The combination technique applied to functionals

Yuancheng Zhou[1] Markus Hegland[2]

(Received 31 December 2020; revised 5 January 2022)

Abstract

Functionals related to a solution of a problem, usually modelled by partial differential equations, can be important quantities used to capture features of the problem. For high dimensional problems the computational cost of the functionals can be large since the numerical solution of a high dimensional partial differential equation is usually expensive to compute. We develop a new sparse grid combination technique to reduce the computational cost of such functionals. Our method is based on error splitting models of the functionals. However, it is hard to obtain a concrete error splitting model for complicated approximations. We show the connection between the decay of the surpluses and the error splitting models. By using the connection, we can also apply our combination technique to functionals when we only know their computed surpluses. Numerical experiments are provided to illustrate our idea and test the performance of our method.

DOI:10.21914/anziamj.v62.16112, Published 2022-02-07, as part of the Proceedings of the 19th Biennial Computational Techniques and Applications Conference. ISSN 1445-8810. (Print two pages per sheet of paper.)

Contents

1 Introduction

The sparse grid method [2, 3] is one of the most popular methods for dealing with many high dimensional problems, for example, solving PDEs and computing integrals. It will mitigate the 'curse of dimensionality' to some extent, if suitable regularity conditions are satisfied. The sparse grid combination technique [4, 2] is a way to compute a solution/integral on a sparse grid using only solutions/integrals computed on full grids. The formulation of the sparse grid combination technique is the inclusion-exclusion principle. Using this principle, one can also generalise the idea of the classical sparse grid combination technique according to different problems. This new method is called the generalised (sparse grid) combination technique [6, 5].

Error splitting models are commonly used as an assumption in proving the convergence of combination techniques. They are also used to design generalised combination techniques. The existence of such error splitting models has been proved for many problems when suitable regularity conditions are satisfied. However, problems arising from real world applications can be far more complicated than those well studied problems. Multistage approximations may be required for such complex problems. In addition, the error splitting model for approximation at each stage can be unknown. Therefore, one basic question is how to determine if the generalised combination technique still works/converges when the error splitting model is not available? In order not to be too general, we consider a kind of two-stage approximation problem: computing a given functional related to a solution of a given PDE/a system of PDEs.

2 Combination technique for functionals

We first state the two-stage approximation problem in detail. Suppose $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^d$ and $f \in U \subset X$ where X is a Banach space. Here f can be either a solution of a PDE/system of PDEs, or a function which needs to be approximated. We consider computing the following functional $T : U \to \mathbb{R}$, $f \mapsto T(f)$. In order to get an approximation to the functional value $T(f)$, we need to know approximations to both f and T. We assume α, β, τ, $\gamma \in \mathbb{N}^d$ are d-dimensional multi-indices. Let f_γ be the approximation of f on a given grid G_γ in domain Ω. Let U_γ be a finite dimensional subspace of U which depends on the choice of the grid G_γ and $f_\gamma \in U_\gamma$. We define an operator $P_\gamma : U \to U_\gamma$, $P_\gamma(f) = f_\gamma$. For example, suppose we take the domain $\Omega = [0, 1]$ and compute the piecewise linear interpolant on the grid G_γ which is an equally spaced grid with spacing $2^{-\gamma}$. The interpolant in this example is

$$f_\gamma = \sum_{i=0}^{2^\gamma} f(x_{\gamma,i}) b_{\gamma,i}\,, \tag{1}$$

where $x_{\gamma,i}$, $i = 0, \dots, 2^\gamma$ are grid points in grid G_γ and $b_{\gamma,i}$, $i = 0, \dots, 2^\gamma$ are linear nodal basis functions. Here U_γ is a finite dimensional space which is spanned by the basis functions and P_γ is a projection from U to its subspace U_γ.

Let $T_\tau : U \to \mathbb{R}$ be the approximation of functional T on a given grid G_τ in Ω. For example, we are particularly interested in the case when

$$T(f) = \int_\Omega f^p(x)\, dx\,, \tag{2}$$

where $p \in \mathbb{N}$, although our method can be applied to more generalised T. In this case, suppose $\Omega = [0, 1]$, f is continuous and G_τ is an equally spaced grid with spacing $2^{-\tau}$. Then

$$T_\tau(f) = \sum_{i=0}^{2^\tau} w_{\tau,i} f^p(x_{\tau,i})\,, \tag{3}$$

where $w_{\tau,i}$ and $x_{\tau,i}$, $i = 0, \ldots, 2^\tau$ are weights and quadrature points, respectively, of the quadrature rule used. With P_γ and T_τ, the approximated value of the functional $T(f)$ is $T_\tau \circ P_\gamma(f)$. For many problems, for simplicity the same grid is used, that is $G_\tau = G_\gamma$. Here we focus on this simple case and further define $Q_\gamma = T_\gamma \circ P_\gamma$. Our aim is to design a combination technique to compute $Q_\gamma(f)$ when the problem is high dimensional.

We say a sequence of grids G_γ, $\gamma \in \mathbb{N}^d$ is hierarchical if $G_\alpha \subset G_\beta$ when $\alpha \leqslant \beta$. Similarly, we say a sequence of finite dimensional spaces U_γ, $\gamma \in \mathbb{N}^d$ is hierarchical if $U_\alpha \subset U_\beta$ when $\alpha \leqslant \beta$. Suppose Q_γ, $\gamma \in \mathbb{N}^d$ is computed on a sequence of hierarchical grids G_γ, $\gamma \in \mathbb{N}^d$. We define the hierarchical surpluses operator

$$\Delta_\alpha := \Delta_{\alpha_1} \otimes \cdots \otimes \Delta_{\alpha_d}\,, \tag{4}$$

where Δ_{α_k}, $k = 1, \ldots, d$ are 1D hierarchical surplus operators

$$\Delta_{\alpha_k} = Q_{\alpha_k} - Q_{\alpha_k - 1}\,. \tag{5}$$

In Figure 1 we show how to compute the 2D surplus $\Delta_{2,2}$.

Using these hierarchical surplus operators, we define a new generalised combination technique for approximating the functional value $T(f)$.

$Q_{0,2}f$	$Q_{1,2}f$	$Q_{2,2}f$
$Q_{0,1}f$	$Q_{1,1}f$	$Q_{2,1}f$
$Q_{0,0}f$	$Q_{1,0}f$	$Q_{2,0}f$

$\Delta_{0,2}f$	$\Delta_{1,2}f$	$\Delta_{2,2}f$
$\Delta_{0,1}f$	$\Delta_{1,1}f$	$\Delta_{2,1}f$
$\Delta_{0,0}f$	$\Delta_{1,0}f$	$\Delta_{2,0}f$

Figure 1: Take $d = 2$ and $\alpha = (2, 2)$. By using the definition (4) we have $\Delta_{2,2} = \Delta_2 \otimes \Delta_2 = (Q_2 - Q_1) \otimes (Q_2 - Q_1) = Q_{2,2} - Q_{1,2} - Q_{2,1} + Q_{1,1}$.

Definition 1. Given $I \in \mathcal{P}(\mathbb{N}^d)$ where $\mathcal{P}(\mathbb{N}^d)$ is the power set of the set of all multi-indices of the form $(\alpha_1, \dots, \alpha_d)$, the generalised combination technique for computing the functional on set I is

$$Q_I(f) = \sum_{\alpha \in I} \Delta_\alpha(f)\,. \tag{6}$$

Although there is no restriction for the choices of the set I, we usually take I as a downset. We say I is a downset if $\alpha \in I$ and $\beta \leqslant \alpha$ implies $\beta \in I$. In particular, if $I = \{\alpha \mid |\alpha| \leqslant n + d - 1\}$, then we get a level n classical sparse grid.

Similarly to the approach used in the generalised combination technique for interpolation and quadrature [6, 5], the summation in (6) can also be written as a linear combination of the approximations of the functional value computed on different regular grids:

$$Q_I(f) = \sum_{\gamma \in I} c_\gamma Q_\gamma(f)\,, \quad c_\gamma = \sum_{\alpha \in C(\gamma)} (-1)^{|\gamma - \alpha|} \chi_I(\alpha)\,, \tag{7}$$

where $C(\gamma) = \{\alpha \mid \gamma \leqslant \alpha \leqslant \gamma + 1\}$ and $\chi_I(\alpha)$ is the characteristic function of I.

In particular, if $d = 2$ and $I = \{\alpha \mid |\alpha| \leqslant n + d - 1\}$, then we have

$$Q_n^c(f) := Q_I(f) = \sum_{\gamma_1 + \gamma_2 = n+1} Q_{\gamma_1,\gamma_2}(f) - \sum_{\gamma_1 + \gamma_2 = n} Q_{\gamma_1,\gamma_2}(f)\,, \tag{8}$$

(a) The classical SGCT Q_3^c. (b) A generalised SGCT Q_I.

Figure 2: The grid (hierarchical structure) used in the sparse grid combination technique (SGCT) to calculate functional approximations using (8) or (7). For each plot, the final approximation is obtained by adding all the red block approximations and subtracting all the blue block approximations. The red contour line shows the blocks with indices included in the set I in (6).

which is the 2D classical sparse grid combination technique. In Figure 2 we compare the classical combination technique (when $n = 3$) with a generalised 2D combination technique.

3 Error splitting model

3.1 Error splitting model and convergence results

The convergence of the sparse grid combination technique is based on the so-called error splitting model. For example, if we consider using a 2D classical sparse grid combination technique to approximate a 2D function f, then we require the error splitting model

$$f(x) - f_\gamma(x) = C_1(x, h_{\gamma_1})h_{\gamma_1}^p + C_2(x, h_{\gamma_2})h_{\gamma_2}^p + D_{1,2}(x, h_{\gamma_1}, h_{\gamma_2})h_{\gamma_1}^p h_{\gamma_2}^p\,, \quad (9)$$

where h_{γ_1} and h_{γ_2} are the spacings for different coordinates for all $\gamma \in \mathbb{N}^2$ and for all $x \in \Omega$. The error of the classical sparse grid combination technique [4]

$$f_n^c := \sum_{\gamma \in I} c_\gamma f_\gamma, \quad I = \{\alpha \mid |\alpha| \leqslant n + 1\}, \quad (10)$$

is bounded by

$$\|f - f_n^c\| \leqslant (3 + (1 + 2^p)n)Kh_n^p\,, \tag{11}$$

where K is the upper bound of the coefficient functions $C_1(\cdot\,,h_{\gamma_1})$, $C_2(\cdot\,,h_{\gamma_2})$ and $D_{1,2}(\cdot\,,h_{\gamma_1},h_{\gamma_2})$. The norm depends on the choice of the space of f. The same error splitting model (9) is required for the convergence of the generalised combination technique [5].

For the two-stage approximation problem and the previously defined generalised combination technique (6) and (7), we require the error splitting model

$$T - Q_\gamma = C_1(h_{\gamma_1})h_{\gamma_1}^p + C_2(h_{\gamma_2})h_{\gamma_2}^p + D_{1,2}(h_{\gamma_1},h_{\gamma_2})h_{\gamma_1}^p h_{\gamma_2}^p \tag{12}$$

to prove convergence in 2D case. Here $C_1(h_{\gamma_1})$, $C_2(h_{\gamma_2})$ and $D_{1,2}(h_{\gamma_1},h_{\gamma_2})$ are elements in the dual space of X, and we require that they are bounded with respect to the norm in the dual space. However, for many real world problems it is hard to obtain such an error splitting model before computation. Therefore, we cannot ensure the convergence of our method.

3.2 Error splitting model and surplus decay

If a numerical problem is approximated by a known numerical scheme with a tensor product structure, then we can obtain a concrete error splitting model from the numerical scheme. Using the error splitting model, we can further build a corresponding model to describe the decay of surpluses. The following theorem explains the idea in 2D.

Theorem 2. *Suppose T and Q_γ are defined as in the Section 2. Take $\Omega = [0,1]^2$. The grid used to compute Q_γ is an equally spaced 2D grid $G_\gamma = G_{\gamma_1} \times G_{\gamma_2}$ with the spacings*

$$h_{\gamma_k} = 2^{-\gamma_k}\,, \quad k = 1,2\,. \tag{13}$$

Suppose Q_γ, for all $\gamma \in \mathbb{N}^2$, satisfies a more general error splitting model

$$T - Q_\gamma = C_1(h_{\gamma_1})h_{\gamma_1}^p + C_2(h_{\gamma_2})h_{\gamma_2}^q + D_{1,2}(h_{\gamma_1},h_{\gamma_2})h_{\gamma_1}^p h_{\gamma_2}^q\,, \tag{14}$$

for $p, q \in \mathbb{N}$ *. Then the surpluses*

$$\Delta_\gamma = \Delta_{\gamma_1,\gamma_2} = Q_{\gamma_1,\gamma_2} - Q_{\gamma_1-1,\gamma_2} - Q_{\gamma_1,\gamma_2-1} + Q_{\gamma_1-1,\gamma_2-1} \tag{15}$$

satisfy

$$\Delta_\gamma = \Theta(h_{\gamma_1}, h_{\gamma_2}) h_{\gamma_1}^p h_{\gamma_2}^q \,, \quad \textit{for all } \gamma \in \mathbb{N}^2 \,, \tag{16}$$

where $\Theta(h_{\gamma_1}, h_{\gamma_2})$ *is an element in the dual space of* X *and*

$$\begin{aligned} \Theta(h_{\gamma_1}, h_{\gamma_2}) = & - D_{1,2}(h_{\gamma_1}, h_{\gamma_2}) + D_{1,2}(h_{\gamma_1-1}, h_{\gamma_2}) 2^p \\ & + D_{1,2}(h_{\gamma_1,}, h_{\gamma_2-1}) 2^q - D_{1,2}(h_{\gamma_1-1}, h_{\gamma_2-1}) 2^{p+q} \,. \end{aligned} \tag{17}$$

Moreover, if $C_1(h_{\gamma_1})$*,* $C_2(h_{\gamma_2})$ *and* $D_{1,2}(h_{\gamma_1,h_{\gamma_2}})$ *are bounded elements in the dual space of* X*, that is* $\|C_1(h_{\gamma_1})\| \leqslant K$ *,* $\|C_2(h_{\gamma_2})\| \leqslant K$ *and* $\|D_{1,2}(h_1, h_2)\| \leqslant K$ *for some* $K > 0$ *, then*

$$\|\Theta(h_{\gamma_1}, h_{\gamma_2})\| \leqslant K(1 + 2^p)(1 + 2^q) \,. \tag{18}$$

Proof: See Appendix A. ♠

Conversely, if we know the model which describes the decay of surpluses, that is,

$$\Delta_\gamma(f)(x) = \Theta(x, h_{\gamma_1}, h_{\gamma_2}) h_{\gamma_1}^p h_{\gamma_2}^q \,, \tag{19}$$

then we can also rebuild the error splitting model.

Theorem 3. *Suppose* T *and* Q_γ *are defined as in the Section 2. Take* $\Omega = [0, 1]^2$ *. The grid used to compute* Q_γ *is an equally spaced 2D grid* $G_\gamma = G_{\gamma_1} \times G_{\gamma_2}$ *with the spacings*

$$h_{\gamma_k} = 2^{-\gamma_k} \,, \quad k = 1, 2 \,. \tag{20}$$

Suppose the surplus operators satisfy

$$\Delta_\gamma = \Theta(h_{\gamma_1}, h_{\gamma_2}) h_{\gamma_1}^p h_{\gamma_2}^q \,, \quad \textit{for all } \gamma \in \mathbb{N}^2 \,, \tag{21}$$

then we have the following error splitting model

$$T - Q_\alpha = C_1(h_{\alpha_1})h^p_{\alpha_1} + C_2(h_{\alpha_2})h^q_{\alpha_2} + D_{1,2}(h_{\alpha_1}, h_{\alpha_2})h^p_{\alpha_1}h^q_{\alpha_2}\,, \tag{22}$$

for all $\alpha \in \mathbb{N}^2$ *,where*

$$\begin{aligned} C_1(h_{\alpha_1}) &= \sum_{\gamma_1>\alpha_1}\sum_{\gamma_2=0}^{\infty}\Theta(2^{(\alpha_1-\gamma_1)}h_{\alpha_1}, 2^{(\alpha_2-\gamma_2)}h_{\alpha_2})2^{(\alpha_1-\gamma_1)p}\,,\\ C_2(h_{\alpha_2}) &= \sum_{\gamma_2>\alpha_2}\sum_{\gamma_1=0}^{\infty}\Theta(2^{(\alpha_1-\gamma_1)}h_{\alpha_1}, 2^{(\alpha_2-\gamma_2)}h_{\alpha_2})2^{(\alpha_2-\gamma_2)q}\,,\\ D_{1,2}(h_{\alpha_1}, h_{\alpha_2}) &= \sum_{\gamma_1>\alpha_1}\sum_{\gamma_2>\alpha_2}\Theta(2^{(\alpha_1-\gamma_1)}h_{\alpha_1}, 2^{(\alpha_2-\gamma_2)}h_{\alpha_2})2^{(\alpha_1-\gamma_1)p}2^{(\alpha_2-\gamma_2)q}\,. \end{aligned} \tag{23}$$

Furthermore, if for some $K > 0$

$$\|\Theta(h_{\gamma_1}, h_{\gamma_2})\| \leqslant K\,, \quad \textit{for all}\; \gamma\,, \tag{24}$$

then $C_1(h_{\gamma_1})$*,* $C_2(h_{\gamma_2})$ *and* $D_{1,2}(h_{\gamma_1,h_{\gamma_2}})$ *are bounded elements in dual space of* X*, that is,*

$$\begin{aligned} &\|C_1(h_{\alpha_1})\| \leqslant \frac{K2^{-p}}{(1-2^{-q})(1-2^{-p})}\,, \quad \|C_2(h_{\alpha_2})\| \leqslant \frac{K2^{-q}}{(1-2^{-q})(1-2^{-p})}\,,\\ &\|D_{1,2}(h_{\alpha_1}, h_{\alpha_2})\| \leqslant \frac{K2^{-p}2^{-q}}{(1-2^{-q})(1-2^{-p})}\,. \end{aligned} \tag{25}$$

Proof: See Appendix B. ♠

Theorem 2 and Theorem 3 give the connection between the error splitting model and the decay of the surpluses. If we know the error splitting model, then the result in Theorem 2 can be used to check the numerical result obtained from computation. On the other hand, if we solve a complicated problem and implement multiple approximations during our computation, then it will be hard for us to get a concrete error splitting model. In this case, we can first compute the surpluses using given numerical schemes and then use Theorem 3 to study the corresponding error splitting model. Also, due to the connection between the error splitting model and the decay of the surpluses, one can check the convergence of the numerical scheme and design a more sophisticated combination technique by directly studying the computed surpluses instead of the unknown error splitting model.

4 Numerical results

In this section, we consider the determination of a quantity of interest for an application from plasma physics. We use the code GENE (Gyrokinetic Electromagnetic Numerical Experiment) which determines the time dependent density $f(x, v, t)$ of charged particles (ions and electrons) from an approximation of the discretised gyrokinetic equations [1]. The code GENE is an open source plasma microturbulence code which is used to efficiently compute gyroradius-scale fluctuations and the resulting transport coefficients in magnetized fusion/astrophysical plasma [7]. Here $x \in \mathbb{R}^3$ denotes the location of particles and $v \in \mathbb{R}^3$ denotes their velocity. Hence the phase space of a particle is 6D. In a GENE simulation, the 6D phase space is reduced into 5D by removing the fast gyromotion from the gyrokinetic equations. We keep using the notation $f(x, v, t)$ for the time dependent density after the fast gyromotion is removed [1]. Now $x \in X \subset \mathbb{R}^3$ denotes the position of the gyrocenter and v is in a 2D space V in which the two coodinates are the parallel velocity and the magnetic moment.

Here we consider the quantity of interest (nrg1[1]) [7]

$$q(t) = T(f(:, :, t)), \tag{26}$$

[1]The main output files of GENE are the nrg, the field and the mom files. The nrg files record the timetrace information, such as density, temperature and transport flux, all spatially averaged. The nrg1 is the first spatially averaged, normalized fluctuating quantity in an nrg file (the first column of the data). The quantity is the velocity space average/moment of the fluctuating part of a time-dependent particle distribution function.

where the functional is defined by

$$T(g) = \int_X \left(\int_V g(x, v)\, dv \right)^2 dx\,, \tag{27}$$

for some integrable function g defined on $X \times V$.

The computation of q at some time t requires a two stage approximation. Code GENE first computes f_γ as an approximation of f for multi-index $\gamma = (\gamma_1, \ldots, \gamma_5)$. Then it uses a quadrature rule T_τ to compute the quantity of interest. These two approximations are treated on the same grid for one simulation in GENE, therefore $\gamma = \tau$. However, it is not easy to obtain an error splitting model for

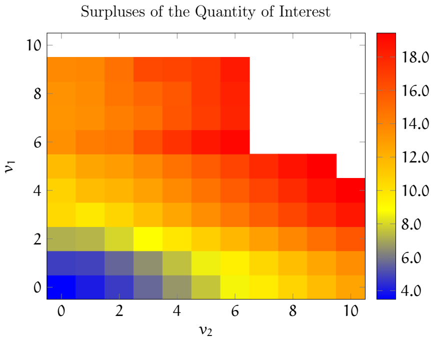
$$T(f(:,:,t)) - T_\gamma(f_\gamma(:,:,t))\,, \tag{28}$$

because the whole computation process in GENE is too complex for us to obtain an accurate analysis. Since GENE provides us with the data on the approximations of the solution of the nonlinear gyrokinetic equations and the quantities of interest on anisotropic full grids, the surpluses can be computed from these data. Therefore, using the equivalence of the error splitting model and surpluses decay model shown in Theorem 2 and 3, we switch to the surpluses decay model for this problem. In order to visualise the surpluses, we only consider the combination in the velocity space, which physicists are most interested in. We fix the real space indices γ_1, γ_2, γ_3 in the multi-index γ and set $\eta = (\gamma_4, \gamma_5)$ to be the new multi-index. The generalised combination technique on the velocity space is

$$Q_I(f(:,:,t)) = \sum_{\eta \in I} c_\eta Q_\eta(f(:,:,t)) = \sum_{\eta \in I} c_\eta T_\eta(f_\eta(:,:,t))\,. \tag{29}$$

By studying the surpluses decay, we are able to design combination techniques with high accuracy.

The absolute values of exponents of all the computed surpluses for the quantity of interest are shown in the Figure 3a. We observe that the absolute value of



(a) Surpluses

nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
14	14	15	16	17	17	18	nan	nan	nan	nan
13	14	15	15	16	17	18	nan	nan	nan	nan
13	13	14	15	16	17	18	nan	nan	nan	nan
13	14	15	16	17	18	19	nan	nan	nan	nan
12	13	13	14	15	16	17	18	19	19	inf
11	12	12	13	14	15	16	17	18	19	19
10	10	11	12	12	13	14	15	16	18	18
7	7	8	9	10	11	12	13	14	15	16
5	5	6	6	7	8	9	10	12	12	13
3	4	5	6	7	8	9	10	11	12	13

(b) Combination techniques

Figure 3: (a) Absolute values of exponents of all the surpluses $|\log(|\Delta_\gamma(T(u))|)/\log 4|$ are computed for the quantity of interest on a sequence of 2D hierarchical grids. The coarsest grid is 5×5 and is in the left bottom corner. The finest grid is 4097×4097 and is in the top right corner. The absolute value of a surplus is around 4^{-c} where c is the value on the colour bar with the corresponding colour. The data of the missing blocks is too expensive to compute and not available. (b) Two (generalised) combination techniques for computing the quantity of interest according to the surpluses plot (a). The number in each block is the absolute value of the exponent of the surpluses, with 'inf' indicating the surplus is very small and 'nan' indicating the surplus is expensive to compute and not available.

the surpluses (almost) decay while we increase the grid points in either the ν_1 or the ν_2 direction. In addition, the absolute values of surpluses decay much faster in the ν_1 direction than in the ν_2 direction. Therefore, the surpluses must satisfy the decay model as shown in (21) with bounded coefficient and different p and q. From the data and the Theorems, we can see the convergence of the generalised combination technique (29). In Figure 3b, we show two generalised combination techniques according to the observation in the Figure 3a. In the first (generalised) combination technique, we combine all the coloured blocks (red with coefficient 1, blue with coefficient -1) when the absolute value of the hierarchical surpluses are greater than 4^{-11}. The downset I (green line) of the combination technique is not symmetric. In the second (generalised) combination technique, we combine these three blocks on the top right corner, which is the best combination for the given data.

5 Conclusions and future work

We discuss a new generalised combination technique to compute a two-stage approximation problem. We show the equivalence between the error splitting model and the surplus decay model. If we can not obtain an accurate error splitting model of the two-stage approximation problem, then we can switch to studying the decay of computed surpluses by using the equivalence. We can also use the information from the computed surpluses to design a generalised combination technique and show the convergence of the method. The idea can also be applied to more complex real world problems which may require multi-stage approximations. By using a properly designed generalised combination technique, we reduce the computational cost to some extents for the complex, high dimensional problem. The dimension adaptive approach will be studied in the future.

References

[1] A. J. Brizard and T. S. Hahm. "Foundations of nonlinear gyrokinetic theory". In: *Rev. Mod. Phys.* 79.2 (2007), pp. 421–468. DOI: 10.1103/RevModPhys.79.421 (cit. on p. C217).

[2] H.-J. Bungartz and M. Griebel. "Sparse grids". In: *Acta Numer.* 13 (2004), pp. 147–269. DOI: 10.1017/S0962492904000182. (Cit. on p. C209).

[3] T. Gerstner and M. Griebel. "Numerical integration using sparse grids". In: *Numer. Algor.* 18 (1998), pp. 209–232. DOI: 10.1023/A:1019129717644. (Cit. on p. C209).

[4] M. Griebel, M. Schneider, and C. Zenger. “A combination technique for the solution of sparse grid problems”. In: *Iterative methods in linear algebra: Proceedings of the IMACS International Symposium on Iterative Methods in Linear Algebra, 1991*. Ed. by P. de Groen and R. Beauwens. North-Holland, Amsterdam, 1992, pp. 263–281. URL: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33.3530 (cit. on pp. C209, C213).

[5] B. Harding. “Fault tolerant computation of hyperbolic partial differential equations with the sparse grid combination technique”. PhD thesis. The Australian National University, 2016. URL: https://openresearch-repository.anu.edu.au/bitstream/1885/101226/1/Harding%20Thesis%202016.pdf (cit. on pp. C209, C212, C214).

[6] M. Hegland. “Adaptive sparse grids”. In: *Proceedings of the 10th Computational Techniques and Applications Conference CTAC-2001*. Ed. by K. Burrage and R. B. Sidje. Vol. 44. 2003, pp. C335–C353. DOI: 10.21914/anziamj.v44i0.685 (cit. on pp. C209, C212).

[7] Gene Development Team; F. Jenko et al. *The Gyrokinetic Plasma Turbulence Code Gene: User Manual*. 2013. URL: http://genecode.org/ (cit. on p. C217).

A Proof of Theorem 1

Proof: Using the definition of the surplus (4) and the error splitting model, we have

$$\begin{aligned}\Delta_\gamma &= \Delta_{\gamma_1,\gamma_2}\\ &= (Q_{\gamma_1,\gamma_2} - Q_{\gamma_1-1,\gamma_2} - Q_{\gamma_1,\gamma_2-1} + Q_{\gamma_1-1,\gamma_2-1}) \qquad (30)\\ &= [(Q_{\gamma_1,\gamma_2} - I) + (I - Q_{\gamma_1-1,\gamma_2}) + (I - Q_{\gamma_1,\gamma_2-1}) + (Q_{\gamma_1-1,\gamma_2-1} - I)]\\ &= \Theta(h_{\gamma_1}, h_{\gamma_2}) h_{\gamma_1}^p h_{\gamma_2}^q\,,\end{aligned}$$

where

$$\begin{aligned}\Theta(h_{\gamma_1}, h_{\gamma_2}) = [&-D_{1,2}(h_{\gamma_1}, h_{\gamma_2}) + D_{1,2}(h_{\gamma_1-1}, h_{\gamma_2})2^p \\ &+ D_{1,2}(h_{\gamma_1,}, h_{\gamma_2-1})2^q - D_{1,2}(h_{\gamma_1-1}, h_{\gamma_2-1})2^{p+q}] \,.\end{aligned} \tag{31}$$

Since the coefficients in the error splitting model are bounded, we have

$$\|\Theta(h_{\gamma_1}, h_{\gamma_2})\| \leqslant K(1+2^p)(1+2^q)\,, \tag{32}$$

by using the triangle inequality. ♠

B Proof of Theorem 2

Proof: Using the inclusion-exclusion principle, we have

$$\begin{aligned}T - Q_\alpha &= T - Q_{\alpha_1,\alpha_2} \\ &= \left(\sum_{\gamma_1>\alpha_1}\sum_{\gamma_2=0}^{\infty} + \sum_{\gamma_1=0}^{\infty}\sum_{\gamma_2>\alpha_2} - \sum_{\gamma_1>\alpha_1}\sum_{\gamma_2>\alpha_2}\right)\Delta_{\gamma_1,\gamma_2} \\ &= \left(\sum_{\gamma_1>\alpha_1}\sum_{\gamma_2=0}^{\infty} + \sum_{\gamma_1=0}^{\infty}\sum_{\gamma_2>\alpha_2} - \sum_{\gamma_1>\alpha_1}\sum_{\gamma_2>\alpha_2}\right)\Theta(h_{\gamma_1}, h_{\gamma_2})h_{\gamma_1}^p h_{\gamma_2}^q \,.\end{aligned} \tag{33}$$

Denote

$$\begin{aligned}\Theta_1(h_{\gamma_1}) &= \sum_{\gamma_2=0}^{\infty}\Theta(h_{\gamma_1}, h_{\gamma_2})h_{\gamma_2}^q\,, \\ C_1(h_{\alpha_1}) &= \sum_{\gamma_1>\alpha_1}\Theta_1(2^{(\alpha_1-\gamma_1)}h_{\alpha_1})2^{(\alpha_1-\gamma_1)p}\,,\end{aligned} \tag{34}$$

and the first term on the right hand side of (33) is

$$\begin{aligned}\sum_{\gamma_1>\alpha_1}\sum_{\gamma_2=0}^{\infty}\Theta(h_{\gamma_1}, h_{\gamma_2})h_{\gamma_2}^q &= \sum_{\gamma_1>\alpha_1}\Theta_1(h_{\gamma_1})h_{\gamma_1}^p \\ &= h_{\alpha_1}^p\sum_{\gamma_1>\alpha_1}\Theta_1(2^{(\alpha_1-\gamma_1)}h_{\alpha_1})2^{(\alpha_1-\gamma_1)p} \\ &= C_1(h_{\alpha_1})h_{\alpha_1}^p\,.\end{aligned} \tag{35}$$

For the second term on the right hand side of (33), we get a similar result. For the last term we denote

$$D_{1,2}(h_{\alpha_1}, h_{\alpha_2}) = \sum_{\gamma_1>\alpha_1} \sum_{\gamma_2>\alpha_2} \Theta(2^{(\alpha_1-\gamma_1)} h_{\alpha_1}, 2^{(\alpha_2-\gamma_2)} h_{\alpha_2}) 2^{(\alpha_1-\gamma_1)} 2^{(\alpha_2-\gamma_2)q}, \tag{36}$$

and then have

$$\sum_{\gamma_1>\alpha_1} \sum_{\gamma_2>\alpha_2} \Theta(h_{\gamma_1}, h_{\gamma_2}) h_{\gamma_1}^p h_{\gamma_2}^q = D_{1,2}(h_{\alpha_1}, h_{\alpha_2}) h_{\alpha_1}^p h_{\alpha_2}^q . \tag{37}$$

If (24) holds, then we have

$$\|\Theta_1(h_{\gamma_1})\| \leqslant K \sum_{\gamma_2=0}^{\infty} h_{\gamma_2}^q = K \sum_{\gamma_2=0}^{\infty} 2^{-\gamma_2 q} = \frac{K}{1-2^{-q}} \tag{38}$$

and

$$\|C_1(h_{\alpha_1})\| \leqslant \frac{K}{1-2^{-q}} \sum_{\gamma_1>\alpha_1} 2^{(\alpha_1-\gamma_1)p} = \frac{K 2^{-p}}{(1-2^{-q})(1-2^{-p})} . \tag{39}$$

Using a similar method to that used to obtain (39), we also compute a bound for $C_2(h_{\alpha_2})$:

$$\|C_2(h_{\alpha_2})\| \leqslant \frac{K 2^{-q}}{(1-2^{-q})(1-2^{-p})} . \tag{40}$$

For $D_{1,2}(h_{\alpha_1}, h_{\alpha_2})$:

$$\|D_{1,2}(h_{\alpha_1}, h_{\alpha_2})\| \leqslant K \sum_{\gamma_1>\alpha_1} 2^{(\alpha_1-\gamma_1)p} \sum_{\gamma_2>\alpha_2} 2^{(\alpha_2-\gamma_2)p} = \frac{K 2^{-p} 2^{-q}}{(1-2^{-q})(1-2^{-p})} . \tag{41}$$

♠

Author addresses

1. **Yuancheng Zhou**, Mathematical Sciences Institute, The Australian National University, Canberra ACT 0200, Australia.
mailto:yuancheng.zhou@anu.edu.au
2. **Markus Hegland**, Mathematical Sciences Institute, The Australian National University, Canberra ACT 0200, Australia.
mailto:markus.hegland@anu.edu.au