ON AGGREGATION METHODS

J.P. QUADRAT

INTRODUCTION.

The purpose of this paper is to give an informal discussion of ideas of aggregation wich appear with different names in many fields : system theory, probability, numerical analysis, econometrics, power systems. The idea is always the same we make a projection on a more or less well approximated invariant subspace of the initial linear system.

After a brief discussion of aggregation introduced by Aoki [2,3] and the dual notion of coherency discussed in Kokotovic [14] we show that the particularization of this notion to Markov chains gives at least conceptually the generalization of this notion to non linear systems. Then we discuss its interest to solve linear systems of equations. We give a new result of convergence of iterative aggregation methods explained for example in Khomyakov [18], which consist in adapting the aggregation in such way that we can prove the convergence to the exact solution.

We recall the results obtained in finite element methods which show a way to build the aggregated operators with nice robustness properties. We show that reversible Markov chains have a variational formulation very similar to the one used for the approximation of elliptic operators. Finally, we discuss multigrids methods (Brandt [23], Nicolaïdes [30], Hackbush [24])

which use the ideas of aggregation to design very fast iterative algorithms to solve some linear systems obtained by discretization of partial differential equations.

2. AGGREGATION AND COHERENCY OF LINEAR SYSTEMS.

The aggregation has been introduced in system theory by Aoki [2]. It is a straight forward application of the observability results of linear system, to their order reduction.

Given the dynamic system :

$$(2.1)\quad \begin{cases} \dot{x} = Ax \\ \\ x(o) \text{ given} \end{cases}$$

where A is a (N,N) matrix, the observation

$$(2.2)\quad y = Cx$$

where C is a (r,N) matrix, $r < N$, we can define an aggregated model of (1) by :

$$(2.3)\quad \begin{cases} \dot{y} = \bar{A}_c y \\ \\ y(o) = C\, x(o) \end{cases}$$

if it exists a (r,r) matrix $\bar{A}$ such that :

$$(2.4)\quad CA = \bar{A}_C C$$

holds ; that is the kernel of C denoted by $\mathcal{N}(C)$ is an invariant subspace of A : $A\,\mathcal{N}(C) \subset \mathcal{N}(C)$. $\mathcal{N}(C)$ is the maximal A-invariant subspace of $\mathbb{R}^N$ unobservable from y. We say that A is C aggregable.

Clearly the dual notion exists it is the coherency point of view Kokotovic [14]. The (N,r) control matrix B defines the aggregated model :

$$(2.5)\quad \begin{cases} \dot{v} = \bar{A}_B\, v \\ \\ v(o) \text{ given} \end{cases}$$

if it exists a (r,r) matrix $\bar{A}_c$, and v(o) such that :

$$(2.6)\quad \begin{cases} AB = B\bar{A}_B \\ \\ x(o) = B\,v(o) \end{cases}$$

that is $R(B)$ is an invariant subspace of A : $A\,R(B) \subset R(B)$. $R(B)$ is the maximal A-invariant subspace controlable by v , of the dynamic system

$$(2.7)\qquad \dot{x} = Ax + Bv.$$

We shall say that A is B coherent.

Clearly, weaker notion of aggregability [resp. coherency] can be introduced by keeping the observable part [resp. the controlable part] of the system (A,C) [resp.(A,B)] to obtain an order reduction A.V. Gaitsgori and A.A. Pervozvanskii [17], E.C.Y. Tse, J.V. Medanic, W.R. Perkins [29].

Specific algorithms for choosing a matrix C are given in Bertrand [19], Commault [20].

3. SPECIALIZATION TO MARKOV CHAINS AND GENERALIZATION TO NON LINEAR SYSTEMS.

We first specialize the results given in the part 2 to stochastic matrices A with B and C formed with vectors having disjoint supports. Then we show that in some sense this particular case is a good conceptual generalization of aggregation to non linear systems.

We consider the recurrent irreductible Markov chain X_t defined on the set $E = \{1,2,\ldots,N\}$, its (N,N) transition matrix is denoted by M whose entry $M_{x'x}$ denotes the probability to go in x' from x, $\forall x,x' \in E$. p^t denotes the N-vector whose entry p^t_x is the probability to be in x at time t. p^t satisfies :

$$(3.1) \qquad p^{t+1} = Mp^t$$

Given a partition $U = \{E_1,E_2,\ldots,E_r\}$ we denote by U the (r,N) characteristic matrix of this partition that is :

$$(3.2) \qquad U_{ix} = \begin{cases} 1 \text{ if } x \in E_i \\ 0 \text{ elsewhere} \end{cases}$$

Then, the U aggregation of the matrix M : UM = $\bar{M}$U is the strong lumpability condition of Kemeny-Snell [5]

$$(3.3) \qquad \sum_{x' \in E_i} M_{x'x} = \bar{M}_{ij} \qquad \forall x \in E_j \ , \ \forall E_i,E_j \in U$$

The interest of this particular case is the stochastic interpretation of the (r,r) matrix $\bar{M}$ as the transition matrix of the aggregated chain Y_t defined by :

$$(3.4) \qquad Y_t = UX_t$$

that is $Y_t = E_i$ when $X_t \in E_i$

The dual notion that is the coherency can be seen as the aggregability of the reversed chain that is of the Markov chain $\hat{X}_t$ of transition matrix $QM'Q^{-1}$, with Q = diag {q}, where q denotes the invariant measure of X_t (which is unique by the irreductibility of X_t). Thus, we can take :

$$(3.5) \qquad B = QU'(UQU')^{-1}$$

B is the conditional probability law q^U where U denotes the partition defined by U indeed

$$(3.6) \qquad B_{xi} = q^U_{xi} = \frac{q_x}{\sum_{x \in E_i} q_x} \quad \forall E_i \in U \quad , \quad \forall x \in E$$

Thus if the Markov chain X_t is B coherent $p^t = q^U \bar{p}^t$ if $p^o = q^U \bar{p}^o$ with $\bar{p}^t$ the probability law of Y_t which is a Markov process of transition matrix $\bar{M}_B = UMB$. A more detailed discussion of this notions, stochastic and network interpretations, can be found in Delebecque-Kokotovic-Quadrat [31].

It is well known that the discretization of non linear stochastic systems leads to a discrete problem which can be interpreted in term of Markov chains. Let us consider the triangular linear system

$$(3.7) \qquad \begin{cases} \dot{x}_1 = \lambda_1 x_1 \\ \dot{x}_2 = a_1 x_1 + a_2 x_2 \end{cases}$$

The corresponding Markov chain is obtained by discretization of the operator :

$$(3.8) \qquad \lambda_1 x_1 \frac{\partial}{\partial x_1} + (a_1 x_1 + a_2 x_2) \frac{\partial}{\partial x_2}$$

which is defined by the transition matrix $A^h : \mathbb{R}^{Z \times Z} \to \mathbb{R}^{Z \times Z}$ whose non zero entries

$$(3.9) \qquad \begin{aligned} A^h_{x_1,x_2;x_1+h,x_2} &= \frac{k}{h} (\lambda_1 x_1)^+ \\ A^h_{x_1,x_2;x_1-h,x_2} &= \frac{k}{h} (\lambda_1 x_1)^- \qquad x_1 \in Zh \\ A^h_{x_1,x_2;x_1,x_2+h} &= \frac{k}{h} (a_1 x_1 + a_2 x_2)^+ \quad x_2 \in Zh \\ A^h_{x_1,x_2;x_1,x_2-h} &= \frac{k}{h} (a_1 x_1 + a_2 x_2)^- \\ A^h_{x_1,x_2;x_1,x_2} &= 1 - \frac{k}{h} \{ |\lambda_1 x_1| + |a_1 x_1 + a_2 x_2| \} \end{aligned}$$

k can be seen as a discretization step in time.

Than the Markov chain living on $Z \times Z$, of transition matrix $A^{h'}$ gives an approximation of the solution (3.7) indeed p^t solution of

$$(3.10) \qquad p^{t+1} = A^{h'} p^t$$

converges in law to δ_x when k and $h \to o$. k/h fixed see for example Stroock-Varadhan [26], Kushner [27], Quadrat [28].

Now if we take for definition of the partition $U = \{E_{x_1}, x_1 \in Zh\}$ with $E_{x_1} = \{(x_1,x_2), x_2 = Zh\}$ the condition of aggregation is satisfied for the corresponding U indeed

$$(3.11) \qquad \sum_{x_2'} A^h_{x_1,x_2;x_1+h_1x_2'} = \frac{k}{h}(\lambda_1 x_1)^+ \qquad \forall x \in E_{x_1}$$

and

$$(3.12) \qquad \sum_{x_2'} A^h_{x_1,x_2;x_1-h,x_2'} = \frac{k}{h}(\lambda_1 x_1)^- \qquad \forall x \in E_{x_1}$$

and then we have seen that associated to the invariant space $x_1 = o$ in the formulation (3.7) corresponds a fibration of the space parallel to this invariant subspace such that there exists an aggregation for the corresponding Markov chain. Clearly the fact that the fibration is linear is not important. We needs only to find a nonlinear system of coordinate such that the dynamic has the form

$$(3.13) \qquad \begin{cases} \dot{x}_1 = a_1(x_1) \\ \dot{x}_2 = a_1(x_1,x_2) \end{cases}$$

Find such a system of coordinate corresponds to find invariant distributions of a vector field Isidori-Krener-Monaco [32]. Thus we see that in some sense the particularization of the aggregation to the characteristic function of a partition is a generalization to nonlinear stochastic systems of the notion of aggregation for the linear ones.

4. ITERATIVE AGGREGATION.

Based on the aggregation ideas of the former paragraph it is possible to design a class of multi-scale algorithm to solve exactly linear systems. This idea has been used in economic litterature for example V.A. Khomyakov [18], I.Y. Vakhutinsky-L.M. Dudkin , A.A. Ryvkin [21]. We adapt one of these algorithm using multigrid ideas to obtain the natural conditions of convergence. This result improves the Khomyakov one .

Let us consider the Markov chain X_t and suppose that we want compute the cost

$$(4.1) \qquad V_x = \mathbb{E}\left\{\sum_{t=o}^{+\infty} \frac{1}{(1+\lambda)^{t+1}} f(X_t) \,|\, X(o) = x\right\} \qquad \forall x \in E$$

where $f : \mathbb{E} \to \mathbb{R}^+$ is seen as a vector belonging to $\mathbb{R}^N$, λ is an actualization rate $\lambda \in \mathbb{R}^+$, $\lambda > o$.

V is solution of the Kolmogorov equation

$$(4.2) \qquad V = M^\lambda V + f^\lambda$$

with

$$(4.3) \qquad M^\lambda = \frac{M'}{1+\lambda}$$

$$(4.4) \qquad f^\lambda = \frac{f}{1+\lambda}$$

' denotes the transposition.

Until the end of this part we do not write the index λ.

The $(k+1)^{th}$ iteration of the algorithm is :

- Given V^k compute B^k define by (3.5) with $q = V^k$
- Aggregation step : compute $\bar{V}^{k+1}$ solution of

$$(4.5) \qquad \bar{V}^{k+1} = UMB^k\bar{V}^{k+1} + Uf$$

- Desaggregation (m relaxation steps)

(4.6) $V_{l+1}^k = MV_l^k + f \qquad l = 1,2,\ldots,m-1$

(4.7) $V_o^k = B^k \overline{V}^{k+1}$

(4.8) $V^{k+1} = V_m^k$

Theorem 1

$\forall \lambda > o$, it exists $m(\lambda)$ such that $V^k \underset{k\to\infty}{\to}$ V solution of (4.2). Moreover we have the relation

(4.9) $\delta^{k+1} = M^{m-1}(I - MP^k)^{-1} M(P^k - I)\delta^k$

with

(4.10) $\delta^k = V^k - V$

(4.11) $P^k = B^k U$

Proof : is given for completness in annex 1, but is straight forward from the Khomiakov one.

Remark 4.1.

The algorithm of Khomyakov [18] is the particular case m = 1, and is applied for matrix M having the same properties than the transposed matrix used here, but the conditions of convergence are much more restrictive.

Remark 4.2.

The one stage method of multigrids is similar to the case where B^k is independant of k, for example W. Hackbush [24].

Remark 4.3.

Neither the aggregation nor the coherency porperty of part 3 are supposed to be true here, and in fact with the definition (4.3), the operator U and B^k are not applied on the right side.

Remark 4.4.

To solve the linear system (4.5) we can reiterate the same method indeed UMB^k and M have the same property. This is the idea of Multigrid methods see part 6.

Remark 4.5.

The formula (4.9), in the context of random walk, shows that :

- $(I - P^k)$ is a projector which plays the role of "slow-pass-filter",

- M plays the role of a "high-pass-filter".

5. ROBUSTNESS AND VARIATIONAL METHODS.

The point of view used in finite element methods to approximate the solution of partial differential equation is very closed of the one discussed in part 3. The main differences are the approximated character of the aggregation and the overlapping support aspect of the columns of B. Nevertheless by this way the "slow part " (restriction to the space generated by the eigenvectors associated to small eigenvalues) is well represented in the approximate model. We recall some results of Strang and Fix [12] on the approximation of eigenvalues and eigenvectors, discuss this result, and show that reversible Markov chains have the property which explains robustness by aggregation based on "grouping matrix".

First let us recall the finite element method on an example.

Given O a bounded open set of $\mathbb{R}^2$ of boundary Γ enough regular we want solve.

$$(5.1)\quad \begin{cases} \Delta W - \mu W = f & x \in O \\ \dfrac{\partial W}{\partial \vec{n}}\Big|_{\Gamma} = o \end{cases}$$

where $\mu \in \mathbb{R}^+$, $\mu > o$, $f \in L^2(O)$, $\vec{n}$ denotes the exterior normal to the boundary of O.

The operator $A = \Delta - \lambda I$ can be seen as a linear bounded operator from $V = H^1(O)^*$ into identified V' dual of V when we have identified $L^2(O)$ with its dual.

Given the operators

$$(5.2)\quad \begin{cases} B : \mathbb{R}^n \to V \\ C : V' \to \mathbb{R}^n \end{cases}$$

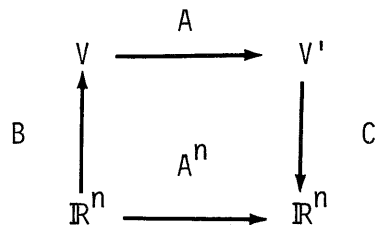
The operator

$$(5.3) \qquad A^n = CAB : \mathbb{R}^n \xrightarrow[A_n]{} \mathbb{R}^n$$

$$H^1(0) = \{v \mid v \in L^2(0), \frac{\partial v}{\partial x_i} \in L^2(0) \; i = 1,2\}$$

defines an approximation of (5.1).

We have the diagramm



In general we choose $C = B'$.

Associated to the symmetric operator A we define the quadratic form

$$(5.4) \qquad a(W,W) = \int_0 \mu W^2 + \nabla W . \nabla W \, d0$$

where ∇ denotes the gradient operator.

The solution of (5.1) minimizes on V, $a(W,W) - (f, W)$ and the solution is unique because a is coercive on V that is it exists $\alpha = \text{Min}\,(\lambda,1) : a(W,W) \geq \alpha \|W\|_V^2$. With the choice $C = B'$ the solution of

$$(5.5) \qquad A^n W^n = f^n$$

with

$$f^n = Cf$$

minimize $a(BW^n, BW^n) - (f, BW^n)$ on $\mathbb{R}^n$. BW^n is the best approximation of (5.1) for the norm defined by (5.4).

In the simplest finite element method we choose B as follows.

We make a partition of Ω with triangular sets after having approximated Ω by a polyhedra Ω^n

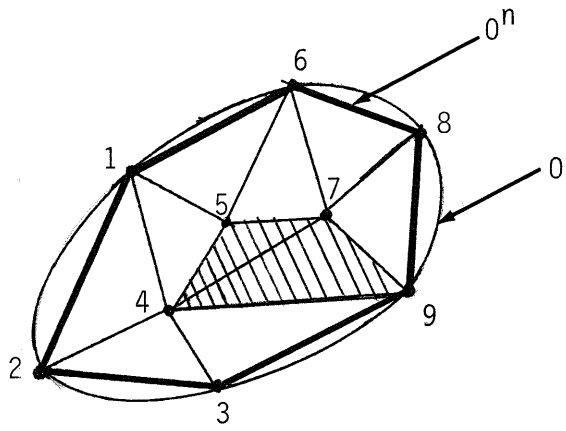


fig. 1

For each nodes of this triangularization we associate a "column of B" : the function piecewise linear taking the value 1 at this node and o elsewhere.

We are almost in the situation where the columns of B have disjoint support. For example on fig. 1 the columns 4 and 8 have disjoint support, but 4 and 7 overlapp on the dash zone. Moreover the condition of aggregation and/or coherency $AB = BA_n$ or $CA = A_nC$ is not satisfied. Nevertheless the eigenvalues and eigenvectors of A and A_n are related in the following way Strang and Fix [12].

$$(5.7) \qquad \lambda^l \leq \lambda_n^l \leq \lambda^l + \rho h^2 (\lambda^l)^2$$

$$(5.8) \qquad a(V^l - V_n^l, V^l - V_n^l) \leq \rho h^2 (\lambda^l)^2$$

where λ^l, V^l denotes the l^{th} eigenvalue and eigenvector of A and λ_n^l, V_n^l denotes the l^{th} generalized eigenvalue and eigenvector of (A_n, CB) that is

$$(5.9) \qquad A_n V_n^l = \lambda_n^l \, CBV_n^l$$

ρ is a constant and h is the largest diameter of the triangles of the partition.

When Ω is an hypercube, h is of order $\frac{1}{\sqrt{n}}$, λ_l increases like $l \in 1,\ldots,n$ (with 1) thus the approximation is good only for the small eigenvalues. This remark shows that on one hand it would be possible to obtain a lower order

model as precise as CAB if we were able to conserve only the invariant space associated to the eigenvalue which are well approximated in CAB, on the other hand this construction gives a way to build a space containing almost the invariant slow spaces of A.

Moreover it is clear that an invariant space of A belonging to $\mathbb{R}(B)$ is the image of a generalized invariant space of CAB. More generallu to have a good approximation of an invariant space X it is necessary that $d(X,\mathbb{R}) = \sup_{\substack{x\in X \\ ||x||=1}} \operatorname{Inf}_{y\in\mathbb{R}(B)} d(x,y)$

The slow eigenspace have a better representation in the aggregated model than the fast one because the eigenvectors associated to large eigenvalue are less "smooth" or oscillate more than the slow one.

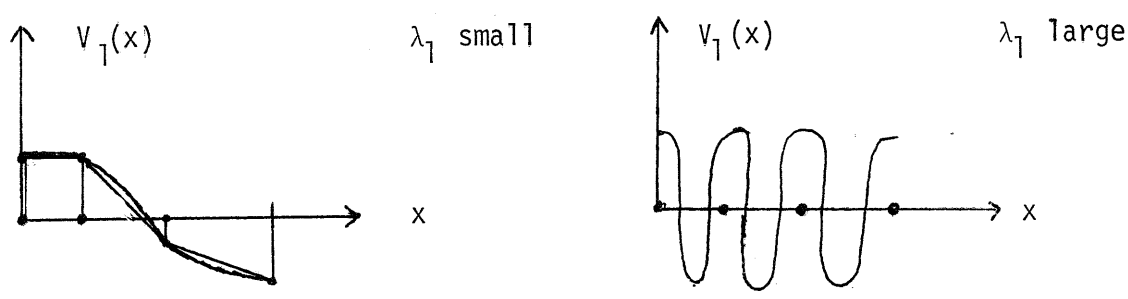
This property is not true in general but can easily be verified on this example indeed by definition of the eigenvalue,the eigenvectors V_l satisfies

$$\frac{a(V_l,V_l)}{|V_l|} = \lambda_l \geq \frac{a(V_{l'},V_{l'})}{|V_{l'}|} = \lambda_{l'} \qquad \text{if } \lambda_l \geq \lambda_{l'}$$

which implies that

$$(4.9) \qquad \frac{\int \nabla V_l \nabla V_l \, d\Omega}{\int V_l^2 d\Omega} \geq \frac{\int \nabla V_{l'} \nabla V_{l'} \, d\Omega}{\int V_{l'}^2 d\Omega}$$

and $\int \nabla V_l \nabla V_l d\Omega$ is a measure of the smootheness of V_l. Moreover it is clear that smooth function will have better approximation in $\mathbb{R}(B)$ than very oscillating ones.



Let us show now that the reversible Markov chains have the same property as Δ that is : it exists a variational formulation which measures the smoothness of the solution.

Let us consider the reversible Markov of transition matrix M it is an irreducible recurrent Markov chain and if q denotes its invariant measure $Mq = q$ and $Q = \text{diag}\{q\}$, M satisfies

$$M = Q^{-1} M' Q \tag{5.10}$$

For such Markov chains to solve

$$AV - \mu V = f \qquad \text{with} \quad A = M - I \tag{5.11}$$

is equivalent to

$$\underset{V}{\text{Min}} \sum_i \sum_{j \neq i} \{\frac{1}{2}(V_j - V_i)\, q_i\, a_{ij}(V_j - V_i) + \mu q_i V_i^2 - f_i q_i V_i\} \tag{5.12}$$

Indeed (5.11) is equivalent to

$$QMV - \mu QV = Qf \tag{5.13}$$

but QM is symetric than (5.13) is equivalent to

$$\underset{V}{\text{Min}} - V'QAV + \mu V'QV - 2V'Qf \tag{5.14}$$

but

$$V'QAV = \sum_i V_i \sum_{j \neq i} q_i\, a_{ij}(V_j - V_i) \tag{5.15}$$

because $\sum_j a_{ij} = o$.

$$V'QAV = \sum_j V_j \sum_{i \neq j} q_i\, a_{ij}(V_i - V_j) \tag{5.16}$$

because $\sum_i q_i a_{ij} = o$ which is true because q is invariant measure of M, thus :

$$(5.17) \qquad V'QAV = -\frac{1}{2} \sum_i \sum_{j \neq i} (V_j - V_i)\, q_i\, a_{ij} (V_j - V_i)$$

and (5.17) is a measure of the smoothness of V.

Then for this Markov chain we are in a situation very close of the P.D.E. one described before and we can hope to have analogous properties when we make an aggregation by grooping matrices, or finite elements (which, clearly, can be also used to aggregate finite dimensional systems).

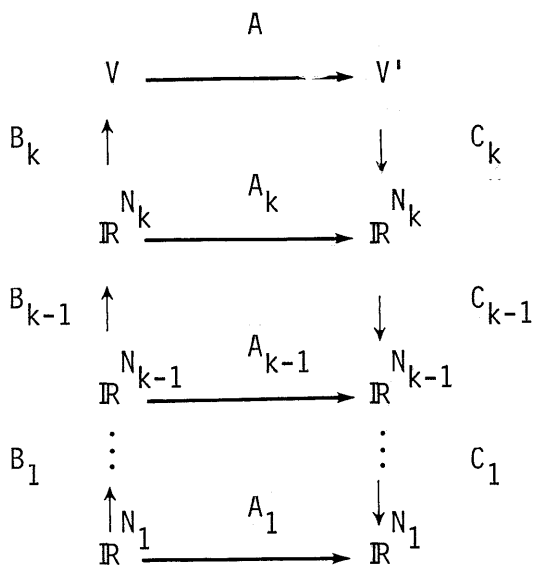
6. MULTIGRIDS METHODS Fedorenko [33], Brandt [23], Nicolaïdes [30], Hackbush [24],etc....

This technique is used in the context of the resolution of large linear systems obtained by discretization of partial differential equations. It is wanted to design an iterative algorithm with a contraction factor independent of the mesh size. This result is obtained using aggregation ideas. We build a pyramidal set of more and more aggregated models, the more aggregated ones giving a good approximation of the "slower part" of the system.

More precisely we have to solve :

(6.0) $AX + u = o$ with $A : V \to V'$ linear, bounded, coercif.

We use the diagram $u \in V'$



$A_k = C_k A B_k$

$A_{k-1} = C_{k-1} A_k B_{k-1}$

$\vdots$

$A_1 = C_1 A_2 B_1$

$N_1 < N_2 < \ldots < N_k$

To solve the approximate problem

(6.1) $A_k X_k + U_k = o$ with $C_k u = U_k$

we use an iterative method which is a numerical integration of the differential equation

(6.2) $\dot{X}_k = A_k X_k + U_k$

By a correct choice of C_k and B_k, A_k keeps the propertise of A : all its eigenvalues are negative. This property insure the stability of (6.2) and

$$\lim_{t\to\infty} X_k(t) = - A_k^{-1} U_k.$$

But the eigenvalues of A_k are spread in general on a very large interval for example in the case

(6.3) $A = \Delta$

(6.4) $O = [o,\pi]^2$

(6.5) $A_k X = \frac{1}{h^2} [X^{i+1,j} + X^{i,j+1} + X^{i-1,j} + X^{i,j-1} - 4X^{i,j}]$

(6.6) $h = \pi / \sqrt{N_k} + 1$

the eigenvalues of A_k are

(6.7) $\lambda^{m,n} = - (\frac{4}{h^2} \sin^2 \frac{mh}{2} + \frac{4}{h^2} \mathrm{Sin}^2 \frac{nh}{2})$ $\quad m = 1,\ldots, \sqrt{N_k}$ $\quad n = 1,\ldots, \sqrt{N_k}$

and the corresponding eigenvectors are

(6.8) $y^{m,n}_{i,j} = \mathrm{Sin}\ m\ ih\ \mathrm{Sin}\ n\ jh$

The eigenvalues are spread on the interval $[-\eta N_k, -\xi]$ where η and ξ are two constants of order 1 and the conditioning factor is $\frac{\xi}{\eta N_k}$. Then the discrete version of (6.2)

$$(6.9) \qquad \frac{X_k^{n+1} - X_k^n}{\rho} = A_k X_k^n + u_k$$

after optimizing ρ has a contraction factor of order $\frac{1}{N_k}$, and (6.9) is an algorithm in $\delta N_k^2 \log N_k$ to solve (6.1) with a precision of order $(\frac{1}{N_k})^\delta$. We remark also that the conjugate gradient method has a computational cost of order $(N_k)^2$ (N_k iterations each one having a cost of order N_k). Direct methods like Gauss elimination using the sparsity of the matrix gives algorithm of order $N_k^{3/2}$.

Let us suppose now that $\mathbb{R}(B_{k-1})$ (resp. $\mathbb{R}(B_{k-1}\ B_{k-2})$ is the right invariant space of A_k containing the N_{k-1} [resp. N_{k-2},...] slowest modes and that $\mathbb{R}(C'_{k-1})$ [resp. $\mathbb{R}(C'_{k-1}\ C'_{k-2})$...] is the left invariant space of A_k containing the N_{k-1} [resp. N_{k-2},...] slowest modes than

$$P_{k-1} = B_{k-1}\ C_{k-1} \text{ [resp. } P_{k-2} = B_{k-1}\ B_{k-2}\ C_{k-2}\ C_{k-1}, \ldots]$$

is a projector on $\mathbb{R}(B_{k-1})$ parallel to $\mathbb{N}(C_{k-1})$.

$$(6.10) \qquad \dot{Z}_k = A_k [I + \alpha_{k-1} P_{k-1} + \alpha_{k-2} P_{k-2} + \ldots \alpha_1 P_1] Z_k + U_k$$

with $\alpha_i \geq o$ has the same properties of stability than (6.2) and $Z_k(t) \underset{t\to\infty}{\to} Z_\infty$ which is such that

$$(6.11) \qquad (I + \alpha_{k-1} P_{k-1} + \ldots + \alpha_1 P_1) Z_\infty = - A_k^{-1} U_k = X_\infty$$

but the conditioning of the operator of $\tilde{A}_k = A_k [I + \ldots + \alpha_1 P_1]$ can be improved by a good choice of $\alpha_i = 1, \ldots, k-1$ indeed the eigenvalue of $\tilde{A}_k$ are

$$(6.12) \qquad \begin{cases} \lambda_i (1 + \sum_{i=1}^{k-1} \alpha_i) & i = 1, N_1 \\ \lambda_i (1 + \sum_{i=2}^{k-1} \alpha_i) & i = N_{1+1}, \ldots N_2 \\ \vdots & \vdots \\ \lambda_i & i = N_{k-1} + 1, \ldots N_k \end{cases}$$

For example in the case $N_{k-1} = N_k/2$, $\alpha_{k-1} = 1$, $\alpha_{k-i} = 2\ \alpha_{k-i+1}$ the conditioning of $\tilde{A}_k \simeq \frac{1}{2}$ independently of k that is the eigenvalues of $\tilde{A}_k$ are spread on the interval $[-\eta N_k, -\frac{\eta}{2} N_k]$. In general it is difficult to know a good set of operators B_j, C_j. Nevertheless using the continuity of the eigenvalues with respect to the perturbation of the operator it is sufficient to find a set of operator B_j, C_j having approximatively the invariance property needed. And, for example, we can take for B_j a linear interpollation and for $C_j = B_j^t$. This choice is enough precise in the case of discretization of elliptic operators to give to $\tilde{A}_k$ a conditioning < 1, independently of N_k, which leads to an algorithm which solve (6.1) in $\delta N_k \log N_k$ with a precision of $(\frac{1}{N_k})^{\delta}$.

The implementation of the algorithm is a discrete version of

$$
(6.13)\quad \left\{ \begin{array}{l}
\dot{Z}_k = A_k \left[Z_k + \alpha_{k-1} B_{k-1} Z_{k-1} + \alpha_{k-2} B_{k-1} B_{k-2} Z_{k-2} + \ldots \right] + U_k \\
\dot{Z}_{k-1} = C_{k-1} A_k Z_k + A_{k-1} \left[\alpha_{k-1} Z_{k-1} + \alpha_{k-2} B_{k-2} Z_{k-2} + \ldots \right] + C_{k-1} U_k \\
\dot{Z}_{k-2} = C_{k-2} C_{k-1} A_k Z_k + \alpha_{k-1} C_{k-2} A_{k-1} Z_{k-1} + A_{k-2} \left[\alpha_{k-2} Z_{k-2} + \ldots \right] + \\
\qquad + C_{k-2} C_{k-1} U_k \\
\quad \vdots
\end{array} \right.
$$

The properties discussed here are proved in Hackbush [24] for example.

To show the meaning of the equation (6.13) let us consider the particular case with only one level of aggregation and $\alpha_{k-1} = 1$

$$
(6.14)\quad \left\{ \begin{array}{l}
\dot{Z}_k = A_k \left[Z_k + B_{k-1} Z_{k-1} \right] + U_k \\
\dot{Z}_{k-1} = C_{k-1} A_k Z_k + A_{k-1} Z_{k-1} + C_{k-1} U_k
\end{array} \right.
$$

which can be written

$$
(6.15)\quad \left\{ \begin{array}{l}
\dot{Z}_k = A_k \left[Z_k + B_{k-1} Z_{k-1} \right] + U_k \\
\dot{Z}_{k-1} = C_{k-1} A_k B_{k-1} Z_{k-1} + C_{k-1} \left[A_k Z_k + U_k \right]
\end{array} \right.
$$

Z_{k-1} is a correction term obtained by solving an aggregate system, the second equation of (6.15), with a forcing term equal to an aggregate of the error between Z_k and the equilibrium point - $A_k^{-1}U_k$. In fact, this is the initial point of view leading to multigrid methods.

7. CONCLUSION.

We have discussed the ideas of aggregation and approximated aggregation and shown its practical interest in designing fast algorithm for solving linear system of equations. The ideas used in multigrid methods can be certainly extended to the resolution of systems of equations appearing in some network problems.

A N N E X

Proof of theorem 1 (adaptation of Khomyakov [18] one) (4.5) and (4.6) gives

$$V_1^k = M\,P^k\,V_1^k + f \tag{A.1}$$

Then

$$V^{k+1} = M^{m-1}\,(I - MP^k)^{-1}\,f + \sum_{q=o}^{m-2} M^q f \tag{A.2}$$

but

$$V = M^{m-1}\,(I - M)^{-1}\,f + \sum_{q=o}^{m-2} M^q f \tag{A.3}$$

then

$$\delta^{k+1} = M^{m-1}\,[(I - MP^k)^{-1} - (I - M)^{-1}]\,f \tag{A.4}$$

We have also by definition of B^k

$$(I - MP^k)V^k = (I - M)V^k \tag{A.5}$$

and denoting

$$f^k = (I - M)V^k \tag{A.6}$$

we have

$$(I - MP^k)\,(I - M)^{-1}\,f^k = f^k \tag{A.7}$$

which proves that

$$[(I - MP^k)^{-1} - (I - M)^{-1}]\,f^k = o \tag{A.8}$$

from (A.4) and (A.8) we have

$$\delta^{k+1} = M^{m-1} \left[(I - MP^k)^{-1} - (I - M)^{-1}\right] (I - M)\, \delta^k \tag{A.9}$$

$$\delta^{k+1} = M^{m-1} \left[(I - MP^k)^{-1} - (I - MP^k)^{-1} M - I\right] \delta^k \tag{A.10}$$

$$\delta^{k+1} = M^{m-1} \left[MP^k + (MP^k)^2 + \ldots - (I - MP^k)^{-1} M\right] \delta^k \tag{A.11}$$

$$\delta^{k+1} = M^{m-1} (I - MP^k)^{-1} M(P^k - I)\, \delta^k \tag{A.12}$$

From (A.12) and the definition of M we have

$$\|\delta^{k+1}\| \leq \left(\frac{1}{1+\lambda}\right)^m \frac{2}{\lambda} \|\delta^k\| \tag{A.13}$$

which proves the convergence of the iteration when $\lambda > o$ for m large enough.

REFERENCES.

[1] H.A. Simon and A. Ando, "Aggregation of Variables in Dynamic Systems", Econometrica, Vol. 29,pp. 11-138, 1963.

[2] M. Aoki, "Control of Large Scale Dynamic Systems by Aggregation", IEEE Trans. on Automatic Control, Vol. AC-13, pp. 246-253, 1968.

[3] M. Aoki, "Some Approximation Methods for Estimation and Control of Large Scale Systems", IEEE Trans. on Automatic Control, Vol. AC-23, pp. 173-181, 1978.

[4] E.C.Y. Tse, J.V. Medanic, and W.R. Perkins, "Generalized Hessenberg Transformations for Reduced-Order Modeling of Large-Scale Systems", Int. J. Control, Vol. 27, N°4, pp. 493-512, 1978.

[5] J.G. Kemeny and J.L. Snell, "Finite Markov Chains", Van Nostrand, Princeton, N.J., 1967.

[6] V.G. Gaitsgori and A.A. Pervozvanskii, "Aggregation of States in a Markov Chain with Weak Interactions", Kybernetika, pp. 91-98, May-June 1975.

[7] A.A. Pervozvanskii and V.G. Gaitsgori, Decomposition, Aggregation, and Approximate Optimization (in Russian), Nauka, Moscow, 1979.

[8] P.J. Courtois, Decomposability : Queuing and Computer System Applications, Academic Press, New York, 1977.

[9] F. Delebecque and J.P. Quadrat, "Optimal Control of Markov-Chains Admitting Strong and Weak Interactions", Automatica, Vol. 17, N° 2, pp. 281-296, 1981.

[10] R.G. Phillips and P.V. Kokotovic, "A Singular Perturbation Approach to Modeling and Control of Markov Chains", IEEE Trans. on Automatic Control, Vol. AC-20, Oct. 1981.

[11] H.T. Vantilborgh, R.L. Garner, and E.D. Lazowska, "Near-Complete Decomposability of Queueing Networks with Clusters of Strongly Interacting Servers", ACM, pp.81-92, 1980.

[12] G. Strang and J. Fix, "An Analysis of the Finite Element Method", Prentice-Hall, 1973.

[13] B. Avramovic, P.V. Kokotovic, J.R. Winkelman and J.H. Chow "Area Decomposition of Electromechanical Models of Power Systems", Automatica, Vol. 16, pp. 637-648, 1980.

[14] P.V. Kokotovic, B. Avramovic, J.H. Chow and J.R. Winkelman, "Coherency Based Decomposition and Aggregation", Automatica, Vol. 17, N° 6, 1981.

[15] S. Sastry and P. Varaiya, "Coherency for Interconnected Power Systems", IEEE Trans. on Automatic Control, Vol. AC-26, pp. 218-226, 1981.

[16] F.P. Kelly, Reversibility and Stochastic Networks, J. Wiley, 1979.

[17] V.G. Gaitsgori and A.A. Pervozvanskii "Aggregation of Linear Control Systems" Automatika i Telemekhanika, N° 8, pp. 88-95, 1980.

[18] V.A. Khomyakov "Generalization of One Proof of Convergence of a Process of Iterative Aggregation for Solving Systems of Linear Equation", Automatika y Telemekhanika, N° 7, pp. 170-173, July 1973.

[19] P. Bertrand, G. Michailesco, J.M. Siret, "Sur la Synthèse de Modèles Réduits par Agrégation", RAIRO, Vol. 10, N° 7, pp. 105-112, Juillet

[20] C. Commault, "Optimal Choice of Modes of Aggregation", Automatica, Vol. 17, N° 2, pp. 397-399, 1981.

[21] B.I. Vakhutinsky , L.M. Dudkin, A.A. Ryvkin, "Iterative Aggregation - A New Approach to the Solution of Large Scale Problems" Econometrica, Vol. 47, N° 4, pp. 821-841, 1979.

[22] W. Whitt, "Approximations of Dynamic Programs I and II", Math. of Operat. Research, Vol. 3 N° 3 August 78
Vol. 4 N° 2 May 79

[23] A. Brandt, "Multilevel Adaptative Computations in Fluid Dynamics", AIAA Journal, Vol. 18 N° 10, Octobre 1980.

[24] W. Hackbush, "On the Convergence of Multigrid Iterations" Berträge zur Numeischen Mathematik, pp. 213-239, Vol. 9, 1981.

[25] G. Peponides, P.V. Kokotovic, J.H. Chow, "Singular Perturbations and Time Scales in Non Linear Models".

[26] D.W. Stroock, J.R.S. Varadhan, "Diffusion with Continuous Coefficients I and II", Com. on pure APpl. Math. 22, pp. 479-530, (1967).

[27] H. Kushner, "Probability Methods for Approximations in Stochastic Control and for Elliptic Equations", Academic Press 77.

[28] J.P. Quadrat, "Existence de Solution et Algorithme de Résolution Numérique de Problèmes de Contrôle Stochastique Dégénéré ou non", SIAM J. Contr. and Ipt., Vol. 18 N° 2, Mars 1980.

[29] D. Lindner, W.R.Perkins, J.V. Medanic, "Chained aggregation on three control component design : a geometric analysis", Int. J. Control, Vol. 35 N° 4, pp 621-635, (82).

[30] R.A. Nicolaïdes, "On the ℓ^2 convergence of an algorithm for solving finite element equations", Math. Comp. 31, (77), pp 892-906.

[31] F. Delebecque, P.V. Kokotovic, J.P. Quadrat, "Aggregation and coherency in Networks and Markov chains to appear.

[32] A. Isidori, A.J. Krener, C. Gori-Giorgi, S. Monaco, "Non linear decoupling via feedback : a differential geometric approach", IEEE, A-C, Vol. AC 26, N° 2, (Avril 81).

[33] R.P. Fedorenko, "The speed of convergence of one iterative process". Vyčisl. Mat. mat., Fiz. 4, (64), pp 599-564.