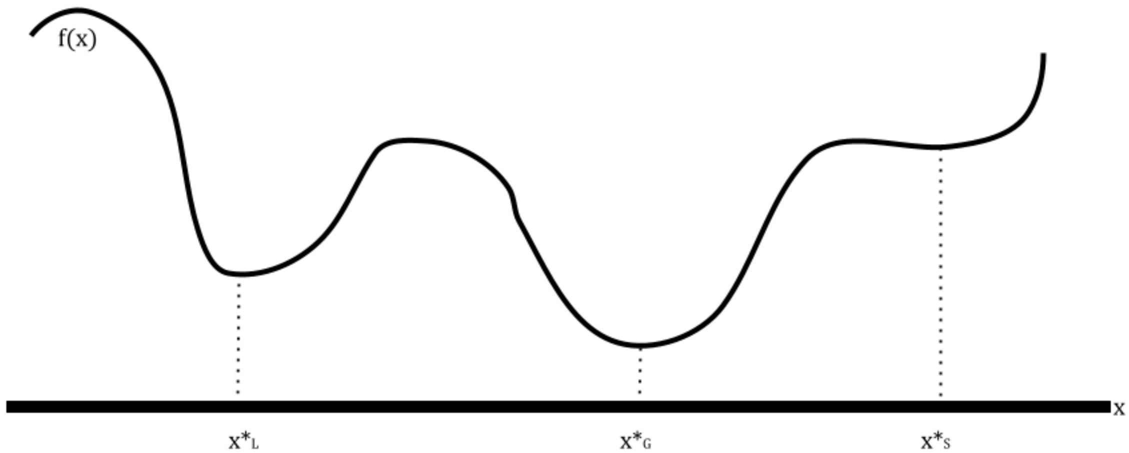
1 Descent Methods (Section 11.2)

Recall the one-dimensional function $f : \mathbb{R} \mapsto \mathbb{R}$. If $f'(x)$ is available, then we seek $f'(x) = 0$.



Newton's method is generally an effective method for finding the zeros of $f'(x)$. Also note, however, that we need f'';

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}, \qquad k = 0, 1, 2, \ldots$$

We can also consider using methods like the bisection methods to find the zero.

1.1 Many Variables Walkthrough

For many variables, we now have the function

$$F : \mathbb{R}^m \mapsto \mathbb{R}.$$

Note that the input is a vector of numbers $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$. The **gradient** of $F(x)$ is given by

$$\nabla F(x) = g(x) = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \vdots \\ \frac{\partial F}{\partial x_m} \end{bmatrix} \in \mathbb{R}^m.$$

The gradient of the gradient, known as the **Hessian matrix**, is then given by

$$\nabla^2 F(x) = H(x) = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1 \partial x_1} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_m} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{\partial^2 F}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 F}{\partial x_2 \partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_m \partial x_1} & \frac{\partial^m F}{\partial x_m \partial x_2} & \cdots & \frac{\partial^2 F}{\partial x_m \partial x_m} \end{bmatrix}.$$

It's typical to use 2 terms in a Taylor expansion $F(x_k + p)$, with $p \in \mathbb{R}^m$. So,

$$F(x_k + p) = F(x_k) + p^T \nabla F(x_k) + \frac{1}{2} p^T \nabla^2 F(x_k) p + \ldots$$

How do we define the update p? Because, that way, we could do $x_{k+1} = x_k + p$.

1.2 Steepest Descent Method

The goal of the steepest descent method is to make $p^T \nabla f(x_k) < 0$ as negative as possible. To do this, we'll consider any p so $||p||_2 = 1$. Then, by the Cauchy-Schwartz inequality,

$$-||p||_2 \cdot ||\nabla f(x_k)||_2 \leq p^T \nabla f(x_k) \leq ||p||_2 \cdot ||\nabla f(x_k)||_2.$$

Then, we're looking for

$$p^T \nabla f(x_k) \geq -||p||_2 \cdot ||\nabla f(x_k)||_2 = -||\nabla f(x_k)||_2.$$

From there, we have

$$p = -\frac{\nabla f(x_k)}{||\nabla f(x_k)||_2}.$$

Note that, in this case, p is a multiple of the negative gradient, i.e., the sharpest descent direction. A line-search strategy to finding this p is

$$\min_{\alpha > 0} F(x_k - \alpha \nabla F(x_k)).$$

1.2.1 Steepest Descent Algorithm

The algorithm for steepest descent takes the following arguments:

- $x_0 \in \mathbb{R}^m$: the initial starting point.
- $\epsilon > 0$: the tolerance.
- M: the maximum number of iterations.

Algorithm 1 Steepest Descent

function $\textsc{Steepest}(x_0, \epsilon, N)$
  $k \leftarrow 0$
  while $||\nabla f(x_k)||_2 > \epsilon$ and $k \leq M$ **do**
    $\alpha_k \leftarrow \min_\alpha F(x_k - \alpha \nabla F(x_k))$
    $x_{k+1} \leftarrow x_k - \alpha \nabla F(x_k)$
    $k \leftarrow k + 1$
  end while
end function

Remarks:

- This algorithm doesn't check if the point that's found is a minimum.
- This algorithm may "zig-zag."

With this in mind, we now think about the directional derivative of $F(x_k + \alpha p)$. In particular, note that

$$0 = F'(x_k + \alpha p)|_{\alpha = \alpha k} = p^T \nabla F(x_k + \alpha_k p) = p^T \nabla F(x_{k+1}).$$

In steepest descent, $p = -\nabla F(x_k)$, so

$$0 = -\nabla F(x_k)^T \nabla F(x_{k+1}).$$

(Example: Test Function.) Consider Rosenbrock's function

$$F(x) = 100(x_1^2 - x_2)^2 + (1 - x_1)^2,$$

with the starting point being $x = \begin{bmatrix} -1.2 \\ 1.0 \end{bmatrix}$.

Additionally, consider Powell's Singular Function,

$$F(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4,$$

with starting point $x = \begin{bmatrix} 3 \\ -1 \\ 0 \\ 1 \end{bmatrix}$.

We have two main strategies to determine iterates:

1. Line search.
   - Fix direction p.
   - Determine step length $\alpha_k = \min_{\alpha > 0} F(x_k + \alpha p)$.
2. Trust-Region
   - Fix step length $\Delta_k > 0$.
   - Determine direction p,

   $$\min_{||p|| \leq \Delta_k} \left(F_k + p^T \nabla F_k + \frac{1}{2} p^T B_k p \right).$$

   Note that $B_k \in \mathbb{R}^{m \times m}$ is a Hessian $H(x_k)$, or an approximation to it.