Convergence analysis of a multi-level relaxation method

Hervé Guillard

UNITÉ DE RECHERCHE
INRIA-SOPHIA ANTIPOLIS

Rapports de Recherche

N°1884

Programme 6

*Calcul scientifique, Modélisation
et Logiciel numérique*

CONVERGENCE ANALYSIS OF A MULTI-LEVEL RELAXATION METHOD

Hervé GUILLARD

Mars 1993

Institut National
de Recherche
en Informatique
et en Automatique

2004 route des Lucioles
B.P. 93
06902 Sophia-Antipolis
France

Analyse d'une méthode de relaxation multi-niveaux

Hervé Guillard

Résumé

On analyse une méthode multi-niveaux où les directions de descente sont projetées sur des espaces vectoriels de dimensions inférieures à celle de l' espace original. L'analyse est basée sur une décomposition de l'espace original en une somme directe (au sens de la norme en énergie) de n+1 espaces orthogonaux où n est le nombre de niveaux. On montre ensuite que les proprietés de convergence de cette méthode sont identiques à celles de méthodes multigrilles plus classiques.

Convergence Analysis of a Multi-level Relaxation Method

Hervé Guillard

abstract

We analyse a multi-level method based on successive projections of descent directions into vector spaces of smaller dimensions than the original space. We first show that the original space can be splitted into a direct sum (in the sense of the energy norm) of n+1 orthogonal spaces where n is the number of levels. Using this result, we then show that the convergence properties of this method are identical to the convergence properties of more classical multigrid methods.

1. Introduction :

This paper analyses a technique to accelerate the convergence rate of iterations of the form $u_{j+1} = u_j + d(u_j, f)$ where d is a "descent" direction, and u_j, u_{j+1} two successive approximations of the solution of a linear system $Au = f$ in a finite dimensional space $\mathcal{H}$. In this technique, the previous iterations are changed for a sequence of iterations of the type :

$$u_{j+1} = u_j + P_j^t P_j d(u_j, f) \tag{1}$$

where each P_j is a projection into a space $\mathcal{H}_j$ whose dimension is smaller than the dimension of $\mathcal{H}$ and P_j^t is its transpose. This technique has been introduced in [1] for non-linear shape optimization problems. Numerical experiments have shown that this procedure is also efficient for linear problems and can be of interest as a linear solver. It is clear that the above procedure has strong connections with the multigrid strategy. Actually, when the dimension of the linear system is large, multigrid techniques are certainly among the best ways to solve a linear system. However there exist some cases where the application of a multigrid strategy is difficult. These cases correspond to situations where the assembly of the coarse grid operators are difficult or even impossible. The use of non-structured finite element type meshes provides a first example of this type of situations : When the solution has to be computed on a given mesh for which no natural coarsening exists, complicated strategies have to be devised to define the coarse levels and the inter-grid transfers. Among the many solutions of this problem (agglomeration methods [6], Algebraic Multigrid [10], Delaunay coarsening [5]) a possibility that has been successfully tested in aerodynamics [8] is the use of totally unrelated meshes that are generated under the sole constraints that they approximately fit the same boundaries. However the complexity and the storage required by this type of algorithm can be prohibitely large and restrict the use of multigrid methods to these problems. A second example occurs when the matrix A is never explicitely built but results from a sequence of elementary simple steps ; for instance A may be the product of simple sparse matrices while A itself is a full matrix that is impossible to store and thus the matrix A is only indirectly known by its product on a given vector. This situation may results from finite element discretization of very large problems or from optimal control problems when the control is on the terminal state while the initial state has to be identified. In such cases, even if inter-grid transfer can be built, it may happen that the assembly of coarse grid operators is simply impossible or that operations on the coarse levels are as costly than the fine grid operations. For such situations, the use of the iterative scheme (1) can be advocated because it never requires the building of the coarse grid operators and thus is much more simpler to implement than a multigrid method. Only straightforward projections of a vector of $\mathcal{H}$ have to be done. Moreover, the construction of the coarse levels never appears explicitely and can be totally fictitious. Of course, it is clear that one of the most important idea of multigrid methods is missing in the multi-level method (1). What is missing is the idea to perform the relaxation on the coarse mesh with coarse grid operators that are less costly to handle. Consequentely while multigrid methods can solve a N degree of freedom problem with an optimal cost (*i.e* with a cost proportional to N or NlogN) this will not happen with scheme (1) where each iteration requires the evaluation of the descent direction d on the fine grid. However, despite this limitation, the multi-level method (1) shares some interesting features with multigrid methods. In particular, it is the purpose of this paper to show that in spite of its simplicity, the iterative scheme (1) exhibits a mesh independent convergence rate as in multigrid method.

2. Multi-level Relaxations :

Let $\mathcal{H}$ be a finite-dimensional Hilbert space. We consider on $\mathcal{H}$ a linear operator equation :

$$A\bar{u} = f \tag{2}$$

where $A \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ is a symmetric positive definite linear operator and $f \in \mathcal{H}$. We introduce a hierarchy of finite dimensional space $\mathcal{H}_0, \mathcal{H}_1, ..., \mathcal{H}_n = \mathcal{H}$ and a family of transfert operators between these spaces represented by full rank matrices : $I_i^{i+1} \in \mathcal{L}(\mathcal{H}_i, \mathcal{H}_{i+1}), I_{i+1}^i \in \mathcal{L}(\mathcal{H}_{i+1}, \mathcal{H}_i)$. We assume that the inter-space transfers satisfy the condition :

$$I_i^{i+1} = (I_{i+1}^i)^t \tag{3}$$

With the help of these operators, transfers $I_i^j \in \mathcal{L}(\mathcal{H}_i, \mathcal{H}_j)$ between any level of the hierarchy $\mathcal{H}_0, \mathcal{H}_1, ..., \mathcal{H}_n = \mathcal{H}$ can be defined by :

$$\begin{aligned} I_i^j &= I_{j-1}^j I_{j-2}^{j-1} \cdots I_i^{i+1} \text{ if } i < j \\ I_i^j &= I_{j+1}^j I_{j+2}^{j+1} \cdots I_i^{i-1} \text{ if } i > j \end{aligned} \tag{4}$$

In particular $I_i^n \in \mathcal{L}(\mathcal{H}_i, \mathcal{H}_n), I_n^i \in \mathcal{L}(\mathcal{H}_n, \mathcal{H}_i)$ define the transfert operators between $\mathcal{H}_i$ and $\mathcal{H}_n = \mathcal{H}$. Note that (3-4) imply that we have :

$$I_i^n = (I_n^i)^t \tag{5}$$

Given a basic relaxation process of the form :

$$u \leftarrow u + B^{-1}(f - Au) \tag{6}$$

we will solve (2) by a sequence of modified relaxation steps where the descent directions $B^{-1}(f - Au)$ are projected into $\mathcal{H}_i$ for $i = 0, \cdots, n$:

$$\begin{array}{ll} 1. & \text{initialize : } u_0 = u \\ 2. & \text{for } j = 0, \cdots, n \\ & \quad \text{perform } \nu \text{ steps of the modified relaxation :} \\ & \qquad u_j \leftarrow u_j + I_j^n I_n^j B^{-1}(f - Au_j) \\ & \quad u_{j+1} \leftarrow u_j \\ 3. & u \leftarrow u_{n+1}, \quad \text{go to 1. until convergence} \end{array} \tag{7}$$

The above algorithm is very much alike a sawtooth coarse to fine multigrid cycle. To point out these connections, let us examine a MG sawtooth cycle when only one relaxation by level is performed. From the definition of a MG sawtooth cycle (see e.g [2]) , it can be infered that the j level correction U^j can be deduced from the $j-1$ level correction U^{j-1} by :

$$U^j = I_{j-1}^j U^{j-1} + (B^j)^{-1}(I_n^j r - A^j I_{j-1}^j U^{j-1}) \tag{8}$$

with $r = f - Au$ and B^j, A^j are the operators corresponding to B, A on the coarse grid. Let us denote $u^j = u + I_j^n U_j$ the corrected approximation of the solution. From (8) we obtain :

$$u^j = u^{j-1} + I_j^n (B^j)^{-1} I_n^j (f - Au^{j-1}) \tag{9}$$

Comparing this relation with (7), we see that the two expressions are identical if $I_j^n (B^j)^{-1} I_n^j = I_n^j I_j^n B^{-1}$. In particular, for Richardson's method where B is a multiple of the identity, the multi-level method (7) is exactly a MG sawtooth cycle. We now proceed to prove that in the general case the multi-level method converge with a mesh-independent rate.

3. Preliminaries :

The Hilbert space $\mathcal{H}$ can be equipped with the energy inner product $:< u, v >_A = < Au, v >$. We denote by $\| u \|_A$ the associated energy norm and we associate to each space $\mathcal{H}_i$ a linear operator $A^i \in \mathcal{L}(\mathcal{H}_i, \mathcal{H}_i)$ defined by:

$$A^i = I_n^i A I_i^n \tag{10}$$

The intergrid transfers I_n^i induce the splitting of $\mathcal{H}$ into two A-orthogonal spaces:

$$\mathcal{H} = \mathcal{S}^i \oplus \mathcal{T}^i \tag{11}$$

where $\mathcal{S}^i = \mathcal{R}(I_i^n)$ and $\mathcal{T}^i = A^{-1}Ker(I_n^i)$. Denote by S^i and T^i the A–orthogonal projectors on $\mathcal{S}^i$ and $\mathcal{T}^i$. The following results on the projectors S^i and T^i are classical in multigrid convergence theory :

Lemma 1. **a.** $\forall i, \forall x \in \mathcal{H} \; \| x \|_A^2 = \| S^i x \|_A^2 + \| T^i x \|_A^2$
b. $S^i x = I_i^n X_i$ where $X_i \in \mathcal{H}^i$ is the solution of the coarse grid problem : $A^i X = I_n^i Ax$.
c. $\forall x \in \mathcal{H}, \| S^i x \|_A = \| X_i \|_{A^i}$ where $\forall X \in \mathcal{H}^i, \| X \|_{A^i}^2 = < A^i X, X >$.
d. $S^i I_i^n = I_i^n$ and $T^i I_i^n = 0$.

proof :(see e.g. [7], [4]).

We now show that $\mathcal{S}^i$ is equal to the following direct sum of A-orthogonal spaces:

Proposition 1. : $\mathcal{S}^i = \mathcal{S}^{i-1} \oplus \mathcal{S}^i \cap \mathcal{T}^{i-1}$

proof : We first remark that we have $A^{i-1} = I_i^{i-1} A^i I_{i-1}^i$ and that each space $\mathcal{H}^i$ can be decomposed into two A^i–orthogonal spaces :

$$\mathcal{H}^i = \mathcal{S}_{i-1}^i \oplus \mathcal{T}_{i-1}^i \tag{12}$$

where here $\oplus$ denote an A^i–orthogonal direct sum, $\mathcal{S}_{i-1}^i = \mathcal{R}(I_{i-1}^i)$ and $\mathcal{T}_{i-1}^i = (A^i)^{-1} Ker(I_i^{i-1})$.

Now, let $x \in \mathcal{S}^i$, then $\exists X_i \in \mathcal{H}^i$ s.t. $x = I_i^n X_i$, by (12) we have that there exist $(Y_i, Z_i) \in \mathcal{S}_{i-1}^i \times \mathcal{T}_{i-1}^i$ s.t. $X_i = Y_i + Z_i$. Because $Y_i \in \mathcal{S}_{i-1}^i$, $\exists Y_{i-1} \in \mathcal{H}^{i-1}$ s.t. $Y_i = I_{i-1}^i Y_{i-1}$ and thus $I_i^n Y_i = I_{i-1}^n Y_{i-1} \in \mathcal{S}^{i-1}$.

To prove that $I_i^n Z_i$ belongs to $\mathcal{T}^{i-1}$, let us show that it is A–orthogonal to any $q = I_{i-1}^n Q_{i-1} \in \mathcal{S}^{i-1}$, we have : $< A I_i^n Z_i, q > = < A^i Z_i, I_{i-1}^i Q_{i-1} > = 0$ because Z_i is A^i–orthogonal to any element of $\mathcal{S}_{i-1}^i$. Thus we have shown: $\mathcal{S}^i \subset \mathcal{S}^{i-1} \oplus \mathcal{S}^i \cap \mathcal{T}^{i-1}$.

The converse follows trivially from : $\mathcal{S}^i = \mathcal{S}^i \cap \mathcal{H} = \mathcal{S}^i \cap (\mathcal{S}^{i-1} \oplus \mathcal{T}^{i-1})$ and

$\mathcal{S}^i \cap \mathcal{S}^{i-1} = \mathcal{S}^{i-1}$.

$\square$

Remark 1. : Note that the previous result can be written : $\mathcal{S}^i = \mathcal{S}^i \cap (\mathcal{S}^{i-1} \oplus \mathcal{T}^{i-1}) = \mathcal{S}^i \cap \mathcal{S}^{i-1} \oplus \mathcal{S}^i \cap \mathcal{T}^{i-1}$.

Corollary 1: The space $\mathcal{H}$ can be decomposed into the following A–orthogonal direct sum :

$$\mathcal{H} = \mathcal{S}^0 \oplus (\mathcal{S}^1 \cap \mathcal{T}^0) \oplus \dots (\mathcal{S}^{n-1} \cap \mathcal{T}^{n-2}) \oplus \mathcal{T}^{n-1} \tag{13}$$

This follows immediately from proposition 1.

Denote by $Q_{i,i-1}$ the projectors that map $\mathcal{H}$ onto $\mathcal{S}^i \cap \mathcal{T}^{i-1}$ the expression of these operators is also a simple consequence of Prop.1.
Corollary 2: Let $x \in \mathcal{H}$ then $Q_{i,i-1}\ x \ = \ (S^i - S^{i-1})x$.
proof : By prop. 1, we have $S^i x = S^{i-1}x + Q_{i,i-1}x$.

Corollary 3: : Let S^i_{i-1} and T^i_{i-1} be the projectors that map $\mathcal{H}^i$ onto $\mathcal{S}^i_{i-1}$ and $\mathcal{T}^i_{i-1}$, let $X_i \in \mathcal{H}^i$ be such that $S^i x = I^n_i X_i$ then we have : $S^{i-1}x = I^n_i S^i_{i-1} X_i$ and $Q_{i,i-1}x = I^n_i T^i_{i-1} X_i$.

The next results establishing the relationships between the projections on two successive levels are easy to obtain :
Lemma 2. **a.** $S^i S^{i-1} = S^{i-1} S^i = S^{i-1}$
b. $T^i S^{i-1} = S^{i-1} T^i = 0$
c. $T^i T^{i-1} = T^{i-1} T^i = T^i$
d. $S^i T^{i-1} = T^{i-1} S^i = S^i - S^{i-1} = T^{i-1} - T^i$
proof : Let us prove for instance **a**, the proofs of the other results will follow immediately :
$\forall x \in \mathcal{H}$ we have :

$$\begin{aligned} S^i S^{i-1} x &= I^n_i (A^i)^{-1} I^i_n A \{ I^n_{i-1} (A^{i-1})^{-1} I^{i-1}_n A x \} \\ &= I^n_i (A^i)^{-1} (I^i_n A I^n_i) I^i_{i-1} (A^{i-1})^{-1} I^{i-1}_n A x \\ &= I^n_{i-1} (A^{i-1})^{-1} I^{i-1}_n A x = S^{i-1} x \qquad \square \end{aligned}$$

We note that we also have :

$$\forall j \leq i, \quad S^i I^n_j = I^n_j \quad \text{and} \quad T^i I^n_j = 0 \tag{14}$$

this follows from $S^i I_j^n x = S^i I_i^n (I_{i-1}^n \cdots I_j^{j+1})\ x$ and lemma 1 d.

4. Convergence proof :

The proof of the convergence of the multi–level method is based on the following A–orthogonal splitting of any vector $x \in \mathcal{H}$ that follows directly from prop.1 :
Let $x \in \mathcal{H}$ then :

$$\begin{aligned} \| x \|_A^2 = \ & \| T^{n-1} x \|_A^2 + \| (S^{n-1} - S^{n-2})(x) \|_A^2 \\ & + \cdots + \| (S^1 - S^0)(x) \|_A^2 + \| S^0(x) \|_A^2 \end{aligned} \tag{15}$$

Let $e^0 = A^{-1} f - u$ be the error at the beginning of (7), we denote by $e^j = A^{-1} f - u_j$ the error after j steps of the cycle (7) and by E^j the element of $\mathcal{H}^j$ s.t. $S^j e^j = I_j^n E^j$. In term of the errors, the iterations performed on the j^{th} level of the cycle (7) becomes :

$$e^j \leftarrow e^j - I_j^n I_n^j B^{-1} A e^j = T^j e^j + I_j^n (E^j - I_n^j B^{-1} A e^j) \tag{16}$$

and thus the relaxations $u \leftarrow u + I_j^n I_n^j B^{-1}(f - Au)$ performed on the j^{th} level modify the component of the error in $\mathcal{S}^j$ but leave unchanged the component in $\mathcal{T}^j$. According to Prop.1, $\mathcal{S}^j$ is the direct A–orthogonal sum of $\mathcal{S}^{j-1}$ and $\mathcal{S}^j \cap \mathcal{T}^{j-1}$. Thus our main assumption about the relaxation process will be that performed on the j^{th} level, it reduces uniformely in j the "oscillatory" part of the component of the error that belongs to $\mathcal{S}^j$ *i.e* the component in $\mathcal{S}^j \cap \mathcal{T}^{j-1}$:
Assumption : There exists β independent of j such that :

$$\| S^j e^{j+1} \|_A^2 \leq \| S^{j-1} e^j \|_A^2 + (1 - \beta) \| Q_{j,j-1} e^j \|_A^2 \tag{17}$$

Before proving that (17) implies the convergence of (7) with a mesh-independent rate, let us show that this inequality is equivalent to the usual assumption done in multigrid convergence studies (see e.g. [3] Lemma 2.2, [9] Assumption 3.1 & 3.2) :
Proposition 2 : The inequality (17) is equivalent to :

$$\| E^{j+1} \|_{A^j}^2 \leq \| E^j \|_{A^j}^2 - \beta \| T_{j-1}^j E^j \|_{A^j}^2 \tag{18}$$

proof : (18) is equivalent to :

$$\| E^{j+1} \|_{A^j}^2 \leq \| S_{j-1}^j E^j \|_{A^j}^2 + (1 - \beta) \| T_{j-1}^j E^j \|_{A^j}^2 \tag{19}$$

and from Corollary 3 we have : $I_j^n S_{j-1}^j E^j = S^{j-1} e^j$ and $I_j^n T_{j-1}^j E^j = Q_{j,j-1} e^j$, moreover for any $W \in \mathcal{H}^j$ $\| W \|_{A^j} = \| I_j^n W \|_A$ and thus (19) is exactly (17).

Remark 2 : (18) follows from the so-called "smoothing" and "approximation" properties. In [9] it is shown that these properties are valid for a large class of relaxation processes and grid transfers.

We now show that (17) implies the convergence of the cycle (7):

Proposition 3. : Let us suppose that the relaxation satisfies (17) for any level j then the cycle (7) converges in energy with a convergence rate bounded at least by $(1-\beta)^{1/2}$.

proof : The error e^2 after 2 steps of (7) satisfies :

$$\| S^1 e^2 \|_A^2 \leq \| S^0 e^1 \|_A^2 + (1-\beta) \| Q_{1,0} e^1 \|_A^2$$

but we have by Corollary 2 : $Q_{1,0} e^1 = (S^1 - S^0) e^1 = (S^1 - S^0) e^0 - (S^1 - S^0) I_0^n W^0$ where W^0 is an element of $\mathcal{H}^0$. Thus $Q_{1,0} e^1 = Q_{1,0} e^0$ because $S^1 I_0^n W^0 = S^0 I_0^n W^0$ by (14).

We introduce the additional assumption that the reduction of the error in $\mathcal{H}^0$ by the first step of (7) is sufficiently efficient to have :

$$\| S^0 e^1 \|_A^2 \leq (1-\beta) \| S^0 e^0 \|_A^2$$

and we now proceed by induction. Assuming that we have :

$$\| S^{j-1} e^j \|_A^2 \leq (1-\beta) \{ \| S^0 e^0 \|_A^2 + \| Q_{1,0} e^0 \|_A^2 + \cdots + \| Q_{j-1,j-2} e^0 \|_A^2 \}$$

we have by (17) :

$$\| S^j e^{j+1} \|_A^2 \leq \| S^{j-1} e^j \|_A^2 + (1-\beta) \| Q_{j,j-1} e^j \|_A^2$$

but using Corollary 2 and recursively (14) we have

$$\begin{aligned} Q_{j,j-1} e^j &= Q_{j,j-1} e^{j-1} - Q_{j,j-1} I_j^n W^j \\ &= Q_{j,j-1} e^{j-1} \\ &= \dots \\ &= Q_{j,j-1} e^0 \end{aligned}$$

and thus :

$$\begin{aligned} \| S^j e^{j+1} \|_A^2 \leq & (1-\beta) \{ \| S^0 e^0 \|_A^2 + \| Q_{1,0} e^0 \|_A^2 + \cdots + \| Q_{j-1,j-2} e^0 \|_A^2 \\ & + \| Q_{j,j-1} e^0 \|_A^2 \} \end{aligned}$$

The last step of (7) gives :

$$\| e^{n+1} \|_A \leq (1-\beta) \| T^{n-1} e^n \|_A + \| S^{n-1} e^n \|_A$$

but by (14) we also have : $T^{n-1}e^n = T^{n-1}e^{n-1} = \cdots = T^{n-1}e^0$ and we finally obtain:

$$\begin{aligned} \| e^{n+1} \|_A^2 \quad & \leq (1-\beta)\{\| S^0 e^0 \|_A^2 + \| Q_{1,0} e^0 \|_A^2 + \cdots + \| Q_{n,n-1} e^0 \|_A^2 \\ & + \| T^{n-1} e^0 \|_A^2 \} \end{aligned}$$

□

We end by noting that the cycle (7) is not the only one that can be built from the idea of using as relaxation a succession of projected descent direction of the form $I_j^n I_n^j d$. Similar cycles to the multigrid W, F , $\cdots$ –cycles can be designed. The convergence of these cycles can also be analysed using the splitting (13) of $\mathcal{H}$ into an A–orthogonal direct sum. Although we do not claim that this expression of $\mathcal{H}$ as a direct sum of A–orthogonal spaces is a new result, it seems that it does not appears explicitely in previous papers on multigrid theory where, instead the recursivity of the multigrid construction is exploited to use only the relation between two successive levels. We believe that the expression (13) is of interest in itself and that it may give a different light on the analysis of multigrid techniques.

Acknowledgements

The work presented here originated from several discussions with A. Dervieux and I acknowledge with pleasure his contribution to this work.

References

[1] F. Beux, A. Dervieux, submitted to Engineering Computations.

[2] S.F. Mc Cormick, SIAM J. Numer. Anal. **21**, (1984), pp 255-263.

[3] S.F. Mc Cormick, SIAM J. Numer. Anal. **22**, (1985), pp 634-643.

[4] S.F. Mc Cormick and J. W. Ruge, SIAM J. Numer. Anal. **19**, (1982), pp 924-929.

[5] H. Guillard, Node-nested Multigrid for non-structured meshes, INRIA research Report in preparation.

[6] M.H. Lallemand, H. Steve, A. Dervieux, Computers & Fluids, **21**, (1992), pp397-433.

[7] J. F. Maitre and F. Musy, SIAM J. Numer. Anal. **21**, (1984) pp 657-671.

[8] D. Mavriplis and A. Jameson, Multigrid solution of the two-dimensional Euler equations on unstructured meshes, AIAA paper 87-0350, (1987).

[9] J. Mandel, S. Mc Cormick and J. Ruge, SIAM J. Numer. Anal. **25**, (1988), pp 91-110.

[10] J. Ruge and K. Stüben, Algebraic Multigrid Methods in *Multigrid Methods* (S. Mc Cormick Ed), Frontiers in Applied Math, Vol 5, SIAM, Philadelphia.