \geq v^0(\mathbf{y}^{(k_i)}) + \mathbf{w}^{(k_i)}(\mathbf{y} - \mathbf{y}^{(k_i)}).$$

The limit $\mathbf{y}^*$ must be feasible in these cuts for all k. Now consider each of the terms in turn: $v^0(\mathbf{y}^{(k_i)})$ converges to $v^0(\mathbf{y}^*) > 0$ while $\mathbf{w}^{(k_i)}$ remain bounded (**A3**) and $\mathbf{y}^* - \mathbf{y}^{(k_i)} \to 0$. Thus in the limit, $0 \geq v^0(\mathbf{y}^*) > 0$ which is a contradiction. □

Lemma 3.2 The value functions $v(\mathbf{y})$ and $v^0(\mathbf{y})$ are convex.

Proof: This follows directly from the convexity of f and $\mathbf{c}$. □

Lemma 3.3 The cuts derived from $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)^+}))$ are underestimators of the value function $v(\mathbf{y})$.

Proof: Assume that $\|\hat{\mathbf{c}}^{(k)^+}\| > 0$, otherwise the result follows from the fact that Benders cuts underestimate the value function. Let $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$ be the solution to $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)^+}))$. The convexity of f, $\mathbf{c}$ implies that

$$\eta \;\geq\; f^{(k)} + \nabla f^{(k)^T} \begin{pmatrix} \mathbf{x} - \mathbf{x}^{(k)} \\ \mathbf{y} - \mathbf{y}^{(k)} \end{pmatrix} \tag{1}$$

$$0 \;\geq\; \mathbf{c}^{(k)} - m_1\hat{\mathbf{c}}^{(k)^+} + \nabla \mathbf{c}^{(k)^T} \begin{pmatrix} \mathbf{x} - \mathbf{x}^{(k)} \\ \mathbf{y} - \mathbf{y}^{(k)} \end{pmatrix} \tag{2}$$

are outer approximations of the feasible region and underestimate the objective. Let $\boldsymbol{\lambda}^{(k)}$ be the multipliers of the "$\mathbf{c}$ constraints" of $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)^+}))$ and let $\boldsymbol{\mu}^{(k)}$ be an element of the normal cone to $X \times Y$ at $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$. Summing the outer approximations (1) and (2) weighted with $(1, \boldsymbol{\lambda}^{(k)})$ gives the valid inequality

$$\eta \geq f^{(k)} + \boldsymbol{\lambda}^{(k)^T}\mathbf{c}^{(k)} + \left[\nabla f^{(k)} + \nabla \mathbf{c}^{(k)}\boldsymbol{\lambda}^{(k)} + \boldsymbol{\mu}^{(k)}\right] \begin{pmatrix} \mathbf{x} - \mathbf{x}^{(k)} \\ \mathbf{y} - \mathbf{y}^{(k)} \end{pmatrix}$$

The Kuhn-Tucker conditions of $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)^+}))$ imply that

$$\nabla_x f^{(k)} + \nabla_x \mathbf{c}^{(k)}\boldsymbol{\lambda}^{(k)} + \boldsymbol{\mu}_x^{(k)} = \mathbf{0}.$$

Hence

$$\eta \geq f^{(k)} + \boldsymbol{\lambda}^{(k)^T}\mathbf{c}^{(k)} + \left[\nabla_y f^{(k)} + \nabla_y \mathbf{c}^{(k)}\boldsymbol{\lambda}^{(k)} + \boldsymbol{\mu}_y^{(k)}\right]^T (\mathbf{y} - \mathbf{y}^{(k)})$$

is a Benders cut which underestimates $v(\mathbf{y})$. The Kuhn-Tucker conditions of $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)^+}))$ imply that the subgradient in this last cut is equal to the multipliers of the constraint which fixes $\mathbf{y} = \mathbf{y}^{(k)}$. □

Now we can prove the main convergence result.

Theorem 3.1 Benders Decomposition with feasibility restoration converges to an optimal solution.

Proof: Lemma 3.1 implies that the limit point exists and is feasible.

Lemma 3.3 ensures that any Benders cut derived from $(R(\mathbf{y}^{(k)}, m_1\hat{\mathbf{c}}^{(k)^+}))$ underestimates the value function $v(\mathbf{y})$. Thus the value of (M^k) always underestimates that of (P).

Now assume that there exists a converging subsequence $\mathbf{y}^{(k)}$ which converges to a nonstationary point $\hat{\mathbf{y}}$ and seek a contradiction. From Lemma 3.1 it follows that $\hat{\mathbf{y}}$ is feasible. Let $\hat{f} = v(\hat{\mathbf{y}})$ be the corresponding objective value. Then it follows that $\hat{f} > f^*$, the optimal value.

The Benders cut added at each iteration has the form

$$\eta \geq f^{(k)} + \boldsymbol{\lambda}^{(k)^T}\mathbf{c}^{(k)} + \hat{\mathbf{w}}^{(k)^T}(\mathbf{y} - \mathbf{y}^{(k)})$$

if $\mathbf{y}^{(k)}$ is infeasible and

$$\eta \geq f^{(k)} + \mathbf{w}^{(k)^T}(\mathbf{y} - \mathbf{y}^{(k)})$$

if it is feasible. Now take the limit as $k \to \infty$ and observe that the complementarity $\boldsymbol{\lambda}^{(k)^T}\mathbf{c}^{(k)} \to 0$ as the auxiliary problem approaches feasibility. Moreover, the subgradients are bounded by assumption **A3** and hence in the limit

$$\eta \geq \lim_{k\to\infty} f^{(k)} = \hat{f} > f^*$$

by continuity of the value function. But this contradicts the fact that the master programs (M^k) are underestimators of the optimal value f^*. □

3.1 Relevance to bundle-based decomposition

Bundle-based decomposition methods build up a bundle of ϵ-subgradients of $v(\mathbf{y})$ and $v^0(\mathbf{y})$. This bundle defines a local piecewise linear model of the nonsmooth master problem (M). This model is then minimized inside a trust-region. The step from the model is either accepted in which case the algorithm moves to a new point, discarding (part of) the bundle or the step is rejected in which case a new cut may be added to the bundle or the trust-region is reduced.

There exist various bundle schemes (e.g. [6], [5], [4] and references therein) which could be adopted to solve (P) by a decomposition scheme. Since bundle methods use the same cuts as Benders Decomposition, the example of Section 2 and the feasibility restoration introduced above are clearly relevant in this situation. Our development shows that it is necessary to include a feasibility restoration in bundle-based decomposition schemes and how this can be done.

References

[1] Fletcher, R. and Leyffer, S. Nonlinear programming without a penalty function. Numerical Analysis Report NA/171, Department of Mathematics, University of Dundee, September 1997.

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

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

[4] A. Grothey and K.I.M. McKinnon. A superlinearly convergent trust region bundle method. Report, Department of Mathematics & Statistics, Edinburgh University, December 1998.

[5] L. Luksan and J. Vlcek. A bundle-Newton method for nonsmooth unconstrained minimization. *Mathematical Programming*, 83(3):373–391, 1998.

[6] H. Schramm and J. Zowe. 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.