The Journal of Online Mathematics and Its Applications
Volume 7, October 2007, Article ID 1611

Some Mathematical and Statistical Aspects of Enzyme Kinetics

by Michel Helfgott and Edith Seier

Abstract

Most calculus or differential equations courses utilize examples taken from physics, often discussing them in great detail. Chemistry, however, is seldom utilized to illustrate mathematical concepts. This tendency should be reversed because chemistry, especially chemical kinetics, provides the opportunity to apply mathematics readily. We will analyze some basic ideas behind enzyme kinetics, which allow us to deal with separable and linear differential equations as well as realize the need to use power series to approximate e^x and $\ln(1+x)$ close to the origin, and to apply the recently defined Lambert W function. The models studied in this context require the estimation of parameters based on experimental data, which in turn allows us to discuss simple and multiple linear regression, transformations and non-linear regression and their implementation using statistical software.

1. Introduction

Enzymes are mainly proteins that catalyze biochemical reactions, which otherwise would proceed very slowly. They are essential to life. Their kinetics began to be understood at the beginning of the 20th century. It was observed that a typical enzyme converts a substrate into a product according to the chemical formula $S + E \to E + P$.

Assuming that we are dealing with a single-step reaction we will have $P'(t) = kS(t)E(t)$. This is due to the law of mass action, which ascertains that the rate at which a single-step chemical reaction proceeds is proportional to the product of the concentration of reactants. Thus increasing S_o, the initial concentration of substrate, and keeping the amount of enzyme concentration constant, we could increase without limits the initial rate v_o at which the product is formed. This conclusion is not in agreement with observations: v_o reaches a value beyond which the addition of more substrate does not increase the rate of initial formation of the product. To circumvent this and other difficulties scientists postulated the existence of an intermediate compound, which achieved rapidly an equilibrium with the reactants and decomposed gradually producing a molecule of the product and regenerating a molecule of enzyme. That is to say, $S + E \leftrightarrow C \to E + P$.

Let us assume that the reversible process has rate constants k_1 and k_{-1} for the forward and backward reaction respectively, while the irreversible process is governed by the rate constant k_2. Due to the above-mentioned equilibrium we have $k_1 S(t)E(t) = k_{-1}C(t)$, so

$E(t) = \frac{k_{-1}}{k_1}\frac{C(t)}{S(t)}$. But the enzyme exists either as free enzyme or forming part of the intermediate compound, thus $E_T = E(t) + C(t)$ where E_T is the total concentration of enzyme. Therefore $C(t) = E_T - E(t) = E_T - \frac{k_{-1}}{k_1}\frac{C(t)}{S(t)}$, which in turn leads to $C(t)(1 + \frac{k_{-1}}{k_1}\frac{1}{S(t)}) = E_T$. Finally we get

$$C(t) = \frac{E_T S(t)}{K + S(t)}$$

where $K = \frac{k_{-1}}{k_1}$. Since the rate of formation of the product is given by $v = P'(t)$ and according to the law of mass action $P'(t) = k_2 C(t)$, we reach the expression

$$v = \frac{k_2 E_T S(t)}{K + S(t)}$$

In particular

$$v_o = \frac{k_2 E_T S_o}{K + S_o} \qquad (1)$$

where $v_o = v(0)$ and $S_o = S(0)$.

A close look at (1) allows us to conclude that if we increase S_o, keeping E_T constant, eventually it will be much greater than K. So v_o will tend to the limiting rate $k_2 E_T$, which we denote $V_{\max}$ following common usage among biochemists. Thus

$$v = \frac{V_{\max} S(t)}{K + S(t)} \qquad (2)$$

This relationship is known as the **Michaelis-Menten equation**, honoring Leonor Michaelis and Maud Menten, who in 1913 published a groundbreaking paper on enzyme kinetics. They were two early pioneers in a relatively new field.

If we consider v_o as a function of S_o (keeping E_T constant), the following graph, shared by all functions of the form $f(x) = \frac{ax}{b + x}$, can be drawn:

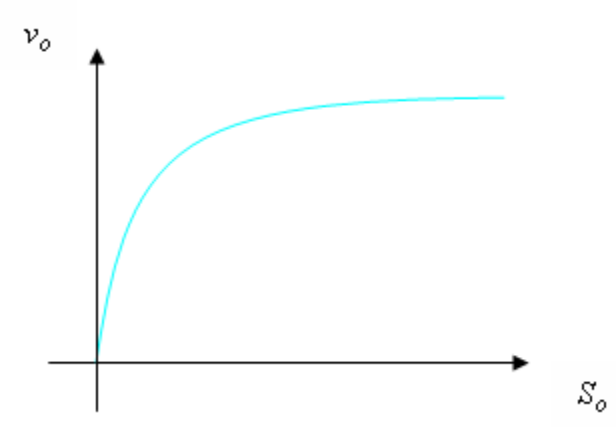


Figure 1. Initial rate as a function of initial concentration of substrate

The reader may note that Michaelis-Menten equation predicts the appearance of the phenomenon of saturation because, no matter how much substrate we add, the initial rate cannot surpass the limiting rate $V_{\max}$. By the early 1920s solid experimental evidence supporting Michaelis-Menten equation had accumulated. But the existence of an equilibrium between reactants and the intermediate compound was challenged by George Briggs and John Haldane (1925) in a remarkable two-page paper. Rather than accepting the equilibrium between substrate, enzyme, and the intermediate compound, they claimed that the rate at which the concentration of the intermediate compound varies is practically zero, except at the very beginning of the reaction. This alternative hypothesis led them to Michaelis-Menten equation, as we will see in the next section.

2. The Steady State Hypothesis

Let us recall that the basic model of enzyme kinetics is given by

$$S + E \leftrightarrow C \rightarrow E + P$$

with rate constants k_1, k_{-1} for the reversible part of the reaction and k_2 for the irreversible part. The substrate S combines with the enzyme E giving birth to an intermediate compound C through a reversible reaction. C decomposes into the product P and regenerates the enzyme E. It should be noted that one usually works with a much higher concentration of substrate than of enzyme.

The law of mass action implies that the rate at which C varies is given by

$$C'(t) = k_1 S(t) E(t) - (k_{-1} + k_2) C(t) \qquad (3)$$

Since $E_T = E(t) + C(t)$ and $S_o = S(t) + C(t) + P(t)$, from (3) it follows that

$$C'(t) = k_1(E_T - C(t))(S_o - C(t) - P(t)) - (k_2 + k_{-1})C(t)$$

Let us recall that $P'(t) = k_2 C(t)$. Thus we have two differential equations in the unknowns $P(t)$ and $C(t)$. Replacing $C(t) = \frac{1}{k_2} P'(t)$ in the first equation we arrive at

$$\frac{1}{k_2} P''(t) = k_1(E_T - \frac{1}{k_2} P'(t))(S_o - \frac{1}{k_2} P'(t) - P(t)) - \frac{k_2 + k_{-1}}{k_2} P'(t)$$

Unfortunately, this is a complicated non-linear differential equation with no known explicit solution. Some sort of qualitative simplification is indeed needed. At the beginning of the experiment substrate and enzyme combine quite rapidly giving birth to the intermediate compound C and thereafter a steady state ensues during which the concentration of C remains practically constant. Each time a molecule of P is formed by a rearrangement of C, a molecule of the enzyme is regenerated and combines rapidly with a molecule of substrate (there is a high affinity between both of them and during most of the process there are many more molecules of substrate than enzyme). This mechanism lasts during considerable time while there is substrate left. Thus we should expect that $C'(t) = 0$ under steady state conditions, which in turn implies

$$k_1 S(t)E(t) - (k_{-1} + k_2)C(t) = 0$$

But E_T, the total concentration of enzyme, equals $E(t) + C(t)$. Therefore

$$k_1 S(t)(E_T - C(t)) - (k_{-1} + k_2)C(t) = 0$$

that is to say $k_1 S(t)E_T - (k_1 S(t) + k_{-1} + k_2)C(t) = 0$.

Consequently $C(t) = \frac{E_T S(t)}{S(t) + K_m}$, where $K_m = \frac{k_{-1} + k_2}{k_1}$.

The rate at which the product is formed is given by $v = P'(t)$. But $P'(t) = k_2 C(t)$, so under steady state we will have

$$v = \frac{k_2 E_T S(t)}{S(t) + K_m} \qquad (4)$$

The reader may observe that this is Michaelis-Menten equation, except that $K_m = \frac{k_{-1} + k_2}{k_1}$ is different from $K = \frac{k_{-1}}{k_1}$ (both coincide when k_{-1} is much bigger than k_2). A similar analysis to the one done in the previous section, right after displaying (1),

leads to the conclusion that at any instant -- during the steady state -- the limiting rate is $k_2 E_T$. Thus we write $V_{\max} = k_2 E_T$ as we did before. A practical task, to be dealt with later in sections 8 - 10, is to estimate the parameters $V_{\max}$ and K_m on the basis of experimental values.

In particular formula (4) is valid at the beginning of the steady state, when we can measure the rate v_o for a certain concentration of substrate S_o. It is to be noted that for most reactions catalyzed by enzymes the stationary state is reached very quickly, in the order of milliseconds, so we may assume that S_o is the concentration of substrate at the beginning of the experiment. Let us pay close attention to the formula

$$v_o = \frac{V_{\max} S_o}{S_o + K_m}$$

For each value of S_o we should expect a different value of v_o. How could we calculate v_o? So far we do not have information about K_m or $V_{\max}$, hence v_o has to be found from experiments. Indeed, under steady state $k_1 E(t)S(t) = (k_{-1} + k_2)C(t)$. So $S'(t) = -k_1 E(t)S(t) + k_{-1}C(t) = -(k_{-1} + k_2)C(t) + k_{-1}C(t) = -k_2 C(t) = -P'(t) = -v$. Thus $v = -S'(t)$, which in turn implies that v can be approximated by the slope of the tangent line to the $S(t)$ curve. In actual practice, to estimate v_o we would have to calculate $\frac{S(t_2) - S(t_1)}{t_2 - t_1}$ where t_1 and t_2 are very close to each other and measurements are made at the beginning of the experiment. Later, once we learn more about $S(t)$ as a function of time, a practical method will be analyzed. It is to be noted that we are analyzing initial rates, but we are not dealing with what is known as the method of initial rates. This is a well-known method to calculate rate laws in chemical kinetics.

Biochemists prefer to perform measurements at the beginning of the steady state, in other words measure S_o and v_o rather than $S(t)$ and v at a later time, because some enzymes may be denaturalized as the process is under way or an appreciable amount of product may inhibit the catalytic role of the enzyme.

The Physical and Theoretical Laboratory at Oxford University (UK) has developed an applet that illustrates quite well how the curves of substrate, enzyme, intermediate compound and product vary across time. An alternative derivation of Michaelis-Menten equation, following the notation of a biochemist rather than a mathematician, has been developed at the Department of Biochemistry at the University of Leicester (UK).

3. The reasonableness of the Steady State Hypothesis

The assumption $C'(t) = 0$ played a critical role in the deduction of the Michaelis-Menten equation. Before the steady state we can write $S_o = S(t)$ instead of $S_o = S(t) + C(t) + P(t)$, in other words approximate $S(t)$ by S_o, since very little intermediate compound and product will have been formed. Then (3) becomes

$$C'(t) = k_1(E_T - C(t))S_o - (k_{-1} + k_2)C(t)$$

Thus $C'(t) + (k_1 S_o + k_{-1} + k_2)C(t) = k_1 E_T S_o$. This is a linear first order differential equation. Taking into consideration that $C(0) = 0$, the use of the integrating factor $e^{(k_1 S_o + k_{-1} + k_2)t}$ leads to the solution

$$C(t) = \frac{k_1 E_T S_o}{k_1 S_o + k_{-1} + k_2} - \frac{k_1 E_T S_o}{k_1 S_o + k_{-1} + k_2} e^{-(k_1 S_o + k_{-1} + k_2)t}$$

$$= \frac{E_T S_o}{S_o + K_m}(1 - e^{-k_1(S_o + K_m)t}) \qquad (5)$$

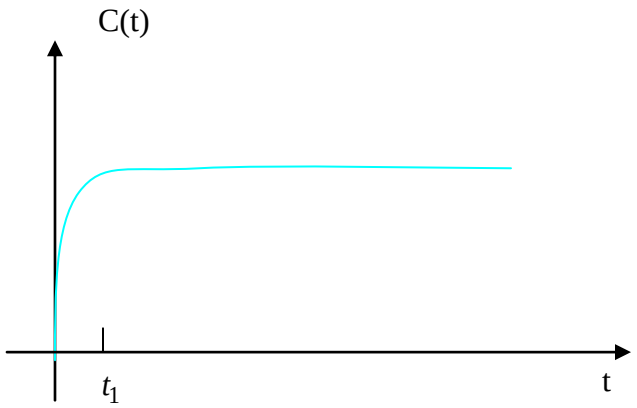


Figure 2. C(t) growth before and during stationary state

This is a function of the well-known type $b(1 - e^{-at})$, which starts from zero and tends to b. If a is a relatively large number, the function will reach its limiting value through a steep ascent. We can note that (5) has been obtained assuming that t is quite small (say, $t < t_1$) which allowed us to ascertain that $S(t)$ is practically S_o. However, if in the laboratory one works with a great excess of substrate, it is to be expected that the

approximation $S_o \approx S(t)$ will be reasonable beyond t_1; thus implying that $C(t)$ tends to the constant value $E_T S_o/(S_o + K_m)$. So, eventually $v = P'(t) = k_2 C(t) = \dfrac{V_{\max} S_o}{S_o + K_m}$.

Cornish-Bowden (1995) ascertains that a reasonable value in practice for $k_1(S_o + K_m)$ is 1000 s^{-1} when one works with a great excess of substrate. Then $e^{-1000t} < 0.01$ provided $-1000t < \ln(1/100)$, i.e. $t > 0.004605$. Thus for t bigger than 5 microseconds $0.99 < 1 - e^{-1000t} < 1$, so

$$0.99\frac{E_t S_o}{S_o + K_m} < \frac{E_T S_o}{S_o + K_m}(1 - e^{-1000t}) < \frac{E_T S_o}{S_o + K_m}.$$

Consequently, after only 5 microseconds $\dfrac{E_T S_o}{S_o + K_m}(1 - e^{-1000t})$ and $\dfrac{E_T S_o}{S_o + K_m}$ practically coincide.

We have reached Michaelis-Menten equation under the assumption that we are working with a much higher concentration of substrate vis a vis enzyme and not much product has been formed. Arguably this equation has been obtained accepting an approximation, but nonetheless it illustrates the fact that quite soon $C(t)$ will adopt a practically constant value in agreement with the steady state hypothesis. Although for quite a while we will concentrate our efforts on the period during which $C'(t) = 0$, in section 7 we will see that analyzing the period $0 < t < t_1$ (during which $C'(t) > 0$) will lead to a method to estimate k_1.

Biochemists have found that the steady state hypothesis is very fruitful. Many consequences of it are in agreement with experimental results. However, it has been challenged by several scientists, who prefer to call it the "pseudo-steady state hypothesis" or "quasi-steady-state approximation" because strictly speaking $C'(t) = 0$ at just one instant. Thereafter the curve that describes $C(t)$ is only approximately constant while there is enough substrate left. Starting in the 1960's a new approach, based on the perturbation theory of differential equations, has been developed (Bartholomay, 1972). This is an advanced branch of mathematics that allows the construction of a framework that somehow supersedes previous theories. But, at an introductory level, the steady state hypothesis is still widely used in enzyme kinetics, and with reasonably good outcomes.

4. Integrated form of Michaelis-Menten equation

We have learned how to calculate K_m and $V_{\max}$ on the basis of experimental values for S_o and v_o. But under certain circumstances it might be difficult to measure initial

velocities. What can be done? There is an alternative way if we can measure $S(t)$ at different values of t during the steady state.

Let us recall that Michaelis-Menten equation ascertains that at any instant t, during the steady state, $v = \frac{V_{\max} S(t)}{S(t) + K_m}$ where $v = P'(t) = -S'(t)$. Thus

$$-S'(t) = \frac{V_{\max} S(t)}{S(t) + K_m}.$$

This is a separable differential equation. Multiplying by $\frac{S(t) + K_m}{S(t)}$ we arrive at:

$$-S'(t) - K_m \frac{1}{S(t)} S'(t) = V_{\max}$$

Integrating with respect to time between 0 and t (considering 0 as the instant when the steady state begins) we get

$$-\int_o^t S'(u)du - K_m \int_o^t \frac{S'(u)}{S(u)}du = \int_o^t V_{\max}\, du.$$

Thus

$$-(S(t) - S(0)) - K_m \ln \frac{S(t)}{S(0)} = V_{\max} t \qquad (6)$$

Therefore

$$\frac{1}{t}\ln \frac{S(0)}{S(t)} = -\frac{1}{K_m}\frac{S(0) - S(t)}{t} + \frac{V_{\max}}{K_m}$$

This is a remarkable identity, known as the **integrated form of Michaelis-Menten,** because it does not involve rates but only experimental values of $S(t)$ during the course of an enzymatic reaction. Moreover, it predicts the appearance of a line if we have $\frac{S(0) - S(t)}{t}$ on the horizontal axis and $\frac{1}{t}\ln \frac{S(0)}{S(t)}$ on the vertical axis. It is a line with slope $-1/K_m$ and vertical intersection $\frac{V_{\max}}{K_m}$.

So, in order to estimate $V_{\max}$ and K_m we need a table of $S(t)$ values obtained at different times t. Then we build a table of two columns with $\frac{S(0)-S(t)}{t}$ on the first column and $\frac{1}{t}\ln\frac{S(0)}{S(t)}$ on the second column. A regression analysis will provide us with an approximation to the slope $\frac{-1}{K_m}$ and the vertical intersection $\frac{V_{\max}}{K_m}$. Finally, a simple arithmetical procedure will lead to the corresponding values of K_m and $V_{\max}$. Section 10 is devoted to analyses of this kind.

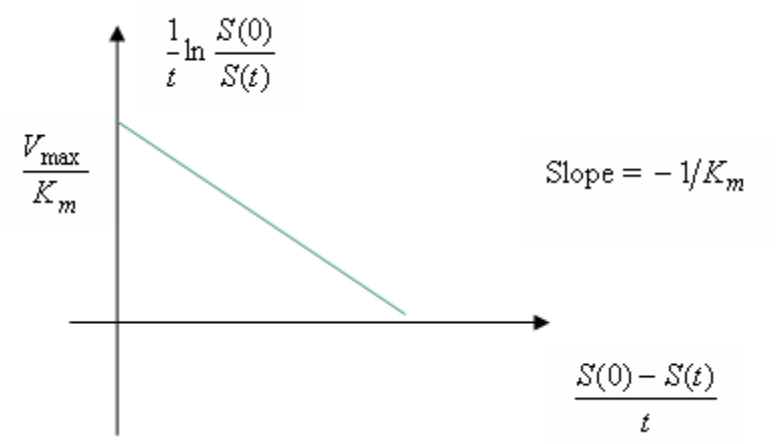


Figure 3. Integrated form of Michaelis-Menten equation

The advantage of this approach to the estimation of K_m and $V_{\max}$ resides on the fact that it does not require measuring rates (often a challenging experimental procedure). A difficulty that may appear when working with the integrated equation is that it requires measuring values of $S(t)$ across time, a situation that could lead to problems because diverse factors may eventually distort the reaction after it has gotten under way. For instance, as we mentioned before, the product might inhibit enzyme activity or the enzyme might become unstable.

5. The Variation of Substrate at the Beginning of the Steady State

The integrated form of the Michaelis-Menten equation provides a useful way of calculating $V_{\max}$ and K_m. Interestingly enough, it is possible to find a simple approximate formula for $S(t)$ at the beginning of the steady state. For this purpose we have to keep in mind a mathematical approximation, namely that $\ln(1+x) \approx x$ whenever $|x|$ is a very small positive number. Thus at the beginning of the steady state, when $S(t)$ does not differ much from $S(0)$,

$\ln\frac{S(t)}{S(0)} = \ln(1+\frac{S(t)}{S(0)}-1) = \ln(1+\frac{S(t)-S(0)}{S(0)}) \approx \frac{S(t)-S(0)}{S(0)}$. Multiplying (6) by -1 we get

$$S(t)-S(0)+K_m \ln\frac{S(t)}{S(0)} = -V_{\max}t\,,$$

which can be replaced by $S(t)-S(0)+K_m\frac{S(t)-S(0)}{S(0)} = -V_{\max}t$ due to the above-mentioned approximation. Therefore

$$S(t) = S(0) - \frac{V_{\max}}{1+\frac{K_m}{S(0)}}t = S(0) - \frac{V_{\max}S(0)}{S(0)+K_m}t \qquad (7)$$

Hence $S(t) = S(0) - v_o t$. Having t on the horizontal axis, we are then dealing with a straight line with slope $-v_o$ and vertical intercept $S(0)$. In other words, at the beginning of the steady state the variation of $S(t)$ is linear. The fact that $-v_o$ is the slope of this line has practical implications because it suggests a way to estimate the initial rate: make several measurements of $S(t)$ at the beginning of the experiment and apply simple regression.

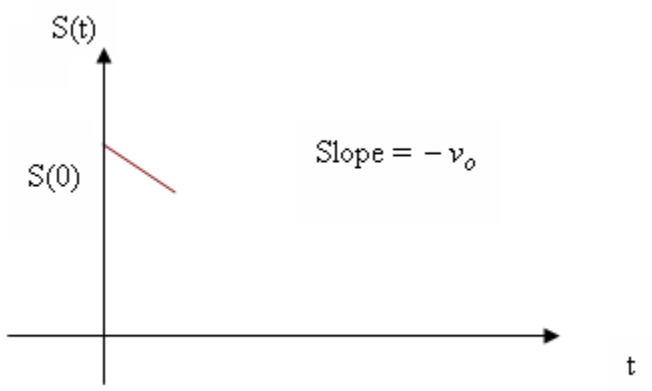


Figure 4. Variation of substrate at the beginning of the steady state

6. Lambert W function

A closed form solution of Michaelis-Menten equation was found by Schnell and Mendoza (1997), using Lambert W function for that purpose. Let $h(x) = xe^x$ for all

$x > 0$. Then $h'(x) = (1+x)e^x$, so $h'(x) > 0$ for every positive x. The inverse of $h : (0,\infty) \to (0,\infty)$ will then exist, which we denote W. This is Lambert W function (actually $h(x)$ is strictly increasing on $[-1,\infty)$, so $W(x)$ is defined on $[-\frac{1}{e},\infty)$, but for the purposes we have in mind it is convenient to restrict W to $(0,\infty)$).

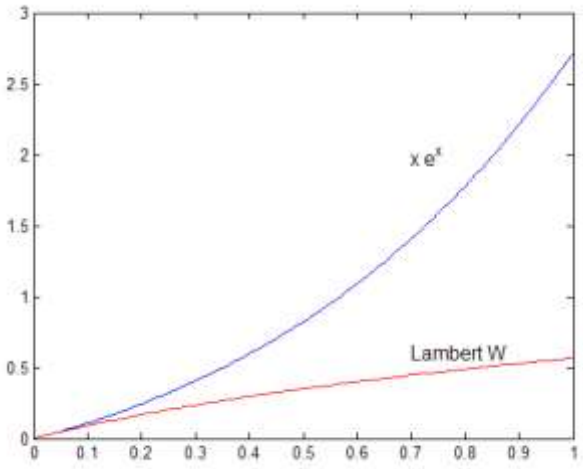


Figure 5. Lambert W function as the inverse of $x \to xe^x$

Therefore $h(W(x)) = x$ for every $x > 0$, i.e. $W(x)e^{W(x)} = x$. Thus

$$\ln W(x) + W(x) = \ln x \tag{8}$$

On the other hand, the integrated form of Michaelis-Menten equation ascertains the validity of (6), i.e. $S(t) - S_o + K_m \ln\frac{S(t)}{S_o} = -V_{\max} t$. Hence

$$\frac{S(t)}{K_m} + \ln\frac{S(t)}{K_m} = \frac{S_o}{K_m} + \ln\frac{S_o}{K_m} - \frac{V_{\max}}{K_m} t \tag{9}$$

Looking closely at (8) and (9), it seems reasonable to suspect that for any t during the steady state phase of the reaction there will exist a positive number x such that $\frac{S(t)}{K_m} = W(x)$. Let us assume that such an x exist. So $\ln\frac{S(t)}{K_m} = \ln W(x)$, which in turn implies $\frac{S(t)}{K_m} + \ln\frac{S(t)}{K_m} = W(x) + \ln W(x)$. Therefore

$\frac{S_o}{K_m} + \ln\frac{S_o}{K_m} - \frac{V_{\max}}{K_m} t = W(x) + \ln W(x)$, i.e.

$$\frac{S_o - V_{\max} t}{K_m} + \ln\frac{S_o}{K_m} = W(x) + \ln W(x)$$

Applying the exponential function to both sides we get $\frac{S_o}{K_m} e^{\frac{S_o - V_{\max} t}{K_m}} = W(x) e^{W(x)}$. But let us recall that $W(x) e^{W(x)} = x$. Therefore $x = \frac{S_o}{K_m} e^{\frac{S_o - V_{\max} t}{K_m}}$.

Next we can prove that under steady state conditions it is true that

$$\frac{S(t)}{K_m} = W\left(\frac{S_o}{K_m} e^{\frac{S_o - V_{\max} t}{K_m}}\right)$$

Indeed, let $x = \frac{S_o}{K_m} e^{\frac{S_o - V_{\max} t}{K_m}}$. Then $\ln x = \ln \frac{S_o}{K_m} + \frac{S_o}{K_m} - \frac{V_{\max}}{K_m} t$, which thanks to (9) leads to the equality $\ln x = \frac{S(t)}{K_m} + \ln \frac{S(t)}{K_m}$. But $\ln x = W(x) + \ln W(x)$, consequently $W(x) + \ln W(x) = \frac{S(t)}{K_m} + \ln \frac{S(t)}{K_m}$. In general, if $a + \ln a = b + \ln b$ then $a = b$ because $r \to r + \ln r$ is a 1:1 function. Therefore $W(x) = \frac{S(t)}{K_m}$.

More about Lambert W function, including an extensive bibliography, can be found in a paper by Hayes (2005).

At the end of section 10 we will show how Lambert W function can help to improve the estimation of $V_{\max}$ and K_m when analyzing data obtained from measurements done across the steady state.

7. A close analysis before the onset of the steady state

Since we are able to estimate $V_{\max}$ from experiments and E_T is known, right away we can estimate the rate constant k_2 (recall that $V_{\max} = k_2 E_T$). How can we estimate k_1 and k_{-1}? Evidently, it is enough to find k_1 because $k_{-1} = k_1 K_m - k_2$ and K_m is found through experiments. We will analyze the basic model of enzyme kinetics before the steady state, a very short period of time at the beginning of the experiment during which it is a good approximation to assume that $S(t)$ can be replaced by S_o.

In section 3 we found that under these circumstances

$$C(t) = \frac{E_T S_o}{S_o + K_m} - \frac{E_T S_o}{S_o + K_m} e^{-k_1(S_o+K_m)t}$$

Since $P'(t) = k_2 C(t)$ we can conclude that

$$P'(t) = \frac{k_2 E_T S_o}{S_o + K_m} - \frac{k_2 E_T S_o}{S_o + K_m} e^{-k_1(S_o+K_m)t}$$

Therefore $P(t) = \int_o^t \frac{k_2 E_T S_o}{S_o + K_m} du - \frac{k_2 E_T S_o}{S_o + K_m} \int_o^t e^{-k_1(S_o+K_m)u} du$

$$= \frac{k_2 E_T S_o}{S_o + K_m} t + [\frac{k_2 E_T S_o}{k_1(S_o + K_m)^2} e^{-k_1(S_o+K_m)u}]_o^t$$

$$= \frac{k_2 E_T S_o}{S_o + K_m} t + \frac{k_2 E_T S_o}{k_1(S_o + K_m)^2} e^{-k_1(S_o+K_m)t} - \frac{k_2 E_T S_o}{k_1(S_o + K_m)^2}$$

During the pre-steady state the values of t are extremely small, so we can approximate $e^{-k_1(S_o+K_m)t}$ by the first three terms of its series expansion, namely

$$1 - k_1(S_o + K_m)t + \frac{{k_1}^2(S_o + K_m)^2}{2} t^2$$

Consequently

$$P(t) = \frac{k_2 E_T S_o}{S_o + K_m} t + \frac{k_2 E_T S_o}{k_1(S_o + K_m)^2} (1 - k_1(S_o + K_m)t + \frac{{k_1}^2(S_o + K_m)^2}{2} t^2)$$

$$- \frac{k_2 E_T S_o}{k_1(S_o + K_m)^2} = \frac{k_1 k_2 E_T S_o}{2} t^2$$

That is to say, $$P(t) = \frac{k_1 V_{\max} S_o}{2} t^2 \qquad (10)$$

Thus, we can predict that if it is possible to measure $P(t)$ before the steady state, and plot points on a graph with time on the horizontal axis and $\frac{2P(t)}{t^2 V_{\max} S_o}$ on the vertical axis, the points should be spread around a line parallel to the horizontal axis. Thereafter we

estimate k_1 through linear regression. It is to be noted that rapid reaction techniques, developed in the 1950's, allow the measurement of $P(t)$ before the onset of the steady state. A great success of the basic model of enzyme kinetics was to make predictions that were later tested with success, within the limits set by experimental errors.

Roughton (1954) reached (10) using an alternative path, which is worth discussing. We start by differentiating (3), keeping in mind that $P'(t) = k_2 C(t)$ as well as the fact that $E(t) = E_T - C(t)$. Thus:

$$\begin{aligned} P''(t) = k_2 C'(t) &= k_2 (k_1 (E_T - C(t)) S(0) - (k_{-1} + k_2) C(t)) \\ &= k_2 k_1 E_T S(0) - (k_1 S(0) + k_{-1} + k_2) k_2 C(t) \\ &= k_2 k_1 E_T S(0) - (k_1 S(0) + k_{-1} + k_2) P'(t) \end{aligned}$$

Hence $P''(t) + a_1 P'(t) = a_2$, where $a_1 = k_1 S(0) + k_{-1} + k_2$ and $a_2 = k_1 S(0) V_{\max}$. This is a second order linear non-homogenous differential equation. Since the solutions of the equation $r^2 + a_1 r = 0$ are 0 and $-a_1$ (the roots of the characteristic polynomial), and a simple inspection allows us to ascertain that $\frac{a_2}{a_1} t$ is a solution of the differential equation, we can conclude that the general solution is

$$P(t) = c_1 + c_2 e^{-a_1 t} + \frac{a_2}{a_1} t .$$

Thus $P'(t) = -c_2 a_1 e^{-a_1 t} + \frac{a_2}{a_1}$. Moreover, we have $P(0) = 0$ and

$P'(0) = k_2 C(0) = k_2 * 0 = 0$. Therefore $c_1 + c_2 = 0$ and $-a_1 c_2 + \frac{a_2}{a_1} = 0$, which lead to

$c_2 = \frac{a_2}{{a_1}^2}$, $c_1 = -\frac{a_2}{{a_1}^2}$. We can conclude that $P(t) = -\frac{a_2}{{a_1}^2} + \frac{a_2}{{a_1}^2} e^{-a_1 t} + \frac{a_2}{a_1} t$. However, taking into consideration that during the pre-steady state the values of t are extremely small, we can approximate $e^{-a_1 t}$ by the first three terms of its series expansion; namely $1 - a_1 t + \frac{{a_1}^2}{2} t^2$. Therefore,

$$P(t) = -\frac{a_2}{{a_1}^2} + \frac{a_2}{{a_1}^2} (1 - a_1 t + \frac{{a_1}^2}{2} t^2) + \frac{a_2}{a_1} t = \frac{a_2}{2} t^2 = \frac{k_1 V_{\max} S(0)}{2} t^2$$

That is to say, $\frac{2P(t)}{t^2 V_{\max} S(0)} = k_1$.

Roughton's approach is found in several works, for instance Bartholomay (1972), Marangoni (2005).

8. Estimation of parameters

Each enzyme has a specific, unique value for K_m and $V_{\max}$, so estimating both constants helps to identify an enzyme. How could we estimate $V_{\max}$ and K_m? Doing experiments we choose different values of S_o and measure the corresponding initial rate v_o. Let us recall that the latter is approximated by the slope of the tangent line to the $S(t)$ curve at the beginning of the experiment. Having a S_o, v_o table of experimentally determined values we could fit a curve as best as possible. The horizontal asymptote would be $V_{\max}$ while K_m is the value of S_o at which the initial rate becomes $V_{\max}/2$. The latter assertion follows from the fact that , using the Michaelis-Menten equation, $\frac{V_{\max}}{2} = \frac{V_{\max} S_o}{S_o + K_m}$ if and only if $S_o = K_m$.

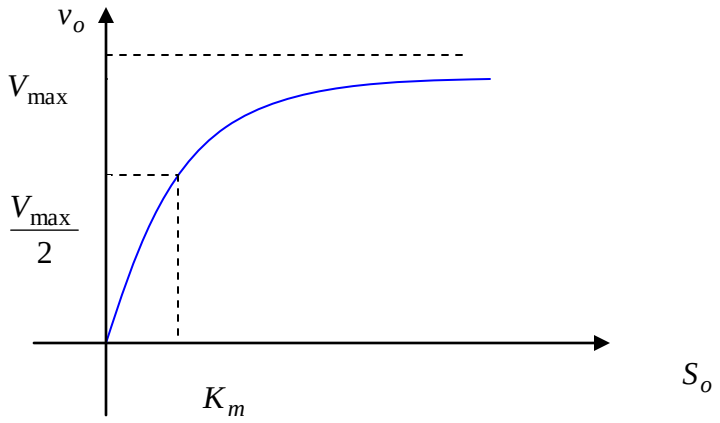


Figure 6. $V_{\max}$ and K_m from the relationship between v_o and S_o.

However, it is not an easy task to fit by hand the above-mentioned curve to experimental values of initial substrate concentrations and initial rates . There is a simple alternative, which we will study next. Taking the converse of the Michaelis-Menten equation we get

$$\frac{1}{v_o} = \frac{S_o + K_m}{V_{\max} S_o}$$

which in turn is equivalent to $\frac{1}{v_o} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}}\frac{1}{S_o}$. Thus, if we choose to have $\frac{1}{S_o}$ on the x-axis and $\frac{1}{v_o}$ on the y-axis , the experimental values should cluster around a straight line (figure 7).

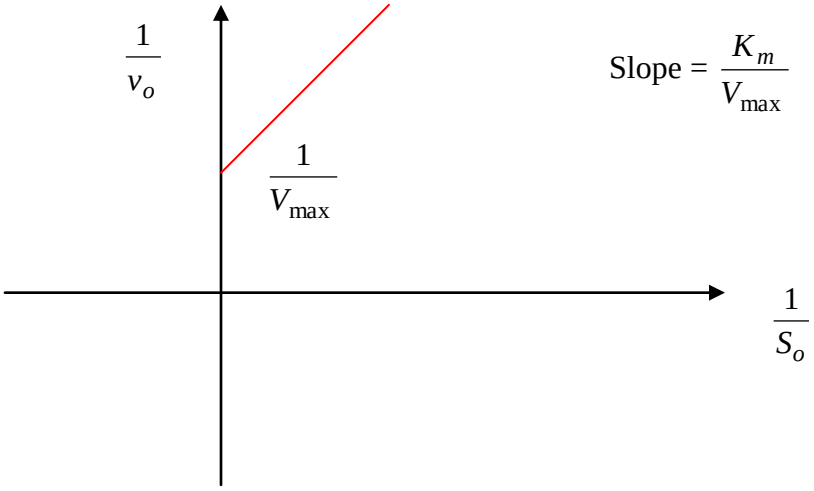


Figure 7. Lineweaver-Burk plot

Thereafter we use calculators or computers to find the least squares line, also called the regression line, which in turn will allow us to calculate $1/V_{\max}$ as the intersection with the y axis and $K_m/V_{\max}$ as the slope. From these values we can easily obtain $V_{\max}$ and K_m. The plot under consideration is known as a Lineweaver-Burk plot in recognition of Hans Lineweaver and Dean Burk who introduced this way of calculating $V_{\max}$ and K_m in 1934. From the statistical point of view, we are doing linear regression on transformations of the original variables. An example will help to understand the procedure. Let us consider the following kinetic data set (Table 1 and Figure 8) related to the hydration of CO_2 utilizing the enzyme carbonic anhydrase (McQuarrie and Simon, 1997):

Table 1. McQuarrie and Simon data

$S_o\ (mol \cdot dm^{-3})$	$v_o\ (mol \cdot dm^{-3} \cdot s^{-1})$
1.25×10^{-3}	2.78×10^{-5}
2.5×10^{-3}	5.00×10^{-5}
5×10^{-3}	8.33×10^{-5}
20×10^{-3}	1.66×10^{-4}

To estimate the parameters of the regression line we will use a graphics calculator (TI-89 or similar calculators) but also statistical software could be used. To work with the calculator we build two lists, one for the $1/S_o$ values and the other for the $1/v_o$ values, and store them as $l1$ and $l2$, i.e.

$$\{\frac{10^3}{1.25},\frac{10^3}{2.5},\frac{10^3}{5},\frac{10^3}{20}\}\rightarrow l1,\quad \{\frac{10^5}{2.78},\frac{10^5}{5},\frac{10^5}{8.33},\frac{10^4}{1.66}\}\rightarrow l2$$

Then we calculate the linear regression line using the commands $LinReg\ \ l1,l2$ and $Showstat$. The following line appears on the screen*:

$$y = 39.934042x + 4023.940015$$

Hence $\frac{1}{\hat{V}_{\max}} = 4023.940015$, which in turn leads to $\hat{V}_{\max} = 0.0002485126$, while $\frac{\hat{K}_m}{\hat{V}_{\max}} = 39.934042$. Replacing the value of $\hat{V}_{\max}$ we get $\hat{K}_m = 0.009924$. That is to say, $\hat{V}_{\max} = 2.4851265\times 10^{-4}$ while $\hat{K}_m = 9.924\times 10^{-3}$.

Multiplying by S_o the Lineweaver-Burk expression $\frac{1}{v_o} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}}\frac{1}{S_o}$ we get a linear model on a different transformation of the variables: $\frac{S_o}{v_o} = \frac{1}{V_{\max}}S_o + \frac{K_m}{V_{\max}}$. A plot of $\frac{S_o}{v_o}$ versus S_o (called Hanes plot) will be linear, with slope $\frac{1}{V_{\max}}$ and y-intercept $\frac{K_m}{V_{\max}}$. Using the same data from Mc Quarrie & Simon, a linear regression leads to the equation $y = 4028.013019x + 39.916148$. Thus $\frac{1}{\hat{V}_{\max}} = 4028.013019$ and $\frac{\hat{K}_m}{\hat{V}_{\max}} = 39.916148$. Therefore the estimations for the parameters are $\hat{V}_{\max} = 2.48\times 10^{-4}$ and $\hat{K}_m = 9.91\times 10^{-3}$.

* The correlation coefficient comes out to be 0.9999999473, thus indicating, based on data, a strong linear association between the two variables.

Another possibility is to multiply the Lineweaver-Burk expression by $v_o V_{\max}$ and thus obtain $v_o = -K_m \frac{v_o}{S_o} + V_{\max}$. Then we choose to have $\frac{v_o}{S_o}$ on the x-axis and v_o on the y-axis. The quantity $-K_m$ will become the slope and $V_{\max}$ the intercept with the y-axis. Plots of this sort are called Eadie-Hofstee plots . A linear regression on the transformed data for the Mc Quarrie & Simon data leads to $y = -0.009914x + 0.000248336$. So $\hat{K}_m = 9{,}914 \times 10^{-3}$ and $\hat{V}_{\max} = 2.48336 \times 10^{-4}$.

We can see that the values of $V_{\max}$ and K_m, estimated using linear regression on the three different sets of transformations of the original variables, do not differ much from each other because there are only four observations to estimate two parameters and the linear correlation in the three cases is very strong. However, when dealing with data from replicates of experiments such that for the same value of S_o not all the values of v_o are equal, the results obtained by the three paths might differ as we will see soon. The three possibilities just examined involve transformations of the original variables in order to convert a non-linear relationship into a linear one. Thanks to statistical software, it is now possible to fit the non-linear model directly and avoid the transformations; this is the main topic in the next section.

9. Non-linear Regression

Before computers became powerful and widely available, linearization of a non-linear model by applying non-linear transformations to the variables, as we did above, was the practical way to solve the problem. However, nowadays there are computer programs available to estimate the parameters of a non-linear model without transforming the variables. In particular **R** is a free software that has a command to do non-linear regression. **R** is available from http://www.r-project.org ; step by step instructions to download the program can be found at http://www.etsu.edu/math/seier/gettingR.doc We will illustrate its use with the same example to which we applied the traditional method of linearization.

The procedure of doing non-linear regression can be summarized in two steps:

a) First we need to come up with initial estimates of the parameters. For this purpose we need to recall the role of the parameters in the curve (Figure 6). If we write the model as $y = \frac{V_{\max} x}{x + K_m}$, $V_{\max}$ is the limiting rate. So it would just make sense to have as initial estimate the maximum rate attained in the experiment or a value close to it. In the example the maximum rate was 0.000166, so we could use its rounded version 0.0002 as an initial estimate for $V_{\max}$.

To get an initial estimate for K_m we must remember that K_m is the value of x (substrate concentration) that corresponds to ½ of $V_{\max}$. In the example, $V_{\max}/2$ is approximately 0.0001. Locating the value 0.0001 on the vertical axis and going to the

right to guess a value of x, we would guess that x = 0.007 when V = 0.0001, so we will take 0.007 as our initial estimate for K_m.

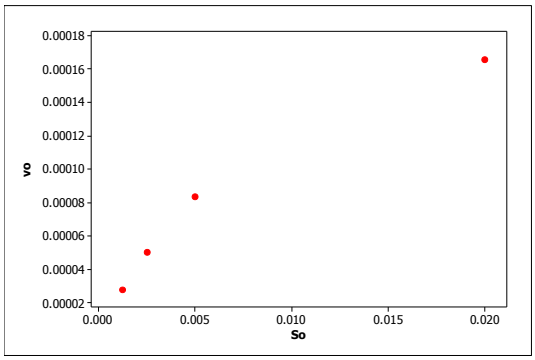


Figure 8. Experimental data from table 1.

b) The program will calculate the sum of squares of residuals from the model assuming the initial estimates are the parameters, then it will change the values of the parameters a little bit and will re-calculate the sum of squares of residuals. The process continues until the reduction in the sum of squares of residuals is negligible. The sum of squares of residuals can be thought of as a function of the values of the parameters; we can think of its graph as a surface and we want to reach to the minimum of that surface. Imagine a valley that can have hills and slopes and we want to reach the location in the valley that has the minimum altitude. In which direction to walk toward the minimum it is important to arrive there soon. Marquadt algorithm searches for the minimum and uses mathematical tools (the Gauss-Newton algorithm) to walk on the route of the steepest slope.

Here we include the commands we need to type in **R** to perform the non-linear estimation for the data in the example.
First we enter data with:

```
> x<-c(0.00125, 0.0025, 0.005, 0.020)
> y<-c(0.0000278,0.00005,0.0000833,0.000166)
```

Then we write the command for nonlinear regression indicating the equation of the model and the values of the initial estimates of the parameters (we are using the symbols a for K_m and b for V_{max} to simplify the notation):

```
nls(y~(b*x)/(x+a), start = list(a = 0.007, b = 0.0002))
```

The output of the program is:

```
Nonlinear regression model
  model:  y ~ (b * x)/(x + a)
   data:  parent.frame()
          a           b
9.9041861615 0.0002482121
 residual sum-of-squares: 1.895619e-15
```

So the estimated value for K_m is 0.0099041861 and the estimated value for $V_{\max}$ is 0.0002482121. That is to say, $K_m = 9.9041861 \times 10^{-3}$ and $V_{\max} = 2.482121 \times 10^{-4}$, quite close to the values found before using linear regression.

When data is such that all the points seem to be on a curve of the type specified by the model, it is likely that the estimated values using transformed variables and linear regression are going to be very similar to the estimated values through non-linear regression. However, one of the principles of experimental design is replication. Experiments should be repeated, and when experiments are repeated under the same conditions not always the same value for the response variable is obtained. This is due to natural randomness of the phenomena (relationships in nature are not exactly deterministic) or to experimental error (measurement errors or involuntary small changes in the conditions or the environment where the experiment is conducted). In those cases the estimated values of the parameters obtained by non-linear estimation and by linearization might differ.

The values of K_m and V_{max} that minimize the sum of squares of residuals $\sum_i^n e_i^2$ in a linear regression with variables 1/V and 1/S might not be the same that minimize $\sum_i^n e_i^2$ in a non-linear regression model with variables V and S. We also need to remember that linear regression makes several assumptions (linearity, constant error variance and normality of errors). If those assumptions are violated by the transformed data we might be better off working with non-linear regression using appropriate software.

In the following examples, the data corresponding to hypothetical replicates have been simulated based on the data from Table 1 (plotted in Figure 8), assuming that the experiment was replicated 5 times at each value of x (S_o) incorporating randomly generated errors using either additive or multiplicative models. The three synthetic data sets appear in Figures 9A and 10A and 12A.

Simulation 1A. **Constant variability for the response variable**.

The variance of the error is intended to be the same for all the values of x (S_o). The artificial errors "e" were added to the experimental value of "V_o"($y = V_o + e$, where the errors have been generated to have mean = 0).

Obs	x	y
1	0.00125	0.0000092
2	0.00125	0.0000363
3	0.00125	0.0000229
4	0.00125	0.0000069
5	0.00125	0.0000627
6	0.00250	0.0000656
7	0.00250	0.0000762
8	0.00250	0.0000530
9	0.00250	0.0000650
10	0.00250	0.0000300
11	0.00500	0.0000687
12	0.00500	0.0001034
13	0.00500	0.0000705
14	0.00500	0.0000289
15	0.00500	0.0001158
16	0.02000	0.0001758
17	0.02000	0.0001574
18	0.02000	0.0001810
19	0.02000	0.0001638
20	0.02000	0.0001038

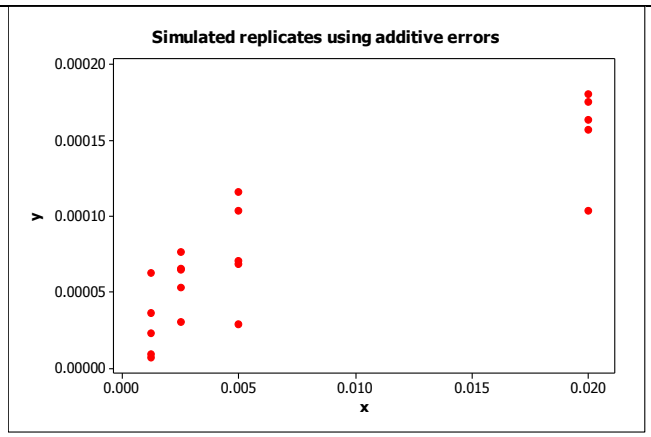


Figure 9.A. Synthetic data with constant variability.

The fact that some of the values of y are quite close to 0 poses a problem for the Lineweaver-Burk linearization approach. The regression equation on the transformed variables is `1/y = - 1934 + 81.4 1/x`. The estimated intercept is negative. This would produce a negative value for $V_{\max}$, something that is not reasonable. However, the values close to 0 do not pose a problem for the non-linear estimation. Using the same initial values as before, namely nls(y1a~(b*x)/(x+a), start = list(a = 0.007,b = 0.0002)), we get the following estimates for the parameters:

$$\hat{K}_m = 0.0080050682\ , \quad \hat{V}_{\max} = 0.0002176077$$

(residual sum-of-squares: 1.184174e-08).

There is an additional problem while working with the linearization approach with the simulated data in Figure 9A; once we plot the values of 1/y and 1/x to fit the regression line, we realize that an assumption of linear regression is being violated (see Figure 9.B): variability is not constant.

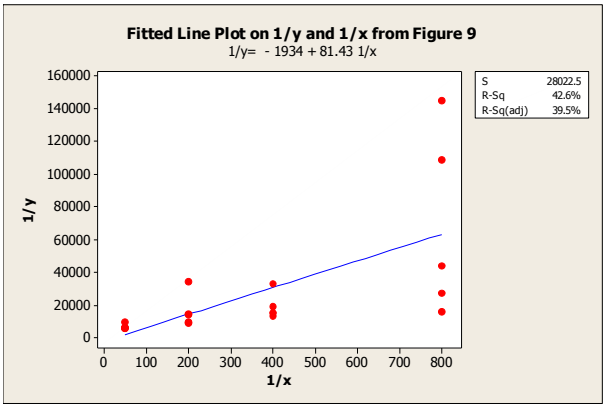


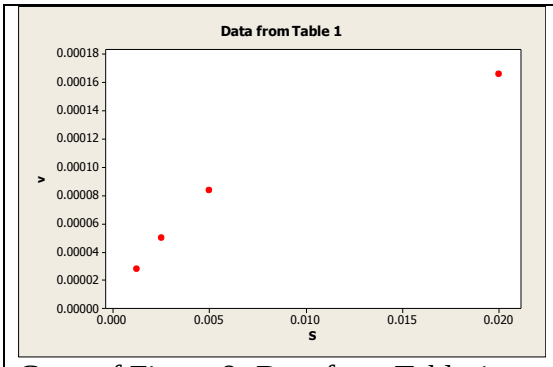
Figure 9.B.Regression on the transformed data from Figure 9A

Figure 9B also helps to understand the origin of the negative estimate of the intercept. A very low value of y for a low value of x in Figure 9A results into a very high value of 1/y for a high value of 1/x, and causes the line to have a steep slope and as a consequence the intercept becomes negative. This could happen even if we had no replicates, it would be enough to have a (negative) error in the measurement of the observation near the origin to produce this problem.

Simulation 1B. No replicates, but there is an error in the first observation

Figure 9C displays the regression line on the transformed variables using the data from Table 1. We introduced an error in the value of v in the first observation replacing 0.0000278 with 0.0000069; the modified data set is displayed on Figure 9D. Figure 9E shows the regression line on the transformed variables for the data in Figure 9D, where the high point to the right is the reason for the negative intercept. The question is: how low the first data point in Figure 9D needs to be to produce a situation like the one we see in Figure 9E?

We will analyze the situation of a negative estimate of V_{max} when using linear regression on the transformed variables in the absence of replicates, but before doing so we should mention that using non-linear regression, for the same example, the estimated value of V_{max} is positive: $\hat{K}_m$ =0.0123880970 $\hat{V}_{\max}$ =0.0002706498.



Copy of Figure 8- Data from Table 1

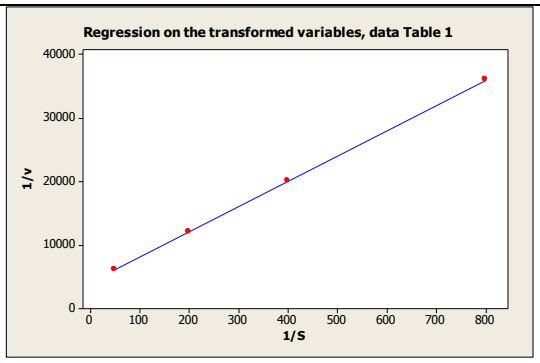


Figure 9C- Transformed data Table 1

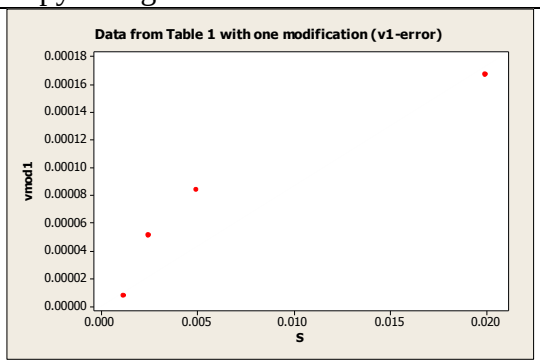


Figure 9D- Data Table 1 with error in first observation

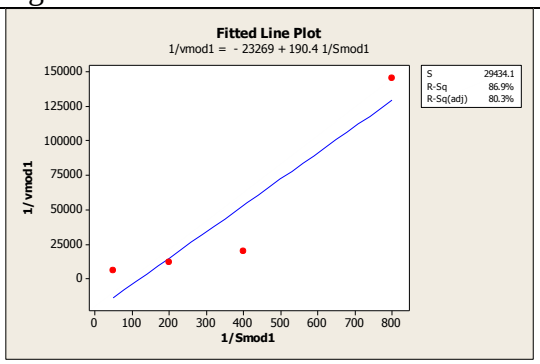


Figure 9 E- Transformed data from Figure 9D

The first point is in a position of high leverage; if it moves down enough, the slope would become steeper and the intercept might become negative. We will analyze the situation for the four observations of Table 1, considering v* = 0.0000278- e_1 for the first observation of the rate variable.

The least squares equations for the simple linear regression are:

$$n\hat{\beta}_o + \hat{\beta}_1 \sum x_i = \sum y_i$$

$$\hat{\beta}_o \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i$$

Since x = 1/S and y = 1/v, the least squares equations with the data in Table 1, except that the first observation for ν, v_1, has been replaced by v* =0.0000278- e_1 are as follows:

$$4\hat{\beta}_o + \hat{\beta}_1 \sum \frac{1}{S_i} = \sum \frac{1}{v_i}$$

$$\hat{\beta}_o \sum \frac{1}{S_i} + \hat{\beta}_1 \sum \frac{1}{S^2_i} = \sum \frac{1}{S_i v_i}$$

or

$$4\hat{\beta}_o + \hat{\beta}_1 \ 1450 = 38028.9 + \frac{1}{0.0000278 - e_1}$$

$$\hat{\beta}_o 1450 + \hat{\beta}_1 842500 = 10702165 + \frac{1}{0.00125(0.0000278 - e_1)}$$

Therefore the solution for the intercept $\hat{\beta}_o$ is

$$\hat{\beta}_o = \frac{842500(38028.9 + \frac{1}{0.0000278 - e_1}) - 1450(10702165 + \frac{1}{0.00125(0.0000278 - e_1)})}{1267500}$$

Thus the condition for the estimated intercept $\hat{\beta}_o$ to be negative can be written as:

$842500(38028.9 + \frac{1}{0.0000278 - e_1}) < 1450(10702165 + \frac{1}{0.00125(0.0000278 - e_1)})$. Thus

$$32039348250 + \frac{842500}{0.0000278 - e_1} < 15518139250 + \frac{1450}{0.0000000347 - 0.00125e_1},$$

i.e. $$16521209000 < \frac{1450}{0.0000000347 - 0.00125e_1} - \frac{842500}{0.0000278 - e_1}$$

It can be perceived from the intermediate equations and from Figure 9E that it is not only the size of the error in the first observation of rate what would determine a negative

estimate for the intercept (and thus a negative estimate for V_{max}) when using linear regression on the transformed variables, but the size of the error in relation to the position of the other data points as well.

Writing the condition for negative intercept in more general terms, we could say that when instead of the true value of the rate v_1 we do the measurement with an error such that we record $v_1 - e_1$ bringing the point closer to v = 0, the problem of the estimated intercept being negative happens when

$$\sum_{i=1}^{n} \frac{1}{S_i^2}\left(\sum_{i\neq 1} \frac{1}{v_i} + \frac{1}{v_1 - e_1}\right) < \sum_{i=1}^{n} \frac{1}{S_i}\left(\sum_{i\neq 1} \frac{1}{S_i v_i} + \frac{1}{S_1(v_1 - e_1)}\right)$$

Simulation 2. **Variability inversely proportional to the value of the response variable**.
The synthetic data appear in the next table and graph. Here we assume that the variability increases when the value of x (S_o) increases.

Obs.	x	y
1	0.00125	0.0000329
2	0.00125	0.0000275
3	0.00125	0.0000341
4	0.00125	0.0000341
5	0.00125	0.0000198
6	0.00250	0.0000591
7	0.00250	0.0000420
8	0.00250	0.0000491
9	0.00250	0.0000448
10	0.00250	0.0000588
11	0.00500	0.0001381
12	0.00500	0.0000803
13	0.00500	0.0001067
14	0.00500	0.0001253
15	0.00500	0.0000842
16	0.02000	0.0001431
17	0.02000	0.0001815
18	0.02000	0.0001839
19	0.02000	0.0001612
20	0.02000	0.0001276

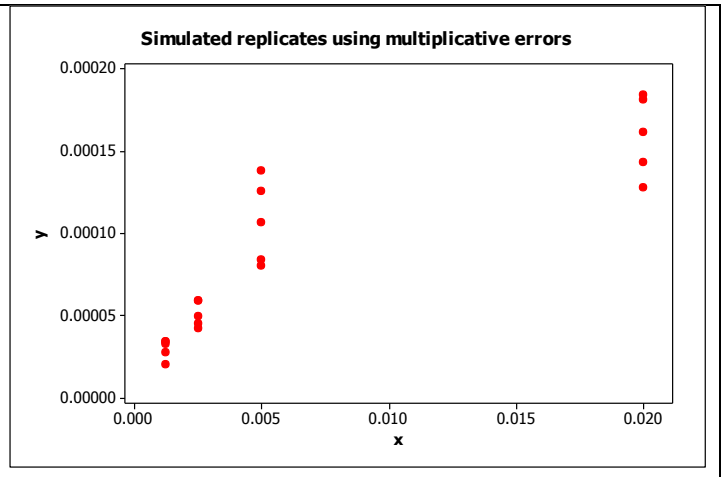


Figure 10A. Synthetic data with variability proportional to level

Lineweaver-Burk's method uses the inverse of the variables. The correlation between $1/v_o$ and $1/S_o$ is 0.926; the fitted line appears in the next figure. Notice how the high variability in v_o when v_o takes higher values and $S_o = 0.02$ gets transformed into small variability and vice-versa, the small variability when $S_o = 0.00125$ and v_o is small (at the left in the preceding figure) gets transformed into the values with high variability at the right of the fitted line plot.

The regression equation in the transformed variables is 1/y = 3497 + 39.63 1/x, thus Lineweaver-Burk's method gives the following estimated values for the parameters:

$\hat{V}_{max}$ =1/3497 = 0.000286, $\hat{K}_m$ = 39.63/3497 = 0.011333

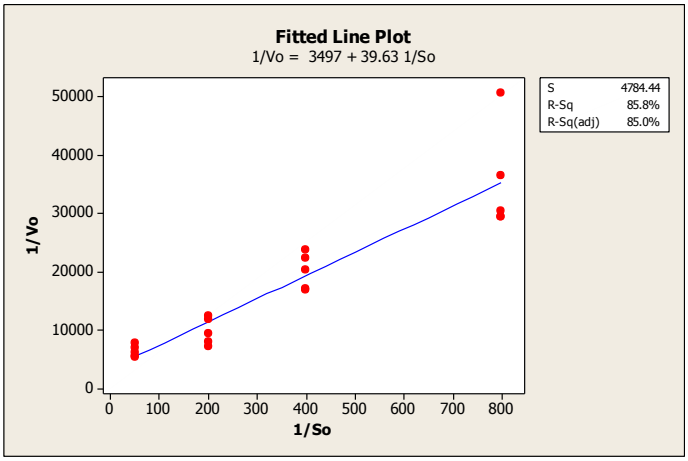


Figure 10B. Linear regression using Lineweaver-Burk transformation

Using **R** to do the estimation of the parameters of the non-linear model, the results are as follows:

nls(y3~(b*x)/(x+a), start = list(a = 0.007, b = 0.0002))

$\hat{K}_m$ = 0.0063723704 $\hat{V}_{max}$ = 0.0002139074 (residual sum-of-squares: 6.769675e-09)

Both models are plotted in the next figure. The values on the plotted curve were obtained in the following way:
Using linearization $\hat{y}$ =(0.000286*x)/(x+0.011334)
Using the non-linear model $\hat{y}$ =(0.0002139074*x)/(x+0.0063723704)

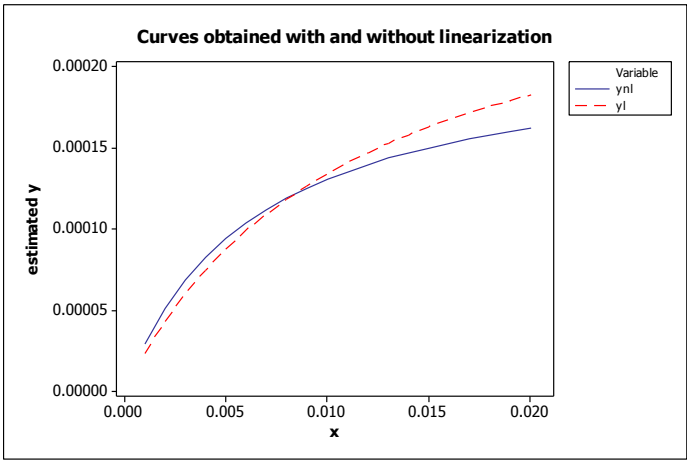


Figure 11. Estimated S_o using linear and non-linear regression

Simulation 3. **Variability inversely proportional to the value of the response variable**.

It was pointed to us that sometimes the larger variation and errors happen in v_o for the low values of S_o, a situation we try to represent in the next simulation (Figure 12A). We were careful to simulate the data so that no value of v_o would be too close to 0 to avoid the problem of a negative estimated value for V_{max}. However, it is clear from Figure 12B that the assumption of constant variance in the transformed variables is not fulfilled.

Row	x	y
1	0.00125	0.0000180
2	0.00125	0.0000363
3	0.00125	0.0000229
4	0.00125	0.0000278
5	0.00125	0.0000627
6	0.00250	0.0000656
7	0.00250	0.0000762
8	0.00250	0.0000530
9	0.00250	0.0000650
10	0.00250	0.0000300
11	0.00500	0.0000687
12	0.00500	0.0001034
13	0.00500	0.0000705
14	0.00500	0.0000700
15	0.00500	0.0000833
16	0.02000	0.0001660
17	0.02000	0.0001574
18	0.02000	0.0001810
19	0.02000	0.0001700
20	0.02000	0.0001543

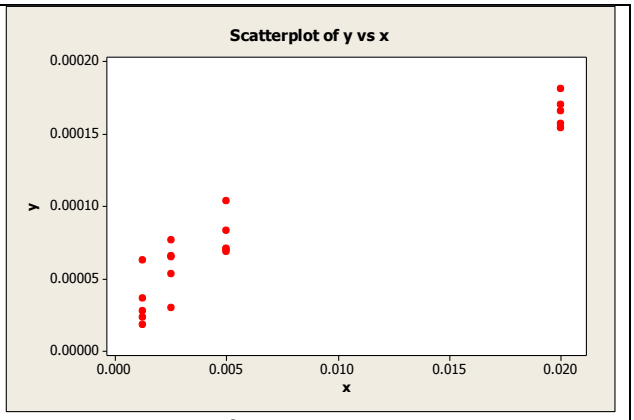


Figure 12A Data from simulation 3

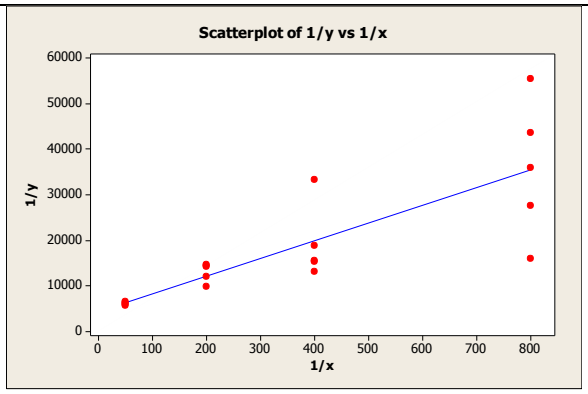


Figure 12B Linear regression on transformed variables

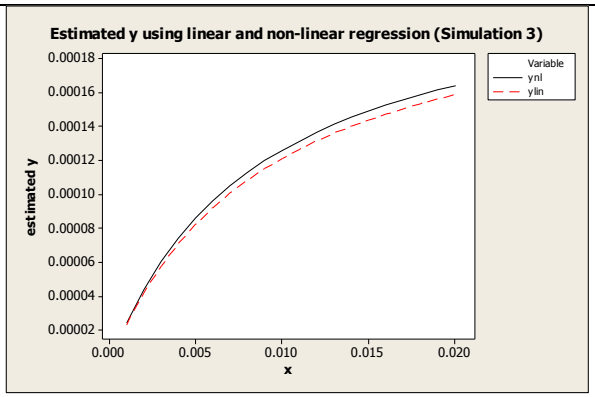


Figure 12C Estimated v_o using non-linear regression and linear regression on transformed variables.

The regression equation on the transformed variables is

$$1/y = 4340 + 39.0\ 1/x$$

Thus, $\hat{V}_{max} = 0.000230415$ and $\hat{K}_m = 0.00898618$

Doing non-linear estimation with **R**, the estimated values of the parameters are:

$\hat{V}_{\max}$ = 0.0002361692 $\hat{K}_m$ =0.0087120580

The estimated values in Figure 12C were calculated as :

Using linearization approach $\hat{y} = \dfrac{0.000230415x}{0.00898618 + x}$

Using the non-linear estimation $\hat{y} = \dfrac{0.0002361692x}{0.0087120580 + x}$

The Hanes linear plot

Next we will use Hanes linear plot. First compare the linear plots for Lineweaver-Burk (Figure 10B), Hanes (Figure 13) and the scatter-plot of v_o versus S_o (Figure 10A) for the second set of simulated data. In Hanes plot the horizontal axis is the same as in the original data and the variability in the response variable is much smaller than in Lineweaver-Burk because $1/v_o$ is being multiplied by S_o and S_o usually takes values smaller than 1. Pearson correlation between S_o/v_o and S_o , the two variables involved in Hanes plot, for this data set is 0.934. A nice feature is that spread or variability in $y = S_o/v_o$ does not look too different for the different values of S_o as it happened in the Lineweaver-Burk plot. The assumption of equal variance is an important one in regression and, working with the variable S_o/v_o as response variable, it is easier to fulfill the equal variance assumption than if we worked with $1/v_o$ as the Lineweaver-Burk method does.

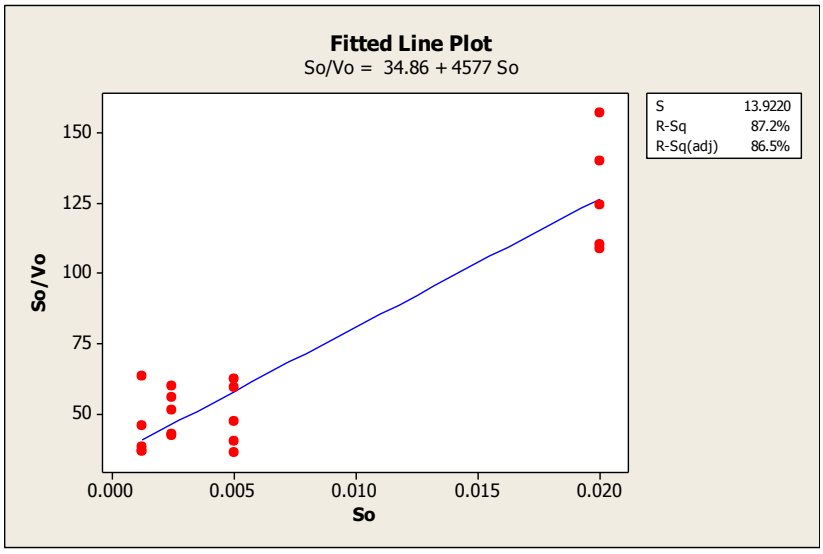


Figure 13. Linear regression using Hanes transformation

Applying Hanes method the Minitab computer output is:

```
The regression equation
So/Vo = 34.9 + 4577 So
```

```
Predictor     Coef  SE Coef      T      P
Constant    34.855    4.306   8.09  0.000
So          4577.3    414.0  11.06  0.000


S = 13.9220   R-Sq = 87.2%   R-Sq(adj) = 86.5%
```

$\hat{V}_m$ =1/4577.3= 0.000218469 and $\hat{K}_m$ =34.855*0=.000218469 = 0.00761475

Another method is Eadie-Hofstee's, which works with v_o and v_o/S_o (Figure 14). We must be aware that v_o is difficult to measure and it appears on both axes of the scatter-plot, so any experimental error will affect both variables. Notice how scattered the dots are in figure 14 . Correlation between v_o and v_o/S_o for this data set is only -0.602.

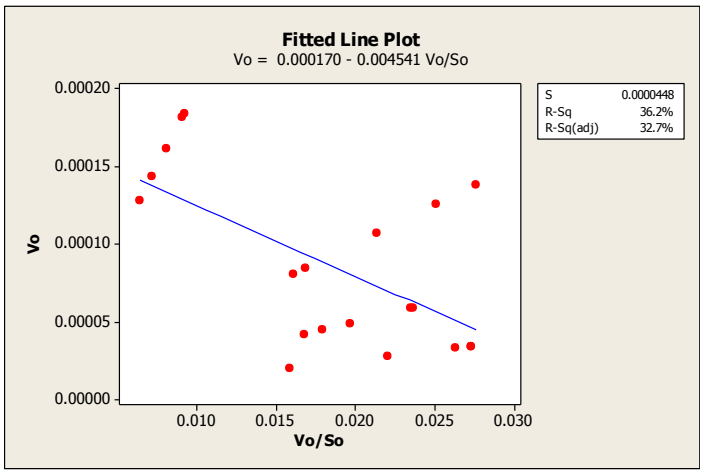


Figure 14. Linear regression using Eadie-Hofstee's transformation

The regression line using Eadie-Hofstee's plot is v_o =0.000170-0.004541 v_o/S_o .
Thus $\hat{V}_{\max} = 0.00017$ and $\hat{K}_m = 0.004541$.

The estimated values through Hanes method are closer to the estimated values obtained through non-linear regression than those obtained with the Lineweaver-Burk or Eadie-Hofstee method. This example illustrates the fact that Hanes method is recommended over Lineweaver-Burk or other linear plots (Cornish-Bowden, 1995). In summary, it is best to use non-linear regression to avoid estimation problems, particularly when some of the observations for v_o (for low values of S_o) are very close to 0 (because a negative estimate for V_{max} might be obtained by linear regression) and when there are replicates and the scatter plots for the transformed variables indicate that the assumption of constant variance is being violated. If software to perform non-linear regression is not available, among linear plots it is best to use Hanes plot but keeping in mind the limitations of all linear plots (Roberts, 1997).

10. Estimation of parameters when data is spread across the steady state

Let us consider the following data from Stern (1936):

Table 2. Stern's data

t (min.)	S(t) ($mol \cdot cm^{-3}$)
0	10.27
3	7.98
6	5.20
9	2.86
12	1.19
15	0.32

Since the data set is given for values of S(t) measured across the steady state, not just at the beginning of the steady state, we will use the Michaelis-Menten integrated form to do a simple regression analysis.

Table 3. Calculations on Stern's data for the integrated form

Row	t	St	So-St	(So-St)/t	ln(So/St)	(1/t)ln(So/St)
1	0	10.27	*	*	*	*
2	3	7.98	2.29	0.763333	0.25229	0.084096
3	6	5.20	5.07	0.845000	0.68057	0.113428
4	9	2.86	7.41	0.823333	1.27841	0.142045
5	12	1.19	9.08	0.756667	2.15527	0.179606
6	15	0.32	9.95	0.663333	3.46866	0.231244

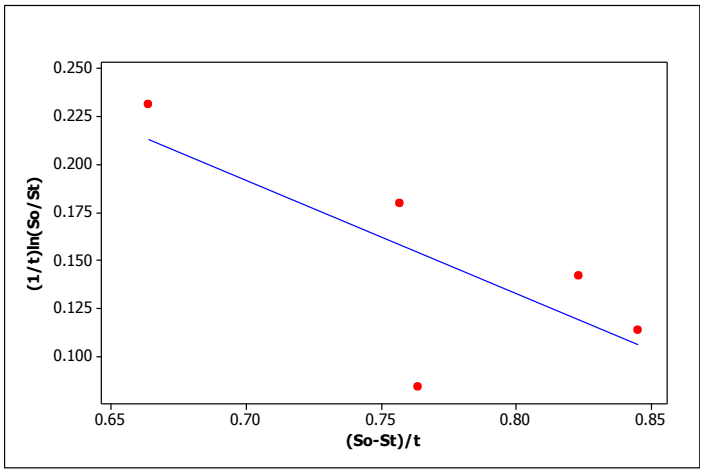


Figure 15. Regression using the integrated form for data in Table 2

Minitab regression output:

```
The regression equation is
(1/t)ln(So/St) = 0.603 - 0.587 (So-St)/t
```

```
Predictor      Coef  SE Coef       T      P
Constant     0.6026   0.2500    2.41  0.095
(So-St)/t   -0.5874   0.3234   -1.82  0.167


S = 0.0458124   R-Sq = 52.4%   R-Sq(adj) = 36.5%
```

$-\frac{1}{\hat{K}_m} = -0.587$ $\qquad \hat{K}_m = 1.70358$

$\hat{V}_{\max} / \hat{K}_m = 0.603$ $\qquad \hat{V}_{\max} = 1.70358 * 0.603 = 1.02726$

The integrated Michaelis-Menten method for calculating in vivo kinetics has been improved by Russell and Drane (1992). They rewrote (6) as

$$S(t) = [S(0) + K_m \ln S(0)] - K_m \ln S(t) - V_{\max} t$$

and then used multiple regression with $\ln S(t)$ and t as the explanatory variables. Let us use this method with Stern's data.

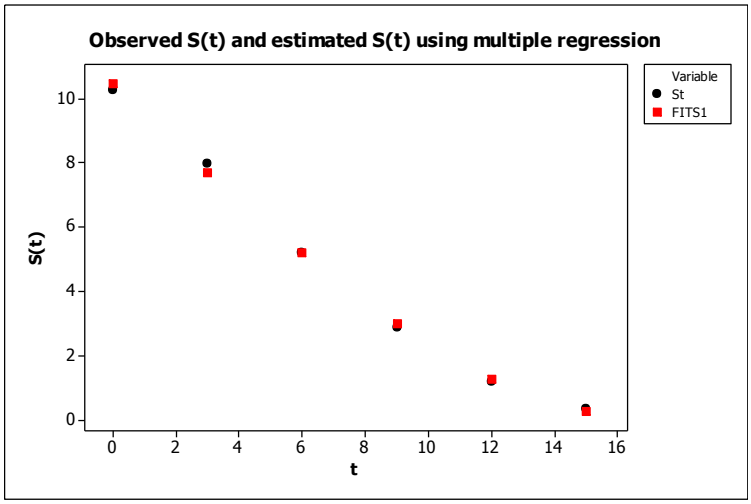


Figure 16. Observed and estimated S(t) using multiple regression

The MINITAB output is:

The regression equation:
St = 14.4 - 1.68 lnSt - 1.07 t

```
Predictor       Coef  SE Coef       T      P
Constant     14.3723   0.7967   18.04  0.000
lnSt         -1.6810   0.2875   -5.85  0.010
t           -1.06872  0.06719  -15.90  0.001
```

```
S = 0.225582   R-Sq = 99.8%   R-Sq(adj) = 99.7%


Analysis of Variance

Source           DF      SS      MS       F      P
Regression        2  76.747  38.373  754.09  0.000
Residual Error    3   0.153   0.051
Total             5  76.899


Source  DF  Seq SS
lnSt     1  63.874
t        1  12.873
```

Thus, the estimated values are $\hat{V}_{\max}$ =1.06872 and $\hat{K}_m$ =1.681

$\hat{V}_{\max}$ and $\hat{K}_m$ are coefficients (or 'slopes') in the multiple regression model and thus they can be read from the regression output directly without any further calculation. That is not true in the case of simple regression. It is to be noted that the adjusted R^2 in the case of simple regression is 36.5% as compared to 99.7% of multiple regression, thus it is advisable to adopt the values provided by the latter, i.e. $\hat{V}_{\max} = 1.06872,\ \hat{K}_m = 1.681.$

Goudar et al. (2004) show how the formula of $S(t)$ in terms of the Lambert W function, namely

$$S(t) = K_m W\left(\frac{S_o}{K_m} e^{\frac{S_o - V_{\max} t}{K_m}}\right)$$

and in conjunction with a nonlinear algorithm for approximation, can be used to estimate $V_{\max}$ and K_m.

Following the framework set by Goudar and his collaborators we wrote a program in MATLAB that includes the command *lambertw*, which provides the value of $W(x)$, and the nonlinear regression command *nlinfit* to apply Goudar's method to Stern's data. We did not use **R** because the Lambert W function has not been written for **R** yet. As starting values we choose $V_{\max}$ = 1.06872 and $K_m = 1.681$, the values gotten through multiple regression, obtaining as a result of the non-linear estimation $\hat{V}_{\max} = 0.9752$ and $\hat{K}_m = 1.2109$. Details about the program can be found in the appendix. The sum of squares of residuals went down from 0.1913 (multiple regression model) to 0.1072 (non-linear model using Lambert W function). Both models are plotted in Figure 17.

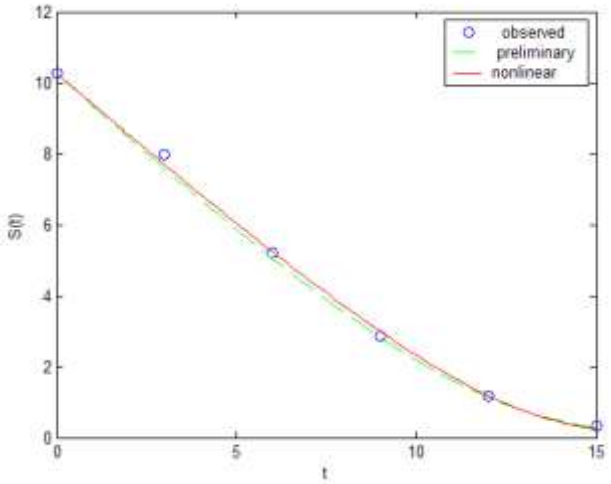


Figure 17. Comparison of observed values and curves estimated with multiple and non-linear regression (using Lambert W function)

Conclusions

We have had the opportunity to work with separable, as well as first and second order linear differential equations. Furthermore, Lambert W function appeared in a natural way when discussing enzyme kinetics under steady state conditions. Lambert W can be considered to be a new "elementary" function since several widely used software packages have incorporated algorithms to calculate it with a high degree of accuracy. Moreover, enzyme kinetics provides the opportunity to stress the importance of making approximations when dealing with differential equations in practical settings. Without approximations it is very difficult to make predictions that can be compared with experimental values

The estimation of $V_{\max}$ and K_m based on experimental data is important for enzyme identification. Several methods are available and the results are similar when experiments are done without replicates and the observations reflect quite strictly the functional relationship between S_o and v_o. However, when there is experimental error, or replicates and variability in v_o, for the same value of S_o results might differ. If observations have been made at the initial stages of the reaction (t close to 0), nonlinear regression (applicable using **R**) works better. If software to do non-linear regression is not available, the regression of S_o/v_o versus S_o (Hanes plot) seems to work better than the other two methods. Non-linear regression using the Lambert W function should be used when observations are done across the steady state. Moreover, multiple regression gives better

results than simple linear regression when software for non-linear regression is not accessible.

References

Bartholomay, A.F. (1972), Chemical Kinetics and Enzyme Kinetics, in *Foundations of Mathematical Biology* (R. Rosen, ed.), Vol. 1, Academic Press, 1972.

Briggs, G.E. and Haldane, J.B.S.(1925), A Note on the Kinetics of Enzyme Action, *Biochemical Journal*, **39**, 338-339.

Cornish-Bowden A. (1995), *Fundamentals of Enzyme Kinetics*, Portland Press, London.

Goudar, C.T., Harris, S.K., McInerney, M.J., Suflita, J.M. (2004), Progress curve analysis for enzyme and microbial kinetic reactions using explicit solutions based on the Lambert W function, *Journal of Microbial Methods*, **59**, 317-326.

Hayes, B. (2005), Why W?, *American Scientist*, 93, 2, 104.

Marangoni, A. (2005), *Enzyme Kinetics*, Wiley-Interscience.

McQuarrie, D.A. and Simon, J.D. (1997), *Physical Chemistry*, University Science Books, Sausalito, California.

Roberts, D.V. (1977), *Enzyme Kinetics*, Cambridge University Press.

Roughton F.J.W. (1954), Rapid Reactions in Biology, *Disc. Faraday Soc*, **17**, 116-120.

Russell, R.W. and Drane, J.W. (1992), Improved Rearrangement of the Integrated Michaelis-Menten Equation for Calculating In Vivo Kinetics of Transport and Metabolism, *Journal of Diary Science*, **75**, No. 12, 3455-3464.

Schnell, S. and Mendoza, C. (1997), Closed Form Solution for Time-dependent Enzyme Kinetics, *Journal of Theoretical Biology*, **187**, 207-212.

Stern, K.G. (1936), A study of the decomposition of monoethyl hydrogen peroxide by Catalase and of an intermediate enzyme-substrate compound, *Journal of Biological Chemistry*, 114, 473-494.

APPENDIX- Matlab commands to estimate $V_{\max}$ and K_m using the Lambert W function

1. Prepare an 'm' file.

Type the following commands using the editor in Matlab and save it as a 'm' file

```
function yhat=STlam(beta,t)
So=10.27;
Kmi=beta(1);
Vmaxi=beta(2);
yhat=Kmi*lambertw((exp((So-Vmaxi*t)./Kmi)).*(So/Kmi));
```

2. **Type the data and initial values of the parameters**

Use as initial values those obtained from the multiple regression or some other source. At the Matlab prompt, type:

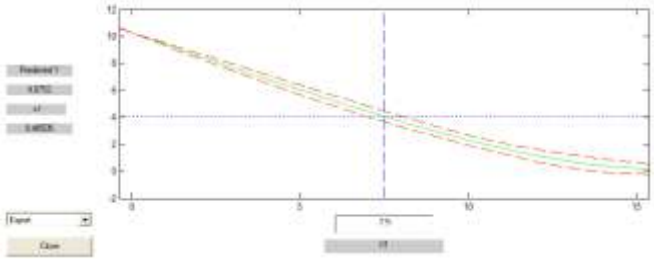
```
global V max Km So
```

```
Vmax=1.06872;Km=1.681;So=10.27;
St= [10.27 ; 7.98  ;5.20 ; 2.86 ; 1.19 ; 0.32];
t=[0 ; 3 ; 6 ; 9 ; 12 ; 15];
beta=[1.681;1.06872];   y=St; x=t;
```

3. **Estimate the parameters using non-linear regression and the W-function**
At the Matlab prompt, type

```
nlintool(x,y,@STlam,beta)
```

The plot of the estimated model will appear on the screen



Click on 'export' and either select all or select certain items. Then typing their names, the output will be shown on the screen. For example, to obtain the estimated parameters type *betafit* and the output will appear on the screen.

```
betafit =
   1.2110
   0.9752
```