# Lecture Notes: Navier-Stokes Equations and Their Relation to LLM Diffusion Models

Francesco Orsi

July 6, 2025

**Abstract**

These lecture notes provide a comprehensive overview of the Navier-Stokes Equations (NSE), fundamental to fluid dynamics, and explore their surprising conceptual and mathematical parallels with modern Large Language Model (LLM) diffusion models. We delve into the derivation and physical interpretation of the NSE, discussing their complexities and the ongoing millennium prize problem. Subsequently, we introduce the theoretical underpinnings of diffusion models, focusing on stochastic differential equations (SDEs) and their role in generative AI. The core of these notes lies in drawing analogies between the flow dynamics described by NSE and the probabilistic flow in diffusion models, highlighting how concepts like conservation laws and transport phenomena find echoes in the denoising process of generative models.

# Contents

# 1 Introduction to Navier-Stokes Equations

The Navier-Stokes Equations (NSE) are a set of partial differential equations (PDEs) that describe the motion of viscous fluid substances. They are central to fluid dynamics, providing a mathematical framework for understanding phenomena ranging from weather patterns and ocean currents to airflow over an aircraft wing and blood flow in arteries.

## 1.1 Historical Context

The equations were independently derived by Claude-Louis Navier in 1822 and George Gabriel Stokes in 1845. They represent a fundamental application of Newton's second law of motion to fluid continua, incorporating concepts of momentum conservation, mass conservation, and energy conservation.

## 1.2 The Millennium Prize Problem

One of the most significant challenges in mathematics and physics is the existence and smoothness of solutions to the Navier-Stokes equations, particularly for three-dimensional incompressible flow. The Clay Mathematics Institute has offered a \$1 million Millennium Prize for a proof of either existence and smoothness or a counterexample.

# 2 Formulation of the Navier-Stokes Equations

The NSE are derived from the application of conservation laws to a continuum fluid. We typically consider two main forms: the momentum equation and the continuity equation (mass conservation).

## 2.1 Conservation of Mass (Continuity Equation)

For an incompressible fluid, the density $\rho$ is constant. The continuity equation states that mass is conserved:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

where $\mathbf{u}$ is the fluid velocity vector. For an incompressible fluid ($\rho =$ constant), this simplifies to:

$$\nabla \cdot \mathbf{u} = 0$$

This condition implies that the fluid flow is divergence-free, meaning there are no sources or sinks of fluid within the domain.

## 2.2  Conservation of Momentum (Momentum Equation)

The momentum equation is essentially Newton's second law ($\mathbf{F} = m\mathbf{a}$) applied to a fluid element. It describes how the change in momentum of a fluid particle is due to forces acting on it. The forces include pressure gradients, viscous stresses, and external body forces (like gravity).

The general form for a Newtonian fluid is:

$$\rho\left(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u}\right) = -\nabla p + \mu \nabla^2 \mathbf{u} + \mathbf{f}$$

Let's break down each term:

- $\rho\frac{\partial \mathbf{u}}{\partial t}$: Local acceleration term. Describes the rate of change of velocity at a fixed point in space.

- $\rho(\mathbf{u} \cdot \nabla)\mathbf{u}$: Convective acceleration term. Describes the acceleration due to the movement of a fluid particle from one point to another where the velocity is different. This term is non-linear and is a major source of complexity in the NSE.

- $-\nabla p$: Pressure gradient term. Pressure forces act from regions of high pressure to low pressure, causing fluid acceleration.

- $\mu \nabla^2 \mathbf{u}$: Viscous term. $\mu$ is the dynamic viscosity of the fluid. This term accounts for internal friction within the fluid due to velocity gradients. For inviscid fluids, $\mu = 0$, and the equations simplify to Euler equations.

- $\mathbf{f}$: External body forces per unit volume (e.g., gravity, electromagnetic forces).

Combining the incompressible continuity and momentum equations, we get the incompressible Navier-Stokes Equations:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{\rho}\nabla p + \nu \nabla^2 \mathbf{u} + \frac{1}{\rho}\mathbf{f}$$

$$\nabla \cdot \mathbf{u} = 0$$

where $\nu = \frac{\mu}{\rho}$ is the kinematic viscosity.

## 2.3  Boundary Conditions

To solve the NSE, appropriate boundary conditions are required. Common types include:

- **No-slip condition**: For viscous fluids, the fluid velocity at a solid boundary is equal to the velocity of the boundary itself. If the boundary is stationary, $\mathbf{u} = 0$.

- **Inlet/Outlet conditions**: Specify velocity or pressure at the fluid's entry or exit points.

- **Free surface conditions**: For interfaces between fluids (e.g., water-air), conditions on stress and continuity of velocity are applied.

# 3 Properties and Challenges of NSE

## 3.1 Non-linearity

The convective term $(\mathbf{u} \cdot \nabla)\mathbf{u}$ is non-linear, which makes the NSE notoriously difficult to solve analytically. This non-linearity is responsible for complex phenomena like turbulence.

## 3.2 Turbulence

At high Reynolds numbers (ratio of inertial forces to viscous forces), fluid flow becomes turbulent, characterized by chaotic, unpredictable, and highly dissipative motion across a wide range of scales. Simulating or predicting turbulent flows is one of the grand challenges in computational fluid dynamics.

## 3.3 Existence and Smoothness

The Millennium Prize Problem specifically asks whether smooth solutions to the 3D incompressible NSE exist for all time given smooth initial conditions, or if singularities (points where solutions become infinite) can form.

# 4 Introduction to Diffusion Models in Machine Learning

Diffusion models are a class of generative models that have recently shown remarkable success in generating high-quality images, audio, and more recently, text. They operate by learning to reverse a gradual diffusion process that transforms data into noise.

## 4.1 The Forward Diffusion Process

In the forward process, data (e.g., an image $\mathbf{x}_0$) is progressively noised over $T$ steps, typically by adding Gaussian noise. This creates a sequence of noisy latent variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$, where $\mathbf{x}_T$ is pure noise. The process can be described by a Markov chain:

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I})$$

where $\beta_t$ is a small noise variance schedule. A key property is that $\mathbf{x}_t$ can be directly sampled from $\mathbf{x}_0$:

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I})$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

## 4.2 The Reverse Diffusion Process

The goal of a diffusion model is to learn the reverse process, i.e., to denoise $\mathbf{x}_t$ back to $\mathbf{x}_{t-1}$ (and eventually to $\mathbf{x}_0$). This reverse process is also a Markov chain, and its transitions are learned by a neural network (often a U-Net for images).

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

The neural network $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ is trained to predict the noise added at each step. The mean $\boldsymbol{\mu}_\theta$ is then derived from this predicted noise.

## 4.3 Stochastic Differential Equations (SDEs) and Ordinary Differential Equations (ODEs)

The discrete diffusion process can be generalized to continuous time using SDEs. The forward process can be written as:

$$dx = f(x, t)dt + g(t)dw$$

where $\mathbf{f}(\mathbf{x}, t)$ is the drift coefficient, $g(t)$ is the diffusion coefficient, and $d\mathbf{w}$ is a Wiener process (Brownian motion). The reverse process is also an SDE:

$$dx = \left( f(x, t) - g(t)^2 \nabla_x \log p_t(x) \right) dt + g(t) d\bar{w}$$

where $d\bar{\mathbf{w}}$ is a reverse-time Wiener process. The term $\nabla_\mathbf{x} \log p_t(\mathbf{x})$ is the score function, which the neural network learns to approximate.

Alternatively, the reverse process can be transformed into an ODE, known as the Probability Flow ODE:

$$dx = \left( f(x, t) - \frac{1}{2} g(t)^2 \nabla_x \log p_t(x) \right) dt$$

This ODE allows for deterministic sampling from the model, enabling faster generation and exact likelihood computation.

# 5 LLM Diffusion Models

While initially popular for images, diffusion models are being adapted for discrete data like text. This often involves:

- **Continuous embeddings**: Mapping discrete tokens to continuous vector spaces where diffusion can occur.

- **Categorical diffusion**: Directly defining diffusion processes on discrete spaces.

- **Latent diffusion**: Applying diffusion in a compressed latent space learned by an autoencoder, which can be applied to text embeddings.

The core idea remains: learn a reverse process to denoise a noisy representation back to a meaningful text sequence.

# 6 The Relation Between Navier-Stokes Equations and Diffusion Models

At first glance, fluid dynamics and generative AI seem disparate. However, a deeper look reveals profound conceptual and mathematical parallels, particularly when considering the continuous-time formulation of diffusion models.

## 6.1 Flow and Transport Phenomena

- **NSE: Physical Flow of Fluid** The NSE describe the physical flow of a fluid, where particles move under the influence of pressure gradients, viscous forces, and external forces. The velocity field $\mathbf{u}(\mathbf{x}, t)$ dictates how fluid elements are transported through space and time. Conservation laws (mass, momentum) govern this transport.

- **Diffusion Models: Probabilistic Flow of Data** Diffusion models describe a probabilistic flow of data in a high-dimensional latent space. The forward process is a "diffusion" of data into noise, while the reverse process is a "denoising flow" that transforms noise back into structured data. The SDEs governing this process describe how the probability density $p_t(\mathbf{x})$ of the data evolves over time.

## 6.2 The Fokker-Planck Equation and Continuity

The evolution of the probability density function $p(\mathbf{x}, t)$ for an SDE is governed by the Fokker-Planck (or Kolmogorov Forward) Equation. For the forward SDE $d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}$, the Fokker-Planck equation is:

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\mathbf{f}p) + \frac{1}{2}\nabla^2(g^2 p)$$

This equation is a conservation law for probability density. It has a striking resemblance to the continuity equation in fluid dynamics, which describes the conservation of mass. If we define a probability current $\mathbf{J} = \mathbf{f}p - \frac{1}{2}\nabla(g^2 p)$, then the Fokker-Planck equation becomes:

$$\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{J} = 0$$

This is precisely the form of the continuity equation for mass, where $p$ acts as the "density" and $\mathbf{J}$ as the "mass flux".

## 6.3 Drift and Velocity Fields

- **NSE: Velocity Field u** The velocity field $\mathbf{u}(\mathbf{x}, t)$ in NSE directly represents the instantaneous velocity of fluid particles.

- **Diffusion Models: Drift Term and Score Function** In diffusion models, the drift term $\mathbf{f}(\mathbf{x}, t)$ in the SDE (or the effective drift in the Probability Flow ODE) can be seen as an "average velocity" or "flow field" that guides the data points. The score function $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ is a crucial component of this drift, pointing towards regions of higher probability density. It acts like a "force" that pulls noisy data points towards the data manifold.

## 6.4 Viscosity and Noise/Diffusion

- **NSE: Viscosity $\nu$** Viscosity in NSE represents the internal friction within the fluid, leading to dissipation of kinetic energy and smoothing of velocity gradients. It acts as a regularizing term.

- **Diffusion Models: Diffusion Coefficient $g(t)$ and Noise** The diffusion coefficient $g(t)$ in SDEs governs the amount of random noise added (forward process) or removed (reverse process). This noise acts as a form of "stochasticity" or "random walk" that allows the model to explore the data space and escape local minima during training. In the context of the reverse process, the noise effectively "smooths" the path from noise to data, analogous to how viscosity smooths fluid flow. The $\nabla^2$ term in NSE (Laplacian) is also related to diffusion processes in physics (e.g., heat equation, Fick's law of diffusion).

## 6.5   Non-linearity and Complexity

- **NSE: Convective Non-linearity** The $(\mathbf{u} \cdot \nabla)\mathbf{u}$ term in NSE is non-linear and leads to complex phenomena like turbulence.

- **Diffusion Models: Neural Network Non-linearity** The neural network $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ that learns the score function is a highly non-linear function. This non-linearity allows diffusion models to capture intricate, high-dimensional data distributions and generate diverse and realistic samples. The complexity of the data manifold in high dimensions can be seen as analogous to the complex flow patterns in turbulent fluids.

## 6.6   Potential for Cross-Pollination

The conceptual links suggest potential for insights to flow between the fields:

- **Fluid Dynamics for AI**: Techniques developed for solving or analyzing NSE (e.g., numerical methods, turbulence modeling, spectral methods) might inspire new architectures or training strategies for diffusion models, especially for high-dimensional data.

- **AI for Fluid Dynamics**: Diffusion models could potentially be used to learn complex fluid dynamics, generate turbulent flow fields, or even aid in solving the NSE by learning the underlying dynamics from data. For instance, diffusion models could generate high-fidelity turbulent flow snapshots, or learn to predict future states in a probabilistic manner.

# 7   Conclusion

The Navier-Stokes Equations and LLM diffusion models, despite originating in vastly different scientific domains, share deep conceptual and mathematical connections rooted in the theory of continuous-time stochastic processes and conservation laws. Both describe complex "flows"—one of physical matter, the other of probability density in abstract data spaces. Understanding these parallels not only enriches our appreciation for the universality of mathematical principles but also opens exciting avenues for interdisciplinary research, potentially leading to breakthroughs in both fluid dynamics and generative AI. The challenges of non-linearity and high-dimensional complexity are central to both fields, making cross-domain insights particularly valuable.