Unconstrained Optimization

Multi-dimensional.

§6.5.2 Steepest Descent Method

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

potent. useful direction to move
but step size ?

Define
$$\phi(\alpha) = f(\vec{x} + \alpha\vec{s})$$
"line search"
use a 1D solver.

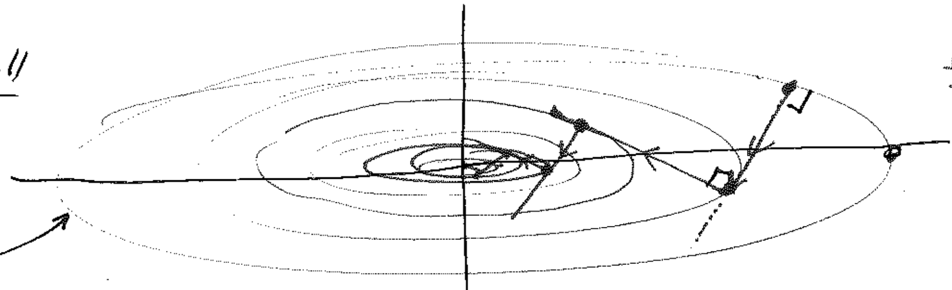
⇝ one-dimensional problem

$$S = -\nabla f$$
"steepest descent method"

x_0 = initial guess
for $k = 0, 1, 2, \ldots$
  $s_k = -\nabla f(x_k)$
  Choose α_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



$$f(x) = 0.5x_1^2 + 2.5x_2^2$$
$$\nabla f = \begin{pmatrix} x_1 \\ 5x_2 \end{pmatrix}$$
$$\vec{x}_0 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

1D opt. ⇒ $\alpha_0 = 1/3$

$$\vec{x}_1 = \vec{x}_0 + \tfrac{1}{3}\vec{s}_0 = \begin{pmatrix} 5 \\ 1 \end{pmatrix} - \tfrac{1}{3}\begin{pmatrix} 5 \\ 5 \end{pmatrix} = \begin{pmatrix} 3.333 \\ -.667 \end{pmatrix}$$

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

- contours where f = constant
- gradient @ $\vec{x}$ normal to level set

- min occurs when $\nabla f(\vec{x} + \alpha\vec{s}) \perp \vec{s}$

Example. 6.12 (Newton's Method)

$$f(\vec{x}) = .5\, x_1^2 + 2.5\, x_2^2$$

$$\nabla f(\vec{x}) = \begin{pmatrix} x_1 \\ 5x_2 \end{pmatrix} \qquad H_f(x) = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$$

$$\vec{x}_0 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \quad \nabla f(x_0) = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$$

$$H_f s = -\nabla f \Rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \vec{s}_0 = \begin{pmatrix} -5 \\ -5 \end{pmatrix} \Rightarrow \vec{s}_0 = \begin{pmatrix} -5 \\ -1 \end{pmatrix}$$

$$\vec{x}_1 = \begin{pmatrix} 5 \\ 1 \end{pmatrix} + \begin{pmatrix} -5 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

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

§6.5.3 Newton's Method

local quadratic approximation:

$$f(x+s) \approx f(x) + \nabla f(x)^T s + \frac{1}{2} s^T H_f(x) s \triangleq g(s)$$

min

$$\nabla g(s) = 0$$

$$\nabla_s g(s) = \nabla f(x)^T + \frac{2}{2} s^T H_f(x) = 0$$

$$\Rightarrow \boxed{H_f(x) s = -\nabla f(x)}$$

(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 zigzags

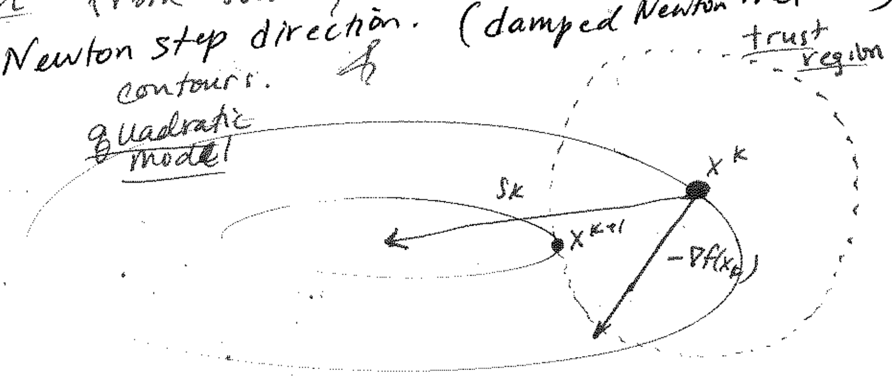
Newton's Method: marble rolls straight to bottom.

- Newton's Method: near soln, no line search required. far from soln, useful to do line search in Newton step direction. (damped Newton method)

END

~~Other Newton-based~~

skipped $\boxed{\text{trust region method}}$



descent direction:

$$\nabla f(x_k)^T s_k < 0$$

How to address issues w/ N.M. far from x^*.

near solution $\quad H_f(x_k) > 0 \Rightarrow s_k$ descent dir

$$H_f(x_k) > 0$$

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

$$\underbrace{s_k^T H_f(x_k)\, s_k}_{>0} = -\, s_k^T \nabla f(x_k)$$

$$\Rightarrow \quad -\, s_k^T \nabla f(x_k) > 0$$

$$\Rightarrow \quad \boxed{s_k^T \nabla f(x_k) < 0}$$

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

$$H_f(x_k) + \mu \mathbb{I} \; > 0.$$

(results in s_k between Newton step + steepest descent)

Skipped

$\boxed{\S\, 6.5.4}$ Quasi-Newton Methods

$$x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f(x_k)$$

secant updating

- more robust
- lower cost/iter — no 2nd deriv. eval.
- superlinear conv. — 1 ∇ eval — $O(n^2)$ for solve (vs. $O(n^3)$)

$\boxed{\S 6.5.5}$ Secant Updating Scheme.

BFGS — preserve symmetry of Hessian
— preserve positive definiteness of Hessian

Broydon's Method

Sec 6.3.

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

not symmetric update

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

$$= B_k \left(I - \frac{s_k s_k^T}{s_k^T s_k} \right) + \frac{(f(x_{k+1}) - f(x_k)) s_k^T}{s_k^T s_k}.$$

BFGS.

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

projection

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

$$B_{k+1} s_k = B_k s_k - \frac{B_k s_k \cancel{s_k^T B_k s_k}}{\cancel{s_k^T B_k s_k}} + \frac{y_k \cancel{y_k^T s_k}}{\cancel{y_k^T s_k}}$$

in skipped

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