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:

$$\vec{x}_{k+1} = \vec{x}_k + \alpha \vec{d}_k.$$

In general (for $f : \mathbb{R}^n \to \mathbb{R}$), we have the linear approximation (as a function of α)

$$f(x_{k+1}) = f(x_k) + \alpha \langle \nabla f(x_k), d_k \rangle.$$

(Assume f has a unique minimum at positive α.)

Now consider the family of lines (as functions of α), parametrized by $\rho \in (0, 1/2)$,

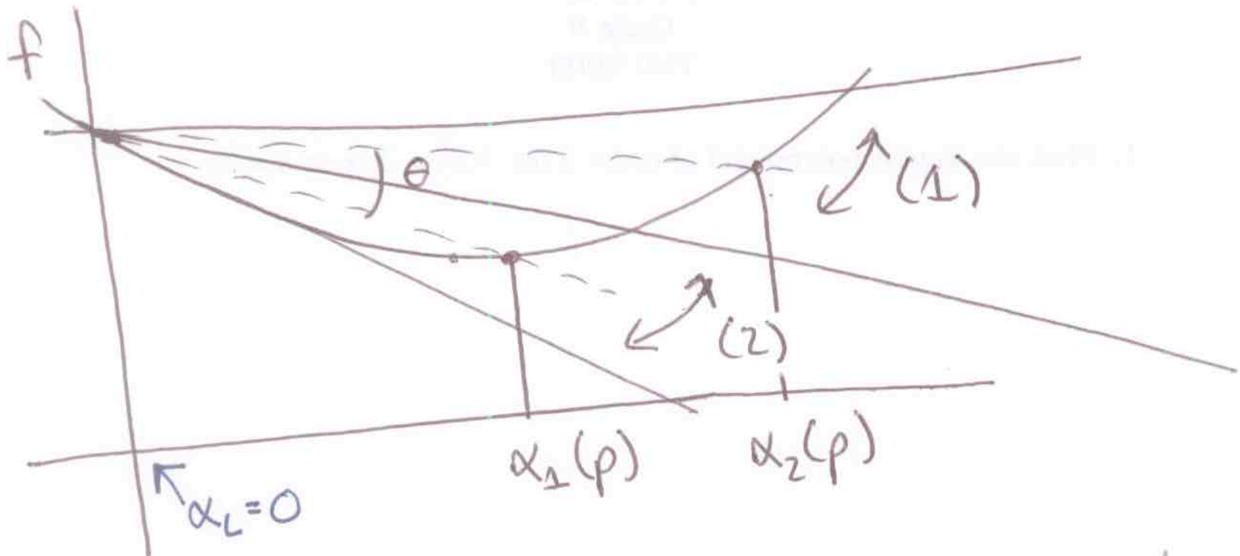
$$(1) \qquad f(x_k) + \rho \left(\langle \nabla f(x_k), d_k \rangle \right) \alpha$$

and the family

$$(2) \qquad f(x_k) + (1-\rho) \left(\langle \nabla f(x_k), d_k \rangle \right) \alpha.$$

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

We have the picture



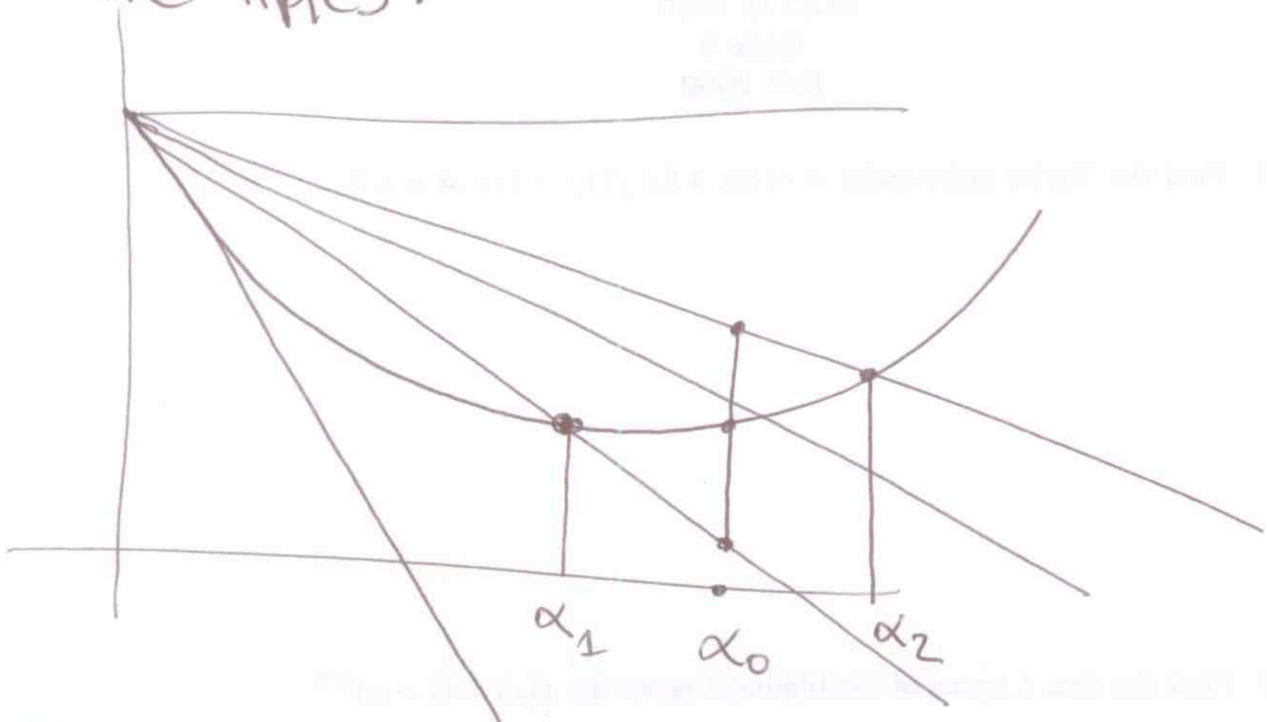
where the dotted lines ~~have~~ correspond to a particular ρ.

Now (trust me)

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

Suppose we guess that α_0 is the minimizer.

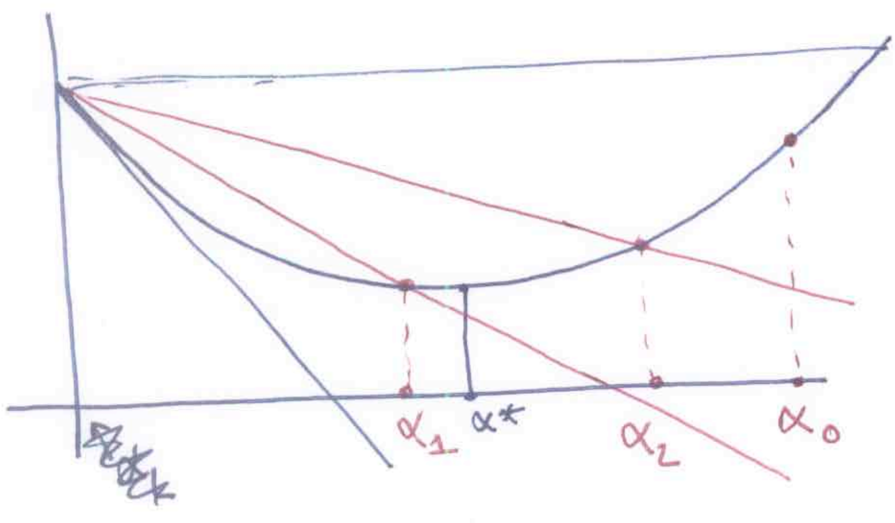
we expect that $(\alpha_0, f(\alpha_0))$ is between the lines:



That is; α_0 satisfies

> Definition: The <u>Goldstein conditions</u> for α_0 are
>
> $$f(x_k) + (1-\rho)\langle \nabla f(x_k), d_k \rangle \alpha_0 \leq f(x_k + \alpha_0 d_k) \leq f(x_k) + \rho \langle \nabla f(x_k), d_k \rangle \alpha_0$$

Our basic method will be to generate "guess" values α_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



if $f(x_k + \alpha_0 d_k) > f(x_k) + \rho \langle \nabla f(x_k), d_k \rangle \alpha_0$,

we know that the min α_* is between ~~x_k and x_k~~ has $\alpha^* < \alpha_0$.

6

We then choose a next guess. Since this is going to be an iterative method, we start at

$$x_k + \alpha_L d_k \leftrightarrow \alpha_L$$

and let

$$f_L = f(x_k + \alpha_L d_k) \quad f'_L = \frac{d}{d\alpha} f(x_k + \alpha d_k)\Big|_{\alpha=\alpha_L}$$

$$f_0 = f(x_k + \alpha_0 d_k) \quad f'_0 = \frac{d}{d\alpha} f(x_k + \alpha d_k)\Big|_{\alpha=\alpha_0}.$$

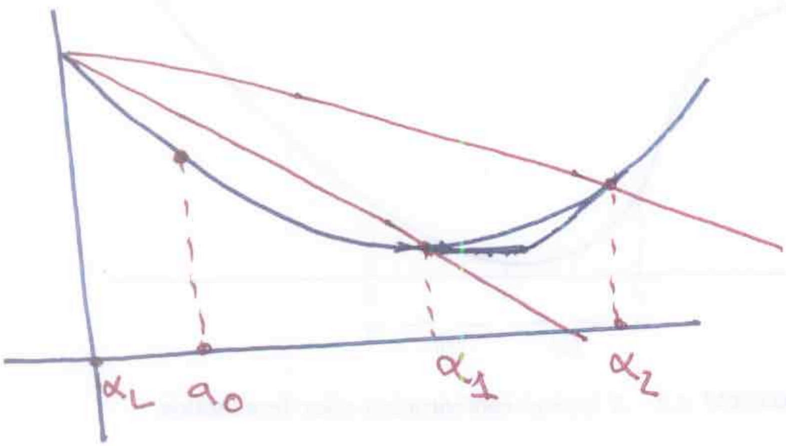
Now we can fit a quadratic to f_0, f_L, f'_L and minimize it, giving us a new estimate of

$$\breve{\alpha} = \alpha_L + \frac{(\alpha_0 - \alpha_L)^2 f'_L}{2[f_L - f_0 - (\alpha_0 - \alpha_L) f'_L]}.$$

On the other hand, if

$$f(x_k + \alpha_0 d_k) < f(x_k) + (1-\rho)\langle \nabla f(x_k), d_k \rangle \alpha_0,$$

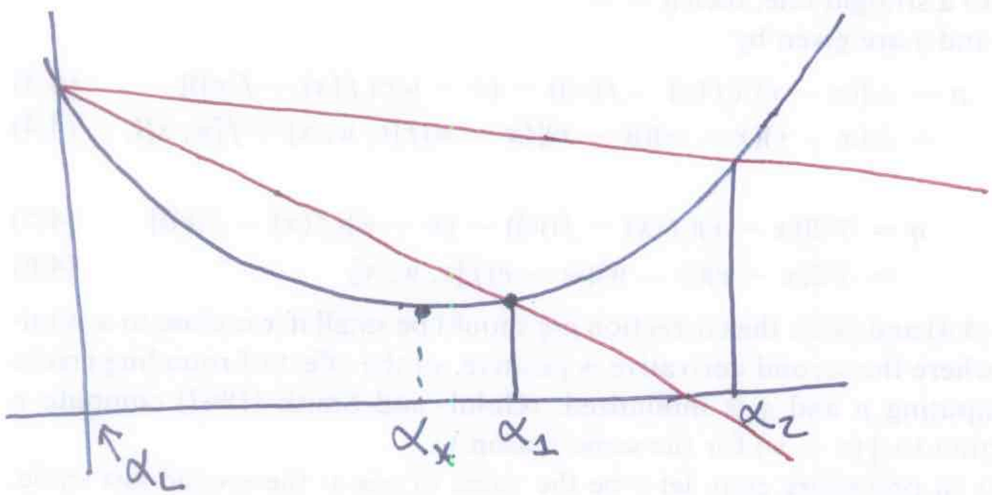
we have



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

$$\breve{\alpha} = \alpha_0 + \frac{(\alpha_0 - \alpha_L)}{(f'_L - f'_0)} f'_0$$

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 α_1 and α_2 bracket the min



In this case, we will never arrive at the true min. So instead of requiring that $f(a_*) >$ the line through α_1, let's require that the derivative of f improve at a_* by a certain fraction.

So we let $f(\alpha) = f(x_k + \alpha d_k)$

and write

Fletcher's Modified Goldstein Conditions.

Choose $\rho \in [0, 1/2)$, $\sigma \in [\rho, 1)$. Then α satisfies the conditions if

$$f'(\alpha) \geq \sigma f'(0)$$

and

$$f(\alpha) \leq f(0) + \rho f'(0) \alpha .$$

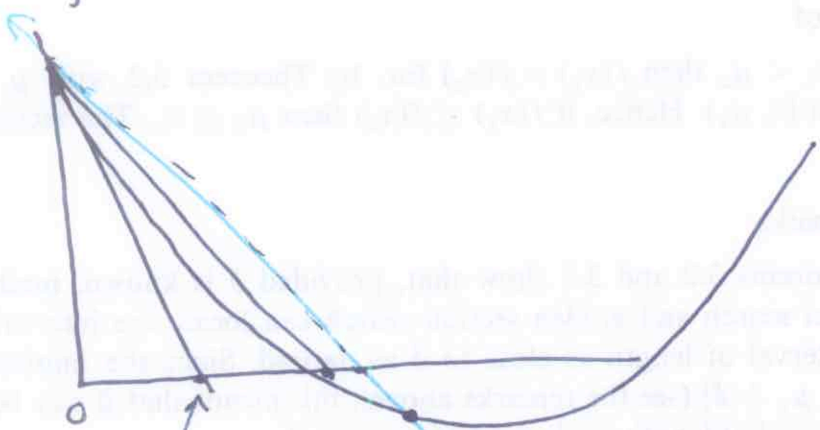
Note to readers: we've switched from multivariable to single-variable notation here (and we should rewrite the whole lecture in this form, but don't have time).

Why does this work? Well,

$\sigma f'(0)$ is still negative, but less negative than $f'(0)$,

so $\sigma f'(0) = f'(\alpha_1)$ for some $\alpha_1 < \alpha_*$.

(Now Antoniou and Lu claim that $\alpha_* < \alpha_2$, but I don't see it.



a line with slope $f'(0)$

a line with slope $> \frac{1}{2} f'(0)$)

The final condition becomes

$$|f'(\alpha)| \leq -\sigma f'(0).$$

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 $[\alpha_L, \alpha_U] = [0, 10^{99}]$,

and a guess for α_0,

(*) for each step:

If $f_0 > f_L + \rho(\alpha_0 - \alpha_L) f_L'$

(that is, (α_0, f_0) is above the line with slope $\rho f_L'$)

use quadratic interpolation on $(\alpha_L, f_L), (\alpha_0, f_0), (\alpha_L^*, f_L')$

to guess $\breve{\alpha}$

if $\alpha_0 < \alpha_u$, reset α_U to α_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 α_0, α_L to guess $\breve{\alpha}$

reset α_L to α_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>