Real and Generic Data without Unconstrained Best-Fitting Verhulst Curves and Sufficient Conditions for Median Mitscherlich and Verhulst Curves to Exist

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.

1. 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” [24, 25]. 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

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

called a Verhulst curve to distinguish it from a logistic curve with $K = 1$. Indeed, the difficulty lies in fitting the carrying capacity $K > 0$ [35, p. 58], [36, p. 697].

Example 1. Figure 1 shows Anderson & Anderson’s data [3, p. 92, Table 1] on the weight vs. age of cactus wrens (*Campylorhynchus brunneicapillus*), and a Verhulst growth curve fitted graphically by Ricklefs by eyeballing the asymptote $K$ [46].

Applications include the identification of parameters of population growth [1], [11], [14], [15], [16], [50] and autocatalysis in chemistry and biology [13, pp. 19–20], [20], [21], [22], [30], [40], [41]. 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 [12], [17].

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 [40]. Approximate formulae for a few equally spaced points followed [49, §3, p. 252], [43, 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 [15, p. 397], [16, p. 149], [17, §II, pp. 4–5], [44]. 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 [29].

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 [41, p. 399, Table 2]. Nevertheless, Examples 27, 29, and 30 in Section 6 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 [9], [35, 36, 37] and cited [33], [42], [51]. Such computations without theoretical support have recently been used for many applications [9], [23], including the U.S. pop-
ulation [7, pp. 13–14], [29], and are used without further ado in recent textbooks [11, 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 [16, 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 $\frac{dy}{dt} = r \cdot y \cdot [1 - (y/K)]$ is also a particular case of the Bernoulli differential equation $\frac{dy}{dt} = r \cdot y \cdot [1 - (y/K)^g]$ [11, 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 function fail. For example, maximum likelihood methods cannot be justified, because “the maximizer of the likelihood function is the vector of parameter estimates” [19, 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 [10, §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 $C$ of curves, such as the class of all straight lines, or the class of all Verhulst curves. For each curve $C \in C$, 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 $C \in C$ 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

\[
F_D(C) := F_D(\tilde{D}).
\]

For each curve \( C \), the map \( D \mapsto F_D(C) \) 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 37 in Section 8 shows that a variety of ordinary and orthogonal regression methods share this feature. Also, for each data sequence \( D \), the map \( C \mapsto F_D(C) \) 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 2 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 3 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 \( \xi_k = (t_k, \eta_k) \), after Verhulst’s transformation \( \eta = \ln[y/(K - y)] \). Gause [16, pp. 149–150] and Gause & Alpatov [17] always pick some value \( K > y_{\text{max}} := \max\{y_1, \ldots, y_N\} > 0 \) for their demonstrations. Yet in several experiments Gause selects a value \( K < y_{\text{max}} \) [16, 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 [18, p. 39], [45, p. 384]. The goal then consists of fitting to the reciprocal data \( \tilde{z}_j = (t_j, q_j) \) the reciprocal of a Verhulst curve, which, after the change of parameters \( B := 1/K \) and \( A := e^r/K \), is a Mitscherlich curve \( M(t) := A \cdot e^{-rt} + B \). To this end, Section 4 reviews the concepts of median points and lines. In Section 5, Theorem 25 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 26 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 6 applies the theory to real examples with real data.

2. 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 4. Also, some median circle passes through two or three data points [34]. Similarly, 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.

2.1. 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 $\zeta_1$ and $\zeta_2$. Three points $\zeta_1, \zeta_2,$ and $\zeta := (t_3, \eta_3)$ 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 $\mathcal{X} \subset (\mathbb{R}^2)^3$ and $\mathcal{Y} \subset (\mathbb{R}^2)^3$ 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:

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

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

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

**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 $\zeta_j := L(z_j) = (t_j, \ln y_j)$ 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

March 2012] MITSCHERLICH AND VERHULST CURVES 215
\( z_2 \): the image by \( E \) of the line through \( \zeta_1 \) and \( \zeta_2 \), with

\[
s = \frac{\ln(y_2) - \ln(y_1)}{t_2 - t_1}, \quad (3)
\]

\[
c = \ln(y_1) - s \cdot t_1. \quad (4)
\]

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 \( t \mapsto e^{t+s+t} \) 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 \( y_2 > \exp_{z_1,z_2}(t_2) \).

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

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

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

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

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

**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 \( L \).

### 2.2. 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.** A Mitscherlich law \([39],[43],[49]\) has the form

\[
M(t) := A \cdot e^{-r-t} + B, \quad (5)
\]

where \( A, B, r \in \mathbb{R} \).

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
\((t_1, q_1^\prime)\) and \((t_2, q_2^\prime)\). Hence formulae (3) and (4) show that \(A, r > 0\) with
\[
 r = r(B; \hat{z}_1, \hat{z}_2) = -\frac{s}{t_2 - t_1} \cdot \ln \left( \frac{q_1 - B}{q_2 - B} \right), \tag{6}
\]
\[
 A = A(B; \hat{z}_1, \hat{z}_2) = e^r = (q_1 - B) \cdot e^{r_1}. \tag{7}
\]
If \(B > q_1 > q_2\), then the shift \(q^\dagger := B - q\) gives \(r < 0\) also defined by formula (6), with \(A\) also defined by formula (7).

If \(q_1 > B > q_2\), then \(A \cdot e^{-r + r_2} = q_2 - B < 0 < q_1 - B = A \cdot e^{-r_2}\) would imply \(\text{A} < 0 < \text{B}\). If \(B \in [q_1, q_2]\), then \(A = 0\), which would imply \(q_1 = B = q_2\).

**Definition 11.** For all points \(\hat{z}_1 = (t_1, q_1^\prime)\) and \(\hat{z}_2 = (t_2, q_2^\prime)\) such that \(t_1 < t_2\) but \(q_1 > q_2\), and \(q_2 > B_1 > B_2\), then \(M_{B; \hat{z}_1, \hat{z}_2}\) be the Mitscherlich curve through \(\hat{z}_1\) and \(\hat{z}_2\) with constant \(B\).

If also \(0 < B < q_2\), then \(V_{K; \hat{z}_1, \hat{z}_2} := 1/M_{B; \hat{z}_1, \hat{z}_2}\) 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 at two common points.

**Lemma 12.** Suppose that \(\hat{z}_1 = (t_1, q_1^\prime)\) and \(\hat{z}_2 = (t_2, q_2^\prime)\), where \(t_1 < t_2\) but \(q_1 > q_2\), and \(q_2 > B_1 > B_2\). Then \(M_{B_1; \hat{z}_1, \hat{z}_2}(t) < M_{B_2; \hat{z}_1, \hat{z}_2}(t)\) for every \(t\) such that \(t_1 < t < t_2\), while \(M_{B_1; \hat{z}_1, \hat{z}_2}(t) > M_{B_2; \hat{z}_1, \hat{z}_2}(t)\) for every \(t \notin [t_1, t_2]\).

**Proof.** To simplify notation, for any \(B\) such that \(q_2 > B \geq B_2\) we write \(M_B\) for \(M_{B; \hat{z}_1, \hat{z}_2}\). Differentiating the definition (5) shows that \(M_B\) 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
\[
 W_B(t) := \ln(M_B(t) - B_2). \tag{9}
\]
Substituting formula (8) into the derivatives of the transformation (9) gives
\[
 W'_B = \frac{M'_B}{M_B - B_2} = r \cdot \frac{B - M_B}{M_B - B_2} = r \cdot \left( \frac{B - B_2}{M_B - B_2} - 1 \right), \tag{10}
\]
\[
 W''_B = r \cdot (B - B_2) \cdot \frac{-M'_B}{(M_B - B_2)^2} = r^2 \cdot (B - B_2) \cdot \frac{M_B - B}{(M_B - B_2)^2}. \tag{11}
\]
By Lemma 10 we have \(A, r > 0\), whence \(M'_B < 0\) by equation (8), and hence by equation (11) \(W'_B\) has the same sign as \(B - B_2\). Taking \(B = B_2\) we see that \(W''_B = 0\). On the other hand, taking \(B = B_1 > B_2\) we see that \(W''_B > 0\). Thus \(W_B\) is a line while \(W_{B_1}\) is convex, and both pass through \(\hat{z}_1 = (t_1, \ln(q_1 - B_2))\) and \(\hat{z}_2 = (t_2, \ln(q_2 - B_2))\). Consequently, \(W_{B_1}(t) < W_{B_2}(t)\) for \(t_1 < t < t_2\), whereas \(W_{B_1}(t) > W_{B_2}(t)\) for \(t \notin [t_1, t_2]\). Finally, we note 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 line $\hat{z}_1, \hat{z}_2$ uniformly on each compact subset of the real line.

Also, $\lim_{B \not\to q_2^-} M_{B;\hat{z}_1,\hat{z}_2}(t) = q_2$ for $t > t_1$, but diverges to $\infty$ for $t < t_1$. Moreover, if $0 < q_2 < q_1$, then $\lim_{K \to \infty} V_{K;\hat{z}_1,\hat{z}_2}(t) = \exp_{\hat{z}_1,\hat{z}_2}(t)$.

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

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

$$= q_1 \cdot e^{-r(t-t_1)} - B \cdot r \cdot \frac{e^{-r(t-t_1)} - 1}{r}.$$  \hfill (12)

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

$$\lim_{B \to \pm\infty} r(B; \hat{z}_1, \hat{z}_2) = \lim_{B \to \pm\infty} \frac{1}{t_2 - t_1} \cdot \ln\left(\frac{q_1 - B}{q_2 - B}\right) = 0.$$  \hfill (14)

With the abbreviation $r(B) := r(B; \hat{z}_1, \hat{z}_2)$, l’Hospital’s rule gives

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

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

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

Also, $M_{B;\hat{z}_1,\hat{z}_2}(t_1) = q_1$, while $\lim_{B \not\to q_2^-} r(B; \hat{z}_1, \hat{z}_2) = \infty$ by formula (6), whence $\lim_{B \not\to q_2^-} M_{B;\hat{z}_1,\hat{z}_2}(t) = q_2$ for $t > t_1$ and diverges to $\infty$ for $t < t_1$ by formula (12). Moreover, if $0 < q_2 < q_1$, then reciprocals yield $\lim_{K \to \infty} V_{K;\hat{z}_1,\hat{z}_2}(t) = 1/M_{0;\hat{z}_1,\hat{z}_2}(t) = \exp_{\hat{z}_1,\hat{z}_2}(t)$. In all cases the convergence is monotonic by Lemma 12 and hence uniform on compacta by Dini’s theorem [47, p. 162].

Theorem 14. For all points $\hat{z}_j = (t_j, q_j)$ with $t_1 < t_2 < t_3$ and $q_1 > q_2 > q_3 > \text{line}_{\hat{z}_1,\hat{z}_2}(t_3)$ there is a unique Mitscherlich curve through $\hat{z}_1, \hat{z}_2,$ and $\hat{z}_3$ 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 as $B < 0$, $B = 0$, 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 $B \not\in [q_2, q_1]$, 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, $\lim_{B \not\to \infty} M_{B;\hat{z}_1,\hat{z}_2}(t_3) = \text{line}_{\hat{z}_1,\hat{z}_2}(t_3) < q_3$ and $\lim_{B \not\to q_2^-} M_{B;\hat{z}_1,\hat{z}_2}(t_3) = q_2 > q_3$. 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.

218 © THE MATHEMATICAL ASSOCIATION OF AMERICA  [Monthly 119
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 $M_{\hat{z}_1, \hat{z}_2}(t_3)$ increases as $B$ increases. Hence the triple $(\hat{z}_1, \hat{z}_2, \hat{z}_3)$ is respectively sub-exponential or super-exponential according as $B < 0$ 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 $z_1 = (t_1, y_1), z_2 = (t_1, y_1)$, and $z_3 = (t_1, y_1)$ are on a Verhulst curve $V(t) = K/(e^{\alpha - t} + 1)$, 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 $A = e^\alpha/K > 0$. 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 $V_{\hat{z}_1, \hat{z}_2, \hat{z}_3}$.

3. 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 triple 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 $F_D(S) < F_D(V)$.

**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 nonzero residual $z_j - \tilde{z}_j \neq 0$, whence $F_D(V) = F_D(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3) > 0$ 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 $X \cap U$ is relatively open, and $F_D : X \to \mathbb{R}_+$ is relatively open by hypothesis. Hence $F_D$ maps $X \cap U$ 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 $(w_1, w_2, w_3) \in X \cap U$ 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, $F_D(S) = \min\{F_D(p_1, p_2, p_3) : (\forall j)(p_j \in S)\}$ by Remark 3. Therefore, $0 < F_D(S) < F_D(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3) = F_D(V)$.

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 5.
4. MEDIAN POINTS AND LINES. This section reviews the concept of medi-
ans without using linear programming, showing that at least one median line passes
through at least two data points.

**Definition 19.** A real number \( r_o \) is a median of real numbers \( r_1, \ldots, r_N \), which are
also called data points, if \( r_o \) minimizes the objective

\[
f_o(r) := \sum_{j=1}^{N} |r - r_j|,
\]

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

\[
N^-(r) := \sum_{r_k < r} 1, \quad N^0(r) := \sum_{r_j = r} 1, \quad N^+(r) := \sum_{r_l > r} 1.
\]

**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
\( f_o(r_n) \leq f_o(r) \). If \( r \notin \{r_1, \ldots, r_N\} \), then \( f_o \) is differentiable at \( r \), so that

\[
f_o(r) = \sum_{r_k < r} (r - r_k) + \sum_{r_l > r} (r_l - r),
\]

\[
f'_o(r) = \sum_{r_k < r} 1 - \sum_{r_l > r} 1 = N^-(r) - N^+(r).
\]

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

If \( f'_o(r) = 0 \), then \( N^-(r) = N^+(r) = N/2 \), whence there is some \( n \) such that
\( r_n < r < r_{n+1} \); in that case \( f'_o(u) = 0 \) for every \( u \) such that \( r_n < u < r_{n+1} \), so that
\( f_o(r_n) = f_o(r) \).

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

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

\[
f_o(A, B) := \sum_{j=1}^{N} |A \cdot x_j + B - y_j|.
\]

Thus \( f_o(A_o, B_o) = \min_{(A,B) \in \mathbb{R}^2} f_o(A, B) \).

**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 $B_1 := y_1 - A \cdot x_1, \ldots, B_N := y_N - A \cdot x_N$. Shifting a line with slope $A$ vertically to the intercept $B_m$ shows that $f_o(A, B_m) \leq f_o(A, B)$ 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 to find a median line it suffices to search among all lines passing through at least one data point (with all slopes).

For each data point $z_m = (x_m, y_m)$ and for each slope $A \in \mathbb{R}$, the line through $z_m$ with slope $A$ has equation $y - y_m = A \cdot (x - x_m)$. For such a line, shifting the origin to $z_m$ changes the objective to $z_m$ changes the objective (17) to

\[
    f_o(A) = \sum_{y_k - y_m < A \cdot (x_k - x_m)} [A \cdot (x_k - x_m) - (y_k - y_m)] \\
    + \sum_{y_k - y_m > A \cdot (x_k - x_m)} [(y_k - y_m) - A \cdot (x_k - x_m)].
\]

For each $j \neq m$ define the slope $A_j := (y_j - y_m)/(x_j - x_m)$. These not necessarily distinct slopes cut the real line into finitely many disjoint open intervals, with a singleton $\{A_j\}$ at each endpoint, in each of which $f'_o$ is constant:

\[
    f'_o(A) = \sum_{A_k < A} |x_k - x_m| - \sum_{A_k > A} |x_k - x_m|.
\]  

If there exists some $j \neq m$ such that $A = A_j$, then the line through $z_m$ with slope $A = A_j$ already passes through at least two distinct data points: $z_m$ and $z_j$.

In the alternative, if $A \neq A_j$ for every $j \neq m$, then the line with slope $A$ through $z_m$ passes through no other data point. Yet in this case $A$ lies in one of the open intervals where $f'_o$ is constant. If $f'_o(A) > 0$, then by equation (18) there exists at least one index $k$ such that $A_k < A$, and decreasing $A$ rotates the line clockwise and decreases the objective $f_o(A)$ until the line touches another data point $z_i$, where $A_i < A$ and $f_o(A_i) < f_o(A)$. A similar argument applies to $f'_o(A) < 0$ by rotating the line counterclockwise until it touches another data point $z_n$, where $A_n > A$ and $f_o(A_n) < f_o(A)$. If $f'_o(A) = 0$, then by equation (18) there are indices $i$ and $n$ such that $A_i < A < A_n$ and $f_o$ is constant on $[z_i, z_n]$; rotating the line in either direction, to $z_i$ or $z_n$, shows that $f_o(A_i) = f_o(A) = f_o(A_n)$.

Therefore, the minimum of $f_o$ on the finite set of lines through any two data points is a global minimum. $lacksquare$

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 $f_o(A_{m,n}, B_{m,n})$ 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 [32, p. 4]. For larger values of $N$ faster linear programming algorithms exist for median points and lines [5, p. 454], [6], [48], [52], and one can also use Korneenko and Martini’s anchored median hyperplane algorithm for orthogonal median regressions [26], [27], [28]. The foregoing discussion also holds for weighted medians, with the objective (17) replaced by $f_o(A, B) := \sum_{j=1}^N w_j \cdot |A \cdot x_j + B - y_j|$. Each positive weight $w_j$ can be adjusted to reflect the accuracy of the $j$th measurement.

5. CONDITIONS FOR MEDIAN MITSCHERLICH AND VERHULST CURVES.

With only positive data ordinates $(y_j)$, a nonpositive carrying capacity $K \leq 0$ never
gives a best-fitting Verhulst curve, since raising \( K \) to \( y_{\text{min}} := \min\{y_1, \ldots, y_N\} > 0 \) 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

\[
f_\circ(K, a, r) := \sum_{j=1}^{N} \left| \frac{1 + e^{a-r-t_j}}{K} - \frac{1}{y_j} \right|.
\]

The change of parameters \( B := 1/K \) and \( A := e^a/K \) changes the objective (19) to

\[
g_\circ(A, B, r) := \sum_{j=1}^{N} \left| B + A \cdot e^{-r \cdot t_j} - q_j \right|.
\]

If the data \( z_j = (t_j, y_j) \) increase, then the reciprocal data \( \hat{z}_j = (t_j, q_j) \) decrease. Subsection 5.1 shows that the search for a minimum of the objective (20) can focus on Mitscherlich curves \( M_{B, \hat{z}_m, \hat{z}_n} \) through two data points \( \hat{z}_m = (t_m, q_m) \) and \( \hat{z} = (t_n, q_n) \) for some \( B \notin [q_n, q_m] \). 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. In Subsection 5.2, Theorem 25 shows that the search for a minimum of the objective (20) can be further narrowed to \( B \) in a compact interval, whence by continuity and compactness the objective (20) has a global minimum. Moreover, \( A > 0, B > 0, \) and \( r > 0 \) at such a minimum. The method fixes pairs of data points and varies \( B \). Theorem 26 then points out that the reciprocal of such a Mitscherlich curve is a Verhulst curve.

Though they occur in Examples 28 and 29, monotonicity and sub-exponentiality are not the rule in practice, but no other existence theorems appear known.

**5.1. Features of median Mitscherlich curves for super-linear data.** For positive decreasing data, there is a Mitscherlich curve with a value of the objective (20) smaller than it is for any horizontal line, as verified in Lemma 23.

**Lemma 23.** For all real sequences \( t_1 < \cdots < t_N \) and \( q_1 > \cdots > q_N > 0 \) with \( N \geq 2, \) and for all coefficients \( A, B, \) and \( r = 0, \) there exist \( A', B', \) and \( r' \neq 0 \) such that \( g_\circ(A', B', r') < g_\circ(A, B, 0), \) with \( g_\circ \) defined by formula (20).

**Proof.** If \( r = 0, \) then the objective (20) has a minimum at any curve for which \( A + B \) is a median of \( q_1, \ldots, q_N, \) and by Theorem 20 there exists \( m \) such that \( A + B = q_m \) is a median. The fitted curve is thus a horizontal line with equation \( q = q_m = q_m + e^{0 \cdot (t - t_m)} \) 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. \( \blacksquare \)

Lemma 23 shows that the search for a minimum can be restricted to nonzero values of the rate \( r. \) 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 $q_1 > \cdots > q_N > 0$ with $N \geq 2$, and for each rate $r \neq 0$, the objective (20) reaches its constrained minimum, with $r$ fixed, at a Mitscherlich curve that passes through at least two data points.

Proof. For each $r \neq 0$, minimizing the objective (20) 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 minimum occurs at a line through two distinct transformed data points. \hfill \blacksquare

Lemmas 23 and 24 thus confirm that for decreasing positive data, a median Mitscherlich curve must pass through two distinct data points $\hat{z}_m = (m, q_m)$ and $\hat{z} = (l_\alpha, q_\alpha)$ with $1 \leq m < n \leq N$ and $r \neq 0$. Lemma 10 reveals that such a curve must be of the form $M_{B; \hat{z}_m, \hat{z}_n}$ for some $B \notin [q_n, q_m]$. Consequently, the search for median Mitscherlich curves can be narrowed to curves of the form $M_{B; \hat{z}_m, \hat{z}_n}$.

5.2. 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 $f_o$ in equation (19). Also, at least one such minimizing Verhulst curve passes through at least two data points. The strategy consists of proving that for the super-exponential reciprocal data, there exists a median Mitscherlich curve, minimizing the objective (20), with $A, B, r > 0$. From Subsection 5.1 it suffices to search for a median Mitscherlich curve $M_{B; \hat{z}_m, \hat{z}_n}$ through two distinct reciprocal data points $\hat{z}_m$ and $\hat{z}_n$.

Theorem 25. 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 for the reciprocal data, minimizing the objective (20), with $A > 0$, $B > 0$, and $r > 0$ for every such curve.

Proof. For all indices $m$ and $n$ such that $1 \leq m < n \leq N$, with $r(B; \hat{z}_m, \hat{z}_n)$ and $A(B; \hat{z}_m, \hat{z}_n)$ defined by formulae (6) and (7), abbreviate the objective (20) by

$$g_o(B) := g_o(A(B; \hat{z}_m, \hat{z}_n), B, r(B; \hat{z}_m, \hat{z}_n)).$$

(21)

By Theorem 14, for each reciprocal data point $\hat{z}_i$ with $1 \leq i \leq N$ but $i \notin \{m, n\}$ there is a unique Mitscherlich curve $M_{B; \hat{z}_m, \hat{z}_n}$ through $\hat{z}_m$, $\hat{z}_n$, and $\hat{z}_i$, with $0 < B_i < q_i$ because (a permutation of) the reciprocal triple $(\hat{z}_i, \hat{z}_m, \hat{z}_n)$ decreases super-exponentially. Let $B_{\text{min}}$ and $B_{\text{max}}$ be the minimum and maximum of $B_i$ for $1 \leq i \leq N$ but $i \notin \{m, n\}$. For each such index $i$, from $\hat{z}_i = M_{B; \hat{z}_m, \hat{z}_n}(t_i)$, if $B^i < B_{\text{min}} \leq B_i \leq B^\dagger < B_i < B^\dagger < q_n < q_m < B^\dagger$, then Lemmas 10 and 12 give

$$M_{B^\dagger; \hat{z}_m, \hat{z}_n}(t_i) < \text{line}_{\hat{z}_m, \hat{z}_n}(t_i) < M_{B^i; \hat{z}_m, \hat{z}_n}(t_i) < M_{\text{B}_{\text{min}}; \hat{z}_m, \hat{z}_n}(t_i) \leq \hat{z}_i \leq M_{\text{B}_{\text{max}}; \hat{z}_m, \hat{z}_n}(t_i) < M_{B^\dagger; \hat{z}_m, \hat{z}_n}(t_i)$$

for $t_i \notin [t_m, t_n]$, with reverse inequalities for $t_m < t_i < t_n$, as in Figure 2.

Thus all the vertical distances between the reciprocal data and a curve $M_{B; \hat{z}_m, \hat{z}_n}$ increase as $B$ escapes from the interval $[B_{\text{min}}, B_{\text{max}}]$, so that $g_o(B) > g_o(B_{\text{min}})$ for $B < B_{\text{min}}$ while $g_o(B) > g_o(B_{\text{max}})$ for $B > B_{\text{max}}$. Hence the search for a minimum of the objective (21) can be further narrowed to $B \in [B_{\text{min}}, B_{\text{max}}]$. Because $g_o$ is continuous on the compact interval $[B_{\text{min}}, B_{\text{max}}]$, it has a minimum at some $B_{m,n} \in [B_{\text{min}}, B_{\text{max}}]$. From $0 < B_{\text{min}} \leq B_{m,n} \leq B_{\text{max}} < q_n$, the Mitscherlich curve $M_{B_{m,n}; \hat{z}_m, \hat{z}_n}$ has parameters $r(B_{m,n}; \hat{z}_m, \hat{z}_n) > 0$ and $A(B_{m,n}; \hat{z}_m, \hat{z}_n) > 0$ by formulae (6) and (7) in Lemma 10.
Among the finitely many pairs $m$ and $n$ with $1 \leq m < n \leq N$, there is a pair $m$ and $n$ with the smallest minimum value $g_{\diamond}(B_{m,n})$, which is then a global minimum of the objective (21). Also, $B_{m,n} > 0$, $A(B_{m,n}; \hat{z}_m, \hat{z}_n) > 0$, $r(B_{m,n}; \hat{z}_m, \hat{z}_n) > 0$.

**Theorem 26.** 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 (19).

**Proof.** Theorem 25 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 $c = \ln(A/B)$.

6. **CASE STUDIES.** Regressions are used to estimate parameters, identify mechanisms, or make predictions.

6.1. **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*” [16]), 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 [16, Ch. IV].

**Example 27.** Figure 3 lists the data from the first experiment [15, p. 395, Table I, Exp. 1], [16, 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 there are no best-fitting Verhulst curves for the regressions described in Remark 3. By Theorem 14 there is a reciprocal Mitscherlich curve with $B < 0 < A$ through all three points.

**Example 28.** Figure 4 lists the data from the second experiment [15, p. 395, Table I, Exp. 2], [16, 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 4. 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.
Population Growth of *Schizosaccharomyces kefir*

\[ y = \frac{301.6}{(e^{2.142-0.02277t} - 1)} \]

**Figure 3.** The third point is above the exponential (···) but below the hyperbola (−−−) through the first two points: there are no best-fitting Verhulst curves, but a reciprocal Mitscherlich curve (−) passes through all data points. Photograph of *Schizosaccharomyces pombe* courtesy Professor Rosa M. Aligué Alemany, Universitat de Barcelona, Facultat de Medicina, Departament de Biologia Cellular, Spain.

\[ y = \frac{924.}{(1 + e^{3.37-0.0484t})} \]

**Figure 4.** The third data point lies below the exponential (···) 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 4 with Matlab’s cftool.

<table>
<thead>
<tr>
<th></th>
<th>Initial Values</th>
<th>Fitted Values*</th>
<th>Goodness of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_0$</td>
<td>$a_0$</td>
<td>$r_0$</td>
</tr>
<tr>
<td>default</td>
<td>0.907</td>
<td>0.862</td>
<td>0.593</td>
</tr>
<tr>
<td>Newton</td>
<td>924.0</td>
<td>3.37</td>
<td>0.0484</td>
</tr>
</tbody>
</table>

*Matlab 7 with its curve-fitting toolbox and default number of iterations on a Power Mac G4.
†Sum of Squares Due to Error.

6.2. 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 [13, p. 10].

Example 29. Figure 5 shows Ostwald’s data from the hydrolysis of ethyl acetate with acetic acid [38, 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 [38, 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 26, there is a median reciprocal Verhulst curve, computed here as

$$y + 1338 = \frac{2601.}{1 + e^{-0.06258 - 0.004146 t}}.$$  \hspace{1cm} (22)

This curve differs from Reed & Berkson’s [44, p. 770] for the following reason.

![Hydrolysis of Ethyl Acetate with Acetic Acid (CH$_3$COOH)](http://en.wikipedia.org/wiki/File:Acetic-acid-3D-balls.png)

Figure 5. Verhulst curves fit to data (+) from Ostwald [38, p. 481, Table XLIII] by Reed & Berkson (−), Matlab’s cftool (−−), and median reciprocal (···). The units and reactions are explained in Example 29. Public-domain picture of an acetic acid molecule by Benjah-bmm27 (http://en.wikipedia.org/wiki/File:Acetic-acid-3D-balls.png).
With water (H$_2$O) and acetic acid (C = CH$_3$COOH) as an auto-catalyst, ethyl acetate (A = CH$_3$COOCH$_2$CH$_3$) decomposes into more acetic acid and ethyl alcohol ("ethanol": E = CH$_3$CH$_2$OH) according to the stoichiometric equation

\[ A + H_2O \rightleftharpoons C + E. \]

(The “ethyl” in A, C, and E is the group C$_2$H$_5$.) With concentrations \([A]_0 = 1370\) and \([C]_0 = 1338\) at \(t = 0\), the concentration of acetic acid \([C] = y + [C]_0\) evolves by the differential equation \(d[C]/dt = r \cdot [C] \cdot ([A]_0 + [C]_0 - [C])\) [38, p. 481]. Hence the sum \(K = [A]_0 + [C]_0\) is the asymptotic least upper bound for \([C]\). With \(K = 1370 + 1338 = 2708\) fixed, Reed & Berkson fitted only the parameters \(a\) and \(r\) [44, pp. 770–771]. Yet the initial concentrations are measured as are the other data points and so are only “approximate” ("rund" and “annähernd” [38, pp. 480, 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 (22). The constant term \([C]_0 = 1338\) in equation (22) could also be estimated, but this is another story for another day.

Table 2. Fitting a Verhulst curve to the data (+) in Figure 5 with Matlab’s cftool.

<table>
<thead>
<tr>
<th>from</th>
<th>Initial Values</th>
<th>Fitted Values*</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(K_0)</td>
<td>(a_0)</td>
<td>(r_0)</td>
</tr>
<tr>
<td>default</td>
<td>0.444</td>
<td>0.478</td>
<td>0.690</td>
</tr>
<tr>
<td>Reed &amp; Berkson</td>
<td>2708</td>
<td>-0.0138</td>
<td>0.00372</td>
</tr>
</tbody>
</table>

*Matlab 7 with its curve-fitting toolbox on a Power Mac G4.
†Sum of Squares Due to Error.

6.3. An example for which a best-fitting Verhulst curve is elusive. In biology, the acid test for a model is its predictive accuracy [19, p. 1252].

Example 30. Figure 6 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 [8, 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 [8], 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 [19, p. 1252].

Accordingly, Figure 6 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 (23):

\[ Y(t) = e^{3.979 + 0.04102 \cdot (t-1970)}. \]
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 (24):

\[ Y(t) = \frac{696.8}{1 + e^{97.19 - 0.04808 t}}. \]  

(24)

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

<table>
<thead>
<tr>
<th>Initial Values</th>
<th>Fitted Values</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_0$</td>
<td>$a_0$</td>
<td>$r_0$</td>
</tr>
<tr>
<td>default</td>
<td>0.0995</td>
<td>0.555</td>
</tr>
<tr>
<td>Allen [1]</td>
<td>500.0</td>
<td>90.6546</td>
</tr>
</tbody>
</table>

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

†Sum of Squares Due to Error.

Table 4 shows that the sums of squared differences between the two models 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 6 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
Table 4. Predictions from exponential and Verhulst’s models

<table>
<thead>
<tr>
<th>Model</th>
<th>Y</th>
<th>SSE*</th>
<th>2004</th>
<th>2005</th>
<th>SSE †</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data [8, Table 1, pp. 12–13], [1, p. 139]</td>
<td>0</td>
<td>217</td>
<td>220</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Exponential (23)</td>
<td>$6 \times 10^3$</td>
<td>216</td>
<td>225</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Verhulst (24)</td>
<td>$6 \times 10^3$</td>
<td>210</td>
<td>218</td>
<td>53</td>
<td></td>
</tr>
</tbody>
</table>

*Sum of Squares Due to Error for the data from 1938 through 2003.

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 [31].

7. 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 nonempty 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 not 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. When no such existential theorems are available, conclusions of studies may have to be based on facts other than a best-fitting curve. 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, such as least absolute deviation or least squares, appears to be an open question.

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

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

so that $F_D(C) := F_D(\tilde{D})$ is well defined. For each curve $C$, the map $D \mapsto F_D(C)$ 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 nonempty topological subspace $(U, E)$ of $\mathbb{R}^N$, such as the plane $U = \mathbb{R}^2$ for linear regression, or the upper half-plane $U = \mathbb{H}_+$ in the context of Verhulst curves, or the space $U = \mathbb{R}^3$ to fit surfaces to data, with the induced Euclidean topology $E$.}

March 2012] MITSCHERLICH AND VERHULST CURVES 229
(B) A nonempty set \( D \) of nonempty 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 nonempty set \( C \) of nonempty closed subsets of \( U \), such as curves or surfaces, but henceforth called curves, with a topology \( T \) on \( C \).

(D) A function \( A \) that assigns to each index \( i \) and 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; the metric may depend on \( i \).

(E) An objective function \( F : D \times C \to \mathbb{R}_+ := [0, \infty) \), or, equivalently, for each sequence \( D \in \mathcal{D} \) a continuous and open map \( F_D : C \to \mathbb{R}_+ \), which factors through \( A \) and a function \( F_D \) in the form \( F_D(C) = F_D[A(1, z_1, C), \ldots, A(N, z_N, C)] = F_D(\tilde{z}_1, \ldots, \tilde{z}_N) \) so that \( F \) depends only on \( D \) and the adjusted points on \( C \).

If \( F_D \) 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 \( F_D \). 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 aggregation, the objective \( F_D \) may depend on a measure of discrepancy \( \rho_C(z, C) = \rho(z, \tilde{z}) \), 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 \( \rho_z \) and \( \rho_w \). The choice of \( F_D \) may arise from the geometry of the adjusted point in \( C \) depends only on the topological features of the objective \( F_D \). 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 31. Orthogonal regression is based on the Euclidean distance \( d \) from each point \( z \in U \) to \( C \subset \bar{U} \). 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 nonempty. 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 \( \rho_C(z, C) := z - \tilde{z} \in U \), and the objective may be \( F_D(C) := F_D(\tilde{D}) := \|d(z_1, \tilde{z}_1), \ldots, d(z_N, \tilde{z}_N)\| \) with any norm \( \| \| \) on \( \mathbb{R}^N \).

Example 32. For ordinary least-squares regression, to each point \( z \) corresponds an affine subspace \( U_z \subset U \) 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 \( \rho_C(z, C) := z - \tilde{z} \in U_z \); the adjusted point in \( U_z \), and the objective may be the squared Euclidean norm \( F_D(C) := F_D(\tilde{D}) := \|d(z_1, \tilde{z}_1), \ldots, d(z_N, \tilde{z}_N)\|_2^2 \) on \( \mathbb{R}^N \).

Example 33. In Example 32, an affine space \( U_z = \{z\} \) corresponds to the constraint that all curves in \( C \) pass through the point \( z \).

Example 34. 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 a modified objective defined by \( \tilde{F}_D(\tilde{W}) := \tilde{\rho}^T \cdot W \cdot \tilde{\rho} \) 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 \( | \cdot |_z \) on \( U_z \), which is thus topologically equivalent to the Euclidean norm \( \| \cdot \|_2 \) on \( U_z \). Also, each object \( C \in C \) intersects every subspace \( U_z \). Hypothesis 35 specifies the types of regression considered here.

**Hypothesis 35.** For the purpose of geometric adjustments of data, the objective

\[
F_D : U = U_{z_1} \times \cdots \times U_{z_N} \rightarrow \mathbb{R}_+
\]

is continuous and maps open subsets of \( U \) 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 \( j \in \{1, \ldots, N\} \). In particular, \( F_D(C) = 0 \) if and only if the curve \( C \in C \) passes through every data point \( z_j \). For each data sequence \( (z_1, \ldots, z_N) \) and for each curve \( C \in C \), there are adjusted points \( (\tilde{z}_1, \ldots, \tilde{z}_N) \in C^N \cap U \) where \( F_D \) has a minimum on \( C^N \cap U \), denoted by

\[
F_D(C) := \min_{(\rho_j)_{j \in C \cap U_j}} F_D(w_1, \ldots, w_N)
= F_D(\tilde{z}_1, \ldots, \tilde{z}_N).
\]

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 36.** Each positive diagonal \( 2 \times 2 \) matrix \( A \) defines a weighted norm \( | \cdot |_A \) on \( \mathbb{R}^2 \) by \( |z|^2_A := z^T \cdot A \cdot z \). For the corresponding \( A \)-orthogonal \( \ell_p \)-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 \( \rho_j := \tilde{e}_j^T \cdot (z_j - w_j) \), with \( \tilde{e}_x := (1, 0) \) for \( x \) vs. \( y \), and \( \tilde{e}_y := (0, 1) \) for \( y \) vs. \( x \):

\[
F_D(w_1, \ldots, w_N) = \tilde{\rho}^T \cdot W \cdot \tilde{\rho}.
\]

More generally, with the norm \( | \cdot |_A \) on every \( U_{z_j} \) and any norm \( \| \cdot \| \) 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, \( F_D(C) = 0 \) if and only if \( z_j \in C \) for every \( j \). For such regression methods, Proposition 37 confirms that increasing points have distinct adjusted points on increasing differentiable functions, where normals have negative slopes.

**Proposition 37.** 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 \( U_{z_k} \cap U_{z_\ell} = \emptyset \), 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 \( z_\ell \) has a positive slope \((y_k - y_\ell)/(t_k - t_\ell) > 0 \). Consequently, \( z_k, \tilde{z}_k \), and \( z_\ell \) are not collinear, whence \( \tilde{z}_k \neq \tilde{z}_\ell \).

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 37 can fail for orthogonal regressions relative to Minkowski norms whose unit discs, although convex, might not have enough symmetries.

Some regressions first transform the ambient space by a homeomorphism \( H : U \to \Gamma \) that leaves all coordinates but the last invariant and preserves the order of the last coordinate; such regressions then map increasing curves and data to increasing curves and data, minimize an objective \( \Phi_\Delta \) of the transformed data \( \Delta := H(D) \), and map the object fitted in \( \Gamma \) back into \( U \), as in logarithmic-linear regression. If \( \Phi_\Delta \) satisfies Hypothesis 35, then so does \( F_D := \Phi_\Delta \circ (H \times \cdots \times H) \), because \( H \) is an order-preserving homeomorphism.

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 29. I also thank the reviewers for Figure 2 and other suggestions that vastly improved the manuscript.

REFERENCES


YVES NIEVERGELT completed his diploma in mathematics from the École Polytechnique Fédérale de Lausanne, Switzerland, in December of 1976, skipping the graduation ceremony early in 1977 to become an officer in the compulsory Swiss military. In 1984 he earned a Ph.D. in several complex variables under the guidance of James R. King at the University of Washington in Seattle, where he also benefited from Caspar R. Curjel’s mentoring in teaching, and then taught mathematics at St. Olaf College for the 1984–1985 academic year. Since 1985 he has been teaching complex and numerical analysis as well as other topics at Eastern Washington University. He also enjoys the gym, coffee, or beer with buddies, wine with ladies, and many adventures in the dark forested mountain wilderness.

Department of Mathematics, Eastern Washington University, Cheney, WA 99004
ynievergelt@ewu.edu

---

**Mathematics Is Poetry**

“For my part,” [says Mary Churchill] “I do not see how you can make mathematics poetical. There is no poetry in them.” “Ah, that is a very great mistake! There is something divine in the science of numbers. Like God, it holds the sea in the hollow of its hand. It measures the earth; it weighs the stars; it illumines the universe; it is law, it is order, it is beauty. And yet we imagine—that is, most of us—that its highest end and culminating point is book-keeping by double entry. It is our way of teaching it that makes it so prosaic.”

From Henry Wadsworth Longfellow’s 1849 novel *Kavanagh*  
—Submitted by Ethan D. Bolker, Boston, Massachusetts