INT. J. SYSTEMS SCI., 1987, VOL. 18, NO. 1, 157–176

Exact augmented lagrangian approach to multilevel optimization of large-scale systems

ALESSANDRO DE LUCA† and GIANNI DI PILLO†

An approach based on the exact augmented lagrangian function is developed for the optimization of large-scale systems composed of interconnected units. The decomposition and coordination strategies are examined and various schemes of upper-level coordination are proposed, all of which are formulated as unconstrained quadratic minimization problems. Convergence analysis is performed exploiting an analogy with minimization by relaxation methods. Numerical examples are reported.

1. Introduction

Optimization methods for large-scale systems are basic tools for solving real-world problems. The wide range of applications covers different engineering areas such as energy production and distribution systems, industrial process control, economic planning systems, oversaturated traffic networks and water resources systems (see, for example, Wismer 1971, Mahmoud 1977, Singh and Titli 1978, Haimes 1982).

The common characteristic of these complex systems is that their mathematical models are of large dimension, but structured. Usually this means that large-scale systems have an underlying physical or functional structure of interacting constituents: a certain number of interconnected subsystems may thus be identified in the analysis stage.

Motivated by the presence of this structure, specific decomposition and coordination methods have been developed for the purpose of optimization. The practical advantages of partitioning optimization tasks rely on the ease of solving, although recursively, smaller and/or simpler subproblems. Early examples of this approach are the decomposition algorithm of Dantzig and Wolfe (1961) in linear programming and the multilevel algorithm for optimal control (Takahara 1965). The first formalization of the multilevel methodology was given by Mesarovic *et al.* (1970), who stated two principles of coordinability of subproblems.

In the mathematical programming framework, the first multilevel methods were based on the lagrangian approach (Lasdon 1970, Schoeffler 1971), since the lagrangian function retains the separability of the problem formulation. A detailed description of the model coordination, goal coordination and mixed methods can be found in Titli (1972) and in Singh and Titli (1978), together with their extension to the dynamical case. Duality gaps limited the effectiveness of these methods in the absence of convexity assumptions. As augmented lagrangian theory was developed (see, for example, Bertsekas 1976), its application to multilevel problems was investigated; although resolving the dual gaps, this approach has the drawback of destroying

Received 12 August 1985.

† Dipartimento di Informatica e Sistemistica, Università di Roma 'La Sapienza', Via Eudossiana 18, 00184 Roma, Italy.

separability, due to the added penalty term. To overcome this difficulty Stephanopoulos and Westerberg (1975) proposed a linear approximation of the non-separable crossterms, while Watanabe *et al.* (1978) transformed these terms into the minimum of a sum of separable terms, obtaining a three-level strategy. No convergence results are reported in these papers, but the latter is supposed to be superior due to the more efficient coordination process.

Findeisen *et al.* (1980), Stoilov (1977), Tatjewski and Michalak (1980) proposed two new mixed methods based on a suitable augmentation of the lagrangian function without approximating terms or perturbing the problem. Both methods coordinate the subproblems, predicting the value of a set of variables (the inputs or the outputs of the subsystems) together with the constraints multipliers and then updating so as to balance the interactions. The coordinator task is formulated as a saddle-point problem and single-level or two-level coordination algorithms are presented. The multipliers are updated making use of the Hestenes–Powell rule. The main disadvantages of this approach are in the difficult definition of feasible sets for the predictions and in the complications that arise for the coordination updates in the presence of local inequality constraints where both interaction variables and control inputs are present.

An alternative augmentation of the lagrangian function was proposed by Bertsekas (1979), where local convexification is obtained simply by duplicating the number of primal variables so as to preserve the separability of the problem. It is worth mentioning that other approaches exist based on penalty functions (Tatjewski 1978) or on decomposing functions to be added to the ordinary lagrangian (Brusilovski and Ostrovski 1983). The first is hampered by the known limitations of penalty methods so that the use is restricted to cases where low accuracy is needed. The second has properties similar to all convexification methods, but requires that the feasible sets of the local variables be convex.

Differences in basic decomposition schemes and applicability conditions, lack of convergence analysis and results and the application-oriented nature of multilevel procedures made the classification and comparison of various methods a difficult task. Three theoretical papers have tried, almost successfully, to fill this gap: Cohen (1978, 1980) imbedded all the previous decomposition–coordination methods in a general framework using the auxiliary problem and the relaxation concepts; Looze and Sandell (1981) provided a local analysis of the iterative behaviour of the various decomposition schemes by the use of non-linear splitting functions.

As a further development we propose in this paper a method based on recent results in non-linear programming, namely, the exact augmented lagrangian approach introduced by Di Pillo and Grippo (1979, 1982) and Lucidi (1985), which consists of adding to the lagrangian function a penalty term on the whole subset of the first-order necessary conditions corresponding to equalities.

In particular, it was shown by Di Pillo and Grippo (1979) that, under suitable hypotheses, the solution of an equality constrained problem of the form:

$$\text{minimize } J(s)$$

$$\text{subject to } g(s) = 0, \quad s \in \mathbb{R}^n$$

and the corresponding Lagrange multiplier $\sigma \in \mathbb{R}^m$, can be found by computing the unconstrained minimum, with respect to s and σ, of the function

$$S(s, \sigma) = J(s) + \sigma^{\mathrm{T}} g(s) + \eta \| g(s) \|^2 + \| M(s)(\nabla J(s) + \nabla g(s)\sigma) \|^2$$

for a value of the penalty coefficient η larger than a threshold value $\eta^* > 0$ and for an appropriate choice of the matrix $M(s)$ such that $M\nabla g$ is an $m \times m$ non-singular matrix, where ∇g denotes the transpose of the Jacobian matrix of the constraints.

The extension of the proposed approach to non-linear programming problems with inequality constraints:

$$\text{minimize } J(s)$$

$$\text{subject to } g(s) \leqslant 0, \quad s \in \mathbb{R}^n$$

was studied in Di Pillo and Grippo (1982) and in Lucidi (1985) by using the device of converting inequalities to equalities via squared slack variables. In particular it was shown in Lucidi (1985) that the solution of the above problem can be obtained by the unconstrained minimization, with respect to $s \in \mathbb{R}^n$ and $\sigma \in \mathbb{R}^m$, of the continuously differentiable function:

$$T(s, \sigma) = J(s) + \sigma^T [g(s) + Y(s, \sigma) y(s, \sigma)] + \eta \| g(s) + Y(s, \sigma) y(s, \sigma) \|^2$$

$$+ \mu \left\| \frac{\partial g(s)}{\partial s} (\nabla J(s) + \nabla g(s) \sigma) + \gamma^2 G^2(s, \sigma) \sigma \right\|^2$$

where

$$y_i^2(s, \sigma) = -\min \left\{ 0, g_i(s) + \frac{\sigma_i}{2\eta} \right\}$$

$$G(s, \sigma) \triangleq \text{diag} \{ g_i(s) \}$$

$$Y(s, \sigma) \triangleq \text{diag} \{ y_i(s, \sigma) \}$$

for a given value of $\mu > 0$, $\gamma \neq 0$ and a value of η larger than a threshold value $\eta^* > 0$.

In both cases, the search for a saddle point of the ordinary lagrangian is replaced with the search for an unconstrained minimum of $S(s, \sigma)$ or $T(s, \sigma)$.

Applications of this approach to the large-scale non-linear problems formulated in §§ 2 and 3, followed by a suitable decomposition, allows us:

(*a*) to solve the problem by a general multilevel method with an efficient coordination process without the need of convexity assumptions;

(*b*) to look at multilevel methods as a straightforward generalization of well-known minimization by relaxation methods;

(*c*) to derive simple convergence analysis and an acceleration procedure for the proposed algorithm.

The paper is organized as follows. In § 2, we consider an exact augmented lagrangian-function approach to the equality constrained large-scale problem. Section 3 describes the proposed method, that is, the decomposition of the exact augmented lagrangian function and the coordination algorithms, for the more general large-scale problem with both equality and inequality constraints. Convergence analysis based on an analogy with minimization algorithms using relaxation is developed in § 4. Section 5 is dedicated to two numerical examples taken from the literature. Conclusions are drawn in § 6 together with a discussion of the obtained results.

2. Large-scale equality constrained problems

We consider a large-scale optimization problem given in the following form.

Problem 1

$$\min f(x, c) = \sum_{i=1}^{N} f_i(x_i, c_i)$$

$$\text{subject to} \quad z_i = t_i(x_i, c_i)$$

$$x_i = \sum_{j=1}^{N} H_{ij} z_j, \quad i = 1, 2, \ldots, N$$

where $x_i \in \mathbb{R}^{n_i}$, $z_i \in \mathbb{R}^{k_i}$ and $c_i \in \mathbb{R}^{m_i}$ are the ith subsystem interaction inputs and outputs and the ith local control vector; the $f_i : \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \to \mathbb{R}$ are local objective functions and $t_i : \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \to \mathbb{R}^{k_i}$ are the input–output subsystem models. The H_{ij} are $(n_i \times k_j)$ interconnection matrices whose elements are 0/1.

Problem 1 can be interpreted as the task of regulating N (usually large) interconnected static systems, minimizing the sum of the local (generally non-linear) cost functions.

Remark 1

The subsystem linear interconnections imply no loss of generality, since all non-linearities can be included in the subsystem models. By a suitable definition of subsystems one can also assume that each (scalar) output is connected to at most one (scalar) input.

Remark 2

Usually it is assumed (Titli 1972) that each and every output is connected to one and only one input of a different subsystem, thus the matrix H, whose blocks are the H_{ij}, is an orthonormal matrix, that is, $H^{-1} = H^{\mathrm{T}}$. Local feedbacks are included in the subsystem models so that $H_{ii} = 0$, for all i.

Problem 1 can be rewritten in a more compact form as

$$\min f(x, c)$$

$$\text{subject to} \quad g(x, c, z) = \begin{bmatrix} t(x, c) - z \\ Hz - x \end{bmatrix} = 0$$

where

$$x \triangleq [x_1^{\mathrm{T}} \quad \ldots \quad x_N^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^n, \quad n = \sum_{i=1}^{N} n_i$$

and analogously, $z \in \mathbb{R}^k$, $c \in \mathbb{R}^m$; furthermore, $t \triangleq [t_1^{\mathrm{T}} \quad \ldots \quad t_N^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^k$ and $H \triangleq \{H_{ij}\}$, an $(n \times k)$ matrix. Following Remark 2, we set from now on $k = n$, $H^{-1} = H^{\mathrm{T}}$.

Let the following assumptions hold for Problem 1, where Ω is a given compact subset of $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$.

Assumption 1. The functions f, t are twice continuously differentiable with respect to the variables x and c on $\mathbb{R}^n \times \mathbb{R}^m$.

Assumption 2. The gradients of the constraints are linearly independent at every point (x, c, z) in the compact subset Ω.

The lagrangian function for Problem 1 is defined as:

$$L(x, c, z, \lambda, p) \triangleq f(x, c) + \lambda^{\mathrm{T}}(t(x, c) - z) + p^{\mathrm{T}}(Hz - x)$$

where $\lambda \in \mathbb{R}^n$, $p \in \mathbb{R}^n$ are the Lagrange multipliers.

An exact augmented lagrangian function for this problem is

$$S(x, c, z, \lambda, p; \eta) \triangleq L(x, c, z, \lambda, p) + \eta(\|t(x, c) - z\|^2 + \|Hz - x\|^2) + \|M(x, c, z)\nabla L(x, c, z, \lambda, p)\|^2 \quad (1)$$

where M is a $2n \times (2n + m)$ matrix whose elements are twice continuously differentiable and ∇L denotes the gradient of L with respect to (x, c, z). Then, under the hypothesis that the matrix $M\nabla g$ is a $2n \times 2n$ non-singular matrix in Ω, the unconstrained minima of function S in $\Omega \times \mathbb{R}^n \times \mathbb{R}^n$ for η larger than a threshold value η^*, which depends on the compact set Ω, yield the solutions of Problem 1 contained in Ω and the corresponding Lagrange multipliers.

A proper choice of the matrix M for this case is indicated by the following proposition.

Proposition 1

If the matrices $\nabla_{c_i} t_i$, $i = 1, \ldots, N$ have full column rank, that is

$$\operatorname{rank}[\nabla_{c_i} t_i] = k_i \leqslant m_i$$

and if the matrix $M(x, c, z)$ is chosen as

$$M(x, c, z) = \mu^{1/2} \begin{bmatrix} 0 & \nabla_c t^{\mathrm{T}} & 0 \\ -I_n & 0 & 0 \end{bmatrix}, \quad \mu > 0 \quad (2)$$

then the matrix $M\nabla g$ is non-singular.

Proof

Since

$$\nabla g = \begin{bmatrix} \nabla_x g \\ \nabla_c g \\ \nabla_z g \end{bmatrix} = \begin{bmatrix} \nabla_x t & -I_n \\ \nabla_c t & 0 \\ -I_n & H^{\mathrm{T}} \end{bmatrix}$$

and $\nabla_c t = \operatorname{diag}(\nabla_{c_i} t_i)$ is of rank $k = \sum_{i=1}^{N} k_i = n$, then by direct computation

$$M\nabla g = \mu^{1/2} \begin{bmatrix} \nabla_c t^{\mathrm{T}} \nabla_c t & 0 \\ \nabla_x t & I_n \end{bmatrix}$$

is a $2n \times 2n$ non-singular matrix. □

The hypothesis of the above proposition is a condition of the local controllability kind: every subsystem has to be regulated by a number of effective control inputs not less than the number of local interconnection outputs. This is a standard hypothesis usually verified in a non-trivial large-scale system; it is the same needed for the classical model coordination approach (Singh and Titli 1978). By straightforward

calculation of $\nabla L = [\nabla_x L^T \nabla_c L^T \nabla_z L^T]^T$ and substitution of (2) in the expression of S, we have:

$$\begin{aligned} S(x, c, z, \lambda, p; \eta, \mu) = {} & f(x, c) + \lambda^T(t(x, c) - z) + p^T(Hz - x) \\ & + \eta(\|t(x, c) - z\|^2 + \|Hz - x\|^2) \\ & + \mu(\|\nabla_c t^T(\nabla_c f + \nabla_c t\lambda)\|^2 + \|\nabla_x f + \nabla_x t\lambda - p\|^2) \end{aligned} \tag{3}$$

The solution of Problem 1 can then be found by minimizing (3) with respect to x, c, z, λ, p. This can be accomplished by the following two-level iterative procedure, where $H_i \triangleq [H_{i1} \quad H_{i2} \quad \dots \quad H_{iN}]$.

(*a*) For fixed values $\bar{z}, \bar{p}$ minimize S with respect to x, c, λ; this problem splits into N independent low-level problems:

$$\min_{x_i, c_i, \lambda_i} S^i(x_i, c_i, \bar{z}, \lambda_i, \bar{p}; \eta, \mu)$$

where

$$\begin{aligned} S^i(x_i, c_i, \bar{z}, \lambda_i, \bar{p}; \eta, \mu) \triangleq {} & f_i(x_i, c_i) + \lambda_i^T(t_i(x_i, c_i) - \bar{z}_i) \\ & + \bar{p}_i^T(H_i\bar{z} - x_i) + \eta(\|t_i(x_i, c_i) - \bar{z}_i\|^2 + \|H_i\bar{z} - x_i\|^2) \\ & + \mu(\|\nabla_{c_i} t_i^T(\nabla_{c_i} f_i + \nabla_{c_i} t_i\lambda_i)\|^2 + \|\nabla_{x_i} f_i + \nabla_{x_i} t_i\lambda_i - \bar{p}_i\|^2) \end{aligned} \tag{4}$$

(*b*) At the coordination level, use the low-level solutions $\bar{x}, \bar{c}, \bar{\lambda}$ and minimize S with respect to z, p.

This way of treating different variables separately allows both a decomposition of S and a very simple second-level task. In addition, due to the quadratic dependence of S on z, p, the coordination is a well-behaved problem. We have the following expressions for the gradients and the Hessian of S with respect to the coordinating variables:

$$\left.\begin{aligned} \nabla_z S &= H^T p - \lambda + 2\eta(2z - t(x, c) - H^T x) \\ \nabla_p S &= Hz - x + 2\mu(p - \nabla_x f(x, c) - \nabla_x t(x, c)\lambda) \\ \nabla^2_{(z,p)} S &= \begin{bmatrix} 4\eta I_n & H^T \\ H & 2\mu I_n \end{bmatrix} \end{aligned}\right\} \tag{5}$$

The Hessian matrix is positive definite for $\mu > 0$ and η sufficiently large, thus implying that the first-order necessary conditions at the coordinating level:

$$\nabla_z S = 0, \qquad \nabla_p S = 0$$

are also sufficient for a minimum with respect to z and p, for every value of the first-level variables $\bar{x}, \bar{c}$ and $\bar{\lambda}$. Furthermore the Hessian matrix (5) is easily invertible. We have analytically:

$$[\nabla^2_{(z,p)} S]^{-1} = \frac{1}{8\mu\eta - 1} \begin{bmatrix} 2\mu I_n & -H^T \\ -H & 4\eta I_n \end{bmatrix}$$

The coordinator task can be accomplished by a Newton iteration which gives new

updates $\hat{z}$ and $\hat{p}$ in the closed form:

$$\begin{bmatrix}\hat{z}\\ \hat{p}\end{bmatrix} = \begin{bmatrix}\bar{z}\\ \bar{p}\end{bmatrix} - [\nabla^2_{(z,p)}S]^{-1}\begin{bmatrix}\nabla_z S(\bar{x},\bar{c},\bar{z},\bar{\lambda},\bar{p})\\ \nabla_p S(\bar{x},\bar{c},\bar{z},\bar{\lambda},\bar{p})\end{bmatrix}$$

$$= \begin{bmatrix}\bar{z}\\ \bar{p}\end{bmatrix} + \frac{1}{8\mu\eta-1}\begin{bmatrix}2\mu I_n & -H^{\mathrm{T}}\\ -H & 4\eta I_n\end{bmatrix}\begin{bmatrix}\bar{\lambda}+2\eta(t(\bar{x},\bar{c})+H^{\mathrm{T}}\bar{x})\\ \bar{x}+2\mu(\nabla_x f(\bar{x},\bar{c})+\nabla_x t(\bar{x},\bar{c})\bar{\lambda})\end{bmatrix}$$

$$-\frac{1}{8\mu\eta-1}\begin{bmatrix}2\mu I_n & -H^{\mathrm{T}}\\ -H & 4\eta I_n\end{bmatrix}\begin{bmatrix}H^{\mathrm{T}}\bar{p}+4\eta\bar{z}\\ H\bar{z}+2\mu\bar{p}\end{bmatrix}$$

$$= \frac{1}{8\mu\eta-1}\begin{bmatrix}2\mu I_n & -H^{\mathrm{T}}\\ -H & 4\eta I_n\end{bmatrix}\begin{bmatrix}\bar{\lambda}+2\eta(t(\bar{x},\bar{c})+H^{\mathrm{T}}\bar{x})\\ \bar{x}+2\mu(\nabla_x f(\bar{x},\bar{c})+\nabla_x t(\bar{x},\bar{c})\bar{\lambda})\end{bmatrix} \tag{6}$$

A basic two-level algorithm for solving Problem 1 by means of the proposed exact augmented lagrangian function may now be given.

Algorithm 1

Step 1. Choose η, μ and a starting point in the extended space of primal and dual variables; label it by 0 and set $k=0$.

Step 2. Solve the N low-level subproblems for fixed z^k, p^k:

$$\min_{x_i,c_i,\lambda_i} S^i(x_i, c_i, z^k, \lambda_i, p_i^k), \quad i=1,2,\ldots,N$$

and denote by x_i^{k+1}, c_i^{k+1}, λ_i^{k+1} the minimizing values.

Step 3. Calculate z^{k+1}, p^{k+1} by (6) for $\bar{x}=x^{k+1}$, $\bar{c}=c^{k+1}$, $\bar{\lambda}=\lambda^{k+1}$.

Step 4. If some stopping criterion is satisfied, then set

$$x^*=x^{k+1}; \qquad c^*=c^{k+1}; \qquad z^*=z^{k+1}; \qquad p^*=p^{k+1}; \qquad \lambda^*=\lambda^{k+1}$$

and stop; else set $k=k+1$ and go to Step 2.

A feature of this algorithm is that it is unaffected by the number N of interconnected systems. Notice also that for decomposing the function S it would be necessary to fix only the interconnection outputs z at the second level; inclusion of p as a coordinating variable allows a more balanced partition of tasks among levels at practically no additional cost for the coordination procedure. As for every multilevel method, it should be noted that independent solving of subproblems at the lower level can take advantage of any special local structure such as linearity or convexity, so that each subsystem can be optimized by the appropriate method (numerical or analytical). The convergence properties of the above algorithm will be discussed in § 4.

Comparing this approach with the one based on augmented lagrangians (Findeisen *et al.* 1980), it should be noted that no direct substitution has been made of x in terms of z (or vice versa) using the interconnection equation explicitly; our choice, despite the increase of dimensionality of the local subproblems, is motivated by the following observations:

(*a*) no such direct substitution is possible when local inequality constraints are

present, unless they were separable in terms of controls and interconnection inputs (Tatjewski and Michalak 1980);

(*b*) the appealing coordination process (6) can only be obtained because of this duplication of variables, which further preserves separability of the problem formulation.

3. Large-scale problems with both equality and inequality constraints

A more general large-scale optimization problem includes, beside the equality constraints considered in Problem 1, additional local inequality constraints of the form:

$$v_i(x_i, c_i) \leqslant 0, \quad i = 1, 2, \ldots, N$$

where $v_i : \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \to \mathbb{R}^{r_i}$. Transforming these inequalities via squared slack variables into

$$v_i(x_i, c_i) + Y_i y_i = 0, \quad i = 1, 2, \ldots, N$$

with $y_i \in \mathbb{R}^{r_i}$, $Y_i \triangleq \text{diag}\,\{y_{ij}, \quad j = 1, \ldots, r_i\}$, and using a compact notation, we are lead to consider the following.

Problem 2

$$\min f(x, c)$$

$$\text{subject to} \quad g(x, c, z, y) = \begin{bmatrix} t(x, c) - z \\ Hz - x \\ \cdots\cdots\cdots \\ v(x, c) + Yy \end{bmatrix} = \begin{bmatrix} g_1(x, c, z) \\ \cdots\cdots\cdots \\ g_2(x, c, y) \end{bmatrix} = 0$$

where $Y \triangleq \text{diag}\,\{Y_i\}$, $y \triangleq [y_1^T \quad \ldots \quad y_N^T]^T$, $v \triangleq [v_1^T \quad \ldots \quad v_N^T]^T \in \mathbb{R}^r$ and $r = \sum_{i=1}^{N} r_i$. The lagrangian function for Problem 2 is defined as:

$$L(x, c, z, y, \lambda, p, \rho) \triangleq f(x, c) + \lambda^T(t(x, c) - z) + p^T(Hz - x) + \rho^T(v(x, c) + Yy)$$

where $\rho \in \mathbb{R}^r$ is the Kuhn–Tucker multiplier.

Given a compact subset Ω of $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$, the following assumptions are assumed to hold for Problem 2.

Assumption 3. The functions f, t, v are twice continuously differentiable with respect to the variables x and c on $\mathbb{R}^n \times \mathbb{R}^m$.

Assumption 4. The gradients of the active constraints are linearly independent at every point (x, c, z) in the compact subset Ω.

Assumption 5. Define the index set $J^i(x, c)$ as

$$J^i(x, c) \triangleq \{j : v_{ij}(x_i c_i) \geqslant 0\}$$

The Mangasarian–Fromowitz regularity assumption then holds on Ω, that is, for any $(x, c, z) \in \Omega$

$$\sum_{k=1}^{2n} \alpha_k \nabla g_{1,k} + \sum_{i=1}^{N} \sum_{j \in J^i} \beta_j^i \nabla v_{ij} = 0$$

with $\beta_j^i \geqslant 0$ for all $j \in J^i$, $i = 1, \ldots, N$, implies that $\beta_j^i = 0$ for all $j \in J^i$, $i = 1, \ldots, N$ and $\alpha_k = 0$ for $k = 1, \ldots, 2n$; here $g_{1,k}$ denotes the kth component of g_1 and the gradients are taken with respect to the primal variables (x, c, z).

Assumption 6. Strict complementarity holds at any Kuhn–Tucker pair $(\bar{x}, \bar{c}, \bar{z})$, $(\bar{\lambda}, \bar{p}, \bar{\rho})$ such that $(\bar{x}, \bar{c}, \bar{z})$ belongs to the compact set Ω.

An exact augmented lagrangian function can be constructed for Problem 2, considering the presence of both equality and inequality constraints. We get:

$$\begin{aligned} T(x, c, z, \lambda, p, \rho) = {} & f(x, c) + \lambda^{\mathrm{T}}(t(x, c) - z) + p^{\mathrm{T}}(Hz - x) \\ & + \rho^{\mathrm{T}}(v(x, c) + Y(x, c, \rho)y(x, c, \rho)) \\ & + \eta(\|t(x, c) - z\|^2 + \|Hz - x\|^2 + \|v(x, c) + Y(x, c, \rho)y(x, c, \rho)\|^2) \\ & + \mu(\|\nabla g_1^{\mathrm{T}} \nabla L\|^2 + \|\nabla g_2^{\mathrm{T}} \nabla L + \gamma^2 V^2(x, c)\rho\|^2) \end{aligned} \tag{7}$$

where

$$\left.\begin{aligned} y_{ij}^2(x_i c_i \rho_i) &= -\min\left\{0, v_{ij}(x_i c_i) + \frac{\rho_{ij}}{2\eta}\right\}, \quad j = 1, \ldots, r_i, \quad i = 1, \ldots, N \\ Y(x, c, \rho) &= \operatorname{diag}\{y_{ij}(x_i c_i \rho_i), \quad j = 1, \ldots, r_i, \quad i = 1, \ldots, N\} \\ \mathrm{V}(x, c) &= \operatorname{diag}\{v_{ij}(x_i c_i), \quad j = 1, \ldots, r_i, \quad i = 1, \ldots, N\} \end{aligned}\right\} \tag{8}$$

∇L denotes the gradient of L with respect to (x, c, z) while ∇g_1 and ∇g_2 denote the transpose of the Jacobian matrix of the equality and inequality constraints. In all cases derivatives are taken with respect to the primal variables of the given problem. We have:

$$\nabla L = \begin{bmatrix} \nabla_x L \\ \nabla_c L \\ \nabla_z L \end{bmatrix} = \begin{bmatrix} \nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p \\ \nabla_c f + \nabla_c t\lambda + \nabla_c v\rho \\ H^{\mathrm{T}} p - \lambda \end{bmatrix}$$

$$\nabla g = \begin{bmatrix} \nabla_x g \\ \nabla_c g \\ \nabla_z g \end{bmatrix} = \begin{bmatrix} \nabla_x t & -I & \nabla_x v \\ \nabla_c t & 0 & \nabla_c v \\ -I & H^{\mathrm{T}} & 0 \end{bmatrix}$$

Finally, back substitution of the explicit expressions of the $y_{ij}(x_i c_i \rho_i)$ (8) into (7) yields:

$$\begin{aligned} T(x, c, z, \lambda, p, \rho; \mu, \eta, \gamma) = {} & f(x, c) + \lambda^{\mathrm{T}}(t(x, c) - z) + p^{\mathrm{T}}(Hz - x) + \rho^{\mathrm{T}} v(x, c) \\ & + \eta(\|t(x, c) - z\|^2 + \|Hz - x\|^2 + \|v(x, c)\|^2) \\ & + \mu(\|\nabla_x t^{\mathrm{T}}(\nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p) \\ & \quad + \nabla_c t^{\mathrm{T}}(\nabla_c f + \nabla_c t\lambda + \nabla_c v\rho) + (\lambda - H^{\mathrm{T}} p)\|^2 \\ & + \|(p - H\lambda) - (\nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p)\|^2 \\ & + \|\nabla_x v^{\mathrm{T}}(\nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p) \\ & \quad + \nabla_c v^{\mathrm{T}}(\nabla_c f + \nabla_c t\lambda + \nabla_c \rho) + \gamma^2 V^2 \rho\|^2) \\ & - \eta \sum_{i=1}^{N} \sum_{j=1}^{r_i} \left[\min\left\{0, v_{ij}(x_i c_i) + \frac{\rho_{ij}}{2\eta}\right\}\right]^2 \end{aligned} \tag{9}$$

Again, as for Problem 1, the unconstrained minima of function T in $\Omega \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^r$, provided that η is larger than a threshold value η^* depending on Ω, yield the solutions of Problem 2 contained in Ω and the associated Kuhn–Tucker multipliers.

The addition of the local inequality constraints implies a relevant complication in the expression of T. Moreover, no analytic coordination is possible. Nevertheless by fixing the vectors z (interconnection outputs), p (interconnection constraints multipliers) and λ (input–output equations multipliers) at the values $\bar{z}$, $\bar{p}$ and $\bar{\lambda}$ respectively, T can be decomposed so that:

$$\min T = \min_{z,p,\lambda} \left\{ \min_{x,c,\rho} T \right\} = \min_{z,p,\lambda} \left\{ \sum_{i=1}^{N} \min_{x_i,c_i,\rho_i} T^i \right\}$$

where

$$\begin{aligned} T^i &= T^i(x_i, c_i, \bar{z}, \bar{p}, \bar{\lambda}, \rho_i; \mu, \eta, \gamma) \\ &\triangleq f_i(x_i, c_i) + \bar{\lambda}_i^{\mathrm{T}}(t_i(x_i, c_i) - \bar{z}_i) + \bar{p}_i^{\mathrm{T}}(H_i\bar{z} - x_i) + \rho_i^{\mathrm{T}} v_i(x_i, c_i) \\ &\quad + \eta(\|t_i(x_i, c_i) - \bar{z}_i\|^2 + \|H_i\bar{z} - x_i\|^2 + \|v_i(x_i, c_i)\|^2) \\ &\quad + \mu(\|\nabla_{x_i} t_i^{\mathrm{T}}(\nabla_{x_i} f_i + \nabla_{x_i} t_i \bar{\lambda}_i + \nabla_{x_i} v_i \rho_i - \bar{p}_i) \\ &\qquad + \nabla_{c_i} t_i^{\mathrm{T}}(\nabla_{c_i} f_i + \nabla_{c_i} t_i \bar{\lambda}_i + \nabla_{c_i} v_i \rho_i) + (\bar{\lambda}_i - H_i^{\mathrm{T}} \bar{p})\|^2 \\ &\quad + \|(\bar{p}_i - H_i\bar{\lambda}) - (\nabla_{x_i} f_i + \nabla_{x_i} t_i \bar{\lambda}_i + \nabla_{x_i} v_i \rho_i - \bar{p}_i)\|^2 \\ &\quad + \|\nabla_{x_i} v_i^{\mathrm{T}}(\nabla_{x_i} f_i + \nabla_{x_i} t_i \bar{\lambda}_i + \nabla_{x_i} v_i \rho_i - \bar{p}_i) \\ &\qquad + \nabla_{c_i} v_i^{\mathrm{T}}(\nabla_{c_i} f_i + \nabla_{c_i} t_i \bar{\lambda}_i + \nabla_{c_i} v_i \rho_i) + \gamma^2 V_i^2 \rho_i\|^2) \\ &\quad - \eta \sum_{j=1}^{r_i} \left[\min \left\{ 0, v_{ij}(x_i c_i) + \frac{\rho_{ij}}{2\eta} \right\} \right]^2 \end{aligned} \tag{10}$$

with $V_i \triangleq \operatorname{diag}\{v_{ij}, \quad j = 1, \ldots, r_i\}$.

Our primary goal is to derive now a coordinator as simple as possible for minimizing T with respect to z, p and λ. For this to be achieved in a multilevel framework, it is useful to introduce as many levels as necessary for having a decomposed solution. Of course, the more the levels used to obtain the solution, the slower the procedure should be. On the other hand, the decomposition should be carried out so that each level subproblem is easy to be solved. Notice that the exact augmented lagrangian function is quadratic with respect to all coordination variables z, p and λ. In principle this $3n$-dimensional quadratic problem can be solved directly, obtaining a one-level coordinator. However, since n is usually large, a decomposition of the coordinator task may result in a computational saving. We propose here two coordination algorithms which differ for the number of required levels, respectively three and two.

First we derive the gradients of T with respect to the coordinating variables:

$$\nabla_z T = H^{\mathrm{T}} p - \lambda + 2\eta(2z - t(x, c) - H^{\mathrm{T}} x)$$

$$\begin{aligned} \nabla_p T = Hz - x + 2\mu(&-(\nabla_x t + H)(\nabla_x t^{\mathrm{T}}(\nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p) \\ &+ \nabla_c t^{\mathrm{T}}(\nabla_c f + \nabla_c t\lambda + \nabla_c v\rho) + (\lambda - H^{\mathrm{T}} p)) \\ &+ 2I(p - H\lambda - (\nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p)) \\ &- \nabla_x v(\nabla_x v^{\mathrm{T}}(\nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p) + \nabla_c v^{\mathrm{T}}(\nabla_c f + \nabla_c t\lambda + \nabla_c v\rho) + \gamma^2 V^2 \rho)) \end{aligned}$$

$$\nabla_\lambda T = t(x, c) - z + 2\mu((I + \nabla_x t^{\mathrm{T}}\nabla_x t + \nabla_c t^{\mathrm{T}}\nabla_c t)(\nabla_x t^{\mathrm{T}}(\nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p) + \nabla_c t^{\mathrm{T}}(\nabla_c f + \nabla_c t\lambda + \nabla_c v\rho) + (\lambda - H^{\mathrm{T}}p))$$
$$- (\nabla_x t + H)^{\mathrm{T}}(p - H\lambda - (\nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p))$$
$$+ (\nabla_x t^{\mathrm{T}}\nabla_x v + \nabla_c t^{\mathrm{T}}\nabla_c v)(\nabla_x v^{\mathrm{T}}(\nabla_x f + \nabla_x t\lambda + \nabla_x v\rho - p) + \nabla_c v^{\mathrm{T}}(\nabla_c f + \nabla_c t\lambda + \nabla_c v\rho) + \gamma^2 V^2\rho))$$

If we are interested in splitting the coordination procedure into three levels, we need to look only at the following second-order expressions:

$$\nabla_z^2 T = 4\eta I$$
$$\nabla_p^2 T = 2\mu(4I + (\nabla_x t + H)(\nabla_x t + H)^{\mathrm{T}} + \nabla_x v\nabla_x v^{\mathrm{T}})$$
$$\nabla_\lambda^2 T = 2\mu((I + \nabla_x t^{\mathrm{T}}\nabla_x t + \nabla_c t^{\mathrm{T}}\nabla_c t)^2 + (\nabla_x t + H)^{\mathrm{T}}(\nabla_x t + H) + (\nabla_x t^{\mathrm{T}}\nabla_x v + \nabla_c t^{\mathrm{T}}\nabla_c v)(\nabla_x t^{\mathrm{T}}\nabla_x v + \nabla_c t^{\mathrm{T}}\nabla_c v)^{\mathrm{T}})$$

each of which is positive definite (as a sum of unit matrices and positive semidefinite matrices) for any positive values of η and μ. By letting the coordinator operate separately on z, p and λ variables in a three-level structure, we are assured that by satisfying the first-order necessary conditions

$$\nabla_z T = 0; \qquad \nabla_p T = 0; \qquad \nabla_\lambda T = 0$$

at each level, the function T is minimized with respect to the corresponding variable. Summarizing, we get the following algorithm.

Algorithm 2

Step 1. Choose η, μ, γ and a starting point in the extended space of primal and dual variables; label it by 0 and set $k = 0$.

Step 2. Solve the N first (low)-level subproblems for fixed z^k, p^k, λ^k:

$$\min_{x_i, c_i, \rho_i} T^i(x_i, c_i, z^k, p^k, \lambda^k, \rho_i), \quad i = 1, \ldots, N$$

and denote by x_i^{k+1}, c_i^{k+1}, ρ_i^{k+1} the minimizing values.

Step 3. Solve the quadratic problem:

$$\min_p T(x^{k+1}, c^{k+1}, z^k, p, \lambda^k, \rho^{k+1})$$

(that is, solve for p the $\nabla_p T = 0$ condition with fixed x^{k+1}, c^{k+1}, ρ^{k+1} and z^k, λ^k) and denote by p^{k+1} the minimizer.

Step 4. Solve the quadratic problem:

$$\min_\lambda T(x^{k+1}, c^{k+1}, z^k, p^{k+1}, \lambda, \rho^{k+1})$$

(that is, solve for λ the $\nabla_\lambda T = 0$ condition with fixed x^{k+1}, c^{k+1}, ρ^{k+1}, p^{k+1} and z^k) and denote by λ^{k+1} the minimizer.

Step 5. Solve the quadratic problem:

$$\min_z T(x^{k+1}, c^{k+1}, z, p^{k+1}, \lambda^{k+1}, \rho^{k+1})$$

that gives explicitly

$$z^{k+1} = \frac{1}{4\eta}(\lambda^{k+1} - H^{\mathrm{T}} p^{k+1}) + \frac{1}{2}(t(x^{k+1}, c^{k+1}) + H^{\mathrm{T}} x^{k+1}) \tag{11}$$

Step 6. If some stopping criterion is satisfied, then set

$$x^* = x^{k+1}; \qquad c^* = c^{k+1}; \qquad z^* = z^{k+1};$$
$$p^* = p^{k+1}; \qquad \lambda^* = \lambda^{k+1}; \qquad \rho^* = \rho^{k+1}$$

and stop; otherwise, set $k = k + 1$ and go to Step 2.

Note that in Step 5, the minimization of T can be carried out analytically giving (11); this expression is a function of terms belonging to different subsystems, so that no elimination of the variable z is possible (as done with the slack variables) without loss of separability. However, we can use (11) during the coordination task, which is not in a decomposed form, thus reducing the number of levels of the coordinator to two. In fact, introducing (11) into the expression for T gives a function $T' = T'(x, c, p, \lambda, \rho; \eta, \mu, \gamma)$; differentiating twice T' with respect to p and λ, we obtain:

$$\nabla_p^2 T' = \nabla_p^2 T - \frac{1}{4\eta} I = \left(8\mu - \frac{1}{4\eta}\right) I + 2\mu((\nabla_x t + H)(\nabla_x t + H)^{\mathrm{T}} + \nabla_x v \nabla_x v^{\mathrm{T}})$$

$$\nabla_\lambda^2 T' = \nabla_\lambda^2 T - \frac{1}{4\eta} I = \left(2\mu - \frac{1}{4\eta}\right) I$$
$$+ 2\mu((\nabla_x t \nabla_x t^{\mathrm{T}} + \nabla_c t \nabla_c t^{\mathrm{T}})^2 + (\nabla_x t + H)^{\mathrm{T}}(\nabla_x t + H)$$
$$+ 2(\nabla_x t \nabla_x t^{\mathrm{T}} + \nabla_c t \nabla_c t^{\mathrm{T}})$$
$$+ (\nabla_x t^{\mathrm{T}} \nabla_x v + \nabla_c t^{\mathrm{T}} \nabla_c v)(\nabla_x t^{\mathrm{T}} \nabla_x v + \nabla_c t^{\mathrm{T}} \nabla_c v)^{\mathrm{T}})$$

Both hessian matrices are positive definite if the couple of penalty coefficients (μ, η) is chosen so that $\eta > 0{\cdot}125/\mu > 0$. We can then solve the first-order necessary conditions

$$\nabla_p T' = 0; \qquad \nabla_\lambda T' = 0$$

in a two-level coordinator structure, where again we are assured that each level minimizes the function T' with respect to the corresponding variable. The procedure outlined leads to the following algorithm.

Algorithm 3

Steps 1–2. As in Algorithm 2.

Step 3. Use (11) to eliminate the variable z from the coordination problem.

Step 4. Solve the quadratic problem:

$$\min_p T''(x^{k+1}, c^{k+1}, p, \lambda^k, \rho^{k+1})$$

and denote by p^{k+1} the minimizer.

Step 5. Solve the quadratic problem:

$$\min_{\lambda} T'(x^{k+1}, c^{k+1}, p^{k+1}, \lambda, \rho^{k+1})$$

and denote by λ^{k+1} the minimizer.

Step 6. Use (11) to update the variable z, obtaining z^{k+1}.

Step 7. If some stopping criterion is satisfied, then set

$$x^* = x^{k+1}; \qquad c^* = c^{k+1}; \qquad z^* = z^{k+1};$$
$$p^* = p^{k+1}; \qquad \lambda^* = \lambda^{k+1}; \qquad \rho^* = \rho^{k+1}$$

and stop; else set $k = k + 1$ and go to Step 2.

Notice that for Algorithm 3, the coordination procedure complexity is exactly the same one present in an augmented lagrangian approach (Findeisen *et al.* 1980). In both Algorithms 2 and 3 the order in which variables are assigned to levels can be interchanged. This possibility and its consequences on speed of convergence will be discussed in a general formulation in the next section.

4. Convergence analysis

We provide here a discussion of convergence for the proposed method. The basic idea is to recognize that the multilevel algorithm for minimizing S or T is nothing but a block relaxation medthod (BRM) (see, for example, Ortega and Rheinboldt 1970). To get a deeper understanding of this analogy the following points should be noted:

(*a*) each level operates a vector minimization which corresponds to a step over a block of variables in the BRM;

(*b*) optimization is carried out iterating sequentially from the first to the last level as in the cyclic exploration of all blocks of variables in the BRM; no inner loops are inserted in the procedure as it is instead done in many multilevel schemes (Mahmoud 1977, Singh and Titli 1978);

(*c*) new level updates are utilized in computations as soon as available, that is, in the optimization at the next level; the algorithm thus operates in a typical Gauss–Seidel mode.

The block relaxation approach for minimizing a function $T(s)$, $s \in \mathbb{R}^n$, needs first a factorization of $\mathbb{R}^n$ as

$$\mathbb{R}^n = \prod_{i=1}^{q} \mathbb{R}^{n_i}; \qquad \sum_{i=1}^{q} n_i = n$$

so that $s^T = (s_1^T \quad s_2^T \quad \dots \quad s_q^T)$ with $s_i \in \mathbb{R}^{n_i}$; $q \geqslant 2$ is in our case the number of levels. The main steps of the BRM are then as follows.

Step 1. Set $h = 0$; choose a starting point $s^{0T} = (s_1^{0T} \quad s_2^{0T} \quad \dots \quad s_q^{0T})$.

Step 2. Set $i = 1$.

Step 3. For given $s_{(1)}^{hT} = (s_1^{hT} \quad s_2^{hT} \quad \dots \quad s_q^{hT})$ solve

$$\min_{s_1} T(s_1 \quad s_2^h \quad \dots \quad s_q^h)$$

and denote by s_1^{h+1} the minimizer.

Step 4. Set $i = 2$.

Step 5. For given $s_{(i)}^{h^T} = (s_1^{h+1^T} \quad \dots \quad s_{i-1}^{h+1^T} \quad s_i^{h^T} \quad \dots \quad s_q^{h^T})$ solve

$$\min_{s_i} T(s_1^{h+1} \quad \dots \quad s_{i-1}^{h+1} \quad s_i \quad s_{i+1}^{h} \quad \dots \quad s_q^{h})$$

and denote by s_i^{h+1} the minimizer.

Step 6. If $i < q$ then set $i = i + 1$ and go to Step 5; else continue.

Step 7. If some stopping criterion is satisfied then stop; else continue.

Step 8. Set $h = h + 1$ and go to Step 2.

When $n_i = 1$ for all i, the BRM becomes a point relaxation method and the ith step of the cycle is just a one-dimensional line search; the well-known univariate method or (cyclic) coordinate descent method operates on this basis. A sufficient convergence condition is established by the theorem of Bazaraa and Shetty (1979) which holds for a general class of algorithms that minimize a function searching along independent directions, and which is recalled here for convenience.

Theorem 1

Let $T: \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and consider an algorithm whose map A gives a vector $\hat{s} \in A(s)$ by minimizing T along the unitary search directions $d_1 \dots d_n$, starting from s. Suppose that:

(*a*) $\exists\, \varepsilon > 0$ such that $|D(s)| \geqslant \varepsilon$ for all $s \in \mathbb{R}^n$, where $D(s)$ is the $n \times n$ matrix whose columns are the search directions d_i (eventually dependent on s);

(*b*) the minimum of T along any line in $\mathbb{R}^n$ is unique.

Then if $s^{h+1} \in A(s^h)$ and the sequence $\{s^h\}$ is contained in a compact subset of $\mathbb{R}^n$, each accumulation point s of $\{s^h\}$ satisfies $\nabla T(s) = 0$.

Since in a multilevel procedure, different variables are treated at different levels, the only thing the algorithms of the previous sections must take care of is that of generating independent search directions locally to each level, since we have:

$$D(s) = \operatorname{diag} \{D_i(s_i), \quad i = 1, \dots, q\}, \quad D_i: n_i \times n_i$$

Furthermore, the first level has this task decomposed in N independent subproblems, so that the particular local structure can be exploited to provide the proper set of search directions (that is, Newton, quasi-Newton, conjugate directions, coordinate directions, etc.).

Notice that hypothesis (*b*) of the above theorem is particularly strong. On the other hand, it has been shown by Powell (1973), referring to the $n_i = 1$ case and for the univariate method ($D(s) = I$ so that hypothesis (*a*) is automatically satisfied), that differentiability alone is not sufficient to avoid failure of the method, cycling along a path with non-zero gradient. Moreover, after initial progress, the univariate method usually tends to slow down in later iterations, especially along valleys stretched other than in the coordinate directions. Several safeguards have been proposed in order to improve the convergence properties of the univariate method; in particular new search directions may be introduced as the minimization goes further. Rosenbrock's method (Bazaraa and Shetty 1979) progressively rotates a starting set of orthogonal search

directions so that the new directions are still mutually orthogonal and fit better with the objective function valleys. Another popular method is that of pattern search performed along directions individuated by some pattern of preceeding iterates, usually the last two as in the Hooke–Jeeves algorithm. All the above discussion can be easily extended to the general case ($n_i > 1$) of block relaxation and hence to multilevel optimization methods.

These similarities between multilevel computation and relaxation methods have been already pointed out by some authors (Looze and Sandell 1981, Xinogalas *et al.* 1983) and follow directly from the theoretical framework of Cohen (1978, 1980). However, with the present approach we are able to extend in a natural way two more features from numerical analysis to multilevel optimization. First, we observe that:

(*a*) Multilevel methods based on lagrangians and augmented lagrangians search for a saddle point and usually operate by treating primal and dual variables separately at different levels. Each level then tries to satisfy a subset of first-order necessary conditions for a saddle point. No explicit search directions are computed at the coordination levels, although some updating formulae bear an implicit extremization, as is typically the case for the multiplier rule of Hestenes–Powell, which takes a fixed step in the steepest ascent direction of a dual function. In general, one has to guarantee that each level computation or update moves toward the optimum, and this check should be included in all methods whose coordinator directly satisfies necessary conditions of optimality.

(*b*) Since the main goal of any multilevel method is to keep the problem decomposed among levels while achieving overall optimality, there is no advantage in extending non-linear programming techniques such as the Rosenbrock method. This would result in a recombination of variables of distinct levels, with an increase in the dimensionality of the subproblems. Moreover, methods based on ordinary or augmented lagrangians treat at different levels variables of the primal and of the dual type whose recombination makes no sense.

If we use instead the exact augmented lagrangian function for solving a large-scale optimization problem in the form of Problems 1 or 2, by means of a multilevel scheme then:

(*a*) the coordination level(s) minimize the exact augmented lagrangian function with respect to the subset of global variables, for every value of the local variables as shown in the previous sections. This enforces stability to the whole multilevel process;

(*b*) it is possible to improve the basic Algorithms 1, 2 and 3 by introducing an additional minimization step across the levels (similar to the one in the Hooke–Jeeves method), thus reducing the sources of inefficiency which otherwise affect all multilevel iterative schemes.

This additional step is just a unidimensional minimization along a direction determined by the last two iterates of the algorithm. In this step, the exact augmented lagrangian function is used as a line search function at little additional computational expense. Introducing this modification for instance in Algorithm 2 we get the following.

Algorithm 4

Steps 1–5. As in Algorithm 2.

Step 6. Set $s^k = ((x^{k+1})^T(c^{k+1})^T(z^{k+1})^T(p^{k+1})^T(\lambda^{k+1})^T(\rho^{k+1})^T)^T$ and

$$d^k = ((x^{k+1} - x^k)^T(c^{k+1} - c^k)^T(z^{k+1} - z^k)^T(p^{k+1} - p^k)^T (\lambda^{k+1} - \lambda^k)^T(\rho^{k+1} - \rho^k)^T)^T$$

and solve

$$T(s^k + \alpha^k d^k) = \min_{\alpha \in \mathbb{R}} T(s^k + \alpha d^k)$$

denoting by $s^{k+1} = s^k + \alpha^k d^k$ the minimizer.

Step 7. Update

$$((x^{k+1})^T(c^{k+1})^T(z^{k+1})^T(p^{k+1})^T(\lambda^{k+1})^T(\rho^{k+1})^T)^T = s^{k+1}$$

Step 8. If some suitable stopping criterion is satisfied then stop; else continue.

Step 9. Set $k = k + 1$ and go to Step 2.

This very simple modification has been tested with good performance in the examples reported in the next section. In particular the direction d^k so obtained gives better results than a steepest descent additional line search, thus confirming the validity of the analogy between coordinate descent methods and multilevel optimization.

We conclude this section with some further remarks on the proposed method, based on the analogies with relaxation methods:

(*a*) It is generally recognized that when the coordinator predictions are far from the solution, it is computationally wasteful to achieve high precision in resolving the local subproblems. This rule of thumb can be automatically included in the proposed algorithms by an appropriate choice of the stopping criterion for the first-level subproblems. A certain kind of inexact minimization, which is asymptotically stringent, has actually been introduced in some implementations reported in the next section.

(*b*) The order of assignment of variables to levels should play no significant role on convergence properties, at least from a theoretical point of view, since it parallels the order in which variables are explored by relaxation methods; this was confirmed by the numerical computations.

5. Numerical examples

The proposed multilevel algorithms have been tested on two examples taken from the literature. The first one, referenced in Stephanopoulos and Westerberg (1975) and in Brusilovski and Ostrovski (1983), is a linearly constrained problem in which a dual gap arises. The second has convex objective functions but non-linear inequality constraints and is referenced in Findeisen *et al.* (1980), where it is used as a benchmark problem for several multilevel methods. Among these methods, that based on augmented lagrangians showed the best performances.

Example 1

Consider the following two single-input/single-output units connected in cascade ($x_2 = z_1$):

$$(\Sigma_1)\begin{cases} z_1 = 3x_1 + 3c_1 \\ v_1(x_1, c_1)^T = (-x_1 \quad -c_1 \quad x_1 - 3 \quad x_1 + 2c_1 - 4)^T \leqslant 0 \end{cases}$$

$$(\Sigma_2)\begin{cases} z_2 = 2x_2 + 2c_2 \\ v_2(x_2, c_2)^T = (-x_2 \quad -c_2 \quad c_2 - 1 \quad x_2 + 2c_2 - 4)^T \leqslant 0 \end{cases}$$

The cost function to be minimized is

$$\hat{f}(x, c, z) = 2c_1 + x_1^{0.6} + 5c_2 + x_2^{0.6} - z_2$$

which can be rewritten as

$$f(x, c) = (2c_1 + x_1^{0.6}) + (3c_2 + x_2^{0.6} - 2x_2)$$

The optimal analytic solution is

$$x_1^* = 4/3; \qquad c_1^* = 0; \qquad x_2^* = z_1^* = 4; \qquad c_2^* = 0; \qquad p^* = \lambda^* = -0{\cdot}178$$

giving the value $f^* = -4{\cdot}52$ for the objective function. Table 1 shows the results obtained using Algorithm 4 for some selected values of the penalties η, μ, with γ^2 held fixed at 4. K is the iteration number at which two successive coordinator predictions differ in norm for less than $\varepsilon = 10^{-5}$; the achieved feasibility is given in terms of the norm of the interaction error $e = x_2^k - t_1(x_1^k, c_1^k)$. The starting point was chosen as in Brusilovski and Ostrovski (1983) for the local variables:

$$x_1^0 = 1{\cdot}03; \qquad c_1^0 = -0{\cdot}01; \qquad x_2^0 = 3{\cdot}06; \qquad c_2^0 = 0{\cdot}35; \qquad \rho^0 \text{ all zero}$$

while the coordinator initial predictions were

$$z_1^0 = 5{\cdot}0; \qquad \lambda^0 = -0{\cdot}2; \qquad p^0 = -0{\cdot}4$$

η	μ	K	f	$\|e\|_2$
10^2	1	57	−4·52	$0{\cdot}2 \times 10^{-3}$
10^3	1	35	−4·53	$0{\cdot}5 \times 10^{-3}$
10^3	10	28	−4·52	$0{\cdot}6 \times 10^{-3}$

Table 1.

Example 2

Given a plant consisting of three interconnected dynamic systems described by their steady-state models as follows:

$$(\Sigma_1)\begin{cases} z_{11} = c_{11} - c_{12} + 2x_{11} \\ v_1(x_1, c_1)^T = (c_{11}^2 + c_{12}^2 - 1 \quad -x_{11} \quad x_{11} - 0{\cdot}5) \leqslant 0 \\ f_1(x_1, c_1) = (x_{11} - 1)^4 + 5(c_{11} + c_{12} - 2)^2 \end{cases}$$

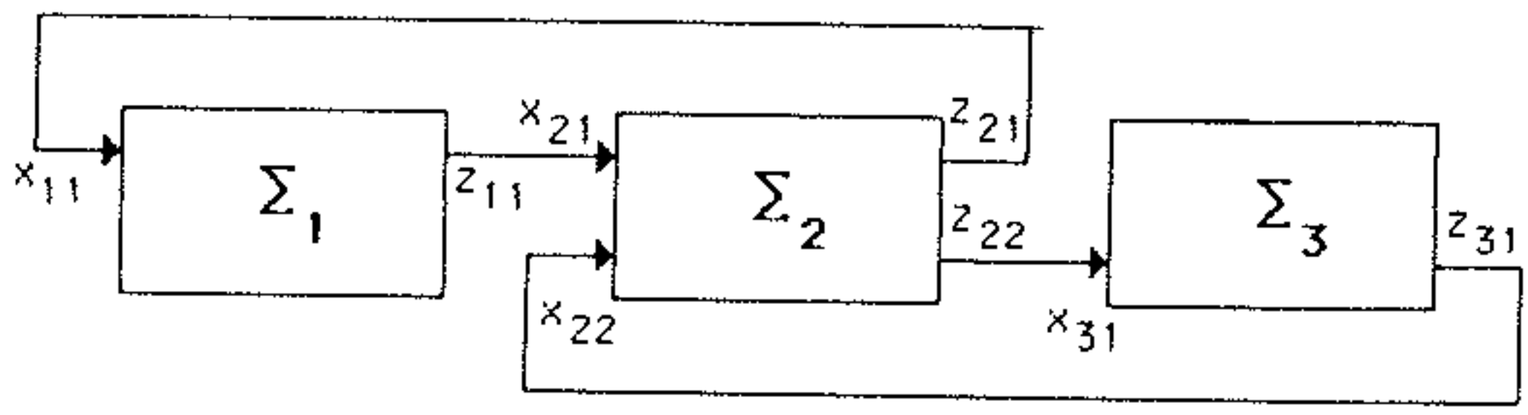
$$(\Sigma_2)\begin{cases} z_{21} = c_{21} - c_{22} + x_{21} - 3x_{22} \\ z_{22} = 2c_{22} - c_{23} - x_{21} + x_{22} \\ v_2(x_2, c_2) = \begin{bmatrix} 0{\cdot}5c_{21} + c_{22} + 2c_{23} - 1 \\ 4c_{21}^2 + 2c_{21}x_{21} + 0{\cdot}4x_{21} + c_{21}c_{23} + 0{\cdot}5c_{23}^2 + x_{21}^2 - 4 \end{bmatrix} \leqslant 0 \\ f_2(x_2, c_2) = 2(c_{21} - 2)^2 + c_{22}^2 + 3c_{23}^2 + 4x_{21}^2 + x_{22}^2 \end{cases}$$

$$(\Sigma_3)\begin{cases} z_{31} = c_{31} + 2{\cdot}5c_{32} - 4x_{31} \\ v_3(x_3, c_3)^{\mathrm{T}} = (-c_{31} - x_{31} - 0{\cdot}5 \quad -c_{32} \quad c_{32} - 1) \leqslant 0 \\ f_3(x_3, c_3) = (c_{31} + 1)^2 + (x_{31} - 1)^2 + 2{\cdot}5c_{32}^2 \end{cases}$$

the design problem is to minimize

$$f(x, c) = \sum_{i=1}^{3} f_i(x_i, c_i)$$

subject to the interconnection constraints ($x = Hz$) shown in the Figure. Notice that the resulting interconnection matrix H is orthonormal.



Starting from the origin of the extended space of primal and dual variables, as in Findeisen *et al.* (1980), several runs were performed with different penalties η, μ and constant $\gamma^2 = 4$ also to compare the relative merits of the various algorithms (Table 2). As before, K is the number of iterations needed so that two successive coordinator predictions differ in norm by less than $\varepsilon = 10^{-5}$. The feasibility of the final point is given in terms of the euclidean norm of the vector

$$e = \begin{bmatrix} x_{11}^k - t_{21}(x_2^k, c_2^k) \\ x_{21}^k - t_{11}(x_1^k, c_1^k) \\ x_{22}^k - t_{31}(x_3^k, c_3^k) \\ x_{31}^k - t_{22}(x_2^k, c_2^k) \end{bmatrix}$$

η	μ	K	f	$\|e\|_2$	Algorithm
10^2	0·1	66	6·1370	$0{\cdot}9 \times 10^{-2}$	4
10^3	0·01	152	6·1182	$0{\cdot}4 \times 10^{-4}$	4
10^3	0·01	242	6·1258	$0{\cdot}6 \times 10^{-3}$	2
5×10^2	0·05	123	6·1270	$0{\cdot}4 \times 10^{-4}$	4

Table 2.

which measures the error in the satisfaction of the interconnection constraints, when the current input–output model is used. The effectiveness of the proposed acceleration step (Algorithm 4) with respect to the basic Algorithm 2 is self-evident. No particular benefits were obtained reducing the number of coordination levels by making use of Algorithm 3. Furthermore, choosing different processing orders for the coordinator variables (namely, with (z, p, λ) or (p, λ, z) treated respectively at the second, third and fourth level) showed different paths from the starting point to the solution, but no significant differences in the terminal figures (number of iterations and feasibility). Subproblem minimization was carried out by a quasi-Newton method with BFS updates; the same was done for the upper level's minimizations, whenever these were not performed analytically.

Finally it is useful to point out some aspect relevant in numerical applications of the method:

(*a*) Although any choice for $\mu > 0$ and for $\eta > \eta^*$ allows us to solve the problem by means of an unconstrained minimization of the exact augmented lagrangian function, there is, for each problem, a preferential range for the penalty parameters η and μ. More specifically the product $\mu\eta$ should lie within a certain interval outside of which we have either convergence which is too slow or divergence of the sequence of generated points.

(*b*) The starting coordinator predictions clearly play an important role for convergence. Note that since the method considered is a non-feasible one, it generates a sequence which usually lies beyond the feasible region, so that problems could arise with functions not defined everywhere, as is the case in Example 1 (the fractional power for x implies non-negativity as a hard constraint). In fact, we needed a proper choice for the coordinator starting point, although far from the true solution.

6. Conclusion

We have developed a new multilevel method for the solution of large-scale structured optimization problems by means of the exact augmented lagrangian approach studied in Di Pillo and Grippo (1979, 1982) and Lucidi (1985). The decomposition of this function is obtained by fixing, in the general case, the subsystems outputs and the equality constraints multipliers. The resulting co-ordinator procedure is organized as a three-level minimization, one of which can be analytically solved, while the other two are positive definite quadratic problems. The efficiency of the coordination task is enhanced when no local inequality constraints are present, being then strictly independent of the number of subsystems. Numerical experience with this method was satisfactory giving results which are competitive with those quoted in the literature. One of the limitations of the present approach is its intrinsic non-feasibility before obtaining the optimal solution, so that no on-line applications are possible. However, this approach gives some further insight into the mechanism of convergence of multilevel methods. The parallelism between some classical minimization methods and those based on decomposition–coordination had already been recognized but could not be stressed in a primal–dual framework. We could instead extend in a multilevel context some simple ideas from general non-linear programming such as the inter-levels acceleration step, whose effectiveness was confirmed by numerical examples.

References

Bazaraa, M. S., and Shetty, C. M., 1979, *Nonlinear Programming: Theory and Applications* (New York: Wiley).

Bertsekas, D. P., 1976, *Automatica*, **12**, 133; 1979, *J. Optim. Theory Applic.*, **29**, 169.

Brusilovski, A. M., and Ostrovski, G. M., 1983, *Int. J. Systems Sci.*, **14**, 1273.

Cohen, G., 1978, *I.E.E.E. Trans. autom. Control*, **23**, 222; 1980, *J. optim. Theory Applic.*, **32**, 277.

Dantzig, G. B., and Wolfe, P., 1961, *Econometrica*, **29**, 767.

Di Pillo, G., and Grippo, L., 1979, *SIAM J. Control Optim.*, **17**, 618; 1982, *J. optim. Theory Applic.*, **36**, 495.

Findeisen W., Bailey, F. N., Brdys, M., Malinowski, K., Tatjewski, P., and Wozniak, A., 1980, *Control and Coordination in Hierarchical Systems* (Chichester: Wiley).

Haimes, Y. Y., (editor), 1982, *Large Scale Systems* (Amsterdam: North-Holland).

Lasdon, L. S., 1970, *Optimization Theory for Large Systems* (New York: Macmillan).

Looze, D. P., and Sandell Jr., N. R., 1981, *J. optim. Theory Applic.*, **34**, 371.

Lucidi, S., 1985, Tech. Rep. IASI-CNR, 116.

Mahmoud, M. S., 1977, *I.E.E.E. Trans. Syst. Man Cyber.*, **7**, 125.

Mesarovic, M. D., Macko, D., and Takahara, Y., 1970, *Theory of Hierarchical Multilevel Systems* (New York: Academic Press).

Ortega, J. M., and Rheinboldt, W. C., 1970, *Iterative Solution of Nonlinear Equations in Several Variables* (New York: Academic Press).

Powell, M. J. D., 1973, *Math. Prog.*, **4**, 193.

Schoeffler, J. D., 1971, In *Optimization Methods for Large-Scale Systems with Applications*, edited by D. A. Wismer (New York: McGraw-Hill).

Singh, M. G., and Titli, A., 1978, *Systems: Decomposition, Optimization and Control* (Oxford: Pergamon).

Stephanopoulos, G., and Westerberg, A. W., 1975, *J. optim. Theory Applic.*, **15**, 285.

Stoilov, E., 1977, *Arch. Autom. Telemech.*, **22**, 210.

Takahara, Y., 1965, A multilevel structure for a class of dynamic optimization problems. M.S. thesis, Case Western Reserve University, Cleveland, U.S.A.

Tatjewski, P., 1978, *RAIRO Automatique*, **12**, 221.

Tatjewski, P., and Michalak, P., 1980, *Systems Sci.*, **6**, 23.

Titli, A., 1972, Contribution a l'étude des structures de commande hierarchisées en vue de l'optimisation de processus complexes. Ph.D. dissertation, Université Paul Sabatier, Toulouse, France.

Watanabe, N., Nishimura, Y., and Matsubara, M., 1978, *J. optim. Theory Applic.*, **25**, 181.

Wismer, D. A. (editor), 1971, *Optimization Methods for Large-Scale Systems with Applications* (New York: McGraw-Hill).

Xinogalas, T. C., Dasigi, S., and Singh, M. G., 1983, *I.E.E.E. Trans. Syst. Man Cyber.*, **13**, 397.