〔特別講演〕

Ideas and Issues behind the Genesis of some Effective CFD Methods and Models

*Metacomp Technologies, Inc., Sukumar Chakravarthy †
Metacomp Technologies, Inc., Uriel Goldberg
Metacomp Technologies, Inc., Sampath Palaniswamy
Metacomp Technologies, Inc., Paul Batten
Metacomp Technologies, Inc., Oshin Peroomian

We explore the nature of how CFD algorithms, methods and mathematical models are created. Sometimes a general scientific principle finds its way into CFD and evolves over time to fulfill a variety of needs. At other times, a CFD researcher faces a specific problem or issue and this leads to an invention or development that helps overcome the problem. We provide many examples of both types of patterns which have motivated, in our own work, the creation of some practical and successful CFD approaches.

1. Introduction

The purpose of this paper is to illustrate some patterns underlying the development of Computational Fluid Dynamics (CFD) methods. We find two interleaving trends. Some methods were first developed by importing some general scientific principle (GSP). Other methods began through a more direct motivation to solve a particular issue seen in the then current methodology (NBD). We consider both mathematical modeling approaches as well as numerical algorithms in this exploration. Independent of the nature of the origin of the specific approach considered, finally the methods contribute to a successful formulation when the GSP-based method helps satisfy some CFD need(s) and when the NBD method is based on a sound scientific approach and is not too ad hoc.

A single paper on this topic cannot cover the entire field. Needless to say, we base our discussions on a personal perspective. Nevertheless, we would like to offer the following list as comprising reasonable examples of this paper's topic.

- The direct use of polynomials in CFD methods, multidimensional polynomials, and least squares interpolation and how these unify the treatment of regular (conformal), patched, and overset meshes.
- The role of the Riemann Initial Value Problem and the Riemann Initial and Boundary Value Problem in the formulation of interior and boundary treatments.
- The need for non-oscillatory interpolation and the development of TVD and ENO schemes.
- The need for dealing with discontinuities and the link between front fitting, capturing and tracking.
- Some of the motivations behind the development of topology-parameter-free (wall-distance-free or pointwise) turbulence models.
- Additional thoughts on the evolution of turbulence models to meet evolving needs as well as observations regarding performance of existing models.
- The role of synthetic turbulence in hybrid RANS/LES approaches.
- The need for various types of hybridization in both numerical and mathematical modeling; e.g. the hybridization of first order and higher order discretizations (locally), the hybridization of the 1-equation and 2-equation turbulence models, hybridization of RANS and LES, etc.
- The evolution of unifying themes and treatments in CFD: the unification of time and space marching; the unified treatment of different meshing approaches; the unification of the treatment of incompressible and compressible flows, etc.

In all these, the paper tries to point out the interweaving connections between the underlying idea, the numerical algorithm or modeling approach and the CFD needs they help satisfy. The two tables given in this paper serve to encapsulate these relationships. We leave it to the reader to guess if the GSP or the NBD came first for each attribute represented.

2. Numerical methodology

Early CFD methods were developed using truncation error analysis and were tailored based on Taylor series expansion of difference representations. Often the same difference representations could be arrived at by beginning with a polynomial representation of the data wherein the polynomial coefficients were constructed from the discrete values available. The direct use of the polynomial representation decidedly got a boost from Bram van Leer's work [1] which connected spurious extrema directly to the underlying phenomenal

*28632 Roadside Drive, #255, Agoura Hills, CA 91301, USA
† E-mail: src@metacomptech.com

representation. In our own work, we preferred the polynomial point of view and developed methods based on multidimensional polynomial representations [2]. These were particularly effective on unstructured meshes.

Another general scientific principle is that of least-squares-interpolation. Standard polynomials of degree *p* would need a suitable set of *p* + 1 data values to determine the coefficients. However, when there is more data available than the number of coefficients, the least-squares fit can be employed. When polynomial representations are sought to be used in CFD, linear polynomials can lead to 2nd order accuracy (at least in an "inviscid" sense). The number of coefficients then is 2 in 1D, 3 in 2D, and 4 in 3D for the simplest representation (e.g. f = a + b x + c y + d z in 3D). These coefficients can be evaluated at each cell by fitting the cell values of the dependent variables along with values at a collection of neighbors. Even on structured grids, the collection of a central cell and its face neighbors can have more cells in the neighborhood than the number of polynomial coefficients. For example, in a 2D structured grid, a cell has 4 face-neighbors (5 cells total) but the number of polynomial coefficients for each variable is only 3. Therefore least squares interpolation can come in very handy.

In unstructured grids, the number in each cell's neighborhood collection can be rather arbitrary and once again the least squares approach works. In our work, we have further used multidimensional least squares interpolation to interpolate across multiblock boundaries by including in the definition of cell neighborhood cells in the vicinity from all overlapping blocks.

Another important trend in the evolution of accurate and robust numerical methods is the role that the mathematical idea of the Riemann Initial Value problem (IVP) for hyperbolic equations began to play in method development. Often shortened as the "Riemann Problem", this IVP was based on an infinite domain split into two (left and right) each with constant states. The Riemann Problem was the IVP with 2 sets of constant states that were allowed to interact at time t = 0. The evolution of the solution (wave interactions) was captured by an analytical solution if possible. This solution procedure was the "Riemann Solver" and various approximate solutions became the "Approximate Riemann Solvers". There was a similarly strong role that the Riemann Initial and Boundary Value Problem (IBVP) could play in developing boundary treatments but this needs more space to elaborate and is left to a future discussion.

The link between the polynomial basis and the Riemann Solver approach is strong. In order to apply the Riemann Solver to the construction of numerical methods, the Riemann problem had to be considered locally where to the left and right of every cell face was a local Riemann problem. This corresponded to a piecewise constant polynomial representation. Once such an analogy was made, it was easy to consider and construct a hierarchy of schemes based on the piecewise constant behavior replaced by polynomial representations of increasing degree and by considering the wave interactions in a semi-discrete sense, in order to somewhat circumvent the need for a detailed consideration of wave interactions for piecewise varying data representations (as opposed to piecewise constant).

It will also be interesting to explore the link between shock-capturing and shock-fitting techniques using the Riemann Solver as a central piece in the discussion. Exact Riemann solvers exactly represent the wave interactions emanating from initial discontinuities and do in fact include and use the Rankine-Hugoniot relationships for shock waves. We leave it to the reader to pursue this line of reasoning further. Another intriguing idea is how all this is related to the use of level sets to advance discontinuity fronts.

3. Turbulence Modeling

In this section, we consider the overall topic of **Realizability enhancements to the Low-Reynolds-Number k-ε Turbulence Model** in order to bring out several of the points enumerated in the Turbulence Modeling Table.

Time-scale realizability

For the past 30 years the following transport equation has been applied to the dissipation rate of the turbulence kinetic energy:

$$\rho\frac{D\varepsilon}{Dt}=\frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial\varepsilon}{\partial x_j}\right]+C_{\varepsilon 1}\frac{\varepsilon}{k}P_k-C_{\varepsilon 2}f_2\rho\frac{\varepsilon^2}{k} \tag{1}$$

where the production term is given by

$$P_k=\left[\mu_t\left(\frac{\partial U_i}{\partial x_j}+\frac{\partial U_j}{\partial x_i}-\frac{2}{3}\frac{\partial U_k}{\partial x_k}\delta_{ij}\right)-\frac{2}{3}\rho k\delta_{ij}\right]\frac{\partial U_i}{\partial x_j} \tag{2}$$

and the eddy viscosity is

$$\mu_t=C_\mu f_\mu\rho\frac{k^2}{\varepsilon} \tag{3}$$

f_μ being a near-wall damping function discussed later. In the immediate vicinity of solid surfaces Eq. (1) reduces to

$$\mu\frac{\partial^2\varepsilon}{\partial y^2}=C_{\varepsilon 2}f_2\rho\frac{\varepsilon^2}{k} \tag{4}$$

where y is normal-to-wall direction. Since $\varepsilon\propto$ const. and $k\propto y^2$, it follows that $f_2\propto y^2$ is necessary to keep Eq. (4) in balance as a wall is approached. On the other hand, $f_2\Rightarrow 1$ toward the boundary layer edge is necessary in order to regain the high-Re form of Eq. (1). Obviously, there is no unique formulation with the above asymptotic limits; a considerable number of f_2 formulations are available in the literature. For example, Myong and Kobayashi [3] suggest

$$f_2=\left[1-\frac{2}{9}e^{-\left(\frac{R_t}{6}\right)^2}\right]\left(1-e^{-\frac{y^+}{5}}\right)^2,\quad R_t=\frac{k^2}{\nu\varepsilon},\quad y^+=\frac{yu_\tau}{\nu},\quad u_\tau=\sqrt{\left(\frac{\tau}{\rho}\right)_w}. \tag{5}$$

Evidently, every such formulation will affect the sink term, $C_{\varepsilon 2} f_2 \rho \varepsilon^2 / k$, in its own unique way, thereby increasing the level of ambiguity for the ε transport equation. However, the non-uniqueness inherent in formulating f_2 is artificial since the necessity for this near-wall correction function is a result of using an *incorrect time-scale*. This is evident when Eq. (1) is rewritten as

$$\rho \frac{D\varepsilon}{Dt} = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right] + \left(C_{\varepsilon 1} P_k - C_{\varepsilon 2} f_2 \rho \varepsilon\right)\frac{\varepsilon}{k}. \qquad 1A)$$

The inverse time-scale, ε/k, is that of the large, energy-containing eddies; it does not represent correctly the time-scale of the dissipative, near-wall eddies. These eddies are represented by the Kolmogorov time-scale, $t_K = \sqrt{\nu/\varepsilon}$. By invoking a time-scale which represents both types of eddies, the 1/k singularity [Eq. (1A)] is eliminated, and so is the ambiguous function f_2. This *realizable* time-scale is given as [2]

$$t_R = \frac{k}{\varepsilon}\max\left\{1, \sqrt{\frac{2}{R_t}}\right\} \qquad (6)$$

and it has the correct asymptotic limits both near and away from walls, with a smooth transition in between. Finally, the version of the ε equation incorporating the realizable time-scale is

$$\rho \frac{D\varepsilon}{Dt} = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right] + \left(C_{\varepsilon 1} P_k - C_{\varepsilon 2} \rho \varepsilon\right)\frac{1}{t_R}, \qquad (1B)$$

and the f_2 function is eliminated.

Improving performance under adverse pressure gradient conditions

The classical k-ε closure tends to under-predict the size of flow separation bubbles due to excessive eddy-viscosity which causes shrinkage of reversed flow regions. This, in turn, results from under-prediction of the magnitude of ε. To correct this tendency and improve the model's performance under adverse pressure gradient flow conditions, an extra source term was proposed for the ε transport equation [5]. There is no unique way of formulating such a source term, however, it must be restricted to the inner portion of boundary layers and *must not* impose any influence on the outer portion; failing the latter will cause unphysical predictions of the viscous/inviscid interface at shear layer edges.

The extra source term in the k-ε closures in CFD++ is as follows.

$$E = A_E \rho \sqrt{\varepsilon t_R}\, \max\left\{\sqrt{k}, (\nu\varepsilon)^{1/4}\right\} \max\left\{\frac{\partial k}{\partial x_j}\frac{\partial (k/\varepsilon)}{\partial x_j}, 0\right\}. \qquad (7)$$

It should be observed that this term includes both time- and velocity-scale realizability attributes (the latter is the $\max\left\{\sqrt{k}, (\nu\varepsilon)^{1/4}\right\}$ portion). Previous proposals (see in [3]) involved the term $\max\left\{\frac{\partial k}{\partial x_j}\frac{\partial (k^{3/2}/\varepsilon)}{\partial x_j}, 0\right\}$ i.e. the gradient of the turbulence length-scale rather than that of the time-scale as in Eq. (7). Fig. (1) shows the effect of such a choice.

The length-scale variant retains too much influence across the boundary layer, resulting in an unphysical kink in the velocity field due to incorrect turbulence entrainment at the shear layer edge (inward rather than outward); the time-scale variant is limited to the inner portion of the boundary layer, thus enabling correct entrainment and avoiding unrealistic kinks. The final form of the ε equation thus reads

$$\rho \frac{D\varepsilon}{Dt} = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right] + \left(C_{\varepsilon 1} P_k + E - C_{\varepsilon 2} \rho \varepsilon\right)\frac{1}{t_R}. \qquad (1C)$$

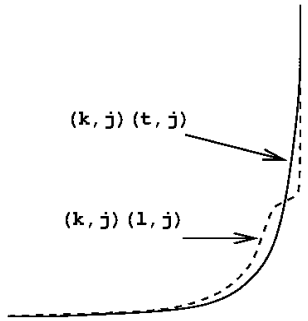

Figure 1 Velocity profiles using $\frac{\partial k}{\partial x_j}\frac{\partial \tau}{\partial x_j}$ (solid) and $\frac{\partial k}{\partial x_j}\frac{\partial \ell}{\partial x_j}$ (dashed)

Eddy-viscosity field and near-wall damping function

The eddy-viscosity is usually expressed in the form

$$\hat{\mu}_t = C_\mu f_\mu \rho \frac{k^2}{\varepsilon} \qquad (8)$$

Where $C_\mu = 0.09$ and f_μ is the near-wall damping function to be discussed below. Eq. (8) gives rise to excessive levels of turbulence kinetic energy (k) in stagnation zones, such as a wing's leading edge. In reality, no k should exist in such flow regions since the solid surface damps out the turbulence in the stagnation zone. To enforce this, yet another realizability constraint, the Schwartz limiter, must be invoked. This limiter, based on the Schwartz inequality, states that no Reynolds shear stress can be larger than the square root of the product of the corresponding normal stresses: $\overline{u'v'} \le \left(\overline{u'u'} \cdot \overline{v'v'}\right)^{1/2}$. Translated into eddy-viscosity language, this becomes

$$\mu_t = \min\left\{\hat{\mu}_t, \varphi \frac{\rho k}{|S|}\right\} \qquad (9)$$

where |S| is mean strain magnitude and φ=1/3 or 2/3, depending on the assumed relationship between k and the three normal stresses in homogeneous turbulence.

Finally, the near-wall damping function also includes the time-scale realizability constraint:

$$f_\mu = \frac{1 - e^{-A_\mu R_t}}{1 - e^{-\sqrt{R_t}}}\max\left\{1, \sqrt{\frac{2}{R_t}}\right\}, \quad R_t = \frac{k^2}{\nu\varepsilon}, \quad A_\mu \cong 0.01. \qquad (10)$$

This damping function has been used successfully in several turbulence models available in CFD++ thanks to the fact that (a) it is a function of only the turbulence Reynolds number, R_t and (b) there is only one calibration parameter, A_μ.

The pointwise approach

It will be observed that the entire formulation discussed above is independent of topography parameters; this is unlike many other models which resort to distance from walls as an integral part of their formulation. The topography-parameter-free approach [6] (also called the pointwise approach) is the underlying method used in all the turbulence closures in CFD++ developed at Metacomp Technologies. Highlights of this modeling approach are given here. Fig. (2) demonstrates the ambiguity of the wall-distance-based approach in the case of intersecting walls. Multiple choices for the representative wall distance are available and each choice will lead to different values for the flow variables between the walls; the situation will get worse as the corner is approached.

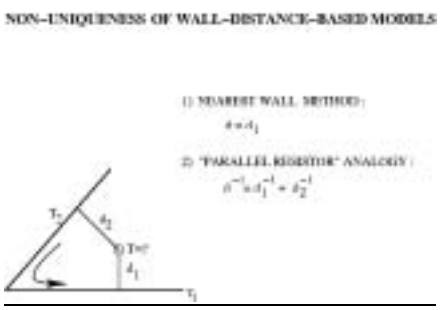

Figure 2 Non-uniqueness of wall-distance-based approach

On the other hand, the wall-distance-free approach, seen in Fig. (3), does not suffer from this problem. Rather than resorting to wall distance, the method is based on several *wall proximity indicators* which are field variables and have unique values at each point in the flowfield. One of these, R_t, was used in the definition of the realizable time-scale {Eq. (6)] and in the near-wall damping function [Eq. (10)]. Another, indicated as "2" in the figure, constitutes a unique measure of distance from walls. Finally, the indicator labeled "3" is very
useful in differentiating between near-wall and "away from wall" regions. Typically the dot product will change sign when going from one region to the other, thus serving as an indicator to activate or inactivate certain terms. In the case of the extra source term for the ε equation [Eq. (7)] this type of indicator was used to ensure that the term was active only in the immediate vicinity of solid surfaces.

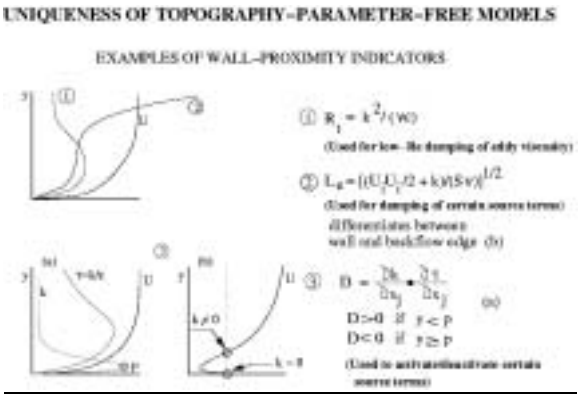

Figure 3 Uniqueness of wall-distance-free approach

4. Conclusion

In this paper, we have seen several examples of general-scientific-principle-based and need-based evolution of numerical methods. In our own work and contemplative musings, we have sought to discover unifying principles and procedures covering a broad spectrum of our activities.

Many of the hybridizations treated above are also examples of unification. Other unifications we mentioned in the introduction. It would be interesting to trace the threads between such unifications. For brevity, we consider only one such tread here. Reimann solvers lead to generalized upwind formulations which resulted in a certain type of non-linear diagonal dominance that could be exploited to build relaxation methods for hyperbolic problems [7]. In conjunction with the possibilities of relaxation methods for elliptic problems, and the use of non-symmetric relaxation (successive or Gauss-Seidel), this leads to the unification of time and space marching. We look forward to expanding on these topics in the future.

5. References

1. B. Van Leer, "Towards the Ultimate Conservative Difference Scheme, V. A Second Order Sequel to Godunov's Method", *J. Com. Phys.*., **32**, 1979, pp. 101–136.
2. O. Peroomian and S. Chakravarthy, "A Grid-Transparent Methodology for CFD", AIAA Paper 97-0724, Jan 1997.
3. H. K. Myong and T. Kobayashi, "Prediction of Three-Dimensional Developing Turbulent Flow in a Square Duct with an Anisotropic Low-Reynolds-Number k-ε Model," *ASME J. Fluid Engrg.,* **113,** Dec. 1991, pp. 608-615.
4. U. Goldberg and D. Apsley, "A Wall-Distance-Free low Re k-ε Turbulence Model," *Computer Methods Appl. Mech. Engrg,* **145,** 1997, pp. 227-238.
5. U. Goldberg, O. Peroomian and S. Chakravarthy, "A Wall-Distance-Free k-ε Model with Enhanced Near-Wall Treatment," *ASME J. Fluid Engrg,* **120,** Sept. 1998, pp. 457-462.
6. U. Goldberg and S. Chakravarthy, "A k-ℓ Turbulence Closure for Wall-Bounded Flows," AIAA Paper No. 2005-4638, 35th AIAA Fluid Dynamics Conference, 6-9 June 2005, Toronto, Canada.
7. S.Chakravarthy and S. Osher, "Computing with High- Resolution Upwind Schemes for Hyperbolic Equations", *Lectures in Applied Mathematics*, Vol 22, 1985, pp. 57-86.

Table 1 Numerical Modeling

ATTRIBUTE	NEED SATISFIED	GENERAL PRINCIPLE
Unstructured mesh	• Complex topology • Short turnaround processes • Mesh adapted to solution	• Local topology defines the mesh • Optimal shapes for filling volume • Grid point addition and deletion impact the mesh locally
Multi-dimensional interpolation	• No co-ordinate directions in unstructured mesh • Field gradients are not always aligned with coordinate directions • Accuracy and robustness	• Multidimensional representations reduce truncation error
Riemann Solver	• Many flow features are masked by numerical smoothing • Spurious oscillations trigger phenomena not otherwise present in the actual solution • Entropy condition satisfaction	• Upwinding and matrix based diffusion provides smoothing just enough for a robust solution • Riemann solver is a systematic way for arriving at the solution that satisfies the entropy condition
Preconditioning	• Numerical diffusion overwhelms the solution when overall flow speed is low compared to the speed of sound (M<<1)	• Reducing the eigenvalue spread of the diffusion operator leads to a higher quality solution
Relaxation with multi-grid acceleration	• Need for many steady state solutions in the design iteration demands tens or hundreds of solutions within short time • Convergence to steady state is mesh dependent and suffers as the volume ratio of largest to the smallest cells increase • Convergence rate independent of mesh size	• Upwinding matrix operator can be used to relax the hyperbolic system of equations to steady state solution • A good multi-grid removes convergence stall in meshes with large volume ratios and in very large meshes
Dual time stepping for implicit time update	• Time march the solution field at time intervals decided by transient events in the field • Time accurate solutions at low mach numbers	• Using a dual-time stepped implicit method removes stability limits of explicit time updates. The solution can evolve at an interval decided purely based on solution transients • Time accurate evolution of preconditioned methods can be achieved with dual time stepping. Only the inner iterations that drive the solution to the next time level are preconditioned
Sliding/moving meshes	• Flow around and within bodies in relative motion	• Generalized interpolation and inter-connection book keeping leads to conservative and continuous representation for fields variables
6-DOF module	• Trajectory of bodies in flight • Unsteady response to control surface actuation • Safety of store separation from aircraft	• Tightly coupled rigid-body and fluid dynamics produces accurate transient response
Sub-models for features that are not resolved	• Stator row to deflect flow as in thrust reversers and swirlers • Mass/Fuel injection • Flame fronts	• Models are needed when the phenomena cannot be resolved due to lack of grid resolution

Table 2 Turbulence Modeling

ATTRIBUTE	NEED SATISFIED	GENERAL PRINCIPLE
Topography-parameter-free turbulence modeling	• Unstructured mesh • Parallel computing • Moving mesh • Corner flow ambiguity	• Real turbulence "detects" wall influence through proximity – NOT by *measuring* distance to wall. Hence "wall proximity indicators"
Separated/reversed flow prediction capability	• Backflow model to augment Baldwin-Lomax • Rt closure • Additional source term in ε-equation (see below)	• No turbulence model works well in all flow types; tailor-made enhancements are needed to improve performance (e.g., wall-bounded/free shear, attached/separated)
Realizability constraints	• Avoid unrealistic Reynolds stresses ($u_i'u_i' < 0$) • Avoid turbulence generation where it doesn't belong (e.g., stagnating flow) • Avoid *ad-hoc* near-wall correction functions	• Schwartz/Bradshaw inequality: $-\overline{u_i'u_j'} \le \sqrt{\overline{u_i'u_i'}\cdot\overline{u_j'u_j'}}$ • Time- and velocity-scale realizability (Kolmogorov limits): $T_t = \max\left\{\frac{k}{\varepsilon}, \sqrt{\frac{\nu}{\varepsilon}}\right\}$ $U_t = \max\left\{\sqrt{k}, (\nu\varepsilon)^{1/4}\right\}$
Synthetic turbulence	• Introduced to prevent turbulence kinetic energy from dissipating into heat when moving from large eddy resolving to non-resolving grid zones (i.e., DES). With synthetic turbulence a portion of the energy is retained in the coarse mesh region – much more realistic (LNS)	• Turbulence energy goes back and forth between generation and dissipation. Synthetic turbulence approximates real life perturbations
Hybridization concepts	• Some models decay freestream turbulence while others don't. A hybrid version combines these attributes to conform to users' needs (e.g., k-ε-R_t vs. k-ε) • Some models are better near walls, others away from walls; a hybrid version combines these strengths (e.g., SST) • RANS/LES hybrids permit faster unsteady flow computation turnaround than pure LES due to limiting fine mesh regions only where needed – not in the entire domain	• Turbulence possesses a vast number of scales; no single closure can resolve them all. A hybrid approach increases the level of scale resolution