Rutgers University Math 251
Solving Systems of Equations

Finding critical points of functions of several variables in order to identify the locations of maximum or minimum values involves solving a system of simultaneous equations. The general theory, even when the equations are given by polynomials is fairly complicated, and the tricks to recognize easily solved systems may not be as clear as in the single-variable case. These notes aim to address difficulties observed in solutions to examination problems in the Multivariable Calculus course.

Computer Programmers have a saying: “First make it work; then make it fast.” There are a few principles that can be used to guide a solution that can be shown to be correct as long as the calculations are performed correctly. However, straightforward application of these principles may be tedious and tedious calculations invite error. In practice, the principle is used in a disguised form that leads to more accurate computation. The temptation to try to reduce the computation to a formula should be resisted. The mixture of a sensible principle guiding an accurate computation will handle a wider scope of problems with less danger of confusion.

Since the goal in this course is to use the solution of the system to locate critical points that will identify extreme values of a function, these algebraic manipulations are only one step in a larger program. Mistakes that escape notice may make further work both difficult and useless, so work should be checked. Evidence of a check will not be requested on exams, so you are free to use any method you choose, but you should have confidence in intermediate results before using them in later work.

An important principle in solving equations is that **linear equations are easy to solve**. You simply put multiples of the variable on one side of the equation and everything else on the other side. The distributive law says that the sum of multiples of the variable is the product of a single expression and the variable, so you can solve the equation by division.

Implicit differentiation is a good example of a problem leading to linear equations. If you make the assumption that one variable is a function of the others, and differentiate with respect to one of the independent variables, the derivative of the dependent variable with respect to an independent variable appears linearly in the equation that results from differentiating the identity giving the implicit definition of the function. This connects the use of the gradient to find a vector perpendicular to a level surface of a function with the method presented earlier for finding the tangent plane to the graph of a function. Treating the level surface as an implicit definition of one coordinate variable as a function of the others gives an equation of the tangent plane in which one of the coefficients has been normalized. In most applications, this normalization is not useful, so the form of the tangent plane produced from the gradient is preferred. Indeed, it is only the existence of a function that justifies the relevance of calculus in some problems, with all the work done by indirect or implicit descriptions of the properties of the function.

Coming back to the solution of systems of equations, if one equation is linear in one of the variables, it is possible to solve that equation for that variable and substitute the result in the remaining equations. This removes one variable from consideration until the the others have been found and **replaces the system by a related system with one fewer equation**. If all equations are linear in all variables, repeated use of this method will find values for all variables that can be found. However, doing the work this way introduces numerical denominators that will usually be multiplied by in later steps and shifts terms to different sides of equations, with a consequent change of sign, many times. Missing just one of these sign changes, or incorrectly combining fractions, can ruin the solution. Thus, as those who are taking Linear Algebra will recognize, a different approach, leading to equivalent results, is used in practice. This approach, called *elimination*, leaves all the variables on the left side of the equations, writing the equations in the form $A = a$, where A is an expression in the variables and a is a constant. If you have two such equations, $A_0 = a_0$ in which the variable x has coefficient c_0, and $A_1 = a_1$ in which the variable x has coefficient c_1, then *clearly*

$$c_1 A_0 - c_0 A_1 = c_1 a_0 - c_0 a_1$$

and x does not appear in this equation. One of the original equations can be use as a definition of x after solving this equation. The method used in Linear Algebra keeps the same number of equations and same number of variables throughout the solution process, even if some equations reduce to the trivial identity $0 = 0$, although more and more equations are used only as definitions of particular variables.

Example: Consider the system

$$4x + 5y = 23$$
$$5x + 6y = 28$$

Multiplying the second equation by 5 and the first by 6 will give y the same coefficient in both equations. Subtracting will give the equation $(5 \cdot 5 - 6 \cdot 4)x = (5 \cdot 28 - 6 \cdot 23)$, which simplifies to $x = 2$, that is a consequence of the given equations. Similarly, multiplying the first by 5 and the second by 4, then subtracting, gives $y = 3$. These values are easily seen to satisfy the given equations.

A form of elimination can also be used for equations of higher degree. Here, it is better to move everything to the left side of the equations, leaving only zero on the right. Concentrating on one of the variables, one can attempt to divide the expression in one of the equations into the expressions in the remaining equations, getting a quotient and a remainder. The **remainders are of lower degree** in the selected variable, but **they give an equivalent statement**. This will introduce fractions whose denominators may be expressions in the remaining variables, so there must be an independent investigation of the consequence of those expressions being zero. Sometimes, though not always, repeated application of this method will produce a linear equation in the selected variable.

Example. Applying the method of Lagrange multipliers to find the maximum of $f(x, y) = x(1-y)$ subject to $x^2 + y^2 = 1$. The constraint says the (x, y) lies on the unit circle, and the function to be maximized (often called the *objective*) is the area of the triangle with vertices $(0, 1)$, (x, y), $(-x, y)$, so this is part of the problem of finding the largest triangle that can be inscribed in a circle. (Because of symmetry, one vertex of the triangle can be fixed at $(0, 1)$, so the general problem would allow two points to move independently around the circle and would seek to maximize the triangle formed by these points and $(1, 0)$. We don't consider that problem only to avoid the distraction of determining the area of such a triangle.) The gradient of the constraint is $\langle 2x, 2y\rangle$ and the gradient of the objective is $\langle 1 - y, -x\rangle$. The method of Lagrange multipliers says that the extreme values of the objective occur at points where these vectors are parallel. Solving for the factor λ that gives one as a scalar multiple of the other, gives that the ration of corresponding entries are equal, and cross-multiplying gives $2x(-x) = 2y(1-y)$. Removing the factor of 2 that serves no useful purpose, and expanding gives a condition for an extremum to be solved together with the constraint. The resulting system is

$$x^2 - y^2 + y = 0$$
$$x^2 + y^2 - 1 = 0$$

Treating x as the variable, the quotient when we divide one of these by the other is 1, and the remainder is simply the difference of the equations, which is $2y^2 - y - 1 = 0$. The roots of this are $y = 1$ and $y = -1/2$. Go back to one of the original equations to determine x (either one — they should say the same thing for these values of y) and evaluate the objective $f(x, y)$. When $y = 1$, $x = 0$ and $f(0, 1) = 0$; when $y = -1/2$, $x^2 = 3/4$, so $x = \pm\sqrt{3}/2$ and $f(\pm\sqrt{3}/2, -1/2) = \pm 3\sqrt{3}/4$. Thus $-3\sqrt{3}/4 \le f(x, y) \le 3\sqrt{3}/4$ on the circle. The triangle that maximizes area is equilateral.

The symmetry of this answer suggests that it may also a solution of the general problem. To approach that problem, not that there is a geometric interpretation of the gradient of the objective: it is the vector perpendicular to the side opposite the point (x, y). The same condition appears in the general problem: in a triangle of maximal area, the tangent to the circle at each point must be parallel to the opposite side for otherwise that point could be moved farther away from an opposite side that is held fixed. This increases the altitude of the triangle while keeping the base fixed, so it increases the area. One application of this geometric principle shows that attention can be confined to the case in which one side is parallel to the tangent at the opposite vertex, but this is the problem that was just solved.

This ends the discussion of the example.

When a variable has been eliminated, it may be useful to collect the reduced system of equations in one place before continuing. That is, consider the reduced system as a replacement for the original problem. Although much of algebra is deductive, in the sense that the task is the accumulation of more and more wonderful consequences of the original system of equations until a solution is revealed, many applications require that we be convinced that all solutions have been found and that the claimed solutions really satisfy the original system. To aid these tasks, as well as to deal with any errors reveled when checking the solutions, it is useful to have a record of the work that is easy to follow.

The principle of factoring is important in recognizing systems that are easy to solve. It says that **a product of two expressions can be zero only if one of the expressions is zero**. An early algebraic method developed calls for recognizing factors of quadratic polynomials in one variable. Such examples, while rare in real life, allow Calculus exercises to be constructed that have integer solutions — so much so that this is expected in such exercises. In solving systems, a factorization of an expression that is to be made zero allows the solution to be split into two parts, connected by the word “or”. (In each part, we are looking for a simultaneous solution, so the different equations are all connected by the word “and”. The full logic of both ways of building compound conditions has become unavoidable. To keep things simple, each “or” clause will be treated as a separate problem containing only conditions combined with “and”.)

Example. Let’s find all critical points of

$$f(x, y) = x^3 + 2xy^2 + 3x^2 + 9y^2.$$

The gradient of this expression is

$$\left\langle 3x^2 + 2y^2 + 6x, 4xy + 18y \right\rangle,$$

which is zero at a critical point. The second component of this vector factors as $(4x + 18)y$, so there are two cases.

Case I: $x = -9/2$. Substitute this into the first equation to get $135/4 + 2y^2$, which has no real roots.

Case II: $y = 0$. Substitute this into the first equation to get $3x^2 + 6x = 3x(x + 2)$. The roots are $x = 0$ and $x = -2$.

Combining the two cases, the only (real) critical points are $(0, 0)$ and $(-2, 0)$.

There is a second derivative test to determine the local behavior of the function near a critical point. This calls for considering the signs of $f_{xx}(x, y) = 6x + 6$ and $f_{xx} f_{yy} - f_{xy}^2 = (6x + 6)(4x + 12) - (4y)^2$ at each critical point. If $(x, y) = (0, 0)$, these are both positive, indicating a local minimum (since the second derivative along any line through the point is positive). If $(x, y) = (-2, 0)$, the second of these is negative, indicating a saddle point (in fact, the coordinate directions already give examples of lines where the second derivative can have either sign).

Although, $(0, 0)$, where $f(0, 0) = 0$, is a local minimum, a graph easily locates points where $f(x, y) < 0$. Indeed, $f(x, 0) = x^2(x + 3)$, so any $x < -3$ gives an example.
