Unconstrained Optimization: Search Directions

AE 6310: Optimization for the Design of Engineered Systems

Spring 2017

Dr. Glenn Lightsey

Lecture Notes Developed By Dr. Brian German

Georgia Institute of Technology

Daniel Guggenheim School of Aerospace Engineering

Search Directions

Now that we have discussed how to search along a given line, we need to understand how to choose search directions.

Consider a first-order Taylor series expansion of the function $f(\boldsymbol{x})$ along a search direction in the neighborhood of the point $\boldsymbol{x_{k-1}}$ (the initial point of our line search):

$$f(\boldsymbol{x_{k-1}} + \alpha \boldsymbol{s_k}) = f(\boldsymbol{x_{k-1}}) + \alpha \boldsymbol{s}_{\boldsymbol{k}}^T \nabla f(\boldsymbol{x_{k-1}})$$

$\boldsymbol{s_k}$ is the search direction, which is typically chosen to obey $||\boldsymbol{s_k}|| = 1$.

Descent Directions

The rate of change of f along the search direction is therefore,

$$\frac{df(\boldsymbol{x})}{d\alpha} = \boldsymbol{s}_{\boldsymbol{k}}^{T} \nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})$$

We typically seek to find a *descent direction*, i.e. one that decreases the value of the objective function for positive α.

The condition for a descent direction is,

$$\frac{df(\boldsymbol{x})}{d\alpha} = \boldsymbol{s}_{\boldsymbol{k}}^{T} \nabla f(\boldsymbol{x}_{\boldsymbol{k-1}}) < 0$$

Descent Directions

We can also write this as

$$\boldsymbol{s}_{\boldsymbol{k}}^{T} \nabla f(\boldsymbol{x}_{\boldsymbol{k-1}}) = ||\boldsymbol{s}_{\boldsymbol{k}}|| \cdot ||\nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})|| \cos\theta < 0$$

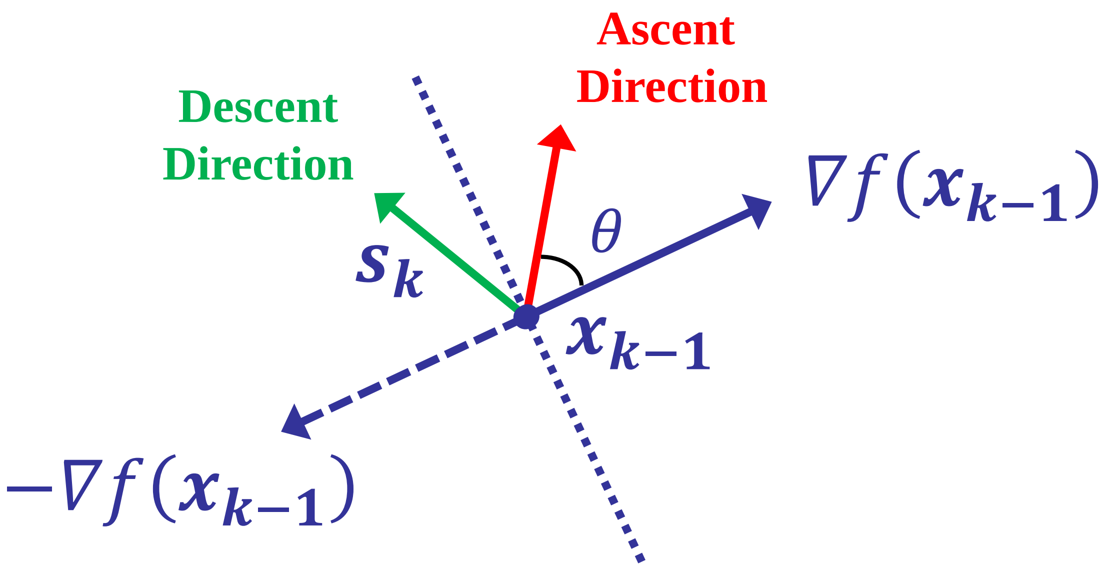
where θ is the angle between the vectors $\boldsymbol{s}_{\boldsymbol{k}}$ and $\nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})$.

Since the vector norms $||\boldsymbol{s}_{\boldsymbol{k}}||$ and $||\nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})||$ are positive, we require $\cos\theta < 0$.

$$\cos\theta < 0 \Rightarrow \quad \pi/2 < \theta < 3\pi/2$$

Descent Directions

We can visualize the situation as follows:



A descent direction forms an acute angle with $-\nabla f(\boldsymbol{x}_{k-1})$.

Steepest Descent Direction

The direction of most rapid decrease in $f(\boldsymbol{x})$ *in a small neighborhood around* $\boldsymbol{x}_{\boldsymbol{k-1}}$ can be found through the problem,

$$\min_{\boldsymbol{s}_{\boldsymbol{k}}} \ \boldsymbol{s}_{\boldsymbol{k}}^T \nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})$$

subject to $||\boldsymbol{s}_{\boldsymbol{k}}|| = 1$

Since $\boldsymbol{s}_{\boldsymbol{k}}^T \nabla f(\boldsymbol{x}_{\boldsymbol{k-1}}) = ||\boldsymbol{s}_{\boldsymbol{k}}|| \cdot ||\nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})|| \cos\theta$, the problem is solved for the value of $\boldsymbol{s}_{\boldsymbol{k}}$ for which $\theta = \pi$ and $||\boldsymbol{s}_{\boldsymbol{k}}|| = 1$:

$$\boldsymbol{s}_{\boldsymbol{k}} = \frac{-\nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})}{||\nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})||}$$

Newton Directions

Consider now a *second-order* Taylor series of the form,

$$f(\boldsymbol{x}_{\boldsymbol{k-1}} + \boldsymbol{s}_{\boldsymbol{k}}) = f(\boldsymbol{x}_{\boldsymbol{k-1}}) + \boldsymbol{s}_{\boldsymbol{k}}^T \nabla f(\boldsymbol{x}_{\boldsymbol{k-1}}) + \frac{1}{2}\boldsymbol{s}_{\boldsymbol{k}}^T H(\boldsymbol{x}_{\boldsymbol{k-1}})\, \boldsymbol{s}_{\boldsymbol{k}}$$

This is a quadratic approximation in the neighborhood of $\boldsymbol{x}_{\boldsymbol{k-1}}$. Note that α does not appear; this implies that $\alpha = 1$.

If the Hessian is positive definite, we can find a search direction that points directly toward the minimum point of this quadratic approximation.

Newton Directions

To do this, we take the derivative of the function with respect to $\boldsymbol{s}_k$ and set it equal zero,

$$\frac{df(\boldsymbol{x}_{k-1}+\boldsymbol{s}_k)}{d\boldsymbol{s}_k} = \nabla f(\boldsymbol{x}_{k-1}) + H(\boldsymbol{x}_{k-1})\,\boldsymbol{s}_k = \boldsymbol{0}$$

We can then solve for $\boldsymbol{s}_k$ to obtain,

$$\boldsymbol{s}_k = -[H(\boldsymbol{x}_{k-1})]^{-1}\,\nabla f(\boldsymbol{x}_{k-1})$$

This $\boldsymbol{s}_k$ is called the **Newton search direction.**

Newton Directions

When $H(\boldsymbol{x_{k-1}})$ is positive definite, the Newton direction is a descent direction.

When $H(\boldsymbol{x_{k-1}})$ is not positive definite, we can have two problems:

- ❖ $[H(\boldsymbol{x_{k-1}})]^{-1}$ may not exist
- ❖ Even if $[H(\boldsymbol{x_{k-1}})]^{-1}$ exists, the Newton direction may not define a descent direction

Algorithms that use Newton's method to determine search directions incorporate ways to modify $\boldsymbol{s_k}$ to deal with these problems.

Quasi-Newton Directions

An advantage of using a Newton direction is that convergence is faster (of second-order). However, the Hessian is expensive to calculate.

A common approach is therefore to approximate the Hessian and then find a **quasi-Newton search direction**. Quasi-Newton methods are sometimes called "variable metric" methods.

To see how this works, let's begin by creating a first-order Taylor series expansion for the gradient:

$$\nabla f(\boldsymbol{x_k}) = \nabla f(\boldsymbol{x_{k-1}}) + H(\boldsymbol{x_{k-1}})[\boldsymbol{x_k} - \boldsymbol{x_{k-1}}]$$

Quasi-Newton Directions

We can rearrange this equation as,

$$H(\boldsymbol{x}_{\boldsymbol{k-1}})[\boldsymbol{x}_{\boldsymbol{k}} - \boldsymbol{x}_{\boldsymbol{k-1}}] = \nabla f(\boldsymbol{x}_{\boldsymbol{k}}) - \nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})$$

This expression is of the form,

$$B_k \boldsymbol{p}_{\boldsymbol{k-1}} = \boldsymbol{y}_{\boldsymbol{k-1}}$$

where B_k is the Hessian (or Hessian-like) matrix and,

$$\boldsymbol{p}_{\boldsymbol{k-1}} = \boldsymbol{x}_{\boldsymbol{k}} - \boldsymbol{x}_{\boldsymbol{k-1}}$$

$$\boldsymbol{y}_{\boldsymbol{k-1}} = \nabla f(\boldsymbol{x}_{\boldsymbol{k}}) - \nabla f(\boldsymbol{x}_{\boldsymbol{k-1}})$$

Quasi-Newton Directions

Quasi-newton methods work by developing approximations for $B_k \approx H(\boldsymbol{x}_{\boldsymbol{k-1}})$ by enforcing the relation,

$$B_k \boldsymbol{p}_{\boldsymbol{k-1}} = \boldsymbol{y}_{\boldsymbol{k-1}}$$

that holds for the true Hessian $H(\boldsymbol{x}_{\boldsymbol{k-1}})$.

The approximations typically enforce additional constraints on B_k:

- Symmetry – because $H(\boldsymbol{x}_{\boldsymbol{k-1}})$ is symmetric
- Low rank of the matrix $[B_k - B_{k-1}]$– implies few rows of the Hessian change in each iteration

Quasi-Newton Directions

A common approach for generating the approximation B_k is the Broyden, Fletcher, Goldfarb, Shanno (BFGS) update formula:

$$B_{k+1} = B_k + \frac{(\boldsymbol{y_k} - B_k \boldsymbol{p_k})(\boldsymbol{y_k} - B_k \boldsymbol{p_k})^T}{(\boldsymbol{y_k} - B_k \boldsymbol{p_k})^T \boldsymbol{p_k}}$$

Set $B_1 = I$ (identity matrix) such that the initial search direction is steepest descent.

Quasi-Newton Directions

Once the approximation B_k has been found, we can use it to define the search direction in a similar way that we used the Hessian to define a Newton direction:

Newton direction:

$$\boldsymbol{s_k} = -[H(\boldsymbol{x_{k-1}})]^{-1} \nabla f(\boldsymbol{x_{k-1}})$$

Quasi-Newton direction:

$$\boldsymbol{s_k} = -B_k^{-1} \nabla f(\boldsymbol{x_{k-1}})$$

Since only B_k^{-1} is needed to define the search direction, some methods compute it directly, without the step of computing B_k .

Conjugate Directions

Let A be an $n \times n$ symmetric matrix. A set of n vectors (or directions) $\{\boldsymbol{s_i}\}$ is said to be conjugate with respect to the matrix A if,

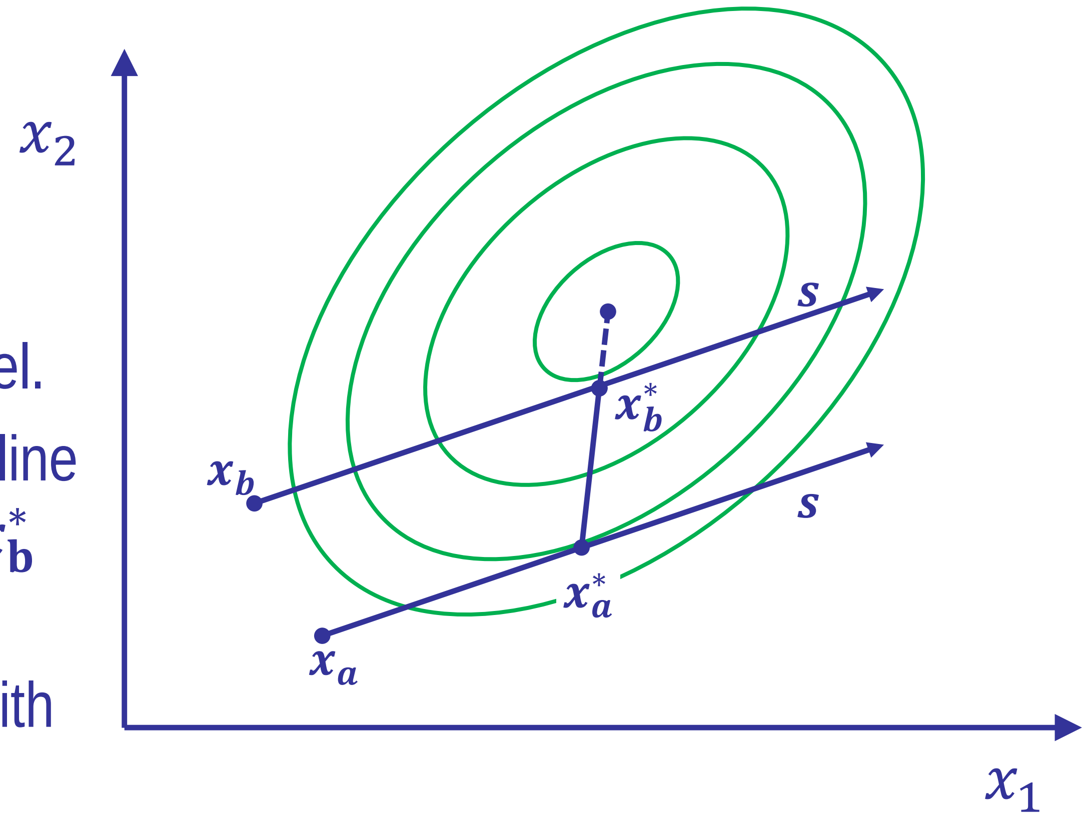
$$\boldsymbol{s}_{\boldsymbol{i}}^{T} A \boldsymbol{s}_{\boldsymbol{j}} = 0$$

for all $i \neq j,\ \ i = 1, \dots, n,\ \ \ j = 1, \dots, n.$

Conjugate Directions

Consider a quadratic function: $f(\boldsymbol{x}) = 1/2\, \boldsymbol{x}^T A \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} + \boldsymbol{c}$

- Draw a line from a point $\boldsymbol{x_a}$ in a direction $\boldsymbol{s}$.
- Draw a 2nd line from a point $\boldsymbol{x_b}$ in the same direction $\boldsymbol{s}$. These first two lines are parallel.
- Find the minimum along each line and call these points $\boldsymbol{x}_\mathbf{a}^*$ and $\boldsymbol{x}_\mathbf{b}^*$
- The direction $(\boldsymbol{x}_\mathbf{b}^* - \boldsymbol{x}_\mathbf{a}^*)$ is <u>conjugate</u> to the direction $\boldsymbol{s}$ with respect to the matrix A.



Adapted from Rao, S., *Engineering Optimization: Theory and Practice*, 3rd Edition, Wiley Interscience, 1996.

Conjugate Directions

Note that mutually perpendicular directions (such as the coordinate directions) are conjugate with respect to the identity matrix, I_n.

Conjugate Directions: So What?

The wonder of conjugate directions is the following theorem, quoted from (Rao, 1996), which we state without proof:

Theorem: If a quadratic function $f(\boldsymbol{x}) = 1/2\,\boldsymbol{x}^T A\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + \boldsymbol{c}$ is minimized sequentially, once along each direction of a set of n mutually conjugate directions, the minimum of the function $f(x)$ will be found at or before the nth step, irrespective of the starting point.

Rao, S., *Engineering Optimization: Theory and Practice*, 3rd Edition, Wiley Interscience, 1996.