EE5606 : Convex Optimization

Srujana B - MA17BTECH11001
Nikhil Gandi - MA17BTECH11002
Vyshnavi - MA17BTECH11005
Venkata Datta Sai - MA17BTECH11007
Aravind Reddy K V - MA17BTECH11010

Mathematics and Computing, IIT-Hyderabad

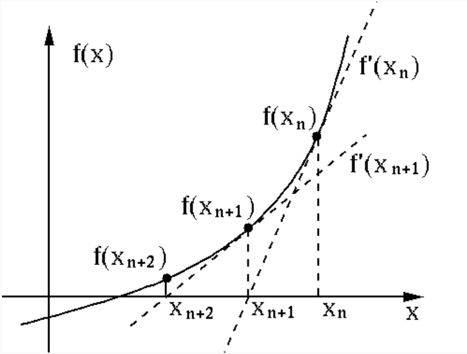
Newton’s Method and it’s variants

Newton's Method

The classic Newton-Raphson method for root finding is :

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

This method approximates $f(x)$ to be **linear** at x_n and finds the root of that function. This root is our new x_n.



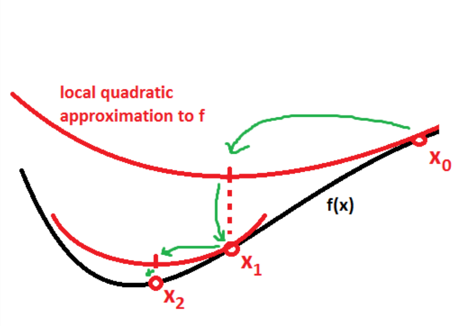
Newton's Method in Optimization

- In optimization, we look for minimum or maximum of an objective function.
- This can be obtained by setting derivative of objective function to 0. So, we are interested in finding roots of $\boldsymbol{f'(x) = 0}$.
- This can be found by using the previously discussed Newton-Raphson Method to $f'(x)$ instead of $f(x)$.

Hence, our iterative scheme changes to the following :

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$$

Graphical Representation of Newton’s Method in Optimization



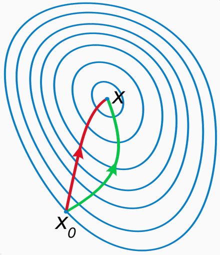
In this method, we approximate f_x to be **quadratic function** at x_n and then try to minimize that function. That point is our new x_n.

Comparison with Gradient Descent

The Gradient Descent Method is a relatively naive approach to solve optimization problems.

$$x_{n+1} = x_n - \alpha_n * \nabla f(x_n)$$

Gradient Descent has **first order** of convergence while the Newton's Method has **second order** of convergence.



A comparison of Gradient Descent and Newton's Method.

Newton’s Method in Multiple Dimensions

Idea : Make a second-order approximation and then minimize that.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be sufficiently smooth.

From **Taylor’s Theorem**, we have

$$f(x) \approx f(a) + g^T(x - a) + \frac{1}{2}(x - a)^T H(x - a)$$

where $g = \nabla f(a)$ and $H = \nabla^2 f(a)$

Newton's Method in Multiple Dimensions

We can rearrange the terms in the Taylor's Theorem to get :

$$f(x) \approx \frac{1}{2}x^T Hx + b^T x + c$$

$$\text{where } b = g - H \cdot a \text{ and } c = f(a)$$

Now, we have to **minimize** $f(x)$. So, $\nabla f(x) = 0$.

$$\nabla f(x) = Hx + b = 0$$

$$\therefore x = -H^{-1}b$$

$$\implies x = -H^{-1}(g - H \cdot a) = a - H^{-1}g$$

x is a minima only if $\nabla^2 f(x) \geq 0$

$$\implies \nabla^2 f(x) = H \geq 0 \text{ i.e H must } \textbf{positive semi-definite}.$$

Newton's Method in Multiple Dimensions

Algorithm :

- Initialize - $x_0 \in \mathbb{R}^n$
- Iterate - $x_{n+1} = x_n - H^{-1}g$.

  where $\boldsymbol{g = \nabla f(a)}$ and $\boldsymbol{H = \nabla^2 f(a)}$

Issues :

Inverting the Hessian may not be easy as the dimension increases.

Rather than finding H^{-1}, we can solve for $H \cdot y = g$.

Now, use the iterative method $x_{n+1} = x_n - \alpha_n \cdot y$, where $\alpha_n > 0$ is the step size introduced to move faster to the optimum.

Affine Invariance of Newton's Method

Given f, non-singular $A \in \mathbb{R}^{n\times n}$. Let $x = Ay$ and $g(y) = f(Ay)$. Newton steps on g are :

$$y^+ = y - \left(\nabla^2 g(y)\right)^{-1}\nabla g(y)$$

$$y^+ = y - \left(A^T\nabla^2 f(Ay)A\right)^{-1}A^T\nabla f(Ay)$$

$$y^+ = y - A^{-1}\left(\nabla^2 f(Ay)\right)^{-1}\nabla f(Ay)$$

Hence,

$$Ay^+ = Ay - \left(\nabla^2 f(Ay)\right)^{-1}\nabla f(Ay)$$

i.e.,

$$x^+ = x - \left(\nabla^2 f(x)\right)^{-1}\nabla f(x)$$

So, we can see that progress is independent of scaling which is why Newton's Method is **Affine Invariant**.

Newton Decrement

At a point x , we define **Newton Decrement** as

$$\lambda(x) = \left(\nabla f(x)^T \left(\nabla^2 f(x)\right)^{-1} \nabla f(x)\right)^{1/2}$$

This relates to the difference between $f(x)$ and the minimum of it's quadratic approximation:

$$\begin{aligned} f(x) - \min_y \left(f(x) + \nabla f(x)^T (y-x) + \tfrac{1}{2}(y-x)^T \nabla^2 f(x)(y-x)\right) \\ = f(x) - \left(f(x) - \tfrac{1}{2}\nabla f(x)^T \left(\nabla^2 f(x)\right)^{-1} \nabla f(x)\right) \\ = \tfrac{1}{2}\lambda(x)^2 \end{aligned}$$

Therefore, we can think of $\frac{1}{2}\lambda(x)^2$ as an an approximate upper bound on the suboptimality gap $f(x) - f^*$.'

Newton Decrement

Another Interpretation of Newton Decrement:

If Newton's direction/step is $v = -\left(\nabla^2 f(x)\right)^{-1} \nabla f(x)$, then

$$\lambda(x) = \left(v^T \nabla^2 f(x) v\right)^{1/2} = \|v\|_{\nabla^2 f(x)}$$

i.e. $\lambda(x)$ is the length of the Newton step in the norm defined by the Hessian $\nabla^2 f(x)$.

Variants of Newton Methods

Let α be a simple zero of a sufficiently differentiable function f.

$$f(x) = f(x_n) + \int_{x_n}^{x} f'(t)dt$$

If we approximate the indefinite integral by the trapezoidal rule and take $x = \alpha$, we obtain

$$0 \approx f(x_n) + \tfrac{1}{2}(\alpha - x_n)(f'(x_n) + f'(\alpha))$$

So, the new approximation x_{n+1} to α is given by

$$x_{n+1} = x_n - \tfrac{2f(x_n)}{f'(x_n)+f'(x_{n+1})}$$

The $(n+1)^{th}$ value of Newton's method is used on the RHS,

$$x_{n+1} = x_n - \tfrac{2f(x_n)}{f'(x_n)+f'(z_{n+1})}$$

$$\text{where } z_{n+1} = x_n - \tfrac{f(x_n)}{f'(x_n)}$$

Variants of Newton Methods

We rewrite the previous equation as

$$x_{n+1} = x_n - \frac{f(x_n)}{(f'(x_n)+f'(z_{n+1}))/2}$$

So, this variant of Newton's method can be viewed as obtained by using arithmetic mean of $f'(x_n)$ and $f'(z_{n+1})$ instead of $f'(x_n)$ in Newton's method.

Therefore, we call this method **Arithmetic Mean Newton (AN) Method**.

If we use **Harmonic Mean** instead of Arithmetic Mean, we get

$$x_{n+1} = x_n - \frac{f(x_n)\left(f'(x_n)+f'(z_{n+1})\right)}{2f'(x_n)f'(z_{n+1})}$$

which we call as **Harmonic Mean Newton (HN) Method**.

Variants of Newton Methods

If we approximate the indefinite integral by the **mid-point** integration rule, instead of trapezoidal rule and take $x = \alpha$, we obtain

$$0 \approx f(x_n) + (\alpha - x_n) f'\left(\frac{x_n + \alpha}{2}\right)$$

and in this case a new approximation x_{n+1} to α is given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'\left((x_n + x_{n+1})/2\right)}$$

The $(n+1)^{th}$ value of Newton's method is used on the RHS,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'\left((x_n + z_{n+1})/2\right)}$$

$$\text{where } z_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

which we call as **Mid-Point Newton (MN) Method**.

Newton’s Method in Image Processing

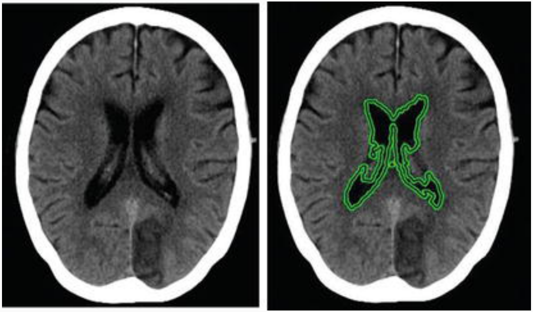
- **Image processing** is a technique which is used to derive information from the images.
- Generally used method for solving the cost function is **Gradient Descent**, which is proper but has lower rate of convergence (linear).
- Newton- type methods, on the other hand, are known to have a rapid (quadratic) convergence. In its classical form, the Newton method relies on the **L2-norm** to define the descent direction.

- In report, we generalize and reformulate this very important optimization method by introducing a novel Newton method based on **general norms**.
- This generalization opens up new possibilities in the extraction of the Newton step, including benefits such as **mathematical stability** and **smoothness constraints**.

Active Contour

- **Segmentation** is a section of image processing for the separation or segregation of information from the **required target** region of the image.
- There are different techniques used for segmentation of pixels of interest from the image.
- **Active contour** is one of the active models in segmentation techniques, which makes use of the **energy constraints** and **forces** in the image for separation of region of interest.
- It defines a separate boundary or curvature for the regions of target object for segmentation

Active Contour



Active Contour

Applications :

- Segmentation of the **medical images** i.e.,
- Various medical applications especially for the **separation of required regions** from the various medical images.
- Different types of images from various 3-D imaging modalities like **MRI**, **CT**, **PET**, **SPECT** scans can be segmented and processed with these active contour models.