Conservative Fluctuations

Hiroaki Nishikawa

March 2004

1 Conservation in Fluctuation-Splitting Schemes

Consider solving the conservation law

$$\mathbf{u}_t + \mathrm{div}\mathcal{F} = 0 \tag{1}$$

where $\mathbf{u}$ is the vector of conservative variables and $\mathcal{F}$ is the flux tensor. In the fluctuation-splitting method, the first step is to evaluate the fluctuation over the cell $T \in T$.

$$\phi^T = -\iiint_T \mathrm{div}\mathcal{F}\; dV. \tag{2}$$

The second step is to distribute the fluctuation to each vertex $j \in \{j_T\}$ of the element,

$$\delta\mathbf{u}_j^n = \delta\mathbf{u}_j^n + \mathcal{B}_j^T\;\phi^T \tag{3}$$

where the matrix $\mathcal{B}_j^T$ is the distribution matrix with the property that the sum over the element reduces to the identity matrix:

$$\sum_{j\in\{j_T\}} \mathcal{B}_j^T = \mathbf{I}. \tag{4}$$

After completing the distribution for all elements, we update the nodal solution

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^n + \frac{\Delta t}{V_j}\,\delta\mathbf{u}_j \tag{5}$$

with an appropriate time step Δt where V_j denotes the median dual volume. Conservation is defined as a property that the sum of the changes of the nodal solutions depends only on boundary data. Consider the sum

$$\sum_{j\in\{J\}} \delta\mathbf{u}_j^n = \sum_{j\in\{J\}} \sum_{T\in\{T_j\}} \mathcal{B}_j^T\;\phi^T \tag{6}$$

where $\{J\}$ is a set of vertices over the entire domain and $\{T_j\}$ is a set of triangles that shares the node j. Manipulating the double sum on the right, we obtain

$$\sum_{j\in\{J\}} \delta \mathbf{u}_j^n = \sum_{T\in\{T\}} \sum_{j\in\{j_T\}} \mathcal{B}_j^T \, \phi^T \tag{7}$$

which by (4) becomes

$$\sum_{j\in\{J\}} \delta \mathbf{u}_j^n = \sum_{T\in\{T\}} \phi^T. \tag{8}$$

Therefore, the conservation is satisfied only if ϕ^T is telescoping, i.e. interior contribution cancels out. This requires that the fluctuation be precisely in the form of flux balance.

$$\phi^T = \oint_{\partial T} \mathcal{F}_n \, dS. \tag{9}$$

where $\mathcal{F}_n$ is the flux vector inwardly normal to the face and ∂T is the cell boundary.

2 General Forms of Conservative Fluctuation

There are two approaches to evaluate the conservative fluctuation.

2.1 Direct Discretization

We may directly discretize the surface integral (9) to evaluate the fluctuation. Over the element faces, we may employ any type of quadrature formula to evaluate the integral. If we assume a simple variation of some variable, the integration may be performed to yield a relatively simple analytic expression. If a numerical quadrature is employed, the result can be written in the general form

$$\phi^T = \sum_{faces} \sum_{k\in\{k\}} w_k \, \mathcal{F}_n(\mathbf{u}_k) \, \Delta S. \tag{10}$$

where $\{k\}$ is the set of quadrature points and w_k is the weight associated with each point. This approach has the advantage that it can be extended easily to higher order.

2.2 Linearization

For the purpose of designing fluctuation-splitting schemes, it is convenient to be able to write the fluctuation in the linearized form. That is, we seek to find an average state $\tilde{\mathbf{u}}$ over the element such that

$$\phi^T = -\mathcal{A}(\tilde{\mathbf{u}}) \cdot \iiint_T \text{grad}\mathbf{u} \, dV = \oint_{\partial T} \mathcal{F}_n \, dS. \tag{11}$$

Note that the linearized form must be *identical* to the flux integral over the element surface for conservation*. For some simple elements, unique average states have been found for the Euler equations. But it is generally difficult to find a practically simple averaged state for arbitrary elements.

3 Linear Elements

3.1 Conservative linearization

Conservative linearization can be derived easily for linear elements. Both approaches discussed above can be used to obtain an identical result. But it is much easier to take the linearization approach in this particular case. Consider

$$\phi^T = -\iiint_T \mathrm{div}\mathcal{F}\, dV. \tag{12}$$

For the Euler equations, the flux tensor is known to be quadratic in the parameter vector $\mathbf{z}$ which is a very convenient set of variables to work with. Then, we write

$$\phi^T = -\iiint_T \mathcal{A}_z \cdot \mathrm{grad}\,\mathbf{z}\; dV \tag{13}$$

where $\mathcal{A}_z = \dfrac{\partial \mathcal{F}}{\partial \mathbf{z}}$. We now assume that $\mathbf{z}$ is linear over the element so that its gradient is constant and can be taken out of the integral.

$$\phi^T = -\left(\iiint_T \mathcal{A}_z\; dV\right) \cdot \mathrm{grad}\,\mathbf{z}. \tag{14}$$

Besides, since $\mathcal{A}_z$ is linear in $\mathbf{z}$, its integral can be performed easily to yield

$$\phi^T = -\,\mathcal{A}_z(\tilde{\mathbf{z}}) \cdot \mathrm{grad}\,\mathbf{z}\, V_T \tag{15}$$

where $\tilde{\mathbf{z}}$ is the arithmetic average of $\mathbf{z}$ over the vertices of the element and V_T is the measure of the element. This has been obtained by *exact integration* under the assumption of linear parameter vector, and therefore this is identical to the surface integral evaluated under the same assumption, thus satisfying conservation. This linearization is called conservative linearization.

3.2 Other Forms

Linearization is convenient because we can work on any form of the governing equations by using transformation matrices evaluated at the same averaged state. For example, if we wish to work with the primitive form of the equations $\mathbf{w}_t + \mathcal{A}_w \cdot \mathrm{grad}\,\mathbf{w} = 0$. Then, using the transformation matrices $\mathbf{T}_u$ and $\mathbf{T}_w$,

$$\partial \mathbf{w} \;=\; \mathbf{M}_u \partial \mathbf{u} \tag{16}$$

$$\partial \mathbf{w} \;=\; \mathbf{M}_z \partial \mathbf{z}, \tag{17}$$

*In one dimension, this requirement reduces to $\mathcal{A}(\tilde{\mathbf{u}})\Delta U = \Delta F$.

we obtain

$$\phi_w^T = \tilde{\mathbf{M}}_u \phi^T = -\tilde{\mathcal{A}}_w \cdot \left(\tilde{\mathbf{M}}_z \, \mathrm{grad}\, \mathbf{z} \right) V_T \tag{18}$$

where tilde indicates that the quantity is evaluated at the averaged state. As this example implies, we may use any form of the governing equations as long as its coefficient matrices are evaluated at the averaged state and the gradient is computed from the gradient of the parameter vector. This gives us, for instance, an opportunity to apply different distribution strategies for different part of the fluctuation as is done in the elliptic-hyperbolic decomposition method.

3.3 Keeping Conservation Property

It is important to note that the fluctuation-splitting shcmes must be carefully implemented to satisfy conservation when we work with the fluctuation in other forms of the equations than the conservation form. For example, in the primitive form, we have

$$\phi_w^T = \tilde{\mathbf{M}}_u \phi^T = \tilde{\mathbf{M}}_u \oint_{\partial T} \mathcal{F}_n \, dS. \tag{19}$$

Then, because of the presence of $\tilde{\mathbf{T}}_u$,

$$\sum_{T \in \{T\}} \phi_w^T = \sum_{T \in \{T\}} \tilde{\mathbf{M}}_u \oint_{\partial T} \mathcal{F}_n \, dS \neq \oint_{\partial \Omega} \mathcal{F}_n \, dS \tag{20}$$

where $\partial\Omega$ denotes the domain boundary. Therefore the fluctuation for the primitive form is not telescoping. This is natural as it reflects the property of the governing equation that there exist no physical fluxes in the primitive form of the governing equations. To check conservation, we need to consider the conservative variables. Suppose we perform the distribution in terms of the primitive variables,

$$\delta \mathbf{w}_j^n = \delta \mathbf{w}_j^n + \mathcal{B}_j^T \, \phi_w^T \qquad \forall \, j \in \{j_T\} \tag{21}$$

and compute the change in the conservative variables at the vertices by the vertex-wise transformation. Then we have

$$\begin{aligned}
\sum_{j \in \{J\}} \delta \mathbf{u}_j^n &= \sum_{j \in \{J\}} (\mathbf{M}_u)_j^{-1} \, \delta \mathbf{w}_j^n && (22)\\
&= \sum_{j \in \{J\}} (\mathbf{M}_u)_j^{-1} \sum_{T \in \{T_j\}} \mathcal{B}_j^T \, \phi_w^T && (23)\\
&= \sum_{T \in \{T\}} \sum_{j \in \{j_T\}} (\mathbf{M}_u)_j^{-1} \mathcal{B}_j^T \, \phi_w^T && (24)\\
&= \sum_{T \in \{T\}} \sum_{j \in \{j_T\}} (\mathbf{M}_u)_j^{-1} \mathcal{B}_j^T \tilde{\mathbf{M}}_u \, \phi^T. && (25)
\end{aligned}$$

Conservation requires

$$\sum_{j\in\{j_T\}} (\mathbf{M}_u)_j^{-1} \mathcal{B}_j^T \tilde{\mathbf{M}}_u = \mathbf{I}, \tag{26}$$

but this is not satisfied in general. Therefore the resulting method is not conservative. On the other hand, if we compute the change in the conservative variables at the vertices by the element-wise transformation,

$$\sum_{j\in\{J\}} \delta \mathbf{u}_j^n = \sum_{j\in\{J\}} \sum_{T\in\{T_j\}} \tilde{\mathbf{M}}_u^{-1} \mathcal{B}_j^T \, \phi_w^T \tag{27}$$

$$= \sum_{T\in\{T\}} \sum_{j\in\{j_T\}} \tilde{\mathbf{M}}_u^{-1} \mathcal{B}_j^T \, \phi_w^T \tag{28}$$

$$= \sum_{T\in\{T\}} \sum_{j\in\{j_T\}} \tilde{\mathbf{M}}_u^{-1} \mathcal{B}_j^T \tilde{\mathbf{M}}_u \, \phi^T. \tag{29}$$

Now, we have

$$\sum_{j\in\{j_T\}} \tilde{\mathbf{M}}_u^{-1} \mathcal{B}_j^T \tilde{\mathbf{M}}_u = \tilde{\mathbf{M}}_u^{-1} \left(\sum_{j\in\{j_T\}} \mathcal{B}_j^T \right) \tilde{\mathbf{M}}_u = \mathbf{I}, \tag{30}$$

and therefore

$$\sum_{j\in\{J\}} \delta \mathbf{u}_j^n = \sum_{T\in\{T\}} \phi^T \tag{31}$$

$$= \oint_{\partial\Omega} \mathcal{F}_n \, dS, \tag{32}$$

thus satisfying conservation. This shows that we must distribute the change for the conservative variables to the vertices and also that we must transform the split fluctuations element-wise before the distribution, not vertex-wise after the distribution.

3.4 Remarks on Unusual Distribution Schemes

Some distribution schemes do not have the property

$$\sum_{j\in\{j_T\}} \mathcal{B}_j^T = \mathbf{I}. \tag{33}$$

The least-squares scheme is one such distribution scheme. Minimizing, with respect to the nodal solutions, the fluctuation in the least-squares norm

$$(\phi^T)^t \, \mathbf{Q}^T \phi^T \tag{34}$$

where $\mathbf{Q}^T$ is a weighting matrix, we obtain the distribution matrix in the form

$$\mathcal{A}_j^T = \left(\mathbf{Q}^T \, \frac{\partial \phi^T}{\partial \mathbf{u}_j} \right)^t \tag{35}$$

where $()^t$ denotes the transpose. This matrix has the property

$$\sum_{j \in \{j_T\}} \mathcal{A}_j^T = \mathbf{0}, \tag{36}$$

which implies that

$$\sum_{j \in \{J\}} \delta \mathbf{u}_j^n = \sum_{T \in \{T\}} \sum_{j \in \{j_T\}} \mathcal{A}_j^T \, \phi^T \tag{37}$$

$$= \mathbf{0}, \tag{38}$$

i.e. the solutions are merely redistributed, and their sum changes only by boundary conditions. Conservation is therefore satisfied, but in an unusual sense.

A problem arises when this type of distribution scheme is used with the conventional one. Suppose we decompose the primitive fluctuation into two parts as follows

$$\phi_w^T = \begin{bmatrix} \phi_H^T \\ \phi_E^T \end{bmatrix}, \tag{39}$$

and wish to distribute ϕ_H^T with the conventional scheme and the other ϕ_E^T with the least-squares scheme,

$$\delta \mathbf{w}_j^n = \delta \mathbf{w}_j^n + \begin{bmatrix} \mathcal{B}_j^T & \mathbf{0} \\ \mathbf{0} & \mathcal{A}_j^T \end{bmatrix} \begin{bmatrix} \phi_H^T \\ \phi_E^T \end{bmatrix}. \tag{40}$$

Then, we have for conservative variables

$$\sum_{j \in \{J\}} \delta \mathbf{u}_j^n = \sum_{T \in \{T\}} \sum_{j \in \{j_T\}} \tilde{\mathbf{M}}_u^{-1} \begin{bmatrix} \mathcal{B}_j^T & \mathbf{0} \\ \mathbf{0} & \mathcal{A}_j^T \end{bmatrix} \tilde{\mathbf{M}}_u \, \phi^T \tag{41}$$

$$= \sum_{T \in \{T\}} \tilde{\mathbf{M}}_u^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tilde{\mathbf{M}}_u \, \phi^T \tag{42}$$

$$= \sum_{T \in \{T\}} \tilde{\mathbf{M}}_u^{-1} \begin{bmatrix} \phi_H^T \\ 0 \end{bmatrix} \tag{43}$$

$$\neq \sum_{T \in \{T\}} \phi^T, \tag{44}$$

and we lose conservation.

4 Linearizing the Euler Equations for Arbitrary Elements

Consider the two-dimensional Euler equations (the extension to three-dimension is straightforward),

$$\mathbf{u}_t + \mathbf{f}_x + \mathbf{g}_y = 0. \tag{45}$$

The flux vectors are quadratic in the parameter vector $\mathbf{z}$, and therefore we can write

$$\mathbf{f} = \mathbf{z}^t \mathbf{C} \mathbf{z} \tag{46}$$

$$\mathbf{g} = \mathbf{z}^t \mathbf{D} \mathbf{z} \tag{47}$$

where $\mathbf{C}$ and $\mathbf{D}$ are constant symmetric 3rd-rank tensors which have been introduced just for convenience and so their precise forms are not important.

Recall that in the linearization approach, we seek to find an average state $\tilde{\mathbf{u}}$ such that the following is satisfied exactly,

$$\phi^T = -\mathcal{A}(\tilde{\mathbf{u}}) \cdot \iiint_T \text{grad}\mathbf{u} \; dV = \oint_{\partial T} \mathcal{F}_n \; dS. \tag{48}$$

For the two-dimensional Euler equations, this is written, as

$$\tilde{\mathbf{z}}^t \mathbf{C} \, \hat{\mathbf{z}}_x + \tilde{\mathbf{z}}^t \mathbf{D} \, \hat{\mathbf{z}}_y = -\oint_{\partial T} \mathcal{F}_n \; dS \tag{49}$$

where we have introduced the notation,

$$\hat{\mathbf{z}}_x = \int_T \mathbf{z}_x \; dV \tag{50}$$

$$\hat{\mathbf{z}}_y = \int_T \mathbf{z}_y \; dV. \tag{51}$$

By symmetry, (49) can be written also as

$$\hat{\mathbf{z}}_x^t \mathbf{C} \, \tilde{\mathbf{z}} + \hat{\mathbf{z}}_y^t \mathbf{D} \, \tilde{\mathbf{z}} = -\oint_{\partial T} \mathcal{F}_n \; dS \tag{52}$$

or

$$\left(\hat{\mathbf{z}}_x^t \mathbf{C} + \hat{\mathbf{z}}_y^t \mathbf{D} \right) \tilde{\mathbf{z}} = -\oint_{\partial T} \mathcal{F}_n \; dS \tag{53}$$

where

$$\hat{\mathbf{z}}_x^t \mathbf{C} + \hat{\mathbf{z}}_y^t \mathbf{D} = \begin{bmatrix} (\hat{z_2})_x + (\hat{z_3})_y & (\hat{z_1})_x & (\hat{z_1})_y & 0 \\ \frac{\gamma-1}{\gamma}(\hat{z_4})_x & \frac{\gamma+1}{\gamma}(\hat{z_2})_x + (\hat{z_3})_y & \frac{\gamma-1}{-\gamma}(\hat{z_3})_x + (\hat{z_2})_y & \frac{\gamma-1}{\gamma}(\hat{z_1})_x \\ \frac{\gamma-1}{\gamma}(\hat{z_4})_y & \frac{\gamma-1}{-\gamma}(\hat{z_2})_y + (\hat{z_3})_x & \frac{\gamma+1}{\gamma}(\hat{z_3})_y + (\hat{z_2})_x & \frac{\gamma-1}{\gamma}(\hat{z_1})_y \\ 0 & (\hat{z_4})_x & (\hat{z_4})_y & (\hat{z_2})_x + (\hat{z_3})_y \end{bmatrix} \tag{54}$$

$$\mathbf{z} = \sqrt{\rho} \begin{bmatrix} 1 \\ u \\ v \\ H \end{bmatrix}, \qquad (\hat{z_i})_x = \int_T (z_i)_x \; dV. \tag{55}$$

This is a linear system for the average state $\tilde{\mathbf{z}}$, and therefore may be solved for the average state, provided the matrix $\hat{\mathbf{z}}_x^t\mathbf{C} + \hat{\mathbf{z}}_y^t\mathbf{D}$ is non-singular. Note that the contour-integral on the right must be computed consistently with the integration of the derivatives $\hat{\mathbf{z}}_x^t$ and $\hat{\mathbf{z}}_y^t$. Hence, *given an assumption on the variation of* $\mathbf{z}$ *over the element (linear, quadratic, etc), we compute* $\hat{\mathbf{z}}_x^t$ *and* $\hat{\mathbf{z}}_y^t$ *and the flux integral, and solve for the averaged state* $\tilde{\mathbf{z}}$*.*

An alternative view is found by writing this linear system as

$$\left[\int \left(\mathbf{z}_x^t\mathbf{C} + \mathbf{z}_y^t\mathbf{D}\right)\, dV\right]\tilde{\mathbf{z}} = \int \left(\mathbf{z}_x^t\mathbf{C} + \mathbf{z}_y^t\mathbf{D}\right)\mathbf{z}\, dV, \tag{56}$$

which is inverted formally to give

$$\tilde{\mathbf{z}} = \left[\int \left(\mathbf{z}_x^t\mathbf{C} + \mathbf{z}_y^t\mathbf{D}\right)\, dV\right]^{-1}\int \left(\mathbf{z}_x^t\mathbf{C} + \mathbf{z}_y^t\mathbf{D}\right)\mathbf{z}\, dV, \tag{57}$$

that is, the state we seek is a gradient-matrix-weighted average. In the case of linear elements, the matrix is constant, and therefore we find

$$\begin{aligned} \tilde{\mathbf{z}} &= \left[\int \left(\mathbf{z}_x^t\mathbf{C} + \mathbf{z}_y^t\mathbf{D}\right)\, dV\right]^{-1}\left[\int \left(\mathbf{z}_x^t\mathbf{C} + \mathbf{z}_y^t\mathbf{D}\right)\, dV\right]\int \mathbf{z}\, dV & (58)\\ &= \int \mathbf{z}\, dV & (59)\\ &= \bar{\mathbf{z}}, & (60) \end{aligned}$$

i.e. the familiar arithmetic average over the element.

Of course, in general, the matrix is not guaranteed to be invertible. It can even be zero. But in such a case, the average state is not needed anyway. It then appears reasonable to believe that the use of the correct average state in the distribution matrix is not important everywhere but important in some limited regions with high solution gradients such as shocks. Using this average state whenever possible, therefore, may prove to be of value.

