Reducing numerical dissipation in smoke simulation

Zhanpeng Huang a,*, Ladislav Kavan b, Weikai Li c, Pan Hui c, Guanghong Gong a

a Department of Automation Science and Electrical Engineering, Beihang University, Beijing, China
b Computer and Information Sciences Department, University of Pennsylvania, PA, USA
c Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong

Abstract

Numerical dissipation acts as artificial viscosity to make smoke viscous. Reducing numerical dissipation is able to recover visual details smeared out by the numerical dissipation. Great efforts have been devoted to suppress the numerical dissipation in smoke simulation in the past few years. In this paper we investigate methods of combating the numerical dissipation. We describe visual consequences of the numerical dissipation and explore sources that introduce the numerical dissipation into course of smoke simulation. Methods are investigated from various aspects including grid variation, high-order advection, subgrid compensation, invariant conservation, and particle-based improvement, followed by discussion and comparison in terms of visual quality, computational overhead, ease of implementation, adaptivity, and scalability, which leads to their different applicability to various application scenarios.

1. Introduction

Smoke is desirable in visual effect and video game industries. It is also one of challenging problems in computer graphics due to its complexity and turbulence. To obtain realistic smoke and gaseous phenomena, physically based methods with Navier–Stoke Equations (NSEs) have been explored to model underlying fluid dynamics. Although numerically integrating NSEs have been studied in computational fluid dynamics (CFD), computer graphics researches focus on simplified discretization and numerical schemes when visual quality matters most. Simplifications make physically based methods possible for smoke simulation but introduce the numerical dissipation. The numerical dissipation increases fluid viscosity to make it appear more viscous than intended. It degrades the visual appearance by smearing out fine details and damping down the motion quickly. The numerical dissipation has been recognized to have substantial visual consequences to the smoke simulation.

Many sources introduce numerical dissipation to the course of the smoke simulation. Coarse spatiotemporal discretization produces numerical truncation errors, which is proven to have a form of viscosity [1]. As fluid quantities are only defined on discrete locations such as grid points and particles, interpolation schemes are required to calculate values at undefined positions, which is equivalent to smoothing operations that produce the numerical dissipation. The semi-Lagrangian method [2] is widely used for the smoke simulation attributed to its unconditional stability and ease of implementation, but it generates a large amount of the numerical dissipation in backward tracing and advection subroutines. Many advanced methods are constructed based on the semi-Lagrangian method to guarantee the unconditional stability. However, they also inherit the disadvantage of massive numerical dissipation.

* This paper has been recommended for acceptance by Peter Lindstrom.
* Corresponding author.

http://dx.doi.org/10.1016/j.gmod.2014.12.002
1524-0703/© 2015 Elsevier Inc. All rights reserved.
Massive effort has been devoted to combat the numerical dissipation from different aspects. Some methods are developed to eliminate sources of the numerical dissipation. For instance, it is straightforward to reduce the numerical dissipation by increasing spatial resolution and reducing time step. However, both approaches increase computational overhead. Adaptive mesh [3], irregular mesh [4,5], and dynamical mesh [6] are proposed to reduce the numerical dissipation without significantly increasing computation. Rather than directly reducing the numerical dissipation, several methods generate artificial details to compensate for visual loss using vorticity confinement [7,8] and subscale turbulence models [9,10]. Particle-based methods only carry quantity but do not dissipate quantity, which does not suffer from the numerical dissipation problem. However, particle methods have problems such as particle redistribution. Hybrid particle and grid methods [11,12] are proposed to leverage advantages of particle and grid to reduce numerical dissipation.

In this paper, we investigate the numerical dissipation in smoke simulation in terms of where it comes out, what impact it has, and how to combat it. The rest of paper is organized as follows: we give a brief introduction to the basic smoke simulation in Section 2 and address the sources of the numerical dissipation in Section 3; in Section 4 we investigate and compare methods of combating numerical dissipation from different aspects, following by a conclusion in Section 5.

2. Background

Smoke and other gaseous phenomena are normally simplified to be incompressible and homogenous, which does not decrease the applicability to model basic dynamical mechanisms. The NSEs to model smoke are derived as:

\[
\begin{align*}
\frac{\partial \mathbf{u}}{\partial t} &= - (\mathbf{u} \cdot \nabla) \mathbf{u} + \nu \nabla^2 \mathbf{u} - \nabla p + \mathbf{f} \quad \text{(a)} \\
\nabla \cdot \mathbf{u} &= 0 \quad \text{(b)}
\end{align*}
\]

where \( \mathbf{u} \) is velocity, \( p \) and \( \rho \) denote pressure and fluid density respectively, \( \nu \) is kinematic viscosity to measure how viscous the fluid is and \( \mathbf{f} \) represents the resultant external force. The two equations indicate that the fluid should conserve both momentum and mass. The first equation is derived from Newton’s second law with left-hand term presenting acceleration and right-hand terms the net force exerted on fluid.

NSEs are too complicated to solve for analytical solution directly. The NSEs usually break down into simple terms including advection, pressure, diffusion, and external force [2]. The simple terms can then be easily solved individually. If we define the terms as operators denoted by \( A, P, D, \) and \( \mathbf{f} \), the operator \( \mathbf{S} \) to solve NSEs can be written as [13]:

\[
\mathbf{S} = \mathbf{P} \circ \mathbf{f} \circ \mathbf{D} \circ \mathbf{A}
\]

where

\[
\begin{align*}
A : \frac{\partial \mathbf{u}}{\partial t} &= -(\mathbf{u} \cdot \nabla) \mathbf{u} \\
P : \frac{\partial \mathbf{u}}{\partial t} &= \nu \nabla^2 \mathbf{u} \\
D : \frac{\partial \mathbf{u}}{\partial t} &= \mathbf{f} \\
\mathbf{f} : \frac{\partial \mathbf{u}}{\partial t} &= \frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot \mathbf{u} = 0
\end{align*}
\]

where \( q \) can be velocity, temperature, or any other fluid quantity.

3. Numerical dissipation as artificial viscosity

Numerical solutions are different from exact solution due to numerical truncation errors. The truncation errors include additional high-order terms which influence fluid motion and appearance. We start with the simple one-dimensional advection to analyze the impact on fluid motion:

\[
\frac{\partial q}{\partial t} + u \frac{\partial q}{\partial x} = 0, \quad u > 0
\]  

(4)

If we discretize it using forward Euler for the time derivative and first-order backward difference for the space derivative we can get:

\[
\frac{q_i^{n+1} - q_i^n}{\Delta t} + u \frac{q_i^n - q_{i-1}^n}{\Delta x} = 0
\]

(5)

We can rearrange it to get

\[
q_i^{n+1} = q_i^n - \Delta t \frac{q_i^n - q_{i-1}^n}{\Delta x} u
\]

(6)

Recalling the Taylor series for \( q_{i-1}^n \) gives

\[
q_i^{n+1} = q_i^n - \Delta t \left( \left( \frac{\partial q}{\partial x} \right)_i^n + \Delta x \left( \frac{\partial^2 q}{\partial x^2} \right)_i^n \right) u + O(\Delta x^3)
\]

(7)

Substituting it into above equation and doing the cancellation gives

\[
q_i^{n+1} = q_i^n - \Delta t \left( \left( \frac{\partial q}{\partial x} \right)_i^n + \Delta t \Delta x \left( \frac{\partial^2 q}{\partial x^2} \right)_i^n \right) u + O(\Delta x^2)
\]

(8)

Deleting the second-order truncation error and rewriting it gets

\[
q_i^{n+1} = q_i^n - \Delta t \left( \left( \frac{\partial q}{\partial x} \right)_i^n + \Delta x \left( \frac{\partial^2 q}{\partial x^2} \right)_i^n \right) u
\]

(9)

Which is the forward Euler in time applied to the modified PDE

\[
\frac{\partial q}{\partial t} + u \frac{\partial q}{\partial x} = u \Delta x \frac{\partial^2 q}{\partial x^2}
\]

(10)

The Laplacian of \( q \) in one dimension is \( \nabla^2 q = \frac{\partial^2 q}{\partial x^2} \). Defining \( \nu = u \Delta x \) and substituting it into the equation gives

\[
\frac{\partial q}{\partial t} + u \frac{\partial q}{\partial x} = \nu \nabla^2 q
\]

(11)
The modified advection equation includes an additional viscosity-like term. It is called artificial viscosity as it is purely of numerical origin without any physical meanings. Even though the fluid is inviscid, the artificial viscosity performs like physical viscosity to make the fluid viscous.

4. Combating numerical dissipation

4.1. Grid variation

The aforementioned artificial viscosity indicates the numerical dissipation is proportional to grid spacing $\Delta x$. In theory, we could reduce the numerical dissipation by refining the grid spacing. However, refining grid dramatically increases computational cost. Taking the widely used MAC grid for example, if the grid resolution increases by $\gamma$ times in three dimensions, the number of grid cells increases up to $\gamma^3$ times. As the grid spacing decreases by $\gamma$, time step has to reduce to $1/\gamma$ to meet Courant–Friedrichs–Lewy (CFL) condition (even for the unconditionally stable semi-Lagrangian method, the time step is usually set to be several times of CFL number in practice). The total computational cost rises up to $\gamma^4$ times, which makes the grid refinement method scale poorly and ill-suited for practical use.

To alleviate the problem, Berger and Oliger [14,15] introduced adaptive mesh refinement (AMR) into CFD to discretize local regions with different grid resolutions. Losasso et al. [3] proposed an octree structure to adaptively resolve irregular boundaries with high-resolution grid while using low-resolution grid at empty space. The method dramatically improves visual effects around irregular boundaries with a small increment of computational cost. Fig. 1 shows visual comparison between the basic method and the octree-based method with similar grid resolutions. As the grid is adaptively refined around irregular boundary, the method captures more visual details than traditional uniform grid method. However, it increases the complexity of grid structure. The additional computational cost pays off only when grid refinements are limited to a few local regions. In addition, non-uniform grid constructs a nonsymmetric linear system to solve for pressure, which may cause potential numerical instability.

As hexahedral grid does not align to horizontal and vertical boundaries, it cannot resolve irregular boundaries that do not align to coordinate axes. Tetrahedral grid is proposed to represent irregular boundaries with much less numerical errors. Elcott et al. [4] discretized and solved NSEs on tetrahedral grid. The method can present complicated boundaries but has problem of energy dissipation, which can be modified using implicit time-reversible velocity integration on the tetrahedral grid [16]. Tetrahedral grid has intrinsic complicated structure to increase the computational cost. A hybrid method is proposed to combine the hexahedral and tetrahedral structures [5]. The tetrahedral grid is used to resolve irregular boundaries while the hexahedral grid is used at empty regions without obstacles. Fig. 2 gives visual comparison of smoke simulations using different discretization schemes. As most interesting features are generated around irregular boundaries, the hybrid grid is able to obtain similar visual quality as tetrahedral method near boundaries. As the hybrid grid uses uniform grid mesh in most empty spaces, it reduces the total computational cost comparable with the uniform grid method.

Instead of the static grid structure, several works reduce computational overhead by translating and reconstructing grid mesh according to boundary conditions. Shah et al. [6] tracked smoke movement with movable and scalable grid mesh. Empty regions without smoke are not discretized in order to decrease computational cost. To guarantee visual consistence between simulation domain and empty regions, Cohen et al. [17] used a simple particle system outside of the simulation domain. Velocity at domain boundary is assigned as initial conditions of the simple particle system. Zhu et al. [18] extended the uniform grid to create a large far-field grid. The uniform grid maintains fine resolution to capture detailed features in interesting regions while extended coarse grid is used to obtain large field view of regions that does not require fine presentation. Fig. 3 compares three different grid configurations. In left figure, the grid is translated along bulk motion of smoke. Only regions filled with smoke are computed to reduce computational overhead. The smoke region expands as smoke moves around, which can be partitioned into two subregions as illustrated in the middle figure. The smoke subregion is calculated using physical methods to capture major body of smoke while the less important subregion is modeled with simple non-physical methods to reduce computation. In order to simulate smoke from a

Fig. 1. Visual comparison of smoke simulations using octree grid (left) [3] structure and MAC (right) [7].
far distance, the right figure extends coarse cells around fine grid to obtain a large-field view while maintaining fine resolution at local regions of interest. Other than translating and extending the grid mesh, Feldman et al. [19] and Klingner et al. [20] dynamically deformed tetrahedral grid mesh to adapt to variable boundaries and moving obstacles. All adaptive grid methods have common drawbacks of high complicated structure and high computational cost.

A full description of truncated error includes both $\Delta x$ and $\Delta t$ terms. Reducing time step also reduces numerical dissipation at the cost of increasing computational overhead. In smoke simulation, we have to carefully choose the time step that does not produce strange results and breaks the numerical stability.

4.2. Advection improvement

The semi-Lagrangian advection is predominant attributed to its simplicity and unconditional stability. The method is widely used as the basic blocks to construct other advanced advection schemes. The basic semi-Lagrangian advection is given as:

$$u'(x, t + \Delta t) = u^n(x - u^n(x, t)\Delta t, t)$$

(12)

which means that the velocity at location $x$ at next time step is the velocity at location $(x - u^n(x, t)\Delta t)$ at current time step $t$. As current velocity is always bounded to previous velocities, it guarantees unconditional stability.

4.2.1. High-order backward tracing

To get the previous position where current quantity ends up at inquired grid point at the next time step, the basic advection employs the forward Euler scheme to trace the trajectory in reverse to calculate previous position, leading to certain numerical dissipation. Some high-order schemes such as Runge–Kutta can be used to improve
the accuracy. Bridson [1] recommended a second-order Runge-Kutta for backward position tracing. Fig. 4 shows the visual results using the first-order forward Euler and the second-order Runge-Kutta. The Runge-Kutta scheme produces more curly details at local regions.

The backward tracing only introduces a fractional part of numerical dissipation to the entire simulation. Boosting accuracy of the backward tracing with high-order schemes limits visual improvement to the final results. In addition, the high-order schemes require more computational cost. The second-order Runge-Kutta is regarded as a default scheme to balance numerical accuracy and computational cost.

4.2.2. High-order interpolation

Most likely the previous position is not on grid points where quantities are defined and stored. We have to interpolate nearby grid points to obtain a good approximation. The linear interpolation is predominant due to its simplicity and stability, but it suffers from massive numerical dissipation because the linear interpolation is similar to the low-pass filter. Catmull–Rom interpolation has second-order numerical accuracy, but it likely overshoots locally and drops into an unstable feedback loop to make simulation blow up finally. Simply preventing overshoot produces additional numerical dissipation in the vicinity of local minima and maxima [1]. Fedkiw et al. [7] clamped slopes to zero where slope signs are flipped. The modified scheme is locally monotone to guarantee stability and second-order accuracy. However, the method is excessive since it flattens overshoots which may be just fine. Huang et al. [21] proposed to adaptively flatten overshoots that break global bounds in order to maintain diversity of fluid features. As it does not require full-scale suppression, the scheme also reduces computational overhead. A visual comparison of smoke using linear, modified, and adaptive Catmull–Rom interpolations is shown in Fig. 5.

There are numerous high-order interpolations in CFD [22], but they are not widely used in smoke simulation for several reasons. As all quantities on grid points require interpolation, a small computational increment of interpolation will significantly increase total computational overhead. In [7], time cost of the monotonic scheme is about 18 times larger than the linear interpolation. Even though the adaptive scheme reduces computational cost, it is about 3–4 times larger than the linear scheme [21]. In addition, the high-order interpolations require more points to construct stencil, which makes it difficult to handle inner boundary conditions. Wide stencil is also complicated to handle adaptive and non-uniform grids. High-order interpolations require limiters to avoid oscillations, new extrema, and possible instability [23]. Similar to backward tracing, interpolation is not the major contribution to the total numerical dissipation, but it likely increases total computational overhead. We have to choose interpolation scheme based on numerical accuracy versus runtime performance trade-off curve.

4.2.3. High-order advection

The basic semi-Lagrangian advection is popular for smoke simulation. However, excessive numerical dissipation makes it ill-suited for simulation that requires high visual quality. Kim et al. [24,25] applied the Back and Forth Error Compensation and Correction (BFECC) to the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.

The backward advection implies that the equation is able to evolve backward, which is inappropriate for parabolic and non-reversible partial differential equations. The BFECC requires great computational cost as it calculates the semi-Lagrangian Courant–Isaacsion–Rees (CIR) to obtain second-order accuracy both in time and space. Forward and backward semi-Lagrangian schemes are compared to estimate numerical error of a single semi-Lagrangian advection, which is then used to compensate a third forward semi-Lagrangian advection. Fig. 6 shows visual comparisons using the semi-Lagrangian and the BFECC advections. The BFECC generates much more detailed features that the basic semi-Lagrangian scheme, especially vortical features that are swept out by the numerical dissipation.
methods are unconditionally stable. However, the back and forth round step makes them inappropriate to work near interface of multiphase flows. It may not conserve momentum if the first forward advection goes across to the other side fluid with different density and then comes back with the same velocity. The same problem happens near boundaries of obstacles and computational domain. It is suitable to replace them with the semi-Lagrangian scheme near these regions. In the semi-Lagrangian method time step is normally set to be less than 5 times of CFL condition number, so it is safe to turn off the BEFCC and the modified MacCormack at 5Ax off the interface and boundaries.

Other than advecting velocity, Molemaker et al. [26] proposed the QUICK scheme to advect velocity fluxes. QUICK [27] is a third-order accurate explicit scheme. It produces numerical dissipation only at finest spatial scale that can be resolved by the grid. Fig. 8 illustrates the visual results with different advects. The QUICK scheme generates much more details than the basic semi-Lagrangian method, especially on high-resolution grid.
The first-order forward Euler is unstable for the QUICK [28]. It can be replaced with other temporal schemes with large stability domains such as second-order Adams–Bashforth (AB2) [29]. As the QUICK is conditionally stable only when it meets CFL condition, it greatly increases computational cost. Attributed to its regular memory access pattern, it normally requires less computational cost than other semi-Lagrangian based schemes with careful memory access.

Apart from advecting velocity flux, several methods use constrained interpolation profile (CIP) [30,31] to advect both the velocity and its derivatives, which can boost numerical accuracy to third order. However, original CIP requires a compact stencil and it is conditionally stable with tight CFL condition. The method does not consider derivatives at farthest cell corner, so it only guarantees numerical accuracy when backward-traced point is not far from the start corner. To improve the numerical stability, Song et al. [32] proposed the monotonic CIP (MCIP) by suppressing potential overshoots along all dimensions. The excessive suppression obtains unconditional stability but introduces some numerical dissipation. The dimension-splitting strategy also dramatically increases computational cost [33]. As splitting computation is related to axis directions, it normally produces some numerical dissipation. The problem is even serious in shear motions [34]. Kim et al. [35] proposed an unsplit semi-Lagrangian CIP (USCIP) to guarantee unconditional stability. Two additional symmetrical terms are added to consider off-axis movements such as shear and rotation to avoid unnecessary wiggles. It is able to reduce the numerical dissipation with relative low computational overhead. Fig. 9 compares visual results by using various advection schemes. The BFECC and the MCIP obtain more compelling visual results than the basic semi-lagrangian schemes, while the USCIP method produces higher visual quality than other methods. CIP-based schemes advect both velocity and derivatives. Their computational cost is comparable with the BFECC method, but much higher than the basic semi-Lagrangian method. As USCIP does not require dimension splitting, it requires less computation than the BFECC and the USCIP methods.

To illustrate how the numerical dissipation disperses and smears out details, we compared several unconditional stable advections using the standard Zalesak disk test [36]. The benchmark is a $100 \times 100$ grid with a slotted disk centered at $(50, 75)$ of 15 cells in radius, 5 cells in width, 25 cells in length of slot and 5 cells of an upper bridge. A constant vorticity velocity field is given as [37]:

\[
\begin{align}
 u &= (\pi/314)(50 - y) \\
 v &= (\pi/314)(x - 50)
\end{align}
\]

so that the slotted disk finishes one revolution every 628 time steps.

Fig. 10 shows results of one revolution with different advection schemes. An accurate advection scheme is able to translate and rotate interface without distortion and deterioration, especially at interface with high curvature. The basic semi-Lagrangian method has too much numerical dissipation to maintain basic shape of the disk. The BFECC and MacCormack schemes keep generally shape but smooth slots and corners. The MCIP scheme conserves the disk shape with certain erosion at inner and outer corners, while USCIP almost suffers no distortion and area loss.

Table 1 lists area loss of the disk with different schemes and grid configurations. The “loss” column represents percentage of the area loss. Negative value means the final disk is smaller than the original disk while positive value indicates the disk grows. With the same grid configuration, high-order schemes such as the USCIP suffer from much less area loss than low-order schemes such as the semi-Lagrangian. With the same advection, the area loss also decreases as grid resolution increases. As high-order schemes are able to suppress numerical dissipation even at relative low-resolution grids, the improvement gain is not significant by increasing grid resolutions when using high-order schemes. While first-order semi-Lagrangian method has too much numerical dissipation on coarse grid, using fine grid significantly reduces total numerical dissipation. We can also find that the amount of loss with the semi-Lagrangian method drops relatively much faster than that of high-order schemes, which indicates that low-order schemes are much more sensitive to the grid resolution than high-order schemes. Table 2 compares discussed advects in terms of numerical stability, numerical accuracy, and computational cost.

Since the backward tracing and interpolation schemes are orthogonal, aforementioned high accurate tracking schemes in Section 4.2.1, high-order interpolations in Section 4.2.2 and high-order advections in Section 4.2.3 can be integrated together to further improve the numerical accuracy. In addition, other smoke quantities including density, temperature, and concentration can also be advected with above methods to reduce total numerical dissipation. However, if high-order schemes are excessively used for all quantities, high-frequency components of quantities will accumulate simultaneously to produce highly sharp

Fig. 8. Simulations using the semi-Lagrangian advection (left) and the QUICK advections with low- and high-resolution grid configurations [26].
interface. For instance, Fig. 11 shows two 2D smoke simulations using different advection configurations. Left image shows result using a combination of the modified MacCormack scheme for velocity and the semi-Lagrangian for density; right image uses the modified MacCormack scheme for both velocity and density. Although the right one has much less numerical dissipation for velocity and density, they accumulate to produce clear sharp interface, which makes it artificial against the real smoke.

4.3. Sub-grid compensation

The numerical dissipation increases the viscosity to make smoke inadequate of highly energetic and turbulent appearance. Rather than directly reducing the numerical dissipation by changing discretization schemes and simulation algorithms, many works focus on how to add small-scale features back with heuristic models.

4.3.1. Vorticity confinement

Fedkiw et al. [7] introduced the vorticity confinement method [39] into smoke simulation. The method injects back a certain amount of energy to recover turbulent features by adding a tuneable body force. The force is perpendicular to the direction of vorticity maximum magnitude to enhance local rotation. It is designed to be linear proportional to the grid spacing so that the modified NSEs degenerate to the original NSEs as the additional term vanishes in the limit. Selle et al. [11] coupled the vorticity confinement with vorticity equation. The vorticity confinement force can be synthesized directly from current vorticity field. The left two images of Fig. 12 compare visual results using the basic and the vorticity confinement methods. Injecting some vortical force back into flow significantly improves visual effects, but it also introduces some random artificial turbulence at local regions.

The vorticity confinement is an inexpensive approach to recover small features eliminated by the numerical dissipation.
dissipation. However, the method cannot determine how much energy should be injected back without causing instability. Fig. 12 gives comparisons by injecting different amount of vorticity confinement forces. The simulation becomes quasi-unstable and degenerates to a random turbulent chaos when too much force is added. The problem lies in the fact that the coefficient is constant over the entire simulation domain. To solve the problem, He et al. [8] presented an adaptive vorticity confinement coefficient according to helicity other than user-defined constant value. Fig. 13 illustrates the visual comparison between the constant and variable vorticity confinement methods. They further developed a robust second vorticity confinement method [40] to guarantee the numerical stability even with large confinement coefficients.

Vorticity confinement methods add kinetic energy to dilute energy diffusion caused by the numerical dissipation, but they cannot fully compensate for excessive numerical dissipation [26]. It is inappropriate for low viscous fluid that inertia is so well conserved that fluid motion can propagate for a significant distance without being damped and broken down.

4.3.2. Subscale turbulence model

Many other methods use high-level turbulence models to compensate for sub-grid details loss due to the numerical dissipation. The basic mean flow is simulated on coarse grid while turbulent details are synthesized using procedural synthesis methods. The methods can be described as [41]:

$$\mathbf{u} \leftarrow NS(\mathbf{U}) \oplus ST(\mathbf{u}')$$  \hspace{1cm} (14)

where $NS$ is a basic fluid simulator to calculate mean velocity $\mathbf{U}$ on the coarse grid, $ST$ is a synthesizer to generate turbulent field $\mathbf{u}'$ and $\oplus$ is the integration operation to couple the two fields together.

Early methods [42–45] used Kolmogorov spectrum to generate pseudo-random turbulence in frequency space. These methods have a common drawback of generating turbulence at incorrect regions. Recently curl operation [46] is applied on Perlin [9,47] and wavelet vector noises [48] to generate turbulent components at various frequencies and scales. Energy cascade is modeled to include spatial distribution and turbulence dynamics using local assembled wavelets [48], linear $k-\varepsilon$ equation [47], one-equation [9], and complete tow-equation of $k-\varepsilon$ [49,10]. Instead of directly layering turbulence components on the basic flow, Zhao et al. [41] modeled fluctuations as controllable turbulence force to agitate the basic flow. The method guarantees temporal consistence with the basic NSEs simulators. Rather than calculating the basic flow, Gregson et al. [50] employed a multi-scale tracking method to reconstruct temporally coherent velocity field from prior low-resolution density images. However, the
method requires fine tuning to obtain good balance between reconstructed details and noise levels. Fig. 14 gives visual comparisons between different turbulence enhancement methods. It shows that more turbulent details are generated when more turbulence components are added to the basic flow. However, smoke appears chaotic if too much turbulence components are added. In order to obtain realistic detailed features, trial and error effort are required to determine the amount of turbulence components.

Procedural turbulence methods decouple high-frequency components from the basic flow and use synthesized small-scale turbulence to compensate for fine feature loss caused by the numerical dissipation. The methods enable animation artists to generate a fast basic simulation and then add details without changing gross motion, which helps the artists to shorten the turnaround. However, as the finest scale is determined by local velocity and temporal averaging rather than the grid resolution, using a much high-resolution grid is able to produce more realistic visual results than the non-physical synthesis methods. Additionally, the temporal averaging operations may smooth out visual appearance to some extent.

4.4. Invariants conservation

Another baseline of reducing the numerical dissipation is to conserve basic fluid invariants. In rigid and deformable body simulations, great improvements can be achieved by following the linear and angular momenta conservation principles. Similar strategy can be applied in fluid simulation by conserving invariants including vorticity, circulation, and energy. Elcott et al. \cite{4,51} proposed to conserve circulation along arbitrary simplicial mesh. As vortex is an important visual cue of smoke, the method uses the back-tracking integrals to solve the vorticity formulation of NSEs. The method preserves discrete circulation on grid mesh to achieve no vorticity numerical

![Fig. 13. Simulations using the vorticity confinement methods with constant (right) \cite{8} and variable coefficients (left) \cite{7}.](image)

![Fig. 14. Visual effects of different turbulent enhancement methods. (a) Original coarse simulation; (b) wavelet subgrid turbulence; (c) controllable and intermittent turbulence method; (d) Add vorticity confinement to (a); (e) wavelet turbulence to (d); (f)–(h) controllable and intermittent turbulence method with $q = 0.8, 0.2$ and $0.1$ respectively, where $q$ is control parameter to calculate resultant velocity field with $qU + (1 - q)u$ \cite{41}.](image)
dissipation. As shown in Fig. 15, the conservation method is able to produce similar visual results as real smoke while the basic semi-Lagrangian method loses most fine features due to the excessive numerical dissipation.

Since the method is constructed based on the semi-Lagrangian advection, it is also unconditional stability. In addition, the vorticity-based simulation requires no pressure calculation. Intrinsic storage without reference to global and local coordinate system constructs a very sparse linear system, which greatly decreases computational cost comparable with the semi-Lagrangian method. However, the method preserves the circulation but does not conserve energy. Although a $L^2$ projection [51] can be employed to preserve the total vorticity energy, the global operation propagates influence to the entire domain. In addition, the projection induces artificial visual effects around boundaries.

In fluid simulation, the energy loss relates to time step, grid spacing, and total frame numbers. The factors are coupled in the simulation course, making it difficult to eliminate energy dissipation by tuning each single term. Besides, the computational cost is sensitive to the spatial resolution and the time step. In order to preserve energy without significantly increasing computational cost, Mullen et al. [16] proposed a fully Eulerian integration scheme independent of spatiotemporal resolutions. The energy conservation scheme [52,53] and the non-dissipative advection ensure no energy loss on arbitrary simplicial grids. In addition, it guarantees unconditional numerical stability. Fig. 16 shows energy dissipation of different schemes over time. The energy is tightly related with the smoke motion. High energetic smoke generally obtains rich and long-time active details. As shows in the figure, the basic semi-Lagrangian method dissipates most energy over time, while the energy conservation method almost diffuses no energy, which indicates it does not have the numerical dissipation problem.

The Eulerian integration preserves energy exactly over time. However, it inevitably suffers from certain numerical diffusion due to the low-pass filtering operations of resampling on the discrete grid, which is a common drawback of grid-based methods.

4.5. Particle based methods

Pure Eulerian methods start with a field sampled on grid points and end with advected field resampled on the grid at each step. The Eulerian methods may reduce numerical dissipation to some extent, but suffer from the fundamental problem of field resampling: the scale of

---

Fig. 15. 2D vortices merging comparisons between the real experimental result (a), the conservation method (b), and the basic semi-Lagrangian method (c) [4].
finest features is limited by the Nyquist frequency of the grid. Even for a pure translation velocity field, the highest spatial frequency that can be reliably advected has period of $4\Delta x$ \cite{1}. A good Eulerian scheme may filter out high-frequency components as a smoothing operation while a bad Eulerian scheme may further generate visual artifacts.

In contrast to the Eulerian methods, the Lagrangian methods employ discrete particles to carry variables along without dissipation. There is no filtering and dissipation loss in advection. From this point of view, Lagrangian particle methods are perfect for advection to eliminate numerical dissipation. In fact, there are already some pure particle methods for fluid simulation, such as smoothed particle hydrodynamics (SPH) \cite{54} and some pure vortex particle methods \cite{55,56,57,58,59,60}. Several vortex particle methods employ geometry structures to describe vortex structure including filament \cite{56,57}, ring \cite{58,59} and sheet \cite{60} to capture various subtle features such as leapfrogging vortex rings and vortex shedding, which are difficult to simulate using the grid-based methods. However, apart from their specified problems such as particle redistribution, pure particle methods cannot enforce incompressible condition as efficient and accurate as the grid methods. Besides, the particle-based methods require sophisticated particle distribution to obtain good performance and quality.

A preferable remedy is to add Lagrangian machinery to the Eulerian scheme in a hybrid way. Selle et al. \cite{11} combined Lagrangian vortex particle and Eulerian grid together to overcome the weakness of both methods. The vortex particles are used to carry and preserve vorticity without dissipation loss while the Eulerian grid is employed to calculate velocity and vorticity confinement force \cite{7} to recover swirling details. As the vortex particles are only used to increase visual details, no particle distribution is required. Yoon et al. \cite{61} proposed a similar scheme with combination of the vortex particles and Eulerian grid. Basic flow is calculated on relative coarse grid to reduce the computational cost. Particle vorticity is transferred to a high-resolution vorticity field to obtain fine details. The method uses a kernel function to calculate vorticity field on high-resolution grid points, which cannot guarantee the vorticity field free of divergence. In addition, as all vortex particles make their contributions to vorticity forces, it likely synthesizes strong vorticity forces to produce visual artifacts. The vortex particles are suitable to preserve swirling features but require additional computations to induce velocity from vorticity. A straightforward modification is to directly store velocity on particles. However, it is far from enough as pressure projection requires velocity information of the whole domain, which means we cannot just consider individual particles but account for particle–particle interactions on top of advections. Particle-in-cell (PIC) \cite{62} is a good strategy to couple particle and grid while solving the pressure projection problem. Velocity and other quantities are defined and advected on particles to eliminate the numerical dissipation, but they are calculated and adjusted by pressure gradients on the grid points. The method reduces the numerical dissipation of advection but introduces even more to the total numerical dissipation of the simulation because quantities are averaged to the grid points from the particles and then interpolated back to the particles from the grid points. Zhu and Bridson \cite{12} proposed the fluid implicit particle (FLIP) method to reduce the total numerical dissipation. The FLIP interpolates the changes rather than the quantity values on the
grid points. It adds the changes to the quantity value other than replacing it. The changes are smoothed but not accumulated to achieve almost no numerical dissipation. Fig. 17 shows the visual comparison of smoke using different methods. The FLIP method obtains much more visual details than both the basic semi-Lagrangian method and the PIC method.

As only changes are transferred in the FLIP method, the velocity fluctuations on particles may average down to zero on the grid points and then show up as unexpected perturbations at other time steps, leading to certain noise. Since PIC has no such problem, a better solution is to combine them together with a tunable regularization parameter to decay the noise and suppress the numerical dissipation.

5. Conclusion

In this paper, we discussed the numerical dissipation problem for smoke simulation in computer graphics community. The numerical dissipation is widely regarded to have visual consequences of visual results. Many methods have been developed to solve the problem. The methods reduce numerical dissipation or compensate for visual loss from various aspects. Most methods fall into three categories: “Decreasing”, “Conserving”, and “Increasing”. Table 3 compares methods in the categories in terms of computational cost, visual quality, implementation, adaptability, and scalability. The adaptability indicates whether the method is adaptive to different application scenarios. The scalability measures how it is easy to integrate the method with existing applications. The following list gives the common characteristics of methods in each category, which has its advantages and disadvantages to make them suitable for different application scenarios including high-quality visual effect, real-time performance, animation control, dynamic domain, obstacle-coupled interaction, and large-scale simulation, as illustrated in Fig. 18.

- “Decreasing”: Solving the problem by reducing the amount of numerical dissipation. The grid variation methods in Section 4.1 and the advection improvement methods in Section 4.2 fall into this category. The “decreasing” methods normally use fine-resolution discretization and high-order numerical schemes to reduce the total amount of the numerical dissipation. As grid
resolution is directly related to the final visual effect, grid variation methods have strong applicability in applications with high-quality effect requirement. The grid-based methods are flexible to handle boundary and obstacle problems, making grid variation methods suitable for applications with dynamic domain and obstacle-coupled interaction. However, great efforts are required to implement spatial discretization and guarantee the numerical stability on irregular grid structures. The heavy computational cost also degrades the runtime performance, making them ill-suited for realtime applications. As computation is exponentially proportional to the domain dimensions, tremendous computations and memory storage are required for large-scale simulations. Besides, most calculations and storage strategy tightly depend on the grid structures. Changing grid structures requires great changes in the simulation, while the advection improvement methods are generally orthogonal to other subroutines in the pipeline, the modifications will not impact other subroutines, so the advection improvement methods are much easier to integrate with existing applications than grid variation methods. In addition, the advection improvement methods are compatible with the grid-based methods. They are feasible to handle the dynamical domain and obstacle boundary problems with the help of grid-based methods. Comparing with the grid variation methods, advection improvement methods improve the visual quality with relative less computational overhead. They are preferable for real-time applications that do not require very high-quality results.

- "Conserving": Solving the problem by conserving indirect fluid quantities. The invariants conservation methods in Section 4.4 and the particle methods in Section 4.5 are two typical "conserving" methods, while they preserve quantities in different ways. The invariants conservation methods integrate indirect fluid quantities (e.g. circulation and energy) on specific structures (e.g. simplicial mesh) to avoid simulating direct quantities that greatly impact fluid motions. As direct quantities such as velocity are not directly involved in calculations that produce the numerical dissipation, the numerical dissipation is significantly reduced, making them especially suitable for applications with high-quality requirement. Especially, as no energy is dissipated for the energy conservation methods, smoke remains energetic to guarantee long-time performances. The invariants conservation methods are usually constructed on the grid structures, so they may inherit the advantages of the grid variation methods. However, they need sophisticated strategies to guarantee conservation, which impeding their adaption and stability to other application scenarios. The particle methods preserve quantity by leveraging the conservation characteristic of the particle. The particle-based methods have totally different simulation pipelines from grid-based methods, leading to different applicabilities. The particle-based methods are difficult to handle boundary problems such as obstacle interaction and dynamic domain. Hybrid particle and grid methods alleviate the problems. The particle based methods do not require some computational intensive subroutines,
which reduces the computational cost to make them suitable for realtime applications. Comparing to the grid point number of grid-based methods, particle number is more relaxant to the domain dimensions, making particle methods applicable to large-scale simulations with less computational cost.

- **“Increasing”:** Solving the problem by increasing additional components. The compensation methods in Section 4.3 are typical “increasing” methods as they add components back into flow to synthesize detailed features. The vorticity confinement methods are usually directly integrated into the grid methods as an additional term. They need no modification to original grid structure and do not impact original simulation. As computational cost is inexpensive, the vorticity confinement methods are able to increase diversity of flow motion with much fewer efforts. They have the same applicability as the grid-based methods. Most sub-grid compensation methods employ the grid-based methods for the basic flow simulations, so they have similar abilities to the grid-based methods. However, the requirement to guarantee consistency between the basic flow and the sub-grid synthesis flow degrades their abilities. Especially, computational cost increases significantly when complicated turbulence methods are used on sub-grid structure to generate turbulent details. A great advantage of the “increasing” methods is that the synthesis procedure is decoupled from the basic simulation. The synthesis procedure can be treated as a postprocessing without influencing the gross motion, which makes the “increasing” methods suitable for animation control. However, as turbulent details are generally generated using synthesis methods, the details may appear artificial and less comparable with the grid-based methods.

In most time, we struggle to eliminate the numerical dissipation, while occasionally we leverage it to guarantee the numerical stability. As smoke is typically simplified as inviscid fluid, there is no viscous friction at the dissipative range of energy cascading. If there was no numerical dissipation acting as artificial viscosity, energy would pile up unbounded at the finest scales (with highest wavenumber as $k = 2\pi/\Delta x$) and cause numerical instability [49]. Additional dissipations are required to maintain numerical stability. The popular semi-Lagrangian scheme is so much dissipative over a wide range of energy spectrum that it degrades the visual quality. As we have observed in the full Eulerian integration method [16], visual results may look unnatural if there is no numerical dissipation. Adding a certain amount of numerical dissipation will make it a little viscous and appear plausible as real smoke.

References


[38] Z.P. Huang, G.H. Gong, I. Han, Physically-based modeling, simulation and rendering of fire for computer animation, Multimedia Tools Appl. (2012).


