Extreme points of the Vandermonde determinant and phenomenological modelling with power exponential functions

Karl Lundengård

Division of Applied Mathematics, UKK, Mälardalen University

September 26, 2019

Information about the dissertation

- Based on 9 papers, referred to as Paper A-I.
- A section that is based on a paper consists of text from the paper unchanged except for minor modifications.
- Contents are rearranged to clarify the relations between different papers and parts of several papers have been omitted to avoid repetition and improve cohesion.
- Two main topics:
  - Optimizing the Vandermonde determinant on a surface
  - Phenomenological modelling with power-exponential functions
    - Approximation of electrostatic discharge currents
    - Approximation of mortality rate curves
- Each slide has the corresponding section (or page) in the dissertation in the header.

Paper A. K. L., Jonas Österberg and Sergei Silvestrov.
Extreme points of the Vandermonde determinant on the sphere and some limits involving the generalized Vandermonde determinant.

Paper B. K. L., Jonas Österberg and Sergei Silvestrov.
Optimization of the determinant of the Vandermonde matrix on the sphere and related surfaces.

Paper C. Asaph Keikara Muhumuza, K. L., Jonas Österberg, Sergei Silvestrov, John Magero Mango, Godwin Kakuba.
Extreme points of the Vandermonde determinant on surfaces implicitly determined by a univariate polynomial.

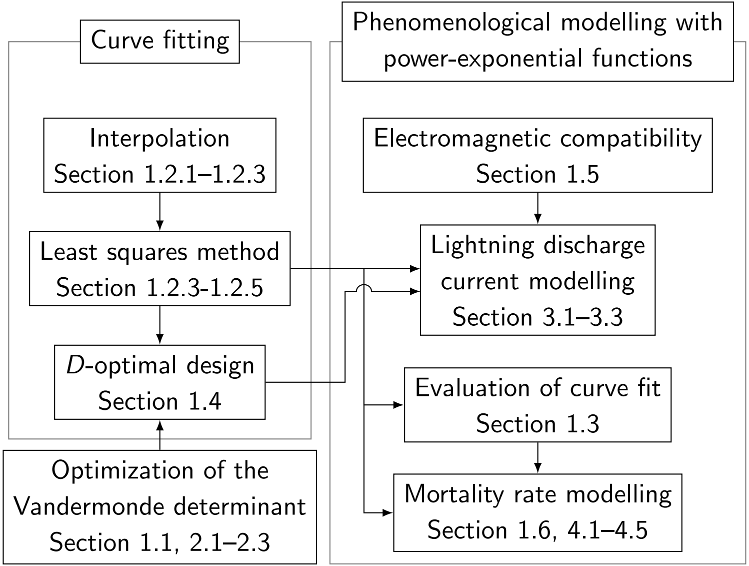
Paper D. Asaph Keikara Muhumuza, K. L., Jonas Österberg, Sergei Silvestrov, John Magero Mango, Godwin Kakuba.
Optimization of the Wishart joint eigenvalue probability density distribution based on the Vandermonde determinant.

Papers: Phenomenological modelling p. 13–14

Paper E. K. L., Milica Rančić, Vesna Javor, Sergei Silvestrov.
On some properties of the multi-peaked analytically extended function for approximation of lightning discharge currents.
Paper F. K. L., Milica Rančić, Vesna Javor, Sergei Silvestrov.
Estimation of parameters for the multi-peaked AEF current functions.
Paper G. K. L., Milica Rančić, Vesna Javor, Sergei Silvestrov.
Electrostatic discharge currents representation using the analytically extended function with p peaks by interpolation on a D-optimal design.

Paper H. K. L., Milica Rančić, Sergei Silvestrov.
Modelling mortality rates using power-exponential functions.
Paper I. Andromachi Boulougari, K. L., Milica Rančić, Sergei Silvestrov, Belinda Straß, Samya Suleiman.
Application of a power-exponential function based model to mortality rates forecasting.

Structure of the dissertation



- A Vandermonde matrix is an $m \times n$ matrix of the form

$$\mathbf{V}_{mn}(\mathbf{x}) = \left[x_j^{i-1}\right]_{i,j}^{m,n} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{m-1} & x_2^{m-1} & \cdots & x_n^{m-1} \end{bmatrix}$$

  where $x_i \in \mathbb{R}$, $i = 1, \ldots, n$. If the matrix is square, $n = m$, the notation $\mathbf{V}_n = \mathbf{V}_{nn}$ will be used.
- Alexandre Théophile Vandermonde (1735–1796) who was a French lawyer, violinist, chemist, politician, economist and (briefly) mathematician.
- The Vandermonde determinant, $v_n(x_1, \ldots, x_n)$, is given by

$$v_n(\mathbf{x}) = \det(\mathbf{V}_n(x_1, \ldots, x_n)) = \prod_{1 \leq i < j \leq n} (x_j - x_i).$$

- In a generalized Vandermonde matrix we allow any sequences of exponents. There are many other generalizations e.g. Alternant matrix, Jacobian matrix, Wronskian matrix, Bell matrix, Moore matrix.
- The Vandermonde determinant appears in many applications, e.g. Lagrange interpolation, Fekete points and Coulomb gas system.
- Important examples of Coulomb gas systems are distributions of charged particles, sphere packing and various types of systems in random matrix theory.
- For example: Wishart ensembles are random matrices whose eigenvalues have a joint probability distribution given by

$$\mathbb{P}_\beta(\boldsymbol{\lambda}) = C_N^{\beta,\alpha} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_i \lambda_i^{\alpha-p} \exp\left(-\frac{1}{2}\sum_{i=1}^N \lambda_i^2\right)$$

where $\alpha = \frac{\beta}{2}m$, $p = 1 + \frac{\beta}{2}(N-1)$ and β is determined by the type of elements in the matrix. It can be shown that maximizing $\mathbb{P}_\beta(\lambda)$ is equivalent to maximize v_n on a sphere.

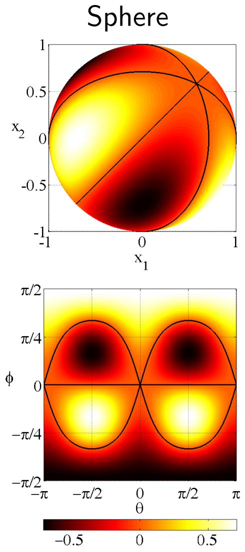
- Here we will focus on the Vandermonde determinant, more specifically we want examine its maximum and minimum values.
- The Vandermonde determinant is a homogeneous polynomial

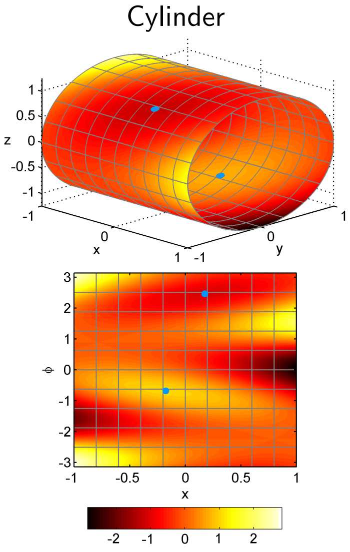
$$v_n(c\mathbf{x}) = \prod_{1 \leq i < j \leq n} (cx_j - cx_i) = c^{\frac{n(n-1)}{2}} v_n(\mathbf{x})$$

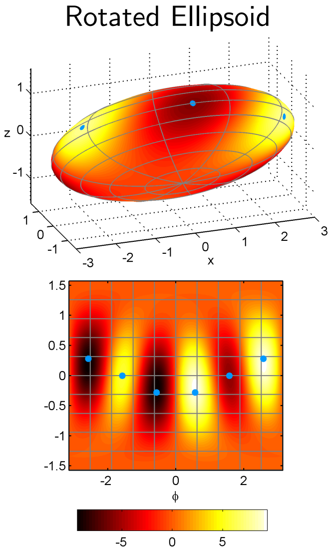
  so it is clearly unbounded and there are no global maximum or minimum.
- If we constrain $\mathbf{x}$ to a bounded volume we can use the homogeneity to show that the extreme points must lie on the surface of the volume.

- For $f(\mathbf{x})$ with $\mathbf{x} \in \{\mathbf{x} \in \mathbb{R}^n | g(\mathbf{x}) = 0\}$ then any $\mathbf{x}$ such that
$$\frac{\partial f}{\partial x_k} = \lambda \frac{\partial g}{\partial x_k}, \ 1 \leq k \leq n.$$
will be stationary points of f.
- The partial derivatives of v_n can be written $\displaystyle\frac{\partial v_n}{\partial x_k} = \sum_{\substack{i=1 \\ i \neq k}}^{n} \frac{v_n(\mathbf{x})}{x_k - x_i}, \ 1 \leq k \leq n.$
- Note that $\displaystyle\sum_{k=1}^{n} \frac{\partial v_n}{\partial x_k} = 0.$
- Combining the equality above and the method of Lagrange multipliers gives that for any stationary point of v_n
$$g(\mathbf{x}) = 0 \text{ and } \sum_{k=1}^{n} \frac{\partial g}{\partial x_k} = 0.$$

Extreme points in 3D







- Find the extreme points of v_n on a surface implicitly defined by
$$g_R(\mathbf{x}) = \sum_{i=1}^{n} R(x_i) = 0, \text{ where } R(x) = \sum_{i=0}^{m} r_i x^i, \quad r_i \in \mathbb{R}.$$
- Let $(x_1, \ldots, x_n)$ be the coordinates of a stationary point and define $f(x) = \prod_{i=1}^{n}(x - x_i)$ and then compare the expression for $\frac{f''(x_j)}{f'(x_j)}$ with the equation system given by applying the method of Lagrange multipliers to our optimization problem we get a differential equation
$$f''(x) - 2\rho R'(x) f'(x) - P(x) f(x) = 0$$
where $P(x)$ is a polynomial of degree $m-2$.
- In some cases finding the coefficients of $P(x)$ and solving the differential equation is easier that solving the equation system given by Lagrange multipliers directly.

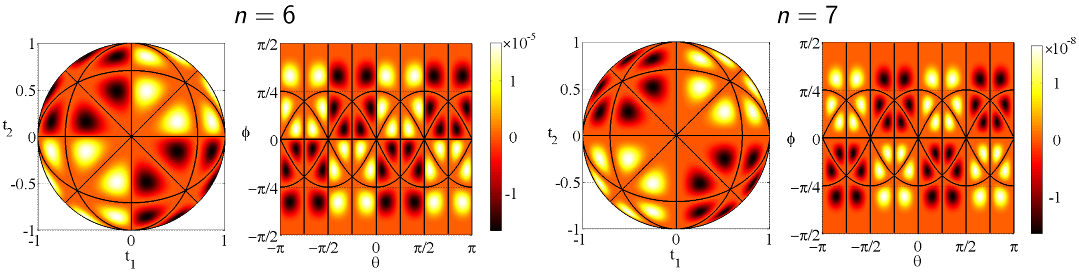
Extreme points as roots of orthogonal polynomials

- With $\sum_{i=1}^{n} \left(\frac{1}{2} x_i^2 + r_1 x_i + r_0 \right) = 0$ the extreme points of v_n are given by the roots of

$$f(x) = H_n \left(\left(\frac{n-1}{2(r_1^2 - 2r_0)} \right)^{\frac{1}{2}} \frac{(x + r_1)}{2} \right) = n! \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^i}{i!} \left(\frac{n-1}{2(r_1^2 - 2r_0)} \right)^{\frac{n-2i}{2}} \frac{(x + r_1)^{n-2i}}{(n-2i)!}$$

  where H_n is the nth (physicist) Hermite polynomial.

- We can use some symmetries of the roots to visualize the results in $n \leq 7$ dimensions.



- Extreme points of v_n on a surface implicitly defined by $\sum_{i=1}^{n} x_i^p = 1$ with even n and p.
- Coefficient matching equations can be reduced to $\frac{n-2}{2}$ equations. For low dimensions the resulting system can be solved.
- General expression unkown. The roots of f_p^n gives the extreme points.

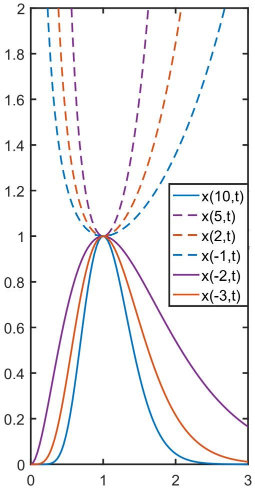
$$f_2^4(x) = x^4 - \frac{1}{2}x^2 + \frac{1}{48},$$
$$f_4^4(x) = x^4 - \frac{\sqrt{6}}{3}x^2 + \frac{1}{12},$$
$$f_6^4(x) = x^4 - \frac{1}{4}(\sqrt{33}+1)^{\frac{1}{3}}x^2 + \frac{1}{96}\left(9-\sqrt{33}\right)\left(\sqrt{33}+1\right)^{\frac{2}{3}}$$
$$f_8^4(x) = x^4 - \frac{\sqrt{3}}{6}(30\sqrt{5}-30)^{\frac{1}{4}}x^2 + \frac{1}{120}\left(\sqrt{5}-5\right)\sqrt{30\sqrt{5}-30}$$

$$f_2^6(x) = x^6 - \frac{1}{2}x^4 + \frac{1}{20}x^2 - \frac{1}{1800}$$
$$f_4^6(x) = x^6 - \frac{\sqrt{50+20\sqrt{5}}}{10}x^4 + \frac{\sqrt{5}}{10}x^2 - \frac{(-4+2\sqrt{5})\sqrt{50+20\sqrt{5}}}{600}$$

$$f_2^8(x) = x^8 - \frac{1}{2}x^6 + \frac{15}{224}x^4 - \frac{15}{6272}x^2 + \frac{15}{1404928},$$
$$f_4^8(x) = x^8 - \frac{\sqrt{140+42\sqrt{6}}}{14}x^6 + \left(\frac{3}{28}+\frac{3\sqrt{6}}{28}\right)x^4 - \left(\frac{-(140+42\sqrt{6})^{\frac{3}{2}}}{16464} + \frac{29\sqrt{140+42\sqrt{6}}}{2352}\right)x^2 - \frac{3}{3136} + \frac{\sqrt{6}}{1568}$$

Phenomenological modelling

- A phenomenological modelling is a model that can approximately describe a phenomena without explaining the phenomena.
- Challenges in engineering
  - A Identify the problem
    - Experiments, Analysis, Modelling, Experience, Simulation *etc*
  - B Understand and describe causes of problem
    - Physics, Chemistry, Other suitable theory, *etc*
  - C Solve the problem
    - Mathematics, Numerical methods, Design, Construction *etc*
  - D Ensure solution is practical
    - Resource constraints, Safety, Noise, Heat, Environmental concerns *etc*
  - E Realize solution
    - Manufacturing, Cost, Shipping, Distribution, Logistics *etc*
- Phenomenological modelling assists in achieving C and D when B is difficult.

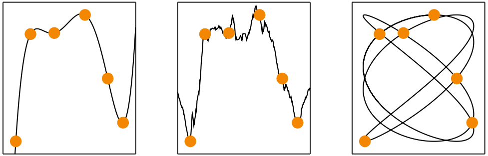


- We will build phenomenological models from what we call the power-exponential function

$$x(\beta; t) = \left(te^{1-t}\right)^{\beta}, \; 0 \leq t.$$

- The phenomenological models will be constructed by linear combinations of piecewise scaled and translated power-exponential functions.
- Two applications
  - We will model electrostatic discharges using power-exponential functions with $\beta > 0$.
  - We will model mortality rates using a linear combination of a power-exponential function with $\beta = -1$ and power-exponential functions with $\beta > 0$.

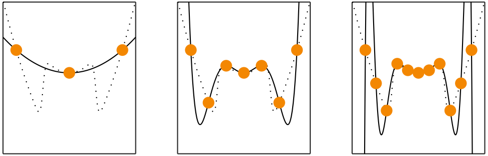
- An interpolation problem is the problem of finding a function that generates a given set of points.
- Many different functions can be used for interpolation.



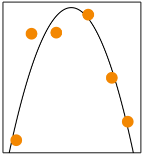
- The Vandermonde matrix appears when interpolating with polynomials.
- Similar approach can be used with other sets of basis function.

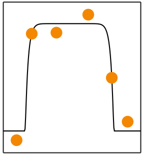
$$\begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

- It is easy to construct an interpolating polynomial but the result can be unstable when interpolating many points unless the points are chosen carefully.



- A least squares fitting does not generate the exact points, instead the sum of the square of the residuals, $\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta}; x_i))^2$, is minimized.



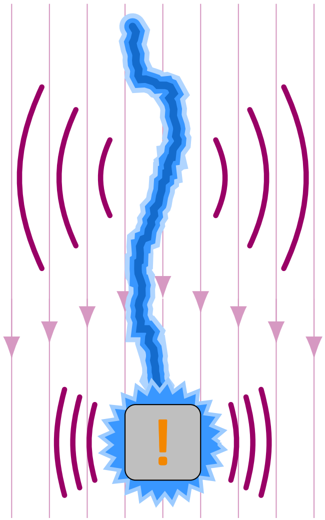


- Least squares fitting is useful when the data is noisy, i.e. the data points $\{(x_i, y_i), i = 1, \ldots, n\}$ are described by $y_i = f(\boldsymbol{\beta}; x_i) + \epsilon_i$ where $f(\boldsymbol{\beta}; x)$ is a given function with parameters $\boldsymbol{\beta}$ and ϵ_i are normally distributed i.i.d. random variables with $\mathrm{E}[\epsilon_i] = 0$, then taking the maximum likelihood estimation of the parameters is the same as find the least squares fit.
- If $f(\boldsymbol{\beta}; x)$ is a polynomial then the least squares fitting problem involves rectangular Vandermonde matrices.

EMC and ESD

- Electromagnetic compatibility (EMC) is the study and design of systems that are not susceptible to disturbances from other systems and does not cause interference with other systems or themselves.
- Important examples include:
  - Communication equipment and standards that do not interrupt each other.
  - Control systems that are resistant to outside influence.
  - Clothing, tools or other equipment can generate charge imbalances or sparks.
- An electrostatic discharge (ESD) is a sudden flow of charge from one object to another, often accompanied by an electrical spark.
- Most familiar examples of ESDs are probably
  - lightning discharges,
  - human-to-object discharges.
- Two approaches for phenomenological modelling
  - Numerically solving a non-linear least squares problem.
  - Interpolation on a D-optimal design.

Electromagnetic disturbances



- Suppose we have an engineered component.
- This component is struck by lightning. What electromagnetic phenomena can cause disturbances?
  1. The discharge current passing through the component.
  2. The component emitting electromagnetic radiation as the current passes through it.
  3. Emission from the lightning channel itself.
  4. Discharge changes electric potential between cloud and ground causing transient changes in the electric field.
- Typically very difficult to observe and model.
- There are standards that describe typical discharge currents and how components should react to them.

The p-peaked AEF

Let $I_{m_q} \in \mathbb{R}$, $t_{m_q} \in \mathbb{R}$, $q = 1, 2, \ldots, p$, $t_{m_0} = 0 < t_{m_1} < t_{m_2} < \ldots < t_{m_p}$ along with $\eta_{q,k}, \beta_{q,k} \in \mathbb{R}$ and $0 < n_q \in \mathbb{Z}$ for $q = 1, 2, \ldots, p+1$, $k = 1, 2, \ldots, n_q$ such that $\sum_{k=1}^{n_q} \eta_{q,k} = 1$.

The *analytically extended function* (AEF), $i(t)$, with p peaks is defined as

$$i(t) = \begin{cases} \left(\sum_{k=1}^{q-1} I_{m_k}\right) + I_{m_q} \sum_{k=1}^{n_q} \eta_{q,k} x_q(t)^{\beta_{q,k}^2+1}, \; t_{m_{q-1}} \leq t \leq t_{m_q}, \; 1 \leq q \leq p, \\ \left(\sum_{k=1}^{p} I_{m_k}\right) \sum_{k=1}^{n_{p+1}} \eta_{p+1,k} x_{p+1}(t)^{\beta_{p+1,k}^2}, \quad t_{m_p} \leq t, \end{cases}$$

where $x_q(t) = \dfrac{t - t_{m_{q-1}}}{\Delta t_{m_q}} \exp\left(\dfrac{t_{m_q} - t}{\Delta t_{m_q}}\right)$, $1 \leq q \leq p$,

$x_{p+1}(t) = \dfrac{t}{t_{m_q}} \exp\left(1 - \dfrac{t}{t_{m_q}}\right)$ and $\Delta t_{m_q} = t_{m_q} - t_{m_{q-1}}$.

Examples of a 2-peak AEF

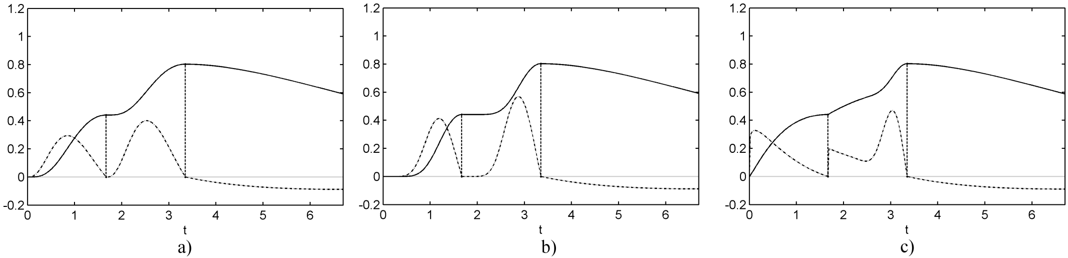
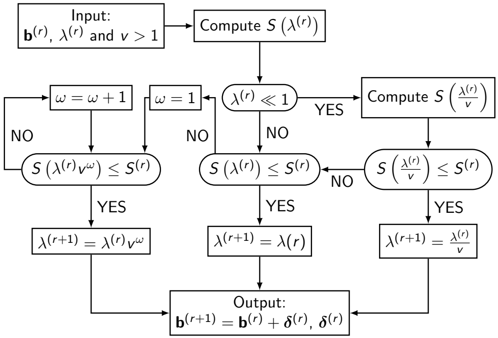
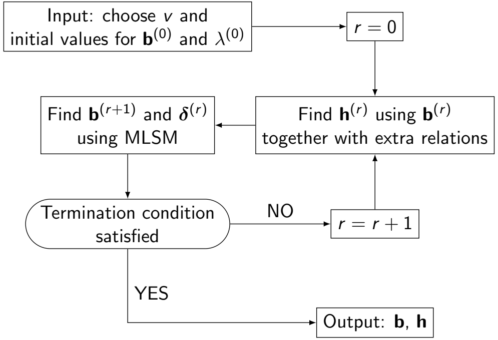


Figure: AEF (solid) and its derivative (dashed) with the same I_{m_q} and t_{m_q}.
(a) $4 < \beta_{q,k} < 5$, (b) $12 < \beta_{q,k} < 13$, (c) a mixture of large and small $\beta_{q,k}$-parameters.

- We will use two approaches to fit the AEF to data.
  - Least squares fitting using the Marquardt Least Squares Method (MLSM).
  - Interpolation on a *D*-optimal design.

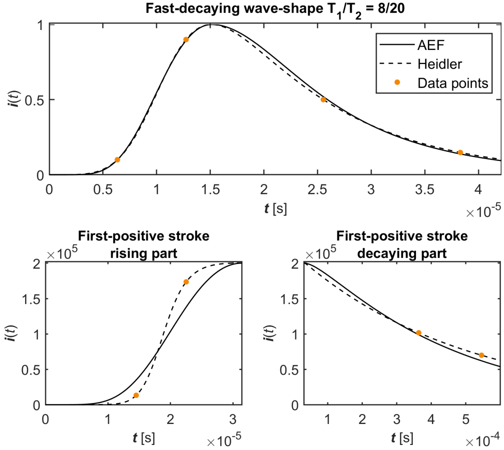


The basic iteration step of the Marquardt least-squares method.

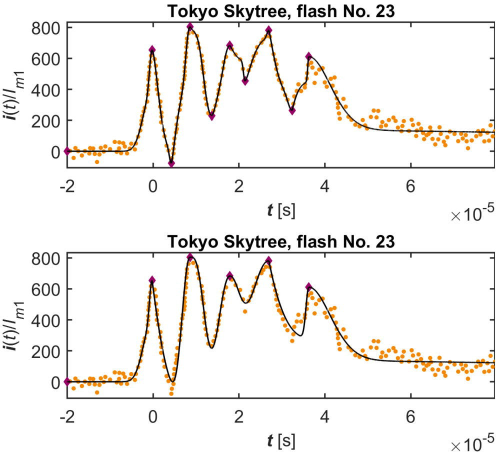


Schematic description of the parameter estimation algorithm.

Fitting to a multi-peaked waveshape



AEF fitted to two waveshapes from the IEC 62305-1 standard.



AEF fitted to measurements of a lightning discharge hitting a skyscraper.

- We will there try another approach with the goal of getting a good and reliable approximation using only a few carefully chosen data points.
- Finding the least squares fitting is equivalent to taking the maximum-likelihood estimation of the parameters that specify the fitted function.
- Thus the result of the fitting is also sensitive to noise in the data.
- The independent coordinates for the data, $\{x_i, i = 1, \ldots, n\}$ are called a design and choosing the design that minimizes the variance of the values predicted by the regression model is called G-optimality.
- The design that minimizes the variance of the parameters of the regression model is called D-optimality.
- The Kiefer–Wolfowitz equivalence theorem says that for a typical linear regression model there exist a D-optimal design which is also G-optimal.

D-optimal experiment design

- A design ξ is said to be D-optimal if it maximizes the determinant of the Fisher information matrix $\mathbf{M}(\beta) = -\mathrm{E}_X\left[\frac{\partial^2}{\partial\beta_i\partial\beta_j}\ln(f(\beta(\xi);X))\right]_{1,1}^{n,n}$.
- For an interpolating polynomial regression problem the Fisher information matrix is given by

$$\mathbf{M}(\beta) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix}$$

  and thus by the Cauchy–Binet formula $\det(\mathbf{M}(\beta)) = v_n(\mathbf{x})^2$.
- Thus finding a D-optimal design for an interpolating polynomial regression problem is equivalent to optimizing the determinant of the Vandermonde matrix in some volume given by the set of possible designs.

Finding a D-optimal on an interval for the AEF

- Consider the AEF between two peaks, $i(t) = \sum_{m=1}^{n} \eta_m t^{\beta_m} e^{\beta_m(1-t)}$.
- Set $\beta_m = \frac{k+m-1}{c}$ and $z(t) = (te^{1-t})^{\frac{1}{c}}$ then $i(t) = \sum_{m=1}^{n} \eta_m z(t)^{k+m-1}$.
- If we have n sample points, t_m, $m = 1, \ldots, n$, then the Fisher information matrix is $\mathbf{M} = \mathbf{U}^\top \mathbf{U}$ where

$$\mathbf{U} = \begin{bmatrix} z(t_1)^k & \ldots & z(t_n)^k \\ \vdots & \ddots & \vdots \\ z(t_1)^{k+n-1} & \ldots & z(t_n)^{k+n-1} \end{bmatrix}.$$

- $\mathbf{U}$ is a generalized Vandermonde matrix and with $z_i = z(t_i)$ it has determinant

$$\det(\mathbf{U}) = \left(\prod_{k=1}^{n} z_k\right)\left(\prod_{1 \leq i < j \leq n} (z_j - z_i)\right) u_n(k; z_1, \ldots, z_n).$$

D-Optimal interpolation on the rising part

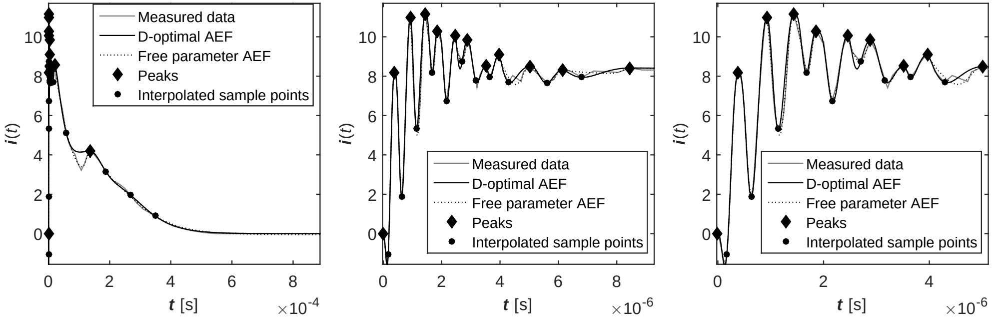
- This determinant can be maximized using a technique similar to the one described previously.
- The determinant

$$u_n(k; z_1, \ldots, z_n) = \left(\prod_{i=1}^{n} z_i^k\right) \left(\prod_{1 \leq i < j \leq n} (z_j - z_i)\right)$$

is maximized or minimized on the cube $[0,1]^n$ when $z_1 < \ldots < z_{n-1}$ are roots of the Jacobi polynomial

$$P_{n-1}^{(2k-1,0)}(1-2z) = \frac{(2k)^{\overline{n-1}}}{(n-1)!} \sum_{i=0}^{n-1} (-1)^n \binom{n-1}{i} \frac{(2k+n)^{\overline{i}}}{(2k)^{\overline{i}}} z^i$$

and $z_n = 1$, or some permutation thereof. Here $a^{\overline{b}}$ is the rising factorial $a^{\overline{b}} = a(a+1)\cdots(a+b-1)$.
- With some modification the same technique also works on the decaying part.



Results of fitting an AEF with 13 peaks and two terms in each interval to lightning discharge data from Mount Säntis in Switzerland.

- Survival function is defined as

$$S_x(\Delta x) = \Pr[T_x > \Delta x].$$

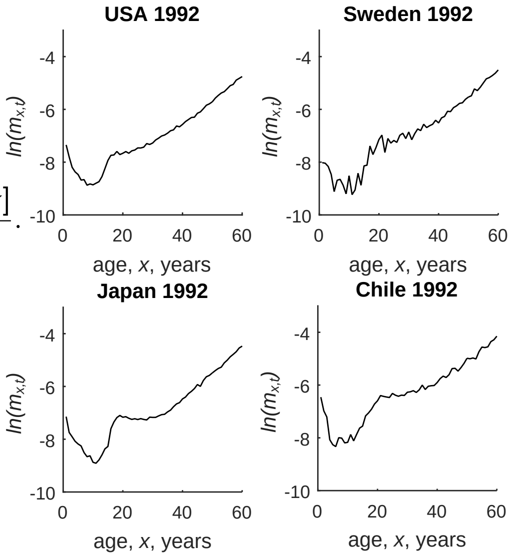
- Mortality rate is defined as

$$\mu(x) = \lim_{\mathrm{d}x \to 0^+} \frac{\Pr[T_0 > x | T_0 \leq x + \mathrm{d}x]}{\Pr[T_0 > x]}.$$

- S_x and $\mu(x)$ relate to each other

$$S_x(\Delta x) = \exp\left(-\int_x^{x+\Delta x} \mu(t)\,\mathrm{d}t\right).$$

- t current year, d_x deaths, L_x living population, central mortality rate is $m_{x,t} = \frac{d_x}{L_x}$ and assume $m_{x,t} \approx \mu(x)$.



Mortality rate models I

Gompertz–Makeham	$\mu(x) = a + be^{cx}$
Weibull	$\mu(x) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1}$
Logistic	$\mu(x) = \frac{ae^{bx}}{1 + \frac{ac}{b}(e^{bx} - 1)}$
Modified Perks	$\mu(x) = \frac{a}{1 + e^{b-cx}} + d$
Gompertz inverse Gaussian	$\mu(x) = \frac{e^{a-bx}}{\sqrt{1 + e^{-c+bx}}}$
Double Geometric	$\mu(x) = a + b_1 b_2^x + c_1 c_2^x$
Thiele	$\mu(x) = a_1 e^{-b_1 x} + a_2 e^{-b_2 \frac{(x-c)^2}{2}} + a_3 e^{b_3 x}$
Heligman–Pollard 1	$\mu(x) = a_1^{(x+a_2)^{a_3}} + b_1 e^{-b_2 \ln\left(\frac{x}{b_3}\right)^2} + c_1 c_2^x$
Heligman–Pollard 2	$\mu(x) = a_1^{(x+a_2)^{a_3}} + b_1 e^{-b_2 \ln\left(\frac{x}{b_3}\right)^2} + \frac{c_1 c_2^x}{1 + c_1 c_2^x}$
Heligman–Pollard 3	$\mu(x) = a_1^{(x+a_2)^{a_3}} + b_1 e^{-b_2 \ln\left(\frac{x}{b_3}\right)^2} + \frac{c_1 c_2^x}{1 + c_3 c_1 c_2^x}$
Heligman–Pollard 4	$\mu(x) = a_1^{(x+a_2)^{a_3}} + b_1 e^{-b_2 \ln\left(\frac{x}{b_3}\right)^2} + \frac{c_1 c_2^{x^{c_3}}}{1 + c_1 c_2^{x^{c_3}}}$

Hannerz $\mu(x) = \dfrac{f(x)}{1+F(x)}$ with $f(x) = \alpha \dfrac{g_1(x)e^{G_1(x)}}{(1+e^{G_1(x)})^2} + (1-\alpha)\dfrac{g_2(x)e^{G_2(x)}}{(1+e^{G_2(x)})^2}$,

$$F(x) = \alpha \frac{e^{G_1(x)}}{1+e^{G_1(x)}} + (1-\alpha)\frac{e^{G_2(x)}}{1+e^{G_2(x)}},$$

$$g_1(x) = \frac{a_1}{x^2} + a_2 x + a_3 e^{cx},\ G_1(x) = a_0 - \frac{a_1}{x} + \frac{a_2 x^2}{2} + \frac{a_3}{c}e^{cx},$$

$$g_2(x) = \frac{a_5}{x^2} + a_2 x + a_3 e^{cx} \text{ and } G_2(x) = a_4 - \frac{a_5}{x} + \frac{a_2 x^2}{2} + \frac{a_3}{c}e^{cx}$$

First Time Exit Model: SKI-6

$$\mu(x) = \frac{g(x)}{\displaystyle\int_x^\infty g(t)\,\mathrm{d}t} \text{ with } g(x) = \frac{k}{\sqrt{x^3}}\exp\left(-\frac{H_x^2}{2x}\right),\ H(x) = a_1 + ax^4 - b\sqrt{x} + lx^2 - cx^3$$

First Time Exit Model: Fractional 1st order approximation

$$\mu(x) = \frac{g(x)}{\displaystyle\int_x^\infty g(t)\,\mathrm{d}t} \text{ where } g(x) = \frac{2|l+(c-1)(bx)^c|}{\sigma\sqrt{2\pi x^3}}\exp\left(-\frac{-(l-(bx)^c)^2}{2\sigma^2 x}\right)$$

First Time Exit Model: Fractional 2nd order approximation

$$\mu(x) = \frac{g(x)}{\displaystyle\int_x^\infty g(t)\,\mathrm{d}t} \text{ where } g(x) = \frac{2}{\sigma\sqrt{2\pi x}}\left(\frac{2|l+(c-1)(bx)^c|}{\sigma\sqrt{2\pi x}} + k\frac{c(c-1)(bx)^c}{2|l+(c-1)(bx)^c|}\right)\exp\left(-\frac{-(l-(bx)^c)^2}{2\sigma^2 x}\right)$$

Power-exponential

$$\mu(x) = \frac{c_1}{xe^{-c_2 x}} + a_1 \left(xe^{-a_2 x}\right)^{a_3}$$

Split power-exponential

$$\mu(x) = \frac{\tilde{c}}{xe^{-c_2 x}} + a_1 \left(xe^{-a_2 x}\right)^{\tilde{a}} + \theta\left(x - \frac{1}{c_2}\right) \cdot c_2 \cdot e \cdot (c_1 - c_3) \text{ where}$$

$$\tilde{c} = \begin{cases} c_1, & x \leq \frac{1}{c_2} \\ c_3, & x > \frac{1}{c_2} \end{cases}, \quad \tilde{a} = \begin{cases} a_3, & x \leq \frac{1}{a_2} \\ a_4, & x > \frac{1}{a_2} \end{cases}, \quad \theta(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}.$$

Adjusted power-exponential

$$\mu(x) = c_1 \left(\frac{e^{c_2 x}}{c_2 x}\right)^{\tilde{c}} + a_1 \left(xe^{-a_2 x}\right)^{\tilde{a}} \text{ where } \tilde{c} = \begin{cases} c_3, & x \leq \frac{1}{c_2}, \\ c_4, & x > \frac{1}{c_2}, \end{cases} \text{ and } \tilde{a} = \begin{cases} a_3, & x \leq \frac{1}{a_2}, \\ a_4, & x > \frac{1}{a_2}. \end{cases}$$

Comparing models with the AIC

- When comparing the different models we need to take into account that the models have different numbers of parameters.

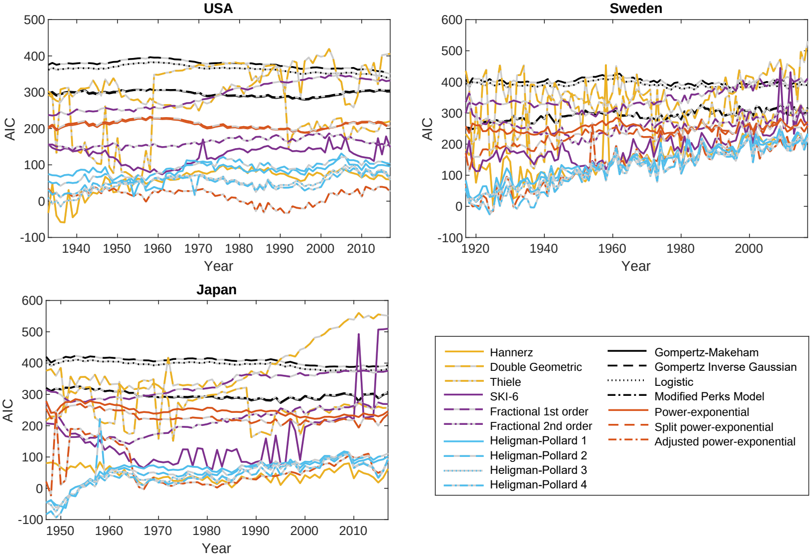
> I remember my friend Johnny von Neumann used to say,
> 'with four parameters I can fit an elephant,
> and with five I can make him wiggle his trunk'.
> - Freeman Dyson, quoting Enrico Fermi

- A common way to do this is to use Akaike's Information Criterion (AIC).
- Let f be a model of some data, y, with k estimated parameters and let $\hat{L}(f|y)$ be the maximum value of the likelihood function for the model. Then the AIC is given by

$$\mathrm{AIC}(f|y) = 2(k+1) - 2\log\left(\hat{L}(f|y)\right).$$

- The previously mentioned mortality rate models were fitted to data from the seven countries and the AIC was computed for each year.

Model fit comparison

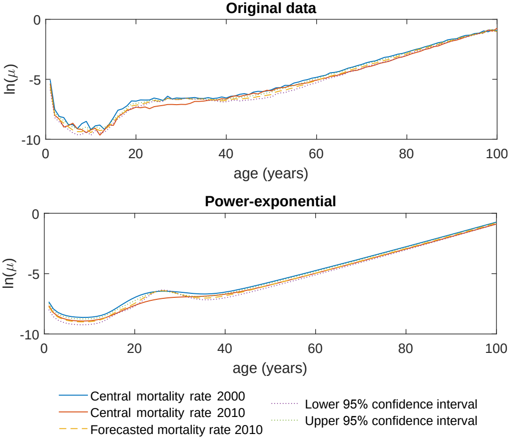


- The Lee–Carter method is based on the assumption that central mortality rates can be fairly accurately approximated by

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t},$$

  where a_x, b_x and k_t are computed from historical central mortality rate.
- Mortality rates are forecasted by assuming that future k_t follow a linear trend.
- Since the Lee–Carter methods uses the logarithm of central mortality rate we can use the previously fitted models to generate corresponding mortality rates and see how this affects the forecast.
- To compare the different forecasts we do two things
  - Estimate the variance of the drift term $\varepsilon_{x,t}$ to compare how well the mortality rates generated by the models match the assumptions of the Lee–Carter model.
  - Estimate the standard error of the forecasted mortality indices to compare how reliable the future forecasts are believed to be.

Model forecast comparison



Estimated variance of ϵ_t	USA	Canada	Switzerland	Japan	Taiwan	Australia
	Country					
Measured data	**0.111**	**0.123**	0.123	0.143	**0.113**	0.0607
Logistic	0.124	0.131	0.140	0.154	0.125	0.0704
Modified Perks	0.122	0.128	0.132	0.149	0.118	0.0695
Power-exponential	0.123	0.130	0.129	0.149	0.141	0.0615
Split power-exp.	0.115	0.125	**0.120**	0.143	0.135	**0.0602**
HP4	0.116	0.134	0.128	**0.142**	0.120	0.0647

Standard error estimate	USA	Canada	Switzerland	Japan	Taiwan	Australia
	Country					
Measured data	0.151	0.199	0.398	0.244	0.299	0.209
Logistic	0.158	0.201	0.345	0.239	0.277	0.238
Modified Perks	0.160	0.204	0.371	0.247	0.294	0.243
Power-exponential	0.157	0.210	0.359	0.235	0.308	0.226
Split power-exp.	0.156	0.209	0.385	0.244	0.297	0.222
HP4	0.152	0.209	0.356	0.245	0.298	0.216

Thank you for your attention!

Questions?