Unconstrained Optimization
Multi-dimensional

86.5.2 Steepest Descent Method

$-\nabla f(x)$ direction of steepest descent (locally)

potentially useful direction to move
but step size?

Define

$$\phi(\alpha) = f(x + \alpha s)$$

one-dimensional problem

$$s = -\nabla f$$

"steepest descent method"

$x_0 = \text{initial guess}$

for $k = 0, 1, 2, \ldots$

$$s_k = -\nabla f(x_k)$$

choose $\alpha_k$ to minimize $f(x_k + \alpha s_k)$ "line search"

$$x_{k+1} = x_k + \alpha_k s_k$$

end

always makes progress, but iterates can zigzag.

Linear conv, w/ factor arbitrarily close to $1$.

Example 6.11

$$f(x) = 0.5x_1^2 + 2.5x_2^2$$

$$\nabla f = \begin{pmatrix} x_1 \\ 5x_2 \end{pmatrix}$$

$$\bar{x}_0 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

$$\Delta D_{opt.} \Rightarrow \alpha_0 = \frac{1}{3}$$

$$\bar{x}_1 = \bar{x}_0 + \frac{1}{3} s_0 = \begin{pmatrix} 5 \\ 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 3.333 \\ 0.667 \end{pmatrix}$$

stop when $\|\nabla f\|$ small.

contour when $f$ constant

gradient @ $x$ normal to level set

min occurs when $\nabla f(x + \alpha s) \perp s$
Example 6.12 (Newton's Method)

\[ f(x) = 0.5x_1^2 + 2.5x_2^2 \]

\[ \nabla f(x) = \begin{pmatrix} x_1 \\ 5x_2 \end{pmatrix}, \quad H_f(x) = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \]

\[ \overrightarrow{x_0} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \quad \nabla f(x_0) = \begin{pmatrix} 5 \\ 5 \end{pmatrix} \]

\[ H_f s = -\nabla f \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} s_0 \end{pmatrix} = \begin{pmatrix} -5 \\ -5 \end{pmatrix} \Rightarrow \begin{pmatrix} s_0 \end{pmatrix} = \begin{pmatrix} -5 \\ -1 \end{pmatrix} \]

\[ \begin{pmatrix} x_1 \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix} + \begin{pmatrix} -5 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix} \]

- converged in a single iteration
- not surprising since \( f \) is quadratic, truncated Taylor series of \( f \) (to \( h^2 \)) is exact.
Newton's Method

Local quadratic approximation:
\[
f(x+s) \approx f(x) + \nabla f(x)^T s + \frac{1}{2} s^T H_f(x) s = g(s)
\]

\[
\min_{s} \quad \nabla g(s) = 0
\]
\[
\nabla g(s) = \nabla f(x)^T + \frac{1}{2} s^T H_f(x) = 0
\]
\[
\Rightarrow H_f(x)s = -\nabla f(x)
\]

(Newton's Method for \( \nabla f(x) = 0 \))

\( x_0 \): initial guess
for \( k = 0, 1, 2, \ldots \)
Solve \( H_f(x_k) s_k = -\nabla f(x_k) \)
\( x_{k+1} = x_k + s_k \)
end

- Steepest descent: marble zig-zags
- Newton's Method: marble rolls straight to bottom.

END

Other methods:
- Trust region method
- Quadratic model
- Damped Newton method
- levenberg-marquardt

Descent direction:
\[
\nabla f(x_k)^T s_k < 0
\]

How to address issues with N.M.: far from \( x^* \).
near solution \[ H_f(x_k) > 0 \Rightarrow s_k \text{ descent dir} \]

\[ H_f(x_k) s_k = -Df(x_k) \]

\[ s_k^T H_f(x_k) s_k = -s_k^T Df(x_k) \]

\[ > 0 \]

\[ \Rightarrow -s_k^T Df(x_k) > 0 \]

\[ \Rightarrow \text{skipped} \]

but away from solution, need alternate choice for \( s_k \), direction of negative curvature.

\[ P_k^T H_f(x_k) P_k < 0 \] (obtain \( P_k \) from symm. indef. factorization of \( H_f \))

modified Hessian \[ H_f(x_k) + \mu I > 0 \] (results in \( s_k \) between Newton step and steepest descent)

\[ \text{(6.5.4)} \text{Quasi-Newton Methods} \]

\[ x_{k+1} = x_k - \alpha_k B_k^{-1} Df(x_k) \text{ secant updating} \]

- more robust
- lower cost/iter \(-\) no 2nd deriv. eval.
- super linear conv. \(-\) \( O(n^2) \) for solve (vs. \( O(n^3) \))

\[ \text{(6.5.5)} \text{Secant Updating Scheme} \]

BFGS - preserve symmetry of Hessian
- preserve positive definiteness of Hess
Broyden's Method

\[ B_{k+1} = B_k + \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k} \]

\[ = B_k \left( I - \frac{s_k s_k^T}{s_k^T s_k} \right) + \frac{y_k s_k^T}{s_k^T s_k} \]

\[ = B_k \left( I - \frac{s_k s_k^T B_k}{s_k^T B_k s_k} \right) + \frac{y_k y_k^T}{y_k^T s_k} \]

\[ B_{k+1} s_k = B_k s_k - B_k s_k \frac{s_k B_k s_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \]

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