Introduction to Data Assimilation
Lecture 9

Quanling Deng*

Abstract

This lecture focuses on linear Gaussian SDEs.

Keywords Rynolds decomposition, statistical equilibrium, decorrelation time

1 Reynolds Decomposition and Time Evolution of Moments

Consider the linear Gaussian stochastic differential equation (SDE)

$$dX_t = (-aX_t + f)\,dt + \sigma\,dW_t, \tag{1.1}$$

where $a > 0$ is the damping rate, $\sigma > 0$ is the noise amplitude, f is a constant external forcing, and W_t denotes the Wiener process.

This model is often referred to as the *Ornstein–Uhlenbeck process* (OU process) when $f = 0$. It is a prototypical example of a linear Gaussian SDE.

1.1 Reynolds decomposition

At any fixed time t, the random variable X_t can be decomposed as

$$X_t = \langle X_t \rangle + X_t', \tag{1.2}$$

where $\langle X_t \rangle$ denotes the ensemble mean (expectation), and X_t' is the fluctuation component satisfying

$$\langle X_t' \rangle = 0, \qquad \mathrm{Var}(X_t) = \langle (X_t')^2 \rangle.$$

The variance can be expressed in terms of the second raw moment as

$$\langle (X_t')^2 \rangle = \langle X_t^2 \rangle - \langle X_t \rangle^2. \tag{1.3}$$

*Yau Mathematical Science Center, Tsinghua University, Beijing, China 100048. E-mail address: qldeng@tsinghua.edu.cn

Proof. Derivation is given below:

Starting from the Reynolds decomposition,

$$X_t = \langle X_t \rangle + X_t', \tag{1.4}$$

we have the fluctuation part

$$X_t' = X_t - \langle X_t \rangle. \tag{1.5}$$

By definition, the variance is the ensemble average of the squared fluctuation,

$$\mathrm{Var}(X_t) = \left\langle (X_t - \langle X_t \rangle)^2 \right\rangle = \langle (X_t')^2 \rangle. \tag{1.6}$$

Expanding the square term gives

$$\begin{aligned} (X_t')^2 &= (X_t - \langle X_t \rangle)^2 \\ &= X_t^2 - 2X_t \langle X_t \rangle + \langle X_t \rangle^2. \end{aligned} \tag{1.7}$$

Taking the ensemble average of both sides yields

$$\langle (X_t')^2 \rangle = \langle X_t^2 \rangle - 2\langle X_t \langle X_t \rangle \rangle + \langle \langle X_t \rangle^2 \rangle. \tag{1.8}$$

Since $\langle X_t \rangle$ is a deterministic number at any fixed time t, it can be taken outside the ensemble operator:

$$\langle X_t \langle X_t \rangle \rangle = \langle X_t \rangle^2, \tag{1.9}$$

$$\langle \langle X_t \rangle^2 \rangle = \langle X_t \rangle^2. \tag{1.10}$$

Substituting these into the expression above gives

$$\begin{aligned} \langle (X_t')^2 \rangle &= \langle X_t^2 \rangle - 2\langle X_t \rangle^2 + \langle X_t \rangle^2 \\ &= \langle X_t^2 \rangle - \langle X_t \rangle^2. \end{aligned} \tag{1.11}$$

Thus, the variance can be expressed in terms of the second raw moment as

$$\boxed{\mathrm{Var}(X_t) = \langle X_t^2 \rangle - \langle X_t \rangle^2.} \tag{1.12}$$

□

1.2 Applications and Background of Reynolds Decomposition

Reynolds decomposition separates a stochastic or turbulent quantity into its mean and fluctuation parts:

$$X_t = \langle X_t \rangle + X_t',$$

where $\langle X_t \rangle$ is the ensemble mean and X_t' is the zero-mean fluctuation. It provides a foundation for analyzing mean dynamics and random variability.

Applications.

- **Turbulence modeling:** Introduced by Osborne Reynolds (1895) in fluid dynamics, leading to the Reynolds-Averaged Navier–Stokes (RANS) equations and the concept of Reynolds stress.
- **Stochastic systems:** Used to derive deterministic moment equations for mean and variance in SDEs, separating predictable trends from random fluctuations.
- **Geophysical and data assimilation models:** Helps describe mean circulation and transient eddies, and underpins ensemble-based filtering methods such as the Kalman filter.

Remark. Reynolds decomposition is conceptually similar to separating "signal" and "noise" in data: the mean captures the systematic dynamics, while the fluctuation quantifies uncertainty or turbulence.

1.3 Equation for the mean

Taking the ensemble average $\langle\cdot\rangle$ of (1.1), and using $\langle dW_t\rangle = 0$, yields

$$\frac{d}{dt}\langle X_t\rangle = -a\langle X_t\rangle + f. \tag{1.13}$$

This is a deterministic ODE whose solution is

$$\langle X_t\rangle = \langle X_0\rangle e^{-at} + \frac{f}{a}(1 - e^{-at}). \tag{1.14}$$

As $t \to \infty$, the mean approaches the equilibrium value

$$\overline{X}_\infty = \frac{f}{a}.$$

1.4 Equation for the variance

Subtracting (1.13) from (1.1), the fluctuation dynamics are

$$dX_t' = -aX_t'\,dt + \sigma\,dW_t. \tag{1.15}$$

We now derive the time evolution of $\langle (X_t')^2\rangle$.

Applying Itô's lemma to $Y_t = (X_t')^2$ gives

$$\begin{aligned} dY_t &= 2X_t'\,dX_t' + (dX_t')^2 \\ &= 2X_t'(-aX_t'\,dt + \sigma\,dW_t) + \sigma^2\,dt \\ &= (-2a(X_t')^2 + \sigma^2)dt + 2\sigma X_t' dW_t. \end{aligned} \tag{1.16}$$

Taking the ensemble average and noting that $\langle X_t' dW_t \rangle = 0$, we obtain

$$\frac{d}{dt}\langle (X_t')^2 \rangle = -2a\langle (X_t')^2 \rangle + \sigma^2. \tag{1.17}$$

Let $\langle (X_0')^2 \rangle = \mathrm{Var}(X_0)$ be the initial variance. Solving (1.17) gives

$$\langle (X_t')^2 \rangle = \mathrm{Var}(X_0)e^{-2at} + \frac{\sigma^2}{2a}(1 - e^{-2at}). \tag{1.18}$$

As $t \to \infty$, the variance approaches the equilibrium value

$$\mathrm{Var}(X)_\infty = \frac{\sigma^2}{2a}.$$

Hence, both the mean and variance evolve deterministically and reach equilibrium exponentially fast.

1.5 Reynolds-Averaged Navier–Stokes (RANS) Equations

Instantaneous equations (incompressible). Let $u_i(x,t)$ and $p(x,t)$ be the instantaneous velocity and pressure fields. The incompressible Navier–Stokes (NS) equations are

$$\frac{\partial u_i}{\partial t} + u_j\frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \nu\frac{\partial^2 u_i}{\partial x_j^2}, \qquad \frac{\partial u_i}{\partial x_i} = 0, \tag{1.19}$$

where ρ is density and ν is kinematic viscosity.

Reynolds decomposition and averaging rules. Decompose each field into mean and fluctuation:

$$u_i = \overline{u_i} + u_i', \qquad p = \overline{p} + p', \qquad \overline{u_i'} = 0,\ \overline{p'} = 0.$$

We assume the averaging operator $\overline{(\cdot)}$ is linear and commutes with space/time derivatives (e.g. time, ensemble, or space average under appropriate conditions). In particular,

$$\overline{\frac{\partial(\cdot)}{\partial t}} = \frac{\partial\,\overline{(\cdot)}}{\partial t}, \quad \overline{\frac{\partial(\cdot)}{\partial x_j}} = \frac{\partial\,\overline{(\cdot)}}{\partial x_j}, \quad \overline{u_i u_j} = \overline{u_i}\,\overline{u_j} + \overline{u_i' u_j'}.$$

A useful identity for the convective term. Using incompressibility $\partial u_j/\partial x_j = 0$,

$$u_j\frac{\partial u_i}{\partial x_j} = \frac{\partial}{\partial x_j}(u_i u_j) - u_i\frac{\partial u_j}{\partial x_j} = \frac{\partial}{\partial x_j}(u_i u_j). \tag{1.20}$$

Average the momentum equation. Apply $\overline{(\cdot)}$ to (1.19) and use (1.20):

$$\frac{\partial\,\overline{u_i}}{\partial t} + \frac{\partial\,\overline{u_i u_j}}{\partial x_j} = -\frac{1}{\rho}\frac{\partial\,\overline{p}}{\partial x_i} + \nu\frac{\partial^2\,\overline{u_i}}{\partial x_j^2}. \tag{1.21}$$

Expand the averaged quadratic term with the decomposition:

$$\overline{u_i u_j} = \overline{u_i}\,\overline{u_j} + \overline{u_i' u_j'} = \overline{u_i}\,\overline{u_j} + R_{ij}, \quad R_{ij} := \overline{u_i' u_j'}.$$

Hence,

$$\frac{\partial\,\overline{u_i}}{\partial t} + \frac{\partial}{\partial x_j}\left(\overline{u_i}\,\overline{u_j}\right) + \frac{\partial R_{ij}}{\partial x_j} = -\frac{1}{\rho}\frac{\partial\,\overline{p}}{\partial x_i} + \nu\frac{\partial^2\,\overline{u_i}}{\partial x_j^2}. \tag{1.22}$$

Recover the mean convective form. Using the product rule and mean incompressibility $\partial\overline{u_j}/\partial x_j = 0$,

$$\frac{\partial}{\partial x_j}\left(\overline{u_i}\,\overline{u_j}\right) = \overline{u_j}\frac{\partial\overline{u_i}}{\partial x_j} + \overline{u_i}\frac{\partial\overline{u_j}}{\partial x_j} = \overline{u_j}\frac{\partial\overline{u_i}}{\partial x_j}.$$

Thus (1.22) becomes the standard RANS momentum equation:

$$\boxed{\frac{\partial\overline{u_i}}{\partial t} + \overline{u_j}\,\frac{\partial\overline{u_i}}{\partial x_j} = -\frac{1}{\rho}\frac{\partial\overline{p}}{\partial x_i} + \nu\frac{\partial^2\overline{u_i}}{\partial x_j^2} - \frac{\partial R_{ij}}{\partial x_j}} \tag{1.23}$$

with the mean continuity equation

$$\boxed{\frac{\partial\overline{u_i}}{\partial x_i} = 0.} \tag{1.24}$$

Here $R_{ij} = \overline{u_i' u_j'}$ is the (specific) Reynolds stress tensor; in conservative form one often writes $-\partial_j(\rho\,\overline{u_i' u_j'})$.

Vector form. Let $\overline{\mathbf{u}}$ be the mean velocity, $\mathbf{R} = \overline{\mathbf{u}'\mathbf{u}'}$ the Reynolds stress:

$$\boxed{\rho\left(\frac{\partial\overline{\mathbf{u}}}{\partial t} + (\overline{\mathbf{u}}\cdot\nabla)\overline{\mathbf{u}}\right) = -\nabla\overline{p} + \mu\nabla^2\overline{\mathbf{u}} - \nabla\cdot\left(\rho\,\mathbf{R}\right), \qquad \nabla\cdot\overline{\mathbf{u}} = 0.} \tag{1.25}$$

Closure problem (remark). Equation (1.23) is not closed because R_{ij} introduces six new unknowns. Turbulence models (e.g. Boussinesq eddy viscosity, k–ε, k–ω, or Reynolds-stress transport models) provide constitutive relations for R_{ij} to close the system.

2 Statistical Equilibrium and Decorrelation Time

The modern understanding of statistical equilibrium, autocorrelation, and decorrelation time stems from foundational developments in physics and statistics. Ludwig Boltzmann (1844–1906) first proposed that microscopic randomness could yield stable macroscopic behavior, introducing the concept of equilibrium as a stationary probability distribution—an idea later reflected in stochastic systems such as the Ornstein–Uhlenbeck process, where noise and damping balance to form a steady Gaussian state. In the 1920s, G. Udny Yule formalized autocorrelation while studying sunspot and economic data, revealing that random processes can possess temporal memory—a concept that led to autoregressive (AR) models, discrete analogs of continuous stochastic systems. Building on this, G. I. Taylor (1935) defined the Lagrangian time scale, now known as the decorrelation time, as the integral of the autocorrelation function, quantifying how quickly a system forgets its past. Together, these ideas describe the long-term balance of randomness, the persistence of memory, and its rate of decay—forming the statistical backbone of modern stochastic dynamics, turbulence modeling, and data assimilation.

2.1 Equilibrium distribution

From (1.14)–(1.18), we observe that as $t \to \infty$,

$$X_t \sim \mathcal{N}\left(\frac{f}{a}, \frac{\sigma^2}{2a}\right),$$

which is the stationary (statistical equilibrium) distribution. Although each sample path remains random, the overall statistical behavior becomes time-invariant.

However, the equilibrium distribution alone does not fully characterize the dynamics: different SDEs may yield identical stationary distributions but have different temporal correlations. This motivates the introduction of the *decorrelation time*.

2.2 Autocorrelation function and decorrelation time

The *autocorrelation function* (ACF) of a stationary process is defined as

$$R(s) = \frac{\mathbb{E}\left[(X_t - \overline{X}_\infty)(X_{t+s} - \overline{X}_\infty)\right]}{\mathrm{Var}(X)_\infty}, \tag{2.1}$$

which measures the correlation between values of X_t separated by a lag s.

The *decorrelation time* (or correlation time) is then defined as

$$\tau_{\mathrm{corr}} = \int_0^\infty R(s)\, ds. \tag{2.2}$$

It quantifies the "memory" of the process — the time scale over which correlations persist.

Notes.

- The ACF and τ_{corr} are meaningful only after the system reaches stationarity.
- Always normalize $R(s)$ so that $R(0) = 1$.
- In numerical data, truncate the integration once $R(s)$ falls below noise level.

2.3 Example: Ornstein–Uhlenbeck process

For the OU process, where $f = 0$, the analytical solution to (1.1) is

$$X_t = X_0 e^{-at} + \sigma \int_0^t e^{-a(t-s)}\, dW_s. \tag{2.3}$$

It can be shown that in the stationary regime,

$$\mathbb{E}[X_t] = 0, \tag{2.4}$$

$$\mathrm{Cov}(X_t, X_{t+s}) = \frac{\sigma^2}{2a} e^{-a|s|}. \tag{2.5}$$

Thus, the normalized autocorrelation function is

$$R(s) = e^{-a|s|}. \tag{2.6}$$

Substituting into (2.2) gives

$$\tau_{\text{corr}} = \int_0^\infty e^{-as}\, ds = \frac{1}{a}. \tag{2.7}$$

Interpretation. The decorrelation time $\tau_{\text{corr}} = 1/a$ is inversely proportional to the damping rate:

- Large $a \Rightarrow$ strong damping, fast loss of memory, rapid decorrelation.
- Small $a \Rightarrow$ weak damping, long memory, slow decorrelation.

This explains why, in simulations, trajectories with small a appear smoother and more predictable, whereas large a yields more noise-dominated, short-memory behavior.

2.4 Summary

- The Reynolds decomposition separates the deterministic mean dynamics from stochastic fluctuations.
- The ensemble mean and variance satisfy deterministic ODEs, solvable analytically for linear SDEs.
- The equilibrium state provides stationary statistics, while the decorrelation time characterizes temporal dependence.
- For the Ornstein–Uhlenbeck process, the decorrelation time is $\tau_{\text{corr}} = 1/a$.