HOMEWORK 11:

Exercise Set 10.1: 8 (part c)

Exercise Set 10.2: 1 (part b)

Exercise Set 10.3: 2 (part b)

Steepest Descent Method

- The steepest descent method converges slower rather than Newton's method.
- This method usually converges even for poor initial approximations. So, steepest descent method is used to find sufficiently accurate starting approximations for Newton-based methods.

The nonlinear system,

$$
\begin{aligned}
f_1(x_1, x_2, \ldots, x_n) &= 0 \\
f_2(x_1, x_2, \ldots, x_n) &= 0 \\
\vdots \qquad\qquad & \quad \vdots \\
f_n(x_1, x_2, \ldots, x_n) &= 0
\end{aligned}
$$

has a solution at $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ precisely when the function g defined by

$$g(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} [f_i(x_1, x_2, \ldots, x_n)]^2$$

has the minimal value 0.

The method of Steepest Descent for finding a local minimum for g can be described as follows:

1. Evaluate g at an initial approximation $\mathbf{x}^{(0)} = \left(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}\right)^t$.

2. Determine a direction from $\mathbf{x}^{(0)}$ that results in a decrease in the value of g.

3. Move an appropriate amount in this direction and call the new value $\mathbf{x}^{(1)}$.

4. Repeat steps 1 through 3 with $\mathbf{x}^{(0)}$ replaced by $\mathbf{x}^{(1)}$.

According to steps 2 and 3, the correct direction and the appropriate distance to move in this direction must be determined.

$$\mathbf{x}^{(\mathbf{k}+1)} = \mathbf{x}^{(\mathbf{k})} + \alpha \mathbf{v}$$

$\mathbf{v}$ and α are correct direction and the appropriate movement distance, respectively.

Suppose that $\mathbf{v} = (v_1, v_2, \ldots, v_n)^t$ is a unit vector in $\mathbb{R}^n$; that is,

$$||\mathbf{v}||_2^2 = \sum_{i=1}^{n} v_i^2 = 1$$

The **directional derivative** of g at $\mathbf{x}$ in the direction of $\mathbf{v}$ measures the change in the value of the function g relative to the change in the variable in the direction of $\mathbf{v}$. It is defined by

$$D_{\mathbf{v}} g(\mathbf{x}) = \lim_{h \to 0} \frac{1}{h}[g(\mathbf{x} + h\mathbf{v}) - g(\mathbf{x})]$$

Using n-dimensional Taylor expansion and ignoring high order terms,

$$\mathrm{g}(\mathrm{x} + \mathrm{hv}) = \mathrm{g}(\mathrm{x}) + \frac{\partial \mathrm{g}}{\partial \mathrm{x}_1} \mathrm{hv}_1 + \ldots + \frac{\partial \mathrm{g}}{\partial \mathrm{x}_n} \mathrm{hv}_n$$

So, we have,

$$D_{\mathbf{v}} g(\mathbf{x}) = \mathbf{v}^t \cdot \nabla g(\mathbf{x})$$

where,

$$\nabla g(\mathbf{x}) = \left(\frac{\partial g}{\partial x_1}(\mathbf{x}), \frac{\partial g}{\partial x_2}(\mathbf{x}), \ldots, \frac{\partial g}{\partial x_n}(\mathbf{x}) \right)^t$$

the direction that produces the maximum value for the directional derivative occurs when $\mathbf{v}$ is chosen to be parallel to $\nabla g(\mathbf{x})$, provided that $\nabla g(\mathbf{x}) \neq \mathbf{0}$.

As a consequence, the direction of greatest decrease in the value of g at $\mathbf{x}$ is the direction given by $-\nabla g(\mathbf{x})$.

The object is to reduce $g(\mathbf{x})$ to its minimal value of zero, so an appropriate choice for $\mathbf{x}^{(1)}$ is to move away from $\mathbf{x}^{(0)}$ in the direction that gives the greatest decrease in the value of $g(\mathbf{x})$.

$$x^{(1)} = x^{(0)} - \alpha \frac{\nabla g(x_0)}{\left\| \nabla g(x_0) \right\|_2} \tag{10.17}$$

The problem now reduces to choosing an appropriate value of α so that $g\left(\mathbf{x}^{(1)}\right)$ will be significantly less than $g\left(\mathbf{x}^{(0)}\right)$. Suppose that,

$$h(\alpha) = g\left(x^{(0)} - \alpha \frac{\nabla g(x_0)}{\left\| \nabla g(x_0) \right\|_2} \right)$$

The value of α that minimizes h is the value needed for Eq. (10.17).

Determination of α is demonstrated in the following example.

Example

Use the Steepest Descent method with $\mathbf{x}^{(0)} = (0, 0, 0)^t$ to find a reasonable starting approximation to the solution of the nonlinear system

$$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2 x_3) - \frac{1}{2} = 0,$$

$$f_2(x_1, x_2, x_3) = x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0$$

$$f_3(x_1, x_2, x_3) = e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0.$$

Solution

$$g(x_1, x_2, x_3) = [f_1(x_1, x_2, x_3)]^2 + [f_2(x_1, x_2, x_3)]^2 + [f_3(x_1, x_2, x_3)]^2$$

$$\begin{aligned}\nabla g(x_1, x_2, x_3) \equiv \nabla g(\mathbf{x}) = \Bigg(& 2 f_1(\mathbf{x}) \frac{\partial f_1}{\partial x_1}(\mathbf{x}) + 2 f_2(\mathbf{x}) \frac{\partial f_2}{\partial x_1}(\mathbf{x}) + 2 f_3(\mathbf{x}) \frac{\partial f_3}{\partial x_1}(\mathbf{x}), \\ & 2 f_1(\mathbf{x}) \frac{\partial f_1}{\partial x_2}(\mathbf{x}) + 2 f_2(\mathbf{x}) \frac{\partial f_2}{\partial x_2}(\mathbf{x}) + 2 f_3(\mathbf{x}) \frac{\partial f_3}{\partial x_2}(\mathbf{x}), \\ & 2 f_1(\mathbf{x}) \frac{\partial f_1}{\partial x_3}(\mathbf{x}) + 2 f_2(\mathbf{x}) \frac{\partial f_2}{\partial x_3}(\mathbf{x}) + 2 f_3(\mathbf{x}) \frac{\partial f_3}{\partial x_3}(\mathbf{x}) \Bigg) \\ = {} & 2\mathbf{J}(\mathbf{x})^t \mathbf{F}(\mathbf{x}).\end{aligned}$$

For $\mathbf{x}^{(0)} = (0, 0, 0)^t$, we have

$$g\left(\mathbf{x}^{(0)}\right) = 111.975 \quad \text{and} \quad z_0 = ||\nabla g\left(\mathbf{x}^{(0)}\right)||_2 = 419.554$$

Let

$$\mathbf{z} = \frac{1}{z_0} \nabla g\left(\mathbf{x}^{(0)}\right) = (-0.0214514, -0.0193062, 0.999583)^t$$

With $\alpha_1 = 0$, we have $g_1 = g(\mathbf{x}^{(0)} - \alpha_1 \mathbf{z}) = g(\mathbf{x}^{(0)}) = 111.975$. We arbitrarily let $\alpha_3 = 1$ so that

$$g_3 = g(\mathbf{x}^{(0)} - \alpha_3 \mathbf{z}) = 93.5649$$

Because $g_3 < g_1$, we accept α_3 and set $\alpha_2 = \alpha_3/2 = 0.5$. Thus

$$g_2 = g(\mathbf{x}^{(0)} - \alpha_2 \mathbf{z}) = 2.53557$$

(If $g_3 > g_1$ then successive divisions of α_3 by 2 are performed until , $g_3 < g_1$)

We now find the quadratic polynomial that interpolates the data (0, 111.975), (1, 93.5649), and (0.5, 2.53557). It is most convenient to use Newton's forward divided-difference inter polating polynomial for this purpose,

$$P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2)$$

This interpolates

$$g(\mathbf{x}^{(0)} - \alpha\nabla g(\mathbf{x}^{(0)})) = g(\mathbf{x}^{(0)} - \alpha\mathbf{z})$$

at $\alpha_1 = 0$, $\alpha_2 = 0.5$, and $\alpha_3 = 1$ as follows:

$$\alpha_1 = 0, \quad g_1 = 111.975,$$

$$\alpha_2 = 0.5, \quad g_2 = 2.53557, \quad h_1 = \frac{g_2 - g_1}{\alpha_2 - \alpha_1} = -218.878,$$

$$\alpha_3 = 1, \quad g_3 = 93.5649, \quad h_2 = \frac{g_3 - g_2}{\alpha_3 - \alpha_2} = 182.059, \quad h_3 = \frac{h_2 - h_1}{\alpha_3 - \alpha_1} = 400.937$$

Thus

$$P(\alpha) = 111.975 - 218.878\alpha + 400.937\alpha(\alpha - 0.5)$$

We have $P'(\alpha) = 0$ when $\alpha = \alpha_0 = 0.522959$.

Since $g_0 = g\left(\mathbf{x}^{(0)} - \alpha_0 \mathbf{z}\right) = 2.32762$ is smaller than g_1 and g_3,

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \alpha_0 \mathbf{z} = \mathbf{x}^{(0)} - 0.522959\mathbf{z} = (0.0112182, 0.0100964, -0.522741)^t$$

and

$$g\left(\mathbf{x}^{(1)}\right) = 2.32762$$

The following table contains the remainder of the results,

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$g(x_1^{(k)}, x_2^{(k)}, x_3^{(k)})$
2	0.137860	−0.205453	−0.522059	1.27406
3	0.266959	0.00551102	−0.558494	1.06813
4	0.272734	−0.00811751	−0.522006	0.468309
5	0.308689	−0.0204026	−0.533112	0.381087
6	0.314308	−0.0147046	−0.520923	0.318837
7	0.324267	−0.00852549	−0.528431	0.287024

A true solution to the nonlinear system is $(0.5, 0, -0.5235988)^t$, so $\mathbf{x}^{(2)}$ would likely be adequate as an initial approximation for Newton's method or Broyden's method.

70 iterations of the Steepest Descent method are required to find

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_\infty < 0.01$$

Picard's method

Problem in particular form $\boxed{\mathbf{A}(\mathbf{x})\mathbf{x} = \mathbf{b}(\mathbf{x})}$

If matrix is inversible, $\boxed{\mathbf{x} = \mathbf{A}(\mathbf{x})^{-1}\mathbf{b}(\mathbf{x})}$

so the iteration scheme is
$$\boxed{\mathbf{x}^{k+1} = \mathbf{A}(\mathbf{x}^k)^{-1}\mathbf{b}(\mathbf{x}^k)}$$

Attention: do not invert matrix!

$$\boxed{\begin{aligned}&\mathbf{x}^0\\ &\mathbf{A}(\mathbf{x}^k)\mathbf{x}^{k+1} = \mathbf{b}(\mathbf{x}^k)\end{aligned}}$$

Practical algorithm

- Solve one linear system per iteration
- Matrix A(x) and vector b(x) one iteration behind

Advantages

- If $A(x)$ has a special structure (e.g. banded , PD), it can be exploited when solving the linear systems

Drawbacks

- Matrix $A(x)$ may be singular for some x

- Convergence is typically linear (if it converges!)

- Computational cost: matrix $A(x)$ and vector $b(x)$ change at every iteration

Example

The nonlinear system

$$3x_1 - \cos(x_2 x_3) - \frac{1}{2} = 0,$$
$$x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0,$$
$$e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0$$

was solved in the previous examples by different methods. Represent the above system in the form A(**x**) **x** = b(**x**).

Solution

The above system can be rewritten in the following form,

$$3x_1 + 0x_2 + 0x_3 = \frac{1}{2} + \cos(x_2 x_3)$$

$$x_1 x_1 + (-81x_2 - 16.2)x_2 + 0x_3 = -\sin x_3 - 0.25$$

$$0x_1 + 0x_2 + 20x_3 = -e^{-x_1 x_2} + \frac{3 - 10\pi}{3}$$

So,

$$\begin{bmatrix} 3 & 0 & 0 \\ x_1 & -81x_2 - 16.2 & 0 \\ 0 & 0 & 20 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} + \cos\left(x_2 x_3\right) \\ -\sin x_3 - 0.25 \\ -e^{-x_1 x_2} + \frac{3 - 10\pi}{3} \end{bmatrix}$$