MA57B Optimización No Lineal. Semestre 2007-1
Profesor: Héctor Ramírez C. **Auxiliar:** Oscar Peredo.

Clase Auxiliar #1
Optimización Irrestricta

15 de Marzo del 2007

1. Optimización Irrestricta

Teorema 1.1. *Sea S convexo abierto no vacío en $\mathbb{R}^n$ y $f : S \to \mathbb{R}$, diferenciable en S. Entonces f es convexa si y solo si $\forall \overline{x} \in S$*

$$f(x) \quad \geq \quad f(\overline{x}) + \nabla f(\overline{x})^t (x - \overline{x}) \qquad \forall x \in S$$

Teorema 1.2 (Valor Medio en $\mathbb{R}^n$)**.** *Sea $f : S \subset \mathbb{R}^n \to \mathbb{R}$ diferenciable en S, con S abierto convexo no vacío. Entonces para todo par de puntos $x_1, x_2 \in S$ existe $\tilde{x} \in]x_1, x_2[$, donde $]x_1, x_2[= \{\lambda x_1 + (1-\lambda)x_2 : \lambda \in (0,1)\}$, tal que:*

$$f(x_2) - f(x_1) \quad = \quad \nabla f(\tilde{x})^t (x_2 - x_1)$$

Problema 1.1. *Utilizando las hipótesis del teorema anterior, pruebe que f es convexa si y solo si $\forall x_1, x_2 \in S$*

$$(\nabla f(x_2) - \nabla f(x_1))^t (x_2 - x_1) \quad \geq \quad 0$$

Solución 1.1. *Probemos las dos implicancias:*

$\Rightarrow$ *Supongamos f convexa. Sean $x_1, x_2 \in S$. Por el teorema 1, se tiene que*

$$\begin{aligned} f(x_1) &\geq f(x_2) + \nabla f(x_2)^t (x_1 - x_2) \\ f(x_2) &\geq f(x_1) + \nabla f(x_1)^t (x_2 - x_1) \end{aligned}$$

Sumando ambas inecuaciones:

$$\begin{aligned} f(x_1) + f(x_2) &\geq f(x_1) + f(x_2) + (\nabla f(x_2) - \nabla f(x_1))^t (x_1 - x_2) \\ 0 &\geq (\nabla f(x_2) - \nabla f(x_1))^t (x_1 - x_2) \\ 0 &\leq (\nabla f(x_2) - \nabla f(x_1))^t (x_2 - x_1) \end{aligned}$$

$\Leftarrow$ *Sean $x_1, x_2 \in S$. Por el teorema del valor medio (teorema 1.2), existe $x = \lambda x_1 + (1-\lambda)x_2 \in S$ tal que*

$$f(x_2) - f(x_1) \quad = \quad \nabla f(x)^t (x_2 - x_1) \qquad (*)$$

Por hipótesis:

$$\begin{aligned} 0 &\leq (\nabla f(x) - \nabla f(x_1))^t (x - x_1) \\ 0 &\leq (\nabla f(x) - \nabla f(x_1))^t (\lambda x_1 + (1-\lambda)x_2 - x_1) \\ 0 &\leq (\nabla f(x) - \nabla f(x_1))^t ((1-\lambda)(x_2 - x_1)) \\ 0 &\leq (\nabla f(x) - \nabla f(x_1))^t (x_2 - x_1) \\ \nabla f(x_1)^t (x_2 - x_1) &\leq \nabla f(x)^t (x_2 - x_1) \\ \nabla f(x_1)^t (x_2 - x_1) &\underbrace{\leq}_{(*)} f(x_2) - f(x_1) \\ f(x_1) + \nabla f(x_1)^t (x_2 - x_1) &\leq f(x_2) \end{aligned}$$

Utilizando el teorema 1.1, se concluye que f es convexa.$\square$

Problema 1.2 (Bertsekas, página 19). *Sea $f:\mathbb{R}^n\to\mathbb{R}$ de clase C^2 que satisface*

$$m\|y\|^2 \leq y^T\nabla^2 f(x)y \leq M\|y\|^2, \qquad \forall x,y\in\mathbb{R}^n$$

donde $\nabla^2 f(x)$ es la matríz hessiana evaluada en x y $m,M\geq 0$. Pruebe que f tiene un único m"inimo global x^* que satisface

$$\frac{1}{2M}\|\nabla f(x)\|^2 \leq f(x)-f(x^*) \leq \frac{1}{2m}\|\nabla f(x)\|^2, \qquad \forall x\in\mathbb{R}^n$$

y

$$\frac{m}{2}\|x-x^*\|^2 \leq f(x)-f(x^*) \leq \frac{M}{2}\|x-x^*\|^2, \qquad \forall x\in\mathbb{R}^n$$

HINT: Use expansión de segundo orden y la siguiente relación:

$$\min_{y\in\mathbb{R}^n}\{\nabla f(x)^T(y-x)+\frac{\alpha}{2}\|y-x\|^2\} = -\frac{1}{2\alpha}\|\nabla f(x)\|^2$$

Solución 1.2. *http://www.athenasc.com/nlpsol1.pdf*

2. Condiciones de Optimalidad y Coercividad

Teorema 2.1 (Condición necesaria de Primer Orden, una vez diferenciable). *Sea $f:\mathbb{R}^n\to\mathbb{R}$ diferenciable en $x\in int(Dom(f))$. Si x es óptimo local de f entonces $\nabla f(x)=0$.*

Teorema 2.2 (Condiciones necesarias de Primer Orden y Segundo Orden, dos veces diferenciable). *Sea $f:\mathbb{R}^n\to\mathbb{R}$ dos veces diferenciable en $x\in int(Dom(f))$. Si x es óptimo local de f entonces $\nabla f(x)=0$ y $H_f(x)$ es semidefinida positiva.*

Teorema 2.3 (Condición Suficiente). *Sea $f:\mathbb{R}^n\to\mathbb{R}$ dos veces diferenciable en $x\in int(Dom(f))$. Si $\nabla f(x)=0$ y $H_f(x)$ es definida positiva, entonces x es mínimo local de f.*

Teorema 2.4 (Coercividad y Continuidad). *Sea $f:\mathbb{R}^n\to\mathbb{R}$ continua y coerciva (i.e. $\lim_{||x||\to\infty} f(x)=\infty$) . Entonces f tiene al menos un mínimo.*

Demostración. *Consideremos la sucesión z_k tal que $\lim_{k\to\infty} f(z_k)=\inf_x f(x)$ (el ínfimo puede ser finito o $-\infty$). Como f es coerciva, la sucesión z_k debe ser acotada, pues de no serlo se tendrá $\lim_{j\to\infty} f(z_{k_j})=\infty$ para alguna subsucesión z_{k_j} que satisface $\|z_{k_j}\|\to\infty$, lo cual indicaría que $\inf_x f(x)=\infty$. Ahora bien, como z_k es acotada, tiene al menos un punto de acumulación, $z\in\mathbb{R}^n$, y por continuidad de f se tiene que $f(z)=\lim_{k\to\infty} f(z_k)=\inf_x f(x)$.*$\square$

3. Discretización del Método de Máximo Descenso

Sea H un Hilbert y Φ función diferenciable y convexa. El Método de Máximo Descenso (Steepest Descent)

- Caso Explícito:

$$x^{k+1}=x^k-\lambda_k\nabla\Phi(x^k) \qquad (E)$$

con $\lambda_k>0$. El principal resultado es el siguiente teorema:

Teorema 3.1. *Sea $\Phi:H\to\mathbb{R}$ una función diferenciable y convexa tal que $Argmin\{\Phi\}$ es distinto de vacío y $\nabla\Phi$ es Lipschitz sobre acotados. Sea la iteracón definida por*

$$x^{k+1}=x^k-\lambda_k\frac{\nabla\Phi(x^k)}{\|\nabla\Phi(x^k)\|}$$

donde $\sum_k\lambda_k=+\infty$ y $\sum\lambda_k^2<+\infty$, entonces

1. *$\forall z\in Argmin\{\Phi\}$, el límite $\lim_{k\to+\infty}\|x^k-z\|$ existe.*
2. *Existe una subsucesión de x^k tal que $x^{k_i}\to x^\infty\in Argmin\{\Phi\}$.*
3. *Si además $dimH<+\infty$, entonces $x^k\to x^\infty$.*

Demostración. *:*

1. *Sea $z \in Argmin\{\Phi\}$. Definamos $\varphi_k := \|x^k - z\|^2$. Notemos que:*

$$\begin{aligned}
\varphi_{k+1} &= \|x^{k+1} - z\|^2 \\
&= \|x^{k+1} - x^k + x^k - z\|^2 \\
&= \|x^{k+1} - x^k\|^2 + 2 < x^{k+1} - x^k, x^k - z > + \varphi_k \\
&= \|\lambda_k \frac{\nabla\Phi(x^k)}{\|\nabla\Phi(x^k)\|}\|^2 + 2 < x^{k+1} - x^k, x^k - z > + \varphi_k \\
&= \lambda_k^2 + 2 < x^{k+1} - x^k, x^k - z > + \varphi_k \\
&= \lambda_k^2 - \frac{2\lambda_k}{\|\nabla\Phi(x^k)\|} < \nabla\Phi(x^k), x^k - z > + \varphi_k
\end{aligned}$$

Por el teorema 1.1 (debido a la convexidad y diferenciabilidad de Φ) se tiene:

$$< \nabla\Phi(x^k), x^k - z > \quad \leq \quad \Phi(x^k) - \Phi(z)$$

Con esto se tiene la siguiente desigualdad:

$$\begin{aligned}
\varphi_{k+1} &\leq \varphi_k + \lambda_k^2 - \frac{2\lambda_k}{\|\nabla\Phi(x^k)\|} \underbrace{(\Phi(x^k) - \Phi(z))}_{\geq 0} \\
&\leq \varphi_k + \lambda_k^2
\end{aligned}$$

Probemos el lema:

Lema 3.1. *Si (φ_k) es una sucesión acotada inferiormente y tal que existe (δ_k) con $\sum \delta_k < +\infty$ de modo que $\varphi_{k+1} \leq \varphi_k + \delta_k$. Entonces φ_k converge.*

Demostración. *Sumando $\sum_{j=k+1}^{\infty} \delta_j$ a ambos lados de la desigualdad, se obtiene $\theta_k \leq \theta_{k+1}$, con $\theta_k := \varphi_k + \sum_{j=k}^{\infty} \delta_j$. Como φ_k es acotado inferiormente, θ_k tambien, luego θ_k converge, y como $\sum_{j=k}^{\infty} \delta_j \to_{k\to\infty} 0$ obtenemos que φ_k converge, y además $\lim_{k\to\infty} \theta_k = \lim_{k\to\infty} \varphi_k$.*$\square$

Como se tiene que $\sum \lambda_k^2 < \infty$ por hipótesis, se concluye que $\lim_{k\to+\infty} \|x^k - z\|$ existe (en particular, x^k es acotado).

2. *Por negación, supongamos que toda sucesión satisface lo contrario, es decir, $\exists \varepsilon > 0$ tal que $\forall k \in \mathbb{N}$, $\Phi(x^k) > \min \Phi + \varepsilon$. De la primera parte:*

$$\varphi_{k+1} \leq \varphi_k + \lambda_k^2 + \frac{2\lambda_k}{\|\nabla\Phi(x^k)\|} (\underbrace{\Phi(z)}_{=\min\Phi} - \Phi(x^k))$$

obtendríamos

$$\frac{2\varepsilon\lambda_k}{\|\nabla\Phi(x^k)\|} \leq \varphi_k - \varphi_{k+1} + \lambda_k^2$$

sumando sobre k se tiene

$$2\varepsilon \sum_k \frac{\lambda_k}{\|\nabla\Phi(x^k)\|} \leq \varphi_0 - \underbrace{\varphi_\infty}_{=\lim_{k\to\infty}\varphi_k} + \underbrace{\sum_k \lambda_k^2}_{<\infty}$$

es decir $\sum_k \frac{\lambda_k}{\|\nabla\Phi(x^k)\|} < \infty$. Por hipótesis, se tiene que $\sum \lambda_k = \infty$, luego $\underline{\lim}_{k\to\infty} \frac{1}{\|\nabla\Phi(x^k)\|} = 0$. Como $\nabla\Phi$ es Lipschitz sobre acotados y x^k es acotada, tenemos

$$\begin{aligned}
\|\nabla\Phi(x^k)\| &\leq L\|x^k\| \\
&\leq \underbrace{LM}_{cte.}
\end{aligned}$$

con $(x^k) \subset B(0, M)$. Es decir, $\nabla\Phi(x^k)$ es uniformemente acotado, lo cual contradice $\underline{\lim}_{k\to\infty} \frac{1}{\|\nabla\Phi(x^k)\|} = 0$ Por lo tanto, existe una subsucesón de x^k convergente a $x^\infty \in Argmin\{\Phi\}$, o equivalentemente $\underline{\lim}_{k\to\infty} \Phi(x^k) = \min \Phi$.

3. *Propuesto.*$\square$

- Caso Implícito:

$$x^{k+1} = x^k - \lambda_k \nabla \Phi(x^{k+1}) \qquad (I)$$

con $\lambda_k > 0$. Si Φ es convexa, la función $\Phi(x) + \frac{1}{2\lambda_k}\|x - x^k\|^2$ es convexa y coerciva, y por teorema anterior, $\text{Argmin}\{\Phi\}$ tiene al menos un elemento (ver Problema siguiente).

El problema (I) anterior es equivalente a

$$x^{k+1} = \text{Argmin}_{x \in H}\{\Phi(x) + \frac{1}{2\lambda_k}\|x - x^k\|^2\}$$

el cual recibe el nombre de Algoritmo del Punto Proximal (PROX).

Problema 3.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función convexa, con derivada continua y $\lambda > 0$. Definamos la función*

$$\phi_\lambda(x) = f(x) + \frac{\lambda}{2}\|x\|^2$$

Demuestre que existe un único mínimo global de ϕ_λ.

Solución 3.1. *La función φ_λ es trivialmente continua. Probemos que es coerciva. Del teorema 1.1, se tiene que*

$$\varphi_\lambda(x) \geq f(0) + \nabla f(0)^T x + \frac{\lambda}{2} x^T x$$

Supongamos $\exists (x_k) \subset \mathbb{R}^n, \|x_k\| \to \infty$ y $\exists M > 0$ tal que $\varphi_\lambda(x_k) \leq M$ para todo k:

$$\begin{aligned} M &\geq f(0) + \nabla f(0)^T x_k + \frac{\lambda}{2} x_k^T x_k \\ \underbrace{M - f(0)}_{\tilde{M}} &\geq \nabla f(0)^T x_k + \frac{\lambda}{2} x_k^T x_k \\ \tilde{M} &\geq (\nabla f(0) + \frac{\lambda}{2} x_k)^T x_k \end{aligned}$$

Geometricamente se tiene que $(\nabla f(0) + \frac{\lambda}{2}x_k)^T x_k = \|\nabla f(0) + \frac{\lambda}{2}x_k\|\|x_k\| \cos\theta_k$, con θ_k el ángulo entre ambos vectores. Como $\|x_k\| \to \infty$, se tendrá que $\cos\theta_k \to_{k\to\infty} 1$, con lo cuál se llega a la contradicción, pues $\exists k_0$ tal que $\forall k \geq k_0, \|\nabla f(0) + \frac{\lambda}{2}x_k\|\|x_k\|\cos\theta_k > \tilde{M} + 1$, luego $\tilde{M} > \tilde{M} + 1$.

Luego, por teorema 2.4, existe al menos un mínimo de φ_λ. La función φ_λ es estrictamente convexa pues si $x, y \in \mathbb{R}^n, x \neq y$:

$$\begin{aligned} \|x - y\|^2 &> 0 \\ \|x\| - 2x^T y + \|y\|^2 &> 0 \\ \frac{\lambda}{2}\|x\|^2 &> -\frac{\lambda}{2}\|y\|^2 + \lambda x^T y \\ f(x) + \frac{\lambda}{2}\|x\|^2 &> f(x) - \frac{\lambda}{2}\|y\|^2 + \lambda x^T y \\ \varphi_\lambda(x) &> f(x) - \frac{\lambda}{2}\|y\|^2 + \lambda x^T y \\ \varphi_\lambda(x) &> f(y) + \nabla f(y)^T(x - y) - \frac{\lambda}{2}\|y\|^2 + \lambda x^T y \\ \varphi_\lambda(x) &> f(y) + \nabla f(y)^T(x - y) + \frac{\lambda}{2}\|y\|^2 + \lambda x^T y - \lambda y^T y \\ \varphi_\lambda(x) &> f(y) + \frac{\lambda}{2}\|y\|^2 + \nabla f(y)^T(x - y) + \lambda y^T(x - y) \\ \varphi_\lambda(x) &> \varphi_\lambda(y) + (\nabla f(y) + \lambda y)^T(x - y) \\ \varphi_\lambda(x) &> \varphi_\lambda(y) + (\nabla f(y) + \nabla(\frac{\lambda}{2}\|y\|^2))^T(x - y) \\ \varphi_\lambda(x) &> \varphi_\lambda(y) + \nabla(f(y) + \frac{\lambda}{2}\|y\|^2)^T(x - y) \\ \varphi_\lambda(x) &> \varphi_\lambda(y) + \nabla\varphi_\lambda(y)^T(x - y) \end{aligned}$$

Por teorema 1.1, se conluye la convexidad estricta. Luego, el mínimo es único.$\square$