Lectures - Week 13
Two Point Boundary Value Problems and Functions of Several Variables

We now want to briefly look at a linear second order BVP which is sometimes called a two-point BVP because we are specifying conditions at the two endpoints of our domain. Specifically we seek $u(x)$ satisfying

$$\begin{aligned} -u''(x) + p(x)u'(x) + q(x)u(x) = g(x) \quad & a < x < b \\ u(a) = \alpha \qquad u(b) = \beta & \end{aligned}$$

This problem is a BVP because we are given the value of the unknown $u(x)$ (or its derivative) at the endpoints of the interval $[a, b]$ and are asked to find $u(x)$ in the interior of the interval which satisfies the given differential equation. To approximate the solution of this equation we can no longer expect to start at a point and "march in time" because the solution is affected by its boundary values at both $x = a$ and $x = b$.

To come up with an approach to approximate this problem we return to Taylor series which we used to originally derive Euler's Method. Recall that

$$u(x + \Delta x) = u(x) + \Delta x u'(x) + \frac{\Delta x^2}{2} u''(x) + \frac{\Delta x^3}{3!} u'''(x) + \cdots$$

In our differential equation we need to replace $u''(x)$ as well as $u'(x)$ with a difference quotient. If we solve the above equation for $u''(x)$ it would be in terms of $u'(x)$ too, which we don't want. However, if we write the Taylor series for $u(x - \Delta x)$ and combine the two then we can get the $u'(x)$ terms to cancel. We have

$$u(x - \Delta x) = u(x) - \Delta x u'(x) + \frac{\Delta x^2}{2} u''(x) - \frac{\Delta x^3}{3!} u'''(x) + \cdots$$

Adding these two series gives

$$\begin{aligned} u(x + \Delta x) + u(x - \Delta x) &= \Big[u(x) + \Delta x u'(x) + \frac{\Delta x^2}{2} u''(x) + \frac{\Delta x^3}{3!} u'''(x) + \cdots\Big] \\ &\quad + \Big[u(x) - \Delta x u'(x) + \frac{\Delta x^2}{2} u''(x) - \frac{\Delta x^3}{3!} u'''(x) + \cdots\Big] \\ &= 2u(x) + (\Delta x)^2 u''(x) + \mathcal{O}(\Delta x^4) \end{aligned}$$

Solving for $u''(x)$ gives

$$u''(x) = \frac{u(x + \Delta x) - 2u(x) + u(x - \Delta x)}{(\Delta x)^2} + \mathcal{O}(\Delta x^2)$$

This is called a *second centered difference* approximation to $u''(x)$ because it can also be found by differencing the forward and backward first order approximations to $u'(x)$. Note that the error term is $\mathcal{O}(\Delta x^2)$. We can now use this approximation in our differential equation to approximate $u''(x)$. However, we must still implement an approximation to $u'(x)$. If we use a forward or backward difference, i.e.,

$$\frac{u(x+\Delta x)-u(x)}{\Delta x} \qquad \text{or} \qquad \frac{u(x)-u(x-\Delta x)}{\Delta x}$$

then we know that the error term is $\mathcal{O}(\Delta x)$ whereas our approximation for $u''(x)$ is $\mathcal{O}(\Delta x^2)$. It would be advantageous if we could approximate $u'(x)$ by a difference quotient that was $\mathcal{O}(\Delta x^2)$. We can easily do this by using a *centered (first) difference*. This is found by subtracting our Taylor series expansions for $u(x+\Delta x)$ and $u(x-\Delta x)$ to get

$$u(x+\Delta x)-u(x-\Delta x) = 2\Delta x u'(x) + \mathcal{O}(\Delta x^3)$$

and thus

$$u'(x) = \frac{u(x+\Delta x)-u(x-\Delta x)}{2x} + \mathcal{O}(\Delta x^2)$$

Our strategy in approximating our two point BVP is to discretize our domain $[a,b]$ by setting

$$x_0 = a, \quad x_1 = a+\Delta x, \quad x_2 = x_1 + \Delta x, \cdots x_i = a + i\Delta x, \quad x_{n+1} = b$$

with $\Delta x = (b-a)/(n+1)$ and obtain an approximation U_i to $u(x_i)$. We know that $U_0 = \alpha$ and $U_{n+1} = \alpha$ and we have n unknowns $U_1, \cdots U_n$. To this end, we write a difference equation at each point x_i, $i = 1, \ldots, n$. Using our difference quotients to approximate $u''(x)$ and $u'(x)$ at x_1 gives

$$-\frac{U_2 - 2U_1 + U_0}{\Delta x^2} + p(x_1)\frac{U_2 - U_0}{2\Delta x} + q(x_1)U_1 = g(x_1)$$

Now $U_0 = \alpha$ is known so terms involving it can be moved to the right hand side but this equation couples two unknowns U_1 and U_2; consequently we can't solve for either one. Writing the equation at x_2 gives

$$-\frac{U_3 - 2U_2 + U_1}{\Delta x^2} + p(x_2)\frac{U_3 - U_1}{2\Delta x} + q(x_2)U_2 = g(x_2)$$

This equation couples three unknowns U_1, U_2 and U_3 so we can't solve for any of them from the first two equations. In general, at the node x_i we have

$$-\frac{U_{i+1} - 2U_i + U_{i-1}}{\Delta x^2} + p(x_i)\frac{U_{i+1} - U_{i-1}}{2\Delta x} + q(x_i)U_i = g(x_i)$$

which again couples three unknowns. So we write our difference equation at each node $x_1, x_2, \ldots, x_n$ so that we have n equations for the n unknowns U_1, $U_2, \ldots, U_n$ but they are all coupled linear equations. This just means that we can't solve for them one at a time but must solve for them simultaneously, i.e., solve a linear system $A\vec{u} = \vec{b}$.

To see this, let's simplify the equation to $-u''(x) = g(x)$, i.e., $p(x) = q(x) = 0$. We then have the ith equation as

$$\frac{-U_{i+1} + 2U_i - U_{i-1}}{\Delta x^2} = g(x_i)$$

Our unknown vector is $(U_1, U_2, \cdots, U_n)^T$, and our system becomes

$$\begin{pmatrix} 2 & -1 & 0 & \cdots & & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ \vdots \\ \vdots \\ U_n \end{pmatrix} = \begin{pmatrix} \Delta x^2 g(x_1) + \alpha \\ \Delta x^2 g(x_2) \\ \Delta x^2 g(x_3) \\ \vdots \\ \vdots \\ \Delta x^2 g(x_n) + \beta \end{pmatrix}$$

Note that the first and last terms in the right hand side vector have been modified due to the $U_0 = \alpha$ term in the first equation and $U_{n+1} = \beta$ term in the last equation. Also the coefficient matrix is symmetric and it can be shown that it is also positive definite. So the discrete solution to our BVP is found by solving a tridiagonal linear system with a method such as Cholesky decomposition which we know requires $\mathcal{O}(n)$ operations. You should contrast this with the case of approximating the solution of an IVP where we "march in time"; for BVPs the solution is coupled and we solve for all unknowns at once. When we briefly look at partial differential equations (PDEs) we will see that these basic ideas carry over to that setting.

We now want to consider functions of more than one independent variable. We have already briefly looked at such functions when we considered the general IVP $y'(t) = f(t,y)$. Here f is a function of two variables and we saw how to take partial derivatives and perform a Taylor series expansion. We want to look at functions of more than one variable in more detail.

Graphically we know that if we plot $y = f(x)$ then we get a curve in R^2. If we have a function $z = f(x,y)$ then for each point (x,y) we have a corresponding z value so we are graphing a surface in R^3. Sometimes one plots *level curves* of the surface, i.e., curves such that $f(x,y)$ is a constant. For example, if we have the surface $z = x^2 + y^2$ then the level curves are $x^2 + y^2 = \alpha$ which are just circles with radius $\sqrt{\alpha}$. If you watch the weather during the winter they will sometimes show plots of the isotherms, i.e., the curves of constant temperatures. These are just level curves.

We have seen that taking the partial derivative of a function of several variables is straightforward, we just hold all variables fixed except the one we are differentiating with respect to. This means we are determining the instantaneous rate of change of the function in that direction only, i.e., along one of the coordinate axis. For example, if $f(x,y)$ then f_x is defined as

$$f_x(x,y) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y) - f(x,y)}{\Delta x}$$

so y is held as a constant. What if we want to find how f changes in a direction other than the coordinate axes? To determine this we first need the *gradient*.

Gradient

Let $f(x_1, x_2, \ldots, x_n)$. Then the *gradient of* f, denoted $\vec{\nabla} f$ (sometimes the arrow is omitted and we just write ∇f) is a vector containing the partial derivatives of f given by

$$\vec{\nabla} f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Example Determine the gradient of $f(x,y,z) = z^3 x \sin \pi y$ at $(1, \frac{1}{2}, 2)$.

Calculating our partial derivatives gives

$$f_x = z^3 \sin \pi y, \quad f_y = \pi z^3 x \cos \pi y, \quad f_z = 3z^2 x \sin \pi y$$

Evaluating at the point we have

$$\nabla f = \begin{pmatrix} 2^3 \sin(\pi/2) \\ \pi 2^3 (1) \cos(\pi/2) \\ 3(2^2)(1) \sin(\pi/2) \end{pmatrix} = \begin{pmatrix} 8 \\ 0 \\ 12 \end{pmatrix}$$

Now suppose that we want to know how a function $f(x, y)$ changes in a direction other than parallel to the x or y axis; for example in the direction of $\vec{u} = u_1\hat{\imath} + u_2\hat{\jmath}$ where $\|\vec{u}\|_2 = 1$ and $\hat{\imath}$ denotes a unit vector in the direction of the x-axis and $\hat{\jmath}$ denotes a unit vector in the direction of the y-axis. For example $\vec{u}$ could be the unit vector in the first quadrant pointing away from the origin at a 45° angle, i.e., $\vec{u} = \frac{1}{\sqrt{2}}\hat{\imath} + \frac{1}{\sqrt{2}}\hat{\jmath}$. We will let $D_{\vec{u}}f$ denote the derivative of f in the direction of the vector $\vec{u}$. If $f(x, y)$ then the definition of $D_{\vec{u}}f$ is given by

$$D_{\vec{u}}f(x, y) = \lim_{h\to 0} \frac{f(x + hu_1, y + hu_2) - f(x, y)}{h}$$

Note that if $\vec{u} = \hat{\imath}$ then this reduces to f_x and if $\vec{u} = \hat{\jmath}$ this reduces to f_y with $\Delta x = h$ and $\Delta y = h$. This is the definition of $D_{\vec{u}}f$ but we don't want to use it every time we need to calculate a directional derivative so we need a simple means to calculate it. If $f(x, y)$ then $D_{\vec{u}}f = f_x\ u_1 + f_y u_2 = (\nabla f)^T\vec{u}$. In general, if f is a function of n variables $f(x_1, x_2, \ldots, x_n)$ then the derivative of f in the direction of the unit vector $\vec{u}$ is just the dot or scalar product of the gradient of f with the directional vector $\vec{u}$.

$$D_{\vec{u}}f = \nabla f \cdot \vec{u}$$

Once again we should make sure that this agrees with our partial derivatives which are just directional derivatives in the direction of the coordinate axes. For example, if f is a function of two variables then $D_{\hat{\imath}}f$ should just be f_x; we have

$$D_{\hat{\imath}}f = f_x = (f_x, f_y) \cdot (1, 0) = f_x$$

For completeness we list some properties of the gradient which should be evident from properties of derivatives.

(1.) $\nabla c = 0$ where c is a constant.
(2.) $\nabla(\alpha f + \beta g) = \alpha\nabla f + \beta\nabla g$ where α, β are scalars
(3.) $\nabla(fg) = g\nabla f + f\nabla g$
(4.) $\nabla\big(\frac{f}{g}\big) = \dfrac{g\nabla f - f\nabla g}{g^2}$
(5.) $\nabla(f^n) = nf^{n-1}\nabla f$

Example Determine $\nabla(f^3)$ where $f(x, y) = x^2y$.

Here we need to use property (5) which is simply the Power Rule for derivatives. We have

$$\nabla f = \nabla[(x^2y)^3] = 3(x^2y)^2\nabla f = 3x^4y^2(2xy\hat{\imath} + x^2\hat{\jmath}) = 6x^5y^3\hat{\imath} + 3x^6y^2\hat{\jmath}$$

Another way to do this is to simply take $\nabla(x^6y^3)$ which is just $6x^5y^3\hat{\imath} + 3y^2x^6\hat{\jmath}$ which is easier in this case but not always. For example, we might want to take ∇f^{10} where $f = (x^5y + 6x^7y^5 + xy)$ which we don't want to raise to the tenth power before we differentiate.

What does the gradient tell us about a function?

Lemma Suppose $f \in C^1$ and let $\nabla f(P)$ denote the gradient of f evaluated at the point P. If $\nabla f(P) \neq 0$ then the largest value of $D_{\vec{u}} f$ is $\|\nabla f(P)\|_2$ and occurs when $\vec{u}$ points in the direction of $\nabla f(P)$.

This lemma tells us that at a point P the function f increases most rapidly in the direction of ∇f and decreases most rapidly in the direction of $-\nabla f$. This fact is the basis for a minimization algorithm called *Steepest Descent* which seeks to minimize a function of several variables.

The proof of this lemma follows from the fact that

$$D_{\vec{u}} f = \nabla f \cdot \vec{u} = \|\nabla f\|_2 \|\vec{u}\|_2 \cos\theta = \|\nabla f\|_2 \cos\theta$$

where we have used the definition of the scalar product and the fact that $\vec{u}$ is a unit vector. Clearly $D_{\vec{u}} f$ has its maximum value when $\cos\theta = 1$, i.e., when $\theta = 0$. Thus the maximum value of $D_{\vec{u}} f$ is $\|\nabla f\|_2$ and the direction of $\vec{u}$ and ∇f are the same.

Example Let $f(x, y) = xe^{2y-x}$. Determine the direction from the point P=$(2, 1)$ that gives the direction in which f decreases most rapidly.

We just need to calculate ∇f, evaluate it at the given point and take the negative of this vector. We have

$$\nabla f = \begin{pmatrix} e^{2y-x} - xe^{2y-x} \\ 2xe^{2y-x} \end{pmatrix} \qquad \nabla f(P) = \begin{pmatrix} e^0 - 2e^0 \\ 4e^0 \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$$

Therefore the direction of maximum decrease is $\hat{\imath} - 4\hat{\jmath}$.

Divergence

The divergence of a vector, denoted $\nabla \cdot \vec{v}$ is a *scalar* which can be thought of as taking the scalar product of the vector operator ∇ with a given vector. So if $\vec{v} = (v_1, v_2, \cdots, v_n)$ then the divergence of $\vec{v}$ is

$$\nabla \cdot \vec{v} = \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \cdots + \frac{\partial v_n}{\partial x_n}$$

So we must take the divergence of a vector and the result is a scalar.

Example Determine the divergence of $\vec{F} = (x^3 y, y^2 + yz, xyz)$.

$$\nabla \cdot \vec{F} = \frac{\partial}{\partial x}(x^3 y) + \frac{\partial}{\partial y}(y^2 + yz) + \frac{\partial}{\partial z}(xyz) = 3x^2 y + 2y + z + xy$$

Curl

The curl of a vector results in a vector; it is denoted $\nabla \times \vec{v}$. For a vector in R^2 or R^3 we can find the curl by evaluating the determinant

$$\nabla \times \vec{v} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ v_1 & v_2 & v_3 \end{vmatrix} = \hat{\imath}\Big((v_3)_y - (v_2)_z\Big) + \hat{\jmath}\Big((v_1)_z - (v_3)_x\Big) + \hat{k}\Big((v_2)_x - (v_1)_y\Big)$$

In $\mathbb{R}^2$ the curl of a vector is just $\hat{k}\big((v_2)_x - (v_1)_y\big)$. In fluid flow the curl of a vector gives its rotation.

Laplacian

When we take the divergence of the gradient of a function then we give this the special name *Laplacian* and denote it by Δf. We have

$$\nabla \cdot \nabla f(x,y,z) = \nabla \cdot \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} = f_{xx} + f_{yy} + f_{zz} = \Delta f$$

When $f(x,y)$ then $\Delta f = f_{xx} + f_{yy}$. The partial differential equation $-\Delta u = g(x,y)$ is the two dimensional analogue of the ODE $-y''(x) = g(x)$.

Vector Identities

There are a myriad of useful vector identities related to the gradient, divergence and curl. One in particular is the divergence of the curl of a vector which the components of the vector are in C^2

$$\nabla \cdot (\nabla \times \vec{F}) = 0$$

This can be shown by explicitly calculating the terms

$$\nabla \cdot (\nabla \times \vec{F}) = (F_3)_{yx} - (F_2)_{zx} + (F_1)_{zy} - (F_3)_{xy} + (F_2)_{xz} - (F_1)_{yz}$$

and because F has components in C^2, the order of differentiation here does not matter so we get zero.

You may explore other identities in the homework.

Newton's Method and the Jacobian matrices

If we have a single nonlinear equation such as $f(x) = \sin \pi x - x^2 = 0$ then we can approximate the solution using Newton's method which if you recall from calculus is an iterative method where the initial guess is given, say x^0 and we determine the remaining iterates from the equation

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$$

For example with $f(x) = \sin \pi x - x^2 = 0$, if $x^0 = 1$ then $x^1 = 1 - (\sin \pi - 1)/(\pi \cos \pi - 2) = 1 - 1/\pi =$ because $f'(x) = \pi \cos \pi x - 2x$.

If we have n nonlinear equations in n independent variables, i.e., n equations of the form

$$f_i(x_1, x_2, \ldots, x_n) = 0$$

then we can use Newton's Method in higher dimensions. Our iterations are now vectors in $\mathbb{R}^n$ and we have a vector $\vec{F}(x^k) = (f_1(\vec{x}^k), f_2(\vec{x}^k), \ldots, f_n(\vec{x}^k))^T$ instead of $f(x)$. But what plays the role of $f'(x^k)$ when we have n functions and n unknowns. Because each f_i is a function of n independent variables, we have n partial derivatives for *each* f_i and because we have n functions we have a total of n^2 first partial derivatives. We represent these as a matrix which we call the *Jacobian* matrix

$$J(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

Newton's method for the nonlinear system then becomes: given $\vec{x}^0$ then for $k = 0, 1, 2, \ldots$

$$\vec{x}^{k+1} = \vec{x}^k - J^{-1}(\vec{x}^k)\vec{F}(x^k)$$

Of course we know that we never actually form the inverse of a matrix so we find the new iterate by solving the linear system for the difference $\vec{\Delta x}^{k+1} = \vec{x}^{k+1} - \vec{x}^k$

$$J(\vec{x}^k)\vec{\Delta x}^{k+1} = -\vec{F}(x^k)$$

and setting

$$\vec{x}^{k+1} = \vec{x}^k + \vec{\Delta x}^{k+1}$$

Example Find the Jacobian matrix for the nonlinear system

$$f_1(x,y) = x^2 \sin \pi y + 2$$

$$f_2(x,y) = e^{xy}$$

and evaluate it at the point (2,1).

The Jacobian is the matrix of partial derivatives so in the first row we put the first partials of f_1 and in the second row we put the first partials of f_2. We have

$$J = \begin{pmatrix} 2x \sin \pi y & \pi x^2 \cos \pi y \\ y e^{xy} & x e^{xy} \end{pmatrix}$$

and J evaluated at (2,1) is

$$\begin{pmatrix} 0 & 4\pi \\ e^2 & 2e^2 \end{pmatrix}$$

The Hessian matrix

Now suppose that we have a single function of n variables and our goal is the minimize (or maximize) this function, say $f(x_1, x_2, \ldots, x_n)$. Recall from calculus that if we have a function $f(x)$ and we want to maximize it, then we find the critical points, i.e., the points where $f'(x) = 0$ or where the $f'(x)$ fails to exist. Then the maximum occurs at a critical point or the boundary of the domain so you simply evaluate f at each critical point and the boundaries and see where it takes on its maximum value.

If we extend this idea to a function $f(x_1, x_2, \ldots, x_n)$ then we would take all n partial derivatives and set the equations to zero and solve. However, the equations would typically be nonlinear equations which we would have to solve by a technique such as Newton's method which means that we would have to form the Jacobian of the equations $\frac{\partial f}{\partial x_1} = 0$, $\frac{\partial f}{\partial x_2} = 0$, etc. which means that we would be calculating a matrix of *second* derivatives of the matrix f. This matrix of second derivatives is called the *Hessian matrix*.

Definition Let $f = f(x_1, x_2, \ldots, x_n)$, then the Hessian H of f is the matrix of second partial derivatives of f where

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Example Let $f(x, y, z) = x^3 z + yz^2$. Calculate ∇f and the Hessian of f.

$$\nabla f = \begin{pmatrix} 3x^2 z \\ z^2 \\ x^3 + 2yz \end{pmatrix}$$

The Hessian matrix is given by

$$\begin{pmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{pmatrix} = \begin{pmatrix} 6x & 0 & 3x^2 \\ 0 & 0 & 2z \\ 3x^2 & 2z & 2y \end{pmatrix}$$