Chapter 3: Line Search Methods

Basic algorithm:

Given x_0, compute p_0 and α_0.

Set $x_1 = x_0 + \alpha_0 p_0$.

$\vdots$

For the $(k+1)$th iterate, compute p_k, α_k, and set: $x_{k+1} = x_k + \alpha_k p_k$

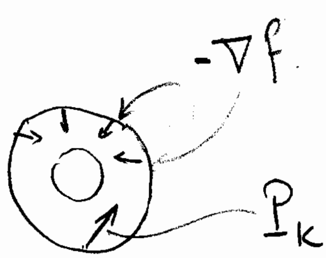
Notation:

α_k is called the step length.

p_k is a descent direction if it satisfies: $p_k^T \nabla f_k < 0$.

Note that if $p_k^T \nabla f_k < 0$, then p_k has a component projection on $-\nabla f_k$, thus moving us in a direction that reduces f.

f:



moving into a descent direction.

A general expression is:

$$p_k = -B_k^{-1} \nabla f_k \quad (*)$$

<u>where:</u>

B_k is a symmetric, nonsingular matrix.

$$B_k = \begin{cases} I, & \text{for } \underline{\text{steepest descent method}}. \\ \nabla^2 f(x_k), & \text{for } \underline{\text{Newton's method}} \\ \approx \nabla^2 f(x_k), & \text{for } \underline{\text{Quasi-Newton methods}} \end{cases}$$

If B_k is <u>positive-definite</u>, then multiply on the left of (*) by ∇f_k^T to get:

$$\nabla f_k^T p_k = -\nabla f_k^T B_k^{-1} \nabla f_k < 0$$

By the definition of what it means to be positive definite

$\Rightarrow$ p_k is a descent direction.

3.1 Step Length

Decreasing $f(\cdot)$ is not enough:

Suppose $f(x) = (x+1)^2$.

Consider $x_k = 1/k$, $k = 1, 2, 3, \ldots$

We have $f(x_k) = (1/k + 1)^2$ decreases to 1,

and $x_k \to 1$.

But, the real minimum is $x_* = -1$, with $f(x_*) = 0$.

Conditions for sufficient decrease.

Require $f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T p_k$ ~ $(*)$

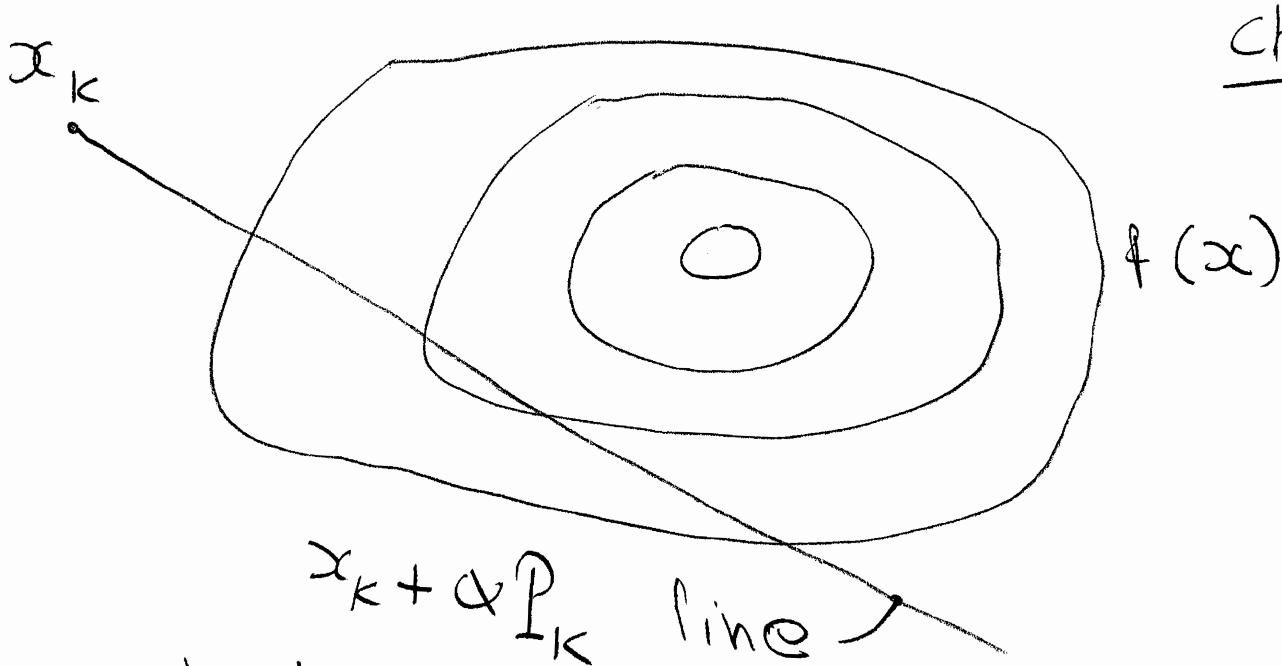
Here: c_1 is constant, $c_1 \in (0,1)$

$\nabla f_k^T p_k$ denotes a directional derivative in the direction of p_k.

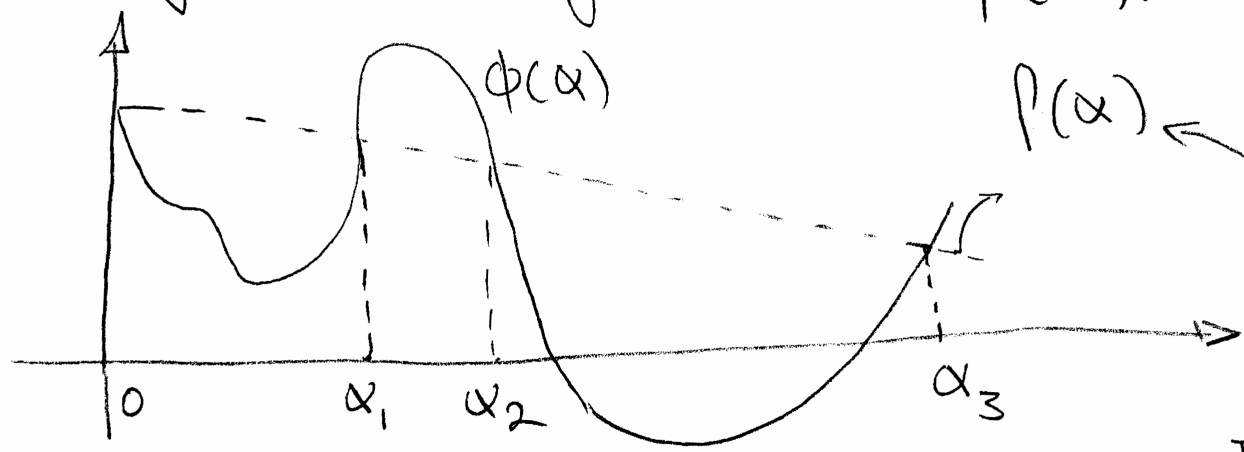
We are "sampling" f from x_k to $x_k + \alpha p_k$ generating a 1-D function of α:

$$\phi(\alpha) = f(x_k + \alpha p_k)$$

Ex 1



Along the line, generate $\phi(\alpha)$:



(*) is often called the Armijo condition. It requires that f stays below the line $l(\alpha)$:

$$l(\alpha) = \underbrace{f(x_k)}_{\text{const.}} + \underbrace{c_1}_{\substack{\text{const} \\ c_1 \simeq 10^{-4}}} \alpha \underbrace{\nabla f_k^T p_k}_{\text{const}}$$

From the graph:

$\alpha \in (0, \alpha_1)$ or $\alpha \in (\alpha_2, \alpha_3)$

<u>will work</u>.

Curvature Condition:

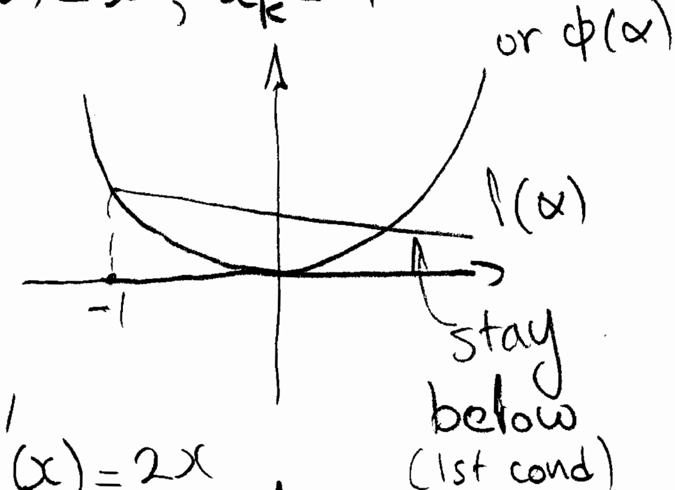
From: $\phi(\alpha) = f(x_k) + \alpha\left(c_1 \nabla f_k^T p_k\right)$

$$\Rightarrow \quad \frac{d\phi(\alpha)}{d\alpha} = c_1 \underbrace{\nabla f_k^T p_k}$$

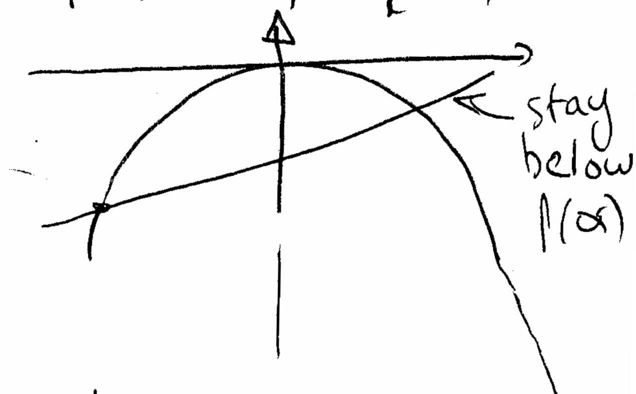
directional derivative along p_k.

Consider two cases:

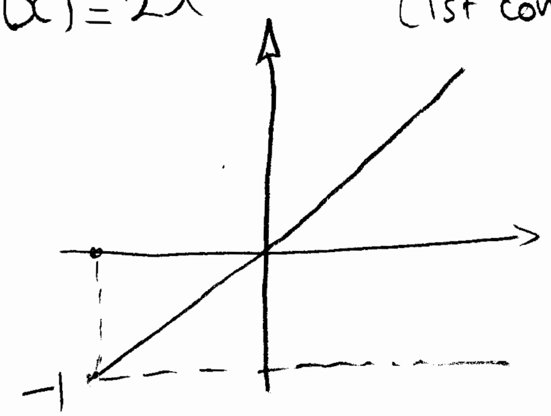
$f(x) = x^2, \; x_k = -1$



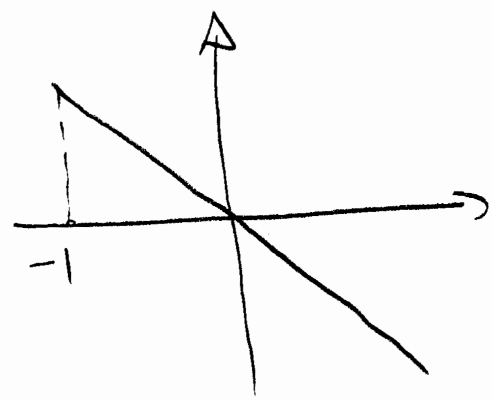
$f(x) = -x^2, \; x_k = -1$



$f'(x) = 2x$



$f'(x) = -2x$



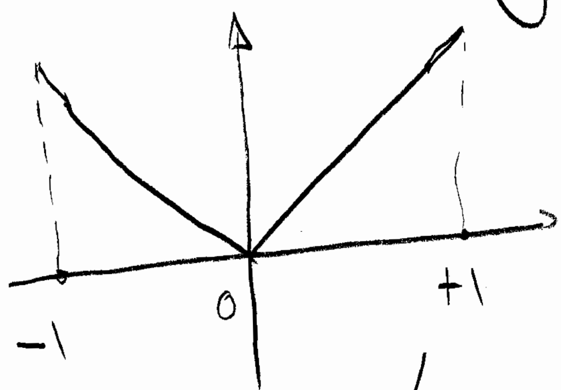
Require slope increase

$\Rightarrow$ Moving to right is ok.

Require slope increase:

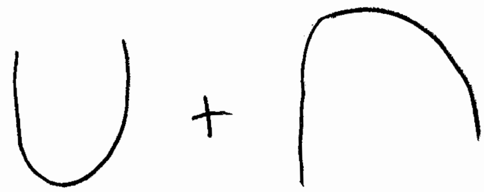
$\Rightarrow$ Moving right is <u>not</u> ok.

Require slope magnitude decrease (strong Wolfe conditions):



← stay within the same stationary point in the strong Wolfe conditions.

magnitude is reduced for both



when moving toward $x=0$.

Wolfe conditions

$$f(x_k+\alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k \quad \text{—} ①$$

$$\nabla f(x_k+\alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k \quad \text{—} ②$$

Want increase from $-ve$ to zero $(+ve)$ slope.

$c_2 \in (c_1, 1)$

Strong Wolfe Conds

Replace ② by:

$$|\nabla f(x_k+\alpha p_k)^T p_k| \leq c_2 |\nabla f_k^T p_k|$$

Require magnitude reduction towards zero.

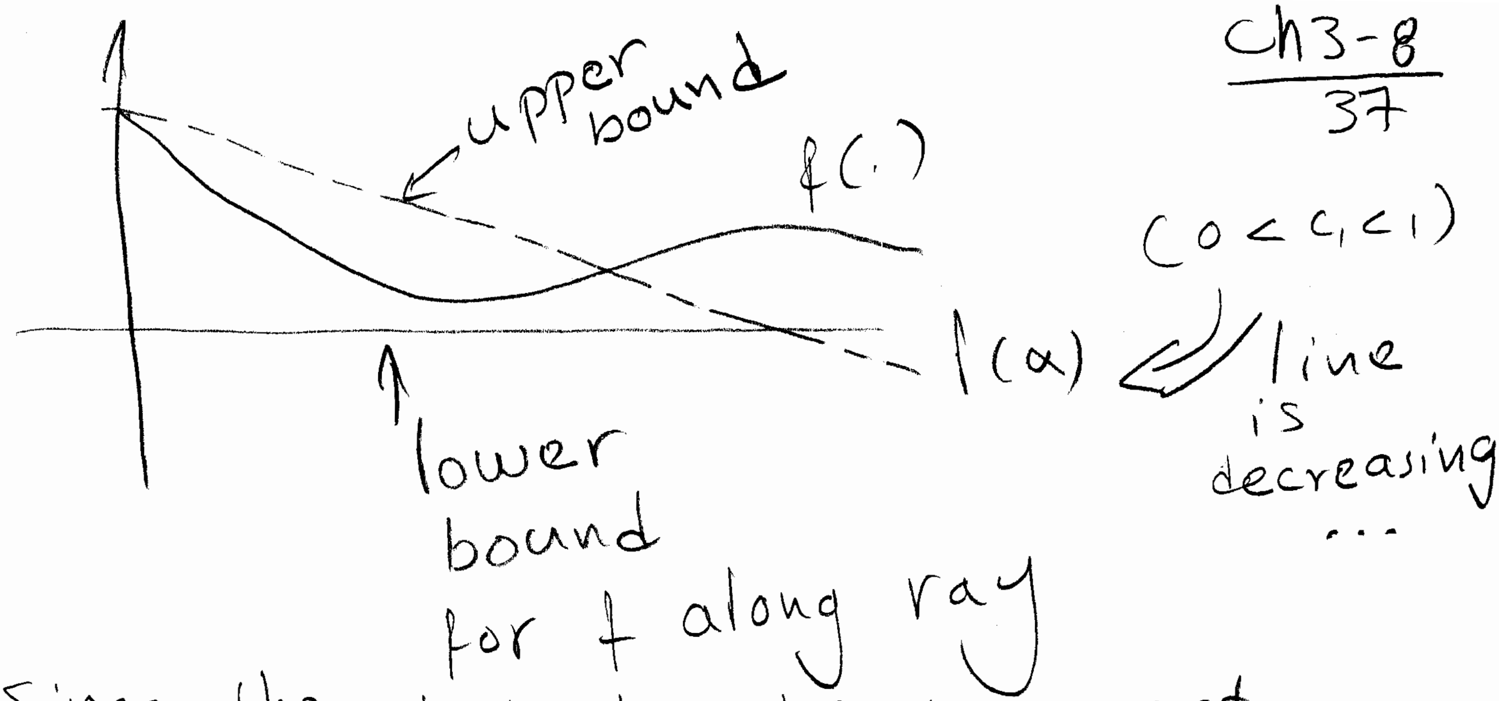
Lemma 3.1

Suppose that $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ is continuously diff'ble and p_k is a descent direction at x_k. (makes less than 90° with $-\nabla f$), and assume that f is bounded below along the ray $\{x_k + \alpha p_k \mid \alpha > 0\}$ (can always be satisfied unless f grows to -infinity along the ray). Then:

If $0 < c_1 < c_2 < 1$, there exist intervals for α satisfying both the Wolfe and strong Wolfe conditions.

Proof:

Since $\phi(\alpha) = f(x_k + \alpha p_k)$ is bounded below, and is also decreasing for small α, we have at-least one point for which $\phi(\alpha) = l(\alpha)$ see next page for figure.



Since the two bounds intersect, there must be at-least one point where $\phi(\alpha)$ meets $l(\alpha)$.

For some α':

$$f(x_k+\alpha' p_k) = f(x_k) + \alpha' c_1 \nabla f_k^T p_k \quad (*)$$

From the figure, it is clear that:

$$f(x_k) + \alpha c_1 \nabla f_k^T p_k > f(x_k+\alpha p_k) \quad (3)$$

for $\alpha \in (0, \alpha')$. $\Rightarrow$ (1) is satisfied here.

From the Mean Value Theorem:

$$f(x+p) = f(x) + \nabla f(x+tp)^T p \ , \overset{\text{some}}{\forall}\ t \in (0,1).$$

set $\begin{cases} p = \alpha' p_k \\ x = x_k \end{cases}$ to get:

$$f(x_k + \alpha' p_k) - f(x_k) = \alpha' \nabla f(x_k + \alpha'' p_k)^T p_k \quad (**)$$

for some $\alpha'' \in (0, \alpha')$.

From (*), (**):

$$\cancel{\alpha'} c_1 \nabla f_k^T p_k = \cancel{\alpha'} \nabla f(x_k + \alpha'' p_k)^T p_k$$

$$\Rightarrow \nabla f(x_k + \alpha'' p_k)^T p_k = c_1 \nabla f_k^T p_k > c_2 \nabla f_k^T p_k \quad (4)$$

since $c_2 > c_1$ (more negative)

$\Rightarrow$ (2) is satisfied strictly here

From (3) + (4) + smoothness

$\Rightarrow$ we can find a neighborhood of α'' where (1) + (2) hold (not just the points).

Backtracking

Choose $\bar{\alpha} > 0$, $\rho, c \in (0,1)$.

$\alpha \leftarrow \bar{\alpha}$

repeat until $f(x_k + \alpha p_k) \leq f(x_k) + c\alpha \nabla f_k^T p_k$

$\alpha \leftarrow \rho\alpha$;

end

Notes:

- $\bar{\alpha} = 1$ for Newton and Quasi-Newton
- Steepest-descent or conjugate gradient use other values...
- guaranteed to converge since α converges to zero, and at a finite number of iterations.
- must have $\rho \in [\rho_{low}, \rho_{hi}]$ for $0 < \rho_{low} < \rho_{hi} < 1$, after some iterations.

good for Newton, not quasi-Newton or conjugate gradient.

no need for curvature (2nd) condition

3.2 Convergence of Line Search Methods

Require $\|\nabla f_k\| \longrightarrow 0$ so that a stationary point is reached as $k \to \infty$.

Define the angles between steepest descent $-\nabla f_k$ and p_k by:

$$\cos\theta_k = \frac{-\nabla f_k^T p_k}{\|\nabla f_k\| \, \|p_k\|}.$$

Thm 3.2

* Let $x_{k+1} = x_k + \alpha_k p_k$, where p_k is a descent direction (has a positive projection on $-\nabla f_k$).
* Suppose α_k satisfies the Wolfe conditions
* f is bounded below, continuously diff'ble in N containing $\mathcal{L} \stackrel{\text{def}}{=} \{x : f(x) \le f(x_0)\}$, where x_0 is the starting point.
* ∇f is Lipschitz conts on N:

  $\exists L > 0$ s.t. $\|\nabla f(x) - \nabla f(\tilde{x})\| \le L\|x - \tilde{x}\|$, for all $x, \tilde{x} \in N$.

Then:

$$\sum_{k \geq 0} \cos^2 \theta_k \, \|\nabla f_k\|^2 < \infty \quad \blacksquare$$

In other words, if $\cos^2 \theta_k \neq 0$, and $\cos^2 \theta_k > \delta > 0$, we must have that $\|\nabla f_k\| \to 0$, (convergence to a stationary point).

Proof: Omitted.

$\Rightarrow$ Steepest descent: $p_k = -\nabla f_k$ is <u>convergent</u> to the local stationary point, wrongly called <u>globally convergent</u> ... (local min or max).

Newton-like method

Suppose that: $x_{k+1} = x_k + \alpha_k p_k$

and: $p_k = -B_k^{-1} \nabla f_k$.

Suppose that B_k are positive definite with uniformly-bounded condition number:

$$\|B_k\| \, \|B_k^{-1}\| \leq M, \quad \text{all } k.$$

Then $\cos \theta_k \geq 1/M$ and

$$\lim_{k \to \infty} \|\nabla f_k\| = 0.$$

For conjugate-gradient methods, can show:

$$\liminf_{k \to \infty} \|\nabla f_k\| = 0.$$

Generating a Globally Convergent Algorithm

Produce:

(i) every iteration reduces the objective function.

(ii) every mth iteration is a steepest descent step that uses step lengths that satisfy the Wolfe or Goldstein conditions.

From the steepest descent steps, we at-least get:

- $$\liminf_{k\to\infty} \|\nabla f_k\| = 0$$

for the sequence $k = m, 2m, 3m, \ldots$

The idea is to take steepest descent steps so as to ensure convergence.

3.3 Convergence of Steepest Descent

Start with quadratic model:

$$f(x) = \tfrac{1}{2} x^T Q x - b^T x$$

where Q is symmetric and positive definite.

$$\Rightarrow \nabla f(x) = Qx - b = 0$$

$$\Rightarrow \quad Qx^* = b.$$

To minimize $f(x_k - \alpha \nabla f_k)$:

Set $f'(\alpha_k) = 0$

$$\Rightarrow \quad \boxed{\alpha_k = \frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k}} \quad (*)$$

Algorithm

$$\boxed{x_{k+1} = x_k - \alpha_k \nabla f_k}$$

where α_k is given by $(*)$.

and $\boxed{\nabla f_k = Qx_k - b.}$

Define: $\|x\|_Q^2 = x^T Q x.$

$$\frac{1}{2}\|x - x^*\|_Q^2 = \frac{1}{2}(x - x^*)^T Q (x - x^*)$$

$$= [x^T Q x - x^T Q x^* - x^{*T} Q x + x^{*T} Q x^*] \cdot \frac{1}{2}$$

Recall: $Qx^* = b \Rightarrow b^T = x^{*T} Q.$

$$\Rightarrow \frac{1}{2}\|x - x^*\|_Q^2 = \frac{1}{2} x^T Q x - \cancel{b^T x} \times \frac{1}{2}$$

$$- \frac{1}{2} x^{*T} Q x^* + \cancel{b^T x^*} \cdot \frac{1}{2}$$

$$- b^T x + b^T x.$$

$$= \underline{\underline{f(x) - f(x^*)}}.$$

<u>Thm 3.3</u> For the guardratic case:

$$\|x_{k+1} - x^*\|_Q^2 \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 \|x_k - x^*\|_Q^2$$

where: $0 < \lambda_1 \leq \cdots \leq \lambda_n$ are the eigenvalues of Q.

Notes:

* Suppose that: $\lambda_1 = \lambda_2 = \cdots = \lambda_n$

$\Rightarrow$ $\boxed{x_{k+1} = x^*, \text{ convergence in one step.}}$

* Recall: the condition number is

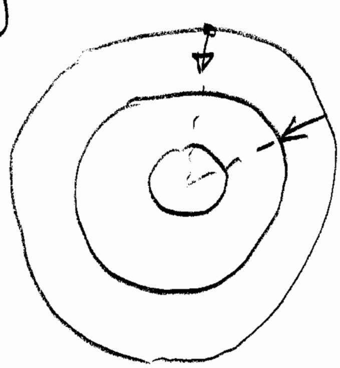
$$\kappa_2(A) = \lambda_n / \lambda_1 .$$

Also:

$$\left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 \times \frac{\lambda_1^2}{\lambda_1^2} = \left(\frac{\kappa - 1}{\kappa + 1}\right)^2$$

For $\kappa(A) = 1000$, reduction is only $\left(\frac{999}{1001}\right)^2$.

$\Rightarrow$ Scale variables to get fast convergence:



"equal eigenvalues"

Quasi-Newton

$$P_k = -B_k^{-1} \nabla f_k.$$

Thm 3.5 Suppose $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is:

* three times continuously diff'ble,
* P_k is a descent direction
* α_k satisfies the Wolfe conditions, with $c_1 \leq 1/2$.
* $x_k \rightarrow x^*$ with $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ pos def.
* $$\lim_{k \to \infty} \frac{\|\nabla f_k + \nabla^2 f_k P_k\|}{\|P_k\|} = 0 \quad (*)$$ ← approx. or...

Then:

* $\alpha_k = 1$ is allowed for $k > k_0$, some k_0.
* if $\alpha_k = 1$, $\forall k > k_0$: $x_k \rightarrow x^*$ superlinearly

* For Quasi-Newton, (*) is:

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla^2 f(x^*))\, p_k\|}{\|p_k\|} = 0$$

So, we require: $B_k \to \nabla^2 f(x^*)$

* Note (*) is both necessary and sufficient for superlinear convergence.

* Recall superlinear convergence:

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0, \quad \text{eg: } 1+k^{-k} \xrightarrow[\text{superlinearly}]{} 1$$

Newton's Method

$$p_k^N = -\nabla^2 f_k^{-1} \nabla f_k$$

Recall:

* Lipschitz continuity:

f is Lipschitz continuous if there is $M > 0$ such that for any two points x_0, x_1 in D:

$$\| f(x_1) - f(x_0) \| \leq M \| x_1 - x_0 \|$$

* Quadratic convergence

$$\frac{\| x_{k+1} - x^* \|}{\| x_k - x^* \|^2} \leq M, \quad k \text{ sufficiently large.}$$

Eg: $1 + (0.5)^{2^k} \longrightarrow 1$ in quadratic convergence.

Thm 3.7

Suppose:

* f is twice diff'ble
* $\nabla^2 f(x)$ is Lipschitz conts in a neighborhood of x^*, where: $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive def.
* $p_k^N = -\nabla^2 f_k^{-1} \nabla f_k$ (Newton's method)

Then:

* if x_0 is sufficiently close to x^*, then $x_k \to x^*$ quadratically and $\|\nabla f_k\| \to 0$ quadratically.

<u>Proof</u>

$$x_{k+1} - x^* = (x_k + p_k^N) - x^*$$

$$= x_k - x^* \underbrace{-(\nabla^2 f_k^{-1}) \nabla f_k}_{\text{def. of } p_k.}$$

$$= \underbrace{(\nabla^2 f_k^{-1})(\nabla^2 f_k)}_{I}(x_k - x^*) - (\nabla^2 f_k^{-1})[\nabla f_k - \underbrace{\nabla f_*}_{=0.}]$$

$$= \nabla^2 f_k^{-1} \underbrace{\left[\nabla^2 f_k (x_k - x^*) - (\nabla f_k - \nabla f_*)\right]}_{\text{We want to bound this second term.}}$$

$(*)$

Recall Taylor's thm:

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)\,p\,dt$$

Set: $p = x^* - x_k$, $x = x_k$

$$\Rightarrow x_k + p = x^*$$

$$\Rightarrow \underbrace{\nabla f(x^*)}_{\nabla f_*} = \underbrace{\nabla f(x_k)}_{\nabla f_k} + \int_0^1 \nabla^2 f(x + t(x^* - x_k)) \times (x^* - x_k)\,dt$$

$$\Rightarrow \nabla f_k - \nabla f_* = \int_0^1 \nabla^2 f(x_k + t(x^* - x_k)) \times (x^* - x_k)\,dt$$

From (*), note that:

$$\nabla^2 f_k (x_k - x^*) = \int_0^1 \underbrace{\nabla^2 f_k (x_k - x^*)}_{\substack{\text{constant,} \\ \text{independent of } t.}} dt$$

$$\Rightarrow \| \nabla^2 f_k (x_k - x^*) - (\nabla f_k - \nabla f(x^*)) \| \sim \text{LHS}$$

$$= \left\| \int_0^1 \left[\nabla^2 f(x_k) - \nabla^2 f(x + t(x^* - x_k)) \right] (x_k - x^*)\,dt \right\|$$

Using the Dominated Convergence thm:

$$\left\| \int f(x)\,dx \right\| \leq \int \| f(x) \|\,dx.$$

to get:

$$LHS \leq \int_0^1 \| \nabla^2 f(x_k) - \nabla^2 f(x_k + t(x^* - x_k)) \| \, \| x_k - x^* \| \, dt$$

From Lipschitz's continuity of $\nabla^2 f(x_k)$:

$$\| \nabla^2 f(x_k) - \nabla^2 f(x_k + t(x^* - x_k)) \|$$

$$\leq L \| \cancel{x_k} - \cancel{x_k} - t(x^* - x_k) \| \text{, some } L.$$

$$= L \| x_k - x^* \| t \quad \text{since } t \in (0,1) \Rightarrow t > 0$$

Thus:

$$LHS \leq L \| x_k - x^* \|^2 \int_0^1 t\,dt = \tfrac{1}{2} L \| x_k - x^* \|^2$$

(**)

For x sufficiently close to x^*, note that:

$$\nabla^2 f_k \rightarrow \nabla^2 f(x^*)$$

Since $\nabla^2 f(x^*)$ is positive definite, it is also invertible:

$$(\nabla^2 f_k)^{-1} \rightarrow (\nabla^2 f(x^*))^{-1}$$

In the neighborhood, we can assume:

$$\|\nabla^2 f_k^{-1}\| \leq 2 \|\nabla^2 f(x^*)^{-1}\| \quad *{*}*$$

Recall $*$, and substitute $**$ + $*{*}*$ to get:

$$\|x_{k+1} - x^*\| \leq \underbrace{\|\nabla^2 f(x^*)^{-1}\| \cdot L}_{\text{constant}} \cdot \|x_k - x^*\|^2$$

$$\leq \tilde{L} \|x_k - x^*\|^2$$

$\Rightarrow$ Convergence to x^* is quadratic.

For $\|\nabla f(x_k)\|$, recall:

$$\begin{cases} x_{k+1} - x_k = p_k^N \\ \nabla^2 f(x_k)\, p_k^N = -\nabla f_k \end{cases}$$ for Newton's method.

We have:

$$\|\nabla f(x_{k+1}) - \vec{0}\| = \|\underbrace{\nabla f(x_{k+1}) - \nabla f_k}\, \overbrace{- \nabla^2 f(x_k) p_k^N}^{=0}\|$$

$$= \|\underbrace{\int_0^1 \nabla^2 f(x_k + t p_k^N)(x_{k+1} - x_k)\, dt}_{\text{as before, from Taylor's thm.}} - \nabla^2 f(x_k)\, p_k^N\|$$

$p_k^N = x_{k+1} - x_k$

$$\leq \int_0^1 \underbrace{\|\nabla^2 f(x_k + t p_k^N) - \nabla^2 f(x_k)\|}_{\leq L\|p_k^N\| t \ \text{ as before.}} \|p_k^N\|\, dt$$

$$\leq \tfrac{1}{2} L \|p_k^N\|^2$$

Continuing:

$$\|\nabla f(x_{k+1})\| \leq \tfrac{1}{2} L \|p_k^N\|^2$$

$$= \tfrac{1}{2} L \underbrace{\|\nabla^2 f(x_k)^{-1}\|^2 \|\nabla f_k\|^2}_{\text{by def of } p_k^N}$$

$$\leq L \underbrace{\|\nabla^2 f(x^*)^{-1}\|^2}_{\text{bounding } \|\nabla^2 f(x_k)^{-1}\|^2 \text{ as before.}} \|\nabla f_k\|^2$$

$$\leq \underset{\substack{\uparrow \\ \text{constant.}}}{C} \|\nabla f_k\|^2$$

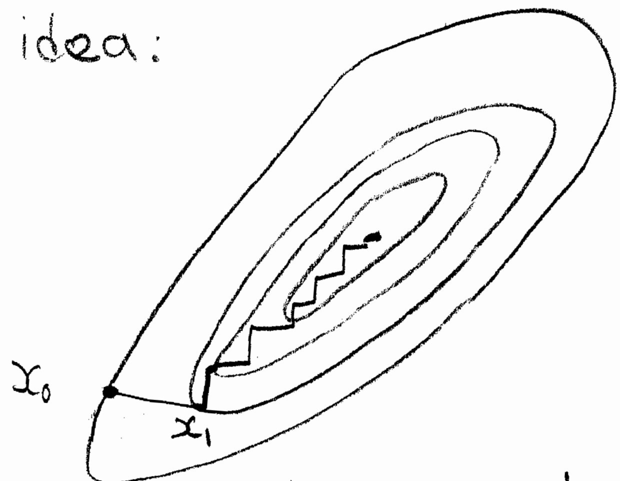
$\Rightarrow$ gradient norm converge to zero quadratically.

Coordinate Descent

* Maybe slow or not converge at-all.
* will converge at an acceptable rate if the variables are "loosely coupled."
* ∇f_k is not needed!
* Globally convergent variants exist

Book suggests Fletcher and Ref. [104].

Basic idea:



Optimize f along each coordinate:

$$p_k = e_1, e_2, \ldots, e_n, e_1, e_2, \ldots, e_n, \ldots$$

where $e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ ith position in the vector

Two variations:

* Minimize along $P_k = \underbrace{e_1, e_2, \dots, e_n, e_{n-1}, \dots, e_2, e_1}_{\text{repeat this pattern}}, \dots$

* Join the first and last point in the cycle, and seach in that direction:

$$x_1 = x_0 + \alpha_0 e_1$$
$$x_2 = x_1 + \alpha_1 e_2$$
$$\vdots$$
$$x_n = x_{n-1} + \alpha_{n-1} e_n$$

and

$$x_{n+1} = x_n + \alpha_n \underbrace{(x_n - x_1)}_{\text{search in this direction.}}$$

repeat

"Loosely coupled":

$$f(x,y) = x^2 + y^2$$

For y fixed $\Rightarrow$ $x = 0.$ $\Rightarrow$ $x_1 = \begin{bmatrix} 0 \\ y_0 \end{bmatrix}$

For x fixed $\Rightarrow$ $y = 0$ $\Rightarrow$ $x_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ ✓

Interpolation

Require: $\phi(\alpha_k) \leq \phi(0) + c_1 \alpha_k \phi'(0)$ — $(*)$

for $c_1 \stackrel{\circ}{=} 10^{-4}$.

Let α_0 be an initial guess.

Check that $\phi(\alpha_0) \leq \phi(0) + c_1 \alpha_0 \phi'(0)$

If ok, we are done, else:

Quadratic interpolation

Given: $\phi(0), \phi'(0), \phi(\alpha_0)$:

$$\phi_q(\alpha) = \left(\frac{\phi(\alpha_0) - \phi(0) - \alpha_0 \phi'(0)}{\alpha_0^2} \right) \alpha^2 + \phi'(0)\alpha + \phi(0)$$

is a quadratic fit, so that:

$$\begin{cases} \phi_q(0) = \phi(0), \quad \phi_q'(0) = \phi'(0), \\ \phi_q(\alpha_0) = \phi(\alpha_0) \end{cases}$$

Minimum point for:

$$\boxed{\alpha_1 = \frac{-\phi'(0)\,\alpha_0^2}{2\left[\phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0\right]}}$$

If (*) is satisfied with α_1, then terminate. Else apply a cubic fit:

$$\phi_c(\alpha) = a\alpha^3 + b\alpha^2 + \alpha\phi'(0) + \phi(0)$$

<u>where:</u>

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\alpha_0^2\alpha_1^2(\alpha_1 - \alpha_0)} \begin{bmatrix} \alpha_0^2 & -\alpha_1^2 \\ -\alpha_0^3 & \alpha_1^3 \end{bmatrix} \begin{bmatrix} \phi(\alpha_1) - \phi(0) - \phi'(0)\alpha_1 \\ \phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0 \end{bmatrix}$$

The minimum is for:

$$\boxed{\alpha_2 = \frac{-b + \sqrt{b^2 - 3a\phi'(0)}}{3a}}$$

* If needed, repeat cubic interpolation with $\phi(0)$, $\phi'(0)$ and two most recent ϕ-values.

* <u>if</u> ($|\alpha_i - \alpha_{i-1}|$ small) or ($\alpha_{i-1} >> \alpha_i$)
  <u>then</u> $\alpha_i \leftarrow \alpha_{i-1}/2$ (safeguard)

Initial Step Length

quasi-Newton: start with $\underline{\underline{\alpha_0 = 1.}}$

Steepest descent, etc

Assume $\alpha_0 \nabla f_k^T p_k = \alpha_{k-1} \nabla f_{k-1}^T p_{k-1}$

$$\Rightarrow \boxed{\alpha_0 = \alpha_{k-1} \frac{\nabla f_{k-1}^T p_{k-1}}{\nabla f_k^T p_k}}$$

or

Newton?

Interpolate $f(x_{k-1}), f(x_k), \phi'(0) = \nabla f_k^T p_k$

using: $$\boxed{\alpha_0' = \frac{2(f_k - f_{k-1})}{\phi'(0)}}$$

or try

$$\boxed{\alpha_0 = \min(1, 1.01\,\alpha_0)}$$

Line Search that satisfies Wolfe Conds

$\alpha_0 \leftarrow 0$

$i \leftarrow 1$

Initialize $\alpha_1 > 0$ and α_{max}

<u>repeat</u>

Evaluate $\phi(\alpha_i)$;

<u>if</u> $(\phi(\alpha_i) > \phi(0) + c_1 \alpha_i \phi'(0))$ <u>or</u>

$(\phi(\alpha_i) \geq \phi(\alpha_{i-1})$ and $i > 1)$

$\alpha_* \leftarrow$ <u>Zoom</u> (α_{i-1}, α_i)

<u>Stop</u>;

(Look for optimal in (α_{i-1}, α_i)
since $\phi(\alpha_i)$ violates decrease)

Evaluate $\phi'(\alpha_i)$;

<u>if</u> $|\phi'(\alpha_i)| \leq -c_2 \phi'(0)$ ← decrease condition <u>already</u> satisfie check curvature condition only.

$\alpha_* \leftarrow \alpha_i$;

<u>Stop</u>;

<u>if</u> $\phi'(\alpha_i) \geq 0$

$\alpha_* \leftarrow$ <u>zoom</u> $\underbrace{(\alpha_i, \alpha_{i-1})}$

stop;

reverse search direction.

Choose $\alpha_{i+1} \in (\alpha_i, \alpha_{max})$

$i \leftarrow i+1$

<u>extrapolate</u>

<u>end</u> (repeat)

* α_i is increasing.

* (α_{i-1}, α_i) contains step lengths satisfying strong Wolfe conds <u>if</u>:

(i) α_i violates the decrease cond.

(ii) $\phi(\alpha_i) \geq \phi(\alpha_{i-1})$

(iii) $\phi'(\alpha_i) \geq 0$

* must converge in a finite number of iterations.

(if not, print an error message)

$\alpha_* = \underline{\text{Zoom}}(\alpha_{low}, \alpha_{hi})$

* α_{low} & a_{hi} are <u>input</u> arguments.

* $(\alpha_{low}, \alpha_{hi})$ or $(\alpha_{hi}, \alpha_{low})$ contains α which satisfies strong Wolfe conds.

$\left\{ \begin{array}{l} \text{Remember that } (0, \alpha_{max}) \text{ covers all} \ldots \\ \& \ (0, \alpha') \text{ some } \alpha' \text{ contains right } \alpha. \end{array} \right.$

* α_{lo} means $f(\alpha_{low}) < f(\underbrace{\alpha_i}_{\text{all other } \alpha_i \text{ exami}})$

* α_{hi} is such that: $\phi'(\alpha_{lo})(\alpha_{hi} - \alpha_{lo}) < 0.$

$\alpha_* = \text{zoom}(\alpha_{lo}, \alpha_{hi})$

<u>repeat</u>

$\alpha_j \leftarrow$ interpolate between α_{lo}, α_{hi}

Evaluate $\phi(\alpha_j)$;

<u>if</u> $\left.\begin{array}{l}\phi(\alpha_j) > \phi(0) + c_1 \alpha_j \phi'(0) \\ \text{or } \phi(\alpha_j) \geq \phi(\alpha_{lo})\end{array}\right\}$ violacter condr

$\alpha_{hi} \leftarrow \alpha_j$;

<u>else</u> (decreasing ...)

Evaluate $\phi'(\alpha_j)$;

<u>if</u> $|\phi'(\alpha_j)| \leq -c_2 \phi'(0)$ } sufficient curvature decrease

$\alpha_* \leftarrow \alpha_j$;

<u>stop</u>,

<u>if</u> $\phi'(\alpha_j)(\alpha_{hi} - \alpha_{lo}) \geq 0$ } Cond. violated reverse.

$\alpha_{hi} \leftarrow \alpha_{lo}$;

$\alpha_{lo} \leftarrow \alpha_j$; ← We are decreasing!

<u>end (repeat)</u>