**Week 7: Wednesday, Mar 14**

# Line search revisited

In the last lecture, we briefly discussed the idea of a *line search* to improve the convergence of Newton iterations. That is, instead of always using the Newton update

$$x^{k+1} = x^k - f'(x^k)^{-1} f(x^k),$$

we allow ourselves to use a scaled version of the step

$$x^{k+1} = x^k - \alpha_k f'(x^k)^{-1} f(x^k),$$

where $\alpha_k$ is chosen to ensure that the iteration actually makes progress. Here, "progress" is typically measured in terms of the residual norm $\|f(x^{k+1})\|$. At the bare minimum, we want to make sure that the residual goes down at each step, but we can prove a bit more with a slightly stricter criterion:

$$\|f(x^{k+1})\| < (1 - \sigma\alpha_k)\|f(x^k)\|$$

where $\sigma$ is chosen to be some small value (say $10^{-4}$). In practice, this looks something like this:

```
% Get Newton step
[f,J] = eval_f(x);
d = J\f;

% Line search
alpha = 1;
for k = 1:maxstep

  % Try step
  xnew = x-alpha*d;
  fnew = eval_f(xnew);

  % Accept if satisfactoy
  if norm(fnew) < (1-sigma*alpha)*norm(f)
    x = xnew;
    f = fnew;
```

```
    break;
  end

  % Otherwise, cut alpha in half and try again
  alpha = alpha/2;

end
```

This line search strategy essentially relies on the fact that we can characterize a solution of $f(x) = 0$ in terms of a minimization of $\|f(x)\|$. Of course, this relationship goes the other way, too: for a differentiable objective function, we can write a nonlinear system of equations that define necessary condtions for a minimum.

# Iterations for optimization

Suppose $g : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable near $x_0$. Then you might remember that Taylor's theorem gives

$$g(x + z) = g(x) + g'(x)z + \frac{1}{2}z^T H_g(x)z + O(\|z\|^3),$$

where $H_g$ is the *Hessian matrix*

$$[H_g(x)]_{ij} = \frac{\partial^2 g(x)}{x_i x_j}.$$

A necessary conditions for $x_*$ to be a local minimum or maximum of $g$ is that $g'(x) = 0$. This suggests one way of trying to find a local minimum of $g$ is simply Newton iteration (with a line search):

$$x^{k+1} = x^k - \alpha_k H_g(x^k)^{-1} \nabla g(x^k).$$

Unfortunately, even if Newton iteration converges to a critical point (a point where the gradient of $g$ is zero), there is nothing to guarantee that this will be a minimum rather than a maximum. In order to make sure that we converge to a minimum, we would like to make sure not that $\|\nabla g\|$ decreases at each step, but that $g$ decreases at each step! There are two ensuring this decrease:

1. We need the Newton direction (or some other search direction) to at least be a *descent* direction. That is, we want
$$x^{k+1} = x^k + \alpha_k d^k$$
where $\nabla g(x^k) \cdot d^k < 0$.

2. Once we have a descent direction, we want to make sure that the steps we take are short enough that we actually decrease $g$ by some sufficient amount. The condition we use might look something like
$$g(x^{k+1}) \leq g(x^k) + \alpha^k \sigma \nabla g(x^k) \cdot d^k$$

Under what conditions can we guarantee that the Newton direction is actually a descent direction? If the Newton direction is
$$d^k = -H_g(x^k)^{-1} \nabla g(x^k),$$
then the descent condition looks like
$$\nabla g(x^k)^T d^k = -\nabla g(x^k)^T H_g(x^k)^{-1} \nabla g(x^k),$$
which is a quadratic form in $H_g(x^k)^{-1}$. So a sufficient condition for the Newton iteration to be a descent direction is that $H_g(x^k)$ is positive definite (and therefore that $H_g(x^k)^{-1}$ is positive definite). This suggests the following modification to the Newton approach to minimizing $g$:

- If the Hessian matrix $H_g(x^k)$ is positive definite, search in the Newton direction
$$d^k = -H_g(x^k)^{-1} \nabla g(x^k).$$
- If the Hessian is not positive definite at $x^k$, use a modified Newton direction
$$d^k = -\hat{H}^{-1} \nabla g(x^k).$$
where $\hat{H}$ is some positive definite matrix. Convergence tends to be fastest when $\hat{H}$ approximates the Hessian in some way (subject to the constraint of being positive definite), but one can also be lazy and just choose $\hat{H} = I$ (i.e. follow the direction of steepest descent).

Note that while it is possible to choose to a local minimum by choosing the steepest descent direction $-\nabla g(x^k)$ at *every* step, this approach can yield painfully slow convergence.

# Problems to Ponder

1. Write a (guarded) Newton iteration to find the intersection of three spheres in three dimensional space, i.e. find $x_*$ such that
$$\begin{aligned} \|x_* - x_a\| &= r_a \\ \|x_* - x_b\| &= r_b \\ \|x_* - x_c\| &= r_c \end{aligned}$$
Assume for the moment that there are exactly two solutions. If you find one, how might you easily find the other?

2. Consider the steepest descent iteration
$$x_{k+1} = x_k - \alpha_k \nabla\phi(x_k)$$
applied to
$$\phi(x) = \frac{1}{2} \begin{bmatrix} x_1 \\ x_t \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & 10^6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_t \end{bmatrix},$$
and suppose that $\alpha_k$ is chosen by *exact line search*: that is $\alpha_k$ is chosen to reduce $\phi(x_{k+1})$ as much as possible. Starting from $\begin{bmatrix} 1 & 1 \end{bmatrix}^T$, what are the iterates produced by this iteration? What can you say about the rate of convergence?

3. What is $\nabla_x \phi(x)$ for $\phi(x) = \|f(x)\|^2$? Argue based on your computation that the Newton direction is a descent direction for this objective function.

4. Write the critical point equations for minimizing $\|f(x) - b\|^2$.

5. The Gauss-Newton iteration for minimizing $\|f(x) - b\|^2$ is
$$\begin{aligned} p_k &= \left(J(x_k)^T J(x_k)\right)^{-1} J(x_k)^T (f(x_k) - b) \\ x^{k+1} &= x^k - \alpha_k p_k \end{aligned}$$
where $J(x_k)$ is the Jacobian of $f$. Argue that $p_k$ is always a descent direction.