AN EXACT AUGMENTED LAGRANGIAN APPROACH TO MULTILEVEL OPTIMIZATION

A. De Luca and G. Di Pillo

Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Via Eudossiana 18, 00184 Roma, Italy

Abstract. An approach based on the exact augmented Lagrangian function is developed for the optimization of large scale systems constituted by 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 a parallel with minimization by relaxation methods. A numerical example is included.

Keywords. Large scale systems, Optimization, Nonlinear programming, Augmented Lagrangian methods, Decomposition, Coordination methods.

INTRODUCTION

A common characteristic of 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 i.e. a certain number of interconnected subsystems may be identified.

Motivated by the presence of this structure, specific decomposition and coordination methods have been developed for the optimization purpose. In the mathematical programming framework the first multilevel methods were based on the Lagrangian approach (Lasdon, 1970), since the Lagrangian function retains the separability of the problem formulation.

Duality gaps limit the effectiveness of these methods in absence of convexity assumptions so that the application of augmented Lagrangians theory to multilevel problems has been considered; due to the added penalty term this approach has the drawback of destroying separability.

Various solutions have been proposed: Stephanopoulos and Westerberg (1975) introduced a linear approximation of the nonseparable crossterms while Watanabe et al. (1978) transformed these terms into the minimum of a sum of separable terms.

The class of methods proposed by Findeisen et al. (1980) are based on a suitable augmentation of the Lagrangian function without approximating terms and using mixed multiplier/interaction variables prediction; the resulting coordinator task is formulated as a saddle-point problem. Finally, in Bertsekas (1979), an alternative augmentation of the Lagrangian function allows local convexification by simply duplicating the number of primal variables.

A clear classification and comparison of these approaches is quite diffucult due to the differences in basic decomposition schemes and applicability conditions, lack of convergence analysis or results and application-oriented nature of multilevel procedures; see however Cohen (1978, 1980).

As a further development we propose in this paper a method based on recent results in nonlinear programming: the exact augmented Lagrangian approach introduced by Di Pillo and Grippo (1979, 1982) and Lucidi (1985), which consists in 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:

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

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

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

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×m nonsingular matrix, where ∇g denotes the transpose of the Jacobian matrix of the constraints.

The extension of the proposed approach to nonlinear programming problems with inequality constraints:

$$\text{minimize } J(s)$$
$$\text{subject to } g(s) \leq 0, \; s \in \mathbb{R}^n$$

was studied by Di Pillo and Grippo (1982) and by Lucidi (1985) by using the device of converting inequalities to equalities via squared slack variables. In particular it was shown by 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

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

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

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 $T(s,\sigma)$.

Application of this approach to large scale nonlinear problems, followed by a suitable decomposition, allows:

- to solve the problem by a general multilevel method with an efficient coordination process without the need of convexity assumptions;
- to look at multilevel methods as a straightforward generalization of well-known minimization by relaxaticn methods;
- to derive simple convergence analysis and an acceleration procedure for the proposed algorithm.

LARGE-SCALE EQUALITY CONSTRAINED PROBLEMS

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

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

$$\text{s.t. } z_i = t_i(x_i,c_i)$$

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

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 i-th subsystem interaction inputs and outputs and the i-th 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 subsystems 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 a large number) interconnected static systems, minimizing the sum of the local (generally nonlinear) cost functions. We can rewrite Problem 1 in a more compact form as:

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

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

where $x \triangleq [x_1^T \ldots x_N^T]^T \in \mathbb{R}^n$, $n = \sum_{i=1}^{N} n_i$ and analogously, $z \in \mathbb{R}^k$, $c \in \mathbb{R}^m$; furthermore, $t \triangleq [t_1^T \ldots t_N^T]^T \in \mathbb{R}^k$ and $H \triangleq \{H_{ij}\}$, an $(n\times k)$ matrix.

Note that the subsystem linear interconnections imply no loss of generality since all nonlinearities can be included in the subsystem models. Moreover each and every output is connected to one and only one input of a different subsystem, thus the matrix H is an orthonormal matrix, i.e. $H^{-1} = H^T$, $k = n$. Local feedbacks are included in the subsystem models so that $H_{ii} = 0$, for all i.

We will admit that the following assumptions hold, where Ω is a given compact subset of $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$:

A1) the functions f,t are twice continuously differentiable with respect to the variables x and c on $\mathbb{R}^n \times \mathbb{R}^m$;

A2) 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:

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

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

An exact augmented Lagrangian function for this problem is

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

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 $2n\times 2n$ nonsingular matrix in Ω, the solutions of Problem 1 contained in Ω and the corresponding Lagrange multipliers coincide with 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.

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

PROPOSITION. *If the matrices* $\nabla_{c_i} t_i$, $i=1,\ldots,N$ *have full column rank,* i.e.

$$\text{rank}[\nabla_{c_i} t_i] = k_i \leq m_i$$

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

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

then the matrix $M\,\nabla g$ *is nonsingular.*

PROOF. See De Luca and Di Pillo (1984). □

The hypothesis of the above proposition means that every subsystem has to be regulated by a number of effective control inputs not less than the number of local interconnection outputs. 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:

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

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

1) For fixed values $\bar{z},\bar{p}$ minimize S w.r.t. x,c,λ; this problem splits into N independent low level ones:

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

where

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

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

Due to the quadratic dependence of S on z,p the coordination is a very simple problem. The gradients and the Hessian of S w.r.t. the coordinating variables are given by:

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

$$\nabla_p S = Hz - x + 2\mu(p - \nabla_x f(x,c) - \nabla_x t(x,c)\lambda)$$

$$\nabla^2_{(z,p)} S = \begin{bmatrix} 4\eta I_n & H^T \\ H & 2\mu I_n \end{bmatrix}. \quad (5)$$

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 ($\nabla_z S = 0$, $\nabla_p S = 0$) are also sufficient for a minimum w.r.t. z and p, for every value of the first level variables $\bar{x}$, $\bar{c}$ and $\bar{\lambda}$. Furthermore the inverse of the Hessian matrix (5) is analytically available. Thus the coordination task can be accomplished by a Newton iteration which gives new updates $\hat{z}$ and $\hat{p}$ in the closed form:

$$\begin{bmatrix} \hat{z} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} \bar{z} \\ \bar{p} \end{bmatrix} - \left[\nabla^2_{(\bar{z},\bar{p})} S \right]^{-1} \begin{bmatrix} \nabla_z S\ (\bar{x},\bar{c},\bar{z},\bar{\lambda},\bar{p}) \\ \nabla_p S\ (\bar{x},\bar{c},\bar{z},\bar{\lambda},\bar{p}) \end{bmatrix} \quad (6)$$

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

A basic two-level algorithm for solving Problem 1 by means of the proposed exact augmented Lagrangian function is then the following.

ALGORITHM 1

1. Choose η,μ and a starting point in the extended space of primal and dual variables; label it by 0 and set $k = 0$.
2. Solve the N low-level subproblems for fixed z^k, p^k:
$$\min_{x_i, c_i, \lambda_i} S^i(x_i, c_i, z^k, \lambda_i, p_i^k), \quad i=1,\dots,N$$
and denote by $x_i^{k+1}, c_i^{k+1}, \lambda_i^{k+1}$ the minimizing values.
3. Calculate z^{k+1}, p^{k+1} by (6) for $\bar{x} = x^{k+1}, \bar{c} = c^{k+1}$, $\bar{\lambda} = \lambda^{k+1}$.
4. If some stopping criterion is satisfied then set $x^* = x^{k+1}, c^* = c^{k+1}, z^* = z^{k+1}, p^* = p^{k+1}, \lambda^* = \lambda^{k+1}$ and stop; else set $k = k+1$ and go to 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 pratically no additional cost for the coordination procedure.

We point out that a direct substitution could be made of x in terms of z (or viceversa) using the interconnection equation explicitly (Findeisen et al., 1980); our choice, despite the increase of dimensionality of the local subproblems, is motivated by the appealing coordination process (6) that can be obtained thanks to this duplication of variables. Moreover no such direct substitution is possible when local inequality constraints are present, unless they were separable in terms of controls and interconnection inputs.

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:

$$v_i(x_i, c_i) \le 0, \quad i = 1,2,\dots,N$$

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

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

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

Problem 2 $\min f(x,c)$

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

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

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

with $\rho \in \mathbb{R}^r$ 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:

A3) the functions f,t,v are twice continuously differentiable with respect to the variables x and c on $\mathbb{R}^n \times \mathbb{R}^m$;

A4) the gradients of the active constraints are linearly independent at every point (x,c,z) in the compact subset Ω;

A5) define the index set $J^i(x,c)$ as
$$J^i(x,c) \triangleq \{j: v_{ij}(x_i, c_i) \ge 0\};$$
then the Mangasarian-Fromowitz regularity assumption holds on Ω, i.e. for any $(x,c,z) \in \Omega$
$$\sum_{k=1}^{2n} \alpha_k \nabla g_{1,k} + \sum_{i=1} \sum_{j \in J^i} \beta_j^i \nabla v_{ij} = 0,$$
with $\beta_j^i \ge 0$ for all $j \in J^i$, $i=1,\dots,N$, implies that $\beta_j^i = 0$ for all $j \in J^i$, $i=1,\dots,N$ and $\alpha_k = 0$ for $k = 1,\dots,2n$; here $g_{1,k}$ denotes the k-th component of g_1 and the gradients are taken w.r.t. the primal variables (x,c,z);

A6) strict complementarity holds at any Kuch-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:

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

where

$$y_{ij}^2(x_i, c_i, \rho_i) = -\min\{0, v_{ij}(x_i, c_i) + \frac{\rho_{ij}}{2\eta}\}, \quad j = 1,\dots,r_i,\ i = 1,\dots,N, \quad (8)$$

$Y(x,c,\rho) = \text{diag}\{y_{ij}(x_i,c_i,\rho_i);\ j=1,\dots,r_i, i=1,\dots,N\}$,

$V(x,c) = \text{diag}\{v_{ij}(x_i,c_i);\ j=1,\dots,r_i,\ i=1,\dots,N\}$.

∇L denotes the gradient of L w.r.t. (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:

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

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

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

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

Again as for Problem 1, the solutions of Problem 2 contained in Ω and the associated Kuhn-Tucker multipliers coincide with 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.

As a matter of fact, the addition of the local inequality constraints implies a relevant complication in the expression of T. 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

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

with $V_i \triangleq \text{diag}\{v_{ij};\ j=1,\dots,r_i\}$.

Our primary goal is to derive now a coordinator as simple as possible for minimizing T w.r.t. z,p and λ. Notice that the exact augmented Lagrangian function is quadratic with respect to all coordination variables z,p and λ. In principle one can solve directly this 3n-dimensional quadratic problem, 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 in fact a three-level coordination algorithm.

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

$$\nabla_z T = H^Tp-\lambda+2\eta(2z-t(x,c)-H^Tx)$$

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

$$\begin{aligned} \nabla_\lambda T &= t(x,c)-z \\ &+2\mu((I+\nabla_x t^T\nabla_x t+\nabla_c t^T\nabla_c t)(\nabla_x t^T\nabla_x L+\nabla_c t^T\nabla_c L+\lambda-H^Tp) \\ &-(\nabla_x t+H)^T(p-H\lambda-\nabla_x L) \\ &+(\nabla_x t^T\nabla_x v+\nabla_c t^T\nabla_c v)(\nabla_x v^T\nabla_x L+\nabla_c v^T\nabla_c L+\gamma^2V^2\rho)). \end{aligned}$$

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

$$\nabla_z^2 T = 4\eta I$$

$$\nabla_p^2 T = 2\mu(4I+(\nabla_x t+H)(\nabla_x t+H)^T+\nabla_x v\nabla_x v^T)$$

$$\begin{aligned} \nabla_\lambda^2 T &= 2\mu((I+\nabla_x t^T\nabla_x t+\nabla_c t^T\nabla_c t)^2 \\ &+(\nabla_x t+H)^T(\nabla_x t+H) \\ &+(\nabla_x t^T\nabla_x v+\nabla_c t^T\nabla_c v)(\nabla_x t^T\nabla_x v+\nabla_c t^T\nabla_c v)^T) \end{aligned}$$

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 $\nabla_z T = 0$, $\nabla_p T = 0$, $\nabla_\lambda T = 0$ at each level the function T is minimized w.r.t. the corresponding variable. Summarizing we get the following algorithm.

<u>ALGORITHM 2</u>

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

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

   and denote by $x_i^{k+1},c_i^{k+1},\rho_i^{k+1}$ the minimizing values.
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 $x^{k+1},c^{k+1},\rho^{k+1}$ and z^k,λ^k) and denote by p^{k+1} the minimizer.
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 $x^{k+1},c^{k+1},\rho^{k+1},p^{k+1}$ and z^k) and denote by λ^{k+1} the minimizer.

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

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

6. If some stopping criterion is satisfied then set

$x^* = x^{k+1}, c^* = c^{k+1}, z^* = z^{k+1}, p^* = p^{k+1}$,

$\lambda^* = \lambda^{k+1}, \rho^* = \rho^{k+1}$

and stop; else set $k = k+1$ and go to 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. Introducing (11) into the expression of T we get a function $T' = T'(x,c,p,\lambda,\rho;\eta,\mu,\gamma)$; the coordinator minimizes then T' w.r.t. p and λ and updates z^{k+1} via (11) before returning to the lowest level.

Notice that with this procedure the coordination complexity is exactly the same one present in an augmented Lagrangian approach (Findeisen et al., 1980). In any case the order in which variables are assigned to levels can be interchanged.

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 method (BRM) (see e.g. Ortega and Rheinboldt, 1970). To get a deeper understanding of this analogy one should note the following points:

- each level operates a vector minimization which corresponds to a step over a block of variables in the BRM;
- 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);
- new level updates are utilized in computations as soon as available, that is in the optimization at the next level: thus the algorithm operates in a typical Gauss-Seidel mode.

The block relaxation approach for minimizing a function $T(s)$, $s \in \mathbb{R}^n$, is a generalization of the well-known univariate method; the one-dimensional line search along a (coordinate) direction, which constitutes a step in this algorithm, is substituted by a minimization of the function on one of the subspaces in which the n-dimensional space has been partitioned. A sufficient convergence condition for this method is established by a 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. *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* $d_1 \ldots d_n$, *starting from s. Suppose that:*

a) $\exists \varepsilon > 0$ s.t. $|D(s)| \geq \varepsilon$ *for all* $s \in \mathbb{R}^n$, *where* D(s) *is the* n×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* $\mathbb{R}$, *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: $D(s) = \mathrm{diag}\{D_i(s_i),\ i = 1,\ldots,q\}$, $D_i: n_i \times n_i$, where $s = (s_1^T s_2^T \ldots s_q^T)^T$, $s_i \in \mathbb{R}^{n_i}$, and $q \geq 2$ is the number of levels. 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 (i.e. Newton, Quasi-Newton, conjugate directions, coordinate directions, etc.).

Notice that hypothesis b) of the above theorem is particularly strong, but it has been shown that differentiability alone may not be sufficient for the method to succeed (Powell, 1973). Moreover, usually the univariate method, after initial progrees, tends to slow down in later iterations especially along valleys stretched not in the coordinate directions. In order to improve the convergence properties new search directions can be introduced as the minimization goes further. A 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; this can be easily extended to the case of block relaxation.

The similarities between multilevel computation and relaxation methods have already been 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 notice that multilevel methods based on Lagrangian and augmented Lagrangians have, in this perspective, two major drawbacks: first, they search for a saddle point, each level trying to satisfy a subset of first order necessary conditions; in general, no explicit search directions are computed at the coordination levels. Hence, one has to check that each level computation or update moves toward the optimum. Second, recombination of variables of distinct levels makes no sense in methods based on ordinary or augmented Lagrangians which treat at different levels variables of the primal and of the dual type.

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

- 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;
- it is possible to improve the basic Algorithm 1 and 2 introducing an additional minimization step

across the level (like the one in the Hooke-Jeeves method), thus reducing the sources of inefficiency which otherwise affect all multi-level 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 line search function at little additional computational expense. Introducing this modification e.g. in Algorithm 2 we get:

ALGORITHM 3

1. - 5. As in Algorithm 2.

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.

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

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

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

This very simple modification has been tested with good performance in the example 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.

NUMERICAL EXAMPLE

The proposed multilevel algorithms have been tested on an example, which has convex objective functions and nonlinear inequality constraints, referenced in Findeisen et al. (1980) where it is used as a benchmark problem for several multilevel methods. For another numerical example see De Luca and Di Pillo (1984). Given a plant consisting of three interconnected dynamic systems described by their steady-state models as follows:

$$z_{11} = c_{11} - c_{12} + 2x_{11}$$

$$(\Sigma_1)\quad v_1(x_1,c_1)^T = (c_{11}^2+c_{12}^2-1\;\; -x_{11}\;\; x_{11}-.5) \le 0$$

$$f_1(x_1,c_1) = (x_{11}-1)^4 + 5(c_{11}+c_{12}-2)^2$$

$$z_{21} = c_{21}-c_{22}+x_{21}-3x_{22}$$

$$z_{22} = 2c_{22}-c_{23}-x_{21}+x_{22}$$

$$(\Sigma_2)\quad v_2(x_2,c_2) = \begin{bmatrix} .5c_{21}+c_{22}+2c_{23}-1 \\ 4c_{21}^2+2c_{21}x_{21}+.4x_{21}+c_{21}c_{23}+.5c_{23}^2+x_{21}^2-4 \end{bmatrix} \le 0$$

$$f_2(x_2,c_2) = 2(c_{21}-2)^2+c_{22}^2+3c_{23}^2+4x_{21}^2+x_{22}^2$$

$$z_{31} = c_{31}+2.5c_{32}-4x_{31}$$

$$(\Sigma_3)\quad v_3(x_3,c_3)^T = (-c_{31}-x_{31}-.5\;\; -c_{32}\;\; c_{32}-1) \le 0$$

$$f_3(x_3,c_3) = (c_{31}+1)^2+(x_{31}-1)^2+2.5c_{32}^2,$$

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 Fig. 1.

Starting from the origin of the extendend 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 schemes (Table I). K is the number of iterations needed in order that two successive coordinator predictions differ in norm by less than $\varepsilon = 10^{-5}$. The feasibility of the

η	μ	K	f	$\|e\|_2$	Algorithm
10^2	.1	66	6.1370	$.9\cdot10^{-2}$	3
10^3	.01	152	6.1182	$.4\cdot10^{-4}$	3
10^3	.01	242	6.1258	$.6\cdot10^{-3}$	2
$5\cdot10^2$	.05	123	6.1270	$.4\cdot10^{-4}$	3

Table I

final point is given in terms of the Euclidean norm of the vector $e = H^Tx^k - t(x^k,c^k)$ 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 3) with respect to the basic Algorithm 2 is self-evident. No particular benefits were obtained reducing the number of coordination levels from three to two as explained in the previous section. 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 levels minimizations, whenever these were not performed analytically. Finally, note that since the method considered is a nonfeasible one, it generates a sequence which usually lies beyond the feasible region, so that troubles could arise with functions not defined everywhere.

CONCLUSION

We 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 coordinator 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 from the number of subsystems. Numerical experience with this method was satisfaxtory giving results which are competitive with the ones quoted in the literature. One of the limitations of the present approach is its intrinsic nonfeasibility before obtaining the optimal solution, so that no on-line applications are possible. However this approach gives some further insight in 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 nonlinear programming such as the intra-level acceleration step, whose effectiveness was confirmed by numerical examples.

REFERENCES

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

Bertsekas, D.P. (1979). Convexification procedures and decomposition methods for nonconvex optimization problems. J. Optimization Theory & Appl., 29, 169-197.

Cohen, G. (1978). Optimization by decomposition and coordination: a unified approach. IEEE Trans. Autom. Control, 23, 222-232.

Cohen, G. (1980). Auxiliary problem principle and decomposition of optimization problems. J. Optimization Theory & Appl., 32, 277-305.

De Luca, A., and G. Di Pillo (1984). An exact augmented Lagrangian approach to multilevel optimization of large scale systems. Università di Roma "La Sapienza", DIS Rep. 13.84.

Di Pillo, G., and L. Grippo (1979). A new class of augmented Lagrangians in nonlinear programming. SIAM J. Control Optim., 17, 618-628.

Di Pillo, G., and L. Grippo (1982). A new augmented Lagrangian function for inequality constraints in nonlinear programming problems. J. Optimization Theory & Appl., 36, 495-519.

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

Lasdon, L.S. (1970). Optimization Theory for Large Sysyems. MacMillan, New York.

Looze, D.P., and N.R. Sandell Jr. (1981). Analysis of decomposition algorithms via nonlinear splitting functions. J. Optimization Theory & Appl. 34, 371-382.

Lucidi, S. (1985). New results on a class of exact augmented Lagrangians. Tech. Rep. IASI-CNR, 116.

Mahmoud, M.S. (1977). Multilevel systems control and applications: a survey. IEEE Trans. Syst., Man, Cybern., 7, 125-143.

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

Powell, M.J.D. (1973). On search direction for minimization algorithms. Math. Programming, 4, 193-201.

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

Stephanopoulos, G., and A.W. Westerberg (1975). The use of Hestenes' method of multipliers to resolve dual gaps in engineering system optimization. J. Optimization Theory & Appl., 15, 285-309.

Watanabe, N., Y. Nishimura, and M. Matsubara (1978). Decomposition in large system optimization using the method of multipliers. J. Optimization Theory & Appl., 25, 181-193.

Xinogalas, T.C., S. Dasigi, and M.G. Singh (1983). Coordination in hierarchical algorithms. IEEE Trans. Syst., Man, Cybernet., 13, 397-406.

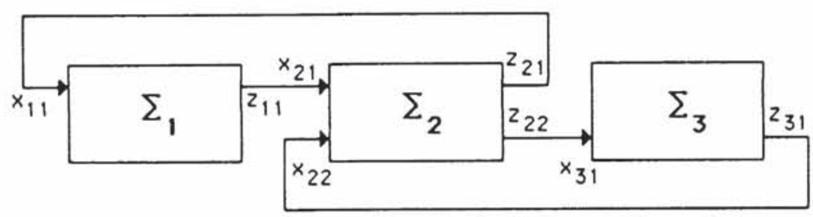

Fig. 1: Subsystem interconnections of the numerical example.