# Minimal residual method applied to the transport equation

**S. Akesbi, E. Maitre**

Laboratoire de Mathématiques et Applications
Université de Haute-Alsace,
4, rue des frères Lumière,
68093 Mulhouse Cedex.
e-mail: `S.Akesbi@univ-mulhouse.fr, E.Maitre@univ-mulhouse.fr`

**Summary** During the last years, various algorithms solving the transport equation in slab geometry have been proposed in the litterature [2][3][4][6][7]. Most of them use a convergence acceleration based on a Picard type algorithm called the source iteration method [4][7]. New non accelerated algorithms have been recently introduced [1][2]. We develop here a non accelerated method based on two points : a splitting of the collision operator and an infinite dimensional adaptation of the minimal residual method. We first prove the theoretical convergence of the method in the frame of non-reflecting boundary conditions. Then we compare numerically this method with existing non accelerated schemes. It gives good results which could even be further improved by adding a DSA kind acceleration.

## 1 Introduction and notations

We consider a one dimensional domain $(0, L)$. The evolution neutrons interacting with each other is described by a function $f(x, \mu)$. $f(x, \mu)$ represents the angular neutrons flux at the $x$ position and traveling in the cosine $\mu \in (-1, 1)$ direction. The cross section $\sigma(x)$ accounts for neutrons-domain interaction, whereas a kernel $k(x, \mu, \mu')$ describes the collisions between neutrons. The function $S(x, \mu)$ represents a neutrons source. We refer to [5] for a more precise introduction.

The function $f$ verifies a linear integro-differential equation. To solve this equation we first prove in this paper the convergence of an infinite dimensional version of minimal residual method. Numerical comparison with non accelerated algorithms developed in [1][2] is given at the end of this work.

### *1.1 Mathematical setting*

Let consider $L > 0$ and $\Omega = (0, L) \times (-1, 1)$. We consider the following problem : given a source term $S \in L^2_+(\Omega)$, find $f : \Omega \to \mathbf{R}$ solution of the transport equation,

$$(P) \quad \begin{cases} Tf(x,\mu) = Kf(x,\mu) + S(x,\mu) & \text{in } \Omega, \\ f(0,\mu) = 0 & \text{for } \mu \in I_1 := (0,1), \\ f(L,\mu) = 0 & \text{for } \mu \in I_2 := (-1,0), \end{cases}$$

where $Tf(x,\mu) = \mu \dfrac{\partial f}{\partial x}(x,\mu) + \sigma(x) f(x,\mu)$ with

$$\mathcal{D}(T) = \left\{ f \in L^2(\Omega) \ : \ \mu \frac{\partial f}{\partial x}(x,\mu) \in L^2(\Omega), \ f(0,\mu) = f(L,-\mu) = 0 \text{ for } \mu > 0 \right\},$$

and $K$ an integral operator of *positive* kernel $k$ :

$$Kf(x,\mu) = \int_{-1}^{1} k(x,\mu,\mu')f(x,\mu')d\mu'.$$

We make the following **assumptions** (where $L^1_+$ denotes the positive cone of $L^1$) :

(A1) $\sigma \in L^1_+(0,L)$.
(A2) $(\mu,\mu') \to k(\mu,\mu') \in L^2_+((-1,1)^2)$.
(A3) $k$ is symmetric and even. If $\sigma$ is a constant (see remark 1), we assume :

$$\exists c < 1,\ \left(\int_{-1}^{1}\int_{-1}^{1} k(\mu,\mu')^2 d\mu d\mu'\right)^{\frac{1}{2}} \le \sigma c,$$

else we assume :

$$\exists c < 1,\ \left(\int_{I_i} \alpha_i(\mu)^2 \left[\int_{I_j} k(\mu,\mu')^2 d\mu'\right] d\mu\right)^{\frac{1}{2}} \le \frac{c}{2},\ \text{for } (i,j) \in \{1,2\}^2, \text{ where}$$

$$\alpha_1(\mu) = \frac{2}{|\mu|} \sup_{x'\in[0,L]} \int_{x'}^{L} e^{-\frac{2}{|\mu|}\int_{x'}^{x}\sigma(y)dy} dx, \quad \alpha_2(\mu) = \frac{2}{|\mu|} \sup_{x'\in[0,L]} \int_{0}^{x'} e^{-\frac{2}{|\mu|}\int_{x}^{x'}\sigma(y)dy} dx.$$

(A4) $k(\mu,\mu') = \sum_{l=1}^{N_k} a_l(\mu)a_l(\mu')$.

*Remark 1* 1. Note that the (A3) assumption may allow $\sigma$ to vanish in $(0,L)$.
2. All forthcoming results remain valid if $k$ depends on $x$ in the following way : $k(x,\mu,\mu') = C(x)k(\mu,\mu')$ with a positive measurable and bounded function $C$. The expression $\frac{c}{2\|C\|_\infty}$ would then appear in the right hand side of (A3).
3. In case of $\sigma$ constant on $[0,L]$, $\alpha_i(\mu) = \dfrac{1}{\sigma}$. This justifies the two proposed forms of (A3). All proofs are made in the general case.
4. The (A4) assumption is not necessary for a theoretical proof of the convergence. However, for a numerical point, the splitting of the collision operator could be exploited only if we assume this expressed form of the kernel $k$.
5. These assumptions (including (A4)) are satisfied for usual neutronic kernels as the constant and Thomson kernels for instance.

### *1.2 Classical and splitting methods*

The standard method for solving $(P)$, and called the source iteration method, is based on a decoupling between the differential and integral parts, through the following iterative scheme : given $f^0 \in \mathcal{D}(T)$, solve

$$(P_s) \quad \begin{cases} Tf^{n+1} = Kf^n + S & \text{in } \Omega, \\ f^{n+1} \in \mathcal{D}(T). \end{cases}$$

In the critical case ($c \approx 1$), this algorithm becomes extremely slow. Diffusion synthetic methods developed in [7][6] and [4] give excellent results. However the are not easy to implement when the kernel strongly depends on $\mu$. Moreover, theoretical results for DSA are achieved on discretized schemes in $\mu$, for reflecting boundary conditions or infinite domain (using Fourier transform).

In contrast, our convergence result is independent of the discretization chosen for the transport equation. It is also proved to be convergent for non-reflecting boundary conditions.

Our method is based on a new algorithm converging faster than $(P_s)$, on which an adapted DSA acceleration by sub-domain could be applicated ([3]). Let us describe the natural splitting of $K$ which leads to this scheme. Let $k_{ij}$, $i,j \in \{1,2\}$ be the positive kernel defined by

$$k_{ij}(x,\mu,\mu') = k(\mu,\mu') \times \mathbf{1}_{\Omega_i}(x,\mu) \times \mathbf{1}_{\Omega_j}(x,\mu'),$$

with $\Omega_1 = (0,L) \times (0,1)$, $\Omega_2 = (0,L) \times (-1,0)$, and $\mathbf{1}_{\Omega_i}$ the indicator function of $\Omega_i$. We introduce the associated integral operator $K_{i,j}$ :

$$K_{ij}(f)(x,\mu) = \int_{-1}^{1} k_{ij}(x,\mu,\mu') f(x,\mu') d\mu'.$$

Since we have $K_{ij}(f) = K_{ij}(f.\mathbf{1}_{\Omega_j})\mathbf{1}_{\Omega_i}$, the operator $K$ splits into $K = K_{11}+K_{12}+K_{21}+K_{22}$. Note that $K_{ij}$ is an operator acting from $L^2(\Omega)$, using only the values of $f$ on $\Omega_j$, such that $K_{ij}f$ has its support in $\Omega_i$.

The solution of $(P)$ is given by $f = f_1 + f_2$ with $f_1, f_2 \in \mathcal{D}(T)$ solution of

$$\begin{pmatrix} T - K_{11} & -K_{12} \\ -K_{21} & T - K_{22} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}. \tag{1}$$

It is easy to prove that $f_i = f\mathbf{1}_{\Omega_i}$, $i = 1,2$. Various methods as Jacobi, Gauss-Seidel and SOR, applied to this matrix of operators, were studied in [3]. The SOR method gives excellent results, but it needs the computation of its optimal parameter, which in turn can be very slow in the critical case. For these reasons we seeked a method that gives good rate of convergence, but do not need any extra parameter calculation.

*Remark 2* Once discretized in $\mu$, the Jacobi method we mentioned corresponds to the so called modified source iteration method MSI [6].

## 2 Minimal residual algorithm

In the following part, we aim at solving the transport equation $(P)$ by a minimal residual method. This method was introduced by O. Axelsson [8], in the finite dimensional case, and proved to converge provided the matrix of the linear system has a definite positive symmetric part.
Using the operator splitting devised by S. Akesbi and M. Nicolet, the transport equation is equivalent to the following system

$$\begin{pmatrix} I - \theta_{11} & -\theta_{12} \\ -\theta_{21} & I - \theta_{22} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} \widetilde{S_1} \\ \widetilde{S_2} \end{pmatrix}, \tag{2}$$

where we applied on both components the operator $T^{-1}$, and set $\theta_{ij} = T^{-1}K_{ij}$, $\widetilde{S_i} = T^{-1}S_i$. Our system matrix of operators will be preconditioned by the inverse of its diagonal i.e.

$$\begin{pmatrix} (I - \theta_{11})^{-1} & 0 \\ 0 & (I - \theta_{22})^{-1} \end{pmatrix},$$

leading to the following system

$$\begin{pmatrix} I & -(I - \theta_{11})^{-1}\theta_{12} \\ -(I - \theta_{22})^{-1}\theta_{21} & I \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} (I - \theta_{11})^{-1}\widetilde{S_1} \\ (I - \theta_{22})^{-1}\widetilde{S_2} \end{pmatrix}. \tag{3}$$

We denote by $\mathcal{A}$ the operator matrix of this system, and

$$F = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \in \mathcal{D}(T) \times \mathcal{D}(T) \qquad B = \begin{pmatrix} (I - \theta_{11})^{-1}\widetilde{S_1} \\ (I - \theta_{22})^{-1}\widetilde{S_2} \end{pmatrix}$$

Let us introduce the minimal residual method to solve $\mathcal{A}F = B$. We denote by $\langle , \rangle$ the scalar product in $L^2(\Omega) \times L^2(\Omega)$, i.e. $\langle F, G\rangle = (f_1, g_1) + (f_2, g_2)$ where $( , )$ is the standard $L^2$ scalar product. Similarly, $\| \|_2$ will represent the norm in $L^2(\Omega) \times L^2(\Omega)$ associated to this scalar product.
The minimal residual method minimizes $\mathcal{E}(F) = \|B - \mathcal{A}F\|_2^2$ by the following algorithm :
Let $f^0 \in \mathcal{D}(T)$, $F^0 = {}^t(f^0\mathbf{1}_{\Omega_1}, f^0\mathbf{1}_{\Omega_2}) \in \mathcal{D}(T)\times\mathcal{D}(T)$, $R^0 = B - \mathcal{A}F^0$, $P^0 = R^0$, $Q^0 = \mathcal{A}P^0$.

$$\text{Compute for } k = 0, 1, \ldots \text{ until } \left\|R^k\right\|_2 < \varepsilon, \qquad \begin{cases} \alpha^k = \frac{\langle R^k, Q^k\rangle}{\langle Q^k, Q^k\rangle} \\ F^{k+1} = F^k + \alpha^k P^k \\ R^{k+1} = R^k - \alpha^k Q^k \\ \beta^{k+1} = -\frac{\langle \mathcal{A}R^{k+1}, Q^k\rangle}{\langle Q^k, Q^k\rangle} \\ P^{k+1} = R^{k+1} + \beta^{k+1} P^k \\ Q^{k+1} = \mathcal{A}R^{k+1} + \beta^{k+1} Q^k \end{cases}$$

where $\varepsilon > 0$ is prescribed.

In the previous algorithm, we have to explicit how we compute the product $\mathcal{A}$ times a vector, since $\mathcal{A}$ contains some inverse operator.
Let $g \in \mathcal{D}(T)$, $g_i = g\mathbf{1}_{\Omega_i}$, then $(g_1, g_2) \in \mathcal{D}(\mathcal{T}) \times \mathcal{D}(\mathcal{T})$. We will describe how to compute

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} I & -(I - \theta_{11})^{-1}\theta_{12} \\ -(I - \theta_{22})^{-1}\theta_{21} & I \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$

Componentwise, this equality means

$$z_1 = g_1 - (I - \theta_{11})^{-1}\theta_{12} g_2,$$
$$z_2 = g_2 - (I - \theta_{22})^{-1}\theta_{21} g_1.$$

Applying $T(I - \theta_{11}) = T - K_{11}$ to the first equation and $T(I - \theta_{22}) = T - K_{22}$ to the second one, we get

$$(T - K_{11})(g_1 - z_1) = K_{12} g_2,$$
$$(T - K_{22})(g_2 - z_2) = K_{21} g_1.$$

The splitting of $K$ allows us to solve these equations numerically and directly (see S.Akesbi and M. Nicolet [3]). We can note from the $K_{ij}$ and $T$ definitions, that these solutions are verifying $z_1 = 0$ on $\Omega_2$ and $z_2 = 0$ on $\Omega_1$. With the same formalism, the equation $R^0 = B - \mathcal{A}F^0$ corresponds to

$$(T - K_{11})(r_1^0 + f_1^0) = S_1 + K_{12} f_2^0,$$
$$(T - K_{22})(r_2^0 + f_2^0) = S_2 + K_{21} f_1^0,$$

whereas $Q^0 = \mathcal{A}P^0$ stands for

$$(T - K_{11})(p_1^0 - q_1^0) = K_{12} p_2^0,$$
$$(T - K_{22})(p_2^0 - q_2^0) = K_{21} p_1^0.$$

At last, the product $\mathcal{A}R^{k+1} =: D^{k+1}$ which of course is calculated only once per iteration, is associated to the following equations :

$$(T - K_{11})(r_1^{k+1} - d_1^{k+1}) = K_{12} r_2^{k+1},$$
$$(T - K_{22})(r_2^{k+1} - d_2^{k+1}) = K_{21} r_1^{k+1}.$$

### *2.1 Rate of residual decreasing*

We apply the elementary analysis of [8], and we get the following estimate on the residual :

**Proposition 1** *Let $F^k$ be constructed by the preceding algorithm, starting from $F^0$. Then for $k \geq 0$,*

$$\mathcal{E}(F^{k+1}) \leq \mathcal{E}(F^k)\left(1 - \frac{\langle R^k, \mathcal{A}R^k\rangle}{\langle R^k, R^k\rangle}\frac{\langle R^k, \mathcal{A}R^k\rangle}{\langle \mathcal{A}R^k, \mathcal{A}R^k\rangle}\right). \tag{4}$$

*Proof* Observe first that $\mathcal{E}(F^0) = \left\|R^0\right\|_2^2$ and

$$\mathcal{E}(F^{k+1}) = \mathcal{E}(F^k + \alpha^k P^k) = \left\|R^k - \alpha^k \mathcal{A}P^k\right\|_2^2 = \left\|R^k - \alpha^k Q^k\right\|_2^2 = \left\|R^{k+1}\right\|_2^2.$$

Thus

$$\mathcal{E}(F^{k+1}) = \left\|R^k\right\|_2^2 + {\alpha^k}^2\left\|Q^k\right\|_2^2 - 2\alpha^k\left\langle R^k, Q^k\right\rangle = \mathcal{E}(F^k) - \frac{\langle R^k, Q^k\rangle^2}{\langle Q^k, Q^k\rangle},$$

owing to the definition of $\alpha^k$. Therefore

$$\mathcal{E}(F^{k+1}) = \mathcal{E}(F^k)\left(1 - \frac{\langle R^k, Q^k\rangle^2}{\langle R^k, R^k\rangle\langle Q^k, Q^k\rangle}\right).$$

On the first hand, using the $\alpha^k$ definition we have

$$\left\langle Q^{k+1}, R^{k+1}\right\rangle = \left\langle \mathcal{A}R^{k+1}, R^{k+1}\right\rangle + \beta^{k+1}\left\langle Q^k, R^{k+1}\right\rangle = \left\langle \mathcal{A}R^{k+1}, R^{k+1}\right\rangle.$$

On the other hand,

$$\left\langle Q^{k+1}, Q^{k+1}\right\rangle = \left\langle \mathcal{A}R^{k+1}, Q^{k+1}\right\rangle + \beta^{k+1}\left\langle Q^k, Q^{k+1}\right\rangle = \left\langle \mathcal{A}R^{k+1}, Q^{k+1}\right\rangle,$$

using the $\beta^{k+1}$ definition. It follows that

$$\begin{aligned}\left\langle Q^{k+1}, Q^{k+1}\right\rangle &= \left\langle \mathcal{A}R^{k+1}, \mathcal{A}R^{k+1}\right\rangle + \beta^{k+1}\left\langle \mathcal{A}R^{k+1}, Q^k\right\rangle \\ &= \left\langle \mathcal{A}R^{k+1}, \mathcal{A}R^{k+1}\right\rangle - (\beta^{k+1})^2\left\langle Q^k, Q^k\right\rangle^2 \\ &\leq \left\langle \mathcal{A}R^{k+1}, \mathcal{A}R^{k+1}\right\rangle.\end{aligned}$$

The announced inequality is achieved.

### *2.2 Theoretical convergence*

First of all, let us prove that our operator $\mathcal{A}$ has a definite positive symetric part, so that for some $\lambda > 0$,

$$\frac{\langle R^k, \mathcal{A}R^k\rangle}{\langle R^k, R^k\rangle} \geq \lambda.$$

**Proposition 2** *Under assumption (A1)-(A3) the operator $\mathcal{A}$ has a definite positive symmetric part and verifies*

$$\langle \mathcal{A}F, F\rangle \geq \frac{1-c}{1-\frac{c}{2}}\|F\|_2^2, \qquad \forall F \in \mathcal{D}(T)\times\mathcal{D}(T). \tag{5}$$

*Proof* We have

$$\langle \mathcal{A}F, F\rangle = \|F_1\|_2^2 + \|F_2\|_2^2 - (F_1, (I-\theta_{11})^{-1}\theta_{12}F_2) - ((I-\theta_{22})^{-1}\theta_{21}F_1, F_2).$$

By definition,

$$\begin{aligned} \|(I-\theta_{11})^{-1}\theta_{12}\|_2 &= \sup_{\|f\|_2=1} \|(I-\theta_{11})^{-1}\theta_{12}f\|_2 \\ &= \sup_{\|f\|_2=1} \left\| \left(\sum_{k=0}^{\infty} \theta_{11}^k\right) \theta_{12} f \right\|_2 \\ &\le \left(\sum_{k=0}^{\infty} \|\theta_{11}\|_2^k\right) \|\theta_{12}\|_2 \\ &\le \frac{\|\theta_{12}\|_2}{1-\|\theta_{11}\|_2}. \end{aligned}$$

We still need to control the norms of $\theta_{11}$ and $\theta_{12}$. So, let us state the following

**Lemma 1** *Under assumptions (A1)-(A3),* $\|\theta_{ij}\|_2 \le \frac{c}{2}$, *for* $(i,j) \in \{1,2\}^2$.

Postponing the proof of this lemma, we get

$$\|(I-\theta_{11})^{-1}\theta_{12}\|_2 \le \frac{\frac{c}{2}}{1-\frac{c}{2}} = \frac{c}{2-c}. \tag{6}$$

Thus

$$\begin{aligned} \langle \mathcal{A}F, F\rangle &\ge \|F_1\|_2^2 + \|F_2\|_2^2 - \frac{2c}{2-c}\|F_1\|_2\|F_2\|_2 \\ &= (1-\frac{c}{2-c})(\|F_1\|_2^2 + \|F_2\|_2^2) + \frac{c}{2-c}(\|F_1\|_2^2 + \|F_2\|_2^2 - 2\|F_1\|_2\|F_2\|_2) \\ &\ge (1-\frac{c}{2-c})(\|F_1\|_2^2 + \|F_2\|_2^2). \end{aligned}$$

This is the expression (5).

*Proof ( of lemma 1.)* Since $\theta_{11}$ is an operator from domain 1 into itself, we can consider that the function $f$ vanishes identically on $\Omega_2$. By definition, $\theta_{11}f$ is the solution $g$ of

$$\mu\frac{\partial g}{\partial x}(x,\mu) + \sigma(x)g(x,\mu) = (K_{11}f)(x,\mu) = \int_0^1 k(\mu,\mu')f(x,\mu')d\mu'$$

with boundary condition $g(0,\mu)=0$. We multiply this equation by $g$ to obtain :

$$\frac{\mu}{2}\frac{\partial g^2}{\partial x}(x,\mu) + \sigma(x)g^2(x,\mu) = g(x,\mu)\int_0^1 k(\mu,\mu')f(x,\mu')d\mu'.$$

If $\Sigma$ is a primitive of $\sigma$,

$$\frac{\partial}{\partial x}\left[e^{\frac{2\Sigma(x)}{\mu}} g^2(x,\mu)\right] = \frac{2}{\mu}e^{\frac{2\Sigma(x)}{\mu}} g(x,\mu)\int_0^1 k(\mu,\mu')f(x,\mu')d\mu'.$$

Integrating from 0 to $x$ and using $g(0,\mu)=0$ gives

$$\begin{aligned} g^2(x,\mu) &= \frac{2}{\mu}e^{-\frac{2\Sigma(x)}{\mu}} \int_0^x e^{\frac{2\Sigma(x')}{\mu}} g(x',\mu)\left[\int_0^1 k(\mu,\mu')f(x',\mu')d\mu'\right]dx' \\ &= \frac{2}{\mu}\int_0^x e^{-\frac{2}{\mu}\int_{x'}^x \sigma(y)dy} g(x',\mu)\left[\int_0^1 k(\mu,\mu')f(x',\mu')d\mu'\right]dx'. \end{aligned}$$

We integrate on $x$ from 0 to $L$ and on $\mu$ from 0 to 1 to get the $L^2$ norm of $g$,

$$\|g\|_2^2 = \int_0^1 \int_0^L \frac{2}{\mu} \int_0^x e^{-\frac{2}{\mu}\int_{x'}^x \sigma(y)dy} g(x',\mu) \left[\int_0^1 k(\mu,\mu')f(x',\mu')d\mu'\right] dx'dxd\mu.$$

Using Fubini's theorem,

$$\begin{aligned}
\|g\|_2^2 &= \int_0^1 \int_0^L \frac{2}{\mu} \int_{x'}^L e^{-\frac{2}{\mu}\int_{x'}^x \sigma(y)dy} g(x',\mu) \left[\int_0^1 k(\mu,\mu')f(x',\mu')d\mu'\right] dxdx'd\mu \\
&= \int_0^1 \frac{2}{\mu} \int_0^L \left[\int_{x'}^L e^{-\frac{2}{\mu}\int_{x'}^x \sigma(y)dy} dx\right] g(x',\mu) \left[\int_0^1 k(\mu,\mu')f(x',\mu')d\mu'\right] dx'd\mu \\
&\leq \int_0^1 \alpha_1(\mu) \int_0^1 k(\mu,\mu') \int_0^L |f(x',\mu')g(x',\mu)|dx'd\mu'd\mu \\
&\leq \left(\int_0^1 \alpha_1(\mu)^2 \int_0^1 k(\mu,\mu')^2 d\mu' d\mu\right)^{\frac{1}{2}} \left[\int_0^1 \int_0^1 \left(\int_0^L |f(x',\mu')g(x',\mu)|dx'\right)^2 d\mu d\mu'\right]^{\frac{1}{2}} \\
&\leq \frac{c}{2}\left[\int_0^1 \int_0^1 \left(\int_0^L f(x',\mu')^2 dx'\right)\left(\int_0^L g(x',\mu)^2 dx'\right) d\mu d\mu'\right]^{\frac{1}{2}} \\
&\leq \frac{c}{2}\|f\|_2\|g\|_2,
\end{aligned}$$

using the (A3) assumption. This gives the bound on $\|\theta_{11}\|_2$. For $\theta_{12}$, the proof is identical except that we are dealing with functions $f$ defined on $\Omega_2$ and vanishing on $\Omega_1$, whereas $g$ is computed on $\Omega_1$. All integrals on $\mu'$ are now from $-1$ to 0, and we end up with

$$\|g\|_2^2 \leq \left(\int_0^1 \alpha_1(\mu)^2 \int_{-1}^0 k(\mu,\mu')^2 d\mu' d\mu\right)^{\frac{1}{2}} \left[\int_{-1}^0 \int_0^1 \left(\int_0^L |f(x',\mu')g(x',\mu)|dx'\right)^2 d\mu d\mu'\right]^{\frac{1}{2}}$$

which has the same bound.
For $\theta_{22}$, we consider the solution $g$ on $\Omega_2$ of

$$\mu\frac{\partial g}{\partial x}(x,\mu) + \sigma g(x,\mu) = K_{22}f = \int_{-1}^0 k(\mu,\mu')f(x,\mu')d\mu'$$

with boundary condition $g(L,\mu) = 0$. We multiplying this equation by $g$ and we use for $\Sigma$ a primitive of $-\sigma$ vanishing in $x = L$,

$$-\frac{\partial}{\partial x}\left[e^{\frac{2\Sigma(x)}{|\mu|}} g^2(x,\mu)\right] = \frac{2}{|\mu|} e^{\frac{2\Sigma(x)}{|\mu|}} g(x,\mu) \int_{-1}^0 k(\mu,\mu')f(x,\mu')d\mu'.$$

We integrating from $x$ to $L$, and we get since $g(L,\mu) = 0$,

$$g^2(x,\mu) = \frac{2}{|\mu|}\int_x^L e^{-\frac{2}{|\mu|}\int_x^{x'} \sigma(y)dy} g(x',\mu) \left[\int_{-1}^0 k(\mu,\mu')f(x',\mu')d\mu'\right] dx'$$

thus using Fubini theorem

$$\|g\|_2^2 = \int_{-1}^0 \int_0^L \left[\frac{2}{|\mu|}\int_0^{x'} e^{-\frac{2}{|\mu|}\int_x^{x'} \sigma(y)dy} dx\right] g(x',\mu) \left[\int_{-1}^0 k(\mu,\mu')f(x',\mu')d\mu'\right] dx'd\mu$$

so that $\alpha_2$ appears in place of $\alpha_1$. The ending of the proof is straightforward.

Now we turn to the second expression appearing in (4), to prove that for some $\nu > 0$ we have

$$\frac{\langle R^k, \mathcal{A}R^k\rangle}{\langle \mathcal{A}R^k, \mathcal{A}R^k\rangle} \geq \nu.$$

We can give an explicit value for $\nu$ :

**Proposition 3** *Under assumptions (A1)-(A3), the matrix of operators $\mathcal{A}$ verifies*

$$\langle \mathcal{A}F, F\rangle \geq \frac{1-\frac{c}{2}}{1+\left(\frac{c}{2}\right)^2}\langle \mathcal{A}F, \mathcal{A}F\rangle, \qquad \forall F \in \mathcal{D}(T)\times\mathcal{D}(T). \tag{7}$$

*Proof* Let $J_1 = (I-\theta_{22})^{-1}\theta_{21}$ and $J_2 = (I-\theta_{11})^{-1}\theta_{12}$, we write :

$$\langle \mathcal{A}F, \mathcal{A}F\rangle = \|F_1\|_2^2 + \|F_2\|_2^2 + \|J_1F_1\|_2^2 + \|J_2F_2\|_2^2 - 2(J_2F_2, F_1) - 2(J_1F_1, F_2)$$

and

$$\langle \mathcal{A}F, F\rangle = \|F_1\|_2^2 + \|F_2\|_2^2 - (J_2F_2, F_1) - (J_1F_1, F_2).$$

Let $\nu > 0$ and consider

$$\begin{aligned}\langle \mathcal{A}F, F\rangle - \nu\langle \mathcal{A}F, \mathcal{A}F\rangle &= (1-\nu)(\|F_1\|_2^2 + \|F_2\|_2^2) - \nu(\|J_1F_1\|_2^2 + \|J_2F_2\|_2^2)\\ &\quad + (2\nu-1)\left[(J_2F_2, F_1) + (J_1F_1, F_2)\right]\\ &\geq \left(1-\nu-2\nu\left(\frac{c}{2-c}\right)^2\right)(\|F_1\|_2^2 + \|F_2\|_2^2) - (2\nu-1)\frac{2c}{2-c}\|F_1\|_2^2\|F_2\|_2^2,\end{aligned}$$

using (6). Thus

$$\langle \mathcal{A}F, F\rangle - \nu\langle \mathcal{A}F, \mathcal{A}F\rangle \geq \left(1-\nu-2\nu\left(\frac{c}{2-c}\right)^2 - (2\nu-1)\frac{c}{2-c}\right)(\|F_1\|_2^2 + \|F_2\|_2^2).$$

We consider $\nu$ such that $1-\nu-2\nu\left(\frac{c}{2-c}\right)^2 - (2\nu-1)\frac{c}{2-c} = 0$ that is $\nu = \frac{1-\frac{c}{2}}{1+\left(\frac{c}{2}\right)^2}$.

We can now state the convergence result.

**Theorem 1** *Under assumptions (A1)-(A3), the minimal residual method converges, i.e. $F^k$ converges toward the unique solution of (3), and the residual is decreasing at the following rate :*

$$\mathcal{E}(F^{k+1}) \leq \mathcal{E}(F^k)\left(1 - \frac{1-c}{1+\left(\frac{c}{2}\right)^2}\right) \tag{8}$$

*Proof* We plug estimations (5) and (7) into (4), to get (8). As $c < 1$, this means that $\mathcal{E}(F^{k+1})$ converges toward 0 when $k$ goes to infinity. We use (5) to get

$$\|F^{k+1} - \mathcal{A}^{-1}B\|_2^2 \leq \frac{2-c}{2(1-c)}\left\langle \mathcal{A}F^{k+1} - B, F^{k+1} - \mathcal{A}^{-1}B\right\rangle$$

so that

$$\|F^{k+1} - \mathcal{A}^{-1}B\|_2 \leq \frac{2-c}{2(1-c)}\mathcal{E}(F^{k+1})^{\frac{1}{2}}$$

which means that $F^{k+1} \to F$ with $\mathcal{A}F = B$.

*Remark 3* When $c$ is equal to zero, the algorithm converges in one step. This is the trivial case $\mathcal{A} = Id$.

## 3 Discretization and numerical results

We only present the resolution of the following problem :

$$\begin{cases}(T-K_{ll})f_l = g,\\ f_l \in \mathcal{D}(T),\end{cases}$$

where $g$ has its support included into $\Omega_l$, and $l \in \{1,2\}$. Indeed we showed that our algorithm only requires the resolution of such equations. Without loss of generality we can consider the case where $l = 1$. The natural space associated to this problem is

$$V = \left\{ f \in L^2(\Omega_1), \quad \mu\frac{\partial f}{\partial x}(x,\mu) \in L^2(\Omega_1) \text{ et } f(0,\mu) = 0 \quad \forall \mu \in (0,1)\right\}.$$

*3.1 Discretization*

For integers $M$ and $N$, we introduce the following mesh on $\Omega_1$ :

$$\overline{\Omega_1} = \bigcup_{i,j} \omega_{ij} \text{ with } \omega_{ij} = (x_i, x_{i+1}) \times (\mu_j, \mu_{j+1}) \text{ for } i = 0, \dots, N-1 \text{ and } j = 0, \dots, M-1$$

where $x_i = i \times \rho$, $\mu_j = j \times \tau$, $\rho = \dfrac{L}{N}$ and $\tau = \dfrac{1}{M}$.
We define $V_h$ as the space of continuous in $x$ functions whose restriction to a given $\omega_{ij}$ is affine in $x$ and constant with respect to $\mu$. We introduce the following projection operator defined for $f \in L^2(\Omega_1)$ by

$$\pi_h(f)_{|\omega_{ij}} = \frac{1}{|\omega_{ij}|} \int_{\omega_{ij}} f(x,\mu)\,dx\,d\mu,$$

our approximated problem is given by

$$(P_h) \quad \begin{cases} \text{Find } f_h \in V_h \text{ such that} \\ A_h f_h = \pi_h(f_h) \end{cases}$$

where

$$A_h f(x,\mu) = \pi_h\left(\mu \frac{\partial f}{\partial x}(x,\mu)\right) + \pi_h(\sigma(x))\pi_h(f(x,\mu)) - \\ \pi_h(C(x)) \sum_{l=1}^{N_k} \pi_h(\alpha_l(\mu)) \int_0^1 \pi_h(\alpha_l(\mu'))\pi_h(f(x,\mu'))d\mu'.$$

Here we assumed the form (A4) for $k$, which is essential for this discrete equation to be solved by an explicit scheme. We refer to [2] and [3] in which the scheme is developed and studied.

*3.2 Numerical results*

We present our numerical results compared to standard, Gauss-Seidel and SOR methods of [3]. We took particular data for which an exact solution $f$ is known :

$$S(x,\mu) = \begin{cases} \mu^2 + \sigma\mu x - \sigma c/4 & \text{if } \mu > 0, \\ \mu^2 + \sigma\mu(x-1) - \sigma c/4 & \text{if } \mu < 0, \end{cases} \qquad f(x,\mu) = \begin{cases} \mu x & \text{if } \mu > 0, \\ \mu(x-1) & \text{if } \mu < 0, \end{cases}$$

and we compared the speed of the various algorithms to reach the exact solution up to $10^{-6}$. As the SOR gives far better result than standard or Gauss-Seidel methods, we first compare the minimal residual method with these two, and then turn to the comparison with SOR. For these two cases, we study the behavior with respect to $c$ and to $\sigma$. On each graph, we report the cpu time to compute the whole test (i.e. to obtain all points of the curve).

One striking property of our scheme, as shown in figure 1, is that it does not blow up in iterations near the critical case ($c \approx 1$). Due to this fact, its computation time is far smaller than the two other schemes.This problem does not occur when $\sigma$ varies at fixed $c$, as all schemes are stable. Nevertheless, our schemes gives better results (figure 2). Turning now to a more serious opponent, the SOR scheme (figure 3), we see that this scheme is still blowing up near the critical case, whereas as already seen above, the minimal residual algorithm remains stable (numerically we can even cross the value $c = 1$), with less computation time. Inspecting now the behavior with respect to $\sigma$ (figure 4), one could think that minimal residual method is not so good for large $\sigma$. But a test for really large values of $\sigma$ reveals that our scheme converges more and more rapidly as $\sigma$ increase, whereas SOR keeps a constant number of iterations (see figure 5 in decimal log scale).

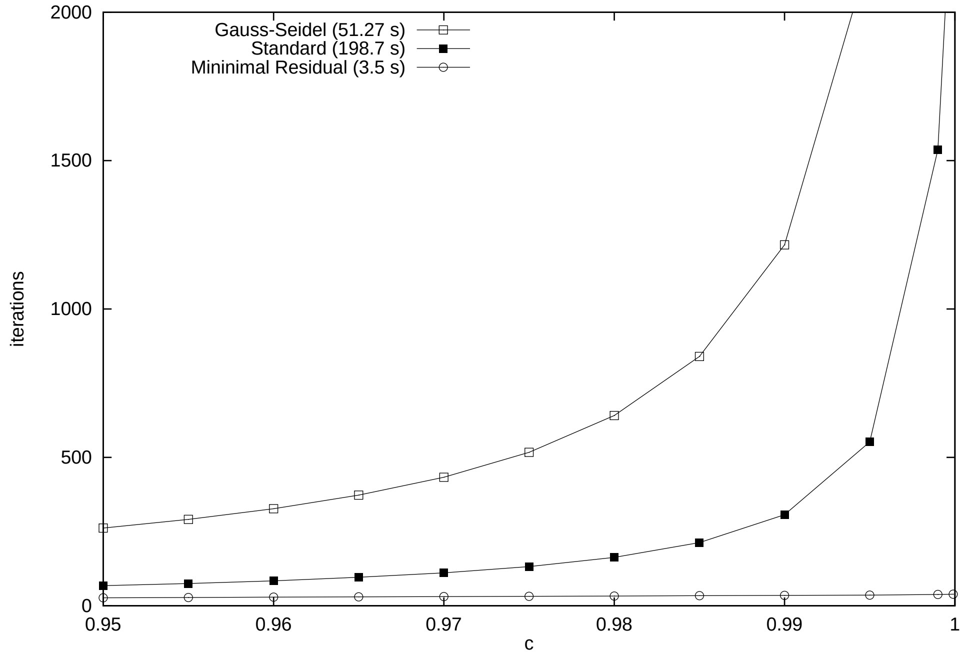


**Fig. 1.** Comparison at fixed $\sigma = 50$, with total computation times

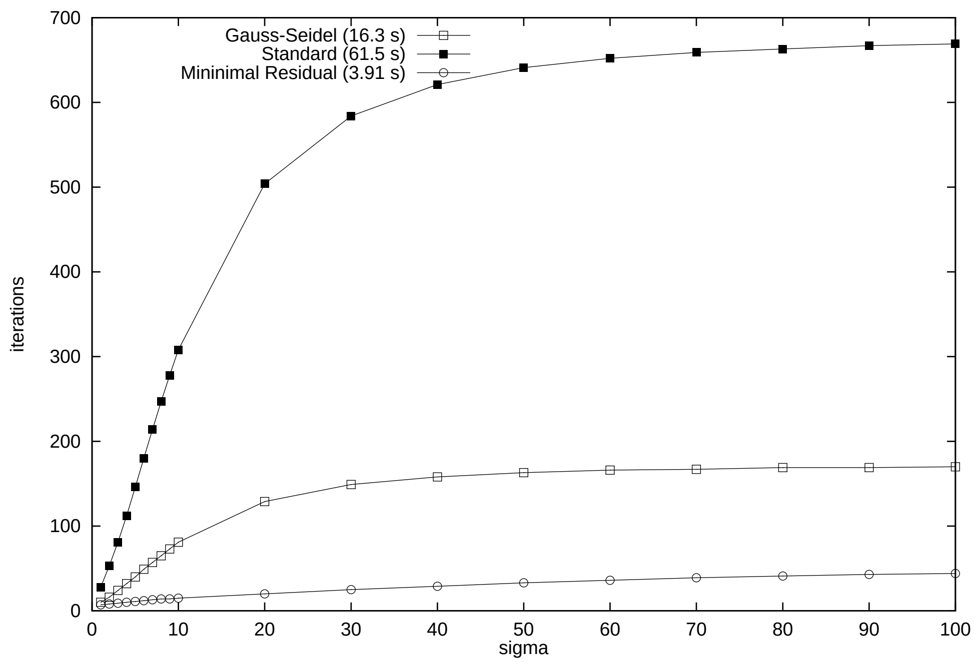


**Fig. 2.** Comparison at fixed $c = 0.98$, with total computation times

## 4 Conclusion

It appears clearly from our numerical results that this new algorithm is very efficient compared to existing schemes studied in [3]. Even if the theoretical limit value of $c$ for convergence is 1, numerically our algorithm works even for values greater than one (as DSA do). Moreover it is easy to implement in dimension one as well as in dimension two (work in progress). Another important point is the existence of a theoretical proof of convergence, independent of discretization, made in the general case for the kernel and with the $\sigma$ parameter which can vanish on some points of the domain. We believe that this proof can be adapted to dimension two, which was not the case for the proofs of convergence

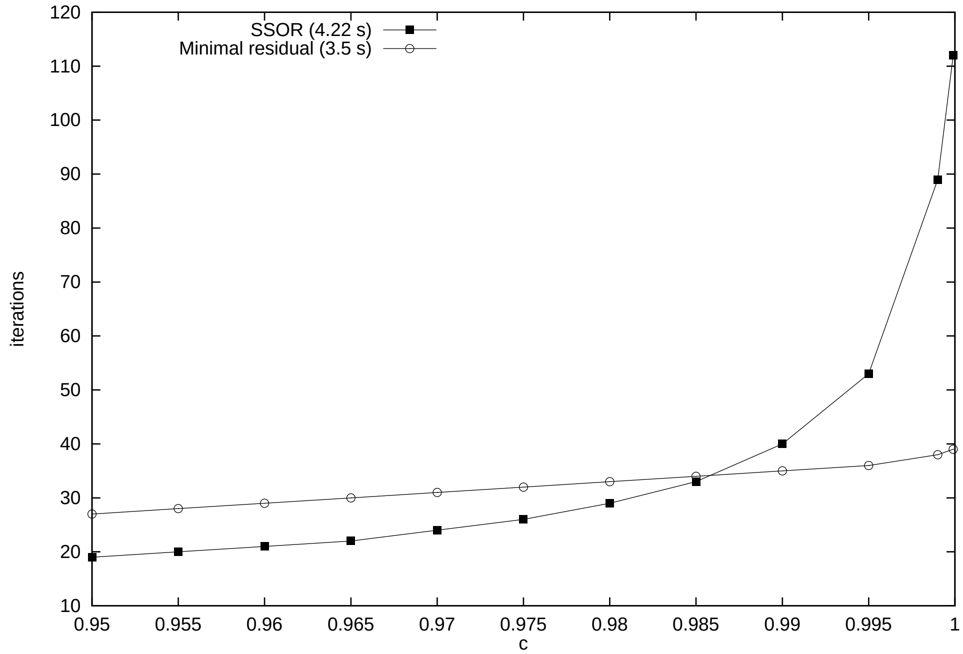


**Fig. 3.** Comparison at fixed $\sigma = 50$, with total computation times

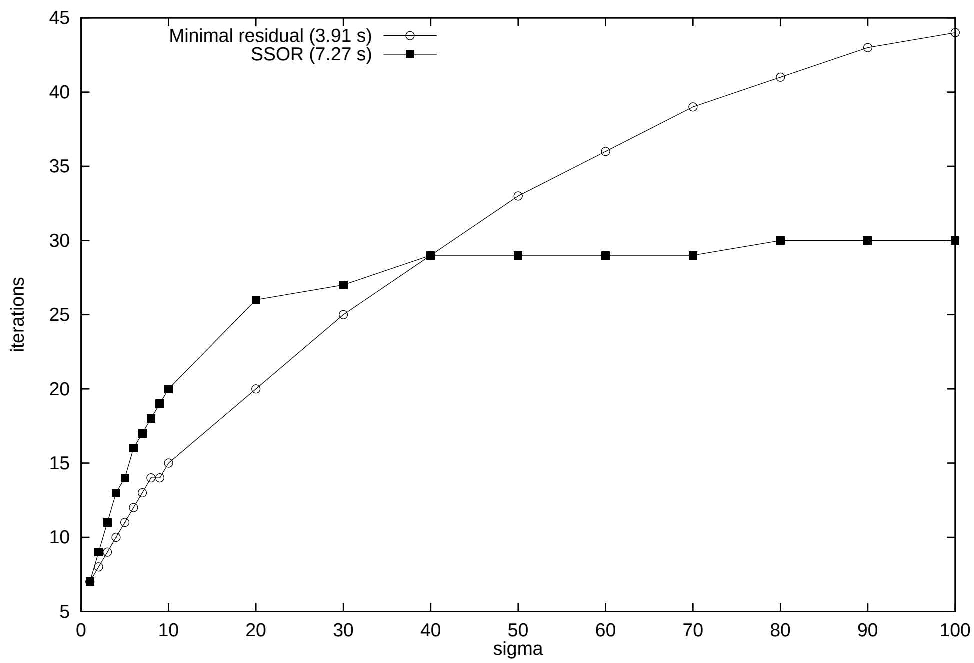


**Fig. 4.** Comparison at fixed $c = 0.98$, with total computation times

of SOR algorithm or so called DSA algorithms (Diffusion Synthetic Acceleration). At last let us mention that the structure of our scheme, constructed upon an operator splitting, is naturally devised for parallelization. □

## References

1. S. Akesbi, M.R. Laydi, and M. Mokhtar-Kharroubi. Décomposition d'opérateurs et accélération de la convergence en neutronique. *C.R. Acad. Sci. Paris*, 319:765–770, 1994.
2. S. Akesbi, M.R. Laydi, and M. Mokhtar-Kharroubi. Schemes and acceleration in transport theory. *To appear in Journ. of Transport Theory and Statistical Physics*, 1999.

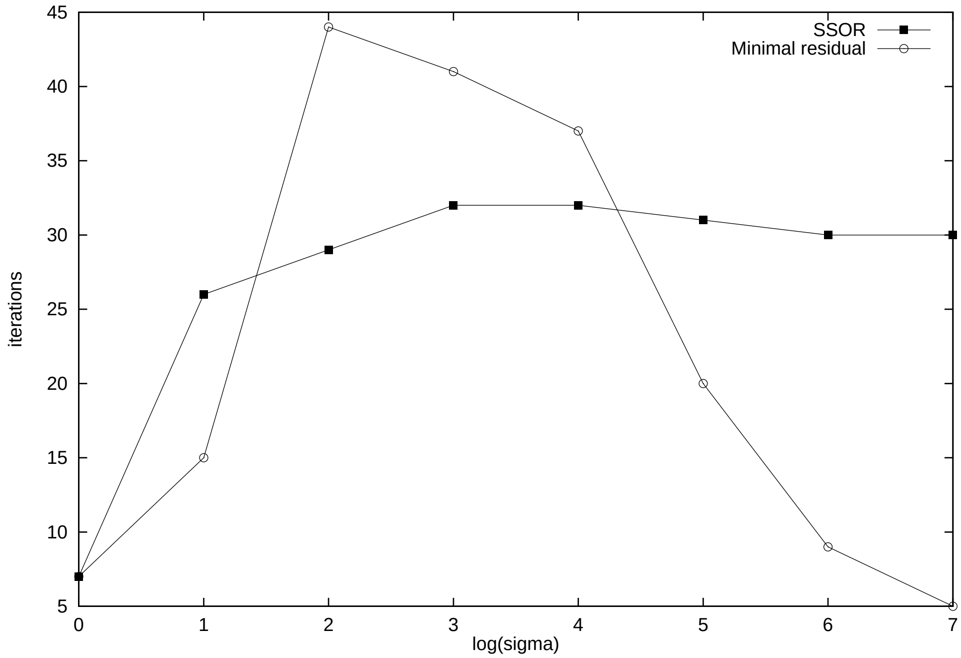


**Fig. 5.** Comparison at fixed $c = 0.98$ for large $\sigma$


3. S. Akesbi and M. Nicolet. Nouveaux algorithmes performants en théorie du transport. *M2AN*, 32(3):341–358, 1998.
4. Alcouffe-Clark-Larsen. *The diffusion synthetic acceleration in multiple time scales.* J. Brackbill, editor Ac. Press, 1985.
5. R. Dautray and J.-L.Lions. *Analyse mathématique et calcul numérique*, volume 9. Masson, 1987.
6. K.M. Khattab and E.W. Larsen. Synthetic acceleration methods for linear transport problems with highly anisotropic scattering. *Nucl. Sci. and Eng.*, 107:217–227, 1991.
7. E.W. Larsen. Unconditionally stable diffusion-synthetic acceleration methods for the slab geometry dicrete-ordinates equations. *Nucl. Sci. and Eng.*, parts I-II, 1988.
8. P. Lascaux and R. Théodor. *Analyse numérique matricelle appliquée à l'art de l'ingénieur*, volume 2. Masson, 1987.
9. I. Marek. Frobenius theory of positive operators, comparison theorems and applications. *Siam. Journ. Appl. Math.*, 19(3), 1970.