Lecture 1 - Part 2:
Singular Value Decomposition & Newton's method

Giancarlo Camarda

Institut national d'études démographiques

Ugofilippo Basellini

Max Planck Institute for Demographic Research

IDEM 117

Advances in Mortality Forecasting

International Advanced Studies in Demography

17 – 21 January, 2022

About SVD

- A SVD decomposition of any $m \times n$ rectangular matrix $\boldsymbol{M}$ is a factorization of the form:

$$\boldsymbol{M} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}'$$

where

  - $\boldsymbol{U}$ is a $m \times m$ orthogonal matrix (left-singular vectors)
  - $\boldsymbol{V}$ is a $n \times m$ orthogonal matrix (right-singular vectors)
  - $\boldsymbol{\Sigma}$ is a $m \times m$ diagonal matrix with $s_{ii} = 0$ if $i \neq j$ (s_{ii}, singular values)

- Equally we can write this decomposition as follows:

$$\boldsymbol{M} = U_{,1}\, s_{1,1}\, V'_{,1} + U_{,2}\, s_{2,2}\, V'_{,2} + U_{,3}\, s_{3,3}\, V'_{,3} + \ldots + U_{,m}\, s_{m,m}\, V'_{,m}$$

SVD: simple example

- Given a matrix:

$$M = \begin{pmatrix} 1 & 5 & 9 & 13 & 17 \\ 2 & 6 & 10 & 14 & 18 \\ 3 & 7 & 11 & 15 & 19 \\ 4 & 8 & 12 & 16 & 20 \end{pmatrix}$$

- Apply the SVD: $M = U \Sigma V'$

$$M = \underbrace{\begin{pmatrix} -0.44 & -0.71 & -0.31 & 0.45 \\ -0.48 & -0.26 & 0.75 & -0.36 \\ -0.52 & 0.18 & -0.56 & -0.62 \\ -0.55 & 0.63 & 0.12 & 0.53 \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} 53.52 & 0.00 & 0 & 0 \\ 0.00 & 2.36 & 0 & 0 \\ 0.00 & 0.00 & \approx 0 & 0 \\ 0.00 & 0.00 & 0 & \approx 0 \end{pmatrix}}_{\Sigma}$$
$$\underbrace{\begin{pmatrix} -0.10 & 0.77 & 0.63 & 0.08 \\ -0.25 & 0.49 & -0.60 & -0.45 \\ -0.39 & 0.21 & -0.40 & 0.75 \\ -0.54 & -0.07 & 0.10 & -0.48 \\ -0.69 & -0.35 & 0.28 & 0.10 \end{pmatrix}'}_{V}$$

- Computing the $m = 4$ matrices:

$$M = \begin{pmatrix} 2.29 & 5.82 & 9.35 & 12.89 & 16.42 \\ 2.48 & 6.31 & 10.13 & 13.96 & 17.78 \\ 2.67 & 6.79 & 10.91 & 15.03 & 19.15 \\ 2.86 & 7.27 & 11.69 & 16.10 & 20.51 \end{pmatrix} + \begin{pmatrix} -1.29 & -0.82 & -0.35 & 0.11 & 0.58 \\ -0.48 & -0.31 & -0.13 & 0.04 & 0.22 \\ 0.33 & 0.21 & 0.09 & -0.03 & -0.15 \\ 1.14 & 0.73 & 0.31 & -0.10 & -0.51 \end{pmatrix} +$$
$$+ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

A fascinating use of the SVD

- Given a linear system of equation: $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$ with $m \times n$ matrix $\boldsymbol{A}$
- We can have two cases:
  1. overdetermined problem ($m > n$):

     classic least-squares solution: $\hat{\boldsymbol{x}} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{b}$

  2. underdetermined problem ($n > m$):
- In both cases we can use $\boldsymbol{A}^{\dagger}$: the *Moore-Penrose* pseudo-inverse

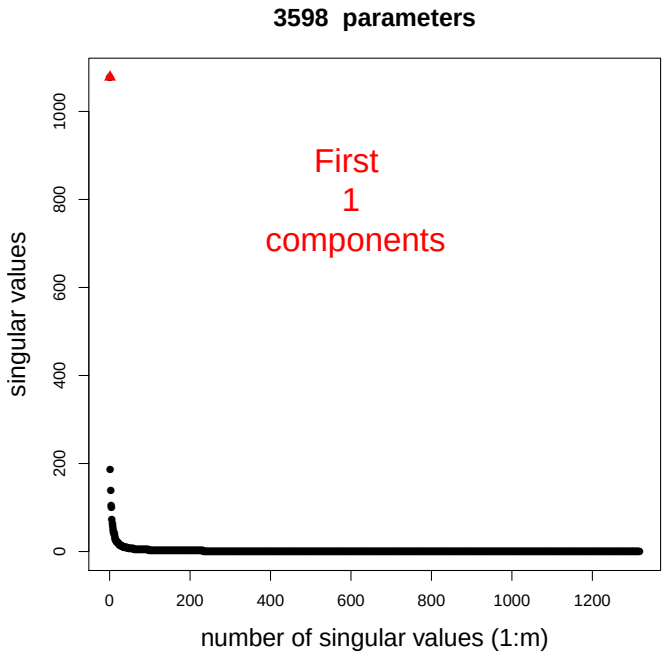
$$
\begin{aligned}
\boldsymbol{A}\,\boldsymbol{x} &= \boldsymbol{b} \quad \text{(apply SVD on } \boldsymbol{A}\text{)} \\
\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'\boldsymbol{x} &= \boldsymbol{b} \quad \text{(invert each element from SVD)} \\
\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'\boldsymbol{x} &= \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}'\boldsymbol{b} \quad (*) \\
\hat{\boldsymbol{x}} &= \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}'\boldsymbol{b} = \boldsymbol{A}^{\dagger}\boldsymbol{b}
\end{aligned}
$$

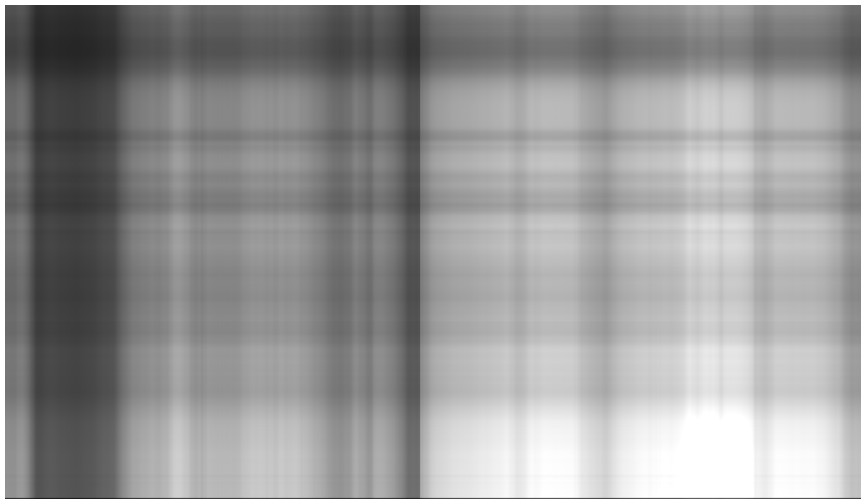
(*) $\boldsymbol{V}\boldsymbol{V}' = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma} = \boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}$

- In our two cases $\hat{\boldsymbol{x}}$ estimated using $\boldsymbol{A}^{\dagger}$ correspond to
  1. the least-squares solution: $\min\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$
  2. among the infinity solutions the one with $\min\|\boldsymbol{x}\|_2$

SVD as Data Reduction tool

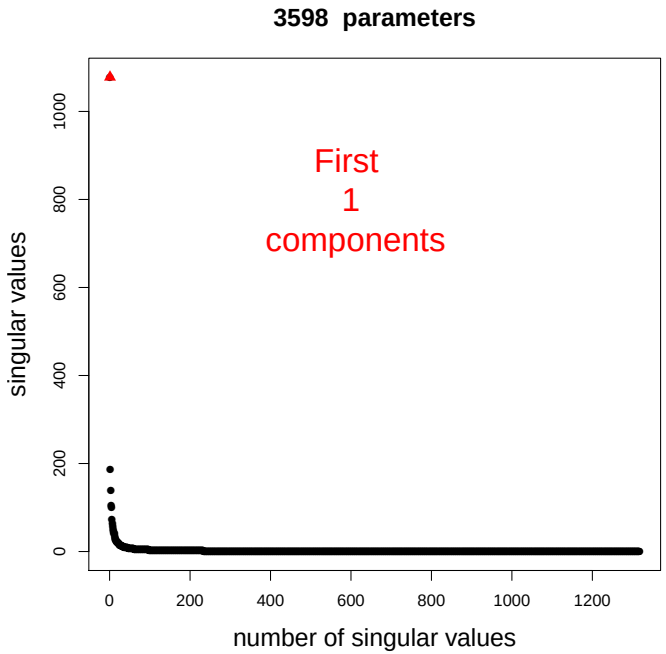
- Note that $s_{1,1} \geq s_{2,2} \geq s_{3,3} \geq \ldots \geq s_{m,m} \geq 0$
- The first singular vector(s)/value(s) are more important in describing the information in $\boldsymbol{M}$
- Here an example of a 1317×2281 (over 3 millions data-points)

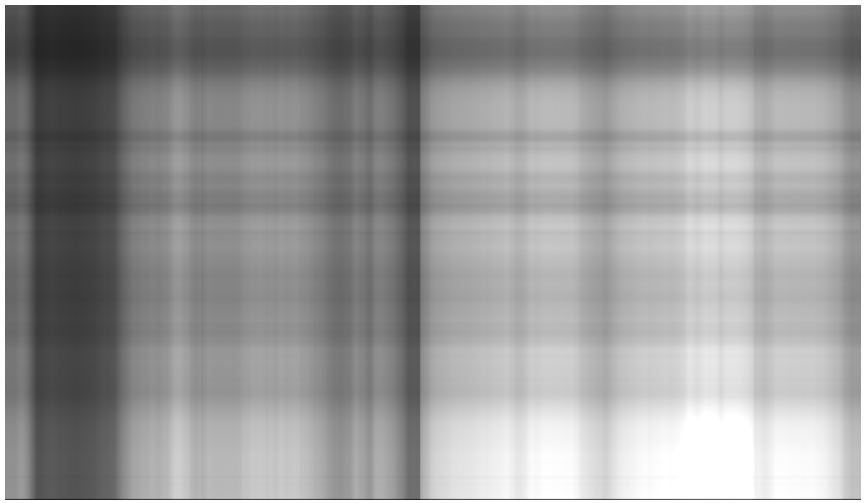




SVD as Data Reduction tool

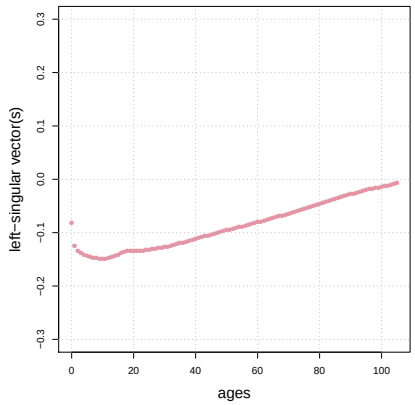
- Note that $s_{1,1} \geq s_{2,2} \geq s_{3,3} \geq \ldots \geq s_{m,m} \geq 0$
- The first singular vector(s)/value(s) are more important in describing the information in $\boldsymbol{M}$
- Here an example of a 1317×2281 (over 3 millions data-points)

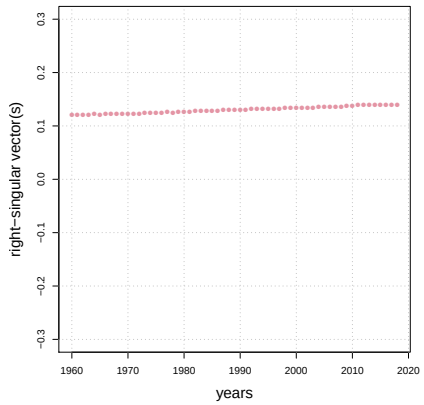


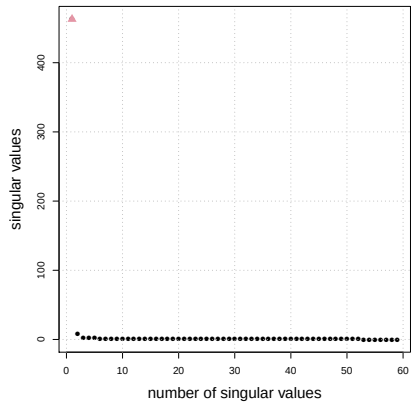


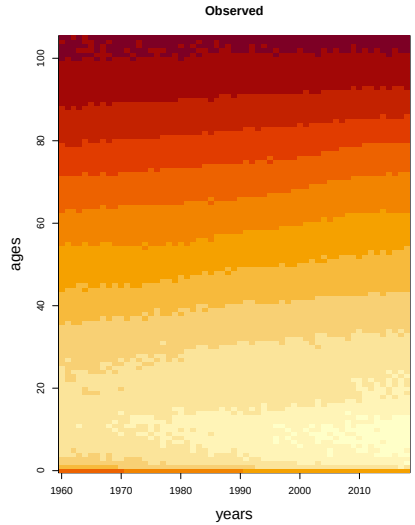
Reducing mortality data

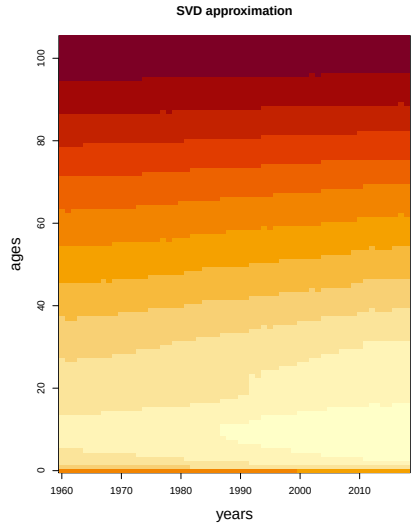
- Let's take the matrix of log-rates over m ages and n years
- And apply the SVD to reduce dimensionality
- Eng & Wal, females, ages 0-105, years 1960-2018 (6254 data-points)

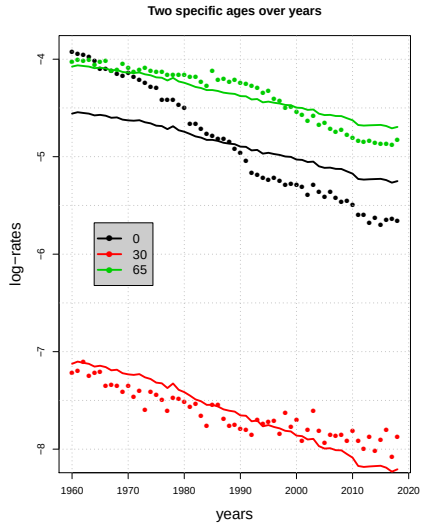












A general intro

- Newton's Method is an iterative equation solver to find the roots of a polynomial function
- It is very powerful and ubiquitous (e.g. the IWLS is based on it)
- Single-variable function: we aim to find x such that $f(x) = 0$
  1. Start with a (good) initial guess x_0
  2. Compute $f(x_0)$ and $f'(x_0)$
  3. Compute the new approximation with

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

  4. Repeat 2 and 3 until convergence, e.g. $f(x)$ close to 0
- Geometrically, $(x_{n+1}, 0)$ is the intersection of the x-axis and the tangent of the graph of $f(\cdot)$ at $(x_n, f(x_x))$: the linear approximation at x_n

A demographic example

- An important demographic relationship: the Lotka equation

$$1 = \int_{\alpha}^{\beta} e^{-ra}\, p(a)\, m(a)\, da$$

- Given $p(a)$ & $m(a)$, survival and maternity schedules over age a, it exists an intrinsic growth rate (r) in which such stable population will growth
- The Lotka equation is satisfied with

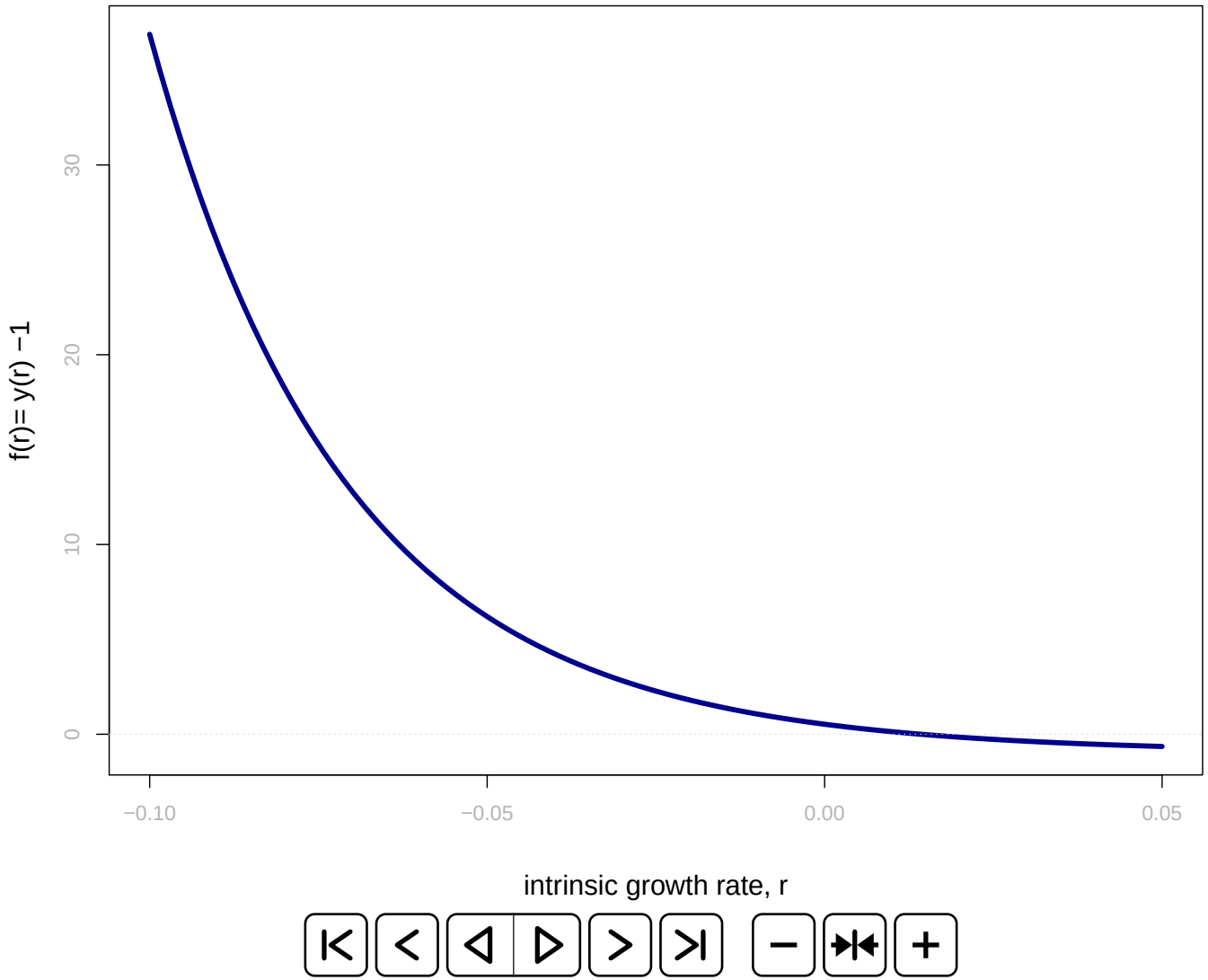
$$r : \quad f(r) = y(r) - 1 = 0$$

i.e. finding the root of $f(r)$ in which

$$f'(r) = y'(r) = -\int_{\alpha}^{\beta} a\, e^{-ra}\, p(a)\, m(a)\, da \qquad [= -\, y(r)\, A_B(r)]$$

- To be operative with discrete data, we need to convert integrals with summations

Estimating r for Egyptian data in 1997



Newton's method and the likelihood

- What happen if we need to maximize the log-likelihood, $\ell(\boldsymbol{\theta})$?
- We need to find when its derivative is zero $\Rightarrow$ Newton's method
  - For each parameter θ one after (and fixing) the other(s):

$$\theta_{n+1} = \theta_n - \frac{\partial \ell_n / \partial \theta}{\partial^2 \ell_n / \partial \theta^2} \quad \text{with} \quad \ell_n = \ell_n(\theta_n)$$

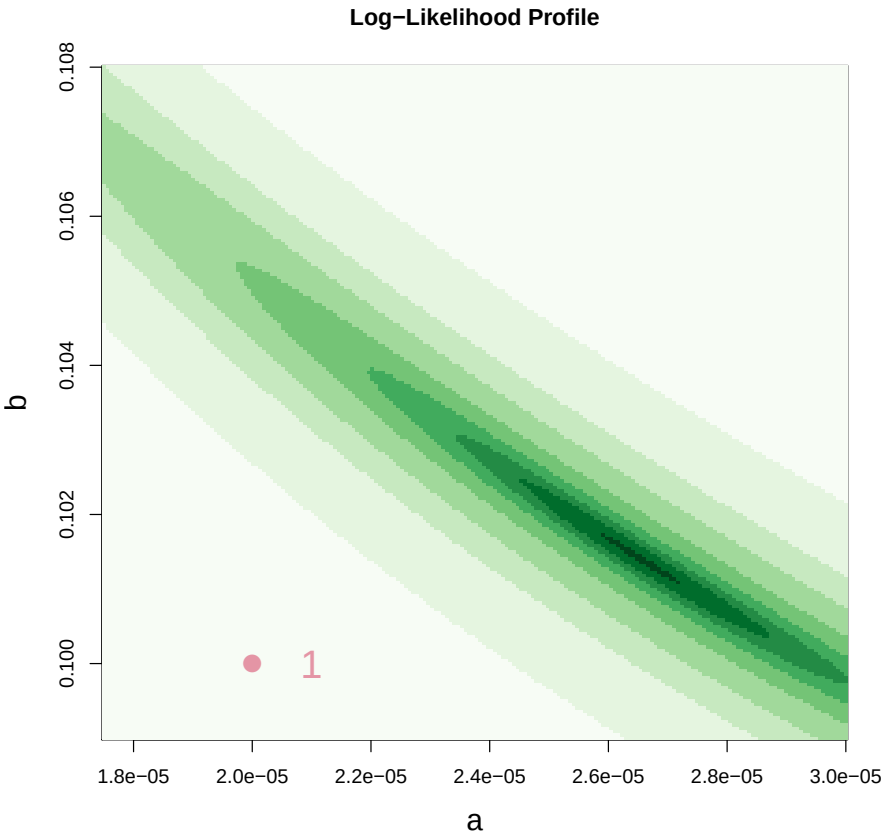
  - For all $\boldsymbol{\theta}$ simultaneously (generalization of the Newton's method to more dimensions):

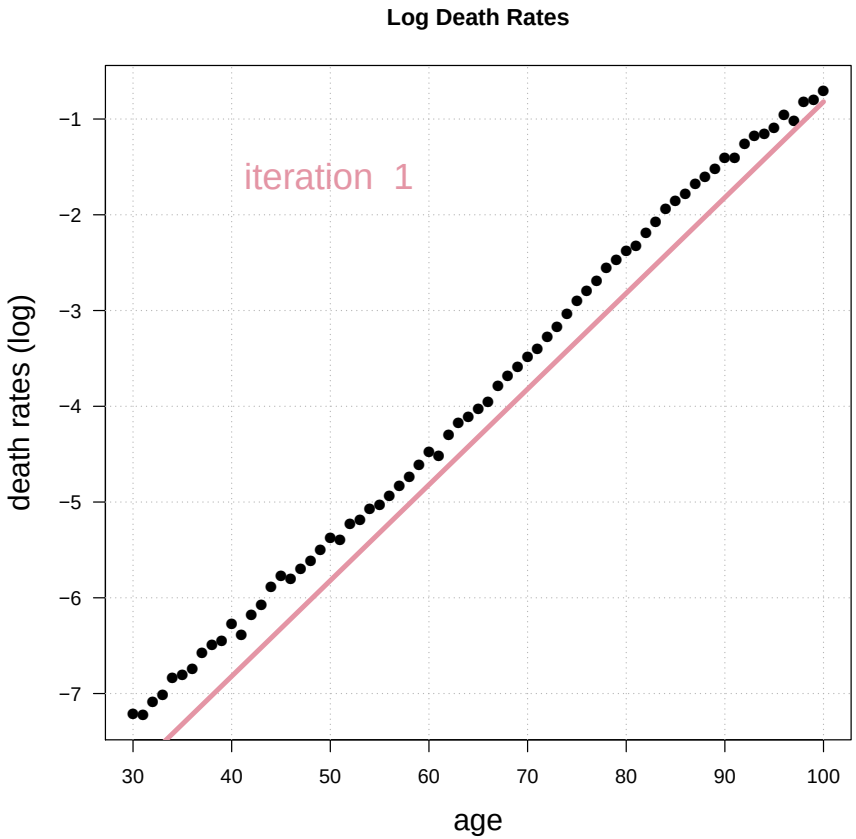
$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - H_{\ell_n}^{-1} \nabla \ell_n$$

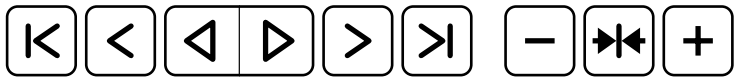
where, for example, in case of the two parameter $\text{Gompertz}(a, b)$:

$$\underbrace{\nabla \ell_n = \begin{pmatrix} \frac{\partial \ell_n}{\partial a} \\ \frac{\partial \ell_n}{\partial b} \end{pmatrix}}_{\text{Gradient}} \quad \text{and} \quad \underbrace{H_{\ell_n} = \begin{pmatrix} \frac{\partial^2 \ell_n}{\partial a^2} & \frac{\partial^2 \ell_n}{\partial a \partial b} \\ \frac{\partial^2 \ell_n}{\partial b \partial a} & \frac{\partial^2 \ell_n}{\partial b^2} \end{pmatrix}}_{\text{Hessian}}$$

Estimating Gompertz







Using the Newton's method to estimate a Gompertz model.
England & Wales, males, 2019, ages 30-100.
Source: Human Mortality Database.