ESAIM: PROCEEDINGS AND SURVEYS, March 2016, Vol. 53, p. 191-210
M. Campos Pinto and F. Charles, Editors

OPTIMIZATION OF THE GYROAVERAGE OPERATOR BASED ON HERMITE INTERPOLATION

Fabien Rozar[1,3], Christophe Steiner[2], Guillaume Latu[3], Michel Mehrenberger[2], Virginie Grandgirard[3], Julien Bigot[1], Thomas Cartier-Michaud[3] and Jean Roman[4]

Abstract. Gyrokinetic modeling is appropriate for describing Tokamak plasma turbulence, and the gyroaverage operator is a cornerstone of this approach. In a gyrokinetic code, the gyroaveraging scheme needs to be accurate enough to avoid spoiling the data but also requires a low computation cost because it is applied often on the main unknown, the $5D$ guiding-center distribution function, and on the $3D$ electric potentials. In the present paper, we improve a gyroaverage scheme based on Hermite interpolation used in the Gysela code. This initial implementation represents a too large fraction of the total execution time. The gyroaverage operator has been reformulated and is now expressed as a matrix-vector product and a cache-friendly algorithm has been setup. Different techniques have been investigated to quicken the computations by more than a factor two. Description of the algorithms is given, together with an analysis of the achieved performance.

1. Introduction

Gyrokinetic modeling is appropriate for describing Tokamak plasma turbulence and the gyroaverage operator is a cornerstone of this approach. This is the model used by Gysela, a simulation code which is used to study the turbulence development in plasma fusion. The gyroaverage operator $\mathcal{J}$ transforms the so-called guiding-center distribution into the actual particle distribution. It enables to take into account effects relative to the finite *Larmor radius*, which is the radius of gyration of the gyro-center (motion which is faster than the turbulence we are looking at). In the present paper, we improve the gyroaverage scheme based on interpolation of the Gysela code to speedup calculations. For the current gyroaverage operator, the Larmor radius ρ is assumed to be independent in space. This code considers a computational domain in five dimensions (3D in space describing a torus geometry, 2D in velocity) [6,7]. Time evolution of the system consists in solving Vlasov equation non-linearly coupled to a Poisson equation (electrostatic approximation, quasi-neutrality is assumed). Routinely, physicists perform large Gysela simulations using from 1k to 16k cores on supercomputers. This work aims at improving the gyroaverage method based on Hermite interpolation which is too slow to be used in production for the moment. In achieving this optimization, we will allow the physicists to access to more accurate simulations and then to better understand the physical processes that arise in the simulations at finest scales.

[1] Maison de la Simulation, USR 3441, CEA / CNRS / Inria / Univ. Paris-Sud / Univ. Versailles, 91191 Gif-sur-Yvette, FRANCE
[2] IRMA, Université de Strasbourg, France
[3] CEA, IRFM, F-13108 Saint-Paul-lez-Durance
[4] Inria, Bordeaux INP, CNRS, FR-33405 Talence

Article published online by EDP Sciences and available at http://www.esaim-proc.org or http://dx.doi.org/10.1051/proc/201653012

The use of the Fourier transform reduces the gyroaveraging operation by a multiplication in the Fourier space by the Bessel function. Good approximations of the Bessel function have been proposed such as the widely used *Padé* expansion. The Padé approximation enables to recover a good approximation for small radii (see Fig. 5) [3]. However, for a larger Larmor radius and ordinary wave-numbers, the Padé approximation truncates the oscillations of the Bessel function, over-damps the small scales, and then introduces bias in the simulation data. Also the use of Fourier transform is not applicable in general geometry as we would like [3]; therefore it can not be employed in realistic tokamak equilibrium. These two limitations are overcome using an interpolation technique on the gyro-circles already introduced in [11] and which is the basis of our study.

In a gyrokinetic code, the gyroaveraging scheme needs to be accurate enough to avoid spoiling the data but also requires low computation cost because it is applied several times per time step on the main unknown, the 5D guiding-center distribution function. The gyroaverage is employed in GYSELA to compute the right-hand side of the Poisson equation, the gyroaveraged electric potentials that are used to get the advection field in the Vlasov solver and several diagnostics that export physical quantities on mass storage. One has to control the cost of this operator without compromising the quality, but the numerical methods, the algorithms and implementations play a major role here. By now, the new implementation of the gyroaveraging is ready for production runs and some numerical results are given.

The optimisation work presented here follows the previous ones [1, 9]. The work we have done on the gyroaverage operator is presented in this article as follows. The numerical method of the gyroaveraging using Hermite interpolation is explained in Section 2. The operator has been reformulated as a matrix-vector product in Section 3. Section 4 details the optimizations that quicken the computations by a factor two. Some cache-friendly algorithms have been setup. Section 5 shows the performance results in the GYSELA code of the different versions of the gyroaverage operator. Section 6 concludes and gives some hints to go forward in the optimization of the gyroaverage operator.

2. The gyroaverage operator based on interpolation method

2.1. Principle of the method

In this section, we describe the computation of the gyroaverage operator in real space developed in [11]. This method implies essentially interpolations over the Larmor circle. Let us consider a function f discretized over the $2D$ plane, on which the gyroaverage operator will be applied in this study. We distribute uniformly N quadrature points on the circle of integration and since the quadrature points do not coincide with grid points, we introduce an interpolation operator $\mathcal{P}$. Fig. 1 gives an example of this approach with $N = 5$. We have evaluated two schemes: Hermite interpolation and cubic spline interpolation. The gyroaverage is then obtained by the rectangle quadrature formula on these points. More precisely, for a given point (r_i, θ_j) in a polar coordinate system, the gyroaverage at this point is approximated by

$$\mathcal{J}_\rho(f)_{r_i,\theta_j} \simeq \frac{1}{2\pi} \sum_{k=0}^{N-1} \mathcal{P}(f)(r_i \cos\theta_j + \rho\cos\alpha_k, r_i \sin\theta_j + \rho \sin\alpha_k)\Delta\alpha, \tag{1}$$

where $\alpha_k = \theta_j + k\Delta\alpha$ and $\Delta\alpha = 2\pi/N$.

When points are outside the domain in the previous sum, we perform a radial projection on the border of the domain:

- if $r < r_{\min}$ then $\mathcal{P}(f)(r, \theta)$ is replaced by $\mathcal{P}(f)(r_{\min}, \theta)$,
- if $r > r_{\max}$ then $\mathcal{P}(f)(r, \theta)$ is replaced by $\mathcal{P}(f)(r_{\max}, \theta)$.

In the following, we describe the interpolation operator $\mathcal{P}$ we used. In practice, the Hermite interpolation method is faster than the cubic spline one [11]. So the optimizations done in this paper focus on the Hermite interpolation method mainly, even if these optimizations can also be applied easily to the cubic spline approach.

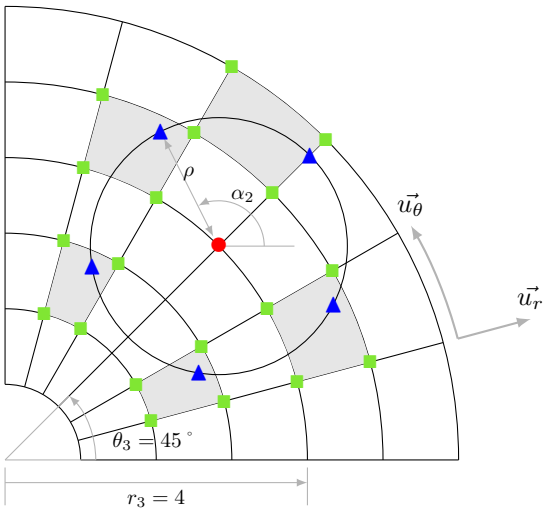

FIGURE 1. Description of the gyroaverage based on an interpolation method. One has to compute the gyroaverage on the red point ●. To do that, one estimates the average of the f function at blue locations ▲. The f values at blue points are obtained thanks to Hermite interpolation using f which is known on the mesh points, here particularly at mesh points ■.

2.2. **Hermite interpolation**

The Hermite interpolation method consists in a reconstruction of a polynomial function of degree 3 over a cell of the mesh (filled in light gray on Fig. 1) such that the value and the first derivative of this polynomial are equal to those of the input discretized function on the edges of the cell.

To do this, the values of the derivatives are reconstructed from the nodal values of the function f a finite difference scheme of arbitrary order d. Regardless of the d value, the Hermite polynomial remains of order 3, which allows a flexible use. It is possible to improve the precision of the interpolation by increasing the arbitrary order of the derivative reconstruction. In addition, unlike cubic splines interpolation, the Hermite interpolation method is very local; it requires only very few points close to the target position of the interpolation.

2.2.1. *Interpolation over a 1D mesh*

We detail here the Hermite interpolation operator in the uni-dimensional case before taking into account the 2D polar coordinate system in the next subsection. Let us consider a domain $\Omega = [a, b] \subset \mathbb{R}$ divided into N cells:

$$C_i = [x_i, x_{i+1}], \quad \text{with } i \in [\![0, N-1]\!].$$

The mesh is assumed uniform, i.e. for any index i the space step Δx verifies

$$\Delta x = x_{i+1} - x_i = \frac{b-a}{N}.$$

Let $\alpha \in [0, 1[$. The reconstruction of f by Hermite interpolation over the cell C_i reads:

$$\begin{aligned} f(x_i + \alpha\Delta x) \approx\ & (2\alpha+1)(1-\alpha)^2 f(x_i) + \alpha^2(3-2\alpha) f(x_{i+1}) + \\ & \alpha(1-\alpha)^2 f'(x_i^+) + \alpha^2(\alpha-1) f'(x_{i+1}^-). \end{aligned}$$

The right and left derivatives are reconstructed by centered finite differences of arbitrary order d (see [10]). In order to compute the left (resp. right) derivative $f'(x_i^-)$ (resp. $f'(x_i^+)$) at the point x_i, we use the nodal

values $f(x_{i+k})$ around the point x_i with a stencil from $k = r_d^- \leq 0$ to $k = s_d^- \geq 0$ (resp. $k = r_d^+ \leq 0$ to $k = s_d^+ \geq 0$). More precisely,

$$f'(x_i^-) \approx \Psi_d^-(f(x_i)), \qquad f'(x_i^+) \approx \Psi_d^+(f(x_i))$$

where the operators $\Psi_d^\pm : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ are:

$$(\Psi_d^-(X))_i = \sum_{k=r_d^-}^{s_d^-} \omega_{k,d}^- X_{i+k}, \qquad (\Psi_d^+(X))_i = \sum_{k=r_d^+}^{s_d^+} \omega_{k,d}^+ X_{i+k}$$

with

$$\omega_{k,d}^\pm = \begin{cases} \displaystyle\prod_{j=r_d^\pm,\ j\neq k,0}^{s_d^\pm} (-j) \Big/ \prod_{j=r_d^\pm,\ j\neq k}^{s_d^\pm} (k-j) & \text{if } k \neq 0 \\ \displaystyle -\sum_{j=r_d^\pm,\ j\neq 0}^{s_d^\pm} \omega_{j,d}^\pm & \text{else.} \end{cases}$$

For an even reconstruction order $d = 2p$, the stencil reads:

$$r_d^- = -p, \qquad s_d^- = p, \qquad r_d^+ = -p+1, \qquad s_d^+ = p+1$$

and for an odd order $d = 2p+1$, we have:

$$r_d^- = r_d^+ = -p, \qquad s_d^- = s_d^+ = p+1.$$

We observe that in the case of an odd (resp. even) order, the reconstruction is uncentered (resp. centered) and the reconstructed derivative function is C^0 (resp. C^1). In the numerical results presented in the following, the default order will be $d = 4$.

2.2.2. *Interpolation over a 2D polar mesh*

We consider now an uniform polar mesh over the domain $[r_{\min}, r_{\max}] \times [0, 2\pi]$ with $\mathrm{N_r} \times \mathrm{N}_\theta$ cells:

$$C_{ij} = [r_i, r_{i+1}] \times [\theta_j, \theta_{j+1}], \qquad \text{with } i \in [\![0, N_r - 1]\!],\ j \in [\![0, N_\theta - 1]\!]$$

where

$$\begin{aligned} r_i &= r_{\min} + i\tfrac{r_{\max}-r_{\min}}{\mathrm{N_r}}, & i &\in [\![0, N_r]\!] \\ \theta_j &= j\tfrac{2\pi}{\mathrm{N}_\theta}, & j &\in [\![0, N_\theta]\!]. \end{aligned}$$

Hermite interpolation over a polar mesh, which corresponds to a tensor product, consists on a succession of one-dimensional Hermite interpolations.

Fig. 2 highlights an intermediate step. On the red points $(\tilde{r}, \theta)$ and $(\tilde{r}, \theta + 1)$, the underlying discretized function f is interpolated to get its value and several derivatives. Then these values are used to evaluate f at the target point $f(\tilde{r}, \tilde{\theta})$ generally inside a cell. The interpolation method is summed up by the following algorithm:

- interpolation of the function $f(\cdot, \theta_j)$ over $[r_i, r_{i+1}]$ in order to evaluate $f(\tilde{r}, \theta_j)$
- interpolation of the function $f(\cdot, \theta_{j+1})$ over $[r_i, r_{i+1}]$ in order to evaluate $f(\tilde{r}, \theta_{j+1})$

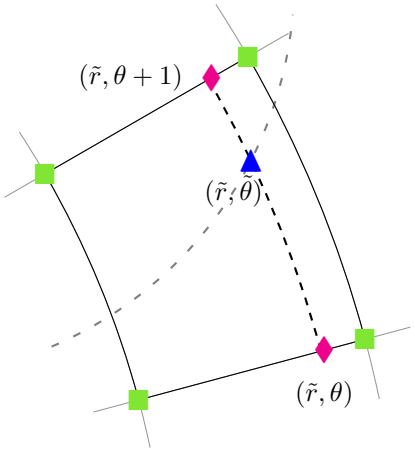

FIGURE 2. Details on the interpolation steps for one point. The point ▲ inside the cell represents the target value. It is obtained thanks to the Hermite interpolation between the intermediate points ♦ from the top and the bottom of the cell.

- interpolation of the function $\partial_\theta f(\cdot, \theta_j^+)$ over $[r_i, r_{i+1}]$ in order to evaluate $\partial_\theta f(\tilde{r}, \theta_j^+)$
- interpolation of the function $\partial_\theta f(\cdot, \theta_{j+1}^-)$ over $[r_i, r_{i+1}]$ in order to evaluate $\partial_\theta f(\tilde{r}, \theta_{j+1}^-)$
- interpolation of the function $f(\tilde{r}, \cdot)$ over $[\theta_j, \theta_{j+1}]$ by using the 4 previous evaluations to calculate $f(\tilde{r}, \tilde{\theta})$.

To achieve these 1D interpolations, we build as a first step the partial derivatives at the cell interfaces:

$$\begin{aligned}
(\partial_r f(r_i^+, \theta_j))_{i\in[\![0,N_r]\!]} &\approx \Psi_d^+(f(r_i,\theta_j)_{i\in[\![0,N_r]\!]}) \qquad \forall j \in [\![0, N_\theta]\!] \\
(\partial_r f(r_i^-, \theta_j))_{i\in[\![0,N_r]\!]} &\approx \Psi_d^-(f(r_i,\theta_j)_{i\in[\![0,N_r]\!]}) \qquad \forall j \in [\![0, N_\theta]\!] \\
(\partial_\theta f(r_i, \theta_j^+))_{j\in[\![0,N_\theta]\!]} &\approx \Psi_d^+(f(r_i,\theta_j)_{j\in[\![0,N_\theta]\!]}) \qquad \forall i \in [\![0, N_r]\!] \\
(\partial_\theta f(r_i, \theta_j^-))_{j\in[\![0,N_\theta]\!]} &\approx \Psi_d^-(f(r_i,\theta_j)_{j\in[\![0,N_\theta]\!]}) \qquad \forall i \in [\![0, N_r]\!],
\end{aligned}$$

then the second derivatives:

$$\begin{aligned}
(\partial^2_{r,\theta} f(r_i^+, \theta_j^+))_{i\in[\![0,N_r]\!]} &\approx \Psi_d^+(\partial_\theta f(r_i,\theta_j^+)_{i\in[\![0,N_r]\!]}) \qquad \forall j \in [\![0, N_\theta]\!] \\
(\partial^2_{r,\theta} f(r_i^-, \theta_j^+))_{i\in[\![0,N_r]\!]} &\approx \Psi_d^-(\partial_\theta f(r_i,\theta_j^+)_{i\in[\![0,N_r]\!]}) \qquad \forall j \in [\![0, N_\theta]\!] \\
(\partial^2_{r,\theta} f(r_i^+, \theta_j^-))_{i\in[\![0,N_r]\!]} &\approx \Psi_d^+(\partial_\theta f(r_i,\theta_j^-)_{i\in[\![0,N_r]\!]}) \qquad \forall j \in [\![0, N_\theta]\!] \\
(\partial^2_{r,\theta} f(r_i^-, \theta_j^-))_{i\in[\![0,N_r]\!]} &\approx \Psi_d^-(\partial_\theta f(r_i,\theta_j^-)_{i\in[\![0,N_r]\!]}) \qquad \forall j \in [\![0, N_\theta]\!].
\end{aligned}$$

In the general case, for any node (r_i, θ_j), the following 9 values need to be recorded:

$$\begin{gathered}
f(r_i,\theta_j), \quad \partial_r f(r_i^+,\theta_j), \quad \partial_r f(r_i^-,\theta_j), \quad \partial_\theta f(r_i,\theta_j^+), \quad \partial_\theta f(r_i,\theta_j^-) \\
\partial^2_{r,\theta} f(r_i^+,\theta_j^+), \quad \partial^2_{r,\theta} f(r_i^-,\theta_j^+), \quad \partial^2_{r,\theta} f(r_i^+,\theta_j^-), \quad \partial^2_{r,\theta} f(r_i^-,\theta_j^-).
\end{gathered}$$

However, in the case of an even order d, as the reconstructed derivative function is C^1, only the four following values are stored since $r^+ = r^-$ and $\theta^+ = \theta^-$:

$$f(r_i,\theta_j), \quad \partial_r f(r_i,\theta_j), \quad \partial_\theta f(r_i,\theta_j), \quad \partial^2_{r,\theta} f(r_i,\theta_j). \tag{2}$$

2.3. **Analytical test case**

To check and verify the gyroaveraging numerical solvers, one can use analytical solutions. For that, we use the Fourier-Bessel functions whose gyroaverage is analytically known [11]. More precisely, for an integer $m \geq 0$, let J_m and Y_m be respectively the Bessel functions of the first kind and of the second kind of order m (see [8]), we consider the following test function which verifies the homogeneous Dirichlet conditions on $r_{\min} > 0$ and $r_{\max}$:

$$f_{m,k} : (r,\theta) \in [r_{\min}, r_{\max}] \times [0, 2\pi] \mapsto \left(J_m(\gamma_{m,k}) Y_m \left(\gamma_{m,k} \frac{r}{r_{\max}} \right) - Y_m(\gamma_{m,k}) J_m \left(\gamma_{m,k} \frac{r}{r_{\max}} \right) \right) e^{im\theta}$$

where $\gamma_{m,k}$ is the k^{th} zero of the function $y \mapsto J_m(y) Y_m \left(y \frac{r_{\min}}{r_{\max}} \right) - Y_m(y) J_m \left(y \frac{r_{\min}}{r_{\max}} \right)$. Its gyroaverage reads:

$$\mathcal{J}_\rho(f_{m,k})(r_0,\theta_0) = J_0 \left(\gamma_{m,k} \frac{\rho}{r_{\max}} \right) f_{m,k}(r_0,\theta_0). \tag{3}$$

The real and imaginary parts of the function $f_{1,1}$ is illustrated at Fig. 3.

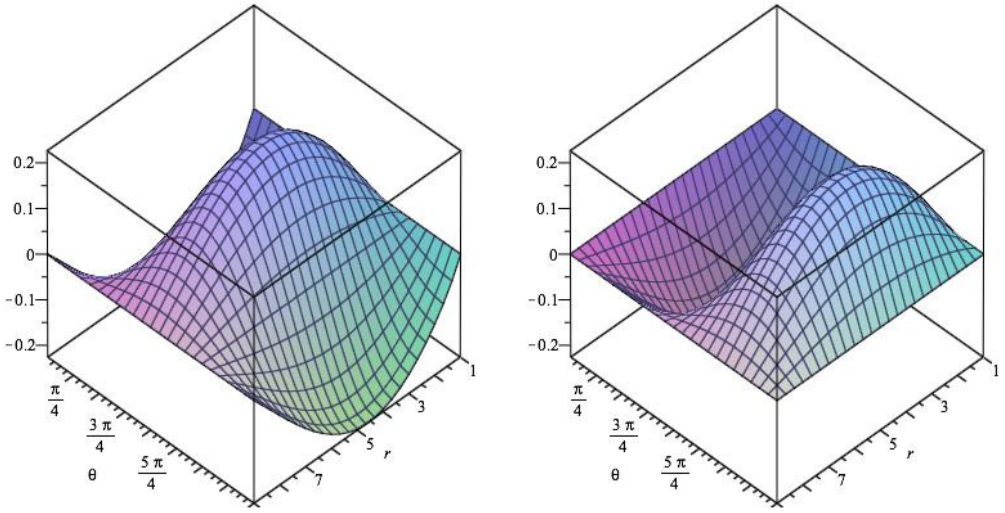

FIGURE 3. Real and imaginary parts of the test function $f_{1,1}$ with $r_{\min} = 1$ and $r_{\max} = 9$.

We designed a unit test which is based on the analysis of the error – the difference between analytical solution and the computed solution (using L^1 or L^2 norm). Setting the gyroradius ρ, the number of points on the Larmor circle N and the interpolation order d, one can draw the numerical behavior of the error and check the order 2 in space; in practice, we use the real part of these functions. The Fig. 4 shows the $L^2 - error$ between the analytical gyroaverage and the numerically computed gyroaverage with the three functions $f_{1,1}$, $f_{3,3}$ and $f_{8,8}$ depending on the resolution of the mesh. The higher the modes (m,k) of a function are, the more this function oscillates, which induces larger approximation errors. This explains the observed differences between the curves on Fig. 4 for the lowest values of N_r, N_θ. Starting from $N_r = N_\theta \geq 10^2$, the three curves are not distinguishable any more, which means the resolution of the mesh is fine enough to catch the oscillations of the three functions and so to ensure a good accuracy of the gyroaverage computation.

Another verification consists, for a function $f_{m,k}$ and a given point (r_0,θ_0), in calculating the ratio

$$\frac{\mathcal{J}_\rho(f_{m,k})(r_0,\theta_0)}{f_{m,k}(r_0,\theta_0)} \approx J_0 \left(\gamma_{m,k} \frac{\rho}{r_{\max}} \right)$$

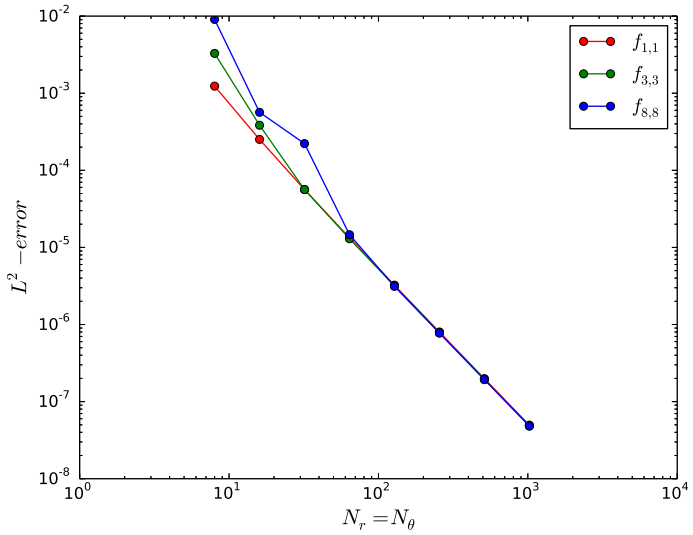

FIGURE 4. L^2-error for the gyroaverage of the real part of the functions $f_{1,1}, f_{3,3}$ and $f_{8,8}$ in function of $N_r = N_\theta$. The gyroaverage is computed using the method based on Hermite interpolation with $N = 1024$ points on the Larmor circle. Parameters : $r_{\min} = 1$, $r_{\min} = 9$, $d = 4$ and $\rho = 1$.

which can be seen as an approximation of the Bessel function J_0 thanks to Eq. (3). Fig. 5 illustrates the reconstruction of J_0 using the function $f_{8,8}$ with $r_{\min} = 1$, $r_{\max} = 9$, $N_r = N_\theta = 1024$, (r_0, θ_0) has coordinates $(470, 470)$ in the mesh, $\gamma_{8,8} \approx 36.0$ and $0 \leq \rho \leq \frac{10 r_{\max}}{\gamma_{8,8}} \approx 2.5$. Since the function $f_{8,8}$ has the highest mode values among the three considred functions, it is the most difficult function to handle by the gyroaverage operator and so it is relevant to study the error depending on the number of interpolation points. The larger the number N is, the better the Hermite curves fit the analytical solution J_0, whereas the Padé approximation appears as rough starting from $x = 2$.

These unit tests allow us to study the precision of the gyroaverage operator depending on the input parameters but it also represents reference cases. The validity of any new implementation was systematically checked by comparing their results with these unit tests against those of the reference implementation. During the development process, this check warns us early if something goes wrong, it facilitates the fixing of bugs.

3. Matrix representation for gyroaveraging

3.1. **Matrix description of the gyroaverage operator**

The previous section described the gyroaverage computed at a given grid point thanks to Eq. (1). In the case of the even order, this formula boils down to a linear combination of 4 values at the nodal points of the polar mesh: the value, the derivative of first order in the radial and in the poloidal directions, and the cross-derivative of second order (see relation (2)). These values are denoted W_f in the following.

In order to figure out how to compute the gyroaverage of all the points in the poloidal plane, a matrix representation of the operator is useful. Using this formulation, we will highlight the properties of the operator

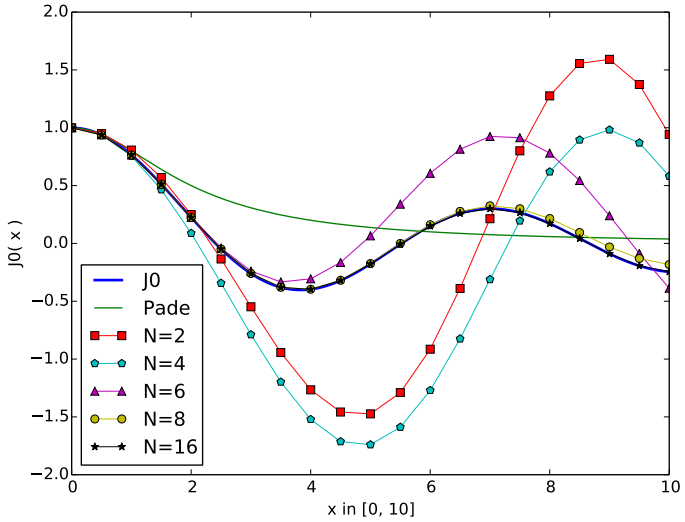

FIGURE 5. Reconstructions of the function J_0 using the function $f_{8,8}$ and its gyroaverage computed by the method based on Hermite interpolation with $N = 2, 4, 6, 8$ or 16 points on the Larmor circle. These reconstructions are functions of $x = \gamma_{8,8}\frac{\rho}{r_{\max}}$.

used for optimization issues. The gyroaverage based on Hermite interpolation can be written as follows:

$$M_{final} = M_{coef} \times M_{fval}$$
$$\text{with} \begin{cases} M_{final} & \in \mathcal{M}_{\mathrm{N_r}+1,\mathrm{N}_\theta}(\mathbb{R}), \\ M_{coef} & \in \mathcal{M}_{\mathrm{N_r}+1,4(\mathrm{N_r}+1)\mathrm{N}_\theta}(\mathbb{R}), \\ M_{fval} & \in \mathcal{M}_{4(\mathrm{N_r}+1)\mathrm{N}_\theta,\mathrm{N}_\theta}(\mathbb{R}). \end{cases}$$

The matrix M_{final} is the result of the gyroaverage applied to all the points of a poloidal cut. Every point on this matrix verifies the relation:

$$M_{final}(i,\, j) = \mathcal{J}_\rho(f)_{r_i,\theta_j}$$
$$\text{with} \begin{cases} r_i = & r_{\min} + i\frac{r_{\max}-r_{\min}}{\mathrm{N_r}}, & i \in [\![0, \mathrm{N_r}]\!], \\ \theta_j = & j\frac{2\pi}{\mathrm{N}_\theta}, & j \in [\![0, \mathrm{N}_\theta - 1]\!]. \end{cases}$$

The last point in the θ direction is excluded as it shares the same position as the first point in this periodical direction. Each row R_i of the matrix M_{coef} contains the contribution coefficients associated to the input plan at a given index i in the radial direction. The matrix M_{fval} contains in each column C_j the required values for Hermite interpolation. The factor 4 which appears in the size of the matrices M_{coef} and M_{fval} is due to the number of required values by the interpolation method over a $2D$ plane – 4 per grid point in the Hermite case.

With this representation, the gyroaverage computed at one grid point can be expressed as the following scalar dot product:

$$\mathcal{J}_\rho(f)_{r_i,\theta_j} = R_i(M_{coef}) \cdot C_j(M_{fval}). \tag{4}$$

3.2. Initial implementation of the gyroaverage operator

For convenience, we will use the matrix-like notations introduced above to describe the initial implementation used in [11]. This implementation does not benefit of the clear identification of the manipulated objects. In the initial implementation of the algorithm, the gyroaverage operator was not seen as a scalar dot product, and so some contribution was not factorized in the most efficient way.

Accordingly to the matrix representation of the gyroaverage, the two main properties of the matrix M_{coef} are *(i)* the sparsity (section 3.3) and *(ii)* and the fact that it remains unchanged during the whole simulation. This matrix is computed once for all in the initialization step, whereas the matrix M_{fval} is computed at each gyroaveraging because the values it contains are determined by the considered input data.

The implementation is composed of two steps: the first at the initialization stage described by Algo. 1 and the second at each call to `gyro_compute` during the simulation run described by Algo. 2.

input : $\mathrm{N_r}$, N_θ, ρ, N
output: A representation of M_{coef} – the weights and the indexes of their corresponding W_f for each i.

for $i \leftarrow 0$ **to** $\mathrm{N_r}$ **do**
  for $k \leftarrow 0$ **to** $N-1$ **do**
    Compute the k^{th} position of the quadrature point pt over the Larmor circle of center $(i,\, j=0)$;
    Compute/store the indexes of the corner points of the cell cell containing pt ;
    Compute/store the weights associated to the contribution of each corner point of cell into a representation of M_{coef};
  end
end

Algorithm 1: Computation of a representation of M_{coef} (initialization step).

input : The f values in the poloidal plane input_plane
output: The gyroaveraged poloidal plane: gyroaveraged values at each grid point of the poloidal plane

for *each grid point* pt $(i,\, j)$ *of* input_plane **do**
  Compute/store the first order derivative of f in each dir. and the cross-derivative at pt $(i,\, j)$ into a representation of $C_{\mathrm{o}}(M_{fval})$;
end
for *each point* pt $(i,\, j)$ *of* input_plane **do**
  Compute the gyroaveraged value at pt $(i,\, j)$– do the sparse dot product $R_i(M_{coef}) \times C_j(M_{fval})$;
  Store the gyroaveraged value into the input_plane at pt $(i,\, j)$;
end

Algorithm 2: Computation of the gyroaverage of f over a poloidal plane taken as input.

3.3. Pattern of the matrix M_{coef}

To figure out the sparsity of the matrix M_{coef}, take a look at Fig. 1 (p. 193). The i-th line of the matrix M_{coef} contains the coefficient contribution of every position of the input poloidal plane for the gyroaverage at a point $(i, j = *)$. As we can see on Fig. 1, to compute the gyroaverage of the point in the center of the circle (●), only a few points of the plane contribute (■). As a consequence, the weight associated to a point of the plan is non-zero for this gyroaverage only if it is used to get the value of an interpolate point (▲). This is the case for any mesh point. This leads us to conclude that the matrix M_{coef} is sparse. The Fig. 6 illustrates the generic appearance of the i-th line vector of M_{coef}.

The value of the contribution coefficients is invariant with respect to the poloidal index j. This is due to the fact that for 2 separate mesh points $(i,\, j_1)$ and $(i,\, j_2)$, the polar coordinates of the interpolation points involved

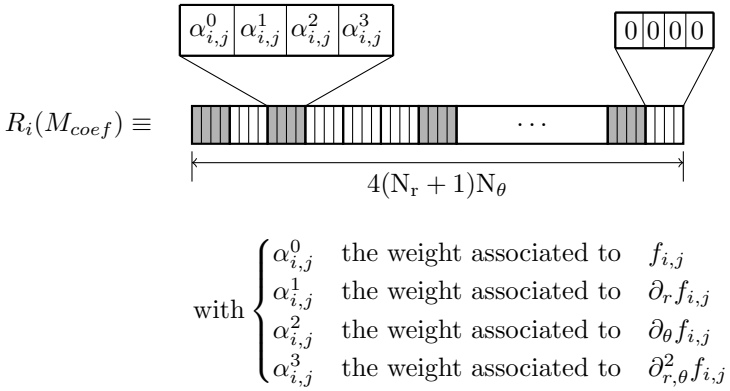

FIGURE 6. Sketch of the sparsity of a row vector $R_i(M_{coef})$. On this example, only the cells filled in grey contain non-zero values $\alpha^*_{i,j}$.

in their respective gyroaverage remain the same in local coordinate system along poloidal direction ($\vec{u_r}$, $\vec{u_\theta}$). The value and the distribution of the $\alpha^*_{i,j}$ coefficients depend only of the radial index i.

The number of non-zero values of 2 separated row vectors may be different. The pattern of the cells which contribute to the computation of the gyroaverage depends on the radial index i of the target point. Generally, the target points close to the center of the poloidal mesh (i.e. i is low) involve a larger number of cells compared to the point near to the external boundary of the plane.

3.4. **Pattern of the matrix** M_{fval}

The matrix M_{fval} stores the needed values for the Hermite interpolation, namely the nodal value and their derivatives. The layout of a column of this matrix is illustrated on Fig. 7.

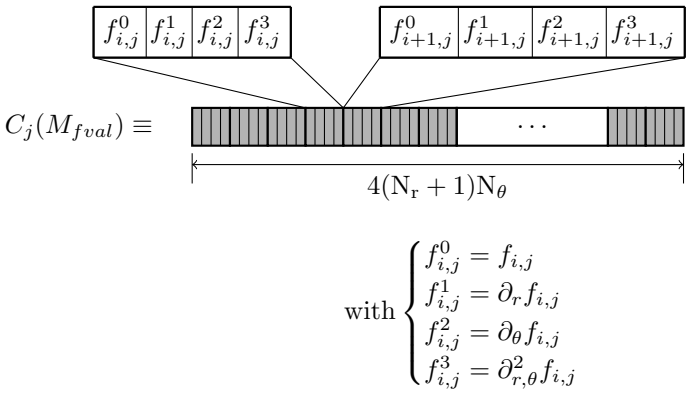

FIGURE 7. Sketch of the layout of a column vectors of M_{fval}.

The length of any vector $C_j(M_{fval})$ is $4(\mathrm{N_r}+1)\mathrm{N}_\theta$ and contains only non-zero values. In the current matrix representation (see Fig. 8), one can notice that M_{fval} has a repetitive pattern. The distribution of the values in a vector $C_j(M_{fval})$ depends on its poloidal index j. The values of the vector $C_{j+1}(M_{fval})$ can be obtained by a circular permutation of $4(\mathrm{N_r}+1)$ positions of the values of $C_j(M_{fval})$. The pattern of M_{fval} is outlined

on Fig. 8. From the 1^{st} column vector, the other ones can be deducted thanks to the following relation:

$$C_{j\neq \mathrm{o}}(M_{fval})[i] = C_{\mathrm{o}}(M_{fval})[i - j \times 4(\mathrm{N_r}+1)] \tag{5}$$

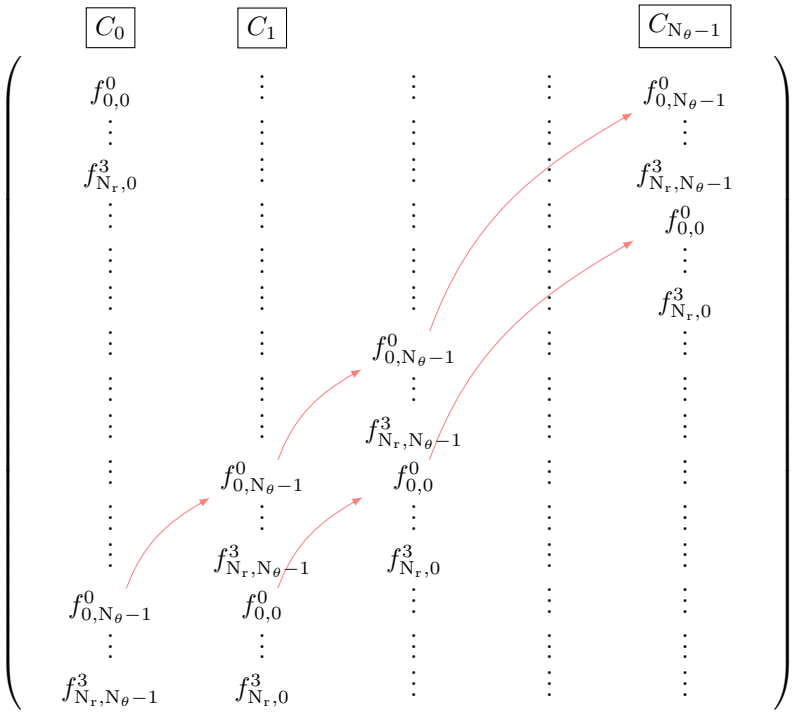

FIGURE 8. Illustration of the value position shift between the columns of M_{fval}.

The matrix representation of the gyroaverage operator provided in this Section gives us a useful description. Section 4 details the different optimizations we have made to quicken the initial computation of the gyroaverage thanks to this matrix representation.

4. Optimize and speedup gyroaveraging

The gyroaverage operator is decomposed into two stages, as already mentioned in section 3.2. Algo. 1 is computed during the initialization step and only once, so its execution time is taken as a payload for the simulation and is not the goal of the optimizations presented here. Algo. 2 is called many times during an iteration of Gysela and is the main kernel of the gyroaverage operator. The optimizations done focus on the reduction of the execution time of this algorithm. To achieve the optimization of the gyroaverage operator, we paid attention to data structures which represent the matrices involved in the computation. In a second step, issues relative to cache memory levels are tackled.

In this Section, we describe several implementations. The first one that is presented in 4.1 must be seen as the starting point for the optimizations introduced in 4.2 and 4.3.

4.1. Compact vector optimization

Obviously, the choice of the data structures for the matrix M_{coef} and M_{fval} has great impact on the performance. As said in 3.3, the matrix M_{coef} is sparse. To benefit of this property, only non-zero values of M_{coef}

are taken into account. By avoiding the storage of null values, the size of vectors involved in Eq. (4) is reduced, but it requires the use of dedicated data structures to achieve the computation of Algo. 2.

The main contribution of this optimization is the choice of the data structure that implements the sparse matrix version of the gyroaverage.

During the development process, different data structures were tested. The data structure which represents the matrix M_{coef} is less challenging than the representation of the matrix M_{fval}, since M_{fval} has a visible impact on performances. Among the different representations of M_{fval}, two of them are detailed and compared below.

4.1.1. *The representation of the matrix M_{coef}*

The goal of the computation of Algo. 1 is to compute the representation of the matrix M_{coef}. As said previously, it is done during the initialization step and only once. To benefit of the sparsity of this matrix M_{coef}, a dedicated data structure `contribution_vector` has been created to handle it. This structure is defined as follows:

```
1     type :: gyro_vector
2        integer, dimension(:), pointer :: ind
3        real(8), dimension(:), pointer :: val
4     end type gyro_vector
5
6     type(gyro_vector), dimension(:), pointer :: contribution_vector
```

The variable `contribution_vector` is an array of length $(N_r + 1)$ and of type `gyro_vector` which is a structure as defined above. Each cell of `contribution_vector` represents a row vector of the matrix M_{coef}. For a given radial index i, the indexes of the underlying mesh points which contribute to the gyroaverage of the point $(i,\ j = 0)$ are computed and saved in `contribution_vector(`i`)%ind`. Its length is denoted by `nb_contribution_pt(`i`)` and may be different for each radial index (see section 3.3). These indexes are used to build the variable which represents the matrix M_{fval} which will be detailed below. The weights associated to these positions are computed and stored in `contribution_vector(`i`)%val`.

4.1.2. *First representation of the matrix M_{fval}*

Algo. 2 which describes the main kernel of gyroaverage can be decomposed into two parts: *(i)* the computation of the derivatives and *(ii)* the computation of the product between the weights and the W_f of the input plane. At each call to `gyro_compute`, the derivatives of the input plane are computed and saved in a variable denoted `rhs_product`. This variable represents the matrix M_{fval}. In this implementation, `rhs_product` is defined as follows:

```
1     type :: small_matrix
2        real(8), dimension(:, :), pointer :: val
3     end type small_matrix
4
5     type(small_matrix), dimension(:), pointer :: rhs_product
```

The variable `rhs_product` is an array of length $(N_r + 1)$; it contains elements of type `small_matrix` which is a structure containing simply a $2D$ pointer. The shape of a $2D$ array `rhs_product(`i`)%val` is $(4 \times$ `nb_contribution_pt(`i`)`, $N_\theta)$. This array is the matrix that contains the subset of values from matrix M_{fval} contributing to the gyroaverage of the points $(i,\ j{=}*)$. Its first column vector `rhs_product(`i`)%val( :,`$j = 0$`)` contains the W_f retrieved from the first column vector $C_o(M_{fval})$ which corresponds to the position recorded in the vector `contribution_vector(`i`)%ind`. The other columns `rhs_product(`i`)%val( :,`$j \neq 0$`)` are deduced thanks to a shift of $j \times 4(N_r + 1)$ positions as one can see in Fig. 8.

Fig. 9 illustrates the matrix-vector product between the weights and the W_f. Here, the W_f from a vector `rhs_product(`i`)%val( :,`j`)` are organized in such a way that their positions match with the correct weight

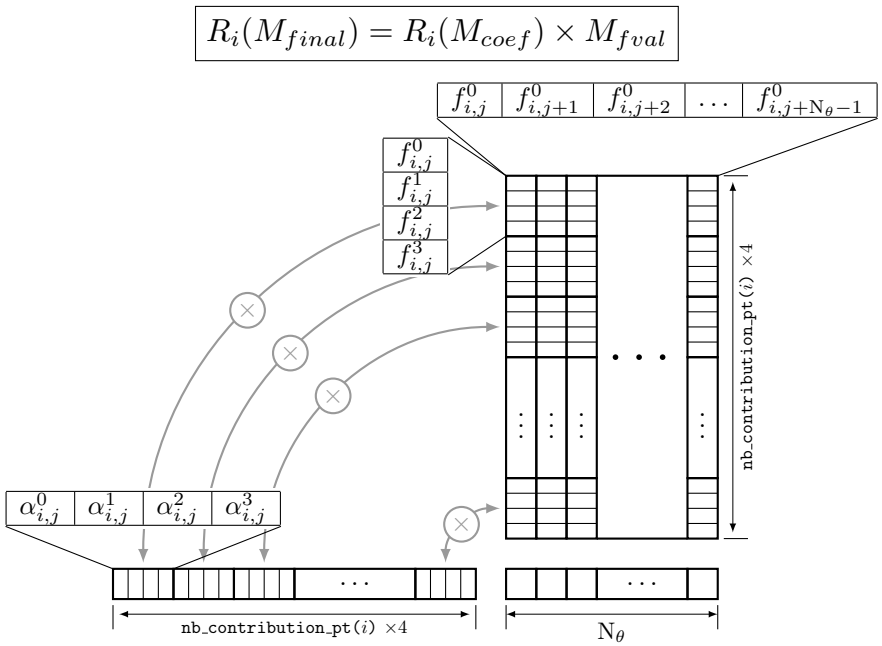

FIGURE 9. Illustration of a matrix-vector product between the weights (`contribution_vector(`i`)%val`) and the value-derivatives of the input plane (`rhs_product(`i`)%val`). One product computes N_θ gyroaverage values.

from the vector `contribution_vector(`i`)%val`. In this way, the gyroaverage can be implemented as matrix-vector products which are efficient operations on modern processors. Each matrix-vector product computes the gyroaverage of N_θ points. The gyroaverage of the whole poloidal plane is achieved with $(\mathrm{N_r}+1)$ matrix-vector products.

4.1.3. *Second representation of the matrix M_{fval}*

In this second approach, the variable `rhs_product` is still the representation of the matrix M_{fval}, but it is defined as follows:

```
1    real(8), dimension(:), pointer :: rhs_product
```

The `rhs_product` variable is here a $1D$ array. It corresponds exactly to the first column vector $C_\mathrm{o}(M_{fval})$. As said in Section 3.4, this vector contains all the W_f deduced from the input plane, so its length is $4(\mathrm{N_r}+1)\mathrm{N}_\theta$. One can see it as the linear flat representation of the input plane.

Fig. 10 shows the product between the weights and the W_f accordingly to this representation. The length of the vectors involved in this product does not match. The association between weights and the W_f is done thanks to the position recorded in the vector `contribution_vector(`i`)%ind`. This association is graphically represented by the arrows. The gyroaverages of points (i, $j \neq 0$) are computed thanks to a shift of $4j(\mathrm{N_r}+1)$ of the indexes from `contribution_vector(`i`)%ind`. This operation will be called *sparse dot product* in the following.

To achieve the gyroaverage of the whole poloidal plane, $(\mathrm{N_r}+1)\mathrm{N}_\theta$ sparse dot products are required.

4.1.4. *Benchmark to compare the data structures*

To compare the 2 data structures introduced above, we used the analytical test case as a benchmark. An execution of this program consists in computing the gyroaverage of a plane initialized by a chosen function and to compare the result with the associated theoretical value, as it is described in section 2.3. Let consider

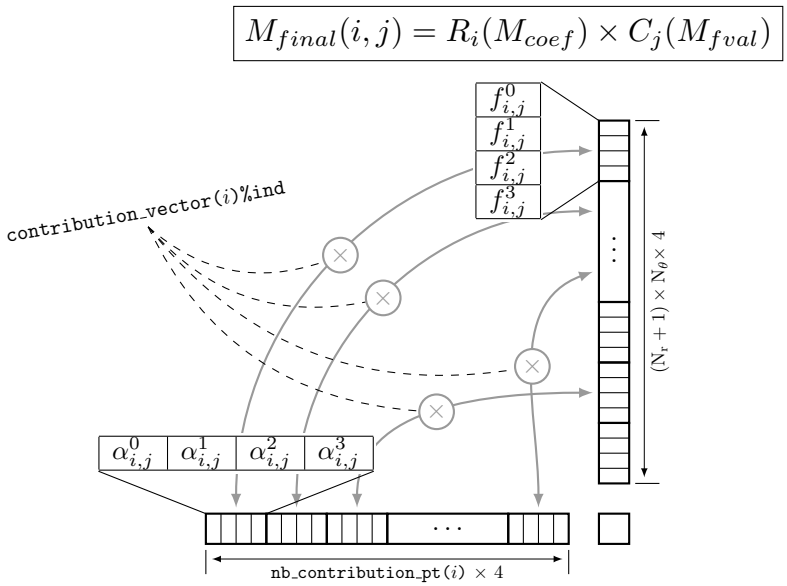

FIGURE 10. Illustration of the sparse dot product between the weigths (`contribution_vector(`i`)%val`) and the value-derivatives of the input plane (`rhs_product`). The matching between these two arrays is done thanks to the position recorded in `contribution_vector(`i`)%ind`.

$N_r = N_\theta$		32	64	128	256	512
Structure 1:	RHS	$1{,}5 \cdot 10^{-3}$	$1{,}1 \cdot 10^{-2}$	$8{,}5 \cdot 10^{-2}$	$5{,}1 \cdot 10^{-1}$	2,3
	Product	$2{,}1 \cdot 10^{-4}$	$1{,}7 \cdot 10^{-3}$	$1{,}4 \cdot 10^{-2}$	$8{,}3 \cdot 10^{-2}$	$3{,}5 \cdot 10^{-1}$
	Total	$1{,}7 \cdot 10^{-3}$	$1{,}3 \cdot 10^{-2}$	$9{,}9 \cdot 10^{-2}$	$5{,}9 \cdot 10^{-1}$	2,7
Structure 2:	RHS	$1{,}7 \cdot 10^{-4}$	$6{,}4 \cdot 10^{-4}$	$2{,}5 \cdot 10^{-3}$	$1{,}0 \cdot 10^{-2}$	$4{,}2 \cdot 10^{-2}$
	Product	$7{,}9 \cdot 10^{-4}$	$5{,}7 \cdot 10^{-3}$	$4{,}6 \cdot 10^{-2}$	$2{,}7 \cdot 10^{-1}$	1,2
	Total	$9{,}7 \cdot 10^{-4}$	$6{,}3 \cdot 10^{-3}$	$4{,}9 \cdot 10^{-2}$	$2{,}8 \cdot 10^{-1}$	1,3

TABLE 1. Comparison of performance of the 2 data structures on the analytical test case. Structure 1 corresponds to the first representation of M_{fval} (described in section 4.1.2) and structure 2 to the second one (described in section 4.1.3). The execution times are given in seconds. They result from the average over 100 runs. ($\rho = 0.05$, $r_{min} = 0.1$, $r_{max} = 0.9$, $N = 32$).

the complexity `comp(`N_r`,` N_θ`)` representing the product between the weights and the W_f as the number of multiplications involved. In our case, it reads as follows:

$$\texttt{comp}(N_r,\ N_\theta) = \sum_{i=0}^{N_r} 4 \times \texttt{nb\_contribution\_pt}(i) \times N_\theta.$$

This relation highlights the strong correlation between the number of grid points and the time complexity of the gyroaverage.

To identify and compare the behavior of 2 implementations of the gyroaverage, a scan on the size of the poloidal plane has been done. Tab. 1 shows the execution time of the gyroaverage for the different sizes of the input plane. The given measurements have been obtained as the average over 100 runs of the case presented in

Section 2.3. The time corresponding to the building of the variable `rhs_product` is given in lines "RHS" and the time for the product between the weights and the W_f in lines "Product".

The second data structure implementation is roughly twice faster than the first one. If you pay attention to the details, the first data structure is competitive during the "Product" step, but the time needed to build the `rhs_product` represents a large overhead. As the second data structure gives shorter execution times, this is the chosen implementation.

With this data structure, the accesses to the W_f are not contiguous during the sparse dot product (see Fig. 10). This is due to the irregular stencil required by the gyroaverage (see Fig. 1) and to the layout of the variable `rhs_product`. This constitutes a limit for this approach from the performance point of view.

The second data structure is trickier to manipulate for the sparse dot product as it is shown in Section 4.1.3, but it allows us to have the control on the internal data distribution which is the key point for the "layout optimization" (details in Section 4.3).

To achieve the gyroaverage of the whole plane, at each grid point, a sparse dot product has to be done. In the implementation on the current "compact vector optimization", the poloidal plan is covered thanks to 2 nested for-loop where j is the index of the internal loop. Geometrically, the plan is explored circle after circle, from the smaller to the larger ones. This access pattern has the property to extensively reuse the weights (`contribution_vector`). Although this property, it constitutes a limit for performance and represents the key point of the next optimization. We will refer to this access pattern to cover the plane as the *mesh_path* in the next sections.

4.2. **Blocking optimization**

As said previously, to achieve the gyroaverage of the whole poloidal plane, a sparse dot product must be done on every grid point. The "blocking optimization" extends the "compact vector optimization" implementation. This optimization is focused on the improvement of the access pattern *mesh_path* to the input plan inspired by the well-known blocking/tiling technique [12].

To improve the performance, the idea here is to follow a path over the grid points *mesh_path* which reduces the frequencies of cache misses during the computation. For any mesh point (i, j), the computation of the gyroaverage required data from `contribution_vector(`i`)` and `rhs_product`. As said in 4.1, doing the gyroaverage circle after circle ensures a good reuse of the weights `contribution_vector(`i`)`, but this is not the case for `rhs_product`. The access to its data is generally not contiguous and differs for 2 distinct mesh points. However, as the data are retrieved by cache line, the cache *hit rate* depends of the *mesh_path*[1]. The goal achieved in this optimization is to increase the reuse of the data available in cache along the *mesh_path*.

Instead of covering the poloidal plan circle after circle, the plan is covered using small $2D$ tiles in this version. These tiles are of size $BLOCK_r \times BLOCK_\theta$ and are numbered following the poloidal direction. Fig. 11 illustrates the tiling decomposition of a part of the poloidal mesh. On this sample, $BLOCK_r = 3$ and $BLOCK_\theta = 4$. In practice, the setting of $BLOCK_r$ and $BLOCK_\theta$ are tuned depending on benchmarks performed on each production machine.

The poloidal plane is partitioned into $NT_r \times NT_\theta$ tiles. The tiles are numbered along the poloidal direction, from the center to the outside of the plane. This *mesh_path* cut in tiles improves the temporal locality of the data from the array `rhs_product`. It increases the cache *hit rate* and so decreases the execution time.

4.3. **Layout optimization**

As for the "blocking optimization", the "compact vector optimization" is the starting point of this implementation. Especially, the *mesh_path* remains the same in the present implementation. Here, this optimization focuses on the order of the data in the array `rhs_product`. For a given *mesh_path*, the arrangement of these values has a strong impact on the execution time. We will refer to this arrangement of data as the *layout* of `rhs_product`.

[1] http://gameprogrammingpatterns.com/data-locality.html

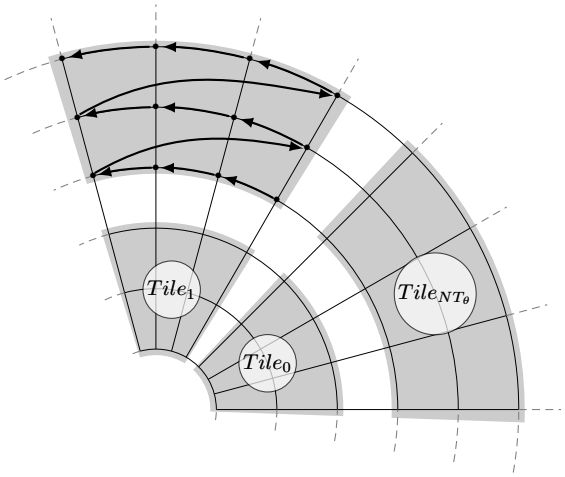

FIGURE 11. Sample of the tiling technique over a part of the poloidal mesh. The tiles which partition the mesh are traveled in ascending order. The nested arrows inside a tile show the underlying order.

The optimization of the layout is done for a given *mesh_path* over the input plane. The access history to the data of `rhs_product` depends on the *mesh_path*. The idea of the optimization presented here is to permute the elements of `rhs_product`, so change its layout, in order to increase the cache *hit ratio* along the *mesh_path*.

The *mesh_path* must be seen as an input parameter of the problem. Depending on the gyroradius ρ, the number of cells $N_r \times N_\theta$, the number of points on the Larmor circle N and the *mesh_path*, the access history to the array `rhs_product` is known. To determine how to change the layout, we need to introduce a metric of temporal distance of access of its data.

contribution point	list of temporal indexes	temporal_average
(0, 0)	[0, 1, 8, 9, 10, 17]	7.0
(0, 1)	[0, 1, 8, 9, 10, 17, 18, 19, 26]	12.0
(0, 2)	[9, 10, 17, 18, 19, 26, 27, 28, 35]	21.0
(0, 3)	[18, 19, 26, 27, 28, 35, 36, 37, 44]	30.0
⋮	⋮	⋮
(1, 0)	[0, 1, 2, 9, 10, 11]	5.0
⋮	⋮	⋮

TABLE 2. For each mesh point, list of the indexes corresponding to their use along the *mesh_path* and average of this temporal indexes.

To compute the gyroaverage at a point (i, j), the W_f values of its neighborhood are required. The points of this neighborhood are the contribution points to the gyroaverage at the point (i, j). To achieve the gyroaverage of a whole poloidal plan, we compute the gyroaverage of each mesh point one after an other. Let the *temporal index* be the index of a grid point along the *mesh_path*. For each point of the plane, we have listed the temporal indexes corresponding to the gyroaverages in which it is involved.

Tab. 2 gives a sample of the association between the mesh points and their temporal indexes. For instance, thanks to this table, one knows when the point `(0,0)` is used along the *mesh_path* during the sparse dot products: 0, 1, 8, 9, 10 and 17. The second information given by this table is the average of the temporal indexes for each

contribution point, which is denoted *temporal_average*. For example, the *temporal_average* of the point `(0,0)` is computed as follows:

$$\text{temporal\_average}((0,0)) = \frac{0+1+8+9+10+17}{6} = 7.$$

Let `pt1` a mesh point. If its *temporal_average* is rather low, this means `pt1` contributes rather in the beginning of the *mesh_path*. On the opposite, if the *temporal_average* of a mesh point `pt2` is rather high compared to the other temporal averages, this means `pt2` contributes rather in the end of the *mesh_path*.

The heuristic implemented in this optimization is to sort the elements of `rhs_product` accordingly to their *temporal_average*. For instance, from the information available on Tab. 2 the resulting partial data layout of `rhs_product` is: [(1, 0), (0, 0), (0, 1), (0, 2), (0, 3), ...]. The main idea is to keep accessed data between 2 successive sparse dot products as near as possible in order to minimize in average the number of cache misses.

In practice, the computation of the new layout of `rhs_product` is done during the initialization step. This implies that the *mesh_path* over the plane must be known at this step. The matching between the weights (`contribution_vector()%val`) and the W_f (`rhs_product`) during a sparse dot product is done thanks to the indexes retrieved from the array `contribution_vector()%ind`. To ensure the validity of the computation, the indexes recorded in array `contribution_vector()%ind` must be updated accordingly to this new layout.

5. Performance results in Gysela

5.1. **Benchmarking small cases**

After integration of the different versions of the gyroaverage operator in Gysela code [5], different runs have been done to compare their effectiveness. To keep a reasonably short execution time, the simulations are composed of only four time steps. These tests were done on the Helios[2] machine. Computation nodes used are equipped with two Intel Xeon E5-2450 2.10GHz processors, so 16 cores a node. On Tab. 3, the different test cases have been performed over the two following meshes (6) and (7):

$$N_r = 256,\ N_\theta = 256,\ N_\varphi = 32,\ N_{v_\parallel} = 16,\ N_\mu = 4 \tag{6}$$

$$N_r = 512,\ N_\theta = 512,\ N_\varphi = 32,\ N_{v_\parallel} = 16,\ N_\mu = 4. \tag{7}$$

To highlight the difference of performance between the different gyroaverage implementations, the size of the poloidal plane is multiplied by 4 between the two meshes. For these runs, every gyroaverage based of Hermite interpolation has $N = 16$ quadrature points on the integration circle. This number of integration points ensures a good presicion of the results as you can see on Tab. 5. The fluctuation of execution times (several percents typically, but rare events can lead to 10% or 20%) due to shared resources such as network and parallel file system is avoided using a specific configuration. Only one computation node of Helios is used for each run in order not to share the network. Also, the writing and reading on the parallel file system is reduced to the minimum. The benchmark configuration of one run is the following: 16 MPI processes composed of only 1 OpenMP thread. This benchmark corresponds to an execution of only four time steps of Gysela of a non physical case. This quite artificial setting will be revised in the next Subsection dedicated to production runs.

Tab. 3 shows some results concerning the Padé approximation compared to 4 versions of the Hermite implementation: original implementation; compact vector (described in Section 4.1), blocking (Section 4.2) and layout (Section 4.3) optimizations. Also, the Gysela code can be executed without any gyroaverage at all (solution named `Disable` gives invalid results of course but consitutes a reference execution time). This last option allows us to determine accurately the fraction of the execution time corresponding to the gyroaverage operator over the total execution time for the other cases.

As expected, the computation time of the gyroaverage operator grows along with the size of the poloidal mesh. The *compact vector optimization* which were an intermediate step during the optimization process is

[2]IFERC-CSC HELIOS super computer in Rokkasho-Japan, http://www.top500.org/system/177449

		Gyroaverage method					
		Padé	Hermite				`Disable`
			Initial	Compact vector	Layout	Blocking	
Mesh (6):	Total execution time (sec.)	110,20	150,60	146,35	132,91	131,39	`106,18`
	Percentage of gyroavg. execution	3.8 %	41.8 %	37.8 %	25.2 %	23.7 %	–
Mesh (7):	Total execution time (sec.)	464,62	629,46	627,80	557,49	550,38	`433,35`
	Percentage of gyroavg. execution	7.2 %	45.3 %	44.9 %	28.6 %	27.0 %	–

TABLE 3. Execution time and percentage over the total time of the gyroaverage operator for the different versions (for two different mesh sizes of poloidal plane with $N = 16$ interpolation points).

already a little bit faster than the original implementation. However, the percentage of total execution time dedicated to gyroaverage operator is about 6 to 10 times higher than for the Padé approximation. The *layout optimization* offers a great speed up compared to this previous step. On the biggest mesh, it reduces the execution by several tens of percents compared to the *compact vector optimization*. This demonstrates the big impact of the data layout in memory.

The best result is obtained with the *blocking optimization* which reduces the accumulated gyroaverage cost by 40% compared to the *initial* version. Nevertheless, it is still approximately 4 times more costly than the Padé approximation on the largest mesh. This is partly due to the number N of points used on the integration circle.

		Gyroaverage method					
		Padé	Hermite				`Disable`
			Initial	Compact vector	Layout	Blocking	
Mesh (6):	Total execution time (sec.)	110,00	131,42	129,11	123,12	120,47	`105,60`
	Percentage of gyroavg. execution	4.2 %	24.4 %	22.3 %	16.6 %	14.1 %	–
Mesh (7):	Total execution time (sec.)	464,20	544,07	538,38	525,21	510,44	`433,49`
	Percentage of gyroavg. execution	7.1 %	25.5 %	24.2 %	21.2 %	17.8 %	–

TABLE 4. Execution time and percentage over the total time of the gyroaverage operator for the different versions (for two different mesh sizes of poloidal plane with $N = 8$ interpolation points).

On Tab. 4, the same simulations have been run with less quadrature points, so $N=8$. As expected, reducing N induces smaller computation costs and decreases the execution times. With twice less quadrature points, the gyroaverage based on Hermite (*blocking*) reduces its execution time by 40%. Also the gap between the *blocking optimization* and the Padé approximation is reduced. However the drawback of taking less interpolation points leads to a subtle degradation of the quality of the gyroaverage operator. The trade-off between precision and computation will be discussed in the Subsection below.

5.2. **Benchmarking a real-life case**

Let us consider a regular domain size and a set of commonly used input parameters to analyze the performance of the new gyroaverage operator on a real-life case. The execution times will be a little bit tainted by network and parallel file system usage, because they are shared by several users. We have launched the simulations several times to ensure the execution times we get are reproducible. We focus on a Cyclone DIII-D base case, with a most unstable mode at $(m, n) = (14, -10)$ and the following parameters: $\mu_{min} = 0.143$, $\mu_{max} = 7.$, radial size $a = 100$ (described in [11], p. 10). The domain size is $N_r = 256$, $N_\theta = 256$, $N_\varphi = 64$, $N_{v_\parallel} = 48$, $N_\mu = 8$. This kind of typical benchmark had already been performed several years ago to validate the GYSELA

code. In order to check the good behavior of the simulation, a classical verification procedure is to focus on the linear phase and to extract the growth rate of the most unstable mode.

In the literature (see for example [2], p. 39-40), some theoretical and applied works have pointed out that taking $N=8$ points to discretize the gyroaverage operator is often sufficient. It has also been shown that this choice is already better than Padé to approximate the Bessel function $J_0(k_\perp \rho_i)$ on the interval $k_\perp \rho_i \in [0,5]$. Working with $N=16$ points to evaluate the gyroaverage extends the capability of the Hermite interpolation which is now able to accurately approximate the Bessel function on a larger interval $k_\perp \rho_i \in [0,10]$. Fig. 5 also corroborates this fact by an illustration on the specific function $f_{8,8}$.

Method	Padé	Hermite					
		N=2	N=3	N=4	N=6	N=8	N=16
Growth rate γ	.10436	.12246	.09420	.09625	.09643	.09644	.09644

TABLE 5. Linear growth rate for the different versions of the gyroaverage operator on a Cyclone test case (normalized growth rate as Fig. 1 of [4]).

On Tab. 5, the linear growth rates (normalized as in [4], Fig. 1) observed in the described use case are presented. They characterize the behavior of all the main components of the code during the linear phase. One can see that the values given by Hermite method for $N=6$, $N=8$, $N=16$ are almost equal, which is a good result that shows that simulations are well converged. The asymptotic value is reached quickly as N grows. However, for $N=2$, $N=3$ and for the Padé method, the growth rate is not well recovered. Practically, $N=8$ should be taken for production runs and this is a good cost-quality balance. The $N=16$ configuration will be useful also in specific cases where great accuracy is wanted by physicists (whenever particles with $k_\perp \rho_i > 5$ play a major role).

Code part \ Method	Padé	Hermite					
		N=8			N=16		
		Initial	Layout	Blocking	Initial	Layout	Blocking
Field solver	28s (0%)	32s (+14%)	31s (+10%)	31s (+11%)	35s (+27%)	32s (+13%)	32s (+16%)
Diagnostics	96s (0%)	123s (+29%)	108s (+12%)	110s (+15%)	147s (+54%)	114s (+19%)	120s (+26%)
Total	629s (0%)	662s (+5.3%)	659s (+4.8%)	651s (+3.6%)	689s (+9.6%)	670s (+6.6%)	666s (+5.8%)

TABLE 6. Cumulative execution time for some parts of the code that are impacted by gyroaverage computation costs. Timings are given for a run of 60 time steps using 512 cores. In parentheses, the percentage of extra time compared to Padé reference time is given.

Several execution times are shown in Tab. 6 that figure out the behavior of the Cyclone case with different versions of the gyroaverage operator. One can see that the *Diagnostics* and *Field solver* parts which use the gyroaverage operator are impacted by the chosen method. The improved versions (Layout and Blocking) divides by a factor 2 the overheads due to the Hermite interpolation method for the *Diagnostics*. On the overall total execution time, the best methods add 4% of extra time calculation for $N=8$ points and only 6% for $N=16$. This overhead is fully acceptable, as the accuracy of numerical results are greatly improved. Furthermore, the new versions described in this paper diminish the execution time of the gyroaverage by at least 40% compared to the initial method (Tab. 4).

6. Conclusion

To achieve the optimization of the gyroaverage operator, 4 steps have been achieved: *(i)* derive numerical approximation of the gyroaverage operator, *(ii)* describe formally the operator thanks to a matrix representation, *(iii)* explore and evaluate different optimization techniques, and *(iv)* integrate and validate these optimizations in the Gysela code. The current best implementation with Hermite interpolation is twice faster than the initial one. This improvement, but also the enhanced accuracy of the gyroaverage based on Hermite interpolation foster us to use the new solution in production runs instead of the Padé approximation.

The Padé gyroaverage implementation required a whole poloidal cut as input, due to a Fourier transform in θ direction and to a finite difference discretization along r direction. This involves sometimes collective communications to redistribute the data before applying the gyroaverage, because the main parallel domain decomposition is along r, θ and μ directions in Gysela. Historically, this was a hard constraint because Padé was the single solution available in Gysela for gyroaveraging. The new gyroaverage based on interpolation methods circumvents this issue. Parallelizing among several MPI processes will, from now on, be greatly eased. Additionally this new way to compute the gyroaverage allows us to consider easily its extension over general geometry meshes (not only polar planes).

References

[1] J. Bigot, V. Grandgirard, G. Latu, Ch. Passeron, F. Rozar, and O. Thomine. Scaling gysela code beyond 32k-cores on bluegene/q. In *CEMRACS 2012*, volume 43 of *ESAIM: Proc.*, pages 117–135, Luminy, France, 2013.

[2] I. Broemstrup. *Advanced Lagrangian Simulation Algorithms for Magnetized Plasma Turbulence*. PhD thesis, Univ. of Maryland, 2008.

[3] N. Crouseilles, M. Mehrenberger, and H. Sellama. Numerical solution of the gyroaverage operator for the finite gyroradius guiding-center model. *Commun. Comput. Phys.*, 8, 2010.

[4] A. M. Dimits et al. Comparisons and physics basis of tokamak transport models and turbulence simulations. *Physics of Plasmas*, 7(3):969–983, 2000.

[5] V. Grandgirard, J. Abiteboul, J. Bigot, T. Cartier-Michaud, N. Crouseilles, Ch. Erhlacher, D. Esteve, G. Dif-Pradalier, X. Garbet, Ph. Ghendrih, G. Latu, M. Mehrenberger, C. Norscini, Ch. Passeron, F. Rozar, Y. Sarazin, A. Strugarek, E. Sonnendrücker, and D. Zarzoso. A 5D gyrokinetic full-f global semi-lagrangian code for flux-driven ion turbulence simulations. working paper or preprint, July 2015.

[6] V. Grandgirard, M. Brunetti, P. Bertrand, N. Besse, X. Garbet, Ph. Ghendrih, G. Manfredi, Y. Sarazin, O. Sauter, E. Sonnendrücker, J. Vaclavik, and L. Villard. A drift-kinetic semi-lagrangian 4d code for ion turbulence simulation. *Journal of Computational Physics*, 217(2):395 – 423, 2006.

[7] V. Grandgirard, Y. Sarazin, P. Angelino, A. Bottino, N. Crouseilles, G. Darmet, G. Dif-Pradalier, X. Garbet, Ph. Ghendrih, S. Jolliet, G. Latu, E. Sonnendrücker, and L. Villard. Global full-f gyrokinetic simulations of plasma turbulence. *Plasma Physics and Controlled Fusion*, 49(12B):B173, 2007.

[8] M. Kreh. Bessel functions. *Lecture Notes, Penn State-Göttingen Summer School on Number Theory*, 2012.

[9] G. Latu, V. Grandgirard, N. Crouseilles, and G. Dif-Pradalier. Scalable quasineutral solver for gyrokinetic simulation. Research Report RR-7611, INRIA, May 2012.

[10] J. Li. General explicit difference formulas for numerical differentiation. *Journal of Computational and Applied Mathematics*, 183(1):29–52, 2005.

[11] C. Steiner, M. Mehrenberger, N. Crouseilles, V. Grandgirard, G. Latu, and F. Rozar. Gyroaverage operator for a polar mesh. *The European Physical Journal D*, 69(1), 2015.

[12] M. E. Wolf and M. S. Lam. A data locality optimizing algorithm. *SIGPLAN Not.*, 26(6):30–44, May 1991.