LINEAR FUNCTIONS: A PRIMER

MATH 196, SECTION 57 (VIPUL NAIK)

Executive summary

Words ...

(1) Linear models arise both because some natural and social phenomena are intrinsically linear, and because they are computationally tractable and hence desirable as approximations, either before or after logarithmic and related transformations.
(2) Linear functions (in the affine linear sense) can be characterized as functions for which all the second-order partial derivatives are zero. The second-order pure partial derivatives being zero signifies linearity in each variable holding the others constant (if this condition is true for each variable separately, we say the function is (affine) multilinear). The mixed partial derivatives being zero signifies *additive separability* in the relevant variables. If this is true for every pair of input variables, the function is completely additively separable in the variables.
(3) We can use logarithmic transformations to study multiplicatively separable functions using additively separable functions. For a few specific functional forms, we can make them linear as well.
(4) We can use the linear paradigm in the study of additively separable functions where the components are known in advance up to scalar multiples.
(5) If a function type is linear in the parameters (not necessarily in the input variables) we can use (input,output) pairs to obtain a system of linear equations in the parameters and determine the values of the parameters. Note that a linear function of the variables with no restrictions on the coefficients and intercepts must also be linear in the parameters (with the number of parameters being one more than the number of variables). However, there are many nonlinear functional forms, such as polynomials, that are linear in the parameters but not in the variables.
(6) Continuing the preceding point, the number of well-chosen (input,output) pairs that we need should be equal to the number of parameters. Here, the "well-chosen" signifies the absence of dependencies between the chosen inputs. However, choosing the bare minimum number does not provide any independent confirmation of our model. To obtain independent confirmation, we should collect additional (input,output) pairs. The possibility of modeling and measurement errors may require us to introduce error-tolerance into our model, but that is beyond the scope of the current discussion. We will return to it later.

Actions ...

(1) One major stumbling block for people is in writing the general functional form for a model that correctly includes parameters to describe the various degrees of freedom. Writing the correct functional form is half the battle. It's important to have a theoretically well-grounded choice of functional form and to make sure that the functional form as described algebraically correctly describes what we have in mind.
(2) It's particularly important to make sure to include a parameter for the *intercept* (or *constant term*) unless theoretical considerations require this to be zero.
(3) When dealing with polynomials in multiple variables, it is important to make sure that we have accounted for all possible monomials.
(4) When dealing with piecewise functional descriptions, we have separate functions for each piece interval. We have to determine the generic functional form for each piece. The total number of parameters is the sum of the number of parameters used for each of the functional forms. In particular, if the nature of the functional form is the same for each piece, the total number of parameters is (number of parameters for the functional form for each piece) $\times$ (number of pieces).

1. A brief summary that reveals nothing

In the natural and social sciences, "linearity" is a key idea that pops up repeatedly in two ways:

(1) A lot of natural and social phenomena are intrinsically linear, i.e., the mathematical formulations of these involve linear functions.
(2) A lot of natural and social phenomena allow for linear approximation. Even though the actual description of the phenomenon is nonlinear, the linear approximation serves well for descriptive and analytic purposes.

Understanding how linear structures behave is thus critical to understanding many phenomena in the natural and social sciences, and more to the point, critical to understanding the mathematical models that have (rightly or wrongly) been used to describe these phenomena.

2. Additively separable, multilinear, and linear

2.1. Terminology: linear and affine.

2.1. Terminology: linear and affine. A function f of one variable x is termed linear if it is of the form $x \mapsto mx + c$ for real numbers m (the slope) and c (the intercept). In later usage, we will sometimes restrict the term "linear" to situations where the intercept is zero. To make clear that we are taking of the more general version of linear that allows a nonzero intercept, it is better to use the term "affine" or "(affine) linear." The choice of word may become clearer later in the course. In contrast, "homogeneous linear" explicitly signifies the absence of a constant term.

2.2. The case of functions of two variables. Consider a function F of two variables x and y.

We say that F is:

- *additively separable* if we can find functions f and g such that $F(x, y) = f(x) + g(y)$. Under suitable connectedness and differentiability assumptions, this is equivalent to assuming that F_{xy} is identically zero. The differentiability assumption is necessary to make sense of F_{xy}, and the connectivity assumption is necessary in order to argue the converse (i.e., going from $F_{xy} = 0$ to F being additively separable).

  Conceptually, additive separability means that the variables do not "interact" in the function. Each variable produces its own contribution, and the contributions are then pooled together. In the polynomial setting, this would mean that there are no polynomials that are mixed products of powers of x with powers of y.
- *(affine) linear in x* if, for each fixed value $y = y_0$, the function $x \mapsto F(x, y_0)$ is a linear function of x. Under reasonable connectivity assumptions, this is equivalent to assuming that F_{xx} is the zero function.

  What this means is that once we fix y, the function has constant returns in x. The graph of the function restricted to $y = y_0$ (so it is now just a function of x) looks like a straight line.
- *(affine) linear in y* if, for each fixed value $x = x_0$, the function $y \mapsto F(x_0, y)$ is a linear function of y. Under reasonable connectivity assumptions, this is equivalent to assuming that F_{yy} is the zero function.

  What this means is that once we fix x, the function has constant returns in y. The graph of the function restricted to $x = x_0$ (so it is now just a function of y) looks like a straight line.
- *(affine) multilinear* if it is linear in x and linear in y.

  This means that the function has constant returns in each variable individually. However, it does not mean that the function has constant returns in the variables together, because the returns to one variable may be influenced by the value of the other.
- *(affine) linear* if it is both (affine) multilinear and additively separable. Under connectivity and differentiability assumptions, this is equivalent to saying that *all* the second-order partial derivatives are zero.

  Additive separability basically guarantees that the returns on one variable are not affected by the other, and therefore, the function must have constant returns in the variables combined.

  F must be of the form $F(x, y) = ax + by + c$ with a, b, c real numbers. The graph of such a function is a plane. The values a and b are the "slopes" in the respective variables x and y and the value c is the intercept. If we are making the graph $z = F(x, y)$, the intersection of the graph (a plane) with the z-axis gives the value of the intercept.

Here are some examples. Note that when we say "linear" below we mean affine linear:

Function	Additively separable?	Linear in x?	Linear in y	Multilinear?	Linear?
$x^2 - \sin y + 5$	Yes	No	No	No	No
xe^y	No	Yes	No	No	No
$(x+1)(y-1)$	No	Yes	Yes	Yes	No
$\cos x + y - 1$	Yes	No	Yes	No	No
$2x + 3y - 4$	Yes	Yes	Yes	Yes	Yes

2.3. **Extending to multiple variables.** A function F on n variables $x_1, x_2, \dots, x_n$ is termed affine linear if there exist real numbers $a_1, a_2, \dots, a_n, c$ such that we can write:

$$F(x_1, x_2, \dots, x_n) := a_1x_1 + a_2x_2 + \cdots + a_nx_n + c$$

The case $c = 0$ is of particular interest because it means that the origin gets sent to zero. We shall talk in more detail about the significance of that condition later.

Once again, linear functions of multiple variables can be broken down in terms of the following key attributes:

- Additive separability in every pair of variables, and in fact, additive separability "overall" in all the variables. This is equivalent (under various connectedness and differentiability assumptions) to the second-order mixed partial derivative in every pair of variables being zero.
- Multilinearity: For each variable, the function is (affine) linear with respect to that variable, holding other variables constant. This is equivalent (under various connectedness and differentiability assumptions) to the second-order pure partial derivative in every variable being zero.

We could have functions that are additively separable and not multilinear. This means that the function is a sum of separate functions of the individual variables, but the separate functions are not all linear. We could also have functions that are multilinear but not additively separable. For instance, $x_1x_2 + x_2x_3 + x_1x_2x_3$ is (affine) multilinear but not additively separable.

2.4. **Real-world importance.** Both additive separability and multilinearity are simplifying assumptions with real-world significance. Suppose we are trying to model a real-world production process that depends on the labor inputs of two individuals. Let x and y be the number of hours each individual devotes to the production process. The output is a function $F(x, y)$.

The additive separability assumption would mean that the inputs of the two workers do not interact, i.e., the hours put in by one worker do not affect the marginal product per hour of the other worker. The multilinearity assumption says that holding the hours of the other worker constant, the output enjoys constant returns to hours for each worker. If we assume both of these, then F is linear, and we can try to model it as:

$$F(x, y) = ax + by + c$$

with a, b, c as real numbers to be determined through empirical observation. c in this situation represents the amount produced if neither individual puts in any work, and we may take this to be 0. a and b represent the "productivity" values of the two respective workers.

Note that both additive separability and the multilinearity are pretty strong assumptions that are not usually satisfied. In the production function context:

- The two alternatives to additive separability (no interaction) are *complementarity* (positive second-order mixed partial) and *substitution* (negative second-order mixed partial).
- The two alternatives to linearity (constant returns) in a particular variable are *increasing returns* (positive second-order pure partial) and *decreasing returns* (negative second-order pure partial).

3. Situations reducible to the linear situation

3.1. Multiplicatively separable functions. If linearity requires such strong assumptions in even simple contexts, can we really expect it to be that ubiquitous? Not in a literal sense. However, some very specific nonlinear functional forms can be changed to linear functional forms by means of the logarithmic transformation.

We say that a function $G(x, y)$ of two variables x and y is *multiplicatively separable* if $G(x, y) = f(x)g(y)$ for suitable functions f and g. Assuming that G, f, and g are positive on our domain of interest, this can be rewritten as:

$$\ln(G(x, y)) = \ln(f(x)) + \ln(g(y))$$

Viewed *logarithmically*, therefore, multiplicatively separable becomes additively separable. Of course, as noted above, additive separability is not good enough for linearity.

Of particular interest are functions of the form below, where c, x, and y are positive:

$$G(x, y) = cx^a y^b$$

Note that *Cobb-Douglas production functions* are of this form.

Taking logarithms, we get:

$$\ln(G(x, y)) = \ln c + a \ln x + b \ln y$$

Thus, $\ln(G)$ is a linear function of $\ln x$ and $\ln y$. Taking logarithms on everything, we thus get into the linear world.

3.2. Additively separable situation where the component functions are known in advance up to scalar multiples. Suppose $f_1, f_2, \dots, f_n$ are known functions, and we are interested in studying the possibilities for a function F of n variables of the form:

$$F(x_1, x_2, \dots, x_n) = a_1 f_1(x_1) + a_2 f_2(x_2) + \dots + a_n f_n(x_n) + c$$

The key is that $f_1, f_2, \dots, f_n$ are known in advance. In this situation, we can "replace" $x_1, x_2, \dots, x_n$ by $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$, treating the latter as the new variables. With these as the new inputs, the output is a linear function, and all the techniques of linearity apply.

4. The computational attraction of linearity

The discussion so far has concentrated on why linearity is a powerful conceptual assumption, and also on why that assumption may or may not be justified. This also gives us some insight into why linear functions might be computationally easier to handle. It's now time to look more explicitly at the computational side.

The introduction to your textbook (Bretscher) mentions the *streetlight strategy*: a person lost his keys, and was hunting for them under the streetlight, not because he thought that was the most likely location he might have dropped them, but because the light was good there, so he had a high chance of finding the keys if indeed the keys were there. Although this is often told in the form of a joke, the streetlight strategy is a reasonable strategy and is not completely baseless. Linear algebra is a very powerful streetlight that illuminates linear functions very well. This makes it very tempting to try to use linear models. To an extent, the best way to deal with this temptation is to yield to it. But not completely. Naive "linear extrapolation" can have tragic consequences.

I'd love to say more about this, but that's what your whole course is devoted to. I will briefly allude to some key aspects of linearity and its significance.

4.1. Parameter determination. So, you've decided on a particular model for a phenomenon. Perhaps, you've decided that the amount produced is a linear function of the hour counts x and y of the form:

$$F(x, y) = ax + by + c$$

The problem, though, is that you don't yet know the values of a, b, and c, the *parameters* in the model. The model is too theoretical to shed light on these values.

In order to use the model for numerically accurate predictions, you need to figure out the values of the parameters. How might you do this? Let's think more clearly. The simplest rendering of a function is:

$$\text{Input variables} \stackrel{\text{the function}}{\rightarrow} \text{Output variables}$$

Now, however, if only a *generic form* for the function is known, and the values of some parameters are not yet known, then the "machine" needs to be fed with both the input variables and the parameter values. We can think of this as:

$$\text{Input variables} + \text{Parameter values} \stackrel{\text{generic form}}{\rightarrow} \text{Output variables}$$

Now, suppose we know the values of the outputs for some specific choices of inputs. Each such piece of information gives an equation in terms of the parameter values. With enough such equations, we hope to have enough information to deduce the parameter values.

As a general rule, keep in mind that:

Dimension of solution space to constraint set = (Dimension of original space) - (Number of independent constraints)

In other words:

Dimension of solution space = (Number of variables) - (Number of equations)

In our case, we are trying to solve equations in the parameters, and we want a "zero-dimensional" solution space, i.e., we want to determine the parameter uniquely. Thus, what we want is that:

0 = (Number of parameters) - (Number of input-output pairs)

Note here that the parameters become the variables that we are trying to find, but they are different from what we usually think of as the variables, namely the inputs to the function. This could be a source of considerable confusion for people, so it's important to carefully understand this fact.

Note also something about the jargon of *input-output pair*: a single input-output pair includes the values of each of the inputs, plus the output. If the function has n inputs and 1 output, an "input-output pair" would be a specification of $n + 1$ values, including n inputs and 1 output.

Once we have found the values of the parameters, we treat them as those constant values and no longer treat them as variables. Our treating them as variables is a provisonal step in finding them.

In other words, in order to determine the parameters, we need to have as many input-output pairs as the number of parameters.

However, if we have exactly as many input-output pairs as the number of parameters, we will get the parameters but we have no additional confirmation, no sanity check, that our model is indeed correct. In order to provide additional confirmation, it will be necessary to have *more* input-output pairs than the number of parameters that need to be determined. If this "overdetermined" system still gives unique solutions, then indeed this provides some confirmation that our model was correct.

A number of issues are not being addressed here. The most important is the issue of modeling errors and measurement errors. The former refers to the situation where the model is an approximation and not exactly correct, while the latter refers to a situation where the input-output pairs are not measured with full accuracy and precision. Usually, both errors are present, and therefore, we will not expect to get exact solutions to overdetermined systems. We need to adopt some concept of error-tolerance and provide an appropriate mathematical formalism for it. Unfortunately, that task requires a lot more work, so we shall set it aside for now.

Instead, let us return to the question of parameter determination, assuming that everything works exactly. Consider our model:

$$F(x, y) = ax + by + c$$

This model has two inputs (x and y) and one output ($F(x, y)$), and it has three parameters a, b, and c. Specifying an input-output pair is equivalent to specifying the output $F(x_0, y_0)$ for a particular input $x = x_0, y = y_0$.

Right now, our goal is to determine the parameters a, b, and c. What we need are specific observations of input-output pairs. Suppose we have the following observations:

$$
\begin{aligned}
F(1,2) &= 12 \\
F(2,3) &= 17 \\
F(3,5) &= 25
\end{aligned}
$$

These observations may have been gathered empirically: these might have been the number of hours put in by the two workers, with the corresponding outputs, in past runs of the production process.

We now plug in these input-output pairs and get a system of linear equations:

$$
\begin{aligned}
a + 2b + c &= 12 \\
2a + 3b + c &= 17 \\
3a + 5b + c &= 25
\end{aligned}
$$

Note that this is a system of linear equations *in terms of the parameters*, i.e., the variables for this system of linear equations are the parameters of our original functional form.

If we solve the system, we will get the unique solution $a = 2$, $b = 3$, $c = 4$.

This gives us the parameters, assuming we are absolutely sure our model is correct. What, however, if we want independent confirmation? In that case, we'd like another data point, i.e., another input-output pair, that agrees with these parameter values. Suppose we were additionally given the data point that $F(1,1) = 9$. This would be a validation of the model.

Suppose, however, that it turns out that $F(1,1) = 10$. That would suggest that the model is wrong. Could it be rescued partially? Yes, it might be rescuable if we assume some errors and approximations in our modeling and measurement process. On the other hand, an observation like $F(1,1) = 40$ (if actually correct and not simply a result of a "typo" or other weird measurement error) would likely mean that the model is close to useless.

We needed three input-output pairs to determine the parameters uniquely because there are three parameters in the generic form. To provide additional confirmation, we need more input-output pairs.

4.2. **Linear in parameters versus linear in variables.** There is a subtle but very important distinction to keep in mind. When we use input-output pairs to form equations in the parameters that we then solve, the nature of those equations depends, obviously, on the nature of the functional form. However, it depends on the way the function depends on the *parameters*, not on the way the function depends on the *input variables*. For instance, consider a function:

$$F(x, y) := axy + be^{x \sin y} + cx^2y^2$$

This function is definitely not linear in x or y, nor is it additively separable. However, the function is linear in the parameters a, b, and c. Thus, the system of equations we set up using (input, output) pairs for the function is a system of linear equations.

If a function is linear in the inputs, then assuming no nonlinear constraints relating the parameters, it is also linear in the parameters (note that the number of parameters is one more than the number of variables, and that's the number of observations we need to determine the parameters uniquely). However, as observed above, it is very much possible to be linear in the parameters but not in the variables. For this reason, linear algebra methods for parameter estimation can be applied to some functional forms that are not linear in the input variables themselves.

4.3. **Generating equations using methods other than input-output values.** The typical method used to generate equations for parameter determination is input-output value pairs. There are, however, other methods, based on other pieces of information that might exist about the function.

These may include, for instance, (input, derivative of output) pairs, or (average of inputs, average of outputs) pairs, or other information.

The case of the differential equation, which we might return to later, is illustrative. The *initial value specification* used for the solution of a differential equation usually involves an input value and the value of the output and many derivatives of the output at a single point.

4.4. **Once the parameters are determined.** Once the parameters for the model have been determined, we can proceed to use the usual methods of multivariable calculus to acquire a deeper understanding of what the function does. If the function is also linear in the input variables, then we can use the techniques of linear algebra. Otherwise, we are stuck with general techniques from multivariable calculus.

5. Creating general functional forms and counting parameters

One major stumbling block for people is in writing the general functional form for a model that correctly includes parameters to describe the various degrees of freedom. Writing the correct functional form is half the battle. It's important to have a theoretically well-grounded choice of functional form and to make sure that the functional form as described algebraically correctly describes what we have in mind.

Note also that it's important to maintain mentally the distinction between the number of parameters and the number of inputs.

5.1. **Some examples for functions of one variable and multiple variables.**

- A polynomial function $f(x)$ of one variable x that is given to have degree $\leq n$: The general functional form is:

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

  The number of parameters here is $n+1$. Therefore, the number of well-chosen input-output pairs we expect to need is $n+1$. In fact, in this case, choosing $n+1$ inputs always allows us to determine the polynomial functon uniquely. This is related to the Lagrange interpolation formula and also to the idea of the Vandermonde matrix and Vandermonde determinant. Note that in order to obtain additional confirmation of the model, we need to have $n+2$ or more input-output pairs.
- A polynomial function $f(x, y)$ of two variables x and y, of total degre $\leq n$.

  The polynomial is obtained by taking linear combinations of monomials of the form x^iy^j where $i + j \leq n$. The number of such monomials depends on n, and the process involves just writing out all the monomials. For instance, in the case $n = 2$, the monomials are 1, x, y, x^2, xy, y^2, and the generic functional form is:

$$a_1 + a_2x + a_3y + a_4x^2 + a_5xy + a_6y^2$$

  The number of parameters is 6 in this case ($n = 2$), so we need 6 well-chosen inputs to find the parameters, and we need 7 or more to both find the parameters and obtain independent confirmation of the model.

  In general, the number of parameters is $(n+1)(n+2)/2$. This formula is not obvious, but we can work it out with some effort. Rather than memorize the formula, try to understand how we would go about writing the generic functional form.
- A polynomial function of multiple variables: The idea of the preceding point generalizes, but the number of parameters now grows considerably. In general, it grows both with the number of variables and with the degree of the polynomial that we are trying to use.
- A trigonometric function of the form $a \sin x + b \cos x + C$. This has three parameters a, b, and C. If $C = 0$ for theoretical reasons, then we have only two parameters.
- A sum of trigonometric functions, such as:

$$f(x) := a_1 \sin(m_1x) + a_2 \sin(m_2x)$$

  This is not linear in the parameters m_1 and m_2, but it is linear in the parameters a_1 and a_2. If we already know m_1 and m_2, we can use (input,output) pairs to find a_1 and a_2.

  This type of function can arise when considering musical sounds that involve multiple frequencies being played simultaneously, as is common with many musical instruments (fundamental and other modes of vibration for guitar strings, for instance, and the use of chords in music, etc.)

- A sum of exponential functions, such as:

$$f(x) := a_1 e^{m_1 x} + a_2 e^{m_2 x}$$

  This is not linear in the parameters m_1 and m_2, but it is linear in the parameters a_1 and a_2. If we already know m_1 and m_2, we can use (input,output) pairs to find a_1 and a_2.

  This type of situation arises when multiple exponential trends arising from different causes or sources get combined.

5.2. **Examples involving piecewise descriptions.** In some cases, instead of assuming a single functional form throughout the domain, we assume a function with a piecewise description, i.e., the function differs on different pieces of the domain. For functions of one variables, these pieces could be intervals. For functions of two variables x and y, these pieces could be rectangular regions, or otherwise shaped regions, in the xy-plane (where the domain of the function resides).

The function description on each piece involves a functional form that has some parameters. The total number of parameters needed for the functional description overall is the sum of the number of parameters needed on each piece. If the functional form is the same on all pieces, then the number of parameters needed is (number of parameters needed for the functional form) $\times$ (number of pieces).

For instance, consider a function of one variable that has a piecewise quadratic description with the domain $[0, 4]$ having four pieces: $[0, 1]$, $[1, 2]$, $[2, 3]$, and $[3, 4]$. For each piece, the functional form needs 3 parameters (on account of being quadratic). There are four such pieces, so we need a total of $3 \times 4 = 12$ parameters. If we assume continuity at the transition points 1, 2, and 3, then each of the continuity conditions gives an equation relating the descriptions, so we effectively have $12 - 3 = 9$ degrees of freedom.

In this case, a particular (input,output) pair would be helpful in so far as it helps pin down the parameters for the interval where the input lives. For inputs that are transition points, we can use its containment in either interval to generate an equation.

There is a lot more that can be said here, but since this topic is relatively tangential to the course as a whole, we will stop here. It may be beneficial to look at the relevant quiz and homework problems for a more detailed discussion of some variations of this setup.