Mathematical tools for pharmacometrics: Calculus

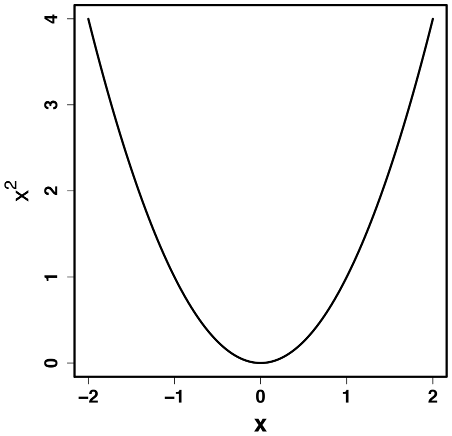
S. Benzekry

sebastien.benzekry@inria.fr

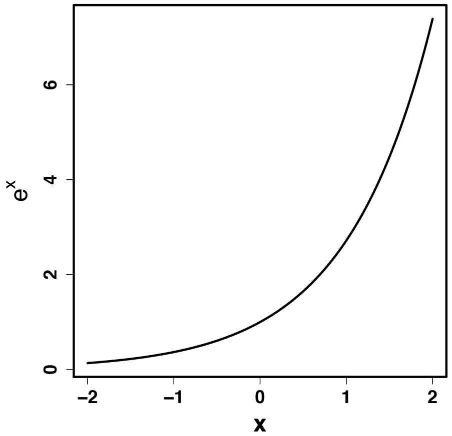
SIMULATION MODELING ADAPTIVE RESPONSE FOR THERAPEUTICS IN CANCER

Basic functions $\mathbb{R} \to \mathbb{R}$

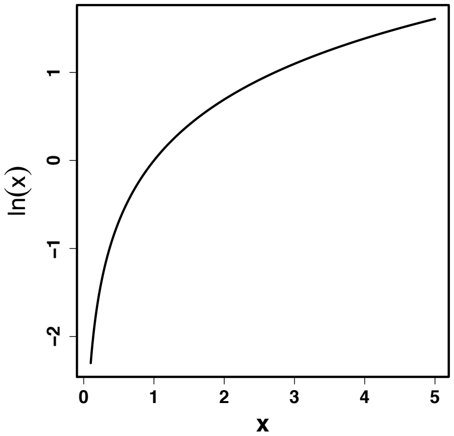
$x \mapsto x^n$



$x \mapsto \exp(x) = e^x$



$x \mapsto \ln(x)$



$$e^{a+b} = e^a \cdot e^b$$

$$\ln(a \cdot b) = \ln(a + b)$$

$$\ln(e^x) = e^{\ln(x)} = x$$

$$\log_{10}(10^x) = 10^{\log_{10}(x)} = x$$

$$\log_{10}(x) = \frac{\ln(x)}{\ln(10)}$$

Derivative

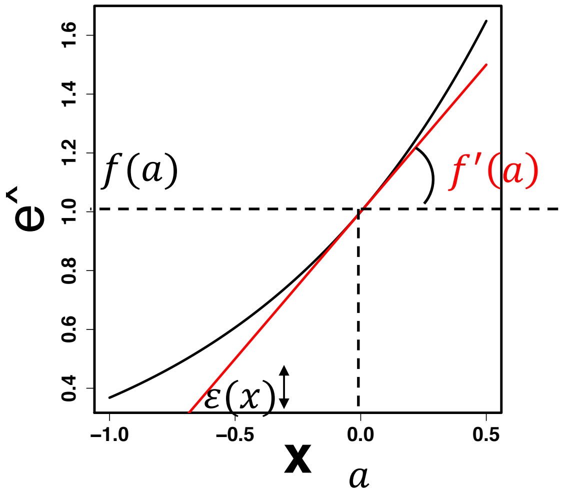
- Linear approximation of a (nonlinear) function $\mathbb{R} \to \mathbb{R}$, in the neighborhood of a

$$f(x) = f(a) + f'(a) \cdot (x - a) + \varepsilon(x)$$

- Formally

$$f'(a) = \lim_{x \to a} \frac{f(x+a) - f(a)}{x - a}$$

$y = f(a) + f'(a) \cdot (x - a)$



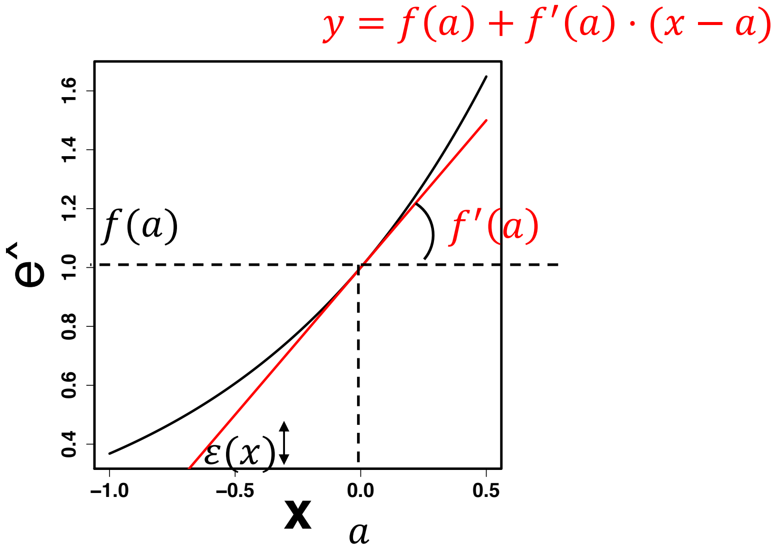
Derivative

- Linear approximation of a (nonlinear) function $\mathbb{R} \to \mathbb{R}$, in the neighborhood of a

$$f(x) = f(a) + f'(a) \cdot (x - a) + \varepsilon(x)$$

- Example, for $x \simeq 0$

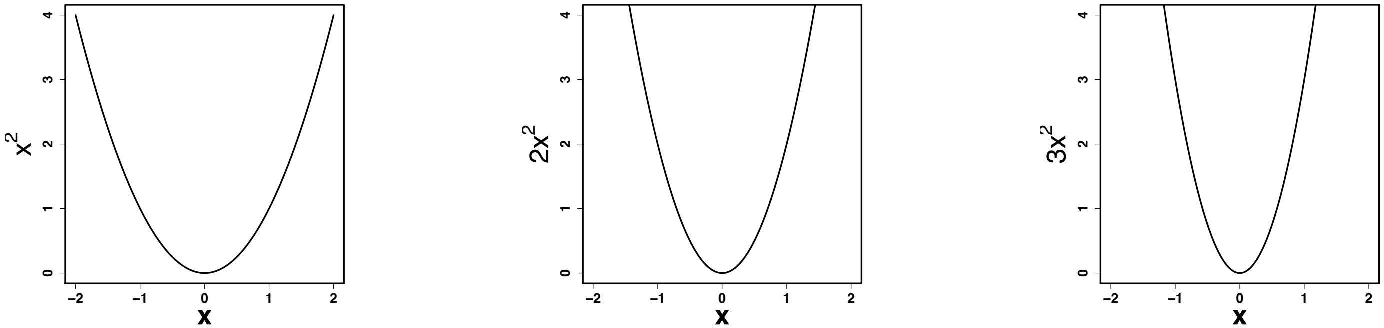
$$e^x \simeq e^0 + (e')(0) \cdot (x - 0)$$
$$e^x \simeq 1 + x$$



Quadratic approximation. Second derivative

- Linear function $f(x) = \theta_0 + \theta_1 x \Leftrightarrow$ line characterized by slope

- Quadratic function $f(x) = \theta_0 + \theta_1 x + \theta_2 x^2 \Leftrightarrow$ characterized by curvature

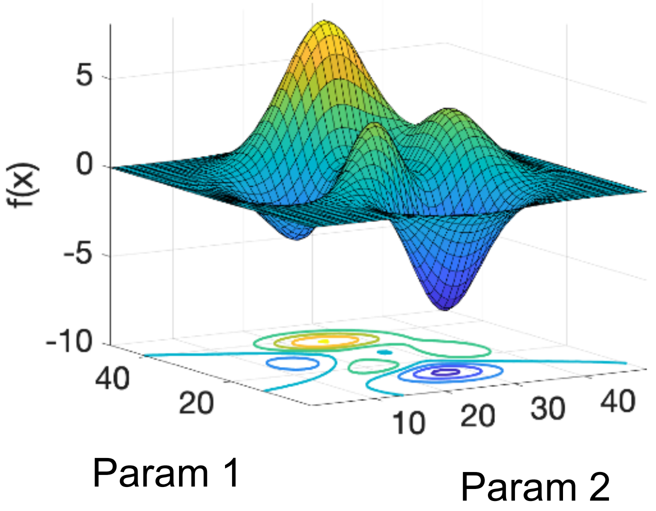


- Same slope as f and same curvature. Taylor's formula

$$f(x) = f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2} \cdot (x - a)^2 + \varepsilon(x)$$

Minimization

Why? fit a model ⟺ **minimize** a function

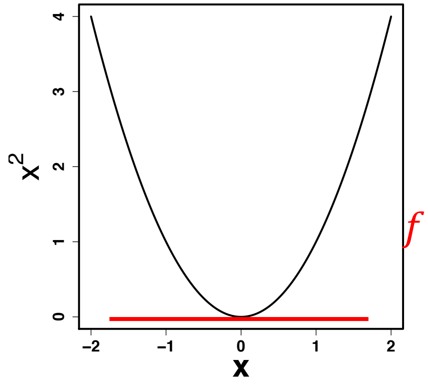


Extremum. Necessary condition

- If f has a minimum or maximum in a

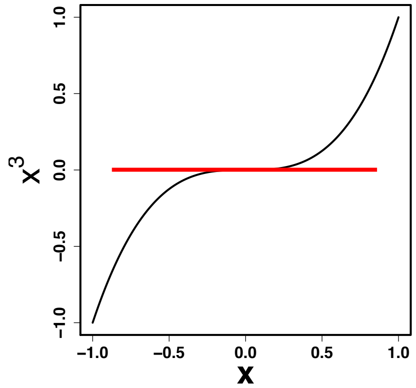
$$f(x) \simeq f(a) + f'(a) \cdot (x - a)$$

$\Rightarrow f'(a) = 0$



$f'(0) = 0$

- But this is not a sufficient condition

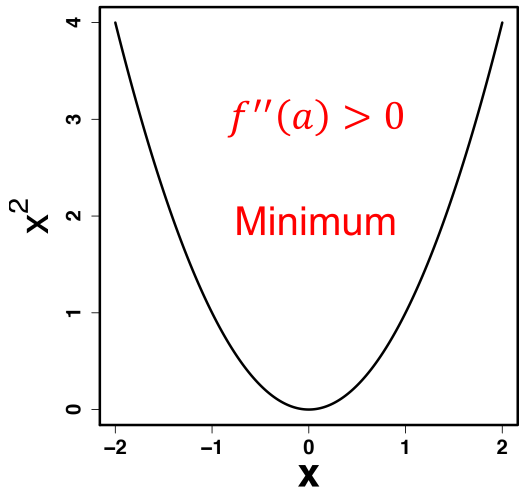


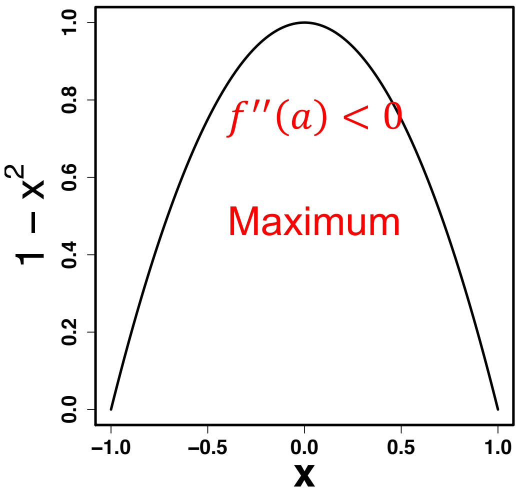
$f'(0) = 0$

Extremum. Sufficient condition

$$f(x) \simeq f(a) + \cancel{f'(a)\cdot(x-a)} + \frac{f''(a)}{2}\cdot(x-a)^2$$

- a extremum





$f: \mathbb{R}^n \to \mathbb{R}$. Gradient

- Linear function $\mathbb{R}^n \to \mathbb{R} \Rightarrow \langle a, x\rangle = a_1 x_1 + \cdots + a_n x_n$

$$\boldsymbol{f(x) \simeq f(a) + \langle \nabla f(a), x - a\rangle = f(a) + \frac{\partial f}{\partial_{x_1}}(a) \cdot (x_1 - a_1) + \cdots + \frac{\partial f}{\partial_{x_n}}(a) \cdot (x_n - a_1)}$$

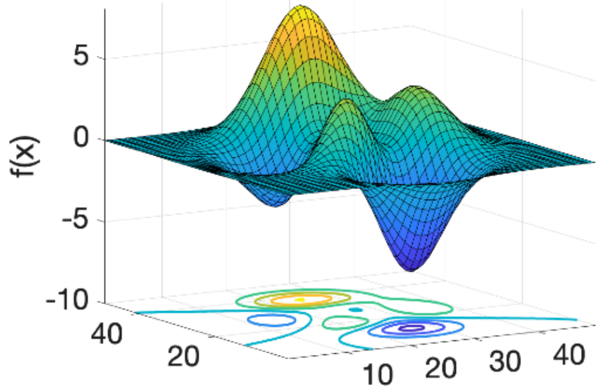
- $\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a)\right) \in \mathbb{R}^n$

- $\frac{\partial f}{\partial_{x_i}}(a)$ is called the partial derivative of f in a in the direction x_i

Gradient descent algorithms

- f has a minimum (or maximum) in $a \Rightarrow \boldsymbol{\nabla f(a) = 0}$, i.e. $\frac{\partial f}{\partial_{x_i}}(a) = 0 \quad \forall i$

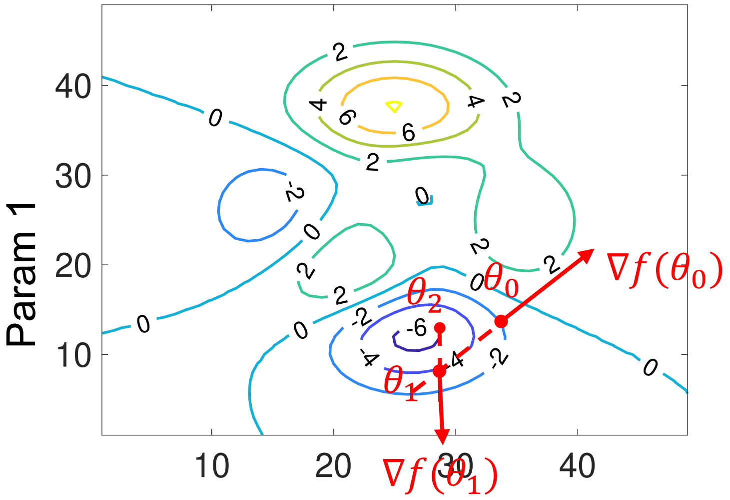
$\Rightarrow$ To minimize f $\qquad \theta_{n+1} = \theta_n - \lambda \nabla f(\theta_n)$



$\nabla f(\theta_n) =$

steepest descent

direction



Optimization algorithms for NLME

- FO = First Order. Linearizes for small values of random effects and residual error
  Sheiner, Comput Biomed Res, 1972, Sheiner and Beal, J Pharmacokinet Biopharm, 1980

⇒ Fast, but inaccurate for large ω or large residual error

- FOCE (NONMEM) = First-Order Conditional Estimation. Improvement over FO in terms of estimation of ω.
  Lindstrom and Bates, Biometrics, 1990

- FOCE-I = FOCE with interaction between intra- (ε) and inter- (η) variability.

⇒ to be used when proportional (or combined) error model but more computationally demanding

- LAPLACE = Same as FOCE except second-derivative (Hessian) approximation. Used for nonnormal (or lognormal) densities.
  Wolfinger, Biometrika, 1993

- SAEM (NONMEM, Monolix) = Stochastic Approximation of Expectation-Maximization
  Delyon, Lavielle, Moulines, Annals of Statistics, 1999

⇒ Slower, but converges better to global minimum

Plan et al., AAPS J, 2012

Bauer, CPT: PSP, 2019

$f\colon \mathbb{R}^n \to \mathbb{R}^m$. Differential

- $$f\colon \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^m \\ x = (x_1, \cdots, x_n) & \mapsto & \begin{pmatrix} f_1(x_1, \cdots, x_n) \\ \vdots \\ f_m(x_1, \cdots, x_n) \end{pmatrix} \end{array}$$

- Linear application $\mathbb{R}^n \to \mathbb{R}^m$? Matrix!

$$Df(a) = \begin{array}{cc} n & \\ \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix} & m \end{array}$$

$$\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{a}) = \boldsymbol{f}(\boldsymbol{a}) + \boldsymbol{Df}(\boldsymbol{a}) \cdot (\boldsymbol{x}-\boldsymbol{a}) + \boldsymbol{\varepsilon}(\boldsymbol{x})$$

Sensitivity matrix

$$y_j = f\left(t_j, \theta^*\right) + \varepsilon_j \Leftrightarrow \begin{pmatrix} y_1 \\ \vdots \\ y_\mathrm{n} \end{pmatrix} = \begin{pmatrix} f(t_1, \theta^*) \\ \vdots \\ f(t_n, \theta^*) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_\mathrm{n} \end{pmatrix} \Leftrightarrow y = f(t, \theta^*) + \varepsilon$$

- Suppose we have a first guess θ_0 close to θ^*.

$$f(t, \theta^*) \simeq f(t, \theta_0) + D_\theta f(t, \theta_0) \cdot (\theta^* - \theta_0)$$

- If there is an exact solution ($\varepsilon = 0$)

$$y \simeq f(t, \theta_0) + S \cdot (\theta^* - \theta_0)$$

$$\theta^* \simeq \theta_0 + S^{-1} \cdot \left(y - f(t, \theta_0)\right)$$

- In general, $\varepsilon \neq 0$ and $n > p \Rightarrow$ least-squares

$$(S^T \cdot S)^{-1} \cdot S^T \cdot \left(y - f(t, \theta_0)\right)$$

$$S = D_\theta f(t, \theta) = \begin{array}{c} \overset{p}{\longleftrightarrow} \\ \begin{pmatrix} \frac{\partial f}{\partial \theta_1}(t_1, \theta) & \cdots & \frac{\partial f}{\partial \theta_p}(t_1, \theta) \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial \theta_1}(t_n, \theta) & \cdots & \frac{\partial f}{\partial \theta_p}(t_n, \theta) \end{pmatrix} \end{array} \updownarrow n$$

= Sensitivity matrix

$f: \mathbb{R}^n \to \mathbb{R}$. Hessian matrix

- Second derivative?

- Quadratic form: $\begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & x^T \cdot M \cdot x \end{array}$, M symmetric matrix

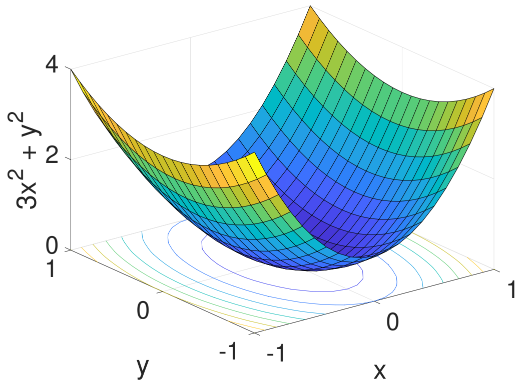
- Matrix of second partial derivatives = Hessian matrix

$$\boldsymbol{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

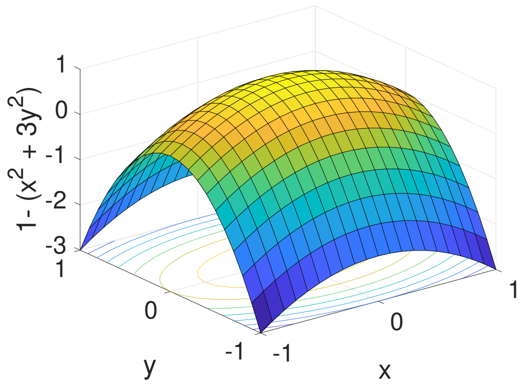
- $\frac{\partial f}{\partial x_i \partial x_j} = \frac{\partial f}{\partial x_j \partial x_i} \Rightarrow H$ is symmetric

Curvature

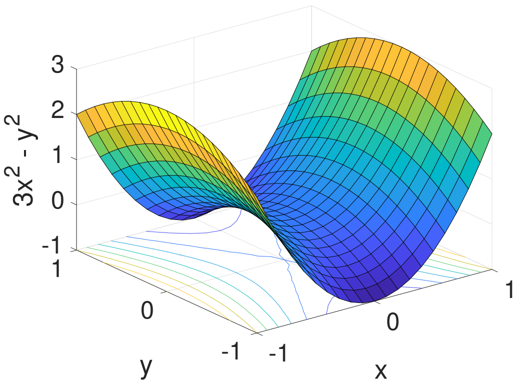
- The Hessian matrix extends the notion of (local) curvature to $f: \mathbb{R}^n \rightarrow \mathbb{R}$
- The eigenvectors give the principal directions and associated eigenvalues the curvatures



$$H = \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}$$



$$H = \begin{pmatrix} -1 & 0 \\ 0 & -6 \end{pmatrix}$$

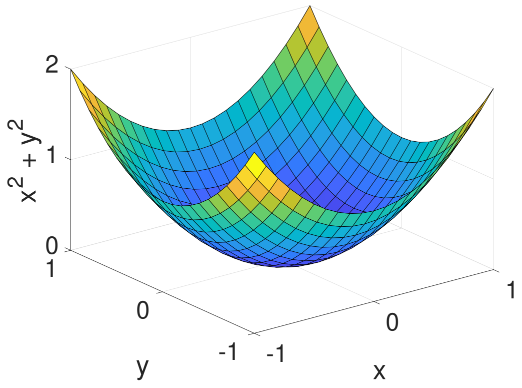


$$H = \begin{pmatrix} 6 & 0 \\ 0 & -1 \end{pmatrix}$$

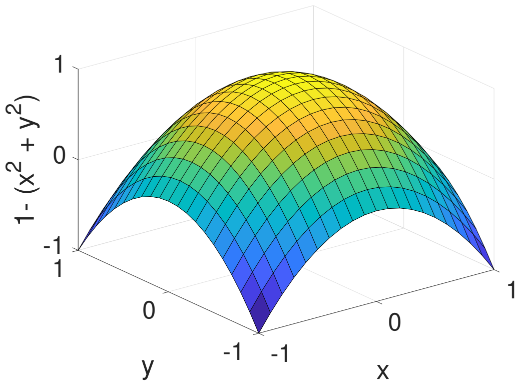
Extremum in dimension n

$$f(x) = f(a) + \langle \nabla f(a), x - a \rangle + \frac{1}{2}(x-a)^T \cdot H(a) \cdot (x-a) + \varepsilon(x)$$

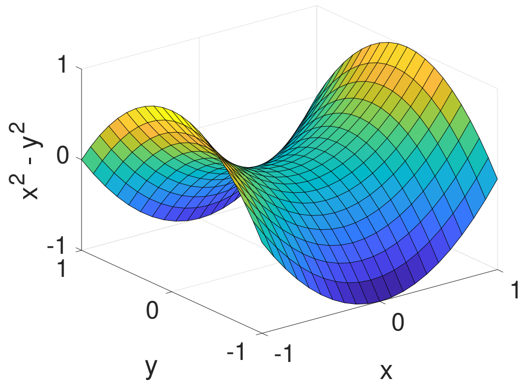
Minimum



Maximum



Saddle



H positive definite

H negative definite

- Note: Hessian of the objective function (-2LL) = R matrix in NONMEM

Differential equations. 1D

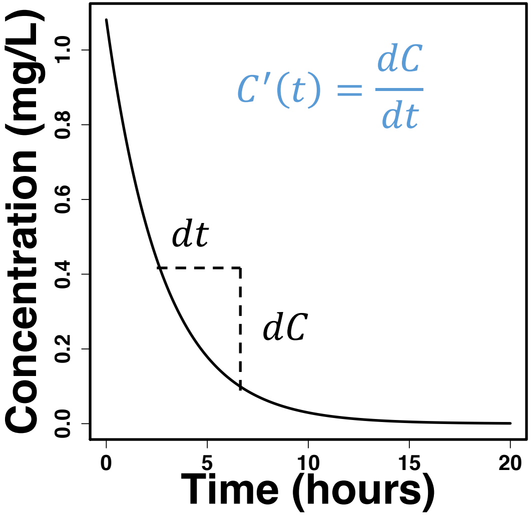
$$y' = f(t, y)$$

- Derivative = rate of variation of one quantity relatively to another

- Historically introduced to describe movement (Newton)

- To determine the solution, we need to specify the differential equation (rate of change) and initial condition. Deterministic system

$$\begin{cases} y' = f(t, y) \\ y(t_0) = y_0 \end{cases}$$



Linear equation in 1D

$y' = f(t, y)$

- The equation is said to be linear when f is linear in y
- The fundamental function with derivative proportional to itself is the exponential $e^{\lambda t}$
- The equation is called homogenous when there is no zero-order term

$$\begin{cases} \dfrac{dC}{dt} = -kC \\ C(0) = \dfrac{D}{V} \end{cases} \quad \Rightarrow \quad C(t) = \frac{D}{V} e^{-kt}$$

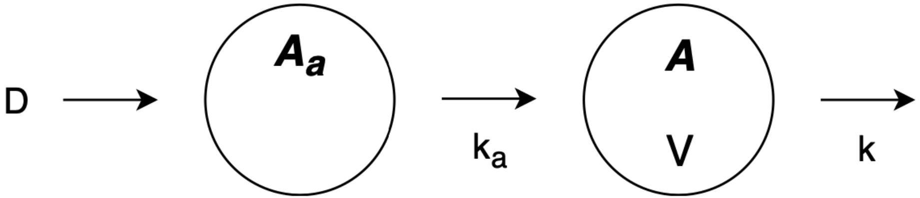
- For a non-homogenous equation, method of « variation of the parameter »

$$\begin{cases} \dfrac{dC}{dt} = -kC + u \\ C(0) = \dfrac{D}{V} \end{cases} \quad C(t) = \lambda(t) e^{-kt} \quad \Rightarrow \quad C(t) = \frac{D}{V} e^{-kt} + \frac{u}{k}\left(1 - e^{-kt}\right)$$

One-compartment model with absorption

$$\begin{cases} \frac{dA_a}{dt} = -k_a A_a \\ \frac{dA}{dt} = k_a A_a - kA \\ A_a(t=0) = D,\ A(t=0) = 0 \end{cases}$$

$$C(t) = \frac{A(t)}{V}.$$



$$A_a(t) = De^{-k_a t}$$

$$\frac{dA}{dt} = k_a D e^{-k_a t} - kA$$

$$\tilde{A}(t) = A(t)e^{kt}$$

$$\frac{d\tilde{A}}{dt} = k_a D e^{-k_a t + kt} - \cancel{kAe^{kt}} + \cancel{kAe^{kt}}$$

$$\tilde{A}(t) = \tilde{A}(0) + k_a D \int_0^t e^{(k-k_a)s} ds = 0 + \frac{k_a D}{k - k_a}\left(e^{(k-k_a)t} - 1\right)$$

$$A(t) = D\frac{k_a}{k_a - k}\left(e^{-kt} - e^{-k_a t}\right)$$

System of linear differential equations

$$\begin{cases} y_1' = a_{1,1}y_1 + a_{1,2}y_2 + \cdots + a_{1,n}y_n \\ y_2' = a_{2,1}y_1 + a_{2,2}y_2 + \cdots + a_{2,n}y_n \\ \vdots \\ y_n' = a_{n,1}y_1 + a_{n,2}y_2 + \cdots + a_{n,n}y_n \end{cases} \Leftrightarrow \begin{pmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\Leftrightarrow \quad y' = M \cdot y$$

$$y'(t) = e^{Mt} y_0 ??$$

- D diagonal

$$D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \quad e^D = \begin{pmatrix} e^{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\lambda_n} \end{pmatrix}$$

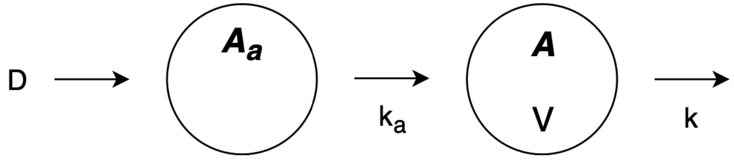
- M diagonalizable $M = P^{-1} \cdot D \cdot P$

$$e^M = P \cdot \begin{pmatrix} e^{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\lambda_n} \end{pmatrix} \cdot P^{-1}$$

$\Rightarrow y_i(t) =$ linear combination of e^{λ_j}

One-compartment model with absorption

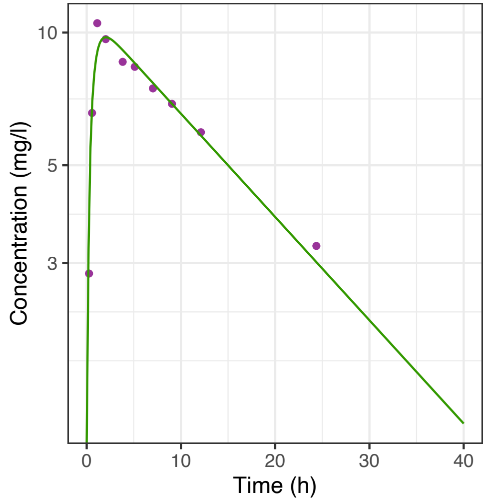
$$\begin{cases} \frac{dA_a}{dt} = -k_a A_a \\ \frac{dA}{dt} = k_a A_a - kA \\ A_a(t=0) = D,\ A(t=0) = 0 \end{cases} \qquad C(t) = \frac{A(t)}{V}.$$



$$A(t) = D\frac{k_a}{k_a - k}\left(e^{-kt} - e^{-k_a t}\right)$$

$$\begin{pmatrix} A_a{}' \\ A' \end{pmatrix} = \begin{pmatrix} -k_a & 0 \\ k_a & -k \end{pmatrix} \cdot \begin{pmatrix} A_a \\ A \end{pmatrix}$$

- Eigenvalues $-k_a$ and $-k$



$\Rightarrow A(t) =$ linear combination of $e^{-k_a t}$ and e^{-kt}

Nonlinear differential equations

$$y' = f(t, y)$$

- No general method of resolution
- Numerical approximation algorithms (solvers) have to be used
  - Euler method
  - Runge-Kutta methods
- Ex: nonlinear elimination (Michaelis – Menten)

$$\frac{dA}{dt} = -V_{max} \frac{A}{VK_m + A}$$