Método del gradiente:

Queremos resolver el problema

$$Ax = b$$

en forma iterativa para A simétrica y definida positiva.

En 1 variable $ax = b$ $a > 0$

$$\Leftrightarrow \quad ax - b = 0$$

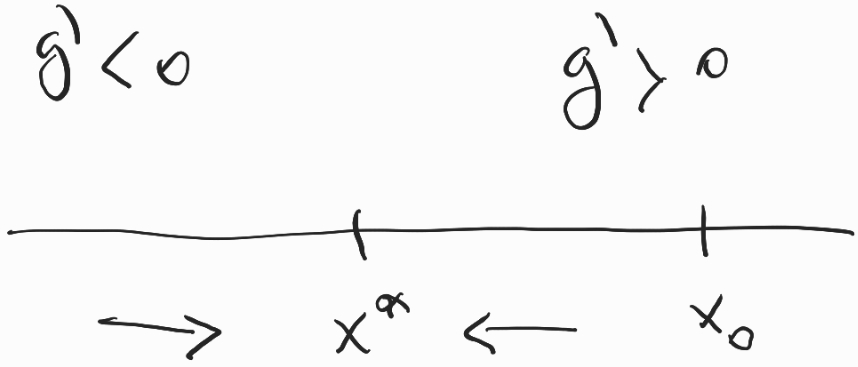
que es la derivada de

$$g(x) = a\frac{x^2}{2} - bx$$

luego, si x^* es mínimo ($a > 0 \Rightarrow$ no tiene máx pero sí min) debe ser

$$g'(x^*) = 0 \quad \Leftrightarrow \quad ax^* - b = 0$$

Dada $g(x)$, para buscar un mínimo en forma iterativa hacemos lo siguiente. Sabemos que



$\Rightarrow$ Queremos definir

$$x_{k+1} = x_k + t_k v_k$$

de manera de ir acercándonos a x^* (en el sentido de las flechas). Una opción es elegir

$v_k = -\text{signo}(g'(x_k))$ y un t_k apropiado.

Más generalmente, dada $A \in \mathbb{R}^{n \times n}$ simétrica y definida positiva

$$A = P D P^t \quad \text{con } d_{ii} > 0 \ \forall i$$

Para resolver $Ax = b$ lo vamos a pensar como encontrar el mínimo de

$$g(x) = \frac{1}{2} x^t A x - b^t x$$

Vemos que son son equivalentes.

Teorema: x^* es solución de $Ax = b \iff x^*$ es el valor mínimo de $g(x)$.

Dem:

Dado $x \in \mathbb{R}^n$, $v \neq 0$, $v \in \mathbb{R}^n$, consideramos

$$h(s) = g(x + s v) \qquad s \in \mathbb{R}$$

$$= \frac{1}{2} (x + s v)^t A (x + s v) - b^t (x + s v)$$

$$= \frac{1}{2} \left[x^t A x + s x^t A v + s v^t A x + s^2 v^t A v \right] - b^t x - s b^t v$$

$$= \frac{1}{2} s^2 v^t A v + s \left[\frac{1}{2} x^t A v + \frac{1}{2} v^t A x - b^t v \right] + \frac{1}{2} x^t A x - b^t x$$

que es una cuadrática en la variable $s \in \mathbb{R}$. Como A es def $+$

$v^t A v > 0 \Rightarrow h$ tiene mínimo

en s^*: $h'(s^*) = 0$

$$h'(s) = s\, v^t A v + \frac{1}{2}\Big[\underbrace{x^t A v}_{\substack{\| \\ (x^tAv)^t = v^t A^t x \\ = v^t A x \\ \downarrow \\ A \text{ simétrica}}} + v^t A x\Big] - \overbrace{b^t v}^{\substack{v^t \cdot b \\ \|}}$$

$$\Rightarrow h'(s) = s\, v^t A v + v^t (Ax - b)$$

Luego, $s^* = - \dfrac{v^t \cdot (Ax-b)}{v^t A v}$

Supongamos que $Ax^* = b$, entonces si construimos la cuadrática con este x^* y $v \neq 0$ cualquiera, tenemos que el mínimo se alcanza en

$$s^* = -\frac{v^t \overbrace{(Ax^*-b)}^{=0}}{v^t A v} = 0$$

$\Rightarrow$ $h(s) \geq h(0)$ $\forall s$ (y para cualquier $v \neq 0$)

$\Rightarrow$ $g(x^* + sv) \geq g(x^*)$ $\quad \forall s$ y $\forall v$

$\Rightarrow$ Dado $y \in \mathbb{R}^n$, $y \neq x^*$, sea $v = y - x^*$ y $s = 1$

$$\Rightarrow \quad g(y) = g(x^* + \underbrace{y - x^*}_{s \cdot v}) \geq g(x^*)$$

Recíprocamente, si x^* : $g(y) \geq g(x^*)$ $\forall y \in \mathbb{R}^n$, entonces

$$\underbrace{g(x^* + sv)}_{h(s)} \geq \underbrace{g(x^*)}_{h(0)} \quad \begin{array}{l} \forall s \in \mathbb{R} \\ \forall v \in \mathbb{R}^n \\ v \neq 0 \end{array}$$

$\Rightarrow$ $s^* = 0$ es valor mínimo de h

$$\Rightarrow 0 = h'(0) = v^t (Ax^* - b) \qquad \forall v \neq 0$$

$$\Rightarrow Ax^* - b = 0 \Rightarrow Ax^* = b.$$

<u>OBS:</u> $g(x) = \frac{1}{2} x^t A x - b^t x$

$$\frac{\partial g}{\partial x_i}(x) = \lim_{s \to 0} \frac{g(x + s e_i) - g(x)}{s}$$

$$= \lim_{s \to 0} \frac{1}{s} \Big[\frac{1}{2} s^2 e_i^t A e_i + s \Big(\overbrace{\frac{1}{2} x^t A e_i + \frac{1}{2} e_i^t A x}^{x^t C_i(A)} - \underbrace{b^t e_i}_{b_i} \Big) + \cancel{\frac{1}{2} x^t A x} - \cancel{b^t x} - \Big(\cancel{\frac{1}{2} x^t A x} - \cancel{b^t x} \Big) \Big]$$

$$= \lim_{s \to 0} \frac{1}{2} s\, a_{ii} + x^t C_i(A) - b_i$$

$$= \underbrace{x^t C_i(A)}_{F_i(A) \cdot x} - b_i = (A \cdot x - b)_i$$

Se puede ver directo que $g(x)$

tendrá mín y que un x^* es valor mín de g $\Leftrightarrow \nabla g(x^*) = Ax^* - b = 0$

$$\Leftrightarrow Ax^* = b.$$

Entonces para buscar x^* sol de $Ax = b$ con A sim y def + buscaremos mín de

$$g(x) = \frac{1}{2} x^t A x - b^t \cdot x$$

Queremos definir una iteración

$$x_{k+1} = x_k + t_{k+1} v_{k+1}$$

para algún $t_{k+1} \in \mathbb{R}$ y $v_{k+1} \in \mathbb{R}^n$ dirección.

de manera que $g(x_{k+1}) \leq g(x_k)$.

Es decir, queremos $t \in \mathbb{R}, v \in \mathbb{R}^n : v \neq 0$

$$g(x_k + t \cdot v) \leq g(x_k)$$

y esos serán el t_{k+1} y v_{k+1} buscados.

Una vez determinada la dirección v por donde "nos moveremos", vimos que el t^* que hace mínimo el valor $g(x_k + t \cdot v)$ viene dado por

$$t^* = - \frac{v^t \cdot (Ax - b)}{v^t A v}$$

Es decir, para cada v_{k+1} nos conviene usar la iteración

$$x_{k+1} = x_k - \underbrace{\frac{v_{k+1}^t (A x_k - b)}{v_{k+1}^t A v_{k+1}}}_{t_{k+1} = -\frac{\langle v_{k+1}, Ax_k - b \rangle}{\langle v_{k+1}, A v_{k+1} \rangle}} \cdot v_{k+1}$$

Lo que resta determinar y que de origen a los distintos métodos es cómo elegir la dirección v_k en cada paso.

Sabemos que en cada x, la dirección de más rápido decrecimiento es $-\nabla g(x)$.

Como estamos buscando un mínimo resulta razonable entonces usar en cada paso esa dirección de búsqueda.

A este método se lo conoce como <u>método de descenso más rápido</u> o método del gradiente.

El algoritmo queda dado por

$$x_{k+1} = x_k + t_{k+1} \cdot v_{k+1}$$

con $v_{k+1} = -\nabla g(x_k)$ y $t_{k+1} = -\dfrac{\langle v_{k+1}, \overbrace{Ax_k - b}^{-r_k} \rangle}{\langle v_{k+1}, A v_{k+1} \rangle}$

Como $\nabla g(x) = Ax - b \Rightarrow$

si llamamos $r_k = \underbrace{b - Ax_k}_{\text{residuo}} = -\nabla g(x_k)$

con lo cual

$$v_{k+1} = r_k = b - Ax_k \quad \text{y} \quad t_{k+1} = \frac{\langle r_k, r_k \rangle}{\langle r_k, A r_k \rangle},$$

$$x_{k+1} = x_k + \frac{r_k^t \, r_k}{r_k^t A r_k} \cdot r_k$$

Supongamos para empezar que

$A = \lambda \, Id_{2\times 2}$ con $\lambda > 0 \Rightarrow$ la solución

x^* de $\lambda x^* = b$ es $x^* = \dfrac{1}{\lambda} b = \dfrac{1}{\lambda} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$

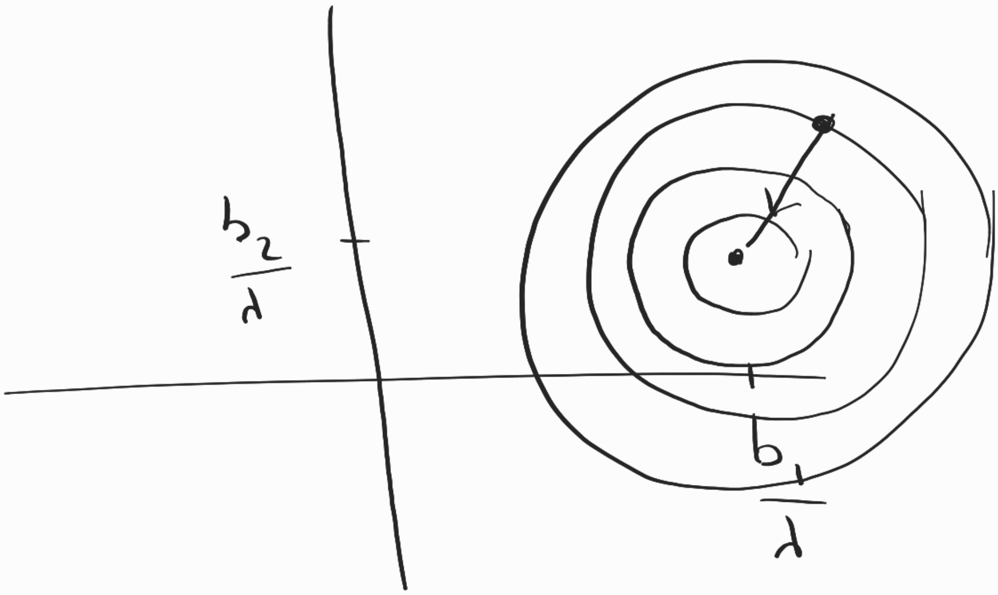
Las curvas de nivel de g son

$$g(x) = \frac{1}{2} x^t \lambda x - b^t x = cte$$

$$\Longleftrightarrow \quad \frac{\lambda}{2}\left(x_1^2+x_2^2\right) - b_1 x_1 - b_2 x_2 = cte$$

$$\Longleftrightarrow \quad \frac{\lambda}{2}\left[\left(x_1 - \frac{b_1}{\lambda}\right)^2 + \left(x_2 - \frac{b_2}{\lambda}\right)^2\right] = \tilde{cte}$$

son circunf con centro $\frac{1}{\lambda}(b_1, b_2)$



En cualquier x_0 que comience, sabemos que x_0 está en la curva de nivel $g(x_0)$. Y que la dirección de más rápido crecimiento viene dada por

$$v = -\nabla g(x_0) = b - \lambda x_0 .$$

Además, como estamos en una circunferencia sabemos que $-\nabla g(x_0) = b - \lambda x_0$ tiene la dirección del rayo que une: x_0 con el centro $\frac{b}{\lambda}$.

Veamos que efectivamente si

$$v_0 = -\nabla g(x_0) = -\lambda x_0 + b$$

y $t_0 = -\dfrac{v_0^t (Ax_0 - b)}{v_0^t A v_0}$ como $A = \lambda Id$

$$\rightarrow \quad t_0 = \frac{(\lambda x_0^t - b^t)(\lambda x_0 - b)}{(\lambda x_0^t - b^t)\lambda(\lambda x_0 - b)} = \frac{1}{\lambda}$$

$\Rightarrow x_1 = x_0 + \frac{1}{\lambda}(-\lambda x_0 + b) = \frac{b}{\lambda}$ es la solución buscada.

¿Esto funcionará para una matriz A cualquiera?

Tomemos ahora

$$A = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$ simétrica y def + (en realidad diagonal!)

y queremos buscar con el método de descenso del gradiente descripto anteriormente la solución de $Ax = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ (que todos sabemos debe ser $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$:))

Definimos

$$f(x) = \frac{1}{2} x^t A x - (0,0) x = \frac{1}{2} x^t A x$$

y buscamos un mínimo (de nuevo todos sabemos que es el $(0,0)$ ☺)

usando la iteración

$$x_{k+1} = x_k - \frac{v_{k+1}^t (A x_k - b)}{v_{k+1}^t A v_{k+1}} \cdot v_{k+1}$$

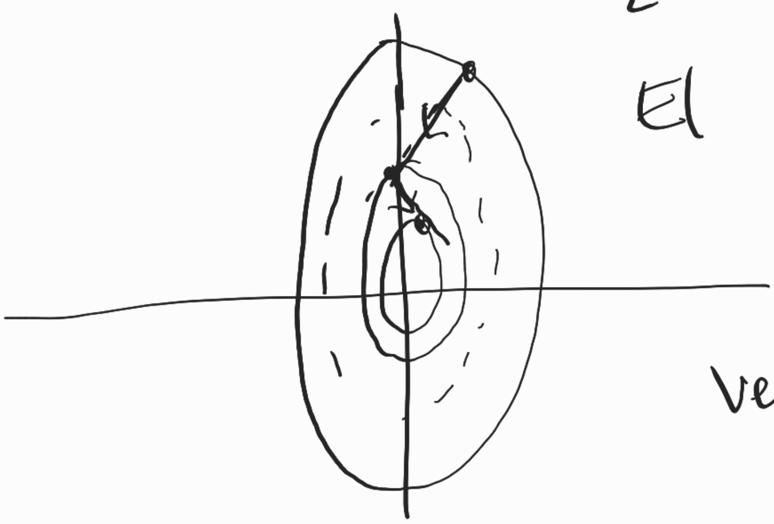
donde $v_{k+1} = -\nabla g(x_k) = -A x_k$

Vemos que $g(x) = \frac{1}{2} (x_1 \; x_2) \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$$= \frac{1}{2}\left(x_1^2 + \frac{1}{2} x_2^2\right)$$

En este caso las curvas de nivel son elipses

$$x_1^2 + \frac{1}{2} x_2^2 = cte$$



El algoritmo hará algo así (zig-zag)

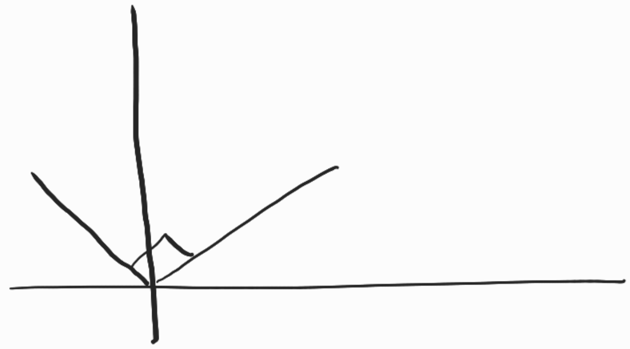
Ver Strang pág 348-349

Podemos mejorar este algoritmo de manera que llegue al cero en menos pasos?

El problema es que, en las elipses, el gradiente no es un rayo que une el punto con el centro de la elipse.

¿Cómo encontrar esas direcciones?

$\mathbb{R}^2$ $\quad \langle x, y \rangle = x^t y = x_1 y_1 + x_2 y_2$



$\langle x, y \rangle = 0 \iff$

$\cos\theta = 0$

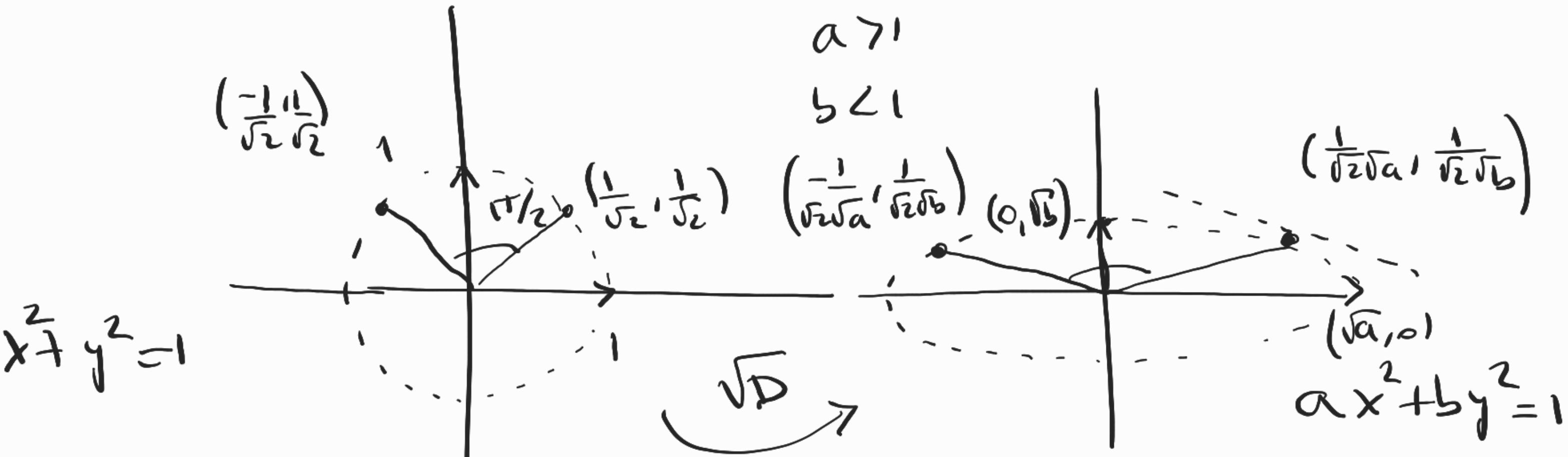
$\iff \theta = \pi/2$

$$\langle x, y \rangle_D = a x_1 y_1 + b x_2 y_2 \qquad a, b > 0$$

$$\langle x, y \rangle_D = x^t \underbrace{\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}}_{D} y = \sqrt{a} x_1 \sqrt{a} y_1 + \sqrt{b} x_2 \sqrt{b} y_2$$

$a > 1$

$b < 1$



$B = \{(\sqrt{a}, 0), (0, \sqrt{b})\}$

$$(x_1, y_1) = x_1 (1, 0) + y_1 (0, 1)$$

$$= \frac{x_1}{\sqrt{a}} (\sqrt{a}, 0) + \frac{y_1}{\sqrt{b}} (0, \sqrt{b})$$

$$(x_1, y_1)_B = \left(\frac{x_1}{\sqrt{a}}, \frac{y_1}{\sqrt{b}}\right)$$

$$(x_1, y_1) \perp_D (x_2, y_2) \iff (x_1, y_1)_B \perp (x_2, y_2)_B$$

En general $A = P D P^t \quad D = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$

$a, b > 0$

$$\langle x, y \rangle_A = x^t P D P^t y = x^t \underbrace{P\sqrt{D}} \underbrace{(P\sqrt{D})^t} y$$

Recordemos que para $ax^2 + by^2 = 1$ una parametrización viene dada por $\sigma(t) = \left(\frac{1}{\sqrt{a}} \cos t, \frac{1}{\sqrt{b}} \operatorname{sen} t\right)$

$\Rightarrow \sigma'(t) = \left(-\frac{1}{\sqrt{a}} \operatorname{sen} t, \frac{1}{\sqrt{b}} \cos t\right)$

con lo cual

$\Rightarrow \langle \sigma(t), \sigma'(t) \rangle$

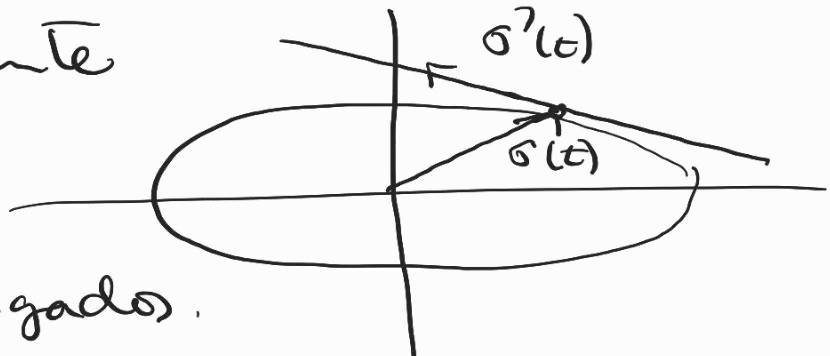
$= -\frac{1}{a} \cos t \operatorname{sen} t + \frac{1}{b} \cos t \operatorname{sen} t \neq 0$ si $a \neq b$

Sin embargo

$\langle \sigma(t), \sigma'(t) \rangle_A = a \cdot \left(-\frac{1}{a} \cos t \operatorname{sen} t\right)$

$+ b \left(\frac{1}{b} \cos t \operatorname{sen} t\right) = 0$!

Geométricamente



$\sigma(t)$ y $\sigma'(t)$

son A conjugados.

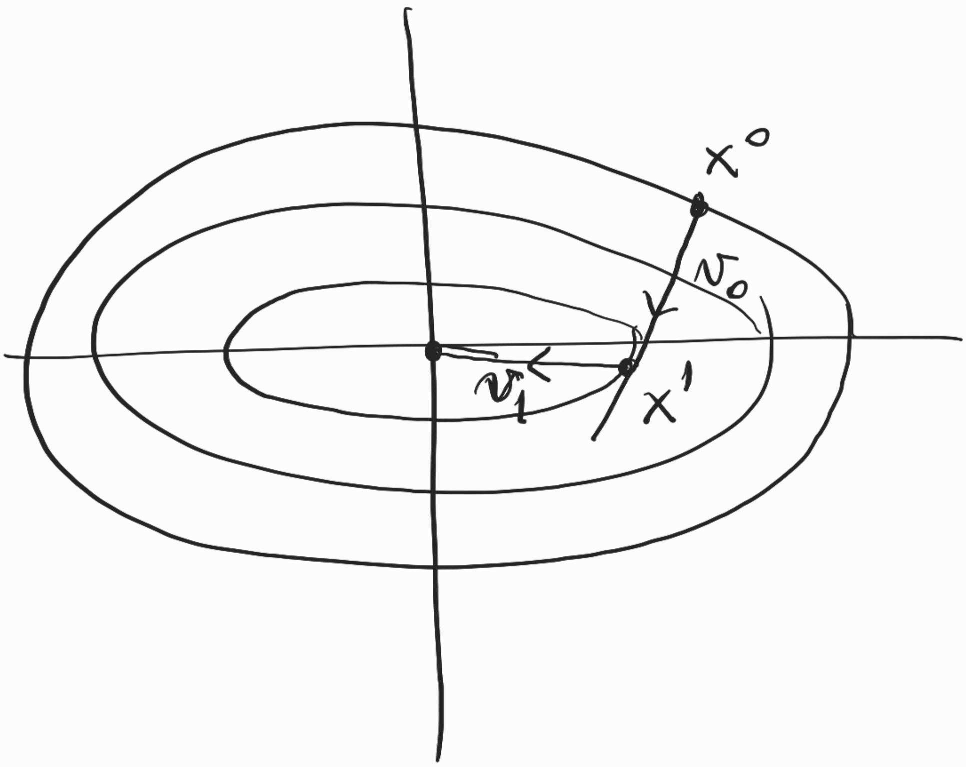
Con lo cual si estamos en una curva de nivel de ax^2+by^2 y queremos movernos en una dirección que sea un rayo al centro, necesitamos movernos por una dirección v que sea "perpendicular a la tangente según el p.i inducido por la matriz A".

Es un p.i en $\mathbb{R}^n$!

$\langle x, y \rangle_A = x^t A y$. Se dice que x e y son A-conjugados o A-ortogonales si $\langle x, y \rangle_A = 0$

La idea en este caso será la siguiente:

Comenzamos con x_0 y $v_1 \neq 0$ un vector cualquiera



$$x_1 = x_0 + t_1 v_1, \text{ con } t_1 = \frac{-v_1^t \overbrace{(Ax_0 - b)}^{-r_0}}{v_1^t A v_1}$$

Luego consideramos $v_2 \neq 0$: $\langle v_1, v_2 \rangle_A = 0$

$$x_2 = x_1 + t_2 v_2, \text{ con } t_2 = -\frac{v_2^t (Ax_1 - b)}{v_2^t A v_2}$$

$$Ax_2 = Ax_1 + t_2 A v_2$$
$$= Ax_0 + t_1 A v_1 + t_2 A v_2$$

$$\Rightarrow Ax_2 - b = Ax_0 - b + t_1 A v_1 + t_2 A v_2$$

$$\Rightarrow \langle v_1, Ax_2 - b \rangle = \langle v_1, Ax_0 - b \rangle$$

$$+ \; t_1 \langle v_1, A v_1 \rangle + t_2 \underbrace{\langle v_1, A v_2 \rangle}_{= \langle v_1, v_2 \rangle_A = 0}$$

Recordando

$$t_1 = -\frac{v_1^t (Ax_0 - b)}{v_1^t A v_1} = -\frac{\langle v_1, Ax_0 - b \rangle}{\langle v_1, A v_1 \rangle}$$

nos queda

$$\langle v_1, \underbrace{Ax_2 - b}_{r_2} \rangle = 0 \qquad (r_2 \perp v_1)$$

De la misma manera (usando t_1)

nos queda

$$\langle v_2, Ax_2 - b \rangle = 0 \qquad (r_2 \perp v_2)$$

Como $\langle v_1, v_2 \rangle_A = 0$ podemos probar que $\{v_1, v_2\}$ son una base de $\mathbb{R}^2$ (conjunto l.i).

Veamos eso: Sean α, β:

$$\alpha v_1 + \beta v_2 = 0$$

$$\Rightarrow \alpha A v_1 + \beta A v_2 = 0$$

$$\Rightarrow \langle v_1, \alpha A v_1 + \beta A v_2 \rangle = 0$$

$$\Rightarrow \alpha \langle v_1, A v_1 \rangle + \beta \underbrace{\langle v_1, A v_2 \rangle}_{=0} = 0$$

$$\Rightarrow \alpha \cdot \underbrace{\langle v_1, A v_1 \rangle}_{>0} = 0$$

>0 ya que A es def + y $v_1 \neq 0$

$$\Rightarrow \alpha = 0. \Rightarrow \beta = 0$$

Luego $\{v_1, v_2\}$ es un conjunto l.i.

Teníamos

$\langle Ax_2 - b, v_1 \rangle = 0$ y $\langle Ax_2 - b, v_2 \rangle = 0$

donde $\{v_1, v_2\}$ son base de $\mathbb{R}^2$

$\Rightarrow \quad Ax_2 - b = 0.$

Este método se conoce como

Método de direcciones conjugadas.

Consiste en lo siguiente:

Dada $A \in \mathbb{R}^{n \times n}$ simétrica y definida positiva se consideran

$\{q_1, \dots, q_n\}$ n direcciones

A- conjugadas y no nulas.

Ejercicio: Si $\{q_1, \dots, q_n\}$ son

vectores A- conjugados no nulos $\Rightarrow$
$\{q_1, \ldots, q_n\}$ es un conjunto l.i.
(Lo hice antes para $n=2$).

Vale el siguiente Teorema:

Sea A una matriz simétrica y def $+$ y sean $\{q_1, \ldots, q_n\}$ un conjunto de vectores A- conjugados y no nulos. Para cualquier $x^0 \in \mathbb{R}$, la iteración

$$x_k = x_{k-1} + t_k \, q_k$$

con $t_k = -\dfrac{\langle q_k, A x_{k-1} - b \rangle}{\langle q_k, A q_k \rangle}$

con $k = 1, \ldots, n$ converge a la solución x^* de $Ax = b$ en a lo sumo n pasos. Es decir $Ax_n = b$.

Dem: ver Burden (pág 357).

Problema: es muy sensible a los errores y hay que calcular a priori los $\{q_1 \cdots q_n\}$

OBS: Si definimos $r_k = b - Ax_k$ se tiene que $\langle r_k, v_j \rangle = 0 \ \forall j = 1, .., k$.

Dem: misma idea de lo hecho en $\mathbb{R}^2$, por inducción.