EASTERN WASHINGTON UNIVERSITY
DEPARTMENT OF MATHEMATICS
30 June 2011

To Dr. Ron Dalla, PhD, Vice Provost, Graduate Education Research, 220 Showalter Hall.
From Yves Nievergelt, Professor of Mathematics, 127 Kingston Hall.
About the report on my 2010 Faculty Research and Creative Works Grant.

Faculty Research and Creative Works Grant: Report for 2010, Part I

Ron:

My Faculty Research and Creative Works Grant for the summer of 2010 so far has produced two talks and two manuscripts. The two talks are

1. "Documented Applications of Curves and Surfaces Fitted to Data" Annual Washington State Community College Mathematics Conference, Yakima Convention Center, Yakima, WA, Saturday 21 May 2010.

2. "On Fitting Curves to Data" Spokane Regional Math Colloquium, Gonzaga University, Spokane, WA, Wednesday 4 May 2011.

Two manuscripts have been submitted for publication in refereed professional journals. The attachments (Parts II and III) form an integral part of this report and provide details of the results. In particular, the abstracts, introductions, and case studies have been written to be accessible to the general audience of scientifically curious readers.

Thank you again for this opportunity for conducting this research.

Sincerely,

Yves Nievergelt.
ynievergelt@ewu.edu

Faculty Research and Creative Works Grant: Report for 2010, Part II

Yves Nievergelt

Abstract

A commonly taught scientific method for building mathematical models uses finite computations to approximate the curve of a specified type that best fits the data, without checking whether any such best-fitting curve exists: not every regression objective need have a global unconstrained minimum. One counterexample will confute its theoretical foundation: any triple of points with super-exponential growth does not admit of any unconstrained best-fitting Verhulst logistic curve, regardless of the regression criterion. Moreover, because the set of all such triples is open, there are still no best-fitting Verhulst curves after sufficiently small but arbitrary perturbations of the data. Nevertheless, the present explanations show that through each triple of points with sub-exponential growth passes a unique Verhulst curve. Furthermore, if every triple of data grows sub-exponentially, then for the reciprocal data there exists a median Mitscherlich curve whose reciprocal is a Verhulst curve. Applications range from alchemy to zoology.

0 Introduction

This article reveals and partially fills a gap in the teaching and practice of mathematical modeling in biology. The mathematical folklore occasionally perpetuates a myth that all problems of one kind or another have a solution. For instance, after decades of investigations producing solutions for every specific partial differential equation, Hans Lewy's counterexample without any solution came as a "considerable surprise" [23, 24]. As a second instance, examined in detail here, it has become a common practice to let machines compute the curve that fits the data "best" by minimizing an objective approximately (for instance, with floating-point arithmetic), without investigating the existence of such a "best-fitting" curve. The counterexamples presented here focus on fitting to data a function V of the type

$$V(t) = \frac{K}{1 + e^{a-r\cdot t}}, \tag{1}$$

called a *Verhulst curve* to distinguish it from a logistic curve with $K = 1$. Indeed, the difficulty lies in fitting the *carrying capacity* $K > 0$ [34, p. 58], [35, p. 697].

Example 1 Figure 1 shows Anderson & Anderson's data [3, p. 92, Table 1] on the weight vs. age of cactus wrens (*Campylorhyncus brunneicapillus*), and a Verhulst growth curve fitted graphically by Ricklefs by eyeballing the asymptote K [45].

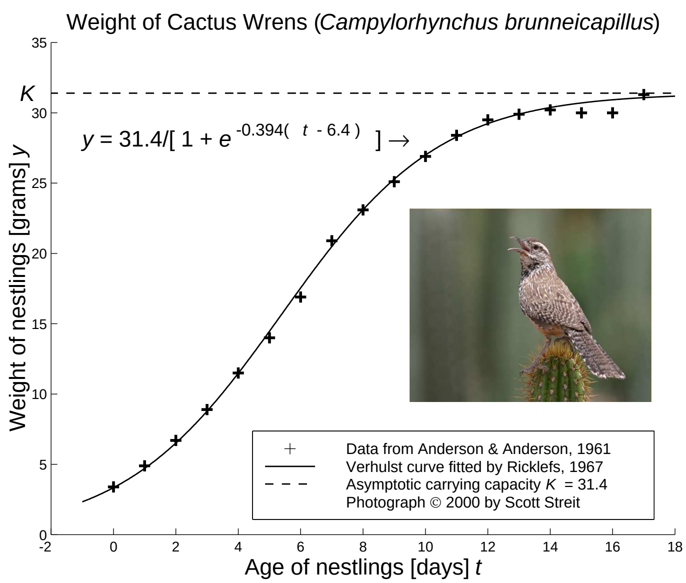

Figure 1: A Verhulst curve (−) fitted graphically by Ricklefs [45] to data (+) by Anderson & Anderson [3]. Photograph of *Campylorhyncus brunneicapillus* © 2000 by Scott Streit (http://www.bird-friends.com), used by permission.

Applications include the identification of parameters of population growth [1], [10], [13], [14], [15], [50], and autocatalysis in chemistry and biology [12, pp. 19–20], [19, 20], [21], [29], [39], [40]. The data and the Verhulst curve can also decrease over time, as in the dependence of animal physiology on temperature; reversing time reverts to increasing data and increasing Verhulst curves [11], [16].

Many attempts to fit Verhulst curves to data have been made. In 1845, Pierre-François Verhulst published an *exact* algebraic formula to find the Verhulst curve that passes through any three data points equally-spaced in time with sub-geometric growth [50, pp. 12–13, 18]. In 1920, without any mention of Verhulst, Raymond Pearl and Lowell J. Reed applied the same formula to the growth of the U.S. population [39]. *Approximate* formulae for a few equally spaced points followed [49, § 3, p. 252], [42, pp. 499–500]. Yet not all data sets are equally spaced. For example, measurements by Alpatov and Pearl occurred at 9:00 in the morning and at 5:00 in the afternoon, separated by 8 and 16 hour intervals [2, p. 41].

Therefore, early investigators merely plotted the data on graph paper and eyeballed ("nach Augenmaß") the carrying capacity, as in Example 1, and then also the line fitted to the data after a logarithmic transformation [14, p. 397], [15, p. 149], [16, § II, pp. 4–5], [43]. An alternative procedure that is popular in the mathematical classroom, though "only for illustration purposes" [1, p. 92], consists of first *guessing* the carrying capacity, and then fitting the initial values and growth rates by least squares or otherwise [28].

In current practice, investigators may have good reasons to impose bounds on the parameters, which guarantee the existence of a minimum for any continuous regression function. For instance, in autocatalysis, the reaction eventually stabilizes, so that the concentrations of reactants remain constant within the measurement accuracy, which effectively yields a measurement of the carrying capacity [40, p. 399, Table 2]. Nevertheless, Examples 28, 30, and 31 in Section 5 of this article show that removing such artificial bounds may yield tighter fits or better predictions.

Algorithms purported to compute the "best" fitting parameters have been published [8], [34, 35, 36], and cited [32], [41], [51]. Such computations without theoretical support have recently been used for many applications [8], [22], including the U.S. population [7, pp. 13–14], [28], and are used without further ado in recent textbooks [10, Example 7.3.2, p. 190]. Yet such results are also consistent with the damning alternative hypothesis that the computers stopped when they ran out of precision, but still nowhere near any "best" fitting curve. Indeed, real experimental data on the growth of *Schizosaccharomyces kefir* [15, p. 143, Table 1, Experiment 1] corroborates this alternative hypothesis. Specifically, the present work reveals real experimental data that do not admit of any unconstrained best-fitting Verhulst curve relative to any objective encompassing ordinary, orthogonal, weighted, correlated, least-squares, or any other power regression. The data are *in general position* in the sense that after any sufficiently small but otherwise arbitrary perturbations, the perturbed data still do not admit of any best-fitting Verhulst curve relative to any such regression. This situation differs from the lack of best-fitting circles for three collinear points: such circles exist after arbitrarily small perturbations of the data.

Remark 2 The Verhulst model with one species is a particular case of the Lotka & Voltera models of competing or prey and predator species [4, § 1.1]. Similarly, the Verhulst differential equation $dy/dt = r \cdot y \cdot [1 - (y/K)]$ is also a particular case of the Bernoulli differential equation $dy/dt = r \cdot y \cdot [1 - (y/K)^{\theta}]$ [10, Example 7.3.2, pp. 190–191]. Consequently, a single counterexample of data without any best-fitting Verhulst growth curve also raises the question whether other data sets may lack any best-fitting Bernoulli or Lotka & Voltera model or yet other models.

However, the implications of this single counterexample are far broader than merely about Verhulst curves: in the absence of any theorem guaranteeing the existence of a best-fitting model, the practice of fitting a mathematical model to data has no theoretical foundations. Consequently, logical inferences based on minimizing the objective fail. For example, maximum likelihood methods cannot be justified, because "the maximizer of the likelihood function is the vector of parameter estimates" [18, p. 1252]. Indeed, maximum likelihood methods depend on the distribution of the parameters of a uniquely identified minimizing curve, for example, the slope and intercept for the ordinary least-squares line [9, § 37.2].

Remark 3 All the regressions just mentioned share the following common features that will preclude the existence of any best-fitting Verhulst curve, as shown in the appendix (Section 8). Experimenters observe and measure a sequence of distinct *data* points $D = (z_1, \ldots, z_N)$, which are *given* (to the next scientists). The next scientists, who need not be distinct from the experimenters, specify a class $\mathcal{C}$ of curves, such as the class of all straight lines, or the class of all Verhulst curves. For each curve $C \in \mathcal{C}$, they identify a sequence of *adjusted* points $\tilde{D} = (\tilde{z}_1, \ldots, \tilde{z}_N)$ on C. For orthogonal regression, the adjusted point $\tilde{z}_j$ is a point on C closest to the data point z_j. For ordinary regression of y vs. t, the adjusted point $\tilde{z}_j$ is a point on C directly above or below the data point z_j, whereas for t vs. y, the adjusted point $\tilde{z}_j$ is a point on C directly to the right or left of the data point z_j. In all such regressions the adjusted point $\tilde{z}_j$ is a point on C but constrained to lie in an affine subspace U_{z_j} containing the data point z_j. The scientists also choose an objective function $F_D : \tilde{D} \mapsto F_D(\tilde{D}) \in \mathbb{R}_+$. For least-squares regression, the objective F_D is the sum of the squared distances, whereas for *median* regression F_D is the sum of the distances; in either case the distance between $\tilde{z}_j$ and z_j is defined by some distance d_j on the affine subspace U_{z_j}. The scientists' goal consists of finding a curve $C \in \mathcal{C}$ on which the adjusted points $\tilde{D}$ minimize the objective F_D. Because at this stage the scientists know neither the curve C nor the adjusted points $\tilde{D}$ on it, the domain of the objective F_D is the entire Cartesian product $\mathbb{U} = U_{z_1} \times \cdots \times U_{z_N}$. Moreover, the objective F_D is a topologically open map and continuous function of $\tilde{D}$ such that $F_D(\tilde{D}) = 0$ if and only if $D = \tilde{D}$. The adjusted points need not be distinct or unique, but if $\tilde{D}' = (\tilde{z}'_1, \ldots, \tilde{z}'_N)$ are other adjusted points on the same curve C for the same data D, then $F_D(\tilde{D}) = F_D(\tilde{D}')$, because

$$F_D(\tilde{D}) = \min\{F_D(p_1, \ldots, p_N) : p_1, \ldots, p_N \in C\}$$

for all the regressions considered here, which allows for the definition

$$\mathcal{F}_D(C) := F_D(\tilde{D}).$$

For each curve C, the map $D \mapsto \mathcal{F}_D(C)$ is a continuous function of the data D. However, to keep the proofs to manageable lengths, for Verhulst curves with *increasing* data sequences, the discussion is restricted to such regression methods for which the adjusted points are also distinct. Proposition 38 in Section 8 shows that a variety of ordinary and orthogonal regression methods share this feature. Also, for each data sequence D, the map $C \mapsto \mathcal{F}_D(C)$ is a continuous function of the parameters specifying the curve C, such as the slope and intercept for lines, or a, K, and r for Verhulst curves. *The issue is that this map may have no unconstrained minimum over the whole space of parameters.*

To establish the absence of best-fitting Verhulst curves for any such regression, Section 1 reviews the concepts of sub-linear and sub-exponential growth. Theorem 16 shows that every triple of points on a Verhulst curve grows sub-exponentially. Conversely, Theorem 17 shows that through each sub-exponential triple of data points passes a unique Verhulst curve. Theorem 18 in Section 2 then proves the absence of any best-fitting Verhulst curve relative to any regression method for triples of data points with exponential or super-exponential growth.

Nevertheless, subsequent sections also show partially what might be needed to fill the gap just described. All the data ordinates considered here are positive and remain below the carrying capacity, so that $0 < y < K$. Therefore, curves may be fitted to the data $z_k = (t_k, y_k)$, or to the transformed data $\zeta_k = (t_k, \eta_k)$, after Verhulst's transformation $\eta = \ln[y/(K-y)]$. Gause [15, pp. 149–150] and Gause & Alpatov [16] always pick some value $K > y_{\max} := \max\{y_1, \ldots, y_N\} > 0$ for their demonstrations. Yet in several experiments Gause selects a value $K < y_{\max}$ [15, pp. 69, 72, 77, 79, 85, 87, 94, 101, 102, 104] without indicating how to handle the undefined transformation $\ln[y_k/(K-y_k)]$. Also, changing K changes the metric in the transformed plane, so that the values of different objectives for different values of K are not readily comparable. For such reasons the present considerations proceed with the reciprocal transformation

$$q := \frac{1}{y}, \tag{2}$$

as done by other authors [17, p. 39], [44, p. 384]. The goal then consists of fitting to the reciprocal data $\hat{z}_j = (t_j, q_j)$ the reciprocal of a Verhulst curve, which, after the change of parameters $B := 1/K$ and $A := e^a/K$, is a *Mitscherlich* curve $M(t) := A \cdot e^{-r \cdot t} + B$. To this end, Section 3 reviews the concepts of median points and lines. In Section 4, Theorem 26 shows that if every triple of reciprocal data points decreases super-exponentially, then there exists a median Mitscherlich curve, which degenerates into a median line as B diverges to $\pm\infty$ while r tends to 0. Theorem 27 shows that if every triple of data points grows sub-exponentially, then there exists a *median reciprocal Verhulst curve*, minimizing the sum of the absolute differences between the reciprocals of the data and the reciprocal of the fitted curve. Section 5 applies the theory to real examples with real data.

1 Verhulst Curves through Two or Three Data Points

For each data set in general position, some median line passes through two data points, as reviewed in Section 3. Also, some median circle passes through two or three data points [33]. Similary, toward a study of median Mitscherlich and Verhulst curves, this section defines criteria that will determine whether three points lie on a Mitscherlich or Verhulst curve. Such criteria will compare the third point with the line and exponential curve through the first two points.

1.1 Exponential Curves through Two Points

This subsection establishes terminology for the position of three points relative to an exponential curve.

Definition 4 For all points $\zeta_1 := (t_1, \eta_1)$ and $\zeta_2 := (t_2, \eta_2)$ with $t_1 \neq t_2$, denote by $\text{line}_{\zeta_1,\zeta_2}$ the affine function passing through ζ_1 and ζ_2. Three points ζ_1, ζ_2, and $\zeta_3 := (t_3, \eta_3)$ with $t_1 < t_2 < t_3$ are, respectively, *sub-linear*, *linear*, or *super-linear*, if $\eta_3 < \text{line}_{\zeta_1,\zeta_2}(t_3)$, or $\eta_3 = \text{line}_{\zeta_1,\zeta_2}(t_3)$, or $\eta_3 > \text{line}_{\zeta_1,\zeta_2}(t_3)$.

Lemma 5 *The sets* $\Xi \subset (\mathbb{R}^2)^3$ *and* $\Upsilon \subset (\mathbb{R}^2)^3$ *of respectively all sub-linear and all super-linear triples* $(\zeta_1, \zeta_2, \zeta_3)$ *with* $t_1 < t_2 < t_3$ *are open in* $(\mathbb{R}^2)^3$.

Also, the following conditions for super-linear triples are mutually equivalent:

$$[\eta_3 > \text{line}_{\zeta_1,\zeta_2}(t_3)] \Leftrightarrow [\eta_1 > \text{line}_{\zeta_2,\zeta_3}(t_1)] \Leftrightarrow [\eta_2 < \text{line}_{\zeta_1,\zeta_3}(t_2)] \,.$$

Moreover, the following conditions for sub-linear triples are mutually equivalent:

$$[\eta_3 < \text{line}_{\zeta_1,\zeta_2}(t_3)] \Leftrightarrow [\eta_1 < \text{line}_{\zeta_2,\zeta_3}(t_1)] \Leftrightarrow [\eta_2 > \text{line}_{\zeta_1,\zeta_3}(t_2)] \,.$$

Proof. The triple $(\zeta_1, \zeta_2, \zeta_3)$ is super-linear if and only if $\det(\zeta_2 - \zeta_1, \zeta_3 - \zeta_1) > 0$. □

Remark 6 Similar considerations apply to exponential curves. Here, a function $g : \mathbb{R} \to \mathbb{R}$ is called *exponential* if there exist scalars $c, s \in \mathbb{R}$ such that $g(t) = e^{c+s\cdot t}$ for every $t \in \mathbb{R}$. To fit exponential curves to data, *logarithmic-linear* regression transforms the open upper-half plane $\mathbb{H}^*_+ := \mathbb{R} \times \mathbb{R}^*_+$, where $\mathbb{R}^*_+ :=]0, \infty[:=$

$\{x \in \mathbb{R} : 0 < x\}$, into the plane $\mathbb{R}^2$ by the map $L : \mathbb{H}_+^* \to \mathbb{R}^2$ and its inverse $E : \mathbb{R}^2 \to \mathbb{H}_+^*$ defined by

$$L(t, y) := (t, \ln y),$$
$$E(t, \eta) := (t, e^{\eta}).$$

The regression transforms each data point $z_j := (t_j, y_j)$ into $\zeta_j := L(z_j) = (t_j, \ln y_j) =: (t_j, \eta_j)$, fits a straight line with equation $\eta = c + s \cdot t$ to the transformed points $\zeta_1, \ldots, \zeta_N$ by any method, and transforms the line back to the exponential curve with equation $y = e^{c+s\cdot t}$. In particular, for all points $z_1 := (t_1, y_1)$ and $z_2 := (t_2, y_2)$ with $t_1 \neq t_2$ in $\mathbb{H}_+^*$, there exists exactly one exponential curve passing through z_1 and z_2: the image by E of the line through ζ_1 and ζ_2, with

$$s = \frac{\ln(y_2) - \ln(y_1)}{t_2 - t_1}, \tag{3}$$
$$c = \ln(y_1) - s \cdot t_1. \tag{4}$$

Also, L and E preserve order, so that $\eta_1 < \eta_2$ if and only if $e^{\eta_1} < e^{\eta_2}$.

Definition 7 For all $t_1 \neq t_2$, with $y_1 > 0$ and $y_2 > 0$, denote by $\exp_{z_1,z_2}$ the exponential function $t \mapsto e^{c+s\cdot t}$ passing through $z_1 := (t_1, y_1)$ and $z_2 := (t_2, y_2)$. Three points z_1, z_2, and $z_3 := (t_3, y_3)$ with $t_1 < t_2 < t_3$ and $0 < y_1 < y_2 < y_3$ are, respectively, *sub-exponential*, *exponential*, or *super-exponential*, if $y_3 < \exp_{z_1,z_2}(t_3)$, or $y_3 = \exp_{z_1,z_2}(t_3)$, or $y_3 > \exp_{z_1,z_2}(t_3)$.

Lemma 8 *The sets $X \subset (\mathbb{R}^2)^3$ and $Y \subset (\mathbb{R}^2)^3$ of respectively all sub-exponential and all super-exponential triples (z_1, z_2, z_3) are open in $(\mathbb{R}^2)^3$.*

The following conditions for super-exponential triples are mutually equivalent:

$$\left[y_3 > \exp_{z_1,z_2}(t_3)\right] \Leftrightarrow \left[y_1 > \exp_{z_2,z_3}(t_1)\right] \Leftrightarrow \left[y_2 < \exp_{z_1,z_3}(t_2)\right].$$

The following conditions for sub-exponential triples are mutually equivalent:

$$\left[y_3 < \exp_{z_1,z_2}(t_3)\right] \Leftrightarrow \left[y_1 < \exp_{z_2,z_3}(t_1)\right] \Leftrightarrow \left[y_2 < \exp_{z_1,z_3}(t_2)\right].$$

Proof. For each $j \in \{1, 2, 3\}$, let $z_j = (t_j, y_j)$, and $\zeta_j := L(z_j) = (t_j, \ln y_j)$. Thus, (z_1, z_2, z_3) is super-exponential if and only if $(\zeta_1, \zeta_2, \zeta_3)$ is super-linear. The conclusions then follow from Lemma 5 by the continuity of $\det$ and L. □

1.2 Mitscherlich and Verhulst Curves through Three Points

This subsection describes the relative positions of $\text{line}_{\hat{z}_1,\hat{z}_2}$, $\exp_{\hat{z}_1,\hat{z}_2}$, and Mitscherlich curves through the same distinct points $\hat{z}_1 = (t_1, q_1)$ and $\hat{z}_2 = (t_2, q_2)$. Thence will follow the uniqueness of a Mitscherlich curve through three points.

Definition 9 For all $A, B, r \in \mathbb{R}$, a *Mitscherlich law* [38], [42], [49] has the form

$$M(t) := A \cdot e^{-r \cdot t} + B. \tag{5}$$

One way to deal with nonlinear regressions investigates whether fixing one parameter (B) leads to a linear regression with the other parameters (A and r).

Lemma 10 *For each constant $B \in \mathbb{R}$ and for all points $\hat{z}_1 = (t_1, q_1)$ and $\hat{z}_2 = (t_2, q_2)$ such that $t_1 < t_2$ but $q_1 > q_2$ with $B \notin [q_2, q_1]$, there exists a unique Mitscherlich curve (5) passing through $\hat{z}_1$ and $\hat{z}_2$ with constant B.*

If $q_1 > q_2 > B$, then $A, r > 0$, whereas if $B > q_1 > q_2$, then $A, r < 0$.

If $q_1 \geq B \geq q_2$, then there is no such Mitscherlich curve through $\hat{z}_1$ and $\hat{z}_2$.

Proof. If $q_1 > q_2 > B$, then the shift $q' := q - B$ transforms any Mitscherlich curve through $\hat{z}_1$ and $\hat{z}_2$ with constant B into the unique exponential curve $\exp_{z'_1,z'_2}$ through (t_1, q'_1) and (t_2, q'_2). Hence formulae (3) and (4) show that $A, r > 0$ with

$$r = r(B; \hat{z}_1, \hat{z}_1) = -s = \frac{1}{t_2 - t_1} \cdot \ln\left(\frac{q_1 - B}{q_2 - B}\right), \tag{6}$$

$$\ln(A) = \ln[A(B; \hat{z}_1, \hat{z}_1)] = c = \ln|q_1 - B| + r \cdot t_1 \in \mathbb{R}. \tag{7}$$

If $B > q_1 > q_2$, then the shift $q^{\dagger} := B - q$ gives $r < 0$ also defined by formula (6), and $\ln(-A) \in \mathbb{R}$, also defined by formula (7).

If $q_1 > B > q_2$, then $A \cdot e^{-r \cdot t_2} = q_2 - B < 0 < q_1 - B = A \cdot e^{-r \cdot t_2}$ would imply $A < 0 < A$. If $B \in \{q_1, q_2\}$, then $A = 0$, which would imply $q_1 = B = q_2$. □

Definition 11 Let M_B denote any Mitscherlich curve with constant B. For all points $\hat{z}_1 = (t_1, q_1)$ and $\hat{z}_2 = (t_2, q_2)$ such that $t_1 < t_2$ but $q_1 > q_2$ with $B \notin [q_2, q_1]$, let $M_{B;\hat{z}_1,\hat{z}_2}$ be the Mitscherlich curve through $\hat{z}_1$ and $\hat{z}_2$ with constant B.

If also $0 < B < q_2$, then let $V_{K;z_1,z_2} := 1/\, M_{B;\hat{z}_1,\hat{z}_2}$ be the Verhulst curve through $z_1 = (t_1, y_1)$ and $z_2 = (t_2, y_2)$ with carrying capacity $K := 1/B$.

One way to deal with nonlinear curves uses the convexity of their logarithm, as in the proof of Lemma 12, which identifies the relative order of two Mitscherlich curves intersecting each other at two common points.

Lemma 12 *Suppose that $\hat{z}_1 = (t_1, q_1)$ and $\hat{z}_2 = (t_2, q_2)$, where $t_1 < t_2$ but $q_1 > q_2$, and $q_2 > B > C$. Then $M_{B;\hat{z}_1,\hat{z}_2}(t) < M_{C;\hat{z}_1,\hat{z}_2}(t)$ for every t such that $t_1 < t < t_2$, while $M_{B;\hat{z}_1,\hat{z}_2}(t) > M_{C;\hat{z}_1,\hat{z}_2}(t)$ for every $t \notin [t_1, t_2]$.*

Proof. For each constant $B \in \mathbb{R}$, differentiating the definition (5) shows that every Mitschlerlich curve M_B with constant B satisfies the differential equation

$$M'_B(t) = -r \cdot [M_B(t) - B] = -r \cdot A \cdot e^{-r \cdot t}. \tag{8}$$

For the purposes of a convex analysis, define the transformation

$$W_B(t) := \ln |M_B(t) - C|. \tag{9}$$

Substituting formula (8) into the derivatives of the transformation (9) gives

$$W'_B = \frac{M'_B}{M_B - C} = r \cdot \frac{B - M_B}{M_B - C} = r \cdot \left(\frac{B - C}{M_B - C} - 1 \right), \tag{10}$$

$$W''_B = r \cdot (B - C) \cdot \frac{-M'_B}{(M_B - C)^2} = r^2 \cdot (B - C) \cdot \frac{M_B - B}{(M_B - C)^2}. \tag{11}$$

Lemma 10 shows that if $t_1 < t_2$ but $q_1 > q_2$ with $B \notin [q_2, q_1]$, then $A \cdot r > 0$, whence $M'_B < 0$ by equation (8), and hence W''_B has the same sign as $r \cdot (B - C)$.

If $B = C$, then $W''_C = 0$ by equation (11), so that W_C is a straight line.

If $q_2 > B$, then $A > 0$ and $r > 0$ by Lemma 10; equations (8) and (11) then show that if $q_2 > B > C$, then $W''_B > 0$.

Thus, if $q_2 > B > C$, then $W_{C;\check{z}_1,\check{z}_2}$ is a line while $W_{B;\check{z}_1,\check{z}_2}$ is convex, and both pass through $\check{z}_1$ and $\check{z}_2$. Consequently, $W_{B;\check{z}_1,\check{z}_2}(t) < W_{C;\check{z}_1,\check{z}_2}(t)$ for $t_1 < t < t_2$, whereas $W_{B;\check{z}_1,\check{z}_2}(t) > W_{C;\check{z}_1,\check{z}_2}(t)$ for $t \notin [t_1, t_2]$.

Finally, from $A > 0$ and $B > C$ follow $M_B > B > C$ and $M_C > C$, so that the transformation (9) preserves order, which completes the proof. □

"Degenerate" cases will help establish the existence or lack of best-fitting curves.

Lemma 13 *As B diverges to $\pm\infty$, the Mitscherlich curve $M_{B;\hat{z}_1,\hat{z}_2}$ tends to $\mathrm{line}_{\hat{z}_1,\hat{z}_2}$ uniformly on each compact subset of the real line.*

Also, $\lim_{B \nearrow q_2^-} M_{B;\hat{z}_1,\hat{z}_2}(t) = q_2$ for $t > t_1$, but diverges to ∞ for $t < t_1$. Moreover, if $0 \notin [q_2, q_1]$, then $\lim_{B \to 0} M_{B;\hat{z}_1,\hat{z}_2}(t) = \exp_{\hat{z}_1,\hat{z}_2}(t)$, and $-s = \lim_{B \to 0} r(B; \hat{z}_1, \hat{z}_1) = \ln(q_1/q_2)/(t_2 - t_1)$, where s is the growth rate of $\exp_{\hat{z}_1,\hat{z}_2}(t) = e^{c+s \cdot t}$, as in formula (3). If $0 < q_2 < q_1$, then $\lim_{K \to \infty} V_{K;z_1,z_2}(t) = \exp_{z_1,z_2}(t)$.

Proof. For $B \notin [q_2, q_1]$, substituting equation (7) for A yields the formulae

$$M_{B;\hat{z}_1,\hat{z}_2}(t) = (q_1 - B) \cdot e^{-r \cdot (t - t_1)} + B \tag{12}$$

$$= q_1 \cdot e^{-r \cdot (t - t_1)} - B \cdot r \cdot \frac{e^{-r \cdot (t - t_1)} - 1}{r}. \tag{13}$$

The term $q_1 \cdot e^{-r\cdot(t-t_1)}$ tends to q_1, because formula (6) for r leads to the limit

$$\lim_{B\to\pm\infty} r(B;\hat{z}_1,\hat{z}_1) = \lim_{B\to\pm\infty} \frac{1}{t_2-t_1}\cdot\ln\left(\frac{q_1-B}{q_2-B}\right) = 0. \tag{14}$$

With the abbreviation $r(B) := r(B;\hat{z}_1,\hat{z}_1)$, l'Hospital's Rule gives

$$\begin{aligned}\lim_{B\to\pm\infty} B\cdot r(B) &= \lim_{B\to\pm\infty}\frac{r(B)}{1/B} = \lim_{B\to\pm\infty}\frac{r'(B)}{-1/B^2}\\ &= \lim_{B\to\pm\infty}\frac{-B^2}{t_2-t_1}\cdot\frac{q_1-q_2}{(q_1-B)\cdot(q_2-B)}\\ &= \frac{q_2-q_1}{t_2-t_1}.\end{aligned} \tag{15}$$

Substituting the limits (14) and (15) into formula (13) yields the convergence

$$\begin{aligned}\lim_{B\to\pm\infty} M_{B;\hat{z}_1,\hat{z}_2}(t) &= q_1 - \frac{q_2-q_1}{t_2-t_1}\cdot\frac{\partial}{\partial r}e^{-r\cdot(t-t_1)}\bigg|_{r=0}\\ &= q_1 + \frac{q_2-q_1}{t_2-t_1}\cdot(t-t_1) = \mathrm{line}_{\hat{z}_1,\hat{z}_2}(t).\end{aligned}$$

Also, $M_{B;\hat{z}_1,\hat{z}_2}(t_1) = q_1$, while $\lim_{B\nearrow q_2^-} r(B;\hat{z}_1,\hat{z}_1) = \infty$ by formula (6), whence $\lim_{B\nearrow q_2^-} M_{B;\hat{z}_1,\hat{z}_2}(t) = q_2$ for $t > t_1$ and diverges to ∞ for $t < t_1$ by formula (12).

Formulae (6), (7), and (13) confirm that $\lim_{B\to 0} M_{B;\hat{z}_1,\hat{z}_2}(t) = \exp_{\hat{z}_1,\hat{z}_2}(t)$. Formula (6) also shows that $\lim_{B\to 0} r(B;\hat{z}_1,\hat{z}_1) = \ln(q_1/q_2)/(t_2-t_1) = -s$. Hence if $0 < q_2 < q_1$, then reciprocals yield $\lim_{K\to\infty} V_{K;z_1,z_2}(t) = \exp_{z_1,z_2}(t)$.

In all cases the convergence is monotonic by Lemma 12 and hence uniform on compacta by Dini's theorem [46, p. 162]. $\square$

Theorem 14 *For all points $\hat{z}_j = (t_j, q_j)$ with $t_1 < t_2 < t_3$ and $q_1 > q_2 > q_3 > \mathrm{line}_{\hat{z}_1,\hat{z}_2}(t_3)$ there is a unique Mitscherlich curve through $\hat{z}_1$, $\hat{z}_2$, $\hat{z}_3$ with $B < q_2$; also, $A, r > 0$ and $B < q_3$. Moreover, the triple $(\hat{z}_1,\hat{z}_2,\hat{z}_3)$ is sub-exponential, exponential, or super-exponential according to $B < 0$, $B = 0$, or $B > 0$.*

Proof. By Lemma 10, for each constant $B \in \mathbb{R}$ and for all points $\hat{z}_1 = (t_1, q_1)$ and $\hat{z}_2 = (t_2, q_2)$ such that $t_1 < t_2$ but $q_1 > q_2$ with $B \notin [q_2, q_1]$, there exists a unique Mitscherlich curve $M_{B;\hat{z}_1,\hat{z}_2}$ through $\hat{z}_1$ and $\hat{z}_2$ with constant B.

For $B < q_2$, by Lemma 13, $\lim_{B\searrow-\infty} M_{B;\hat{z}_1,\hat{z}_2}(t_3) = \mathrm{line}_{\hat{z}_1,\hat{z}_2}(t_3) < q_3$ and $\lim_{B\nearrow q_2^-} M_{B;\hat{z}_1,\hat{z}_2}(t_3) = q_2 > q_3$. By the Intermediate-Value Theorem, there exists $B < q_2$ such that $M_{B;\hat{z}_1,\hat{z}_2}(t_3) = q_3$. This value of B is unique by Lemma 12.

Also, from $B < q_2$ follows $A > 0$ and $r > 0$ by Lemma 10. Hence $q_3 = A \cdot e^{-r \cdot t_3} + B > B$. By uniqueness, the triple $(\hat{z}_1, \hat{z}_2, \hat{z}_3)$ is exponential if and only if $B = 0$. If $B < q_2$, then Lemma 12 shows that $M_{B;\hat{z}_1,\hat{z}_2}(t_3)$ increases as B increases. Hence the triple $(\hat{z}_1, \hat{z}_2, \hat{z}_3)$ is respectively sub-exponential or super-exponential according to $B < 0$ or $B > 0$. □

Definition 15 For all points $\hat{z}_j = (t_j, q_j)$ with $t_1 < t_2 < t_3$ and $q_1 > q_2 > q_3 > \text{line}_{\hat{z}_1,\hat{z}_2}(t_3)$ let $M_{\hat{z}_1;\hat{z}_2,\hat{z}_3}$ be the Mitscherlich curve through $\hat{z}_1$, $\hat{z}_2$, $\hat{z}_3$ with $B < q_2$.

Theorem 16 *Every triple of points on a Verhulst curve is sub-exponential.*

Proof. If $z_1 = (t_1, y_1)$, $z_2 = (t_1, y_1)$, $z_3 = (t_1, y_1)$ are on a Verhulst curve $V(t) = K/(e^{a-r \cdot t} + 1)$, then the transformed points $\hat{z}_j = (t_j, q_j)$ with $q_j = 1/y_j$ for $j \in \{1, 2, 3\}$ are on the Mitscherlich curve $M(t) = A \cdot e^{-r \cdot t} + B$ with $B = 1/K > 0$ and $A = e^a/K > 0$. By Theorem 14, the triple $(\hat{z}_1, \hat{z}_2, \hat{z}_3)$ is then super-exponential, so that the triple (z_1, z_2, z_3) is sub-exponential. □

Theorem 17 *Through each sub-exponential triple passes exactly one Verhulst curve.*

Proof. If the triple (z_1, z_2, z_3) is sub-exponential, then the triple $(\hat{z}_1, \hat{z}_2, \hat{z}_3)$ is super-exponential. By Theorem 14 there is a unique Mitscherlich curve through $\hat{z}_1$, $\hat{z}_2$, $\hat{z}_3$ with $0 < B < q_2$ and hence $A > 0$. Its reciprocal is the unique Verhulst curve through z_1, z_2, z_3, which may be denoted by $V_{z_1;z_2,z_3}$. □

2 Generic Data without Best-Fitting Verhulst Curves

Theorem 18 shows that for each data set D in the open subspace Y of super-exponential triples of points, there is no best-fitting Verhulst curve for any of the regression methods described in Remark 3.

Theorem 18 *For each increasing exponential or super-exponential triples of points D, for each objective F_D described in Remark 3, and for each Verhulst curve V, there exists another Verhulst curve S such that $\mathcal{F}_D(S) < \mathcal{F}_D(V)$.*

Proof. By the hypotheses on the regression methods in Remark 3, for each Verhulst curve V, and each *increasing* exponential or super-exponential triple of points $D := (z_1, z_2, z_3)$, there is a triple of *distinct* adjusted points $\tilde{D} := (\tilde{z}_1, \tilde{z}_2, \tilde{z}_3)$ on V. By Theorem 16, the three adjusted points $\tilde{D}$ on V are sub-exponential, and the three exponential or super-exponential points D do not all lie on V. Hence there is at least one non-zero residual $z_j - \tilde{z}_j \neq 0$, whence $\mathcal{F}_D(V) = F_D(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3) > 0$ by the hypothesis on F_D that $F_D(\tilde{D}) = 0$ if and only if $\tilde{D} = D$.

By Lemma 8, the set X of sub-exponential triples is open, so $X \cap \mathbb{U}$ is relatively open, and $F_D : X \to \mathbb{R}_+$ is relatively open by hypothesis. Hence F_D maps $X \cap \mathbb{U}$ onto an open neighborhood $U \subseteq \mathbb{R}_+^* =]0, \infty[$ of $F_D(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3) > 0$. In particular, there exists $w \in U$ such that $0 < w < F_D(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3)$. Consequently, there exists $(w_1, w_2, w_3) \in X \cap \mathbb{U}$ such that $0 < w = F_D(w_1, w_2, w_3) < F_D(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3)$.

By Theorem 17, there is a Verhulst curve S passing through all three points (w_1, w_2, w_3). Also, $\mathcal{F}_D(S) = \min\{F_D(p_1, p_2, p_3) : (\forall j)(p_j \in S)\}$ by Remark 3. Therefore, $0 \leq \mathcal{F}_D(S) \leq F_D(w_1, w_2, w_3) < F_D(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3) = \mathcal{F}_D(V)$. $\square$

If K is constrained below or at a maximum M, then for each exponential triple D the convergence of Verhulst curves to exponentials in Lemma 13 shows that F_D has its minimum at $K = M$ for the "median" regression in Section 4.

3 Median Points and Lines

This section reviews the concept of medians without using linear programming, showing that at least one median line passes through at least two data points.

Definition 19 A real number $r_\diamond$ is a *median* of real numbers $r_1, \ldots, r_N$, which are also called *data points*, if $r_\diamond$ minimizes the objective

$$f_\diamond(r) := \sum_{j=1}^{N} |r - r_j|, \tag{16}$$

so that $f_\diamond(r_\diamond) = \min_{r \in \mathbb{R}} f_\diamond(r)$. Moreover, for each r, define the numbers of data points smaller than (N^-), equal to (N^0), or larger than (N^+) the number r:

$$N^-(r) := \sum_{r_k < r} 1, \qquad N^0(r) := \sum_{r_j = r} 1, \qquad N^+(r) := \sum_{r_\ell > r} 1.$$

Theorem 20 *For all real numbers $r_1 \leq \cdots \leq r_N$ there is at least one median. Moreover, at least one median coincides with a data point.*

Proof. This proof shows that for each $r \in \mathbb{R}$ there exists a data point r_n such that $f_\diamond(r_n) \leq f_\diamond(r)$. If $r \notin \{r_1, \ldots, r_N\}$, then $f_\diamond$ is differentiable at r, so that

$$\begin{aligned} f_\diamond(r) &= \sum_{r_k < r} (r - r_k) + \sum_{r_\ell > r} (r_\ell - r), \\ f'_\diamond(r) &= \sum_{r_k < r} 1 - \sum_{r_\ell > r} 1 = N^-(r) - N^+(r). \end{aligned}$$

If $f'_\diamond(r) > 0$, then $N^-(r) \geq N^-(r) - N^+(r) = f'_\diamond(r) > 0$, whence there exists at least one data point to the left of r; in this case, let $n := \max\{j \, : \, r_j < r\}$. Then $f'_\diamond(u) = f'_\diamond(r) > 0$ for every u such that $r_n < u \leq r$, and hence $f_\diamond(r_n) < f_\diamond(r)$.

Similarly, if $f'_\diamond(r) < 0$, then there is some r_n with $r < r_n$ and $f_\diamond(r) > f_\diamond(r_n)$.

If $f'_\diamond(r) = 0$, then $N^-(r) = N^+(r) = N/2 > 0$, whence there are two data points such that $r_n < r < r_{n+1}$; in that case $f'_\diamond(u) = 0$ for every u such that $r_n < u < r_{n+1}$, so that $f_\diamond(r_n) = f_\diamond(r)$.

Thus, if $r \notin \{r_1, \ldots, r_N\}$, then there is a data point r_n where $f_\diamond(r_n) \leq f_\diamond(r)$. The minimum of $f_\diamond$ on the finite set $\{r_1, \ldots, r_N\}$ is then a global minimum. □

Definition 21 A line with equation $y = A_\diamond \cdot x + B_\diamond$ is a *median* line for data points $z_1 = (x_1, y_1), \ldots, z_N = (x_N, y_N)$ if $(A_\diamond, B_\diamond)$ minimizes the objective

$$f_\diamond(A, B) := \sum_{j=1}^{N} |A \cdot x_j + B - y_j|. \tag{17}$$

Thus $f_\diamond(A_\diamond, B_\diamond) = \min_{(A,B) \in \mathbb{R}^2} f_\diamond(A, B)$.

Theorem 22 *For all data $z_1, \ldots, z_N$ such that $x_1 < \cdots < x_N$ with $N \geq 2$, there exists at least one median line that passes through at least two distinct data points.*

Proof. For each slope $A \in \mathbb{R}$, by Theorem 20 there is an index m such that $B_m := y_m - A \cdot x_m$ is a median of the numbers $B_1 := y_1 - A \cdot x_1$, ..., $B_N := y_N - A \cdot x_N$. Shifting a line with slope A vertically to the intercept B_m shows that $f_\diamond(A, B_m) \leq f_\diamond(A, B)$ for every intercept $B \in \mathbb{R}$. Because $B_m = y_m - A \cdot x_m$, the line with slope A and intercept B_m passes through the data point $z_m = (x_m, y_m)$. Thus it suffices to search for median lines passing through at least one data point.

If there is another index $n \neq m$ such that $B_n = B_m$, then the line already passes through at least two distinct data points z_m and z_n. If there is no such index $n \neq m$ with $B_n = B_m$, then there are indices i and n such that $B_i < B_m < B_n$, or, equivalently, $(y_i - y_m)/(x_i - x_m) < A < (y_n - y_m)/(x_n - x_m)$.

Shifting the origin to z_m changes the objective (17) to

$$\begin{aligned} f_\diamond(A) = & \sum_{y_k - y_m < A \cdot (x_k - x_m)} [A \cdot (x_k - x_m) - (y_k - y_m)] \\ & + \sum_{y_\ell - y_m > A \cdot (x_\ell - x_m)} [(y_\ell - y_m) - A \cdot (x_\ell - x_m)]. \end{aligned}$$

For each $j \neq m$ define the slope $A_j := (y_j - y_m)/(x_j - x_m)$. If the line with slope A through z_m passes through no other data point, then $f'_\diamond$ is piecewise constant:

$$f'_\diamond(A) = \sum_{A_k < A} (x_k - x_m) - \sum_{A_\ell > A} (x_\ell - x_m).$$

If $f'_\diamond(A) > 0$, then decreasing A rotates the line clockwise and decreases the objective $f_\diamond(A)$ until the line touches another data point z_i, where $f_\diamond(A_i) < f_\diamond(A)$. A similar argument applies to $f'_\diamond(A) < 0$ by rotating the line counterclockwise until it touches z_n. If $f'_\diamond(A) = 0$, then rotate the line in either direction, to z_i or z_n.

Therefore, the minimum of $f_\diamond$ on the finite set of lines through any two data points is a global minimum. □

Theorem 22 generalizes to data $x_1, \ldots, x_N$ with at least two distinct abscissas $x_m \neq x_n$. Because there are at most $N \cdot (N-1)/2$ lines through at least two data points, for small values of N, it suffices to compute the sum of the distances $f_\diamond(A_{m,n}, B_{m,n})$ for each line through two distinct data points $z_m \neq z_n$ and pick for a median line any such line with the smallest sum of the distances [31, p. 4]. For larger values of N faster linear programming algorithms exist for median points and lines [5, p. 454], [6], [47], [52], and Korneenko and Martini's *Anchored Median Hyperplane Algorithm* for orthogonal median regressions [25], [26], [27]. The foregoing discussion also holds for weighted medians, with the objective (17) replaced by $f_\diamond(A, B) := \sum_{j=1}^{N} w_j \cdot |A \cdot x_j + B - y_j|$. Each positive weight w_j can be adjusted to reflect the accuracy of the j-th measurement.

4 Conditions for Median Mitscherlich and Verhulst Curves

With only positive data ordinates (y_j), a non-positive carrying capacity $K \leq 0$ never gives a best-fitting Verhulst curve, because raising K to $y_{\min} := \min\{y_1, \ldots, y_N\} > 0$ decreases all vertical distances to all the data points. Consequently, the search for a median Verhulst curve can focus on positive carrying capacities $K > 0$. After the reciprocal transformation (2) applied to the data ordinates and the Verhulst curve, a *median reciprocal Verhulst* curve has parameters K, a, and r minimizing

$$f_\diamond(K, a, r) := \sum_{j=1}^{N} \left| \frac{1 + e^{a - r \cdot t_j}}{K} - \frac{1}{y_j} \right|. \tag{18}$$

The change of parameters $B := 1/K$ and $A := e^a/K$ changes the objective (18) to

$$g_\diamond(A, B, r) := \sum_{j=1}^{N} \left| B + A \cdot e^{-r \cdot t_j} - q_j \right|. \tag{19}$$

If the data $z_j = (t_j, y_j)$ increase, then the reciprocal data $\hat{z}_j = (t_j, q_j)$ decrease. Subsection 4.1 then shows that the objective (19) does not have a minimum where $A \leq 0$ or $r \leq 0$. The method fixes r and regresses A and B.

If the data $z_j = (t_j, y_j)$ increase sub-exponentially, then the reciprocal data $\hat{z}_j = (t_j, q_j)$ decrease super-exponentially. Subsection 4.2 then shows that the objective (19) does not have a minimum where $B \leq 0$. The method fixes B and regresses A and r. Theorem 26 also shows that the objective (19) does not have a minimum outside a compact interval, whence by continuity and compactness the objective (19) has a global minimum, with $A > 0$, $B > 0$, and $r > 0$. Theorem 27 then points out that the reciprocal of such a Mitscherlich curve is a Verhulst curve.

Though they occur in Examples 29 and 30, monotonicity and sub-exponentiality are not the rule in practice, but no other existence theorems appear known.

4.1 Features of Median Mitscherlich Curves for Super-Linear Data

For positive decreasing data, a horizontal line is not a median Mitscherlich curve, which excludes all cases with $A = 0$ or $r = 0$, as verified in Lemma 23.

Lemma 23 *For all real sequences $t_1 < \cdots < t_N$ and $q_1 > \cdots > q_N > 0$ with $N \geq 2$, the objective* (19) *does* not *reach a minimum where $A = 0$ or $r = 0$.*

Proof. If $A = 0$ or $r = 0$, then the objective (19) has a minimum where $A + B$ is a median of $q_1, \ldots, q_N$, and by Theorem 20 there is an index m such that $A + B = q_m$. The fitted curve is thus a horizontal line with equation $q = q_m = q_m \cdot e^{-0 \cdot (t - t_m)}$ through (t_m, q_m), which passes below every data point (t_j, q_j) with $j < m$, and above every data point (t_n, q_n) with $n > m$. Increasing r from 0 to $r > 0$ then decreases all the vertical distances to all the data points, until the curve with equation $q = q_m \cdot e^{-r \cdot (t - t_m)}$ passes through a second data point. □

For each $r \neq 0$, the change of variable $p := e^{-r \cdot t}$ with $p_j := e^{-r \cdot t_j}$ transforms median Mitscherlich curves into median lines with equation $q = A \cdot p + B$, to which all the concepts and methods of median lines apply.

Lemma 24 *For all real sequences $t_1 < \cdots < t_N$ and $q_1, \ldots, q_N$ with $N \geq 2$, and for each rate $r \neq 0$, there exist indices m and n with $1 \leq m < n \leq N$ such that the objective (19) reaches its constrained minimum, with r fixed, at the slope $A_{m,n}(r)$ and intercept $B_{m,n}(r)$ defined by*

$$A_{m,n}(r) := \frac{q_m - q_n}{p_m - p_n}, \tag{20}$$

$$B_{m,n}(r) := q_n - A_{m,n}(r) \cdot p_n. \tag{21}$$

Furthermore, if $q_1 > \cdots > q_N$, then r and $A_{m,n}(r)$ have the same sign.

Proof. For each $r \neq 0$, minimizing the objective (19) amounts to fitting a median line to the transformed data $s_k = (p_k, q_k) = (e^{-r \cdot t_k}, 1/y_k)$. By Theorem 22, the constrained minimum occurs at a line through two distinct points (p_m, q_m) and (p_n, q_n), with slope and intercept as in equations (20) and (21). If $q_1 > \cdots > q_N$, and if $r > 0$, then $A_{m,n}(r) > 0$ because p_j and q_j decrease as j increases, whereas if $r < 0$, then $A_{m,n}(r) < 0$ because p_j increases while q_j decreases. $\square$

If the data $z_j = (t_j, y_j)$ increase sub-exponentially, then the reciprocal data $\hat{z}_j = (t_j, q_j)$ decrease super-exponentially, which excludes concave Mitscherlich curves, for which $A < 0$ and $r < 0$. For use in Example 28, it suffices that for each triple of data (z_h, z_k, z_ℓ) with $t_h < t_k < t_\ell$, the third point z_ℓ lies on or below the rectangular hyperbola through z_h and z_k, or, equivalently, that the reciprocal data $\hat{z}_j = (t_j, q_j)$ decrease linearly or super-linearly, as verified in Lemma 25.

Lemma 25 *For all data $\hat{z}_j = (t_j, q_j)$, such that $t_1 < \cdots < t_N$ and $q_1 > \cdots > q_N > 0$ with $N \geq 3$, if every triple $(\hat{z}_h, \hat{z}_k, \hat{z}_\ell)$ is linear or super-linear, then the objective* (19) *does* not *reach a minimum where $A < 0$ or $r < 0$.*

Proof. If $r < 0$, then by Lemma 24 every constrained minimum of the objective (19) passes through two data points (t_m, q_m) and (t_n, q_n) with $t_m < t_n$ and a negative slope $A_{m,n}(r) < 0$. Thus the corresponding Mitscherlich curve is strictly concave and passes below every data point (t_j, q_j) with $j < m$ or $j > n$, and above every data point with $m < j < n$, because the linear or super-linear data lie on or above the line through (t_m, q_m) and (t_n, q_n) for $j < m$ or $j > n$, and on or below the same line for $m < j < n$. Increasing r then decreases all the vertical distances to all the other data points, for every $j \notin \{m, n\}$. $\square$

"Degenerate" cases also play a role in fitting curves to data. Thus, the "degenerate" cases in Lemma 13 help prove the existence of interpolating curves in Theorem 14, and also explain and remedy the absence of best-fitting curves.

Indeed, the proofs of Lemmas 23 and 25 reveal why some data might not admit of any best-fitting curve and then how to find a remedy. Still with $t_m < t_n$ and $q_m > q_n > 0$, if $r < 0$, then $B > q_m > q_n$ by Lemma 10. Equation (6) shows that for $r < 0$, if r increases to 0, then B diverges to $+\infty$. Hence Lemma 13 shows that the Mitschelich curve $M_{B;\hat{z}_m,\hat{z}_n}$ tends to $\text{line}_{\hat{z}_1,\hat{z}_2}$. If all the data points are also collinear, then $\text{line}_{\hat{z}_1,\hat{z}_2}$, is also the line supporting all the data. *Thus for collinear data the absence of any best-fitting Mitscherlich curve arises from an inadequate parametrization of the space of curves under consideration.*

One remedy re-parametrizes the space of Mitscherlich curves for r near 0. Multiplying tops and bottoms by r in formulae (20) and (21) leads to

$$M_{B;\hat{z}_m,\hat{z}_n}(t) = \frac{r \cdot (q_m - a_n)}{e^{-r \cdot t_m} - e^{-r \cdot t_n}} \cdot \frac{e^{-r \cdot t} - e^{-r \cdot t_m}}{r} + q_m.$$

The coefficients have a removable singularity at $r = 0$, defined in terms of

$$\varphi_{r;\hat{z}_m,\hat{z}_n}(t) := \frac{e^{-r\cdot t} - e^{-r\cdot t_m}}{r} = (t_m - t) + \sum_{k=2}^{\infty}(-1)^k \cdot \frac{t^k - t_m^k}{k!} \cdot r^{k-1},$$

so that the Mitscherlich curve with rate r through $\hat{z}_m$ and $\hat{z}_n$ is given by the formula

$$M_{r;\hat{z}_m,\hat{z}_n}(t) = \varphi_{r;\hat{z}_m,\hat{z}_n}(t) \cdot (q_n - q_m)/\varphi_{r;\hat{z}_m,\hat{z}_n}(t_n) + q_m. \tag{22}$$

With m and n fixed, substituting formula (22) for $B + A \cdot e^{-r\cdot t}$ shows that the objective (19) has a continuous extension through $r = 0$, which coincides with (19) for r near 0, but has a minimum value 0 at the line supporting collinear data. This extension with formula (22) also holds for data on a horizontal line, with $q_m = q_n$.

Similar considerations apply to $r > 0$ and $B < q_n < q_m$ by Lemma 10. If $r > 0$ and r decreases to 0, then B diverges to $-\infty$ by equation (6). The extension through $r = 0$ from both sides amounts to a one-point compactification at $B = \pm\infty$, preventing B from drifting away in theory or computations.

A two-point compactification occurs at $B \in \{q_m, q_n\}$, or, equivalently, $r = \pm\infty$, by Lemma 13. Thus, all data admit of a "generalized median Mitscherlich curve," which may be a line or a half-line.

Another remedy identifies conditions for a best-fitting curve, in Subsection 4.2.

4.2 Sufficient Conditions for Median Mitscherlich and Verhulst Curves

For monotonic positive data increasing sub-exponentially, this subsection proves that there is at least one Verhulst curve minimizing the objective $f_\diamond$ in equation (18). Also, at least one such minimizing Verhulst curve passes through at least two data points. From $q_n = A_{m,n}(r) \cdot e^{-r\cdot t_n} + B_{m,n}(r)$ with $A_{m,n}(r) > 0$ follows $B_{m,n}(r) < q_n$. Thus it suffices to search for a median Mitscherlich curve $M_{B;\hat{z}_m,\hat{z}_n}$ through two reciprocal data points $\hat{z}_m$ and $\hat{z}_n$ with $B < q_n$.

Theorem 26 *For each sequence of $N \geq 3$ positive data points where each triple of data points increases sub-exponentially, there exists a median Mitscherlich curve, minimzing the objective* (19)*, with $A > 0$, $B > 0$, and $r > 0$ for every such curve.*

Proof. If the triple (z_h, z_k, z_ℓ) increases sub-exponentially, then the reciprocal triple $(\hat{z}_h, \hat{z}_k, \hat{z}_\ell)$ decreases super-exponentially. By Theorem 14, there is a unique Mitscherlich curve $M_{\hat{z}_h;\hat{z}_k,\hat{z}_\ell}$ through $\hat{z}_h$, $\hat{z}_k$, $\hat{z}_\ell$ with a constant $B_{\hat{z}_h,\hat{z}_k,\hat{z}_\ell}$ such that $0 < B_{\hat{z}_h,\hat{z}_k,\hat{z}_\ell} < q_\ell$. Let $B_{\min} := \min_{h,k,\ell} B_{\hat{z}_h,\hat{z}_k,\hat{z}_\ell} > 0$. The proof first shows that the objective $g_\diamond(A(B;\hat{z}_h,\hat{z}_k), B, r(B;\hat{z}_h,\hat{z}_k))$ has no minima where $B < B_{\min}$.

An upper bound $B_{\min} < q_{\min} := \min_j q_j = q_N$ follows from the Mitscherlich curve $M_{\hat{z}_h;\hat{z}_k,\hat{z}_\ell}$ through $\hat{z}_1$, $\hat{z}_2$, $\hat{z}_N$ with constant $B < q_N$ by Theorem 14.

Recall from Lemma 10 that for each $B \notin [q_k, q_h]$ there is a unique Mitscherlich curve $M_{B;\hat{z}_h,\hat{z}_k}$ through any two reciprocal points $\hat{z}_h$ and $\hat{z}_k$ with constant B.

If $t_h < t_k$ and $t_\ell \notin [t_h, t_k]$ with $B < B_{\min}$, then $M_{B;\hat{z}_h,\hat{z}_k}(t_\ell) < M_{B_{\min};\hat{z}_h,\hat{z}_k}(t_\ell) \leq M_{\hat{z}_h;\hat{z}_k,\hat{z}_\ell}(t_\ell) = q_\ell$, by Lemma 12 applied to $q_{\min} > B_{\min} > B$. Also, every data points $\hat{z}_\ell$ with $t_\ell \notin [t_h, t_k]$ lies on or above $M_{B_{\min};\hat{z}_h,\hat{z}_k}$.

If $t_h < t_k$ and $t_h < t_\ell < t_k$ with $B < B_{\min}$, then $M_{B;\hat{z}_h,\hat{z}_k}(t_\ell) > M_{B_{\min};\hat{z}_h,\hat{z}_k}(t_\ell) \geq M_{\hat{z}_h;\hat{z}_k,\hat{z}_\ell}(t_\ell) = q_\ell$, again by Lemma 12 applied to $q_{\min} > B_{\min} > B$. Moreover, every data points $\hat{z}_\ell$ with $t_h < t_\ell < t_k$ lies on or below $M_{B_{\min};\hat{z}_h,\hat{z}_k}$.

Thus as B decreases but remains below $B_{\min}$, the vertical distances to the reciprocal data increase, and so does the objective $g_\diamond(A(B; \hat{z}_h, \hat{z}_k), B, r(B; \hat{z}_h, \hat{z}_k))$.

For all h and k with $1 \leq h < k \leq N$, similar arguments yield an upper bound for B. By Lemmas 23 and 25, the objective (19) has no minima where $r \leq 0$. Thus it suffices to search where $r > 0$. By Lemma 10 there are no Mitscherlich curves through $\hat{z}_h$ and $\hat{z}_k$ with $B \geq q_k$ and $r > 0$. Therefore it suffices to search where $B < q_k$. To this end, let $B_{h,k,\max} := \max_{\ell \notin \{h,k\}}\{B_{\hat{z}_h,\hat{z}_k,\hat{z}_\ell} : B_{\hat{z}_h,\hat{z}_k,\hat{z}_\ell} < q_k\}$, so that $B_{\min} \leq B_{\hat{z}_h,\hat{z}_k,\hat{z}_\ell} \leq B_{h,k,\max}$ for each $\ell \notin \{h, k\}$.

If $t_h < t_k$ and $t_\ell \notin [t_h, t_k]$ with $B_{h,k,\max} < B < q_k$, then $M_{B;\hat{z}_h,\hat{z}_k}(t_\ell) > M_{B_{h,k,\max};\hat{z}_h,\hat{z}_k}(t_\ell) \geq M_{\hat{z}_h;\hat{z}_k,\hat{z}_\ell}(t_\ell) = q_\ell$, by Lemma 12 applied to $C := B_{h,k,\max}$. Also, every data points $\hat{z}_\ell$ with $t_\ell \notin [t_h, t_k]$ lies on or below $M_{B_{h,k,\max};\hat{z}_h,\hat{z}_k}$.

If $t_h < t_k$ and $t_h < t_\ell < t_k$ with $B_{h,k,\max} < B < q_k$, then $M_{B;\hat{z}_h,\hat{z}_k}(t_\ell) < M_{B_{h,k,\max};\hat{z}_h,\hat{z}_k}(t_\ell) \leq M_{\hat{z}_h;\hat{z}_k,\hat{z}_\ell}(t_\ell) = q_\ell$, again by Lemma 12 with $C := B_{h,k,\max}$. Moreover, every data points $\hat{z}_\ell$ with $t_h < t_\ell < t_k$ lies on or above $M_{B_{h,k,\max};\hat{z}_h,\hat{z}_k}$.

Thus every data points $\hat{z}_\ell$ lies on or between $M_{B_{\min};\hat{z}_h,\hat{z}_k}$ and $M_{B_{h,k,\max};\hat{z}_h,\hat{z}_k}$. As B increases and $B_{h,k,\max} < B < q_k$, the vertical distances to the reciprocal data increase, and so does the objective $g_\diamond(A(B; \hat{z}_h, \hat{z}_k), B, r(B; \hat{z}_h, \hat{z}_k))$.

By formulae (6) and (7) in Lemma 10, the objective $g_\diamond(A(B; \hat{z}_h, \hat{z}_k), B, r(B; \hat{z}_h, \hat{z}_k))$ is a continuous function of $B < q_k$ and thus reaches a local minimum on the compact interval $[B_{\min}, B_{h,k,\max}]$ at some $B_{h,k} \in [B_{\min}, B_{h,k,\max}]$.

Among the finitely many pairs h and k with $1 \leq h < k \leq N$, there is a pair m and n with the smallest minimum value $g_\diamond(A(B_{m,n}; \hat{z}_m, \hat{z}_n), B, r(B_{m,n}; \hat{z}_m, \hat{z}_n))$, which is then a global minimum of the objective (19). Moreover, $B_{m,n} > 0$, $A(B_{m,n}; \hat{z}_m, \hat{z}_n) > 0$, and $r(B_{m,n}; \hat{z}_m, \hat{z}_n) > 0$. □

Theorem 27 *For each finite sequence of positive data, if every triple of data points increases sub-exponentially, then there exists a Verhulst curve minimizing $f_\diamond$ in (18).*

Proof. Because the reciprocal data decrease super-exponentially, Theorem 26 yields a median Mitscherlich curve with $A > 0$, $B > 0$, and $r > 0$, the reciprocal of which is a Verhulst curve, with $K = 1/B$ and $c = \ln(A/B)$. □

5 Case Studies

Regressions are used to estimate parameters, identify mechanisms, or predictions.

5.1 Examples With and Without Best-Fitting Verhulst Curves

To estimate parameters, G. F. Gause measured the growth of the yeast *Schizosaccharomyces kefir* (also spelled "*kephir*" [15]), reporting exactly three data points from each of two experiments. Each data point (t_j, y_j) consists of a time t_j and the average number y_j of cells per square of a Thoma counting chamber [15, Ch. IV].

Example 28 Figure 2 lists the data from the first experiment [14, p. 395, Table I, Exp. 1], [15, p. 143, Table 1, Exp 1]. The third point lies above the exponential curve, but below the hyperbola, through the first two points: the data are super-exponential but their reciprocals are super-linear. By Theorem 18 are *no* best-fitting Verhulst curves relative to the regressions described in Remark 3. By Theorem 14 there is a reciprocal Mitscherlich curve with $B < 0 < A$ through all three points.

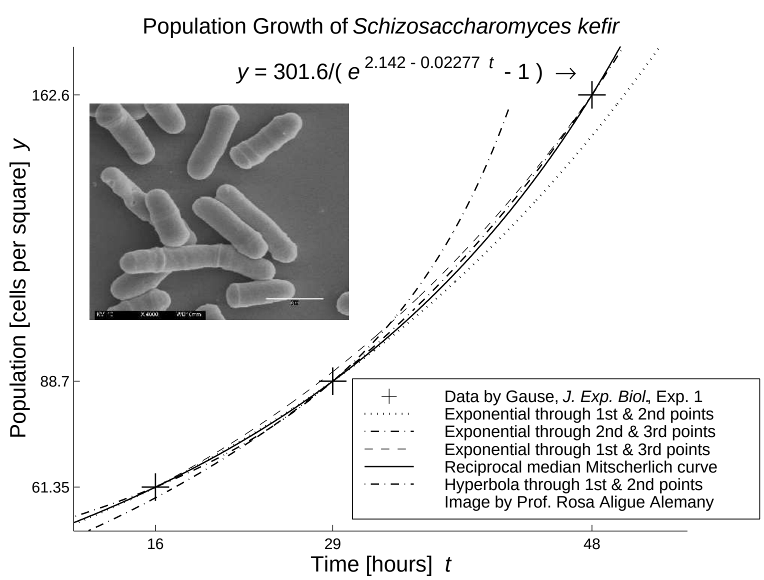

Figure 2: The third point is above the exponential $(\cdots)$ but below the hyperbola $(- \cdot -)$ through the first two points: there are no best-fitting Verhulst curves, but a reciprocal Mitscherlich curve $(-)$ passes through all data points. Photograph of *Schizosaccaromyces pombe* courtesy Professor Rosa M. Aligué Alemany, Universitat de Barcelona, Facultat de Medicina, Departament de Biologia Cellular, Spain.

Example 29 Figure 3 lists the data from the second experiment [14, p. 395, Table I, Exp. 2], [15, p. 143, Table 1, Exp 2], showing that the third data point lies below the exponential curve through the first two data points. By Theorem 17, there exists exactly one Verhulst curve through all three data points, computed by Newton's Method and displayed in figure 3. This curve is also the best-fitting Verhulst curve relative to any regression criterion, because all the residuals vanish. Yet for classroom demonstrations table 1 shows that software straight out of the box may still need initial values from the user, even for finding an exact fit.

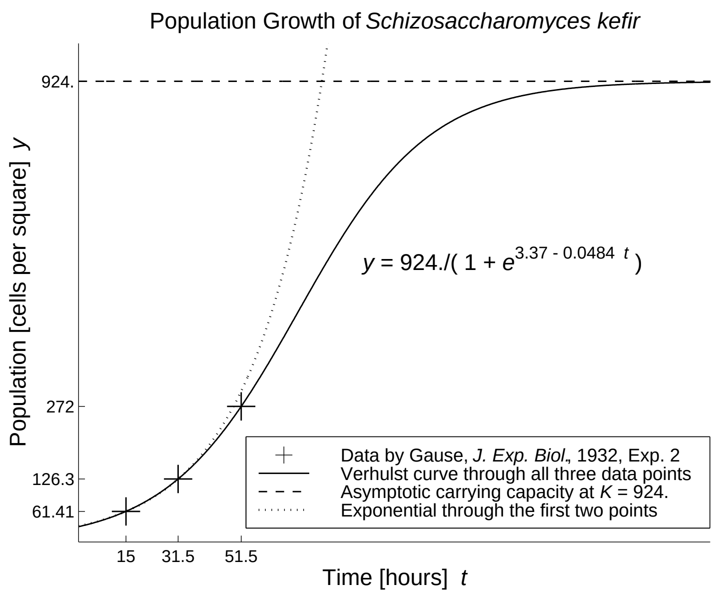

Figure 3: The third data point lies below the exponential $(\cdots)$ through the first two data points. By Theorem 17, there exists a unique best-fitting Verhulst curve $(-)$ through all three data points, computed here by Newton's Method.

Table 1: Fitting a Verhulst curve to the data in figure 3 with Matlab's `cftool`.

INITIAL VALUES				FITTED VALUES *			GOODNESS OF FIT	
FROM	K_0	a_0	r_0	$\hat{K}$	$\hat{a}$	$\hat{r}$	R^2	SSE†
default	0.907	0.862	0.593	153.3	1.042	0.09519	0.3206	1.6×10^4
Newton	924.0	3.37	0.0484	923.9	3.369	0.04844	1	8.9×10^{-17}

*Matlab 7 with its curve-fitting toolbox and default number of iterations on a Power Mac G4
†Sum of Squared Errors

5.2 An Example With a Median Reciprocal Verhulst Curve

In chemistry, the type of the fitted curve, rather than the values of its parameters, is "the principal evidence for the postulated mechanisms" of a reaction [12, p. 10].

Example 30 Figure 4 shows Ostwald's data from the hydrolysis of ethyl acetate with acetic acid [37, p. 481, Table XLIII]. The units are the volumes of barium hydroxide ($Ba(OH)_2$) necessary to titrate the acetic acid, with time in minutes [37, p. 451, Table I]. The concentration is the sum $y + 1338$ of the measured excess y over the initial concentration 1338. Each triple of data increases sub-exponentially. By Theorem 27, there is a median reciprocal Verhulst curve, computed here as

$$y + 1338 = \frac{2601.}{1 + e^{-0.06258 - 0.004146 \cdot t}}. \tag{23}$$

This curve differs from Reed & Berkson's [43, p. 770] for the following reason.

With water (H_2O) and acetic acid ($C = CH_3COOH$) as an auto-catalyst, ethyl acetate ($A = CH_3COOCH_2CH_3$) decomposes into more acetic acid and ethyl alcohol ("ethanol": $E = CH_3CH_2OH$) according to the *stoichiometric* equation

$$A + H_2O \quad \overset{C}{\rightleftharpoons} \quad C + E. \tag{24}$$

(The "ethyl" in A, C, and E is the group C_2H_5.) Hence the sum $K = [A]_0 + [C]_0$ of the initial concentrations is an upper bound for the concentration [C] of acetic acid. In this example, $[A]_0 = 1370$ and $[C]_0 = 1338$, so that $K = 1370 + 1338 = 2708$. With $K = 2708$ fixed, Reed & Berkson fitted only the parameters a and r [43, pp. 770–771]. Yet the initial concentrations are measured as are the other data points and so are only "approximate" ("rund" and "annähernd" [37, p. 480 & p. 482]), so that K may also be estimated. Table 2 shows that the Verhulst curve fitted by Matlab's `cftool` with nonlinear least-squares agrees to three significant digits with the median reciprocal Verhulst curve (23). The constant term $[C]_0 = 1338$ in equation (23) could also be estimated, but this is another story for another day.

Table 2: Fitting a Verhulst curve to the data ($+$) in figure 4 with Matlab's `cftool`.

	Initial Values			Fitted Values *			Fit	
from	K_0	a_0	r_0	$\hat{K}$	$\hat{a}$	$\hat{r}$	R^2	SSE†
default	0.444	0.478	0.690	1920	0.4781	0.69	0	$4. \times 10^5$
Reed & Berkson	2708	-0.0138	0.00372	2602	-0.0641	0.004125	1	15.28

*Matlab 7 with its curve-fitting toolbox on a Power Mac G4

†Sum of Squared Errors

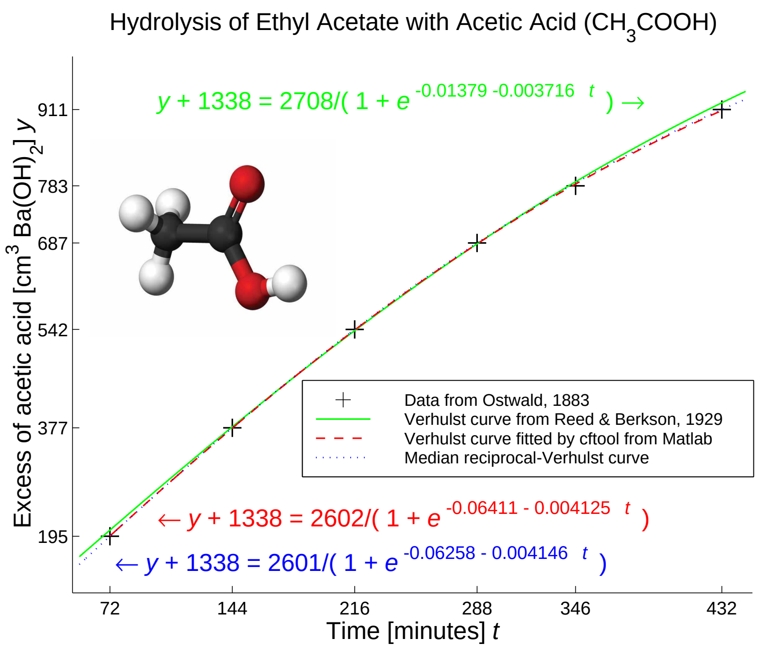

Figure 4: Verhulst curves fit to data $(+)$ from Ostwald [37, p. 481, Table XLIII] by Reed & Berkson $(-)$ [43, pp. 770–771], Matlab's `cftool` $(--)$, and median reciprocal $(\cdots)$. The units and reactions are explained in Example 30. Public-domain picture of an acetic acid molecule by Benjah-bmm27 (http://en.wikipedia.org/wiki/File:Acetic-acid-3D-balls.png).

5.3 An Example for Which a Best-Fitting Verhulst Curve is Elusive

In biology, the acid test for a model is its predictive accuracy [18, p. 1252].

Example 31 Figure 5 displays data on the growth of the population of whooping cranes (*Grus americana*) in the flock that migrates between the Wood Buffalo National Park in the Canadian Northwest Territories and the Aransas National Wildlife Refuge in Texas [48, Table 1, pp. 12–13], an example suggested by Allen [1, Table 3.3, p. 139]. For these data, the existence of a best-fitting Verhulst curve remains an open question. However, two final data points for 2004 and 2005 appeared in the more recent report from 2007 [48], which was not available in 2006 for Allen's text [1]. Such additional data provide a test for any fitted model:

> A good model not only describes and explains, but also predicts; otherwise modeling is merely a curve-fitting exercise. Model validation is about testing model predictability on a data set not used to estimate

> the parameters. The most convincing models are those further tested by making *a priori* predictions that are then borne out by new experiments [18, p. 1252].

Accordingly, figure 5 shows several models fitted to the data for 1938–2003 only. The exponential model fitted by logarithmic-linear regression as in Remark 6 is used only for initial values and refined by Matlab's `cftool` into the model (25):

$$Y(t) = e^{3.979+0.04102\cdot(t-1970)}. \tag{25}$$

The Verhulst curve fitted by Allen "only for illustration purposes" [1, p. 92] also gives starting values and is refined by Matlab's `cftool` into the model (26):

$$Y(t) = \frac{696.8}{1 + e^{97.19-0.04808\cdot t}}. \tag{26}$$

Table 3 again shows that software may still benefit from guidance by the user.

Table 4 shows that the sums of squared differences between either model and the 1938–2003 data have the same magnitude. Yet in comparing the predictions for 2004 and 2005 from these models, table 4 and figure 5 show how in 2004 and 2005 the population was still growing faster than the nonlinear least-squares Verhulst model but straddles the nonlinear least-squares exponential model with a smaller sum of squared errors in the predicted values. Matlab's `cftool` also gives the confidence interval $15.6 \leq K \leq 1378$, reinforcing Allen's statement that "an estimate for K is not known" [1, p. 92]. Nevertheless, reflecting the convergence from Lemma 13, the growth rates of the nonlinear least-squares exponential and Verhulst models agree with each other to the first significant digit (they are both within 0.005 away from 0.045), which suggests that the carrying capacity *need not* be estimated while the growth rate r *can* be estimated. That, however, is yet another story for yet another day [30].

Table 3: Fitting a Verhulst curve to the 1938–2003 data $(+)$ in figure 5 with Matlab's `cftool`.

	INITIAL VALUES			FITTED VALUES *			FIT	
FROM	K_0	a_0	r_0	$\hat{K}$	$\hat{a}$	$\hat{r}$	R^2	SSE†
default	0.0995	0.555	0.0466	74.06	0.5554	0.0466	0	2.0×10^5
Allen [1]	500.0	90.6546	0.04497	696.8	97.19	0.04808	0.97	6.0×10^3

*Matlab 7 with its curve-fitting toolbox and "Robust" option on a Power Mac G4

†Sum of Squared Errors

Table 4: Predictions from exponential and Verhulst's models

MODEL Y	SSE*	2004	2005	SSE†
Data [48, Table 1, pp. 12–13]	0	217	220	0
Exponential (25)	6×10^3	216	225	26
Verhulst (26)	6×10^3	210	218	53

*Sum of Squared Errors for the data from 1938 through 2003.
†Sum of Squared Errors for the predictions: $[Y(2004) - 217]^2 + [Y(2005) - 220]^2$.

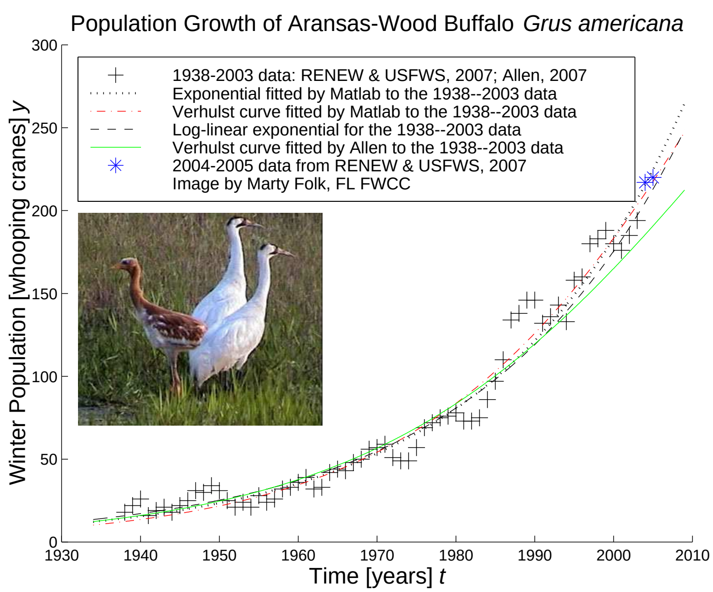

Figure 5: Comparisons of predictions for the 2004 and 2005 data $(*)$ from models fitted to the 1938–2003 data $(+)$: the Verhulst curve $(-)$ from [1, p. 92] with its refinement $(- \cdot -)$ by Matlab's `cftool`, and the logarithmic-linear least-squares exponential $(--)$ with its refinement $(\cdots)$ by Matlab's `cftool`,

6 Conclusions

The present considerations prove that there exist data that do not admit of any best-fitting curve of a specified type, regardless of perturbations of the data within a specific non-empty open domain. This counterexample demonstrates the necessity of theorems to establish the existence of a best-fitting curve or surface for each situation. For instance, for data with positive ordinates, the present considerations provide such a theorem and method to fit a Verhulst curve to three non-necessarily equally spaced points in the plane, another theorem guaranteeing the existence of a median Mitscherlich curve provided that all triples of data decrease super-exponentially, and another theorem guaranteeing the existence of a median reciprocal Verulst curve provided that all triples of data increase sub-exponentially. Barring such theorems arises the challenge to base conclusions, for example, the probabilistic distribution of fitted parameters, on facts other than a best-fitting curve. In particular, the existence of a best-fitting Verhulst curve relative to criteria other than least absolute deviations, such as least squares, appears to be an open question. More generally, the existence of a best-fitting curve, surface, or model, of specified classes other than Verhulst, such as Lotka-Voltera models, and relative to any metric, also appears to be an open question.

7 Acknowledgments

This work was supported in part by a Faculty Research and Creative Works Grant from Eastern Washington University. I thank Dr. John Douglas, retired professor of chemistry, for his insight into the chemistry of Example 30. I also thank the reviewers for suggestions that vastly improved the manuscript.

8 Appendix: Geometric Adjustments of Data

This section sets up a framework for adjustments of data by many types of regression, which share one common feature that suffices to establish the absence of any best-fitting Verhulst curve relative to any such regression, as described in Remark 3: For each curve C of a specified class, and for each sequence of distinct data points $D = (z_1, \ldots, z_N)$, all such regressions identify adjusted points $\tilde{D} = (\tilde{z}_1, \ldots, \tilde{z}_N)$ on C, and then minimize an objective $F_D : \tilde{D} \mapsto F_D(\tilde{D}) \in \mathbb{R}_+$ that is a topologically open map and continuous function of $\tilde{D}$ such that

$$F_D(\tilde{D}) = \min\{F_D(p_h, p_k, p_\ell) : p_h, p_k, p_\ell \in C\},$$

so that $\mathcal{F}_D(C) := F_D(\tilde{D})$ is well defined. For each curve C, the map $D \to \mathcal{F}_D(C)$ is a continuous function of the data D. Also, $F_D(\tilde{D}) = 0$ if and only if $D = \tilde{D}$.

The types of regression considered here involve the following objects:

(A) A non-empty topological space $(U, \mathcal{E})$, such as the real line $U = \mathbb{R}$ or plane $U = \mathbb{R}^2$, or any subset thereof, with the Euclidean topology $\mathcal{E}$.

(B) A non-empty set $\mathcal{D}$ of non-empty finite sequences in U, called *data sets* or *data sequences* $D := (z_1, \ldots, z_N)$; each $z_i \in U$ is called a *data point*.

(C) A non-empty set $\mathcal{C}$ of non-empty *closed* subets of U, such as curves or surfaces, but henceforth called *curves*, with a topology $\mathcal{T}$ on $\mathcal{C}$.

(D) A function A that assigns to each data point z_i in each data sequence D and to each curve C an *adjusted point* $\tilde{z}_i \in C$; the adjusted point $\tilde{z}_i$ may be the point on C closest to z_i relative to a metric, which may depend on i.

(E) An *objective* function $\mathcal{F} : \mathcal{D} \times \mathcal{C} \to \mathbb{R}_+ := [0, \infty[$, or, equivalently, for each set $D \in \mathcal{D}$ a continuous and *open* map $\mathcal{F}_D : \mathcal{C} \to \mathbb{R}_+$, which factors through A and a function F_D in the form $\mathcal{F}_D(C) = F_D[A(z_1, C), \ldots, A(z_N, C)] = F_D(\tilde{z}_1, \ldots, \tilde{z}_N)$ so that $\mathcal{F}$ depends only on D and the adjusted points on C.

If $\mathcal{F}_D$ reaches a minimum at an element $C \in \mathcal{C}$, then C is called a *best-fitting* point, curve, surface, manifold, or variety, relative to $\mathcal{F}_D$. The regression then consists of identifying such a best-fitting object $C \in \mathcal{C}$. For each $C \in \mathcal{C}$, through summation, integration, or other agregation, the objective $\mathcal{F}_D$ may depend on a measure of discrepancy $\rho_z(z, C) = \rho_z(z, \tilde{z})$, called a *residual*, between each data point $z \in D$ and the adjusted point $\tilde{z} \in C$. Different points z and w may be subject to different residual functions ρ_z and ρ_w. The choice of $\mathcal{F}_D$ may arise from geometry or statistics that will not matter here: the existence of a best-fitting object $C \in \mathcal{C}$ depends only on the topological features of the objective $\mathcal{F}_D$. The framework just described extends to multiple adjusted points for some data points, provided that the residuals are independent of the choice of the adjusted point.

Example 32 *Orthogonal* regression is based on the Euclidean distance d from each point $z \in U$ to $C \subset U$. In particular, there exists at least one point $\tilde{z} \in C$ where $d(z, C) = d(z, \tilde{z})$, because C is closed and non-empty. Another closest point $\tilde{z}' \in C$ may exist, but then $d(z, \tilde{z}') = d(z, C) = d(z, \tilde{z})$. The residual is then the difference $\rho_z(z, C) := z - \tilde{z} \in U$, and the objective may be $\mathcal{F}_D(C) := F_D(\tilde{D}) := \|d(z_1, \tilde{z}_1), \ldots, d(z_N, \tilde{z}_N)\|$ with any norm $\| \ \|$ on $\mathbb{R}^N$.

Example 33 For *ordinary least-squares* regression, to each point z corresponds an affine subspace $U_z \subseteq U$ through z with a Euclidean distance d_z on U_z, and each object $C \in \mathcal{C}$ intersects U_z. For instance, if $z = (t, y) \in \mathbb{R}^2$, then regression of y versus t corresponds to the vertical line $U_z = \{(t, q) : q \in \mathbb{R}\}$ with $d_z[(t, y), (t, q)] = |y - q|$. Similarly, regression of t versus y corresponds to the horizontal line $U_z = \{(h, y) : h \in \mathbb{R}\}$ with $d_z[(t, y), (h, y)] = |t - h|$. In either case, there exists a point $\tilde{z} \in C \cap U_z$ where $d_z(z, C) = d_z(z, \tilde{z})$. The residual is again the difference $\rho_z(z, C) := z - \tilde{z} \in U_z$, and the objective may be the squared Euclidean norm $\mathcal{F}_D(C) := F_D(\tilde{D}) := \|d(z_1, \tilde{z}_1), \ldots, d(z_N, \tilde{z}_N)\|_2^2$ on $\mathbb{R}^N$.

Example 34 In Example 33, an affine space $U_z = \{z\}$ corresponds to the constraint that all curves in $\mathcal{C}$ pass through the point z.

Example 35 *Weighted* ordinary least-squares regression uses a symmetric positive definite matrix W, for instance, the inverse of the correlation matrix of the data, and minimizes $F_D(\vec{\rho}) := \vec{\rho}^T \cdot W \cdot \vec{\rho}$ for residuals $\rho_k = z_k - \tilde{z}_k$. If W is diagonal, then the regression is *uncorrelated*.

In a common framework accommodating all the foregoing examples, to each point $z \in U$ corresponds an affine space $U_z \subseteq U$ through z, endowed with a distance d_z defined by a norm $| \ |_z$ on U_z, which is thus topologically equivalent to the Euclidean norm $\| \ \|_2$ on U_z. Also, each object $C \in \mathcal{C}$ intersects every subspace U_z. Hypothesis 36 specifies the types of regression considered here.

Hypothesis 36 For the purpose of geometric adjustments of data, the objective

$$F_D : \mathbb{U} = U_{z_1} \times \cdots \times U_{z_N} \to \qquad \mathbb{R}_+$$

is continuous and maps open subsets of $\mathbb{U}$ to relatively open subsets of $\mathbb{R}_+$, with $F_D(w_1, \ldots, w_N) = 0$ if and only if $w_j = z_j$ for every $z_j \in D$. In particular, $\mathcal{F}_D(C) = 0$ if and only if the curve $C \in \mathcal{C}$ passes through every data point z_j. For each data sequence $(z_1, \ldots, z_N) \in \mathbb{U}$ and for each curve $C \in \mathcal{C}$, there are *adjusted* points $(\tilde{z}_1, \ldots, \tilde{z}_N) \in C^N \cap \mathbb{U}$ where F_D has a minimum on $C^N \cap \mathbb{U}$, denoted by

$$\begin{aligned} \mathcal{F}_D(C) &:= \min_{(\forall j)(w_j \in C \cap U_{z_j})} F_D(w_1, \ldots, w_N) \\ &= F_D(\tilde{z}_1, \ldots, \tilde{z}_N). \end{aligned}$$

Finally, in the particular case of Verhulst curves with exactly three data points $(N = 3)$ growing super-exponentially, the adjusted points are also distinct.

Example 37 Each positive diagonal 2×2 matrix A defines a weighted norm $| \ |_A$ on $\mathbb{R}^2$ by $|z|_A^2 := z^T \cdot A \cdot z$. For the corresponding A-orthogonal ℓ_p-regression,

$$F_D(w_1, \ldots, w_N) = \sum_{j=1}^{N} \left[(z_j - w_j)^T \cdot A \cdot (z_j - w_j)\right]^{p/2}.$$

For weighted ordinary regression with a positive definite matrix W, let $\rho_j := \vec{e}_i^T \cdot (z_j - w_j)$, with $\vec{e}_i := (1,0)$ for x vs. y, and $\vec{e}_i := (0,1)$ for y vs. x:

$$F_D(w_1, \ldots, w_N) = \vec{\rho}^T \cdot W \cdot \vec{\rho}.$$

More generally, with the norm $|\ |_A$ on every U_{z_j} and any norm $\|\ \|$ on $\mathbb{R}^N$,

$$F_D(w_1, \ldots, w_N) = \|(|z_1 - w_1|_A, \ldots, |z_N - w_N|_A)\|.$$

Because all norms are topologically equivalent on Euclidean spaces, each such objective F_D is open, even if some (but not all) subspaces U_{z_j} reduce to singletons, corresponding to curves constrained to pass through some (but not all) data points. Also, $\mathcal{F}_D(C) = 0$ if and only if $z_j \in C$ for every j. For such regression methods, proposition 38 confirms that increasing points have distinct adjusted points on increasing differentiable functions, where normals have negative slopes.

Proposition 38 *For each ordinary or orthogonal regression (weighted, unweighted, correlated, or uncorrelated), each* increasing *data sequence $D = (z_1, \ldots, z_N)$, such that $z_j = (t_j, y_j)$ with $t_1 < \cdots < t_N$ and $y_1 < \cdots < y_N$, has* pairwise distinct *adjusted points $\tilde{D} = (\tilde{z}_1, \ldots, \tilde{z}_N)$ on each curve C that is the graph of a differentiable function V with $V' > 0$ everywhere.*

Proof. Each adjusted point $\tilde{z}_j$ lies in the subspace $U_{z_j} \subseteq \mathbb{R}^2$, which is a vertical line for ordinary regression of y vs. t, or a horizontal line for ordinary regression of t vs. y. In either case if $k \neq \ell$, then $U_{z_k} \cap U_{z_\ell} = \varnothing$, whence $\tilde{z}_k \neq \tilde{z}_\ell$. For orthogonal regression, $U_{z_j} = \mathbb{R}^2$, but each adjusted point $\tilde{z}_j = (\tilde{t}_j, \tilde{y}_j)$ lies on the normal from z_j to V, which has a *negative* slope $-1/V'(\tilde{t}_j) < 0$. In contrast, the line through any two distinct data points z_k and z_ℓ has a *positive* slope $(y_k - y_\ell)/(t_k - t_\ell) > 0$. Consequently, z_k, $\tilde{z}_k$, and z_ℓ are not collinear, whence $\tilde{z}_k \neq \tilde{z}_\ell$. $\square$

In orthogonal regressions, each data point z_j lies on the normal to the curve at the adjusted points $\tilde{z}_j$ only because the Euclidean disc centered at z_j and passing through $\tilde{z}_j$ is tangent to the curve at $\tilde{z}_j$. Therefore, proposition 38 can fail for orthogonal regressions relative to Minkowski norms whose unit discs, although convex, might not have anough symmetries.

Some regressions first transform the ambient space by an order-preserving homeomorphism $H : U \to \Gamma$, mapping increasing curves and data to increasing curves and data, which may be the identity $I : U \to U$, minimize an objective Φ_Δ of the transformed data $\Delta := H(D)$, and map the object fitted in Γ back into U, as in logarithmic-linear regression. If Φ_Δ satisfies hypothesis 36, then so does $F_D := \Phi_{H(D)} \circ (H \times \cdots \times H)$, because H is an order-preserving homeomorphism.

References

[1] Linda J. S. Allen. *An Introduction to Mathematical Biology*. Pearson/Prentice Hall, Upper Saddle River, NJ, 2007. ISBN 0-13-035216-0; QH323.5.A436 2007; LCCC No. 2006042585; 570.1'5118–dc22.

[2] W. W. Alpatov and Raymond Pearl. Experimental studies on the duration of life. XII. Influence of temperature during the larval period and adult life on the duration of the life of the imago of *Drosophila melanogaster*. *The American Naturalist*, 63(684):37–67, January–February 1929.

[3] Anders H. Anderson and Anne Anderson. Life history of the Cactus Wren. part IV: Development of nestlings. *The Condor*, 63(1):87–94, January-February 1961.

[4] Mary M. Ballyk and Gail S. K. Wolkowicz. Classical and resource-based competition: a unifying graphical approach. *Journal of Mathematical Biology*, 16 February 2010.

[5] I. Barrodale and F. D. K. Roberts. Application of mathematical programming to ℓ_p approximation. In J. B. Rosen, O. L. Mangasarian, and K. Ritter, editors, *Nonlinear Programming*, volume 25 of *Mathematics Research Center The University of Wisconsin*, pages 447–464, New York, NY, 1970. Academic Press. QA3.U45 no. 25; LCCC No. 75-1132012.

[6] Manuel Blum, Vaughan Pratt, Robert E. Tarjan, Robert W. Floyd, and Ronald L. Rivest. Time bounds for selection. *Journal of Computer and System Science*, 7(4):448–461, August 1973.

[7] Nicholas Ferris Britton. *Essential Mathematical Biology*. Springer-Verlag, London, UK, 2003. ISBN 1-85233-536-X; QH 323.5.B747 2002; LCCC No. 2002036455; 570'.1'51–dc21.

[8] Fabio Cavallini. Fitting a logistic curve to data. *College Mathematics Journal*, 24(3):247–253, May 1993.

[9] Harald Cramér. *Mathematical Methods of Statistics*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1999. ISBN 0-691-000547-8; QA276.C72 1999; 99-12156; 519.5–dc21; first edition: Almqvist & Wiksells, Uppsala, Sweden, 1945.

[10] Gerda de Vries, Thomas Hillen, Mark Lewis, Johannes Müller, and Birgitt Schönfisch. *A Course in Mathematical Biology*, volume 12 of *Mathemati-*

cal Modeling and Computation. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2006. ISBN 0-89871-612-8; QH323.5.C69 2006; 570.1'5118–dc22; LCCC No. 2006044305.

[11] Ernest C. Driver. The temperature-effective period — the key to eye facet number in *Drosophila*. *Journal of Experimental Zoology*, 46(3):317–332, November 1926.

[12] Arthur A. Frost and Ralph G. Pearson. *Kinetics and Mechanism: A Study of Homogeneous Chemical Reactions*. Wiley, New York, NY, second edition, 1961. LCCC No. 61-6773.

[13] G. F. Gause. Ecology of populations. *Quarterly Review of Biology*, 7(1):27–46, March 1932.

[14] G. F. Gause. Experimental studies on the struggle for existence. *Journal of Experimental Biology*, IX:389–402, 1932.

[15] G. F. Gause. *The Struggle for Existence*. Dover, Mineola, NY, 2003. QH375.G35 2003; 576.8'2–dc21; LCCC No 2002037091. First published by Hafner in 1969.

[16] G. F. Gause and W. W. Alpatov. Die logistische Kurve von Verhulst-Pearl und ihre Anwendung im Gebiet der quantitativen Biologie. *Biologishes Zentralblatt*, 51(1/2):1–14, November-December 1931.

[17] H. O. Hartley. An analysis of variance test for normality (complete samples). *Biometrika*, 35(1/2):32–45, May 1948.

[18] Shandelle M. Henson and James L. Hayward. The mathematics of animal behavior: An interdisciplinary dialogue. *Notices of the American Mathematical Society*, 57(10):1248–1258, November 2010.

[19] M. Kunitz. Formation of trypsin from trypsinogen by an enzyme produced by a mold of the genus *Penicillium*. *Journal of General Physiology*, 21(5):601–620, 20 May 1938.

[20] M. Kunitz. Formation of trypsin from crystalline trypsinogen by means of enterokinase. *Journal of General Physiology*, 22(4):429–446, 20 March 1939.

[21] M. Kunitz and John H. Northrop. Isolation from beef pancreas of crystalline trypsinogen, trypsin, a trypsin inhibitor, and an inhibitor-trypsin compound. *Journal of General Physiology*, 19(6):901–1007, 20 June 1936.

[22] Donald Leach. Re-evaluation of the logistic growth curve for human populations. *Journal of the Royal Statistical Society. Series A. General*, 144(1):94–103, 1981.

[23] Hans Lewy. An example of a smooth linear partial differential equation without solution. *The Annals of Mathematics, Second Series*, 66(1):155–158, July 1957. Erratum: Vol. 68, No. 1 (July 1958), p. 202.

[24] Hans Lewy. Erratum: An example of a smooth linear partial differential equation without solution. *The Annals of Mathematics, Second Series*, 68(1):202, July 1958.

[25] H. Martini and N. M. Korneenko. Approximating finite weighted point sets by hyperplanes. In J. R. Gilbert aned R. Karlsson, editor, *SWAT 90: 2nd Scandinavian Workshop on Algorithm Theory, Bergen, Sweden, July 11–14, 1990 Proceedings*, volume 447 of *Lecture Notes in Computer Science*, pages 276–286, Berlin Heidelberg New York, 1993. Springer-Verlag. ISBN 3-540-52846-6; QA76.L42 v. 447.

[26] Horst Martini and Nikolai M. Korneenko. Chapter VI. hyperplanes approximation and related topics. In János Pach, editor, *New Trends in Discrete and Computational Geometry*, pages 135–161, Berlin Heidelberg New York, 1993. Springer-Verlag. ISBN 3-540-55713-X; QA448.D38N49 1993; LCCC 92-23684.

[27] Horst Martini and Anita Schöbel. Median hyperplanes in normed spaces — a survey. *Discrete Applied Mathematics*, 89:181–195, 1998.

[28] John H. Mathews. Bounded population growth: A curve fitting lesson. *Mathematics and Computer Education*, 26(2):169–176, Spring 1992. Supplementary *Mathematica* notebook: `http://mathews.ecs.fullerton.edu/n2003/LogisticEquationMod.html`.

[29] Margaret R. McDonald and M. Kunitz. The effect of calcium and other ions on the autocatalytic formation of trypsin from trypsinogen. *Journal of General Physiology*, 25(1):53–73, 20 September 1941.

[30] Hongyu Miao, Xiaohua Xia, S. Perelson, and Hulin Wu. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review*, 53(1):3–39, March 2011.

[31] James G. Morris and John P. Norback. A simple approach to linear facility location. *Transportation Science*, 14(1):1–8, 1980.

[32] J. C. Nash. A discrete alternative to the logistic growth function. *Applied Statistics*, 26(1):9–14, 1977.

[33] Yves Nievergelt. Median spheres: Theory, algorithms, applications. *Numerische Mathematik*, 114(4):573–606, February 2010.

[34] F. R. Oliver. Methods of estimating the logistic growth function. *Applied Statistics*, 13(2):57–66, 1964.

[35] F. R. Oliver. Aspects of maximum likelihood estimation of the logistic growth function. *Journal of the American Statistical Association*, 61(315):697–705, September 1966.

[36] F. R. Oliver. Notes on the logistic curve for human populations. *Journal of the Royal Statistical Society. Series A. General*, 145(3):359–363, 1982.

[37] Wilh. Oswald. Studien zur chemischen Dynamik. zweite Abhandlung: Die Einwirkung der Säuren auf Methylacetat. *Journal für Praktische Chemie*, 28(1):449–495, 30 June 1883.

[38] H. D. Patterson. A simple method for fitting an asymptotic regression curve. *Biometrics*, 12(3):323–329, September 1956.

[39] Raymond Pearl and Lowell J. Reed. On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Sciences*, 6(6):199–272, 15 June 1920.

[40] Jean-François Pechère and Hans Neurath. Studies on the autocatalytic activation of trypsinogen. *Journal of Biological Chemistry*, 229(1):389–407, 1 November 1957.

[41] D. F. Phipps. The logistic curve: A fitting technique. *The Statistician*, 24(2):129–136, June 1975.

[42] Frederico Pimentel Gomes. The use of Mitscherlich's regression law in the analysis of experiments with fertilizers. *Biometrics*, 9(4):498–516, December 1953.

[43] Lowell J. Reed and Joseph Berkson. The application of the logistic function to experimental data. *The Journal of Physical Chemistry*, 33(5):760–779, January 1929.

[44] E. C. Rhodes. Population mathematics — III. *Journal of the Royal Statistical Society*, 103(3):362–387, May 1940.

[45] Robert E. Ricklefs. A graphical method of fitting equations to growth curves. *Ecology*, 48(6):978–983, November 1967.

[46] H. L. Royden. *Real Analysis*. Macmillan, New York, NY, 2nd edition, 1968.

[47] A. Schönhage, M. Paterson, and N. Pippenger. Finding the median. *Journal of Computer and System Science*, 13(2):184–199, October 1976.

[48] Canadian Wildlife Service and U.S. Fish and Wildlife Service. International recovery plan for the whooping crane (*Grus americana*). Technical report, Recovery of Nationally Endangered Wildlife (RENEW) and U.S. Fish and Wildlife Service, Ottawa, Canada, and Albuquerque, NM, March 2007.

[49] W. L. Stevens. Asymptotic regression. *Biometrics*, 7(3):247–267, September 1951.

[50] Pierre-François Verhulst. Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux Memoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18(1):1–45, 1845.

[51] S. Viera and R. Hoffmann. Comparison of the logistic and the Gompertz growth functions considering additive and multiplicative error terms. *Applied Statistics*, 26(2):143–148, 1977.

[52] Eitan Zemel. An O(n) algorithm for the linear multiple choice knapsack problem and related problems. *Information Processing Letters*, 18(3):123–128, 30 March 1984.

Department of Mathematics, Eastern Washington University, 216 Kingston Hall, Cheney, Washington 99004-2418, USA
ynievergelt@ewu.edu

FACULTY RESEARCH AND CREATIVE WORKS GRANT: REPORT FOR 2010, PART III

YVES NIEVERGELT*

Abstract. Despite two centuries of study, practical difficulties occasionally still arise in fitting Verhulst logistic curves to data, or, equivalently, Mitscherlich's shifted exponential curves to the reciprocals of the data. Typical obstacles include divergent iterations, negative values for theoretically positive parameters, or the absence of any solution. The analysis presented here reveals that such difficulties occur near removable singularities of the objective functions to be minimized for the regression. Moreover, such singularities lie at the transition to different types of curves, including exponentials, hyperbolae, horizontal lines, and generalized Verhulst logistic curves. Removing the singularities — in the sense of complex analysis — encompasses all such curves into a compactified topological space, which guarantees the existence of a global minimum for the continuous objective function, and which also provides a smooth and transparent transition from one type of curve to another: The type of the fitted curve is automatically determined by the location of the minimum. However, the location and type may be an artifact of the objective selected for the regression. Examples range from agronomy to zoology.

Key words. Asymptotic regression, parameter identification, Verhulst logistic curve.

AMS subject classifications. 41A27, 41A30, 65K10, 80A30, 92C45, 92D25

0. Introduction.

The present article determines the existence, or the absence, of "best" translated exponential curves (1), also called Mitscherlich's laws, in the sense of weighted or correlated least squares, also called "asymptotic regression" [27], [31], [36]. The problem consists in fitting to data-points $(t_1, q_1), \ldots, (t_N, q_N)$ the parameters A, B, C of equation (1):

$$q = A \cdot e^{C \cdot t} + B. \tag{1}$$

If $A > 0$, $B > 0$, and $q > 0$, then setting $K := 1/B$ and $y := 1/q$ gives a Verhulst curve (2):

$$y = \frac{K}{1 + K \cdot A \cdot e^{C \cdot t}}. \tag{2}$$

If $A \neq 0 \neq K$, then equation (2) may be called a generalized Verhulst equation. Applications of generalized Verhulst's logistic equation (2) and Mitscherlich's law (1) include

- agronomy, to evaluate and optimize the efficiency of fertilizers [31], [36].
- chemical kinetics [14, p. 20, eq. (29)], [20, p. 993], [24, p. 65], [25], [26], [30, p. 393, eq. (1)], to identify the type of chemical reaction occuring in experiments;

*Department of Mathematics, Eastern Washington University, 216 Kingston Hall, Cheney, Washington 99004-2418, USA (ynievergelt@ewu.edu)

- population growth [2, p. 92], [6], [10, p. 191], [16], [19], [22], [28], [33], [42], [43];
- physiology, to model the growth over time of plants and animals [34] or individual organs, for example, wing span [13, p. 808];
- rheology [7].

Hence arises the need to estimate the parameters A, C, B or K that fit the data best in the sense of generalized (ordinary, weighted, or correlated) least squares [41]. Despite so many applications, except for special cases with a few $(N \leq 7)$ equally spaced abscissae $t_1, \ldots, t_N$ [31], [36], [42], [43], no theorems about the existence of such optimal curves seem known.

The problem addressed here is that generalized regressions based on equation (1) or (2) need not have any solution, or numerical iterations with finite precision computations may then let any of the parameters diverge to infinity, stopping at some value near $+\infty$ or $-\infty$ merely for the lack of additional precision, or computed solution yields negative values for theoretically positive parameters:

> A successful fit results in positive values for $[K]$ and $[-C]$ but data which deviate too far from the model result in either a fit to a curve with a negative $[K]$ or negative $[-C]$ [22, p. 94, § 2].

The present analysis reveals that such situations reflect features of the modelled phenomenon. Replacing equation (1) or (2) by the model ordinary differential equation

$$\frac{dy}{dt} = \frac{C}{K} \cdot y \cdot (K - y), \tag{3}$$

L. J. Reed and J. Berkson [33, p. 765] had already pointed out that the curve becomes a hyperbola if $A \cdot B < 0$ and C/K remains constant while K tends to 0, though they did not mention the possibility that the solution becomes a Malthusian [23, II.7] exponential growth curve as K diverges to $+\infty$. In either case C/K need not remain constant but need only converge to a limit. A similar situation occurs as A tends to 0, or $D = \ln(A)$ diverges to $-\infty$. Neither did Reed and Berkson [33] mention why or how parameters would diverge, converge to anything, or change in the first place. Indeed, they selected the type of curve to be fitted *before* starting the regression, which never failed for their examples [33, p. 769–779]. The present analysis will reveal that the limits happen to lie exactly at a removable singularity of the objective function for the generalized least-squares regression.

To this end, the strategy adopted here consists of fixing one parameter, C for Mitscherlich's equation (1), and then in analyzing the explicit formulae for the solutions of the generalized *linear* least-squares regression with the other two parameters A and B in terms of the fixed parameter C.

1. Notation for Generalized Least-Squares. All the regressions considered here reduce in part to generalized least squares to estimate two parameters, leaving only one parameter for estimation by non-linear least squares. Though the numerical solution of least-squares problems proceeds more accurately with other algorithms [3], [11], [21], the present analysis and resolution of singularities uses explicit formulae for the solutions. Moreover, examples from section 5 show that the data may be highly correlated, in the sense that in the covariance matrix entries are larger off than on the diagonal. Therefore the present analysis allows for a generalized regression.

The situation just outlined arises in fitting a least-squares line to data (x_1, y_1), ..., (x_N, y_N) in the plane as described by [12, p. 35–42 & 112–114]. Each point (x_k, y_k) may be the average of repeated measurements $(x_k, y_{k,1})$ ..., $(x_k, y_{k,M})$, with the same absissa x_k but different ordinates $y_{k,1}$, ..., $y_{k,M}$ with mean $y_k = (y_{k,1} + \cdots + y_{k,M})/M$. Consequently, the measurements $y_{k,1}$, ..., $y_{k,M}$ are modeled as values of a random variable Y_k with expectation $\mathcal{E}Y_k$. The random variables $Y_1, \ldots, Y_N$ may be correlated with a covariance matrix V defined by

$$V_{k,\ell} = \mathcal{E}[(Y_k - \mathcal{E}Y_k) \cdot (Y_\ell - \mathcal{E}Y_\ell)].$$

The matrix V is symmetric and non-negative [8, § 22.3, p. 296]. A linear model for such data consists of N linear equations $Y_k = x_k \cdot \alpha + \beta + \epsilon_k$, where each ϵ_k is a random variable with $\mathcal{E}\epsilon_k = 0$. With the column vectors $\vec{\theta} = (\alpha, \beta)^T$, $\vec{Y} = (Y_1, \ldots, Y_N)^T$, $\vec{x} = (x_1, \ldots, x_N)^T$, $\vec{0} = (0, \ldots, 0)^T$, $\vec{1} = (1, \ldots, 1)^T$, $\vec{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)^T$, and the *design* matrix $X = (\vec{x}, \vec{1})$, the model becomes

$$\vec{Y} = X \cdot \vec{\theta} + \vec{\epsilon}.$$

The problem of least squares is then to find the vector of parameters $\vec{\theta}$ that minimizes the variance of $\vec{Y}$. A linear transformation into uncorrelated variables simplifies matters. To this end, if V is invertible, then let W be *any* matrix such that $W^T \cdot W = V^{-1}$, so that $V = W^{-1} \cdot (W^{-1})^T$, and define

$$\vec{Z} = W \cdot \vec{Y}.$$

Under the hypothesis that $\mathcal{E}\epsilon = \vec{0}$, the random vector $\vec{Z}$ has expectation

$$\mathcal{E}\vec{Z} = W \cdot \mathcal{E}\vec{Y} = W \cdot X \cdot \vec{\theta}$$

and covariance matrix $W \cdot V \cdot W^T = I$ [18, p. 410, ex. 12.4]. With the Euclidean norm $\|\vec{p}\|$ defined by $\|\vec{p}\|^2 = \vec{p}^T \cdot \vec{p}$, the best linear unbiased estimate of the vector of parameters $\vec{\theta} = (\alpha, \beta)^T$ is the solution $\vec{\theta}_* = (\alpha_*, \beta_*)^T$ minimizing the objective [32, § 3.3, p. 36–37], [39, p. 144]

$$\mathcal{L}(\vec{\theta}) := \frac{1}{2} \cdot \|W \cdot X \cdot \vec{\theta} - W \cdot \vec{y}\|^2 = \frac{1}{2} \cdot (X \cdot \vec{\theta} - \vec{y})^T \cdot W^T \cdot W \cdot (X \cdot \vec{\theta} - \vec{y}). \quad (4)$$

The single factor $1/2$ avoids numerous factors 2 in the partial derivatives.

The literature formulates the generalized least-squares solution $\vec{\theta}_* = (\alpha_*, \beta_*)^T$ from the normal equations in terms of $(X^T W^T W X)^{-1}$ or with pseudo-inverses [1], [3, p. 7], [12, p. 108–116], [21, p. 37], [32, p. 47–50], [37, p. 5], [38, p. 246], but here a different notation is convenient [5]. To simplify matters, assume V invertible, so that $W^T \cdot W = V^{-1}$ is positive definite, which defines an inner product

$$\langle \vec{p}, \vec{q} \rangle := \vec{p}^T \cdot W^T \cdot W \cdot \vec{q}. \quad (5)$$

The inner product (5) reduces generalized least squares to ordinary least squares: for all sequences $\vec{p}, \vec{q} \in \mathbb{R}^N$, re-define their mean $\text{mean}_V(\vec{p})$, variance $\text{var}_V(\vec{p})$, covariance $\text{cov}_V(\vec{p}, \vec{q})$, and correlation coefficient $\text{cor}_V(\vec{p}, \vec{q})$ by

$$\text{mean}_V(\vec{p}) := \frac{\langle \vec{p}, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle} := \frac{\vec{p}^T \cdot W^T \cdot W \cdot \vec{1}}{\vec{1}^T \cdot W^T \cdot W \cdot \vec{1}}, \quad (6)$$

$$\text{cov}_V(\vec{p}, \vec{q}) := \frac{\langle \vec{p} - \text{mean}_V(\vec{p}) \cdot \vec{1}, \vec{q} - \text{mean}_V(\vec{q}) \cdot \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle}, \quad (7)$$

$$\text{var}_V(\vec{p}) := \text{cov}_V(\vec{p}, \vec{p}), \quad (8)$$

$$\text{cor}_V(\vec{p}, \vec{q}) := \frac{\text{cov}_V(\vec{p}, \vec{q})}{\sqrt{\text{var}_V(\vec{p}) \cdot \text{var}_V(\vec{q})}}. \quad (9)$$

DEFINITION 1. *For each finite squence $\vec{p}$, the **centered** sequence is defined by $\check{p} := \vec{p} - \text{mean}_V(\vec{p}) \cdot \vec{1}$, so that $\check{p}_k := p_k - \text{mean}_V(\vec{p})$ for each k.*

In particular, the following algebraic identities will lead to further simplifications:

$$\text{mean}_V(\vec{1}) := \frac{\langle \vec{1}, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle} = 1, \quad (10)$$

$$\begin{aligned} \text{mean}_V(\check{p}) &= \frac{\langle \vec{p} - \text{mean}_V(\vec{p}) \cdot \vec{1}, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle} = \frac{\langle \vec{p}, \vec{1} \rangle - \text{mean}_V(\vec{p}) \cdot \langle \vec{1}, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle} \\ &= \text{mean}_V(\vec{p}) - \text{mean}_V(\vec{p}) = 0. \end{aligned} \quad (11)$$

2. Solutions for Generalized Least-Squares. With the Euclidean inner product replaced by the inner product (5) defined by V^{-1}, the same calculations as for ordinary least squares show that the line of generalized (correlated) least squares, which minimizes the objective (4), passes through the correlated mean $(\text{mean}_V(x), \text{mean}_V(y))$ with slope $\alpha_* = \text{cov}_V(\vec{x}, \vec{y})/\text{var}_V(\vec{x})$. To this end, the constant column $\vec{1}$ in the design matrix $X = (\vec{1}, \vec{x})$ allows for simplifications. Subtracting and adding the means in the objective function (4) shows that the fitted line passes through the correlated mean of the data, defined by formula (6):

$$\left(\vec{x}, \vec{1}\right) \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \vec{y}$$

$$= \left(\vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1}, \vec{1} - \text{mean}_V(\vec{1}) \cdot \vec{1}\right) \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \left(\vec{y} - \text{mean}_V(\vec{y}) \cdot \vec{1}\right) \tag{12}$$

$$+ \left[\left(\text{mean}_V(\vec{x}), \text{mean}_V(\vec{1})\right) \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \text{mean}_V(\vec{y})\right] \cdot \vec{1}. \tag{13}$$

Yet $\vec{1} - \text{mean}_V(\vec{1}) \cdot \vec{1} = \vec{0}$ by equation (10), with $\langle \vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1}, \vec{1} \rangle = 0$ and $\langle \vec{y} - \text{mean}_V(\vec{y}) \cdot \vec{1}, \vec{1} \rangle = 0$ by equation (11). Hence the two summands (12) and (13) are mutually perpendicular relative to the inner product (5), whence

$$\mathcal{L}(\alpha, \beta)$$

$$= \frac{1}{2} \cdot \left\| W \cdot \left[\left(\vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1}, \vec{0}\right) \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \left(\vec{y} - \text{mean}_V(\vec{y}) \cdot \vec{1}\right)\right] \right\|^2 \tag{14}$$

$$+ \frac{1}{2} \cdot \left\| W \cdot \vec{1} \left[\left(\text{mean}_V(\vec{x}), \text{mean}_V(\vec{1})\right) \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \text{mean}_V(\vec{y})\right] \right\|^2 \tag{15}$$

by the Pythagorean Theorem. Moreover, regardless of α, the summand (14) does not depend on β because of its factor $\vec{0}$, while the summand (15) vanishes for

$$\beta_* := \text{mean}_V(\vec{y}) - \alpha \cdot \text{mean}_V(\vec{x}). \tag{16}$$

Consequently, equation (16) must hold at a minimum of the objective function (14)–(15), which may thus be reformulated as [8, § 21.6]

$$\mathcal{L}(\alpha, \gamma_*) := \frac{1}{2} \cdot \left\| W \cdot \left[\left(\vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1}\right) \cdot \alpha - \left(\vec{y} - \gamma_* \cdot \vec{1}\right)\right] \right\|^2, \tag{17}$$

where equations (6) and (16) give $\gamma_* = \text{mean}_V(\vec{y})$. For the objective function in the form (17), the "normal" equation for α becomes

$$0 = \frac{\partial \mathcal{L}}{\partial \alpha}(\alpha, \gamma_*) = \left\langle \left(\vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1}\right) \cdot \alpha - \left(\vec{y} - \gamma_* \cdot \vec{1}\right), \vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1} \right\rangle. \tag{18}$$

By hypothesis $\vec{x}$ is not a multiple of $\vec{1}$, so that $\text{var}_V(\vec{x}) > 0$. Solving the normal equation (18) for α, substituting $\gamma_* = \text{mean}_V(\vec{y})$, and using $\langle \vec{1}, \vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1} \rangle = 0$ in formula (20), by equation (11), gives the following solution, with all formulae gathered here at the same place:

$$\gamma_* = \text{mean}_V(\vec{y}) = \frac{\langle \vec{y}, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle} = \left\langle \vec{y}, \frac{\vec{1}}{\langle \vec{1}, \vec{1} \rangle} \right\rangle; \tag{19}$$

$$\alpha_* = \frac{\langle \vec{y} - \text{mean}_V(\vec{y}) \cdot \vec{1}, \vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1}, \rangle}{\langle \vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1}, \vec{x} - \text{mean}_V(\vec{x}) \cdot \vec{1}, \rangle} = \frac{\text{cov}_V(\vec{x}, \vec{y})}{\text{var}_V(\vec{x})}; \tag{20}$$

$$\beta_* = \text{mean}_V(\vec{y}) - \alpha_* \cdot \text{mean}_V(\vec{x}); \tag{21}$$

$$y = \alpha_* \cdot (x - \text{mean}_V(\vec{x})) + \text{mean}_V(\vec{y}). \tag{22}$$

$$\mathcal{L}(\alpha_*, \gamma_*) = \frac{1}{2} \cdot \langle \vec{1}, \vec{1} \rangle \cdot \text{var}_V(\vec{y}) \cdot \left[1 - \text{cor}_V(\vec{x}, \vec{y})^2\right]. \tag{23}$$

For *uncorrelated* data, theorem 2 confirms that the line of weighted least squares has a positive slope for increasing data, and a negative slope for decreasing data.

THEOREM 2. *If W is* diagonal *and invertible, and if the finite sequences $\vec{x}$ and $\vec{y}$ are monotonic, either both non-decreasing, or both non-increasing, then* $\text{cov}_V(\vec{x}, \vec{y}) \geq 0$.

If $\vec{x}$ or $\vec{y}$ is non-decreasing while the other is non-increasing, then $\text{cov}_V(\vec{x}, \vec{y}) \leq 0$.

Equality holds if and only if at least one of $\vec{x}$ or $\vec{y}$ is constant.

Proof. If $\vec{x}$ and $\vec{y}$ are both monotonic non-decreasing, then so are their centered sequences $\check{x}$ and $\check{y}$. In particular, $\check{y}_k - \check{y}_1 \geq 0$ for every k. Let $m := \min\{k \in \{1, \dots, N\} : \check{x}_k \geq 0\}$. With a regular diagonal matrix $W = \text{diag}(w_1, \dots, w_N)$, equation (11) leads to

$$\begin{aligned}
\langle \check{x}, \check{y} \rangle &= \sum_{k<m} w_k^2 \cdot \check{x}_k \cdot (\check{y}_k - \check{y}_1) + \sum_{k \geq m} w_k^2 \cdot \check{x}_k \cdot (\check{y}_k - \check{y}_1) \\
&\geq \sum_{k<m} w_k^2 \cdot \check{x}_k \cdot (\check{y}_m - \check{y}_1) + \sum_{k \geq m} w_k^2 \cdot \check{x}_k \cdot (\check{y}_m - \check{y}_1) \\
&= (\check{y}_m - \check{y}_1) \cdot \text{mean}_V(\check{x}) = 0,
\end{aligned}$$

with equality if and only if $\vec{x}$ or $\vec{y}$ is constant. If $\vec{x}$ and $\vec{y}$ are monotonic non-increasing, then the same proof applies to $-\vec{x}$ and $-\vec{y}$. If $\vec{x}$ or $\vec{y}$ is non-decreasing while the other is non-increasing, then the proof applies to $\vec{x}$ and $-\vec{y}$. □

Theorem 2 shows that if the data increase and W is diagonal, then the line of weighted least squares has a positive slope. By continuity, the slope remains positive after sufficiently small perturbations of W, which may yield a non-diagonal matrix. Diagonality is a sufficient but not necessary condition for the line of generalized least squares to have a positive slope.

Some non-linear least-squares problems arise as parametric sets of linear least-squares problems, where the design matrix $X(\vec{S})$ and the observations $\vec{q}(\vec{S})$ depend on an array $\vec{S}$ of one or more parameters. Then the objective function (**??**) becomes

$$\mathcal{L}(\vec{S},\vec{P}) := \frac{1}{2}\cdot\left\|W\cdot\left[X(\vec{S})\cdot\vec{P}-\vec{q}(\vec{S})\right]\right\|_2^2. \tag{24}$$

If there exist parameters $\vec{S}_*$ and $\vec{P}_*$ where $\mathcal{L}(\vec{S}_*,\vec{P}_*)=0$, so that all the data points satisfy the system

$$W\cdot\left(X(\vec{S})\cdot\vec{P}-\vec{q}(\vec{S})\right)=\vec{0}, \tag{25}$$

then $\mathcal{L}$ has a global minimum there, because $\mathcal{L}\geq 0$ everywhere, whence also

$$\vec{0}=\frac{\partial\,\mathcal{L}}{\partial\vec{S}}(\vec{S}_*,\vec{P}_*)=\left\langle X(\vec{S})\cdot\vec{P}-\vec{q}(\vec{S}),\frac{\partial X}{\partial\vec{S}}(\vec{S})\cdot\vec{P}-\frac{\partial\vec{q}}{\partial\vec{S}}(\vec{S})\right\rangle, \tag{26}$$

whence substituting equation (26) into the Hessian matrix annihilates one summand and leaves

$$\frac{\partial^2\,\mathcal{L}}{\partial\vec{S}^2}(\vec{S}_*,\vec{P}_*)=\left\langle\frac{\partial X}{\partial\vec{S}}(\vec{S})\cdot\vec{P}-\frac{\partial\vec{q}}{\partial\vec{S}}(\vec{S}),\frac{\partial X}{\partial\vec{S}}(\vec{S})\cdot\vec{P}-\frac{\partial\vec{q}}{\partial\vec{S}}(\vec{S})\right\rangle. \tag{27}$$

Thus if

$$\frac{\partial}{\partial\vec{S}}\left[X(\vec{S})\cdot\vec{P}-\vec{q}(\vec{S})\right]\neq 0, \tag{28}$$

then the Hessian of $\mathcal{L}$ relative to $\vec{S}$ is symmetric positive definite. By transversality, it follows that after sufficiently small perturbations of the data $\mathcal{L}$ still has a unique global minimum, which lies at some point $(\vec{S},\vec{P})$ in a neighborhood of $(\vec{S}_*,\vec{P}_*)$.

3. Asymptotic Regression. This section focuses on fitting to data the parameters of Mitscherlich's law (1), with equation $q=A\pm B\cdot e^{C\cdot t}$, including the limiting cases where C either tends to 0 or diverges to $\pm\infty$ with "exponential" non-linear regression.

The main results are that the objective function has a continuous extension and hence a global minimum on the extended real line $[-\infty,+\infty]$ with the topology of the two-point compactification, but that its removable singularity at 0 differs from the shortest least-squares solution.

To avoid repetition, this section sets up some common notation for generalized least squares [1], subject to the following hypotheses:

HYPOTHESES 3. With $\mathbb{R}$ denoting the set of all real numbers, the present theory and applications pertain to a weight matrix $W \in \mathbb{R}^{N\times N}$ and data points (t_k, q_k) for $k \in \{1,\dots,N\}$ such that

(3.1) W is non-singular,

(3.2) $q_k > 0$ for every $k \in \{1,\dots,N\}$;

(3.3) there exist $k, \ell \in \{1,\dots,N\}$, with $t_k \neq t_\ell$; in particular, $N \geq 2$;

(3.4) for all $k, \ell \in \{1,\dots,N\}$, if $k < \ell$, then $t_k \leq t_\ell$, so that $t_1 \leq \cdots \leq t_N$.

In other words, the matrix $W^T W$ is symmetric positive definite, all the ordinates (measured data) are strictly positive, and there are at least two different abscissae (times, temperatures, etc.); the last hypothesis (3.4) that all the abscissae (times, temperatures, etc.) are in non-decreasing order will serve only to shorten the notation.

3.1. Asymptotic Regression by Exponential Non-Linear Least Squares. A common method of fitting Mitscherlich's law (1) to data proceeds by ordinary (unweighted) non-linear least-squares regression [31, p. 499], [27, p. 328], [36, p. 250], minimizing

$$F(A,B,C) := \frac{1}{2} \cdot \sum_{k=1}^{N} \left[B + A \cdot e^{C\cdot t_k} - q_k\right]^2. \tag{29}$$

Yet although its Hessian matrix has been derived in the literature [36, p. 250, eq. (2.3)], the objective function F has a singularity, hitherto seemingly unnoticed, along the planes $B = 0$ or $C = 0$, where its Hessian matrix is singular. Also, the details presented below (in remark 4) reveal that the shortest least-squares solution does not remove this singularity. To remove the singularity, define the change of variable

$$x_k^C := e^{C\cdot t_k}. \tag{30}$$

Then for each C the non-linear regression amounts to a linear regression for q_k versus x_k^C. Several cases can arise, according to whether $C \neq 0$, $\lim_{C\to 0}$, $\lim_{C\to+\infty}$ or $\lim_{C\to-\infty}$.

3.1.1. Finite Non-Zero Growth Rate. If $C \neq 0$, then $x_k^C \neq x_\ell^C$ for all k and ℓ, whence the variance of x^C is positive, by the hypothesis (3.3) that $N \geq 2$. Hence the results from section 1 with $X = (\vec{x}^C, \vec{1})$ and $\vec{P} = (A,B)^T$ in the objective function (**??**) yield the minimizing constant and slope in the form

$$z = Z_C + A_C \cdot \left(x^C - \overline{\vec{x}^C}\right), \tag{31}$$

$$Z_C = \overline{\vec{q}}, \tag{32}$$

$$A_C = \frac{\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{x}^C - \overline{\vec{x}^C}\vec{1}, \rangle}{\langle \vec{x}^C - \overline{\vec{x}^C}\vec{1}, \vec{x}^C - \overline{\vec{x}^C}\vec{1}, \rangle} = \frac{\text{cov}_V(\vec{x}^C, \vec{q})}{\text{var}_V(\vec{x}^C)}, \tag{33}$$

$$B_C = Z_C - A_C \cdot \overline{\vec{x}^C}. \tag{34}$$

Substituting these results back into the objective function (**??**) gives an objective function F of the single variable C:

$$F(C) := \mathcal{L}(B_C, A_C) = \frac{1}{2} \cdot \left\| W \cdot \left[\overline{\vec{q}} \cdot \vec{1} + A_C \cdot \left(\vec{x}^C - \overline{\vec{x}^C} \cdot \vec{1} \right) \right] \right\|_2^2. \tag{35}$$

Upper and lower bounds for a global minimum C_* of F have not been derived yet, so that the search for C_* has not yet been automated. Nevertheless, in section 5, plots of F will illustrate the phenomena that occur at the singularities.

Closed-form solutions to $F'(C) = 0$ seem unlikely: already for the special case of $N = 4$ equally spaced abscissae, the equation $F'(C) = 0$ is a polynomial of degree 6 in $z = e^C$ [31, p. 499].

3.1.2. Growth Rate Near Zero. If $C = 0$, then $x_k^C = 1 = x_\ell^C$ for all k and ℓ, whence the variance of x^C vanishes. For C near 0 but $C \neq 0$, however, section 1 applies; with $o(h)$ denoting any term such that $\lim_{h \to 0} o(h)/h = 0$, formula (30) leads to

$$x^C := e^{C \cdot t} = 1 + C \cdot t + o(C), \tag{36}$$

$$x_k^C = e^{C \cdot t_k} = 1 + C \cdot t_k + o(C), \tag{37}$$

$$x_k^C - \overline{x^C} = C \cdot (t_k - \overline{\vec{t}}) + o(C) \tag{38}$$

Substituting these results into formula (33) gives $\lim_{C \to 0} A_C = 0$, but dividing top and bottom in by C equation (39) shows that $\lim_{C \to 0} A_C \cdot (x^C - \overline{x^C}) = Q_* \cdot (t - \overline{\vec{t}})$:

$$\lim_{C \to 0} A_C \cdot (x^C - \overline{x^C}) = \lim_{C \to 0} \frac{\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{x}^C - \overline{\vec{x}^C}\vec{1}, \rangle}{\langle \vec{x}^C - \overline{\vec{x}^C}\vec{1}, \vec{x}^C - \overline{\vec{x}^C}\vec{1}, \rangle} \cdot (x^C - \overline{x^C}) \tag{39}$$

$$= \frac{\text{cov}_V(\vec{t}, \vec{q})}{\text{var}_V(\vec{t})} \cdot (t - \overline{\vec{t}}) \tag{40}$$

$$= Q_* \cdot (t - \overline{\vec{t}}). \tag{41}$$

Thus formulae (32) and (41) show that the fitted equation (31) converges to the line of ordinary least squares for z versus t:

$$\lim_{C\to 0} z = \bar{\vec{q}} + Q_* \cdot (t - \bar{\vec{t}}) \tag{42}$$

Also, substituting the same limit (41) back into formula (35) for the objective shows that

$$\lim_{C\to 0} F(C) = \lim_{C\to 0} \frac{1}{2} \cdot \left\| W \cdot \left[A_C \cdot \left(\vec{x}^C - \overline{\vec{x}^C} \right) - \left(\vec{q} - \bar{\vec{q}} \cdot \vec{1} \right) \right] \right\|_2^2 \tag{43}$$

$$= \frac{1}{2} \cdot \left\| W \cdot \left[Q_* \cdot \left(\vec{t} - \bar{\vec{t}} \right) - \left(\vec{q} - \bar{\vec{q}} \cdot \vec{1} \right) \right] \right\|_2^2 \tag{44}$$

$$= \frac{\langle \vec{1}, \vec{1} \rangle}{2} \cdot \mathrm{var}_V(\vec{q}) \cdot \left[1 - \mathrm{cor}_V(\vec{t}, \vec{q})^2 \right], \tag{45}$$

by equation (23). Consequently, F has an analytic extension, also denoted by F, across $C = 0$; hence F is continuous on $\mathbb{R}$.

Thus the regression automatically fits the line

$$\frac{1}{y} = -(Q_* + P_* \cdot t), \tag{46}$$

or, equivalently, the hyperbola

$$y = \frac{1}{-(Q_* + P_* \cdot t)}. \tag{47}$$

In the case with $Q_* = 0 \neq P_*$, this hyperbola also includes the horizontal straight line $y = -1/P_*$, which also corresponds to $B = 0$ in $y = M/(1 + B \cdot e^{C \cdot t})$.

The same limits hold with $u_{M,k} = \ln(K \cdot q_k - 1)$ replaced by $u_{M,k} = f(M, q_k)$ such that $f(M, q_k) = K \cdot q_k + o(K)$, or, more generally, $e^{D + C \cdot t_k} - u_{M,k} = g(D, C, M, t_k, q_k) = (D + C \cdot t_k) - K \cdot q_k + o(K) + o(D + C \cdot t_k)$.

REMARK 4. The limit just obtained differs from the limit of the shortest (smallest norm) least-squares solutions. For instance, with un-weighted least squares, $W = I$, $\langle \vec{1}, \vec{1} \rangle = N$, and with † denoting the Moore-Penrose pseudo-inverse, define

$$X_C := \left(\vec{1}, \vec{x}^C \right), \tag{48}$$

$$X_0 := \lim_{C\to 0} X_C = \left(\vec{1}, \vec{1} \right), \tag{49}$$

$$X_0^\dagger = \frac{1}{2N} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot (1, \ldots, 1), \tag{50}$$

$$\begin{pmatrix} A \\ B \end{pmatrix} = X_0^\dagger \vec{q} = \frac{\bar{q}}{2N} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \neq \begin{pmatrix} P_* \\ Q_* \end{pmatrix}. \tag{51}$$

In other words, although for each $C \neq 0$ the least-squares solution is unique and hence coincides with the shortest least-squares solution, so that $\vec{P}_* = X_C^\dagger \vec{q}$, substituting the singular matrix $X_0 = (\vec{1}, \vec{1})$ into the initial objective function (**??**) and computing the shortest least-squares solution $\vec{P}^\dagger := X_0^\dagger \vec{q}$, does *not* produce a continuous solution. Indeed, as the smallest singular value of X_C tends to 0, the pseudo-inverse $X_C^\dagger$ becomes unbounded and consequently need not be a continuous function of the entries of X_C [21, p. 44].

3.1.3. Growth Rate Diverging to Infinity. For C diverging to $+\infty$, factoring out the largest term x_N^C in formulae (30) and (34) gives the limits for $1 \leq k \leq N-1$

$$\lim_{C \nearrow +\infty} A_C \cdot x_k^C = 0 \qquad\qquad k < N. \tag{52}$$

For $k = N$, define the temporary abbreviation

$$L_N := \lim_{C \nearrow +\infty} \frac{\overline{\vec{x}^C}}{x_N^C} = \lim_{C \nearrow +\infty} \frac{1}{x_N^C} \cdot \frac{\langle \vec{x}^C, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle} = \frac{\langle \vec{e}_N, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle}. \tag{53}$$

Hence the limit (53) combined with equation (11) leads to

$$\lim_{C \nearrow +\infty} A_C \cdot x_N^C = \lim_{C \nearrow +\infty} \frac{\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{x}^C - \overline{\vec{x}^C}\vec{1} \rangle}{\langle \vec{x}^C - \overline{\vec{x}^C}\vec{1}, \vec{x}^C - \overline{\vec{x}^C}\vec{1} \rangle} \cdot x_N^C \tag{54}$$

$$= \frac{\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{e}_N - L_N\vec{1} \rangle}{\langle \vec{e}_N - L_N\vec{1}, \vec{e}_N - L_N\vec{1} \rangle} \tag{55}$$

$$= \frac{\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{e}_N \rangle}{\langle \vec{e}_N - L_N\vec{1}, \vec{e}_N - L_N\vec{1} \rangle}, \tag{56}$$

$$\lim_{C \nearrow +\infty} A_C \cdot \overline{\vec{x}^C} = \lim_{C \nearrow +\infty} \frac{\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{x}^C - \overline{\vec{x}^C}\vec{1} \rangle}{\langle \vec{x}^C - \overline{\vec{x}^C}\vec{1}, \vec{x}^C - \overline{\vec{x}^C}\vec{1} \rangle} \cdot \frac{\langle \vec{x}^C, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle} \tag{57}$$

$$= \frac{\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{e}_N \rangle}{\langle \vec{e}_N - L_N\vec{1}, \vec{e}_N - L_N\vec{1} \rangle} \cdot \frac{\langle \vec{e}_N, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle}, \tag{58}$$

$$\lim_{C \nearrow +\infty} A_C \cdot (x_N^C - \overline{\vec{x}^C}) = \frac{\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{e}_N \rangle}{\langle \vec{e}_N - L_N\vec{1}, \vec{e}_N - L_N\vec{1} \rangle} \cdot \frac{\langle \vec{1} - \vec{e}_N, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle}. \tag{59}$$

Therefore the following limit exists:

$$\lim_{C \nearrow +\infty} F(C) = \lim_{C \nearrow +\infty} \frac{1}{2} \cdot \left\| W \cdot \left[A_C \cdot \left(\vec{x}^C - \overline{\vec{x}^C}\vec{1} \right) - \left(\vec{q} - \overline{\vec{q}}\vec{1} \right) \right] \right\|_2^2. \tag{60}$$

Similar results hold as C diverges to $-\infty$ with (t_1, q_1) instead of (t_N, q_N).

Hence F has a continuous extention, also denoted by F, to the extended real line $[-\infty, +\infty]$ with the topology of the two-point compactification. Consequently, F has a global minimum.

For $t < t_N$,

$$\lim_{C \nearrow +\infty} z = \lim_{C \nearrow +\infty} \overline{\vec{q}} + A_C \cdot \left(e^{C \cdot t} - \overline{\vec{x}^C} \right) \tag{61}$$

$$= \overline{\vec{q}} + 0 - \frac{\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{e}_N \rangle}{\langle \vec{e}_N - L_N \vec{1}, \vec{e}_N - L_N \vec{1} \rangle} \cdot \frac{\langle \vec{e}_N, \vec{1} \rangle}{\langle \vec{1}, \vec{1} \rangle}, \tag{62}$$

by equation (58). Thus the fitted curve is a horizontal straight half-line.

Yet for $t > t_N$,

$$\lim_{C \nearrow +\infty} z = \lim_{C \nearrow +\infty} \overline{\vec{q}} + A_C \cdot \overline{\vec{x}^C} \cdot \left(e^{C \cdot t} / \overline{\vec{x}^C} - 1 \right) \tag{63}$$

exists if and only if $\lim_{C \nearrow +\infty} A_C \cdot \overline{\vec{x}^C} = 0$, which occurs if and only if $\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{e}_N \rangle = 0$ by equation (58). In the particular case where $W^T W$ is diagonal and, $q_1 \leq \cdots \leq q_N$, then $\langle \vec{q} - \overline{\vec{q}}\vec{1}, \vec{e}_N \rangle = q_N - \overline{\vec{q}} = 0$, so that all the data points lie on the same straight line with equation $q = \overline{\vec{q}}$, or on a Heavyside step function.

4. Sufficient Criteria for the Existence of Specific Best-Fitting Curves. This section establishes sufficient criteria on the data for the best-fitting Mitscherlich curve in the sense of uncorrelated weighted least-squares to have parameters B, B, and C with specific signs.

THEOREM 5. *With weighted but uncorrelated least-squares regression, for all finite sequences of data* $t_1 < \cdots < t_N$ *and* $q_1 > \cdots > q_N > 0$*, if every triple of data points decreases super-linearly, then* $B_* \geq 0$ *and* $C_* < 0$*. If every triple of data points decreases super-exponentially, then* $B_* > 0$ *and* $C_* < 0$*.*

Proof. If $C < 0$, then the transformed data increase, because $x_1^C > \cdots > x_N^C > 0$ and $q_1 > \cdots > q_N > 0$; consequently the fitted line has a positive slope $B_C > 0$.

The proof proceeds to exclude $C > 0$. If $C > 0$, then the transformed data decrease, because $0 < x_1^C < \cdots < x_N^C$ and $q_1 > \cdots > q_N > 0$. Consequently the fitted line has a negative slope $B_C < 0$, so that the fitted Mitscherlich curve $M_C(t) = B_C \cdot e^{C \cdot t} + B_C$ is decreasing and concave. If $C = 0$, then the initial data decrease, and the fitted line $M_0(t) = B_0 \cdot t + B_0$ is decreasing and weakly concave. A half line is also weakly concave.

If $q_1 \leq M_C(t_1)$, then let $k := 1$. If $q_1 > M_C(t_1)$, then denote by k the last index such that $(t_1, q_1), \ldots, (t_k, q_k)$ lie on or above the fitted curve:

$$k := \max\bigl\{ j \in \{1, \ldots, N\} : \ q_1 \geq M_C(t_1), \ldots, q_j \geq M_C(t_j) \bigr\}.$$

Thus $q_k \geq M_C(t_k)$ but $M_C(t_{k+1}) > q_{k+1}$. Consequently, there exists $\tau_k \in [t_k, t_{k+1}]$ where $M_C(\tau_k) = \varrho_k$ intersects the line segment between (t_k, q_k) and (t_{k+1}, q_{k+1}).

If $q_N \leq M_C(t_N)$, then let $\ell := N$. If $q_N > M_C(t_N)$, then denote by ℓ the first index such that $(t_\ell, q_\ell), \dots, (t_N, q_N)$ lie on or above the fitted curve:

$$\ell := \min\bigl\{j \in \{1, \dots, N\} : \ q_j \geq M_C(t_j), \dots, q_N \geq M_C(t_N)\bigr\}.$$

Thus $q_{\ell-1} < M_C(t_{\ell-1})$ but $M_C(t_\ell) \leq q_\ell$. Consequently, there exists $\tau_\ell \in [t_{\ell-1}, t_\ell]$ where $M_C(\tau_\ell) = \varrho_\ell$ intersects the line segment between (t_ℓ, q_ℓ) and $(t_{\ell-1}, q_{\ell-1})$.

If $q_1 \leq M_C(t_1)$ and $q_N \leq M_C(t_N)$, then the line $\text{line}_{(t_1,q_1),(t_N,q_N)}(t) = A \cdot t + B$ is decreasing, with $B < 0$ and $C = 0$.

Because the data decrease super-linearly, the piecewise linear interpolant is convex while M_C is concave and they intersect each other at (τ_k, ϱ_k) and $(\tau_\ell, \varrho_\ell)$:

If $k < j < \ell$, then $q_j < \text{line}_{(\tau_k,\varrho_k),(\tau_\ell,\varrho_\ell)}(t_j) < M_C(t_j)$.

If $j \notin [k, \ell]$, then $q_j > \text{line}_{(\tau_k,\varrho_k),(\tau_\ell,\varrho_\ell)}(t_j) \geq M_C(t_j)$.

By uniform convergence on compacta of Mitscherlich curves to line, the same inequalities hold for exponential curves with parameters B, B, and C in a neighborhood of the line $\text{line}_{(\tau_k,\varrho_k),(\tau_\ell,\varrho_\ell)}$ for which $C = 0$, in particular, for some exponential curves with $B > 0 > C$.

Therefore M_C is not a best-fitting curve.

By the existence of a best-fitting curve, it follows that a best-fitting curve has parameters $B_* > 0$ and $C_* < 0$. □

THEOREM 6. *With weighted but uncorrelated least-squares regression, for all finite sequences of data* $t_1 < \cdots < t_N$ *and* $q_1 > \cdots > q_N > 0$*, if every triple of data points decreases super-exponentially, then* $B_* > 0$*,* $B_* > 0$*, and* $C_* < 0$*.*

Proof. Theorem 5 already shows that. $B_C > 0$ and $C < 0$. The proof proceeds to exclude $B \leq 0$.

If $C = 0$, then the initial data decrease, and the fitted line $M_0(t) = B_0 \cdot t + B_0$ is decreasing and weakly concave. A half line is also weakly concave.

If $q_1 \leq M_C(t_1)$, then let $k := 1$. If $q_1 > M_C(t_1)$, then denote by k the last index such that $(t_1, q_1), \dots, (t_k, q_k)$ lie on or above the fitted curve:

$$k := \max\bigl\{j \in \{1, \dots, N\} : \ q_1 \geq M_C(t_1), \dots, q_j \geq M_C(t_j)\bigr\}.$$

Thus $q_k \geq M_C(t_k)$ but $M_C(t_{k+1}) > q_{k+1}$. Consequently, there exists $\tau_k \in [t_k, t_{k+1}]$ where $M_C(\tau_k) = \varrho_k$ intersects the exponential curve between (t_k, q_k) and (t_{k+1}, q_{k+1}).

If $q_N \leq M_C(t_N)$, then let $\ell := N$. If $q_N > M_C(t_N)$, then denote by ℓ the first index such that $(t_\ell, q_\ell), \dots, (t_N, q_N)$ lie on or above the fitted curve:

$$\ell := \min\bigl\{j \in \{1, \dots, N\} :\ q_j \geq M_C(t_j), \dots, q_N \geq M_C(t_N)\bigr\}.$$

Thus $q_{\ell-1} < M_C(t_{\ell-1})$ but $M_C(t_\ell) \leq q_\ell$. Consequently, there exists $\tau_\ell \in [t_{\ell-1}, t_\ell]$ where $M_C(\tau_\ell) = \varrho_\ell$ intersects the exponential between (t_ℓ, q_ℓ) and $(t_{\ell-1}, q_{\ell-1})$.

If $q_1 \leq M_C(t_1)$ and $q_N \leq M_C(t_N)$, then the exponential $\exp_{k,\ell}(t) = B \cdot t + B$ is decreasing, with $B < 0$ and $C = 0$.

Because the data decrease super-exponentially, the piecewise exponential interpolant is logarithmically convex while M_C is logarithmically concave:

$$\begin{aligned} W(t) &:= \ln\left(B \cdot e^{C \cdot t} + B\right), \\ W'(t) &= C \cdot \left(1 - \frac{B}{B \cdot e^{C \cdot t} + B}\right), \\ W''(t) &= \frac{B \cdot B \cdot C^2 \cdot e^{C \cdot t}}{\left(B \cdot e^{C \cdot t} + B\right)^2} \leq 0 \end{aligned}$$

because $B > 0 \geq B$, and they intersect each other at (τ_k, ϱ_k) and $(\tau_\ell, \varrho_\ell)$:

If $k < j < \ell$, then $q_j < \exp_{(\tau_k,\varrho_k),(\tau_\ell,\varrho_\ell)}(t_j) < M_C(t_j)$, from the same inequalities with their $\ln$.

If $j \notin [k, \ell]$, then $q_j > \exp_{(\tau_k,\varrho_k),(\tau_\ell,\varrho_\ell)}(t_j) \geq M_C(t_j)$, from the same inequalities with their $\ln$.

Therefore M_C is not a best-fitting curve.

By the existence of a best-fitting curve, it follows that a best-fitting curve has parameters $B_* > 0$ and $C_* < 0$. □

REMARK 7. Similar results hold for positive data $(t_1, p_1), \dots, (t_N, p_N)$ *increasing sub-linearly or sub-exponentially*: for any $M > \max_j p_j$, setting $q := M - p$ produces transformed positive data that decrease super-linearly or super-exponentially. Transforming the fitted Mitscherlich equation $q = A \cdot e^{C \cdot t} + B$ with $A > 0 > C$ and $B > 0$ back to $p = M - q$ gives $p = -A \cdot e^{C \cdot t} + (M - B)$. with $-A < 0$, $C < 0$, and $M - B > 0$, otherwise the fitted curve would lie in the fourth quadrant while all the data lie in the upper half plane.

5. Case Studies. The case studies presented here demonstrate how the resolution of singularities just established also resolves the difficulties encountered by previous authors.

5.1. Mitscherlich Curves Fitted by Ordinary and Weighted Least Squares. This subsection demonstrates Mitscherlich curves fitted to correlated data by ordinary and weighted least squares.

EXAMPLE 8. F. Pimentel Gomes fitted Mitscherlich's shifted exponential curves to the data included in table 5.1 [31, p. 510]. The sums of squares, weights, and Shapiro-Wilk's W were computed separately and are not in [31, p. 510].

TABLE 5.1
Yield of Irish red potato crops with various levels of triple superphosphate fertilizer [31, p. 510].

	TRIPLE SUPERPHOSPHATE [LBS/ACRE]				
	0	40	80	120	160
	199.2	292.3	387.2	491.4	387.2
1945 YIELD	210.3	387.2	418.8	504.5	629.2
[65 LBS/ACRE]	253.1	396.6	508.3	446.8	523.1
	268.0	400.2	508.2	523.1	506.5
MEANS	232.650	369.075	455.625	491.450	511.500
SUMS OF SQUARES	3286.2500	7949.3075	11576.8475	3165.6500	29463.3400
WEIGHTS*	0.351259	0.145211	0.099710	0.364641	0.039178
SHAPIRO-WILK'S W	0.902	0.721	0.833	0.943	0.970
	103.0	193.5	215.5	205.5	217.0
1946 YIELD	110.5	188.5	227.5	234.5	243.0
[65 LBS/ACRE]	103.5	194.0	201.0	206.5	227.5
	102.0	178.5	203.0	224.0	237.0
MEANS	104.750	188.625	211.750	217.625	231.125
SUMS OF SQUARES	45.2500	155.1875	454.2500	596.1875	388.1875
WEIGHTS	0.631447	0.184119	0.062901	0.047926	0.073606
SHAPIRO-WILK'S W	0.776	0.851	0.908	0.877	0.976

*Reciprocals of Sums of Squares normalized to add up to 1.

Based on the methods of section 3, figure 1 displays a plot of the objective functions for the weighted and unweighted mean squared error, which are not everywhere convex, but show a minimum near $C = -0.015$. Table 5.2 displays the fitted parameters. As a check, the parameters fitted by ordinary (unweighted) least squares agree to three significant digits with Pimentel Gomes's values [31, p. 511]. The parameters fitted by weighted least squares agree to two significant digits with the previous values, but with a smaller weighted sum of squared errors. Figure 2 shows the fitted curves with Pimentel Gomes's data. Because the data are positive and increase sub-linearly, $A < 0$ and $C < 0$ but $B > 0$, conforming to remark 7. The 1946 data lead to similar results, which are omitted.

The selection of ordinary and weighted least squares, as opposed to any other objective,

TABLE 5.2
Mitscherlich parameters fitted by least squares to Pimentel Gomes's 1945 data [31, p. 510].

SOURCE	A	B	C	WSSE*
[31, p. 511]	-307.5	539.1	-0.0155	14.0325
Equal weights $1/5$	-307.608	539.162	-0.01551	11.224
Weights from table 5.1	-310.893	542.215	-0.01498	6.011745

*Weighted Sum of Squared Errors.

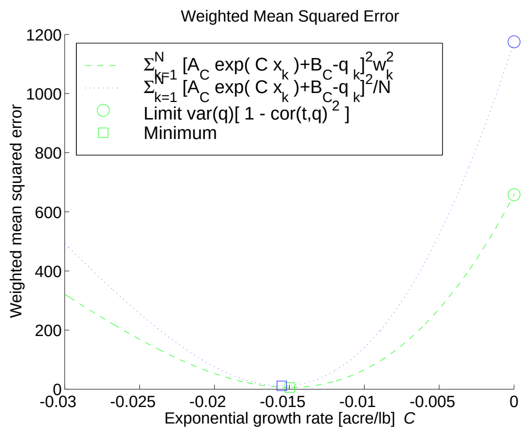

FIG. 1. *Mean squared error for F. Pimentel Gomes's 1945 data [31, p. 510] with weighted* $(--)$ *and unweighted* $(\cdots)$ *least squares.*

in this example serves as a comparison with Pimentel Gomes's original results [31, p. 510–112] but does not imply any Gaussian normality in the distribution of errors. Whereas the Shapiro-Wilk's W statistics for normality [9, § 9.3.4.1, p. 393–395], [35] for the last columns in each year of data exceeds 0.970, which is above the 50% level, for the second column in the 1945 data W has the value 0.721, which is below the 05% level, and hence cautions against using any hypothesis of normality in the data.

Also, the covariance matrix is far from diagonal, which suggests the use of generalized least-squares, [39, § 2.5, p. 137–145]. Yet exact rational arithmetic with *Mathematica*™ reveals that for each year the covariance matrix is exactly singular, and hence not invertible.

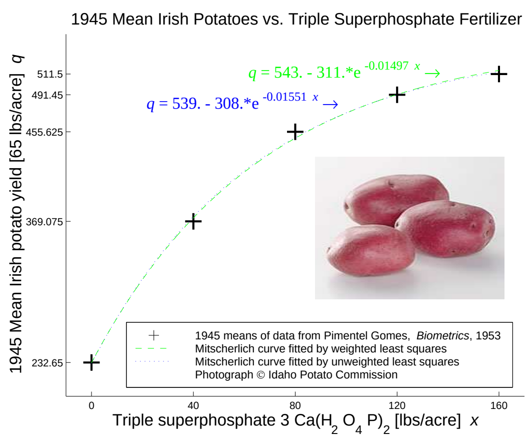

FIG. 2. *Mitscherlich's shifted exponential fitted to F. Pimentel Gomes's 1945 data* $(+)$ *[31, p. 510] by weighted* $(--)$ *and unweighted* $(\cdots)$ *least squares.*

5.2. Curves Other than Verhulst Curves. This subsection presents examples that challenge the belief that all data must conform to a Verhulst curve.

EXAMPLE 9. Joseph Pearson measured the age, diameters, and weight of *Placuna placenta* [29]. For diameter (y) vs. weight (x), Pearson fitted the hyperbola $y = 98 + 90 \cdot (x - 45)/(x + 125)$ [29, p. 16; p. 19, Fig. 3]. However, the physical relations between length and weight with volume suggest that volume varies as the cube of the length. Indeed, figure 3 shows that a cube-root fits diameters vs. weight better than a hyperbola does.

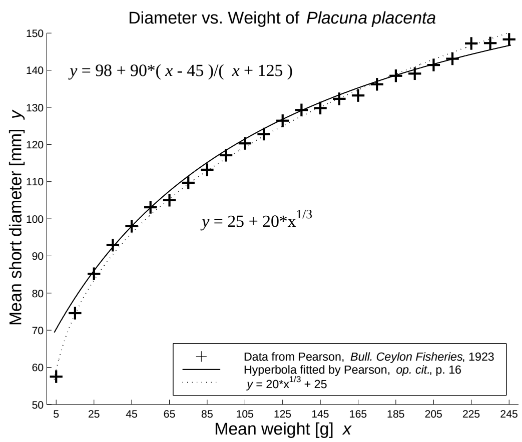

FIG. 3. *Hyperbola* $(-)$ *fitted by Pearson to data* $(+)$ *[29, p. 16].*

EXAMPLE 10. Without providing a formula, Pearson fitted a curve to diameter (y) vs. age (t) [29, p. 17, Fig. 1], adapted by D'Arcy Wentworth Thompson [40, p. 166, Fig. 36]. Fitting a rectangular hyperbola by an algebraic method similar to Gander, Golub, and Strebel's [15] gives the equation $y = 251.38 - 12\,025/(x + 50.37)$ and a plot similar to Pearson's and Thompson's, as shown here in figure 4. Pearson does not provide any reason for a hyperbola other than it "is a close approximation to the actual results" [29, p. 16]. In contrast, Thompson states a heuristic argument to fit a Verhulst curve to the same data:

> The window-pane oyster in Ceylon (*Placuna placenta*) has been kept under observation for eight years, during which it grows from two inches long to six (Fig. 36). The young grow quickly, and slow down asymptotically towards the end; an S-shaped beginning to the growth-curve has not been seen, but would probably be found in the growth of the first year. Changes of shape as growth goes on are hard to see in this and other shells; rather is it characteristic of them to keep their shape from first to last unchanged. — D'Arcy Wentworth Thompson [40, p. 166]

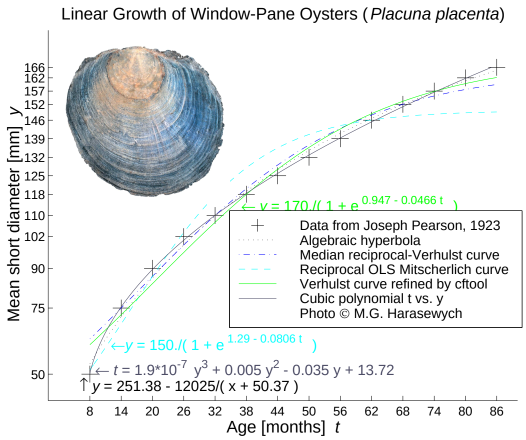

FIG. 4. *Curves fitted to data $(+)$ by Joseph Pearson [29, p. 16]. Photograph © M. G. Harasewych [17] used by permission.*

Calculations confirm that all triples of data increase sub-exponentially, so that their reciprocals decrease super-exponentially. The Mitscherlich curve fitted to the reciprocal data by ordinary (unweighted) least squares has parameters $A = 0.0242 > 0$, $B = 0.00668$, and $C = -0.0806$, with signs conforming to theorem 6. The reciprocal of this Mitscherlich curve is a Verhulst curve with equation $y = 150./(1 + e^{1.29-0.0806 \cdot t})$. Figure 4 reveals that the hyperbola fits the data better than any Verhulst curve found so far, and, again, as indicated in example 9, that the inverse function of a cubic polynomial fits the data better yet.

REMARK 11. Bradley [4] argues by dimensional analysis that the ratio $R := (K - y)/y = (q - B)/B = q/B - 1$ is a quantity more natural than y. However, large positive values of y have small positive reciprocals q, with yet smaller differences of positive reciprocals, so that fitting curves to reciprocal data amounts to assigning smaller weights to larger data. Indeed, figure 4 illustrates how the reciprocal of the Mitscherlich curve fitted to the reciprocal data passes closer to the smaller values of the data but farther away from the larger values.

6. Acknowledgments. This work was supported in part by a Faculty Research and Creative Works Grant from Eastern Washington University.

REFERENCES

[1] A. C. Aitken. On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh, Section A*, 55:42–47, 1935.

[2] Linda J. S. Allen. *An Introduction to Mathematical Biology*. Pearson/Prentice Hall, Upper Saddle River, NJ, 2007. ISBN 0-13-035216-0; QH323.5.A436 2007; LCCC No. 2006042585; 570.1'5118–dc22.

[3] Åke Björck. *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1996. ISBN 0-89871-360-9; QA214.B56 1996; LCCC No. 96-3908; 512.9'42–dc20.

[4] David Bradley. Verhulst's logistic curve. *The College Mathematics Journal*, 32(2):94–98, March 2001.

[5] A. Buse. Goodness of fit in generalized least squares estimation. *The American Statistician,*, 27(3):106–108, June 1973.

[6] Fabio Cavallini. Fitting a logistic curve to data. *College Mathematics Journal*, 24(3):247–253, May 1993.

[7] Wen M. Chow. Parameter estimation for simple nonlinear models. *Communications of the ACM*, 2(7):28–29, July 1959.

[8] Harald Cramér. *Mathematical Methods of Statistics*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1999. ISBN 0-691-000547-8; QA276.C72 1999; 99-12156; 519.5–dc21; first edition: Almqvist & Wiksells, Uppsala, Sweden, 1945.

[9] Ralph B. D'Agostino. Tests for the normal distribution. In Ralph B. D'Agostino and Michael A. Stephens, editors, *Goodness-of-Fit Techniques*, volume 68 of *Statistics: Textbooks and Monographs*, pages 367–419, New York, NY, 1986. Marcel Dekker. ISBN 0-8247-7487-6; QA277.G645 1986; 519.5'6; LCCC No. 86-4571.

[10] Gerda de Vries, Thomas Hillen, Mark Lewis, Johannes Müller, and Birgitt Schönfisch. *A Course in Mathematical Biology*, volume 12 of *Mathematical Modeling and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2006. ISBN 0-89871-612-8; QH323.5.C69 2006; 570.1'5118–dc22; LCCC No. 2006044305.

[11] J. E. Dennis, Jr. and Robert B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, volume 16 of *Classics In Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1996. ISBN 0-89871-364-1; QA402.5.D44 1996; 95–51776.

[12] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley, New York, NY, 1966. QA276.D68; 66-17641.

[13] V. Elangovan, H. Raghuram, E. Yuvana Satya Priya, and G. Marimuthu. Wing morphology and flight performance in *Rousettus leschenaulti*. *Journal of Mammalogy*, 85(4):806–812, August 2004.

[14] Arthur A. Frost and Ralph G. Pearson. *Kinetics and Mechanism: A Study of Homogeneous Chemical Reactions*. Wiley, New York, NY, second edition, 1961. LCCC No. 61-6773.

[15] Walter Gander, Gene H. Golub, and Rolf Strebel. Least-squares fitting of circles and ellipses. *BIT*, 34:558–578, 1994.

[16] G. F. Gause. *The Struggle for Existence*. Dover, Mineola, NY, 2003. QH375.G35 2003; 576.8'2–dc21; LCCC No 2002037091. First published by Hafner in 1969.

[17] M. G. Harasewych and Fabio Moretzsohn. *The Book of Shells: A Life-Size Guide to Identifying and Classifying Six Hundred Seashells*. University of Chicago Press, Chicago and London, 2010. ISBN 0226315770.

[18] Robert V. Hogg and Allen T. Craig. *Introduction to Mathematical Statistics*. Macmillan, New York, NY, fourth edition, 1978. ISBN 0-02-355710-9; QA276.H59 1978; 77-2884; 519.

[19] H. P. Kuang. Forecasting future enrollment by curve-fitting techniques. *The Journal of Experimental Educa-*

tion, 23(3):271–274, March 1955.

[20] M. Kunitz and John H. Northrop. Isolation from beef pancreas of crystalline trypsinogen, trypsin, a trypsin inhibitor, and an inhibitor-trypsin compound. *Journal of General Physiology*, 19(6):901–1007, 20 June 1936.

[21] Charles L. Lawson and Richard J. Hanson. *Solving Least Squares Problems*, volume 15 of *Classics In Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1995. 0-89871-356-0; QA275.L38 1995; 95-35178; 511'.42–dc20.

[22] Donald Leach. Re-evaluation of the logistic growth curve for human populations. *Journal of the Royal Statistical Society. Series A. General*, 144(1):94–103, 1981.

[23] Thomas Robert. Malthus. *An Essay on the Principle of Population, as it affects the Future Improvement of Society, etc.* J. Johnson, London, UK, 1798. http://www.econlib.org/library/Malthus/malPop.html.

[24] Margaret R. McDonald and M. Kunitz. The effect of calcium and other ions on the autocatalytic formation of trypsin from trypsinogen. *Journal of General Physiology*, 25(1):53–73, 20 September 1941.

[25] Wilh. Oswald. Studien zur chemischen Dynamik; erste Abhandlung: Die Einwirkung der Säuren auf Acetamid. *Journal für Praktische Chemie*, 27(1):1–39, 11 January 1883.

[26] Wilh. Oswald. Studien zur chemischen Dynamik. zweite Abhandlung: Die Einwirkung der Säuren auf Methylacetat. *Journal für Praktische Chemie*, 28(1):449–495, 30 June 1883.

[27] H. D. Patterson. A simple method for fitting an asymptotic regression curve. *Biometrics*, 12(3):323–329, September 1956.

[28] Raymond Pearl and Lowell J. Reed. On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Sciences*, 6(6):199–272, 15 June 1920.

[29] Joseph Pearson. Statistics dealing with the growth-rate, &c., of *Placuna placenta*. *Bulletin of the Ceylon Fisheries*, I(1–3):13–64, 17 January 1923.

[30] Jean-François Pechère and Hans Neurath. Studies on the autocatalytic activation of trypsinogen. *Journal of Biological Chemistry*, 229(1):389–407, 1 November 1957.

[31] Frederico Pimentel Gomes. The use of Mitscherlich's regression law in the analysis of experiments with fertilizers. *Biometrics*, 9(4):498–516, December 1953.

[32] R. L. Plackett. *Principles of Regression Analysis*. Oxford University Press, London, UK, 1960. QA276.P5.

[33] Lowell J. Reed and Joseph Berkson. The application of the logistic function to experimental data. *The Journal of Physical Chemistry*, 33(5):760–779, January 1929.

[34] Robert E. Ricklefs. A graphical method of fitting equations to growth curves. *Ecology*, 48(6):978–983, November 1967.

[35] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, December 1965.

[36] W. L. Stevens. Asymptotic regression. *Biometrics*, 7(3):247–267, September 1951.

[37] G. W. Stewart. Errors in variables for numerical analysts. In Sabine Van Huffel, editor, *Recent Advances in Total Least Squares and Errors-In-Variables Models: Proceedings of the Second International Workshop on Total Least Squares and Errors-in-Variables Modeling, Leuven, Belgium, August 21–24, 1996*, pages 3–10, Philadelphia, PA, 1997. Society for Industrial and Applied Mathematics. ISBN 0-89871-393-5; 96-72486.

[38] Josef Stoer and Roland Bulirsch. *Introduction to Numerical Analysis*. Springer-Verlag, New York, NY, third edition, 2002. ISBN 0-387-95452-X; QA297.S8213 2002; 2002-019729; 519.4–dc21.

[39] Gilbert Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, MA, 1986. ISBN 0-9614088-0-4; QA37.2.S87 1986; 84-52450; 510.

[40] D'Arcy Wentworth Thompson. *On Growth and Form*. Cambridge University Press, Cambridge, UK, new edition, 1942.

[41] Malcolm E. Turner, Robert J. Monroe, and Henry L. Lucas, Jr. Generalized asymptotic regression and non-linear path analysis. *Biometrics*, 17(1):120–143, March 1961.

[42] Pierre-François Verhulst. Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathématique et Physique*, 10:113–121, 1838.

[43] Pierre-François Verhulst. Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux Memoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18(1):1–45, 1845.