Multivariate Newton's Method

Two topics the book doesn't cover:

(1) Multivariable root finding

(2) Optimization

(1) Suppose we want a solution to

$$f_1(x_1, x_2) = 0$$
$$f_2(x_1, x_2) = 0$$

What is the equivalent Newton method? 1D: $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$

We are now looking for a vector: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \vec{x}$

Begin with the multi-variable Taylor's Thm:

$$f(x,y) = f(x_0, y_0) + \left[\frac{\partial f}{\partial x}(x_0,y_0)\right](x-x_0) + \left[\frac{\partial f}{\partial y}(x_0,y_0)\right](y-y_0)$$

2nd order terms

$$\left[\begin{aligned} &+\frac{1}{2}\left[\frac{\partial^2 f}{\partial x \partial y}(x_0-y_0)\right](x-x_0)(y-y_0) + \frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(x_0,y_0)\right](x-x_0)^2 \\ &+ \frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(x_0,y_0)\right](y-y_0)^2 + \ldots \end{aligned}\right.$$

$\boxed{\Rightarrow \text{Write out Hessian}}$

Note: f is scalar valued ← Not our scenario!

Our problem: $f_1(x_1,x_2) = 0$, $f_2(x_1,x_2) = 0$ $\iff$ $\vec{f}(\vec{x}) = 0$

vector vector

What does Taylor's Thm. say now?

$$\vec{f}(\vec{x}) = \vec{f}(\vec{x}_0) + \left[\frac{\partial \vec{f}}{\partial x_1}(\vec{x}_0)\right](x_1-x_{10}) + \left[\frac{\partial \vec{f}}{\partial x_2}(\vec{x}_0)\right](x_2-x_{20}) + \ldots$$

$$= \vec{f}(\vec{x}_0) + \begin{pmatrix} \frac{\partial f_1}{\partial x_1} \\ \frac{\partial f_2}{\partial x_1} \end{pmatrix}_{\vec{x}_0}(x_1-x_{10}) + \begin{pmatrix} \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_2} \end{pmatrix}_{\vec{x}_0}(x_2-x_{20}) + \ldots$$

$$= \vec{f}(\vec{x}_0) + \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix}_{\vec{x}_0} \begin{pmatrix} x_1 - x_{10} \\ x_2 - x_{20} \end{pmatrix} + \ldots$$

$$\approx \vec{f}(\vec{x}_0) + \overbrace{\underbrace{(D\vec{f})}_{\substack{\text{matrix of} \\ \text{mixed partial} \\ \text{derivatives}}}(\vec{x}_0)\;(\vec{x}-\vec{x}_0)}^{\text{matrix vector multiply}}$$

And for the second derivative?

Next terms in the expansion should be:

$$\frac{1}{2}\left[\frac{\partial^2 \vec{f}}{\partial x_1^2}(\vec{x}_0)\right](x_1 - x_{10})^2 + \frac{2\cdot 1}{2}\left[\frac{\partial^2 \vec{f}}{\partial x_1 \partial x_2}(\vec{x}_0)\right](x_1 - x_{10})(x_2 - x_{20})$$

$$+ \frac{1}{2}\left[\frac{\partial^2 \vec{f}}{\partial x_2^2}(\vec{x}_0)\right](x_2 - x_{20})^2 + \dots$$

$$= \frac{1}{2}\begin{pmatrix} \frac{\partial^2 f_1}{\partial x_1^2} \\ \frac{\partial^2 f_2}{\partial y_1^2} \end{pmatrix}(x_1 - x_{10})^2 + \frac{1}{2}\begin{pmatrix} \frac{\partial^2 f_1}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f_2}{\partial x_1 \partial x_2} \end{pmatrix}(x_1 - x_{10})(x_2 - x_{10}) + \frac{1}{2}\begin{pmatrix} \frac{\partial^2 f_1}{\partial x_2 \partial x_1} \\ \frac{\partial^2 f_2}{\partial x_2 \partial x_1} \end{pmatrix}(x_2 - x_{20})(x_1 - x_{10})$$

$$+ \frac{1}{2}\begin{pmatrix} \frac{\partial^2 f_1}{\partial x_2^2} \\ \frac{\partial^2 f_2}{\partial x_2^2} \end{pmatrix}(x_2 - x_{20})^2$$

= third order tensor: $\vec{H}(\vec{f}) = \left(\vec{H}(f_1) \quad \vec{H}(f_2)\right)$

Back to the vector approximation:

$$\vec{f}(\vec{x}) \approx \vec{f}(\vec{x}_0) + (D\vec{f})(\vec{x}_0)(\vec{x}-\vec{x}_0)$$

In particular, if $\vec{x}^*$ solves $\vec{f}(\vec{x}) = \vec{0}$, then

$$\vec{0} \approx \vec{f}(\vec{x}_0) + (D\vec{f})(\vec{x}_0)(\vec{x}^*-\vec{x}_0)$$

Solving for $\vec{x}^*$, we have

$$\vec{x}^* \approx \vec{x}_0 - \underbrace{(D\vec{f})}^{-1}(\vec{x}_0)\,\vec{f}(\vec{x}_0)$$

$D\vec{f}$ is known as the <u>Jacobian</u>

It might <u>not</u> be invertible all the time...

<u>Example</u>

$$\vec{f}(\vec{x}) = \begin{pmatrix} x^2+y^2+z^2-3 \\ x^2+y^2-z-1 \\ x+y+z-3 \end{pmatrix}$$

What is the Jacobian $(D\vec{f})(\vec{x})$?

$$(D\vec{f})(\vec{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \\ \frac{\partial f_2}{\partial x} & \ddots & \\ \frac{\partial f_3}{\partial x} & & \frac{\partial f_3}{\partial z} \end{pmatrix}$$

$$(D\vec{f})(\vec{x}) = \begin{pmatrix} 2x & 2y & 2z \\ 2x & 2y & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

Newton says:

$$\vec{x}_{k+1} = \vec{x}_k - (Df)^{-1}(\vec{x}_k)\, \vec{f}(\vec{x}_k)$$

Or equivalently; solve:

$$(Df)(\vec{x})\, \underbrace{(\vec{x}_{k+1} - \vec{x}_k)}_{\vec{y}} = -\vec{f}(\vec{x}_k)$$

then $\vec{x}_{k+1} = \vec{y} + \vec{x}_k$

Initial guess: $\vec{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

① $\Rightarrow$ solve $\begin{pmatrix} 2 & 0 & 2 \\ 2 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix} \vec{y} = -\begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

$\Rightarrow \vec{x}_1 = \left(\frac{3}{2}, \frac{1}{2}, 1\right)^t$

② $\Rightarrow$ solve $\begin{pmatrix} 3 & 1 & 2 \\ 3 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_2 - \frac{3}{2} \\ y_2 - \frac{1}{2} \\ z_2 - 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{pmatrix}$

$$\Rightarrow \quad \vec{x}_2 = \begin{pmatrix} 9/8 \\ 7/8 \\ 1 \end{pmatrix}. \qquad \vec{x}_k \text{ converges to } \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

What if $\vec{x}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$?

Then $(D\vec{f})(\vec{x}_0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix}$ Not invertible!

Newton fails.

Later on: optimization

(Most algorithms) are very close to Newton for root finding:

Problem: Minimize $f(\vec{x})$ (Simplest, "global" problem)

$$\Leftrightarrow \quad f'(\vec{x}) = \vec{0}$$

$$\Leftrightarrow \quad (Df)(\vec{x}) = \vec{0}$$

$$\Leftrightarrow \quad \underbrace{\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}}_{\vec{F} = \nabla f}(\vec{x}) = 0$$

$\Rightarrow$ Perform "Newton" on $\vec{F}$, a vector-valued function.