Inexact minimization. (with derivatives)

Brent's method is a very fast and stable way to converge on a minimum to high precision. However, in the methods for multidimensional minimization we will study soon, hundreds of thousands of line minimizations may be required, and they don't have to be that accurate. We now present a fast, approximate, minimizer.

Suppose we are stepping from $x_k$ in direction $d_k$:

$$\hat{x}_{k+1} = x_k + \alpha d_k.$$
In general (for \( f: \mathbb{R}^n \rightarrow \mathbb{R} \)), we have the linear approximation (as a function of \( \alpha \))

\[
\begin{align*}
f(x_{k+1}) &= f(x_k) + \alpha \langle \nabla f(x_k), d_k \rangle.
\end{align*}
\]

(Assume \( f \) has a unique minimum at positive \( \alpha \).)

Now consider the family of lines (as functions of \( \alpha \)), parametrized by \( p \in (0, \frac{1}{2}) \),

\[
\begin{align*}
(1) \quad f(x_k) + p \langle \nabla f(x_k), d_k \rangle \alpha  \\
\end{align*}
\]

and the family

\[
\begin{align*}
(2) \quad f(x_k) + (1-p) \langle \nabla f(x_k), d_k \rangle \alpha.
\end{align*}
\]

Clearly the first family has slopes in \((0, \frac{1}{2} \langle \nabla f(x_k), d_k \rangle)\) and the second has slopes in \((\frac{1}{2} \langle \nabla f(x_k), d_k \rangle, \langle \nabla f(x_k), d_k \rangle)\).
We have the picture

where the dotted lines correspond to a particular \( p \).

Now (trust me)

\[
\Theta = \arctan \left[ \frac{-(1-2p) \langle \nabla f(x_k), d_k \rangle}{1 + \rho (1-p) \langle \nabla f(x_k), d_k \rangle^2} \right]
\]

Suppose we guess that \( x_0 \) is the minimizer.
we expect that \((x_0, f(x_0))\) is between the lines:

That is; \(x_0\) satisfies

**Definition**: The **Goldstein conditions**

for \(x_0\) are

\[
f(x_k) \leq f(x_k + x_0d_k) \leq f(x_k) + p \langle \nabla f(x_k), d_k \rangle x_0 + (1-p) \langle \nabla f(x_k), d_k \rangle x_0
\]
Our basic method will be to generate "guess" values $x_0$ based on some criteria and then check whether they satisfy the Goldstein conditions. If so, the algorithm terminates. If not, we are in the situation

\[ f(x_k + \alpha_0 d_k) > f(x_k) + \rho \langle \nabla f(x_k), d_k \rangle \alpha_0, \]

we know that the minimum is $x^*$ between $\alpha_1$ and $\alpha_0$ has $\alpha^* < \alpha_0$. 
We then choose a next guess. Since this is going to be an iterative method, we start at

\[ x_k + \alpha_L d_k \Rightarrow \alpha_L \]

and let

\[ f_L = f(x_k + \alpha_L d_k) \quad f_L' = \left. \frac{d}{d\alpha} f(x_k + \alpha d_k) \right|_{\alpha = \alpha_L} \]
\[ f_0 = f(x_k + \alpha_0 d_k) \quad f_0' = \left. \frac{d}{d\alpha} f(x_k + \alpha d_k) \right|_{\alpha = \alpha_0}. \]

Now if we can fit a quadratic to \( f_0, f_L, f_L' \) and minimize it, giving us a new estimate of

\[ \alpha = \alpha_L + \frac{(\alpha_0 - \alpha_L)^2 f_L'}{2([f_L - f_0 - (\alpha_0 - \alpha_L) f_L']}. \]
On the other hand, if
\[ f(x_k + \alpha_0 d_k) < f(x_k) + (1-p) \langle \nabla f(x_k), d_k \rangle \alpha_0, \]
we have

In this case, we must extrapolate forward to guess where the min might be. We simply use the secant method on \( f' \) to guess

\[ \hat{\alpha} = \alpha_0 + \frac{(\alpha_0 - \alpha_L)}{(f'_L - f'_0)} f'_0. \]
Now there are a few problems with this, so in practice we fix it up a little. First, we note there's no guarantee that $\alpha_1$ and $\alpha_2$ bracket the minimum.

In this case, we will never arrive at the true minimum. So instead of requiring that $f(\alpha_*) >$ the line through $\alpha_1$, let's require that the derivative of $f$ improve at $\alpha_*$ by a certain fraction.
So we let \( f(\alpha) = f(x_k + a\alpha_k) \)
and write

Fletcher's Modified Goldstein Conditions.

Choose \( p \in [0, \frac{1}{2}], \sigma \in [p, 1) \). Then \( \alpha \)
satisfies the conditions if

\[
f'(\alpha) \geq \sigma f'(0)
\]

and

\[
f(\alpha) \leq f(0) + p f'(0) \alpha.
\]

Note to readers: we've switched from multivariable to single-variable notation here (and we should rewrite the whole lecture in this form, but don't have time).
Why does this work? Well, \( \sigma f'(0) \) is still negative, but less negative than \( f'(0) \), so \( \sigma f'(0) = f'(a_1) \) for some \( a_1 < a_* \).

(Now Antoniou and Lu claim that \( a_* < a_2 \), but I don't see it.)

The final condition becomes

\[ |f'(a)| \leq -\sigma f'(0). \]

This ensures that the effective upper
bound is converging to $a^*$ as well. When all of this is combined, we get

Fletcher's Inexact Line Search.

Start with a bracket of $[a_L, a_U] = [0, 10^{99}]$, and a guess for $a_0$.

(*) for each step:

If $f_0 > f_L + p(a_0 - a_L) f_L'$

(that is, $(a_0, f_0)$ is above the line with slope $pf_L'$)

use quadratic interpolation on $(a_L, f_L), (a_0, f_0), (a_0^*, f_L')$

to guess $a$

if $a_0 < a_U$, reset $a_U$ to $a_0$

goto (*)
If $f'_0 < \sigma f'_L$

(that is, if we fell short of the point where $f'$ is $> \sigma f'_L$)

use linear interpolation on $f'$ at $\alpha_0, \alpha_L$

to guess $\alpha$

reset $\alpha_L$ to $\alpha_0$

goto (*)

Terminate.

As $\sigma \to 0$, this becomes an exact line search. Fletcher recommends $\rho = 0.1$ and $\sigma = 0.7$ for a relatively good search.

<mathematica demo>