Examen Parcial I
Esteban Reyes Saldaña

Problema 1

a) Sea $f: \mathbb{R}^n \to \mathbb{R}$ una función convexa. Definamos la función $g: \mathbb{R}^m \to \mathbb{R}$ como sigue

$$g(x) = f(x_0 + Dx)$$

donde $x \in \mathbb{R}^m$, $x_0 \in \mathbb{R}^n$, $D \in \mathbb{R}^{n \times m}$. Muestre que la función $g(x)$ es convexa.

Sean $x, y \in \mathbb{R}^m$. Hay que demostrar que

$$g(\alpha x + (1-\alpha) y) \leq \alpha g(x) + (1-\alpha) g(y)$$

Demostración. Sea $\alpha \in (0,1)$ y $x, y \in \mathbb{R}^n$. Como $f(x)$ es convexa,

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

Ahora, Sean $x, y \in \mathbb{R}^m$. Luego,

$$\begin{aligned}
g(\alpha x + (1-\alpha) y) &= f(x_0 + D[\alpha x + (1-\alpha) y]) \\
&= f(x_0 + \alpha D x + (1-\alpha) D y) \\
&= f([\alpha + (1-\alpha)] x_0 + \alpha D x + (1-\alpha) D y) \\
&= f(\alpha x_0 + (1-\alpha) x_0 + \alpha D x + (1-\alpha) D y) \\
&= f(\alpha x_0 + \alpha D x + (1-\alpha) x_0 + (1-\alpha) D y) \\
&= f(\alpha (x_0 + D x) + (1-\alpha)(x_0 + D y)) \\
&\leq \alpha f(x_0 + D x) + (1-\alpha) f(x_0 + D y) \\
&= \alpha g(x) + (1-\alpha) g(y) \quad \square
\end{aligned}$$

Uso el hecho que $1 = \alpha + (1-\alpha)$

convexidad de $f(x)$

definición de $g(x)$

b) Muestre que la función $f(x)=||x||^2$ con $x \in \mathbb{R}^n$ es convexa.

Demostración.

Lema. Sean $f(z): \mathbb{R}^n \to n$ y $g(x): \mathbb{R} \to \mathbb{R}$ convexas y $g(x)$ no decreciente. Entonces

$$g \circ f(z) \text{ es convexa.}$$

Demostración (del lema). Sean $x, y \in \mathbb{R}^n$ y $\alpha \in (0,1)$. Luego, por convexidad de $f(z)$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

Notemos que

$$f(\alpha x + (1-\alpha) y \in \mathbb{R}$$
$$\alpha f(x) \quad \alpha f(x) \in \mathbb{R}$$
$$(1-\alpha) f(y) \in \mathbb{R}$$

Y usando el hecho que $g(x)$ es no decreciente,

$$g(f(\alpha x + (1-\alpha) y)) \leq g(\alpha f(x) + (1-\alpha) f(y))$$

como $g(x)$ es convexa,

$$g \circ f(\alpha x + (1-\alpha) y) \leq \alpha g \circ f(x) + (1-\alpha) g \circ f(y)$$

por lo tanto, $g \circ f(z)$ es convexa. □

Ahora, para la demostración del problema. Sea

$$f(z) = ||z|| : \mathbb{R}^n \to \mathbb{R} \quad \text{y} \quad g(x) = x^2 : \mathbb{R} \to \mathbb{R}$$

Notemos que ambas son convexas y $g(x)$ es no decreciente en $[0, \infty)$.

En efecto,

PROBLEMA 1 (Inciso B) continuación

(i) $f(z) = \|z\| : \mathbb{R}^n \to [0,\infty)$ es convexa.

Dado que si $x, y \in \mathbb{R}^n$ y $\alpha \in (0,1)$

$$
\begin{aligned}
f(\alpha x + (1-\alpha) y) &= \|\alpha x + (1-\alpha) y\| \\
&\leq \|\alpha x\| + \|(1-\alpha) y\| \quad \to \text{desigualdad del triángulo} \\
&= |\alpha| \|x\| + |1-\alpha| \|y\| \quad \to \text{Propiedad de norma} \\
&= \alpha \|x\| + (1-\alpha) \|y\| \quad \to \alpha > 0,\ 1-\alpha > 0 \\
&= \alpha f(x) + (1-\alpha) f(y).
\end{aligned}
$$

(ii) $g(x) = x^2 : \mathbb{R} \to \mathbb{R}$ es convexa.

$$g'(x) = 2x$$
$$g''(x) = 2.$$

Notemos que como $g''(x) = 2 > 0$ entonces es convexa

(iii) $g(x) = x^2 : [0,\infty) \to [0,\infty)$ no decreciente.

Sean $x \leq y$ con $x \geq 0$ y $y \geq 0$ entonces

$$x^2 \leq y^2$$

Luego,

$$g(x) \leq g(y).$$

Así que por el lema anterior,

$$
\begin{aligned}
g(f(x)) &= g(\|x\|) \\
&= \|x\|^2
\end{aligned}
$$

es convexa. $\square$

Esteban Reyes Saldaña

Sea la función $f: \mathbb{R}^m \to \mathbb{R}$ definida a continuación

$$f(x) = \frac{1}{2}\|Ax - b\|^2 + \frac{\lambda}{2}\|Dx\|^2,$$

con $\lambda > 0$, $x \in \mathbb{R}^m$, $b \in \mathbb{R}^n$, $D \in \mathbb{R}^{p \times m}$, $A \in \mathbb{R}^{n \times m}$ una matriz de rango completo.

(a) $f(x)$ es una función convexa?

(b) Calcula y clasifica los puntos críticos de $f(x)$.

(a) $f(x)$ es convexa.

<u>Demostración</u>. Sean $x, y \in \mathbb{R}^n$. Por el problema 1(b) sabemos que $\|x\|^2$ es convexa, i.e., para todo $\alpha \in (0,1)$,

$$\|\alpha x + (1-\alpha)y\|^2 \leq \alpha\|x\|^2 + (1-\alpha)\|y\|^2 \quad (1)$$

Ahora,

$$f(\alpha x + (1-\alpha)y) = \frac{1}{2}\|A[\alpha x + (1-\alpha)y] - b\|^2 + \frac{\lambda}{2}\|D[\alpha x + (1-\alpha)y]\|^2$$

$$= \frac{1}{2}\|\alpha Ax + (1-\alpha)Ay - b\|^2 + \frac{\lambda}{2}\|\alpha Dx + (1-\alpha)Dy\|^2$$

Uso el hecho que $1 = \alpha + (1-\alpha)$

$$= \frac{1}{2}\|\alpha Ax + (1-\alpha)Ay - [\alpha + (1-\alpha)]b\|^2 + \frac{\lambda}{2}\|\alpha Dx + (1-\alpha)Dy\|^2$$

reordeno términos

$$\underbrace{= \frac{1}{2}\|\alpha Ax - \alpha b + (1-\alpha)Ay - (1-\alpha)b\|^2 + \frac{\lambda}{2}\|\alpha Dx + (1-\alpha)Dy\|^2}_{(*)}$$

$$(*) = \frac{1}{2}\|\alpha(Ax-b)+(1-\alpha)(Ay-b)\|^2 + \frac{\lambda}{2}\|\alpha Dx + (1-\alpha)Dy\|^2$$

$$\leq \frac{1}{2}\left[\alpha\|Ax-b\|^2 + (1-\alpha)\|Ay-b\|^2\right] + \frac{\lambda}{2}\left[\alpha\|Dx\|^2 + (1-\alpha)\|Dy\|^2\right] \leftarrow$$

usando convexidad de $\|\cdot\|^2$ del problema 1(b)

$$= \frac{1}{2}\alpha\|Ax-b\|^2 + \frac{1}{2}(1-\alpha)\|Ay-b\|^2 + \frac{\lambda}{2}\alpha\|Dx\|^2 + \frac{\lambda}{2}(1-\alpha)\|Dy\|^2$$

$$= \frac{1}{2}\alpha\|Ax-b\|^2 + \frac{\lambda}{2}\alpha\|Dx\|^2 + \frac{1}{2}(1-\alpha)\|Ay-b\|^2 + \frac{\lambda}{2}(1-\alpha)\|Dy\|^2$$

reordeno términos

$$= \alpha\left[\frac{1}{2}\|Ax-b\|^2 + \frac{\lambda}{2}\|Dx\|^2\right] + (1-\alpha)\left[\frac{1}{2}\|Ay-b\|^2 + \frac{\lambda}{2}\|Dy\|^2\right]$$

$$= \alpha f(x) + (1-\alpha)f(y).$$

Por lo tanto

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y).$$

así que $f(x)$ es convexa. $\square$

PROBLEMA 2 inciso (b). Esteban Reyes Saldaña

(b)
$$f(x) = \frac{1}{2}\|Ax - b\|^2 + \frac{\lambda}{2}\|Dx\|^2$$

reescribiendo $f(x)$ con producto punto obtenemos

$$f(x) = \frac{1}{2}(Ax-b)^T(Ax-b) + \frac{\lambda}{2}(Dx)^T Dx$$

Ahora, usando la regla del producto para derivadas se obtiene,

$$\nabla f(x)^T = \frac{1}{2}\left[\left(D_x(Ax-b)^T\right)\cdot(Ax-b) + (Ax-b)^T D_x(Ax-b)\right] + \frac{\lambda}{2}\left[\left(D_x(Dx)^T\right)\cdot Dx + (Dx)^T D_x(Dx)\right]$$

$\boxed{\text{transposición}}$
$$= \frac{1}{2}\left[A^T(Ax-b) + (Ax-b)^T A\right] + \frac{\lambda}{2}\left[D^T Dx + (DX)^T D\right]$$
$$= (Ax-b)^T A + \lambda D^T Dx$$

Luego, $\nabla f(x) = 0 \iff (Ax-b)^T A + \lambda D^T D x = 0$

Entonces
$$A^T(Ax-b) + \lambda D^T D x = 0$$
$$A^T A x - A^T b + \lambda D^T D x = 0$$
$$(A^T A + \lambda D^T D)x = A^T b$$

Notemos que dado que A es de rango completo y asumiendo que D también $A^TA + \lambda D^T D$ es invertible.

Luego,
$$x^* = (A^TA + \lambda D^T D)^{-1} A^T b$$

Ahora,
$$\nabla^2 f(x)^T = D_x(A^TAx - A^Tb + \lambda D^T D x)$$
$$= A^TA + \lambda D^T D$$

que es definida positiva dado que $\lambda > 0$ y A^TA y D^TD son definidas positivas (por el rango completo).

Luego x^* es mínimo y además como f es convexa, x^* es mínimo global.

Esteban Reyes Saldaña

Problema 3.

Sea $f(x) = \frac{1}{2}x^TQx - b^Tx$ donde $Q \succ 0$, i.e. Q es positiva definida. Si se usa un algoritmo de búsqueda en línea con tamaño de paso exacto, muestra que se satisface la siguiente condición de Goldstein:

$$f(x_{k+1}) \leq f(x_k) + c\alpha_k \nabla f(x_k)^T d_k.$$

donde $0 < c < \frac{1}{2}$, d_k es la dirección de descenso en el punto x_k y α_k es el tamaño de paso exacto.

Demostración. Primero calculemos el tamaño de paso, i.e.

$$\alpha_k = \operatorname*{armin}_{\alpha > 0} f(x_k + \alpha d_k)$$

Notemos que

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

Entonces

$$\nabla f(x_k + \alpha d_k) = Q(x_k + \alpha d_k) - b$$

Por otro lado, si

$$\phi(\alpha) = f(x_k + \alpha d_k)$$
$$\phi'(\alpha) = \nabla f^T(x_k + \alpha d_k)\, d_k$$

Así que,

$$\phi'(\alpha) = 0 \Leftrightarrow -\nabla f^T(x_k + \alpha d_k)\, d_k = 0$$

$$\Leftrightarrow [Q(x_k + \alpha d_k) - b]^T d_k = 0$$

$$\Leftrightarrow [(x_k^T + \alpha d_k^T)Q - b^T]\, d_k = 0$$

$$\Leftrightarrow x_k^T Q d_k + \alpha d_k^T Q d_k - b^T d_k = 0$$

$$\Leftrightarrow -\alpha\, d_k^T Q d_k = x_k^T Q d_k - b^T d_k$$

$$\alpha = -\frac{(x_k^T Q - b^T)\, d_k}{d_k^T Q d_k} = -\frac{g_k^T d_k}{d_k^T Q d_k}$$

$$\Leftrightarrow \boxed{\alpha = -\frac{g_k^T d_k}{d_k^T Q d_k}} \quad (1)$$

Ahora

$$f(x_{k+1}) = \frac{1}{2} x_{k+1}^T Q x_{k+1} - b^T x_{k+1}$$

$$= \frac{1}{2}(x_k + \alpha_k d_k)^T Q (x_k + \alpha_k d_k) - b^T(x_k + \alpha_k d_k)$$

$$= \frac{1}{2}(x_k^T + \alpha_k d_k^T) Q (x_k + \alpha_k d_k) - b^T(x_k + \alpha_k d_k)$$

$$= \frac{1}{2}\left[x_k^T Q x_k + \alpha_k x_k^T Q d_k + \alpha_k d_k^T Q x_k + \alpha_k^2 d_k^T Q d_k\right]$$

$$- b^T x_k - \alpha_k b^T d_k$$

$$= \frac{1}{2}\left[x_k^T Q x_k - b^T x_k\right] + \frac{1}{2}(\alpha_k d_k^T Q x_k + \alpha_k x_k^T Q d_k$$

$$+ \alpha_k^2 d_k^T Q d_k) - \alpha_k b^T d_k$$

Producto punto $\rightarrow$

$$= f(x_k) + \frac{1}{2}(\alpha_k x_k^T Q d_k + \alpha x_k^T Q d_k + \alpha^2 d_k^T Q d_k)$$

$$- \alpha_k b^T d_k$$

$$= f(x_k) + \frac{1}{2}(2\alpha_k x_k^T Q d_k) + \frac{1}{2}\alpha^2 d_k^T Q d_k - \alpha_k b^T d_k$$

$$= f(x_k) + \alpha_k x_k^T Q d_k - \alpha_k b^T d_k + \frac{1}{2}\alpha_k^2 d_k^T Q d_k$$

$$= f(x_k) + \alpha_k (x_k^T Q d_k - b^T d_k) + \frac{1}{2}\alpha_k^2 d_k^T Q d_k$$

$$= f(x_k) + \alpha_k (x_k^T Q - b^T) d_k + \frac{1}{2}\alpha_k^2 d_k^T Q d_k$$

tamaño de paso (1) $\rightarrow$

$$= f(x_k) + \alpha_k \nabla f(x_k)^T d_k + \frac{1}{2}\alpha_k \left(-\frac{g_k^T d_k}{d_k^T \cancel{Q d_k}}\right) d_k^T \cancel{Q d_k}$$

$$= f(x_k) + \alpha_k \nabla f(x_k)^T d_k - \frac{1}{2}\alpha_k g_k^T d_k$$

$$= f(x_k) + \alpha_k g_k^T d_k - \frac{1}{2}\alpha_k g_k^T d_k$$

$$= f(x_k) + \frac{1}{2}\alpha_k g_k^T d_k$$

hipótesis $0 < c \leq \frac{1}{2}$ $\rightarrow$

$$\leq f(x_k) + c\,\alpha_k \nabla f(x_k)^T d_k \;\therefore\; \square$$