# HYPOTHESIS TESTING, RANK, AND OVERDETERMINATION

MATH 196, SECTION 57 (VIPUL NAIK)

## Executive summary

Words ...

(1) In order to test a hypothesis, we must conduct an experiment that could conceivably come up with an outcome that would falsify the hypothesis. This relates to Popper's notion of falsifiability.
(2) In the setting where we use a model with a functional form that is linear in the parameters, the situation where the coefficient matrix (dependent on the inputs) does *not* have full row rank is the situation where we can use consistency of the system to obtain additional confirmation of the hypothesis that the model is correct. If the coefficient matrix has full column rank, we can determine the parameters uniquely assuming consistency. The ideal situation would be that we choose inputs such that the coefficient matrix has full column rank but does not have full row rank. In this situation, we can obtain verification of the hypothesis *and* find the parameter values.
(3) In order to test a hypothesis of a function of multiple variables being affine linear, we could choose three points that are collinear in the input space and see if the outputs behave as predicted. If they do, then this is evidence in favor of linearity, but it is not conclusive evidence. If they do not, this is conclusive evidence against linearity.
(4) If the goal is to find the coefficients rather than to test the hypothesis of linearity, we should try picking independent inputs (in general, as many inputs as the number of parameters, which, for an affine linear functional form, is one more than the number of variables). Thus, the choice of inputs differ for the two types of goals. If, however, we are allowed enough inputs, then we can both find all the coefficients *and* test for linearity.

## 1. Testing the hypothesis of linearity: the case of functions of one variable

We begin by understanding all the ideas in the context of a function of one variable. We will then proceed to look at functions of more than one variable.

1.1. **Preliminary steps.** Suppose $f$ is a function of one variable $x$, defined for all reals. Recall that $f$ is called a *linear* function (in the affine linear, rather than the homogeneous linear, sense) if there exist constants $m$ and $c$ such that $f(x) = mx + c$. Suppose the only thing you have access to is a black box that will take an input $x$ and return the value $f(x)$. You have no information about the inner working of the program.

You do not know whether $f$ is a linear function, but you believe that that may be the case.

(1) Suppose you are given the value of $f$ at one point. This is useful information, but does not, in and of itself, shed any light on whether $f$ is linear.
(2) Suppose you are given the values of $f$ at two points. This would be enough to determine $f$ uniquely under the assumption that it is linear. However, it will not be enough to provide evidence in favor of the hypothesis that $f$ is linear. Why? Because *whatever* the values of the function, we can always fit a straight line through them.
(3) Suppose you are given the values of $f$ at three points. This allows for a serious test of the hypothesis of linearity. There are two possibilities regarding how the three points on the graph of $f$ look:
   - The points on the graph are non-collinear: This is *conclusive* evidence *against* the hypothesis of linearity. We can definitely *reject* the hypothesis that the function is linear.
   - The points on the graph are collinear: This is evidence *in favor of* the hypothesis of linearity, but it is not conclusive. We can imagine a function that is not linear but such that the outputs

for the three specific inputs that we chose happened to fall in a straight line. As we will discuss below, there are good reasons to treat the collinearity of points on the graph as *strong evidence* in favor of linearity. But it is not conclusive. For instance, for the function $f(x) = x^3 + x$, the points $x = -1$, $x = 0$, and $x = 1$ appear to satisfy the collinearity condition, despite the function not being linear.

(4) Suppose you are given more than three points and the values of $f$ at all these points. This allows for an even stronger test of the hypothesis of linearity. There are two possibilities regarding how the corresponding points on the graph of the function look:
- The points on the graph are not all collinear: This is *conclusive* evidence *against* the hypothesis of linearity. We can definitely *reject* the hypothesis that the function is linear.
- The points on the graph are collinear: This is evidence *in favor of* the hypothesis of linearity, but is not conclusive. After all, we can draw curves that are not straight lines to fill in the unknown gaps between the known points on the graph. However, it does constitute strong evidence in favor of the linearity hypothesis. And the more points we have evaluated the function at, the stronger the evidence.

1.2. **Popper's concept of falsifiability.** Karl Popper, a philosopher of science, argued that scientific knowledge was built on the idea of *falsifiability*. A theory is falsifiable if we can envisage an experiment with a possible outcome that could be used to definitively show the theory to be false. Popper took the (arguably extreme) view that:

- A scientific theory *must* be falsifiable, i.e., until we come up with a possible way to falsify the theory, it isn't a scientific theory.
- A scientific theory can never be demonstrated to be true. Rather, as we construct more and more elaborate ways of trying to falsify the theory, and fail to falsify the theory for each such experiment, our confidence in the theory gradually increases. We can never reach a stage where we are *absolutely* sure of the theory. Rather, we become progressively more and more sure as the theory resists more and more challenges.

Popper's stringent criteria for what theories count as scientific have been subjected to much criticism. We do not intend to take a side in the debate. Rather, we will make the somewhat more elementary point at the heart of Popper's reasoning: *an experimental outcome can be viewed as evidence in favor of a theory only if there was some alternative outcome that would have caused one to reject (definitely or probabilistically) the theory.* Another way of putting this is that if the opposition to a theory isn't given a serious chance to make its case, the theory cannot be declared to have won the debate.[1] Further, the fairer and more elaborate the chance that is given to the opposition, the more this should boost confidence in our theory. Also, for any theory that makes assertions about infinite sets where only finitely many things can be checked at any given time, the asymmetry alluded to above exists: it may be possible to falsify the theory, but one can never be absolutely sure of the truth of the theory, though as we check more and more points and continue to fail to falsify the theory, our confidence in the theory improves.

Consider the preceding example where we are testing the hypothesis that a given function of one variable is linear. Replace the term "theory" by the term "hypothesis" in the preceding paragraph to better understand the subsequent discussion. Our hypothesis on the function $f$ is that $f$ is linear. The "experiments" that we can perform involve collecting the values of $f$ at finitely many known inputs.

As per the preceding discussion:

(1) If we evaluate $f$ at only one input, the information at hand is insufficient to come to any conclusions regarding whether $f$ is linear. Another way of putting it is that we have had no opportunity to falsify the linearity hypothesis with just one point.
(2) The same holds if we evaluate $f$ at only two inputs.
(3) If we evaluate $f$ at three inputs, we have a serious opportunity to falsify the linearity hypothesis. Thus, this is the first serious test of the linearity hypothesis. Two cases:
- The points on the graph are non-collinear: This definitively falsifies the linearity hypothesis.

[1]You could argue that there are cases where the opposition doesn't exist at all, and the theory is *prima facie* true. But insofar as there is doubt about the truth of the theory, the point made here stands.

- The points on the graph are collinear: This is *consistent* with the linearity hypothesis, hence does not falsify it. It is therefore evidence in favor of the linearity hypothesis (note that it is evidence in favor because there was *potential* for the evidence to go the other way).

(4) Similar remarks apply to the evaluation of $f$ at more than three points. Two cases:
- The points on the graph are not all collinear: In this case, the linearity hypothesis is definitively rejected, i.e., falsified.
- The points on the graph are all collinear: This is consistent with the linearity hypothesis, and can be construed as evidence in favor of the hypothesis. The greater the number of points, the stronger the evidence in favor of the linearity hypothesis.

1.3. **Interpretation of the above in terms of rank of a linear system.** When fitting a linear model $f(x) = mx + c$ using input-output pairs, we use the input-output pairs to generate a system of simultaneous linear equations in terms of the parameters $m$ and $c$ (these parameters become our "variables" for the purpose of solving the system). Explicitly, for each input-output pair $(x_i, y_i)$ (with $y_i = f(x_i)$) we get a row of the following form in the augmented matrix:

$$\left[\begin{array}{cc|c} 1 & x_i & y_i \end{array}\right]$$

where the first column corresponds to the parameter $c$ and the second column corresponds to the parameter $m$. Note that we choose the first column for $c$ because using this ordering makes the process of computing the reduced row-echelon form easier. The row for the coefficient matrix reads:

$$\begin{bmatrix} 1 & x_i \end{bmatrix}$$

We can now formulate the earlier results in terms of ranks of matrices:

(1) If we evaluate $f$ at only one input, the information at hand is insufficient to come to any conclusions regarding whether $f$ is linear. It is also insufficient to determine $f$ even assuming $f$ is linear. Another way of putting it is that we have had no opportunity to falsify the linearity hypothesis with just one point. If the input-output pair is $(x_1, y_1)$, we obtain the augmented matrix:

$$\left[\begin{array}{cc|c} 1 & x_i & y_1 \end{array}\right]$$

The rank of the coefficient matrix is 1. Note that the system has *full row rank* and is therefore consistent by definition (therefore, it cannot be used to falsify the hypothesis of linearity). The system does not have full column rank, so it does not give a unique solution even if we assume linearity.

(2) If we evaluate $f$ at two inputs, we obtain $f$ uniquely subject to the assumption that it is linear, but we do not obtain any verification of the linearity of $f$. Explicitly, if $(x_1, y_1)$ and $(x_2, y_2)$ are the input-output pairs, then the augmented matrix (with the first column for the variable $c$ and the second column for the variable $m$; remember that we use this ordering to make the matrix easier to convert to rref) is:

$$\left[\begin{array}{cc|c} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{array}\right]$$

The rank of the coefficient matrix is 2. Note that the system has *full row rank* and is therefore consistent by definition. The system also has *full column rank*. This, along with consistency, implies a unique solution. The "unique solution" here refers to a unique straight line passing through the two points. Note, however, that since the coefficient matrix has full row rank, the system has no potential to be inconsistent regardless of the choice of outputs. Therefore, since there is no potential for falsification of the linearity hypothesis, we cannot use this as evidence in favor of the linearity hypothesis.

(3) If we evaluate $f$ at three inputs, we have a serious opportunity to falsify the linearity hypothesis. Thus, this is the first serious test of the linearity hypothesis. Suppose the three input-output pairs are $(x_1, y_1)$, $(x_2, y_2)$, and $(x_3, y_3)$. The augmented matrix is:

$$\begin{bmatrix} 1 & x_1 & | & y_1 \\ 1 & x_2 & | & y_2 \\ 1 & x_3 & | & y_3 \end{bmatrix}$$

The coefficient matrix has rank 2. Therefore it has full column rank but does not have full row rank. What this means is that *if* a solution exists, the solution is unique (geometrically, if a line passes through the three points, it is unique) but a solution need not exist.

There are two cases:

- The points on the graph are non-collinear. This corresponds to the case that the system of equations is inconsistent, or equivalently, when we row reduce the system, we get a row with zeros in the coefficient matrix and a nonzero augmenting entry. This definitively falsifies the linearity hypothesis.
- The points on the graph are collinear. This corresponds to the case that the system of equations is consistent. This is *consistent* with the linearity hypothesis, hence does not falsify it. There was *potential* for falsification, so this is therefore evidence in favor of the linearity hypothesis.

(4) Similar remarks apply to the evaluation of $f$ at more than three points. The coefficient matrix now has $n$ rows and rank 2. It therefore has full column rank (so unique solution if there exists a solution) but does not have full row rank (so that the system may be inconsistent). Explicitly, the augmented matrix is of the form:

$$\begin{bmatrix} 1 & x_1 & | & y_1 \\ 1 & x_2 & | & y_2 \\ \cdot & \cdot & | & \cdot \\ \cdot & \cdot & | & \cdot \\ \cdot & \cdot & | & \cdot \\ 1 & x_n & | & y_n \end{bmatrix}$$

Two cases:

- The points on the graph are not all collinear. This corresponds to the case that the linear system is inconsistent. In this case, the linearity hypothesis is definitively rejected, i.e., falsified.
- The points on the graph are all collinear. This corresponds to the case that the linear system is consistent. This is consistent with the linearity hypothesis, and can be construed as evidence in favor of the hypothesis. The greater the number of points, the stronger the evidence, because the greater the potential for falsification.

1.4. **Two different pictures.** It's worth noting that there are two different types of pictures we are using:

- One picture is the $xy$-plane in which we are trying to fit existing data points to obtain the graph $y = f(x)$ of the function $f$. The two axes here are the axis for $x$ and the axis for $y = f(x)$.
- The other picture is a plane, but the coordinates for the plane are now the parameters $m$ and $c$, viewed as variables. Every *point* in this plane corresponds to a particular choice of *function* $f(x) = mx + c$ and therefore a *line* in the earlier plane.

The switching back and forth between these two different geometric structures that we use can be a bit confusing. Note also that in this case, both the pictures are in two dimensions, but there is no intrinsic reason why the two kinds of pictures should involve the same number of dimensions. The number of dimensions used for the first picture is 2 because the function has one input and one output. The number of dimensions used for the second picture is 2 because that is the number of parameters. These numbers coincide because we are using an affine linear model, where the number of parameters is one more than the number of inputs. When we are using polynomial models, then the dimensions do not match. In any case, there is no intrinsic reason why the dimensions should be related.

1.5. **Short generic summary.** The following general principles relating the linear algebra setting and hypothesis testing will be useful for the:

- The coefficient matrix has full row rank: The system is always consistent (regardless of the values of the outputs), so we cannot use consistency as evidence in favor of the hypothesis.

- The coefficient matrix does not have full row rank: There is potential for inconsistency, so consistency is evidence in favor of the hypothesis.
- The coefficient matrix has full column rank: The system has a unique solution if consistent, so can be used to find the parameters assuming the truth of the hypothesis.
- The coefficient matrix does not have full column rank: We cannot find the solutions uniquely even if the system is consistent.
- The coefficient matrix has full row rank and full column rank (the situation arises when number of parameters = number of input-output pairs, and the inputs are chosen to give independent information): We can find the parameters uniquely assuming the truth of the hypothesis, but we cannot verify the truth of the hypothesis.
- The coefficient matrix has full column rank but does not have full row rank: We can find the parameters uniquely and also obtain independent confirmation of the truth of the hypothesis.

1.6. **Type I and Type II errors and $p$-values.** *Note: Some of this is quite confusing, so I don't really expect you to understand all the terminology if you haven't taken statistics, and perhaps even if you have.*

Let's borrow a little terminology from statistics to get a better sense of what's going on. Suppose there are two hypotheses at hand regarding whether the function $f$ of one variable is linear. One hypothesis, called the *null hypothesis*, states that $f$ is not linear. The other hypothesis, which is precisely the logical opposite of the null hypothesis, is that $f$ is in fact linear.

Let's say we conduct our "hypothesis testing" by looking at $n$ points, calculating the function values, and then looking at the points in the graph of the function and checking if they are collinear. The rule is that we reject the null hypothesis (i.e., conclude that the function is linear) if the $n$ points are collinear. Otherwise, we do not reject the null hypothesis, i.e., we conclude that the function is not linear.

In principle, any procedure of this sort could involve two types of errors:

- *Type I error*, or "false positive" error, refers to a situation where we incorrectly reject the null hypothesis. In this case, it refers to a situation where we incorrectly conclude that the function is linear, even though it is not.
- *Type II error*, or "false negative" error, refers to a situation where we incorrectly fail to reject the null hypothesis. In ths case, it refers to a situation where we incorrectly conclude that the function is not linear, even though it is linear.

Now, in the hypothesis test we have designed here, Type II errors are impossible. In other words, if the function is linear, we will always conclude that "it is linear" if we use the above test. Type I errors, however, are possible. For $n = 1$ and $n = 2$, the test is useless, because we'll always end up with Type I errors for non-linear functions. For $n \geq 3$, Type I errors are possible in principle. But how likely are they to occur in practice? Quantifying such probabilities is very difficult. There are several different approaches one could use, including $p$-values and the $\alpha$ and $\beta$ values, all of which mean different but related things. It's probably not worth it to go into these at this point in the course, but we'll get back to it later in the course in a setting more amenable to probability.

Instead of going into the details, I will explain briefly why, barring some very special information about the nature of the function, we should loosely expect Type I errors to be quite unlikely for this setup. In fact, in many models, the associated $p$-value for this setup is 0. Here's the intuitive reasoning: let's say we calculated the function values at two points, then fitted a line through them. Now, we want to look at the value of the function at a third point. If the function is an arbitrary function in a loose sense, the probability of its value at the third point just "happening" to be the right value for linearity seems very low: it's effectively asking for the probability that a randomly chosen real number takes a particular predetermined value. Thus, getting collinear points in the graph for three or more inputs is strong evidence in favor of linearity: it would be highly unlikely that you'd just hit upon such points by chance.

1.7. **Bayesian reasoning.** In practice, the reasoning above does not occur in a vacuum. Rather, it occurs in a *Bayesian* framework. Namely, you already have some *prior* ideas regarding the nature of the function. Data that you collect then lets you update your priors.

Below, we discuss some examples of priors under which collecting data of this sort does not provide very strong evidence of linearity. Suppose your prior belief is that the function is a *piecewise linear continuous function*: we can break the reals into contiguous intervals such that the restriction to each interval is linear, but the function as a whole is not linear. For instance, $|x|$ is a piecewise linear function of $x$, because it has the definition:

$$|x| = \begin{cases} -x, & x < 0 \\ x, & x \geq 0 \end{cases}$$

There are reasonable prior probability distributions on collections of piecewise linear functions where information about three points being collinear is partial, but not near-conclusive, evidence of the function overall being linear. "Piecewise linearity" offers an alternative hypothesis to global linearity that can parsimoniously explain the observed data of collinear points on the graph. Namely: perhaps all your observed points just happened to be in one piece of the piecewise linear description! Piecewise linearity is not merely a hypothetical scenario. Many functions in finance and taxation are designed to be piecewise linear. If we restrict all the inputs as being close to each other, we may be fooled into thinking the function is linear.

For useful hypothesis testing of linearity that could distinguish it from piecewise linearity, we want to pick our points spaced away as far as possible, so that it is harder to explain away linearity of the points on the graph by saying that we got lucky about all points being in the same piece.

Another type of situation where it may be hard to reject alternatives to linearity is a situation where all our observations are at integer inputs, and we believe the function may have the form:

$$f(x) := mx + A\sin(\pi x) + c$$

Note that the $A\sin(\pi x)$ is not detected by the choices of inputs, all of which are integers. Again, useful testing of this alternative hypothesis can happen only if we choose non-integer inputs.

Everything boils down to what type of prior we have, based on broad theoretical considerations, about the model of the function we are dealing with.

## 2. Functions of multiple variables

For functions of one variable, we can perform the entire analysis by imagining the nature of the graph of the function. For functions of two variables, graphing already becomes more tricky: the graph of the function is now a surface in three-dimensional space. For functions of three or more variables, the graph becomes too tricky to even consider.

Fortunately, we have another option: setting up a linear system.

2.1. **The linear system setup: three input-output pairs for a function of two variables.** Suppose we have a function $f$ of two variables $x$ and $y$. If $f$ is (affine) linear (i.e., we allow for a nonzero intercept), it has the form:

$$f(x, y) := ax + by + c$$

where $a$, $b$, and $c$ are constants.

We do not, however, know for sure whether $f$ is linear. We have a black box that outputs the values of $f$ for various inputs.

If we provide three inputs $(x_1, y_1)$, $(x_2, y_2)$, and $(x_3, y_3)$ to $f$, with outputs $z_1$, $z_2$, and $z_3$ respectively, the augmented matrix we get (with the columns corresponding to $c$, $a$, and $b$ in that order) is:

$$\left[\begin{array}{ccc|c} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \end{array}\right]$$

There are two cases now:

- The inputs are collinear: In this case, the row rank of the system is 2. In other words, the system does not have full row rank. One of the rows of the coefficient matrix is redundant. Thus, the system may or may not be consistent. Which case occurs depends on the outputs. We consider both cases:
  - The system is inconsistent: This means that we have falsified the hypothesis of linearity. Pictorially, this means that in $\mathbb{R}^3$, the three points $(x_1, y_1, z_1)$, $(x_2, y_2, z_2)$, and $(x_3, y_3, z_3)$ are non-collinear.

- The system is consistent: This means that the data are consistent with the hypothesis of linearity. Pictorially, this means that in $\mathbb{R}^3$, the three points $(x_1, y_1, z_1)$, $(x_2, y_2, z_2)$, and $(x_3, y_3, z_3)$ are collinear. In other words, we cannot reject the hypothesis of the function being linear. Thus, we have collected strong evidence in favor of the linearity hypothesis. Note that even though we have collected evidence in favor of linearity, the system not having full column rank means that we cannot determine $a$, $b$, and $c$ uniquely.

- The inputs are non-collinear: In this case, the row rank of the system is 3. In other words, the system has full row rank, and also full column rank. Thus, whatever the output, there exists a unique solution, i.e., we uniquely determine $a$, $b$, and $c$. In fact, we can find unique values of $a$, $b$, and $c$ even if the function is not linear. In other words, if the inputs are non-collinear, then we do not get any information that could potentially falsify the linearity hypothesis, hence we cannot conclude anything in favor of the linearity hypothesis.

In other words, if we are only allowed three inputs, then we have to choose between either attempting to test the hypothesis or attempting to find $a$, $b$ and $c$ uniquely (conditional to the truth of the hypothesis).

2.2. **More than three input-output pairs for a function of two variables.** If, however, we are allowed four or more inputs, we can have our cake and eat it too. As long as three of our inputs are non-collinear, we can use them to obtain the coefficients $a$, $b$, and $c$ assuming linearity. Pictorially, this would be fixing the plane that is the putative graph of the function in three dimensions. Further input-output information can be used to test the theory. Pictorially, this is equivalent to testing whether the corresponding points in the graph of the function lie in the plane that they are allegedly part of.

2.3. **Functions of more than two variables.** Suppose $f$ is a function of more than two variables that is alleged to be affine linear (i.e., we allow for an intercept), i.e., it is alleged to be of the form:

$$f(x_1, x_2, \dots, x_n) = a_1 x_1 + a_2 x_2 + \dots + a_n x_n + c$$

If you are skeptical of the claim of linearity, and/or want to convince somebody who is skeptical, the easiest thing to do would be to pick three points which are collinear in $\mathbb{R}^n$, thus getting a $3 \times (n+1)$ coefficient matrix of rank 2. If the system is inconsistent, that is definite evidence against linearity. If the system is consistent, that is evidence in favor of linearity, but it is not conclusive.

If the goal is to *find* the values $a_1$, $a_2$, $\dots$, $a_n$, and $c$ then we should pick $n+1$ different points that are *affinely independent*. This is a concept that we will define later. It is beyond the current scope. However, a randomly selected bunch of points will (almost always) work.

If, however, we want to find the coefficients *and* obtain independent confirmation of the theory, then we need to use at least $n+2$ observations, with the first $n+1$ being affinely independent. The more observations we use, the more chances we have given to potential falsifiers, and therefore, if the linearity hypothesis still remains unfalsified despite all the evidence that could falsify it, that is very strong evidence in its favor.

2.4. **Bayes again!** Recall that, as mentioned earlier, how strongly we view our evidence as supporting the function being linear depends on our prior probability distribution over alternative hypotheses. Consider, for instance, the function $f(x, y) := x + y^2 - 2$. For any fixed value of $y$, this is linear in $x$. Thus, if we choose our three collinear points on a line with a fixed $y$-value, then this function will "fool" our linearity test, i.e., we will get a Type I error.

Therefore, as always, the original prior distribution matters.

## 3. Verifying a nonlinear model that is linear in parameters

Many similar remarks apply when we are attempting to verify whether a particular nonlinear model describes a function, if the model is linear in the parameters.

3.1. **Polynomial function of one variable.** Consider the case of a function of one variable that is posited to be a polynomial of degree at most $n$, knowing the function value at $n+1$ distinct points allows us to determine the polynomial, but does not allow any verification of whether the function is indeed polynomial, because we can fit a polynomial of degree $\leq n$ for *any* collection of $n+1$ input-output pairs. If, however, we have $n+2$ input-output pairs and we get a polynomial of degree $\leq n$ that works for them, that is strong evidence in favor of the model being correct.

Let's make this more explicit. Consider a model with the functional form:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n$$

This is a generic polynomial of degree $\leq n$. There are $n+1$ parameters $\beta_0, \beta_1, \ldots, \beta_n$. Given $m$ input-output pairs $(x_1, y_1)$, $(x_2, y_2)$, $\ldots$, $(x_m, y_m)$, the augmented matrix that we obtain is:

$$\left[\begin{array}{ccccc|c} 1 & x_1 & x_1^2 & \ldots & x_1^n & y_1 \\ 1 & x_2 & x_2^2 & \ldots & x_2^n & y_2 \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ 1 & x_m & x_m^2 & \ldots & x_m^n & y_m \end{array}\right]$$

Note that the coefficient matrix, as always, depends only on the inputs. The coefficient matrix is a $m \times (n+1)$ matrix:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^n \\ 1 & x_2 & x_2^2 & \ldots & x_2^n \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 1 & x_m & x_m^2 & \ldots & x_m^n \end{bmatrix}$$

By standard facts about rank, the rank of the coefficient matrix is at most $\min\{m, n+1\}$. It turns out that, for matrices of this type, the rank always achieves its maximum value as long as the values $x_1, x_2, \ldots, x_m$ are all distinct. This is not completely obvious, and in fact depends on some algebraic manipulation. But it should not be *surprising* per se, because we know that the values at different points are "independent" pieces of data and therefore should give independent equations to the extent feasible, so we expect the coefficient matrix to have the largest possible rank.

Essentially, this means that:

- If $m \leq n$, then we do not have enough information either to verify the hypothesis of being a polynomial of degree $\leq n$ or to find the coefficients assuming the truth of the hypothesis. Our inability to use the data to verify the hypothesis arises because the system has full row rank $m$, so it is *always* consistent, regardless of output. Our inability to find the solution uniquely arises from the fact that the dimension of the solution space is $n+1-m > 0$, so the solution is non-unique.
- If $m = n+1$, then we do *not* have enough information to verify the hypothesis of being a polynomial of degree $\leq n$, but we *do* have enough information to find the polynomial assuming the truth of the hypothesis. Our inability to use the data to verify the hypothesis arises because the coefficient matrix is now a square matrix of full row rank $m = n+1$, so the system is always consistent, regardless of output. On the other hand, since the coefficient matrix has full column rank, the solution, if it exists, is unique, so we can find the polnyomial uniquely.
- If $m > n+1$, then we have information that can be used both to verify the hypothesis of being polynomial of degree $\leq n$ (albeit not concislvey) and we also have enough information to find the polynomial assuming the truth of the hypothesis. Our ability to use the data to verify the hypothesis arises because the coefficient matrix no longer has full row rank (the rank is $n+1$, and is less than $m$) so there is potential for inconsistency, therefore, consistency provides evidence in favor of the hypothesis. We can find the polynomial uniquely assuming the truth of the hypothesis because the matrix has full column rank $n+1$, so the solution, if it exists, is unique.

3.2. **Polynomial function of more than one variable.** The ideas discussed above can be applied to the case where the functional form is polynomial of bounded degree in more than one variable. For instance, a functional form for a polynomial of total degree at most 2 in the variables $x$ and $y$ is:

$$f(x, y) = a_1 + a_2x + a_3y + a_4x^2 + a_5xy + a_6y^2$$

We notice the following:

- In order to *find* the parameters $a_1, a_2, \dots, a_6$ uniquely, we would need to determine the outputs for a well-chosen collection of six inputs. By "well-chosen inputs" we mean that the inputs should satisfy the property that the coefficient matrix (which is a $6 \times 6$ square matrix) has full rank.

  Note that, when we were talking of an affine linear function of two variables, the condition for being well-chosen was that the inputs are not collinear as points in the $xy$-plane. It is possible to find a similar geometric constraint on the nature of the inputs for the coefficient matrix here to have full rank. However, finding that constraint would take us deep into realms of higher mathematics that we are not prepared to enter.
- In order to *find* the parameters and *additionally to obtain verification of the model*, we would need to determine the outputs for a well-chosen collection of more than six inputs. Here, "well-chosen" means that the coefficient matrix would still have rank 6, so it has full column rank (allowing us to find the polynomial uniquely if it exists) but does not have full row rank (creating a potential for inconsistency, and therefore allowing us to use consistency to obtain evidence in favor of the claim).

## 4. Measurement and modeling errors

Most of the discussion above is utopian because it ignores something very real in most practical applications: *error*, both *measurement error* and *modeling error*. Measurement error means that the inputs and outputs as measured are only approximately equal to the true values. Modeling error means that we are not claiming that the function is actually linear. Rather, we are claiming that the function is only approximately linear, even if we use the "true values" of the variables rather than the measured values.

We use the term *overdetermined* for a linear system that has more than the required minimal equations to determine the parameter values. In the absence of measurement error, overdetermined systems will be consistent assuming the hypothesis is true, i.e., the extra equations allow us to "test" the solution obtained by solving a minimal subset.

However, measurement errors can ruin this, and we need to adopt a more error-tolerant approach. The methods we use for that purpose fall broadly under the category of *linear regression*. We will discuss them later in the course.