Introduction to Mechanistic Models

Jianbin Tan

School of Mathematics
Sun Yat-sen University

Table of Contents

Mechanistic models

In many fields, mechanistic models are parametric or semi-parametric models that describe the evolution of curves (on temporal domain) for system variables.

- Intrinsic behavior among system variables.
  - Functional relationship.
  - Interdependency of curves.
  - Causality/network.
- Infer parameter or pattern.

Smallpox in Montreal

Buffering process:

$$Dx(t) = -\beta x(t).$$

Describe the negative feedback of curve's dynamics.

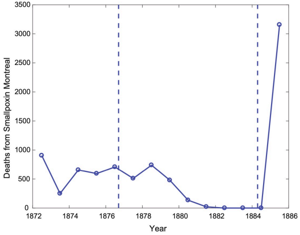


Figure: The number of smallpox deaths per year in Montreal.

Spread of disease

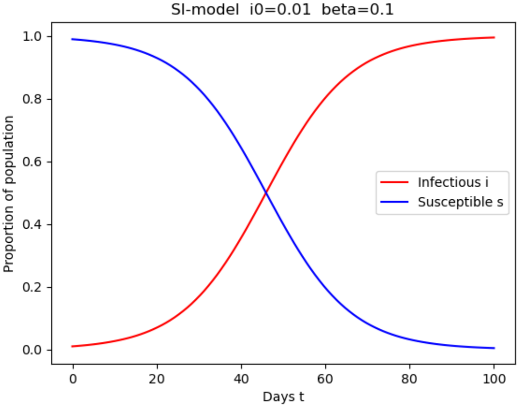


Figure: Dynamic curves of SI model.

Spread of disease

Infection process:

$$DS(t) = -\beta S(t)I(t) + \delta I(t),$$
$$DI(t) = \beta S(t)I(t) - \delta I(t).$$

Describe the interdependency of curves by non-linear functional relationship.

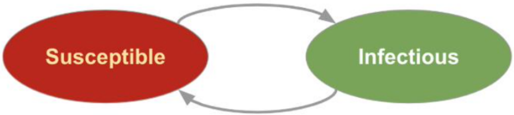


Figure: Directed graph of SI model.

Network of genes

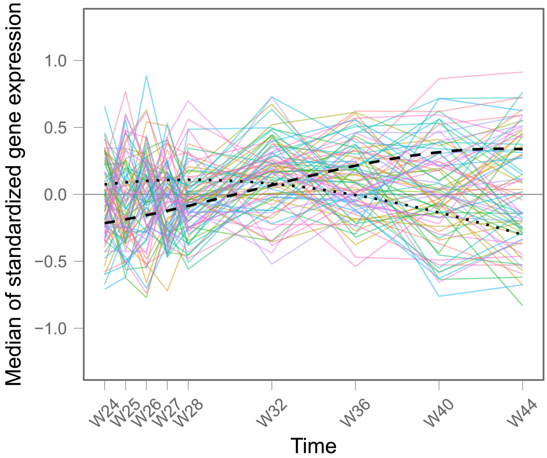


Figure: Curves of genes' products.

Network of genes

- High dimension.
- Complex functional relationship.
- Latent network.

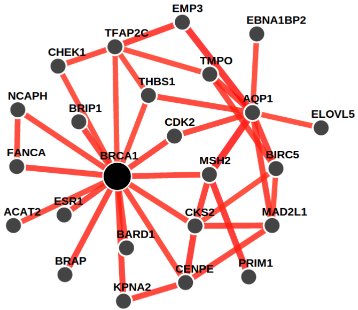


Figure: Network of genes.

Head impact and brain Acceleration

Damped harmonic equation:

$$D^2x(t) = -\beta_0 x(t) - \beta_1 Dx(t) + \alpha u(t).$$

Harmonic motion + external force + damping.

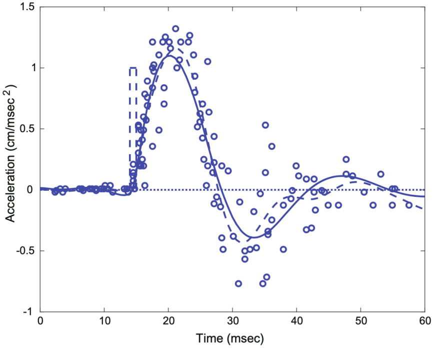
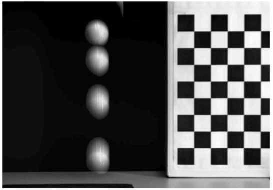


Figure: The motion of brain tissues.

Contact dynamics

Infer the environment properties from motion.



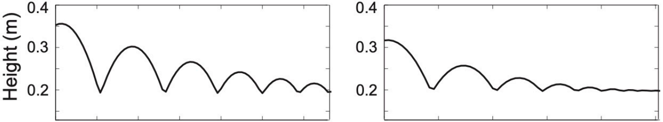
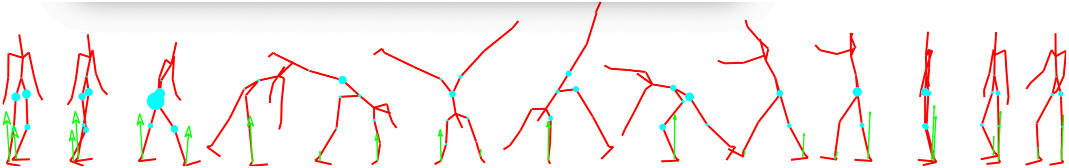


Figure: The motion of a ball bouncing onto a hard surface (left) and onto a soft mouse pad (right).

Motion capture data



Chinese handwriting

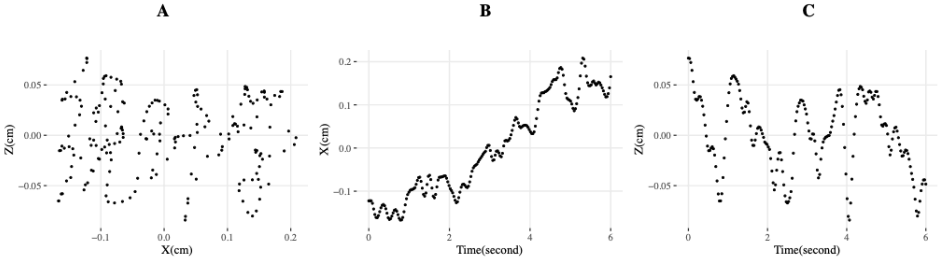


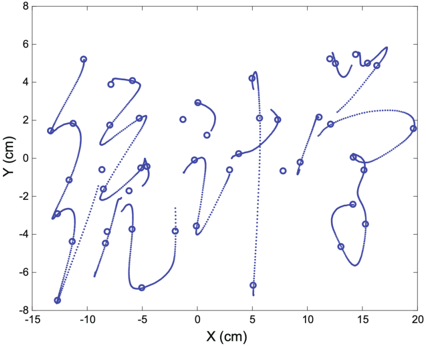
Figure: The production of a Chinese script.

Chinese handwriting

Harmonic equations:

$$D^2 X(t) = -\beta_X X(t) + \alpha_X(t),$$
$$D^2 Y(t) = -\beta_Y Y(t) + \alpha_Y(t),$$

where $\alpha_X(t)$ and $\alpha_Y(t)$ are step functions (representation of the Chinese script).



General framework

Let $\boldsymbol{x}(t) \in \mathbb{R}^p$ and $x_i(t)$ be the ith element of $\boldsymbol{x}(t)$, we assume that

$$\mathcal{D}_i x_i(t) = f_i\left(\boldsymbol{x}(t), t; \boldsymbol{\beta}\right), \quad i = 1, \cdots, p. \tag{1}$$

- $x_i(t)$ is the smooth curve on $[0, C]$.
- $\mathcal{D}_i$ is a known or unknown differential operator (e.g. D).
- $f_i(\cdot, t; \boldsymbol{\beta})$ is a forcing function onto the original curves to construct the drive items (fixed or random), which is indexed by parameter $\boldsymbol{\beta}$ and t (time-varying forcing).

Main problems

Trade-off between the statistical and computational efficiency.

- Estimate the parameters β (and unknown differential operators).
- Estimate the unknown f_i.
  - Non-parametric or semi-parametric models for f_i: the trajectories are in Gaussian processes/reproducing kernel Hilbert space, or $f_i(\boldsymbol{x}(t)) = \boldsymbol{c}^T\boldsymbol{x}(t) + g_i(\boldsymbol{x}(t))$.
  - Sparse additive model for $f_i((\boldsymbol{x}(t))) = \sum_{j\in\mathcal{S}} g_{ij}(x_i(t))$.
  - Neural network or tree representation for f_i.

In practice, we usually don't observe the dynamic curves directly. Instead, we obtain a collection of discrete noise data $Y_{it_j} \sim L(\cdot; x_i(t_j), \sigma_i)$, where $t_j \in [0, C]$ and $L(\cdot; \mu, \sigma_i)$ is a mass function with mean μ and dispersion parameter σ_i.

- Robustness.
- Missing data.
- Sparse and irregularly observed data.

Table of Contents

Linear system

More generally, a multi-equation first-order system with p variables will allow for each variable to force or input each other variable, so that

$$D\mathbf{x}(t) = \mathbf{B}\mathbf{x}(t)$$

where the square matrix $\mathbf{B}$ will have arbitrary real-valued entries. The solution to satisfy the above equation can be expressed with the matrix exponential function

$$\mathbf{x}(t) = \exp(\boldsymbol{B}t)\boldsymbol{c}$$

where $\mathbf{c}$ is a vector of initial value defining the state of the system at time 0. For example, if $p = 1$, $x(t) = Ce^{Bt}$.

Linear system

If a forcing term $\mathbf{u}$ is added to this equation to make $\mathbf{Dx}(t) = \mathbf{Bx}(t) + \mathbf{u}(t)$, then

$$\mathbf{x}(t) = \exp(\boldsymbol{B}t)\boldsymbol{c} + \int_0^t \exp(\mathbf{B}(t-s))\mathbf{u}(s)\,\mathrm{d}s.$$

For higher-order equations, for example, $\mathbf{D}^2\mathbf{x}(t) + \omega\mathbf{Dx}(t) = \mathbf{Bx}(t) + \mathbf{u}(t)$, we define $\boldsymbol{x}_*(t) = \mathbf{Dx}(t)$, and re-write the equations as

$$\begin{aligned}\mathbf{Dx}_*(t) &= -\omega\mathbf{x}_*(t) + \mathbf{Bx}(t) + \mathbf{u}(t),\\ \mathbf{Dx}(t) &= \boldsymbol{x}_*(t),\end{aligned}$$

which is again the linear case. The closed-form of $\boldsymbol{x}(t)$ can be calculated similarly if B is time-varying.

Table of Contents

Non-linear system

Typically, the parameter of the mechanistic model with a closed-form solution (denoted as $x_i(\cdot; \boldsymbol{\beta})$) can directly estimated by

$$\arg\min_{\boldsymbol{\beta}} \prod_{i,j} L(Y_{it_j}; x_i(t_j; \boldsymbol{\beta}), \sigma_i).$$

While for the non-linear system cases and also the unknown forcing function f_i, we require to discuss the existence and uniqueness of its solution, and it may also have no closed-form expression for model inference.

Existence and uniqueness

We now denote the differential equations as $\mathcal{D}\boldsymbol{x}(t) = \boldsymbol{f}(\boldsymbol{x}(t), t)$, where $\mathcal{D}$ is a general-order differential operator.

- Local existence: $\boldsymbol{f}(\boldsymbol{x}(t), t)$ is Lipschitz-continuous w.r.t. its first component over a neighbourhood of $(\boldsymbol{x}(t_0), t_0)$ (denoted as U), i.e.,

$$||\boldsymbol{f}(\boldsymbol{x}_1, t) - \boldsymbol{f}(\boldsymbol{x}_1, t)|| \leq K||\boldsymbol{x}_1 - \boldsymbol{x}_2||$$

  $\forall\ (\boldsymbol{x}_1, t),\ (\boldsymbol{x}_2, t) \in U$. A sufficient condition is the operator norm of $\partial_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}, t)$ is less than K for all $(\boldsymbol{x}, t) \in U$.

Existence and uniqueness

We now denote the differential equations as $\mathcal{D}\boldsymbol{x}(t) = \boldsymbol{f}(\boldsymbol{x}(t), t)$, where $\mathcal{D}$ is a general-order differential operator.

- Uniqueness: there is only one solution $\boldsymbol{x}$ s.t. $\mathcal{D}\boldsymbol{x}(t) = \boldsymbol{0}$. If the ith operator of $\mathcal{D}$ is of form $D^K + \sum_{k=0}^{K-1} \omega_{ik} D^k$, the null space of the operator is a K-dimensional function space. Therefore, we may need K constraint for x_i, e.g., we know $D^k x_i(0)$ for $k = 0, \cdots, K-1$.

In practice, we may not know the constraints to determine the forward solution, therefore, we can treat these unknown constraints as the nuisance parameters.

Parameter identifiability

Another crucial problem for statistical inference is that whether the observed data can identify the parametric or non-parametric $\boldsymbol{f}$. Assume that $\boldsymbol{f}(\boldsymbol{x}(t), t; \boldsymbol{\beta})$ is Lipschitz continuous for all t as mentioned before.

Essentially, we need to ensure that $\boldsymbol{x}(t; \boldsymbol{\beta}_2) = \boldsymbol{x}(t; \boldsymbol{\beta}_2)$, $\forall t$ iff $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, where $\boldsymbol{x}(t; \boldsymbol{\beta}_2) = \boldsymbol{x}(t; \boldsymbol{\beta}_2)$ implies that

$$f_i(\boldsymbol{x}(t), t; \boldsymbol{\beta}_1) = f_i(\boldsymbol{x}(t), t; \boldsymbol{\beta}_2), \ \forall t, \ i.$$

$\Rightarrow f_i(\boldsymbol{x}(t), t; \boldsymbol{\beta}_1) = f_i(\boldsymbol{x}(t), t; \boldsymbol{\beta}_2), \ \forall t, \ i$ iff $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. More often, $f_i(\boldsymbol{x}(t), t; \boldsymbol{\beta})$ could be represented as $\boldsymbol{g}_i^T(\boldsymbol{x}(t))\boldsymbol{l}_i(\boldsymbol{\beta})$, it may be more easy to check that:

$$\boldsymbol{l}_i(\boldsymbol{\beta}_1) = \boldsymbol{l}_i(\boldsymbol{\beta}_2), \ \forall i,$$

iff $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

Table of Contents

Numerical solution

Ordinary differential equation (ODE) models are a powerfully expressive class of models for describing the way systems evolve over time. However, outside a fairly narrow class of equations, we cannot obtain solutions to these ODEs analytically, and therefore directly understand the behavior that these equations produce.

Runge–Kutta Methods

- Euler methods: By Taylor expansion,

$$\begin{aligned} x_i(t+h) &= x_i(t) + hDx_i(t) + O(h^2), \\ &= x_i(t) + hf_i(\boldsymbol{x}(t), t; \boldsymbol{\beta}) + O(h^2), \end{aligned}$$

  as $h \to 0$.

- Higher-order approximation:

$$\begin{aligned} x_i(t+h) &= x_i(t) + hDx_i(t) + \frac{h^2}{2}D^2x_i(t) + O(h^3), \\ &= x_i(t) + hf_i(\boldsymbol{x}(t), t; \boldsymbol{\beta}) + \frac{h^2}{2}Df_i(\boldsymbol{x}(t), t; \boldsymbol{\beta}) + O(h^3), \end{aligned}$$

  as $h \to 0$.

Numerical problem

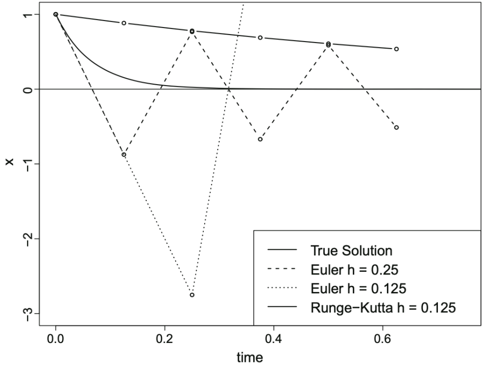


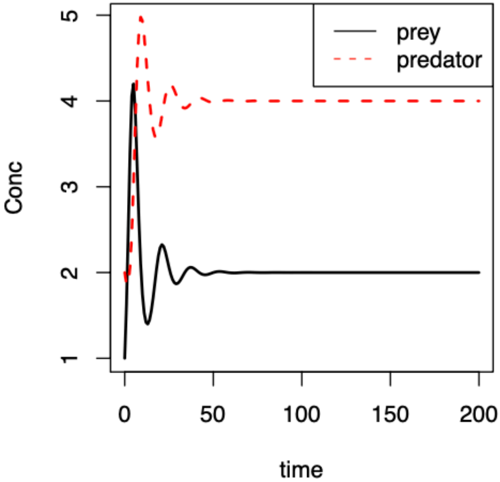
Figure: Example of a stiff system $Dx = 15x$. Solid blue line produces exact solution $x(t) = \exp(15t)$; the dashed and dotted lines are solutions by Euler methods with $h = 0.125$ and 0.25, respectively and the solid black line with circles is the second-order Runge–Kutta solution.

R package deSolve

```
R> LVmod0D <- function(Time, State, Pars) {
+    with(as.list(c(State, Pars)), {
+      IngestC <- rI * P * C
+      GrowthP <- rG * P * (1 - P/K)
+      MortC   <- rM * C
+
+      dP    <- GrowthP - IngestC
+      dC    <- IngestC * AE - MortC
+
+      return(list(c(dP, dC)))
+    })
+  }
```

```
R> pars <- c(rI = 0.2, rG = 1.0, rM = 0.2, AE = 0.5, K  = 10)
R> yini <- c(P = 1, C = 2)
R> times <- seq(0, 200, by = 1)
R> print(system.time(
+    out <- ode(func = LVmod0D, y = yini, parms = pars, times = times)))
```

R package deSolve



Qualitative Behavior

Within the discipline of applied mathematics, a large part of dynamical systems theory is concerned with the description of the qualitative behavior of dynamical systems.

- Converge to a fixed point.
- Diverge to infinity.
- Produces a consistent pattern of oscillations.
- Something more complicated.

ODE models – particularly those that produce complex dynamics – have poor quantitative agreement with observed data.

Table of Contents

Fixed points

The first point of analysis in any system is to understand its fixed points – that is, at what values will the system remain? In the context of ODEs, the fixed point $\boldsymbol{x}^*$ is the solution of

$$\boldsymbol{f}(\boldsymbol{x}; \boldsymbol{\beta}) = \mathbf{0},$$

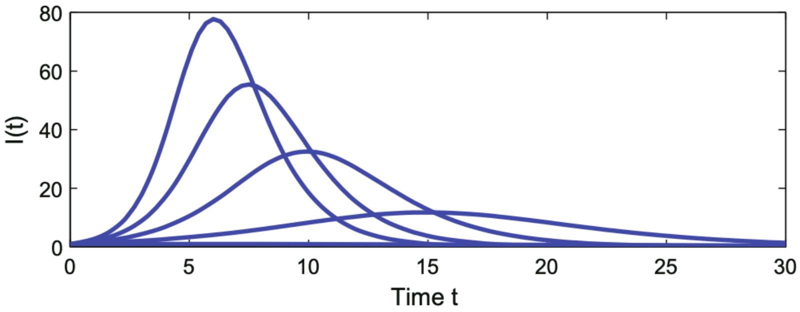
w.r.t. $\boldsymbol{x}$ given the parameter $\boldsymbol{\beta}$.

As we will see below, neither fixed point tells us much about trajectories that are close to them in this case. For this we require some stability analysis.

SIR model

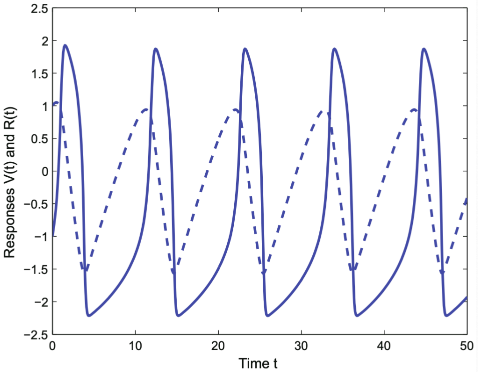
$$
\begin{aligned}
DS(t) &= -\beta I(t)S(t), \\
DI(t) &= \beta I(t)S(t) - \gamma I(t), \\
DR(t) &= \gamma I(t),
\end{aligned}
$$

where $R(t) = \frac{\beta S(t)}{\gamma}$ is the basic reproduction number.



FitzHugh-Nagumo model

$$DV(t) = c(V(t) - V^3(t) - R(t)),$$
$$DR(t) = (V(t) - a + bR(t)).$$



Linear cases

Recall that the equation

$$\mathbf{D}\mathbf{x} = \mathbf{B}\mathbf{x}$$

has solutions given by

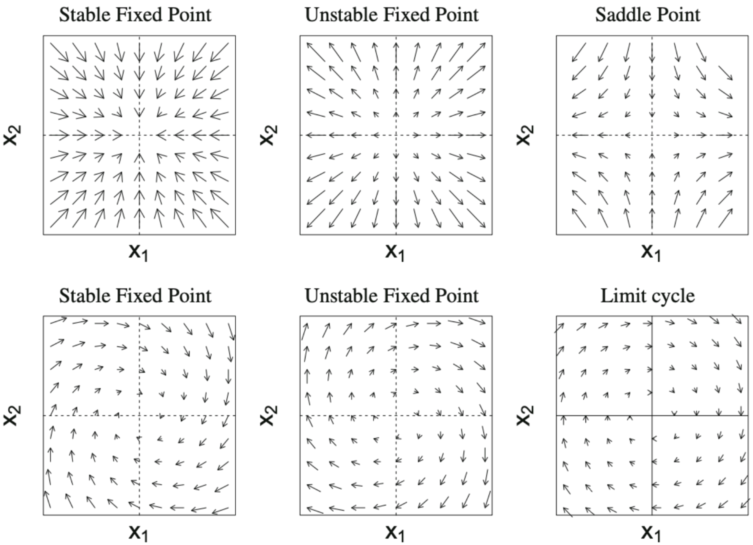
$$\mathbf{x}(t) = \exp(\boldsymbol{B}t)\boldsymbol{c} = \sum_{i=1}^{p} c_i e^{u_i t} v_i$$

where the u_i and v_i are eigenvalues and eigenvectors of $\mathbf{B}$ and the c_i are constants that depend on initial conditions. Now, define $u_i = a_i + ib_i$ we have

$$e^{(a_i+ib_i)t} = e^{a_i t}\left(\cos\left(b_i t\right) + i \sin\left(b_i t\right)\right).$$

- $\mathbf{0}$ is a fixed point.
- If $a_i < 0$ for all i, $\boldsymbol{x}(t) \to \mathbf{0}$ as $t \to \infty$.

Vector fields



Non-linear cases

We want to examine whether $\boldsymbol{x}^*$ is a stable fixed point for $\boldsymbol{Dx} = \mathbf{f}(\mathbf{x}; \boldsymbol{\beta})$. Note that

$$\mathbf{Dx} = \mathbf{f}(\mathbf{x}^*, \boldsymbol{\beta}) + \partial_{\mathbf{x}}\mathbf{f}(\mathbf{x}^*, \boldsymbol{\beta})(\mathbf{x} - \mathbf{x}^*) + o\left(\|\mathbf{x} - \mathbf{x}^*\|^2\right)$$

Here the Jacobian $\mathbf{J} = \partial_{\mathbf{x}}\mathbf{f}(\mathbf{x}^*, \boldsymbol{\theta})$ is a square matrix with derivatives along columns and components of $\mathbf{f}$ down rows:

$$[\mathbf{J}]_{ij} = \frac{df_i(\mathbf{x}^*, \boldsymbol{\beta})}{dx_j}.$$

Substituting $\mathbf{z}(t) = \mathbf{x}(t) - \mathbf{x}^*$ yields

$$\mathbf{Dz} = \mathbf{Jz} + o\left(\|\mathbf{z}\|^2\right)$$

If $\mathbf{0}$ is a stable fixed point of $\mathbf{Dz} = \mathbf{Jz} \Rightarrow \boldsymbol{x}^*$ is a stable fixed point for $\boldsymbol{Dx} = \mathbf{f}(\mathbf{x}; \boldsymbol{\beta})$.

Global analysis

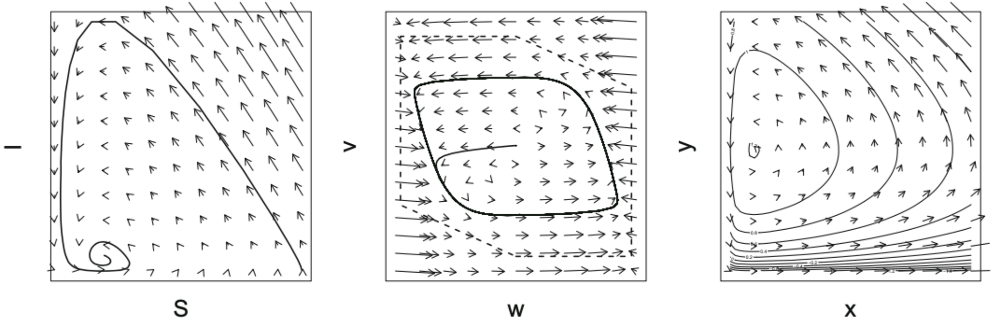


Table of Contents

Bifurcations

Bifurcation analysis refers to changes in qualitative behavior as parameters of the system change. The very simplest bifurcation can be observed in the linear equation

$$Dx = \beta x,$$

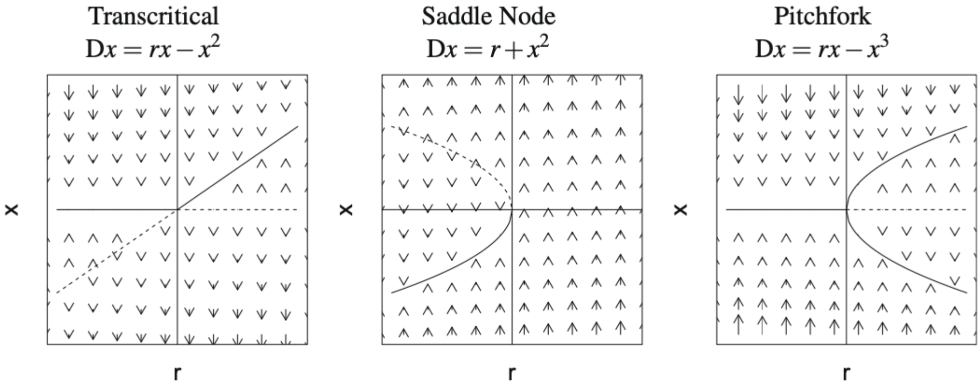
with solutions $x(t) = c\exp(\beta x)$ where we see that there is a fixed point at $x = 0$ which is stable if $\beta < 0$ and unstable if $\beta > 0$.

In fact, for linear systems the only real changes in behavior are from stable to unstable fixed points and in the direction (or existence) of circular motion. However, nonlinear systems can exhibit a variety of behaviors as a parameter is changed.

Why should a statistician care?

- The statistical properties of parameter estimates can change depending on system behavior. For a system with stable fixed points, for example, there are generally no consistent estimators of initial conditions if observations are taken over an increasing time domain.
- Frequently, the parameter that is being varied can be controlled in an experiment.
- In some systems it can be useful to think of a slow-moving state variable as a parameter and examine the behavior of the fast-moving variables with the slow-moving variable fixed. This can allow us to understand the global behavior of a system by breaking it down into lower-dimensional counterparts.

Transcritical bifurcations



Hopf bifurcations

Van der Pol system:

$$Dx = -y + (\mu - y^2)x$$
$$Dy = x.$$

These describe a stable fixed point becoming unstable and producing a stable limit cycle along the way.

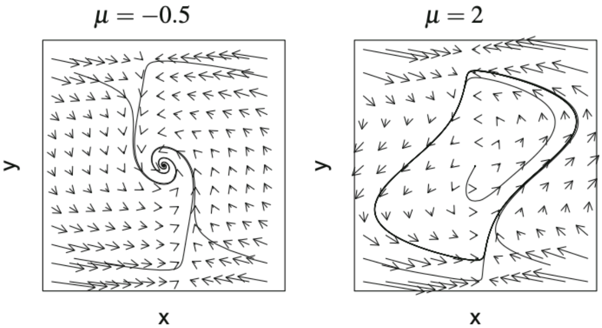


Table of Contents

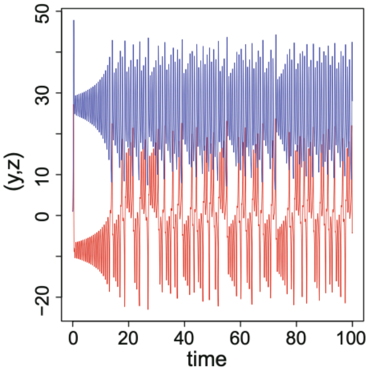
Chaos

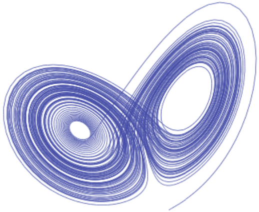
The notion of chaos has been one of the most influential developments in 20th century mathematics. Few have not heard of the "butterfly effect," – that a butterfly flapping its wings in Brazil can cause a hurricane in Florida – coined by Edward Lorenz as a description of the consequences of chaotic systems.

Butterfly effect

Lorenz system:

$$\begin{aligned} Dx &= \sigma(y - x), \\ Dy &= x(\rho - z) - y, \\ Dz &= xy - \beta z. \end{aligned}$$





Implication

- There is estimation error for the initial values, therefore, we can only predict very short time-periods ahead,
- The introduction of numerical errors in solving differential equations can mean that we end up quite far from the true trajectory of the system.
- Sensitivity to initial conditions should mean that we can estimate these very well indeed.