EMALCA 2013 – SALTA

Optimización no lineal

usando el método de Newton

Damián Fernández[1]
dfernandez@famaf.unc.edu.ar

[1]FaMAF–Universidad Nacional de Córdoba. CIEM–CONICET.

Notación:

Usaremos $\mathbb{R}$ para denotar el conjunto de números reales. Sucesiones de escalares se denotarán con subíndices, *verbi gratia* (v.g.), $\{\alpha_k\}$.

Vectores

Usaremos $\mathbb{R}^n$ para denotar el espacio de vectores reales n-dimensionales. Los elementos de $\mathbb{R}^n$ serán denotados con letras minúsculas y dado $x \in \mathbb{R}^n$ denotaremos por x_i su i-ésima componente. Todo $x \in \mathbb{R}^n$ será considerado un vector columna, o sea, $x \in \mathbb{R}^{n\times 1}$. Por lo tanto, si $x \in \mathbb{R}^n$ entonces

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i \in \mathbb{R}.$$

Denotaremos por $x^\top$ al vector fila (o sea, $x \in \mathbb{R}^{1\times n}$) que es transpuesto de $x \in \mathbb{R}^n$. Sucesiones de vectores se denotarán con supraíndices, o sea, $\{x^k\}$.

Sean $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$ y $\alpha \in \mathbb{R}$, entonces:

$$\begin{array}{lll} z = \alpha x & \Longrightarrow & z_i = \alpha x_i, \\ z = x + y & \Longrightarrow & z_i = x_i + y_i, \\ z = x \circ y & \Longrightarrow & z_i = x_i y_i \qquad \text{(producto de Hadamard)}, \\ c = \langle x, y\rangle = x^\top y & \Longrightarrow & c = \displaystyle\sum_{i=1}^{n} x_i y_i \qquad \text{(producto interno)}, \end{array}$$

Además, usaremos la norma vectorial $\|x\| := \sqrt{\langle x, x\rangle}$.

Matrices

Usaremos $\mathbb{R}^{n\times m}$ para denotar el espacio de matrices reales de n filas y m columnas. Las matrices serán denotadas con letras mayúsculas (A, B, C, etc.) y se usará su respectiva minúscula con subíndice ij para denotar el elemento (i, j) de la misma (a_{ij}, b_{ij}, c_{ij}, etc.). Por lo tanto, si $A \in \mathbb{R}^{n\times m}$ entonces

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \qquad a_{ij} \in \mathbb{R}.$$

Dada una matriz $A \in \mathbb{R}^{n\times m}$, denotaremos por A_{*j} la j-ésima columna de A y por A_{i*} la i-ésima fila de A. Denotaremos por $A^\top$ la transpuesta de $A \in \mathbb{R}^{n\times m}$. Sucesiones de matrices se denotarán con subíndices, o sea, $\{M_k\}$.

Sean $A, \tilde{A} \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{m\times p}$ y $\alpha \in \mathbb{R}$, entonces:

$$\begin{array}{lll} C = \alpha A & \Longrightarrow & c_{ij} = \alpha a_{ij}, \\ C = A + \tilde{A} & \Longrightarrow & c_{ij} = a_{ij} + \tilde{a}_{ij}, \\ C = A \circ \tilde{A} & \Longrightarrow & c_{ij} = a_{ij}\tilde{a}_{ij} \qquad \text{(producto de Hadamard)}, \\ C = AB & \Longrightarrow & c_{ij} = \displaystyle\sum_{k=1}^{m} a_{ik} b_{kj}, \end{array}$$

Además, usaremos la norma matricial $\|A\| := \max_{\|x\|=1} \|Ax\|$.

Diremos que $A \in \mathbb{R}^{n\times n}$ es definida positiva si $\langle Ax, x\rangle > 0$ para todo $x \in \mathbb{R}^n$, $x \neq 0$.

Diremos que $A \in \mathbb{R}^{n\times n}$ es simétrica si $A^\top = A$ y antisimétrica si $A^\top = -A$.

Conjuntos

Usaremos $\mathcal{P}(\mathbb{R}^n)$ para denotar el espacio formado por todos los subconjuntos de $\mathbb{R}^n$. Los elementos de $\mathcal{P}(\mathbb{R}^n)$ serán denotados con letras mayúsculas caligráficas y todo $x \in \mathbb{R}^n$ es considerado un elemento de $\mathcal{P}(\mathbb{R}^n)$, o sea, usaremos x y $\{x\}$ de manera indistinta.

Para $\mathcal{U} \in \mathcal{P}(\mathbb{R}^n)$ usaremos $\mathrm{int}(\mathcal{U})$ y $\mathrm{cl}(\mathcal{U})$ para denotar el interior y la clausura de $\mathcal{U}$.

Denotaremos por $\mathcal{B}_r(x)$ a la bola unitaria cerrada de radio $r > 0$ y centro $x \in \mathbb{R}^n$, i.e., $\mathcal{B}_r(x) = \{y \in \mathbb{R}^n \mid \|y - x\| \leq r\}$. Denotaremos por $\mathbb{R}^n_+$ al ortante no negativo n dimensional, i.e., $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_i \geq 0,\, i = 1 \ldots, n\}$.

Sean $\mathcal{U} \in \mathcal{P}(\mathbb{R}^n)$, $\mathcal{V} \in \mathcal{P}(\mathbb{R}^n)$ y $\alpha \in \mathbb{R}$, entonces:

$$
\begin{array}{lcll}
\mathcal{W} = \alpha\mathcal{U} & \Longrightarrow & \mathcal{W} = \{\alpha u \mid u \in \mathcal{U}\}, & \\
\mathcal{W} = \mathcal{U} \pm \mathcal{V} & \Longrightarrow & \mathcal{W} = \{u \pm v \mid u \in \mathcal{U},\, v \in \mathcal{V}\} & \text{(suma de Minkowski)}, \\
\mathcal{W} = \mathcal{U} \setminus \mathcal{V} & \Longrightarrow & \mathcal{W} = \{w \mid w \in \mathcal{U},\, w \notin \mathcal{V}\} & \text{(diferencia de conjuntos)},
\end{array}
$$

Funciones

Usaremos la notación $\psi(t) = o(t)$ para cualquier función $\psi : \mathbb{R}_+ \to \mathbb{R}^m$ tal que $\displaystyle\lim_{t\to 0^+} \frac{1}{t}\psi(t) = 0$, o sea, si dado $\varepsilon > 0$ existe $\delta > 0$ tal que $\|\psi(t)\| \leq \varepsilon t$ para todo $t \in (0, \delta]$.

Diremos que $F : \mathbb{R}^n \to \mathbb{R}^m$ es Lipschitz contínua en $D \subset \mathbb{R}^n$ si existe $L \geq 0$ tal que

$$\|F(y) - F(x)\| \leq L\|y - x\| \quad \text{para todo} \quad x, y \in D.$$

Diremos que $F : \mathbb{R}^n \to \mathbb{R}^m$ es diferenciable en $\bar{x} \in \mathbb{R}^n$ si existe $M \in \mathbb{R}^{m\times n}$ tal que

$$F(x) = F(\bar{x}) + M(x - \bar{x}) + o(\|x - \bar{x}\|).$$

En tal caso diremos que M es la derivada de F en $\bar{x}$ y la denotaremos por $F'(\bar{x})$. Para ser coherentes con la notación, cuando $m = 1$ tomaremos $F'(\bar{x}) = M^\top$.

Por lo tanto $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable en $\bar{x}$ si y solo si

$$f(x) = f(\bar{x}) + \langle f'(\bar{x}), x - \bar{x}\rangle + o(\|x - \bar{x}\|);$$

f es dos veces diferenciable en $\bar{x}$ si y solo si $f' : \mathbb{R}^n \to \mathbb{R}^n$ es diferenciable en $\bar{x}$; en este caso $f''(\bar{x}) \in \mathbb{R}^{n\times n}$ y

$$f(x) = f(\bar{x}) + \langle f'(\bar{x}), x - \bar{x}\rangle + \frac{1}{2}\langle f''(\bar{x})(x - \bar{x}), x - \bar{x}\rangle + o(\|x - \bar{x}\|^2);$$

Precaución al momento de derivar

La imposición de identificar todo vector como vector columna (i.e., $\mathbb{R}^n \equiv \mathbb{R}^{n\times 1}$) surge como solución para hacer compatible la notación estándar del *Cálculo en varias variables* con la notación estándar del *Álgebra Lineal Numérica*. Esta simple modificación en la notación impacta al momento de realizar la derivada de una composición de funciones. Al realizar la *regla de la cadena* las derivadas de las composiciones deben multiplicarse a izquierda luego de una transposición.

Ejemplo. Considere $A \in \mathbb{R}^{n\times m}$, $x \in \mathbb{R}^m$, $b \in \mathbb{R}^n$ y $f : \mathbb{R}^m \to \mathbb{R}$ tal que $f(x) = \frac{1}{2}\|Ax - b\|^2$. Si se desea encontrar $\bar{x}$ solución de minimizar $f(x)$ para todo $x \in \mathbb{R}^m$, es necesario y suficiente que $f'(\bar{x}) = 0$. Para calcular esta derivada se escribe $f = h(g(x))$ con $h : \mathbb{R}^n \to \mathbb{R}$ t.q. $h(y) = \frac{1}{2}\|y\|^2$ y $g : \mathbb{R}^m \to \mathbb{R}^n$ t.q. $g(x) = Ax - b$. Usando la regla de la cadena estándar del Cálculo en varias variables obtenemos que

$$f'(x) = h'(g(x))g'(x) = g(x)A,$$

pues $h'(y) = y \in \mathbb{R}^n$ y $g'(x) = A \in \mathbb{R}^{n\times m}$. Para que este producto matricial tenga sentido, se debe considerar $g(x) \in \mathbb{R}^{1\times n}$ y por lo tanto no respeta la notación adoptada en este curso de considerar $\mathbb{R}^n$ como $\mathbb{R}^{n\times 1}$. Si no respetamos esa identificación obtendríamos

$$0 = f'(\bar{x}) = (\bar{x}^\top A^\top - b^\top)A = \bar{x}^\top \tilde{A}^\top - \tilde{b}^\top,$$

donde $\tilde{A} = A^\top A$ y $\tilde{b} = A^\top b$. Note que éste es un sistema lineal de vectores fila y el producto matriz vector se realiza a izquierda. Para ser coherente con la notación establecida, la regla de la cadena que se usará es

$$f'(x) = g'(x)^\top h'(g(x)) = A^\top(Ax - b) = \tilde{A}x - \tilde{b}.$$

De esta forma obtendremos sistemas lineales en el formato estándar de Álgebra Lineal.

Capítulo 1

Método de Newton

En 1669 Isaac Newton escribió su trabajo *De Analysi per Equationes Numero Terminorum Infinitas*, donde, entre otras cosas, describió un procedimiento iterativo para aproximar raíces reales de un polinomio de tercer grado. En 1690 Joseph Raphson propuso un procedimiento iterativo similar para resolver ecuaciones polinomiales más generales, dándole el crédito a Newton. En 1740 Thomas Simpson escribió el método tal como lo conocemos hoy, para ecuaciones no necesariamente polinomiales; él también propuso usar el método para resolver problemas de optimización encontrando un cero del gradiente.

1.1. Formulación del método

Sea $F : \mathbb{R}^n \to \mathbb{R}^n$ contínuamente diferenciable en $\mathbb{R}^n$ y se desea resolver el siguiente problema:

$$\text{hallar} \quad \bar{x} \in \mathbb{R}^n \quad \text{tal que} \quad F(\bar{x}) = 0.$$

Como F es diferenciable tenemos que $F(y) = F(x) + F'(x)(y-x) + o(\|y-x\|)$, por lo tanto para encontrar una raíz $\bar{x}$ de la función no lineal F resolveremos sucesivamente funciones lineales de la forma $y \mapsto F(x) + F'(x)(y-x)$.

Método de Newton. Dado $x^0 \in \mathbb{R}^n$ generar $\{x^k\}$ por el siguiente proceso:

$$\text{hallar} \quad x^{k+1} \in \mathbb{R}^n \quad \text{tal que} \quad F(x^k) + F'(x^k)(x^{k+1} - x^k) = 0.$$

Numéricamente puede escribirse de esta forma: dado $x^0 \in \mathbb{R}^n$ hacer $k = 0$;

1. calcular $F(x^k) \in \mathbb{R}^n$ y $F'(x^k) \in \mathbb{R}^{n \times n}$,
2. si $\|F(x^k)\| = 0$ PARAR,
3. hallar d^k tal que $F'(x^k)d^k = -F(x^k)$,
4. definir $x^{k+1} = x^k + d^k$, hacer $k = k+1$ y regresar al paso 1.

1.2. Convergencia del método

Veremos una demostración de la convergencia del método desde el punto del análisis variacional, pensando en perturbaciones de la función. Con el fin de presentar un resultado totalmente fundamentado, enunciaremos y demostraremos versiónes adaptadas del *Teorema de la aplicación contractiva* y del *Teorema de la Función Inversa* .

Teorema 1 (Aplicación Contractiva) *Sea $\bar{x} \in \mathbb{R}^n$ y $\Phi : \mathbb{R}^n \to \mathbb{R}^n$. Si existen $a > 0$ y $\lambda \in [0,1)$ tales que* *(i)* $\|\Phi(\bar{x}) - \bar{x}\| \leq a(1-\lambda)$*;*

(ii) $\|\Phi(y) - \Phi(x)\| \leq \lambda\|y - x\|$ *para todo* $x, y \in \mathcal{B}_a(\bar{x})$*.*

Entonces existe un único $x \in \mathcal{B}_a(\bar{x})$ *tal que* $\Phi(x) = x$*.*

Demostración. Veamos primero que Φ mapea elementos de $\mathcal{B}_a(\bar{x})$ en $\mathcal{B}_a(\bar{x})$. Sea $x \in \mathcal{B}_a(\bar{x})$ arbitrario, entonces

$$\begin{aligned} \|\Phi(x) - \bar{x}\| &\leq \|\Phi(x) - \Phi(\bar{x})\| + \|\Phi(\bar{x}) - \bar{x}\| \\ &\leq \lambda\|x - \bar{x}\| + a(1-\lambda) \\ &\leq a\lambda + a(1-\lambda) = a, \end{aligned}$$

donde usamos (i) y (ii) en la segunda desigualdad y el hecho de que $x \in \mathcal{B}_a(\bar{x})$ en la última desigualdad. Por lo tanto $\Phi(x) \in \mathcal{B}_a(\bar{x})$.

Veamos la existencia del punto fijo. Dado $x^0 \in \mathcal{B}_a(\bar{x})$, genero una sucesión $\{x^k\}$ tal que $x^{k+1} = \Phi(x^k) \in \mathcal{B}_a(\bar{x})$ para todo k. Tomo k, j arbitrarios con $k > j$, entonces

$$\begin{aligned} \|x^k - x^j\| &= \|\Phi(x^{k-1}) - \Phi(x^{j-1})\| \leq \lambda\|x^{k-1} - x^{j-1}\| \\ &\leq \lambda^2\|x^{k-2} - x^{j-2}\| \leq \ldots \leq \lambda^j\|x^{k-j} - x^0\| \\ &\leq \lambda^j(\|x^{k-j} - \bar{x}\| + \|\bar{x} - x^0\|) \leq 2a\lambda^j. \end{aligned}$$

Usando que $\lambda < 1$, obtenemos que $\lim_{k,j\to\infty} \|x^k - x^j\| = 0$. Como $\{x^k\}$ es una sucesión de Cauchy, entonces existe $x \in \mathcal{B}_a(\bar{x})$ tal que $x^k \to x$. Ahora, (ii) implica que Φ es contínua en $\mathcal{B}_a(\bar{x})$, por lo tanto

$$x = \lim_{k\to\infty} x^{k+1} = \lim_{k\to\infty} \Phi(x^k) = \Phi(x).$$

Para ver la unicidad, supongamos que existe $y \in \mathcal{B}_a(\bar{x})$ tal que $\Phi(y) = y$. Usando (ii), obtenemos que

$$\|y - x\|(1-\lambda) = \|\Phi(y) - \Phi(x)\| - \lambda\|y - x\| \leq 0.$$

Como $1 - \lambda > 0$, concluímos que $y = x$. ∎

Además del Teorema anterior, necesitaremos el siguiente resultado.

Proposición 2 *Si* $F : \mathbb{R}^n \to \mathbb{R}^n$ *es diferenciable en un entorno de* $\bar{x}$ *y su derivada* F' *es contínua en* $\bar{x}$*, entonces para cada* $\varepsilon > 0$ *existe* $\delta > 0$ *tal que*

$$\|F(y) - F(x) - F'(\bar{x})(y - x)\| \leq \varepsilon\|y - x\| \quad \textit{para todo} \quad x, y \in \mathcal{B}_\delta(\bar{x}). \tag{1.1}$$

Más aún, obtenemos que

$$\|F(y) - F(x) - F'(x)(y - x)\| \leq 2\varepsilon\|y - x\| \quad \textit{para todo} \quad x, y \in \mathcal{B}_\delta(\bar{x}).$$

Demostración. Fijo $\varepsilon > 0$. Como F' es contínua en $\bar{x}$ existe $\tilde{\delta} > 0$ tal que $\|F'(x) - F'(\bar{x})\| \leq \varepsilon$ para todo $x \in \mathcal{B}_{\tilde{\delta}}(\bar{x})$. Defino $\delta = \tilde{\delta}/3$ y tomo $x, y \in \mathcal{B}_\delta(\bar{x})$.

Usando la función $\phi(t) = F(x + t(y - x))$ tenemos

$$F(y) - F(x) = \phi(1) - \phi(0) = \int_0^1 \phi'(t)dt = \int_0^1 F'(x + t(y - x))(y - x)dt.$$

Por lo tanto

$$\begin{aligned} \|F(y) - F(x) - F'(\bar{x})(y - x)\| &= \left\|\int_0^1 (F'(x + t(y - x)) - F'(\bar{x}))(y - x)dt\right\| \\ &\leq \int_0^1 \|F'(x + t(y - x)) - F'(\bar{x})\| \, \|y - x\|dt \\ &\leq \sup_{t\in[0,1]} \|F'(x + t(y - x)) - F'(\bar{x})\|\|y - x\|. \end{aligned}$$

Para cualquier $t \in [0,1]$ tenemos que

$$\|x + t(y-x) - \bar{x}\| \le \|x - \bar{x}\| + \|y - x\| \le 2\|x - \bar{x}\| + \|y - \bar{x}\| \le 3\delta = \tilde{\delta}.$$

Por la contiunidad de F' obtenemos que $\sup_{t\in[0,1]} \|F'(x + t(y-x)) - F'(\bar{x})\| \le \varepsilon$, demostrando así (1.1).

Para demostrar la segunda afirmación, usamos que

$$\begin{aligned}
\|F(y) - F(x) - F'(x)(y-x)\| &\le \|F(y) - F(x) - F'(\bar{x})(y-x)\| \\
&\quad + \|(F'(\bar{x}) - F'(x))(y-x)\| \\
&\le \varepsilon\|y-x\| + \varepsilon\|y-x\| = 2\varepsilon\|y-x\|.
\end{aligned}$$

■

Teorema 3 (Función Inversa) *Sea $F : \mathbb{R}^n \to \mathbb{R}^n$ contínuamente diferenciable en un entorno de $\bar{x}$ y sea $\bar{p}$ tal que $F(\bar{x}) + \bar{p} = 0$. Si $F'(\bar{x})$ es no singular, entonces existen $a > 0$, $b > 0$ y una única función $x : \mathcal{B}_b(\bar{p}) \to \mathcal{B}_a(\bar{x})$ tal que $x(\bar{p}) = \bar{x}$ y $F(x(p)) + p = 0$ para todo $p \in \mathcal{B}_b(\bar{p})$. La función $p \mapsto x(p)$ es continua con*

$$\|x(q) - x(p)\| \le 2c\|q - p\| \quad \textit{para todo} \quad p, q \in \mathcal{B}_b(\bar{p}).$$

donde $c = \|F'(\bar{x})^{-1}\|$. Mas aún, es contínuamente diferenciable con

$$x'(p) = -F'(x(p))^{-1} \quad \textit{para todo} \quad p \in \mathcal{B}_b(\bar{p}).$$

Demostración. Sea $c = \|F'(\bar{x})^{-1}\|$. Usando (1.1) tenemos que existe $a > 0$ tal que

$$\|F(y) - F(x) - F'(\bar{x})(y-x)\| \le \frac{1}{4c}\|y - x\| \quad \text{para todo} \quad x, y \in \mathcal{B}_a(\bar{x}). \tag{1.2}$$

Defino $b = a/(4c)$. Fijo $p \in \mathcal{B}_b(\bar{p})$ y defino $\Phi_p : \mathbb{R}^n \to \mathbb{R}^n$ tal que

$$\Phi_p(x) = x - F'(\bar{x})^{-1}(F(x) + p).$$

Entonces

$$\|\Phi_p(\bar{x}) - \bar{x}\| = \| - F'(\bar{x})^{-1}(p - \bar{p})\| \le cb = \tfrac{a}{4} < a(1 - \tfrac{1}{4}).$$

Por otro lado, para cualquier $x, y \in \mathcal{B}_a(\bar{x})$ y usando (1.2) tenemos que

$$\begin{aligned}
\|\Phi_p(y) - \Phi_p(x)\| &= \|y - x - F'(\bar{x})^{-1}(F(y) - F(x))\| \\
&\le \|F'(\bar{x})^{-1}\|\|F(y) - F(x) - F'(\bar{x})(y-x)\| \\
&\le c\frac{1}{4c}\|y - x\| = \frac{1}{4}\|y - x\|.
\end{aligned}$$

Por lo tanto, valen las condiciones (i) y (ii) del Teorema 1 para $\lambda = 1/4$. Entonces existe un único $x(p) \in \mathcal{B}_a(\bar{x})$ tal que $\Phi_p(x(p)) = x(p)$, que por definición de Φ_p, equivale a $F(x(p)) + p = 0$.

Veamos que $p \mapsto x(p)$ es localmente Lipschitz contínua. Tome $p, q \in \mathcal{B}_b(\bar{p})$ y defina $x = x(p)$, $y = x(q)$. Como

$$x = -F'(x)^{-1}(F(x) + p - F'(x)x), \quad y = -F'(x)^{-1}(F(y) + q - F'(x)y),$$

entonces

$$y - x = -F'(x)^{-1}(F(y) - F(x) - F'(x)(y-x) + q - p). \tag{1.3}$$

Usando (1.2) y la segunda afirmación en Proposición 2, concluímos que

$$\|y - x\| \le c\|F(y) - F(x) - F'(x)(y-x)\| + c\|q - p\| \le \frac{1}{2}\|y - x\| + c\|q - p\|.$$

Entonces $\|y - x\| \leq 2c\|q - p\|$. Por lo tanto

$$\|x(q) - x(p)\| \leq 2c\|q - p\| \quad \text{para todo} \quad p, q \in \mathcal{B}_b(\bar{p}). \tag{1.4}$$

Para ver la diferenciabilidad, tomo $p \in \text{int}(\mathcal{B}_b(\bar{p}))$ y fijo $\varepsilon > 0$. Como F es diferenciable en $x = x(p)$, existe $\delta > 0$ tal que

$$\|F(y) - F(x) - F'(x)(y - x)\| \leq \frac{\varepsilon}{2c^2}\|y - x\| \quad \text{para todo} \quad y \in \mathcal{B}_\delta(x). \tag{1.5}$$

Tomo $\tau > 0$ suficientemente pequeño tal que $\tau < \delta/(2c)$ y $p + h \in \mathcal{B}_b(\bar{p})$ para todo $h \in \mathbb{R}^n$ con $\|h\| \leq \tau$. Por (1.4), tenemos que

$$\|x(p + h) - x(p)\| \leq 2c\|h\| \leq 2c\tau < \delta.$$

Como $F(x(p + h)) + p + h = 0$ y $F(x(p)) + p = 0$, usando (1.3) para $y = x(p + h)$, $x = x(p)$ y $q = p + h$, obtenemos

$$y - x + F'(x)^{-1}h = -F'(x)^{-1}(F(y) - F(x) - F'(x)(y - x)).$$

Usando (1.5), obtenemos que

$$\begin{aligned}
\|x(p + h) - x(p) + F'(x(p))^{-1}h\| &= \|y - x + F'(x)^{-1}h\| \\
&\leq c\|F(y) - F(x) - F'(x)(y - x)\| \\
&\leq \frac{\varepsilon}{2c}\|y - x\| = \frac{\varepsilon}{2c}\|x(p + h) - x(p)\| \\
&\leq \varepsilon\|h\|,
\end{aligned}$$

para todo $\|h\| \leq \tau$. Obteniendo así que $x'(p) = -F'(x(p))^{-1}$. ∎

Note que al momento de resolver $F(x) = 0$ computacionalmente, siempre estaremos sujetos a errores propios de la aritmética finita del computador, por lo tanto uno siempre termina resolviendo $F(x) + p = 0$ para algúna perturbación $p \in \mathbb{R}^n$. El resultado anterior nos dice que cerca de la solución $\bar{x}$ de la ecuación no perturbada ($\bar{p} = 0$), si existe $F'(\bar{x})^{-1}$ entonces

$$\|x(p) - \bar{x}\| \leq 2c\|p\|,$$

donde $c = \|F'(\bar{x})^{-1}\|$. O sea, la solución $x(p)$ del problema perturbado tiene un error con respecto a la solución real $\bar{x}$ del mismo orden de la perturbación p. Con esto, podemos demostrar la convergencia del método de Newton.

Teorema 4 (Convergencia local del método de Newton) *Considere $F : \mathbb{R}^n \to \mathbb{R}^n$ contínuamente diferenciable en un entorno de $\bar{x}$ con $F(\bar{x}) = 0$ y $F'(\bar{x})$ no singular. Entonces existe $\varepsilon > 0$ tal que para cualquier punto inicial $x^0 \in \mathcal{B}_\varepsilon(\bar{x})$, la suceción $\{x^k\}$ generada por el método de Newton converge a $\bar{x}$. Más aún, existe $\gamma_k \to 0^+$ tal que*

$$\|x^{k+1} - \bar{x}\| \leq \gamma_k\|x^k - \bar{x}\|.$$

Si además F' es Lipschitz continua en un entorno de $\bar{x}$, existe $\gamma > 0$ tal que

$$\|x^{k+1} - \bar{x}\| \leq \gamma\|x^k - \bar{x}\|^2.$$

Demostración. Sean $a > 0$ y $b > 0$ dados por el Teorema 3 y $c = \|F'(\bar{x})^{-1}\|$. Por Proposición 2 existe $\delta_0 > 0$ tal que

$$\|F(y) - F(x) - F'(x)(y - x)\| \leq \frac{1}{6c}\|y - x\| \quad \text{para todo} \quad x, y \in \mathcal{B}_{\delta_0}(\bar{x}). \tag{1.6}$$

Por otro lado, por contunuidad, existe $\delta_1 > 0$ tal que

$$\|F'(x)^{-1}(F(x) - F(\bar{x}) - F'(x)(x - \bar{x}))\| \leq \text{mín}\{\delta_0, 3bc\} \quad \text{para todo} \quad x \in \mathcal{B}_{\delta_1}(\bar{x}). \tag{1.7}$$

Tomo $\varepsilon \leq \text{mín}\{\delta_0, \delta_1, a, 3bc\}$ y $x^0 \in \mathcal{B}_\varepsilon(\bar{x})$. Veamos por inducción que $x^k \in \mathcal{B}_\varepsilon(\bar{x})$ para todo $k \geq 0$.

Supongamos que $x^k \in \mathcal{B}_\varepsilon(\bar{x})$. De acuerdo al método de Newton vale que $0 = F(x^k) + F'(x^k)(x^{k+1} - x^k)$, junto al hecho de que $F(\bar{x}) = 0$ se obtiene

$$x^{k+1} - \bar{x} = -F'(x^k)^{-1}\left(F(x^k) - F(\bar{x}) - F'(x^k)(x^k - \bar{x})\right).$$

Como $x^k \in \mathcal{B}_{\delta_1}(\bar{x})$, por (1.7) obtenemos que $\|x^{k+1} - \bar{x}\| \leq \text{mín}\{\delta_0, 3bc\}$. Usando (1.6) se concluye que

$$\begin{aligned} \|F(x^{k+1}) - F(x^k) - F'(x^k)(x^{k+1} - x^k)\| &\leq \frac{1}{6c}\|x^{k+1} - x^k\| \\ &\leq \frac{1}{6c}(\|x^{k+1} - \bar{x}\| + \|\bar{x} - x^k\|) \\ &\leq \frac{1}{6c}(3bc + 3bc) = b.. \end{aligned} \tag{1.8}$$

Defino $p^k = F(x^k) + F'(x^k)(x^{k+1} - x^k) - F(x^{k+1})$. Entonces $F(x^{k+1}) + p^k = 0$ con $p^k \in \mathcal{B}_b(0)$ (por (1.9)). Por Teorema 3 para $\bar{p} = 0$ tenemos que $x^{k+1} = x(p^k)$ con $x(\bar{p}) = \bar{x}$ y por lo tanto

$$\|x^{k+1} - \bar{x}\| \leq 2c\|p^k\| \leq \frac{1}{3}\|x^{k+1} - x^k\|,$$

donde en la última relación usamos (1.8). De la desigualdad $\|x^{k+1} - x^k\| \leq \|x^{k+1} - \bar{x}\| + \|\bar{x} - x^k\| \leq \frac{1}{3}\|x^{k+1} - x^k\| + \|\bar{x} - x^k\|$ se obtiene que

$$\|x^{k+1} - x^k\| \leq \frac{3}{2}\|x^k - \bar{x}\|. \tag{1.9}$$

Por lo tanto

$$\|x^{k+1} - \bar{x}\| \leq \frac{1}{3}\|x^{k+1} - x^k\| \leq \frac{1}{2}\|x^k - \bar{x}\| \leq \frac{\varepsilon}{2}.$$

Demostrando así que $x^{k+1} \in \mathcal{B}_\varepsilon(\bar{x})$. Como el proceso anterior vale para todo $k \geq 0$, se tiene que

$$\|x^{k+1} - \bar{x}\| \leq \frac{1}{2}\|x^k - \bar{x}\| \leq \frac{1}{2^2}\|x^{k-1} - \bar{x}\| \leq \ldots \leq \frac{1}{2^{k+1}}\|x^0 - \bar{x}\|.$$

Por lo tanto $\{x^k\}$ converge a $\bar{x}$.

Como F es diferenciable, existe $t_k \in [0, 1]$ tal que

$$F(x^{k+1}) - F(x^k) = F'(x^k + t_k(x^{k+1} - x^k))(x^{k+1} - x^k).$$

Defino $\gamma_k = 3c\|F'(x^k + t_k(x^{k+1} - x^k)) - F'(x^k)\|$, como $\{x^k\}$ converge a $\bar{x}$ obtenemos que $\{c_k\}$ converge a 0. Por lo tanto

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &\leq 2c\|p^k\| \\ &\leq 2c\|F(x^{k+1}) - F(x^k) - F'(x^k)(x^{k+1} - x^k)\| \\ &\leq \frac{2}{3}\gamma_k\|x^{k+1} - x^k\| \\ &\leq \gamma_k\|x^k - \bar{x}\|, \end{aligned}$$

donde en la última desigualdad usamos (1.9).

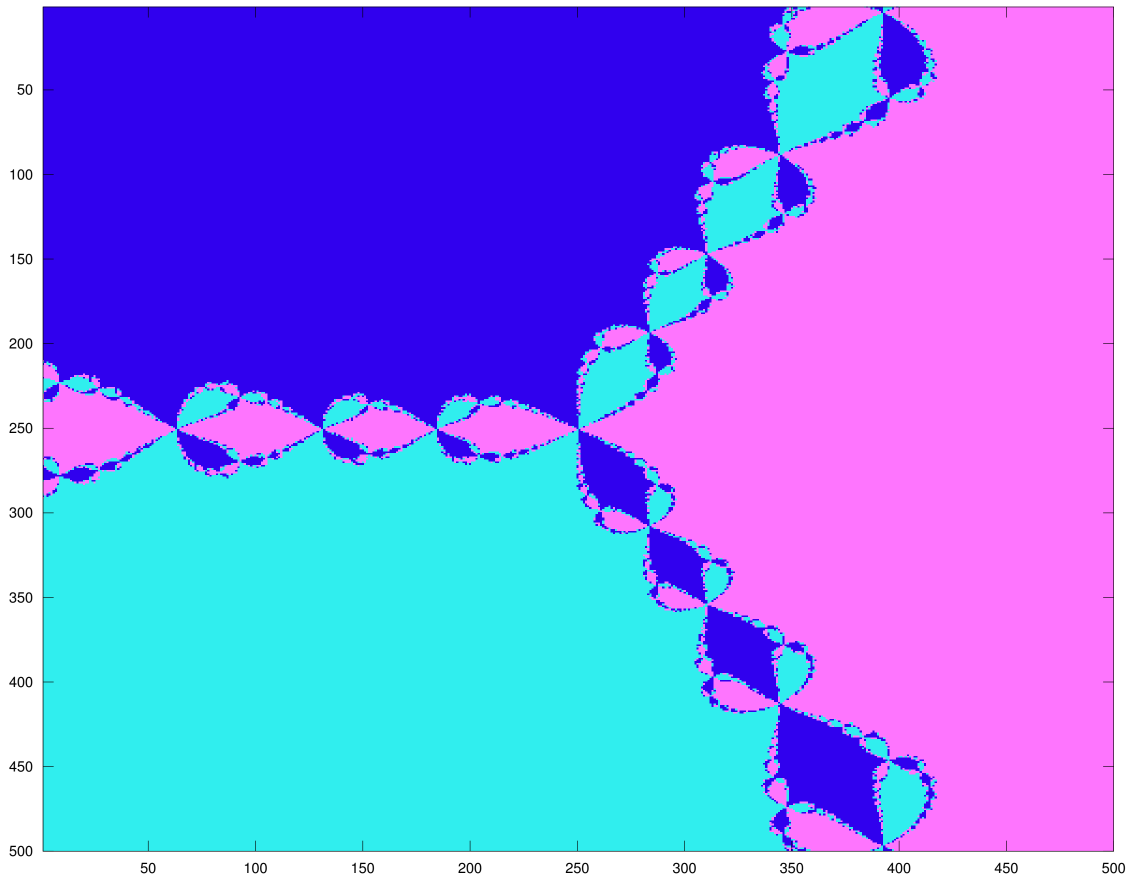


Figura 1.1: Cuencas de atracción de $F(x, y) = (x_1^3 - 3x_1x_2^2 - 1, 3x_1^2x_2 - x_2^3)$

Si Además F' es Lipschitz continua en un entorno de $\bar{x}$, obtenemos que

$$\begin{aligned}\gamma_k &= 3c\|F'(x^k + t_k(x^{k+1} - x^k)) - F'(x^k)\| \\ &\leq 3cLt_k\|x^{k+1} - x^k\| \leq 3cL\|x^{k+1} - x^k\|.\end{aligned}$$

Por lo tanto

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &\leq \frac{2}{3}\gamma_k\|x^{k+1} - x^k\| \\ &\leq 2cL\|x^{k+1} - x^k\|^2 \\ &\leq \gamma\|x^k - \bar{x}\|^2,\end{aligned}$$

donde en la última desigualdad usamos (1.9) y tomamos $\gamma = 9c/2$. ∎

Ejemplo 1 *Se desea encontrar y y x tales que $(x + iy)^3 - 1 = 0$. Desarrollando la potencia y separando parte real e imaginaria, tenemos que hallar un cero de $F(x, y) = (x_1^3 - 3x_1x_2^2 - 1, 3x_1^2x_2 - x_2^3)$. Las cuencas de atracción del método de Newton se observan en la figura 1.1.*

Capítulo 2

Optimización

Suponga que tenemos un conjunto $\mathcal{D} \subset \mathbb{R}^n$, una función $f : \mathcal{D} \to \mathbb{R}$ y se desea resolver el problema de

$$\begin{array}{rl} \text{minimizar} & f(x) \\ \text{sujeto a} & x \in \mathcal{D}. \end{array} \tag{2.1}$$

Definición 1 *Diremos que $\mathcal{D}$ es el conjunto factible de* (2.1).

Diremos que $\bar{v} \in [-\infty, \infty]$ es el valor óptimo *de* (2.1) *si $\bar{v} = \inf_{x \in \mathcal{D}} f(x)$.*

Diremos que $\bar{x} \in \mathcal{D}$ es:

- minimizador global *de* (2.1) *si $f(\bar{x}) \leq f(x)$ para todo $x \in \mathcal{D}$.*
- minimizador local *de* (2.1) *si existe $\delta > 0$ tal que $f(\bar{x}) \leq f(x)$ para todo $x \in \mathcal{D} \cap \mathcal{B}_\delta(\bar{x})$.*

Ejemplo 2 *Sea $f(x) = e^x$. Si $\mathcal{D} = \mathbb{R}$, entonces $\bar{v} = 0$ y no existen minimizadores. Si tomamos $\mathcal{D} = [0, +\infty)$, entonces $\bar{v} = 1$ y $\bar{x} = 0$ es minimizador global.*

2.1. Minimizadores globales

Veamos algunas condiciones que garantizan la existencia de un minimizador global.

Teorema 5 (Teorema de Weierstrass) *Si $\mathcal{D}$ es compacto no vacío y $f : \mathcal{D} \to \mathbb{R}$ es continua, entonces existe $\bar{x} \in \mathcal{D}$ minimizador global de* (2.1).

Corolario 6 *Si $f : \mathcal{D} \to \mathbb{R}$ es continua y existe $c \in \mathbb{R}$ tal que el conjunto $\{x \in \mathcal{D} \mid f(x) \leq c\}$ es compacto no vacío, entonces existe $\bar{x} \in \mathcal{D}$ minimizador global de* (2.1).

Ejemplo 3 *Considere $\mathcal{D} = (0, 1)$ y $f(x) = \frac{1}{x} + \frac{1}{1-x}$. Como se ve en la Figura 2.1, para cualquier $c \geq 4$ el conjunto $\{x \in \mathcal{D} \mid f(x) \leq c\}$ es compacto no vacío.*

Ejemplo 4 *Considere $\mathcal{D} = \{x \in \mathbb{R}^2 \mid x_1 + x_2 > 0\}$ y $f(x) = x_1^2 + x_2 + \frac{1}{x_1+x_2}$. Como se ve en la Figura 2.2, para cualquier $c \geq 0$ el conjunto $\{x \in \mathcal{D} \mid f(x) \leq c\}$ es compacto no vacío.*

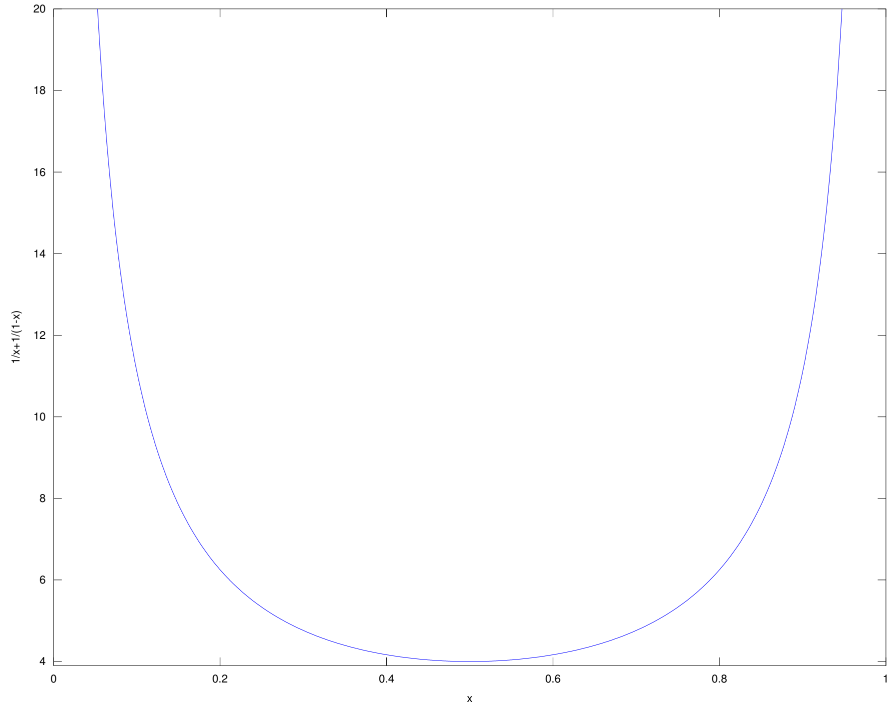


Figura 2.1: Gráfico de $\frac{1}{x} + \frac{1}{1-x}$

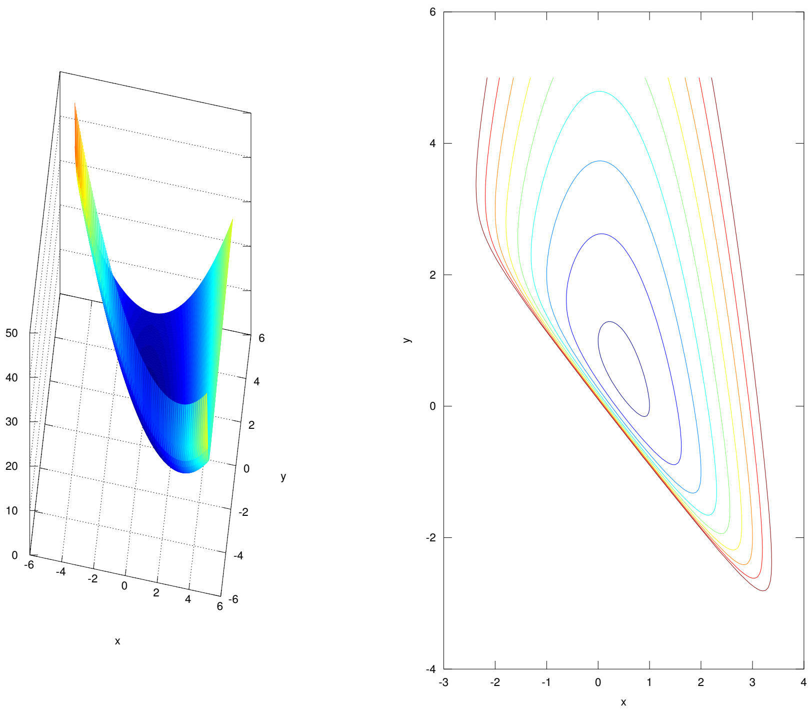


Figura 2.2: Gráfico de $x_1^2 + x_2 + \frac{1}{x_1+x_2}$ y curvas de nivel

2.2. Optimización sin restricciones

Consideremos que el conjunto de restricciones es todo el espacio,i.e., $\mathcal{D} = \mathbb{R}^n$. Por lo tanto el problema a resolver es

$$\begin{array}{rl} \text{minimizar} & f(x) \\ \text{sujeto a} & x \in \mathbb{R}^n. \end{array} \tag{2.2}$$

Ejemplo 5 ***Cartera de acciones.*** *Para una cartera de n acciones, si invierto x_i pesos en la empresa i, el retorno será*

$$\sum_{i=1}^{n} r_i x_i,$$

donde $r_i = r_i(\omega)$ variable aleatoria. Para manejar la aleatoriedad, trabajaremos con el retorno esperado

$$\mathbb{E}\left(\sum_{i=1}^{n} r_i x_i\right) = \sum_{i=1}^{n} \mathbb{E}(r_i) x_i.$$

Además, mediremos el riesgo de la inversión con su varianza

$$\mathbb{V}\left(\sum_{i=1}^{n} r_i x_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij} x_i x_j,$$

donde σ_{ij} es la covarianza entre las variables aleatorias r_i y r_j. Con esto, el problema que deseamos resolver es

$$\underset{x\in\mathbb{R}^n}{\text{minimizar}} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij} x_i x_j - \alpha \sum_{i=1}^{n} \mathbb{E}(r_i) x_i,$$

para $\alpha > 0$. Note que si $\alpha \approx 0$ importa el riesgo y si $\alpha \approx \infty$ importa el retorno. Para poder resolver este problema, tomaremos m muestras $r_i(\omega_l)$ con $l = 1, \ldots, m$ para cada variable aleatoria con $i = 1, \ldots, n$ y aproximaremos

$$\mathbb{E}(r_i) \approx \mu_i = \frac{1}{m}\sum_{l=1}^{m} r_i(\omega_l) \quad y \quad \sigma_{ij} \approx s_{ij} = \frac{1}{m-1}\sum_{l=1}^{m} (r_i(\omega_l) - \mu_i)(r_j(\omega_l) - \mu_j).$$

Por lo tanto el problema que resolveremos es

$$\underset{x\in\mathbb{R}^n}{\text{minimizar}} \quad \frac{1}{2}\langle Sx, x\rangle - \alpha\langle \mu, x\rangle.$$

Para poder hallar el minimizador, tenemos que saber las condiciones que deben cumplirse.

Teorema 7 (Condiciones necesarias de optimalidad) *Sea $f : \mathbb{R}^n \to \mathbb{R}$ y $\bar{x}$ minimizador local de f en $\mathbb{R}^n$.*

1. *Si f es diferenciable en $\bar{x}$, entonces $f'(\bar{x}) = 0$.*
2. *Si f es dos veces diferenciable en $\bar{x}$, entonces $f'(\bar{x}) = 0$ y $f''(\bar{x})$ es semidefinida positiva.*

Demostración. 1. Tomo $d = -f'(\bar{x})$ y t suficientemente pequeño tal que $f(\bar{x}) \leq f(\bar{x} + td)$, como

$$f(\bar{x} + td) = f(\bar{x}) + t\langle f'(\bar{x}), d\rangle + o(t),$$

entonces dividiendo por $t > 0$ se obtiene $0 \leq \langle f'(\bar{x}), d\rangle + o(t)/t$. Tomando límite para $t \to 0^+$,

$$0 \leq \langle f'(\bar{x}), d\rangle = -\|f'(\bar{x})\| \leq 0.$$

Por lo tanto $f'(\bar{x}) = 0$.

2. Tomo $d \in \mathbb{R}^n$ y $t > 0$ suficientemente pequeño. Como

$$f(\bar{x}) \leq f(\bar{x} + td) = f(\bar{x}) + t\langle f'(\bar{x}), d\rangle + \frac{1}{2}t^2\langle f''(\bar{x})d, d\rangle + o(t^2),$$

usando que $f'(\bar{x}) = 0$ y dividiendo por $t^2 > 0$, obtenemos que $0 \leq \langle f''(\bar{x})d, d\rangle + o(t^2)/t^2$. Tomando límite para $t \to 0^+$,

$$0 \leq \langle f''(\bar{x})d, d\rangle,$$

donde d fue elegido arbitrario. ■

Definición 2 *Decimos que $\bar{x} \in \mathbb{R}^n$ es un punto estacionario del problema* (2.2) *si $f'(\bar{x}) = 0$.*

En la literatura, la ecuación $f'(\bar{x}) = 0$ se conoce como *condición de Fermat.*

2.2.1. Usando el método de Newton.

Para hallar $\bar{x}$ tal que $f'(\bar{x}) = 0$, el método de Newton genera $\{x^k\}$ tal que

$$0 = f'(x^k) + f''(x^k)(x^{k+1} - x^k).$$

Desde el punto de vista de optimización, podemos decir que en la k-ésima iteración el método de Newton halla x^{k+1} punto estacionario del problema

$$\underset{x\in\mathbb{R}^n}{\text{minimizar}} \;\; f(x^k) + \langle f'(x^k), x - x^k\rangle + \frac{1}{2}\langle f''(x^k)(x - x^k), x - x^k\rangle.$$

Para garantizar la convergencia del método de Newton a $\bar{x}$ tal que $f'(\bar{x}) = 0$, debemos tener $f''(\bar{x})$ no singular, lo que es equivalente a decir que $x = \bar{x}$ es la única solución de $0 = f'(\bar{x}) + f''(\bar{x})(x - \bar{x})$. Por el teorema anterior, esto equivale a decir que $x = \bar{x}$ es la única solución del problema

$$\underset{x\in\mathbb{R}^n}{\text{minimizar}} \;\; f(\bar{x}) + \langle f'(\bar{x}), x - \bar{x}\rangle + \frac{1}{2}\langle f''(\bar{x})(x - \bar{x}), x - \bar{x}\rangle.$$

Como $f'(\bar{x}) = 0$ y el valor optimal de este problema es $f(\bar{x})$, debemos tener

$$\langle f''(\bar{x})(x - \bar{x}), x - \bar{x}\rangle > 0 \quad \text{para todo} \quad x \neq \bar{x}.$$

Veamos que esta condición nos garantiza que en un entorno de $\bar{x}$, f está acotada inferiormente por un paraboloide.

Teorema 8 (Condición suficiente de optimalidad) *Si f es dos veces diferenciable en $\bar{x}$ con $f'(\bar{x}) = 0$ y $f''(\bar{x})$ definida positiva, entonces $\bar{x}$ es un minimizador local de* (2.2). *Más aún, existen $\gamma > 0$ y $\varepsilon > 0$ tales que*

$$f(x) \geq f(\bar{x}) + \frac{\gamma}{2}\|x - \bar{x}\|^2 \quad \textit{para todo} \quad x \in \mathcal{B}_\varepsilon(\bar{x}). \tag{2.3}$$

Demostración. Supongamos que no vale (2.3), o sea, podemos generar $\{x^k\}$ tal que $\|x^k - \bar{x}\| \leq \frac{1}{k}$ y

$$f(x^k) < f(\bar{x}) + \frac{1}{2k}\|x^k - \bar{x}\|^2.$$

Entonces

$$\begin{aligned}
\frac{1}{2k}\|x^k - \bar{x}\| &> f(x^k) - f(\bar{x}) \\
&= \langle f'(\bar{x}), x^k - \bar{x}\rangle + \frac{1}{2}\langle f''(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|^2) \\
&= \frac{1}{2}\langle f''(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|^2). \qquad (2.4)
\end{aligned}$$

Defino $d^k = \frac{x^k - \bar{x}}{\|x^k - \bar{x}\|}$. Como $\|d^k\| = 1$, entonces existe una subsucesión $\{d^{k_j}\}$ convergente a d con $\|d\| = 1$. De (2.4), obtenemos que

$$\frac{1}{k_j} > \langle f''(\bar{x})d^{k_j}, d^{k_j}\rangle + 2\frac{o(\|x^{k_j} - \bar{x}\|^2)}{\|x^{k_j} - \bar{x}\|^2}.$$

Tomando límite, deducimos que

$$0 \geq \langle f''(\bar{x})d, d\rangle,$$

con $d \neq 0$, contradiciendo el hecho de que $f''(\bar{x})$ es definida positiva. ■

2.3. Optimización con restricciones abstractas

Consideremos un conjunto abstracto $\mathcal{D} \subset \mathbb{R}^n$ y el problema 2.1, o sea

$$\begin{array}{rl} \text{minimizar} & f(x) \\ \text{sujeto a} & x \in \mathcal{D}. \end{array} \tag{2.5}$$

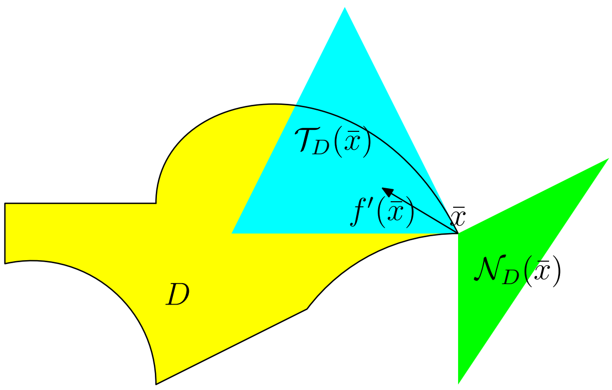
Vamos a aproximar el conjunto $\mathcal{D}$ localmente por un cono, o sea, un conjunto $\mathcal{K}$ tal que si $d \in \mathcal{K}$ entonces $\alpha d \in \mathcal{K}$ para todo $\alpha \geq 0$.

Definición 3 *Para $\bar{x} \in \mathcal{D}$, el cono tangente a $\mathcal{D}$ en $\bar{x}$ es*

$$\mathcal{T}_{\mathcal{D}}(\bar{x}) = \left\{ d \in \mathbb{R}^n \,\middle|\, \begin{array}{c} \exists \alpha_k \to +\infty,\ \{x^k\} \subset \mathcal{D} \ \textit{tales que} \\ x^k \to \bar{x} \quad y \quad \alpha_k(x^k - \bar{x}) \to d \end{array} \right\},$$

el cono normal a $\mathcal{D}$ en $\bar{x}$ es

$$\mathcal{N}_{\mathcal{D}}(\bar{x}) = \left\{ v \in \mathbb{R}^n \,\middle|\, \begin{array}{c} \forall \{x^k\} \subset \mathcal{D} \setminus \{\bar{x}\} \quad \textit{con} \quad x^k \to \bar{x} \\ \textit{vale que} \quad \limsup_{k \to +\infty} \frac{\langle v, x^k - \bar{x}\rangle}{\|x^k - \bar{x}\|} \leq 0 \end{array} \right\}$$



Proposición 9 *Para cualquier $\bar{x} \in \mathcal{D}$, vale que*

$$\mathcal{N}_{\mathcal{D}}(\bar{x}) = \{v \in \mathbb{R}^n \mid \langle v, d\rangle \leq 0 \ \forall\, d \in \mathcal{T}_{\mathcal{D}}\}.$$

Demostración. Tomo $v \in \mathbb{R}^n$ tal que $\langle v, d\rangle \leq 0$ para todo $d \in \mathcal{T}_{\mathcal{D}}$. Para cualquier $\{x^k\} \subset \mathcal{D} \setminus \{\bar{x}\}$ con $x^k \to \bar{x}$, defino $\alpha_k = 1/\|x^k - \bar{x}\|$. Tomando subsucesiones si fuera necesario obtengo $\alpha_k(x^k - \bar{x}) \to d$. Por lo tanto $d \in \mathcal{T}_{\mathcal{D}}(\bar{x})$. Entonces

$$\limsup_{k \to +\infty} \frac{\langle v, x^k - \bar{x}\rangle}{\|x^k - \bar{x}\|} = \limsup_{k \to +\infty} \langle v, \alpha_k(x^k - \bar{x})\rangle = \langle v, d\rangle \leq 0.$$

Por lo tanto $v \in \mathcal{N}_{\mathcal{D}}(\bar{x})$.

Por otro lado, si $v \in \mathcal{N}_{\mathcal{D}}(\bar{x})$, tomo $d \in \mathcal{T}_{\mathcal{D}}(\bar{x})$ arbitrario. Entonces existen $\alpha_k \to 0^+$, $\{x^k\} \subset \mathcal{D}$ tales que $x^k \to \bar{x}$ y $\alpha_k(x^k - \bar{x}) \to d$. Por definición de $\mathcal{N}_{\mathcal{D}}(\bar{x})$ tenemos

$$\begin{aligned} 0 &\geq \limsup_{k\to+\infty} \frac{\langle v, x^k - \bar{x}\rangle}{\|x^k - \bar{x}\|} \\ &= \limsup_{k\to+\infty} \frac{\langle v, \alpha_k(x^k - \bar{x})\rangle}{\|\alpha_k(x^k - \bar{x})\|} \\ &= \frac{\langle v, d\rangle}{\|d\|}. \end{aligned}$$

Concluyendo que $\langle v, d\rangle \leq 0 \; \forall \, d \in \mathcal{T}_{\mathcal{D}}(\bar{x})$. ■

Ejercicio 1 *Sea* $\mathcal{D} = \{x \in \mathbb{R}^2 \mid x_2 \geq \sqrt{|x_1|}\}$ *y* $\bar{x} = (0,0)$*. Demuestre que* $\mathcal{T}_{\mathcal{D}}(\bar{x}) = \{d \in \mathbb{R}^2 \mid d_1 = 0, \, d_2 \geq 0\}$ *y* $\mathcal{N}_{\mathcal{D}}(\bar{x}) = \{v \in \mathbb{R}^2 \mid v_2 \leq 0\}$*.*

Ejercicio 2 *Sea* $\mathcal{D} = \{x \in \mathbb{R}^2 \mid x_2 \leq \sqrt{|x_1|}\}$ *y* $\bar{x} = (0,0)$*. Demuestre que* $\mathcal{T}_{\mathcal{D}}(\bar{x}) = \mathbb{R}^2$ *y* $\mathcal{N}_{\mathcal{D}}(\bar{x}) = \{0\}$*.*

Ejercicio 3 *Demuestre que si* $\bar{x} \in \operatorname{int}(\mathcal{D})$*, entonces* $\mathcal{T}_{\mathcal{D}}(\bar{x}) = \mathbb{R}^n$ *y* $\mathcal{N}_{\mathcal{D}}(\bar{x}) = \{0\}$*.*

Proposición 10 *Sea* $\bar{x} \in \mathcal{D}$ *con* $\mathcal{D} = \{x \in \mathbb{R}^n \mid Ax = b\}$ *donde* $A \in \mathbb{R}^{m\times n}$ *y* $b \in \mathbb{R}^m$*. Entonces* $\mathcal{T}_{\mathcal{D}}(\bar{x}) = \operatorname{Ker}(A)$ *y* $\mathcal{N}_{\mathcal{D}}(\bar{x}) = \operatorname{Im}(A^\top)$*.*

Demostración. Comencemos con el cono tangente. Tomo $d \in \operatorname{Ker}(A)$, para cualquier $\alpha_k \to +\infty$ defino $x^k = \bar{x} + \frac{1}{\alpha_k} d$. Como

$$Ax^k = A\bar{x} + \frac{1}{\alpha_k} Ad = b + 0 = b,$$

entonces $\{x^k\} \subset \mathcal{D}$ con $x^k \to \bar{x}$ y $\alpha_k(x^k - \bar{x}) = d$. Por lo tanto $d \in \mathcal{T}_{\mathcal{D}}(\bar{x})$.

Si $d \in \mathcal{T}_{\mathcal{D}}(\bar{x})$, existen $\alpha_k \to +\infty$ y $\{x^k\} \subset \mathcal{D}$ con $x^k \to \bar{x}$ y $\alpha_k(x^k - \bar{x}) \to d$. Como

$$A\left(\alpha_k(x^k - \bar{x})\right) = \alpha_k(Ax^k - A\bar{x}) = \alpha_k(b - b) = 0,$$

tomando límites obtenemos que $Ad = 0$. Por lo tanto $d \in \operatorname{Ker}(A)$. Demostrando así que $\mathcal{T}_{\mathcal{D}}(\bar{x}) = \operatorname{Ker}(A)$.

Veamos el cono normal. Tomo $v \in \operatorname{Im}(A^\top)$, o sea, existe $\lambda \in \mathbb{R}^m$ tal que $v = A^\top \lambda$. Tomo $\{x^k\} \subset \mathcal{D} \setminus \{\bar{x}\}$ arbitrario tal que $x^k \to \bar{x}$, entonces

$$\langle v, x^k - \bar{x}\rangle = \langle A^\top \lambda, x^k - \bar{x}\rangle = \langle \lambda, Ax^k - A\bar{x}\rangle = \langle \lambda, b - b\rangle = 0.$$

Tomando $\limsup$ para $k \to +\infty$ obtenemos que $v \in \mathcal{N}_{\mathcal{D}}(\bar{x})$.

Si $v \in \mathcal{N}_{\mathcal{D}}(\bar{x})$, tomo $d \in \operatorname{Ker}(A)$ arbitrario con $d \neq 0$ y $t_k \to 0^+$. Defino $x^k = \bar{x} + t_k d$ y como $Ax^k = b$, obtenemos que $\{x^k\} \subset \mathcal{D} \setminus \{\bar{x}\}$ con $x^k \to \bar{x}$. Usando que

$$0 \geq \limsup_{k\to+\infty} \frac{\langle v, x^k - \bar{x}\rangle}{\|x^k - \bar{x}\|} = \frac{\langle v, d\rangle}{\|d\|},$$

obtenemos que $\langle v, d\rangle \leq 0$ para todo $d \in \operatorname{Ker}(A)$, y como $-d \in \operatorname{Ker}(A)$ también tendremos $-\langle v, d\rangle \leq 0$. Por lo tanto $\langle v, d\rangle = 0$ para todo $d \in \operatorname{Ker}(A)$. Entonces $v \in \operatorname{Ker}(A)^\perp = (A^\top)$. Demostrando así que $\mathcal{N}_{\mathcal{D}}(\bar{x}) = \operatorname{Im}(A^\top)$. ■

Definición 4 *Decimos que* $\mathcal{D}$ *es convexo si dados* $x, y \in \mathcal{D}$ *vale* $(1-\alpha)x + \alpha y \in \mathcal{D}$ *para todo* $\alpha \in [0,1]$*.*

Proposición 11 *Si $\mathcal{D} \subset \mathbb{R}^n$ es convexo y $\bar{x} \in \mathcal{D}$, entonces*

$$\mathcal{N}_{\mathcal{D}}(\bar{x}) = \{v \in \mathbb{R}^n \mid \langle v, x - \bar{x}\rangle \leq 0 \,\forall\, x \in \mathcal{D}\}.$$

Demostración. Sea $v \in \mathcal{N}_{\mathcal{D}}(\bar{x})$. Tomo $x \in \mathcal{D}$ con $x \neq \bar{x}$, $t_k \to +\infty$ y defino $x^k = \bar{x} + t_k(x - \bar{x})$. Como $t_k \in (0, 1]$ para k suficientemente grande, por la convexidad de $\mathcal{D}$ obtenemos $x^k = (1 - t_k)\bar{x} + t_k x \in \mathcal{D}$. Entonces $\{x^k\} \subset \mathcal{D} \setminus \{\bar{x}\}$ con $x^k \to \bar{x}$, por lo tanto

$$0 \geq \limsup_{k\to+\infty} \frac{\langle v, x^k - \bar{x}\rangle}{\|x^k - \bar{x}\|} = \frac{\langle v, x - \bar{x}\rangle}{\|x - \bar{x}\|},$$

obteniendo que $\langle v, x - \bar{x}\rangle \leq 0$ para todo $x \in \mathcal{D}$.

Si $\langle v, x - \bar{x}\rangle \leq 0$ para todo $x \in \mathcal{D}$, para cualquier sucesión arbitraria $\{x^k\} \subset \mathcal{D} \setminus \{\bar{x}\}$ con $x^k \to \bar{x}$ tendremos $\langle v, x^k - \bar{x}\rangle \leq 0$. Por lo tanto

$$\limsup_{k\to+\infty} \frac{\langle v, x^k - \bar{x}\rangle}{\|x^k - \bar{x}\|} \leq 0.$$

Demostrando así que $\mathcal{N}_{\mathcal{D}}(\bar{x}) = \{v \in \mathbb{R}^n \mid \langle v, x - \bar{x}\rangle \leq 0 \,\forall\, x \in \mathcal{D}\}$. ■

Teorema 12 (Condición necesaria de optimalidad) *Sea $\mathcal{D} \subset \mathbb{R}^n$ y $f : \mathbb{R}^n \to \mathbb{R}$ diferenciable en $\bar{x} \in \mathcal{D}$. Si $\bar{x}$ es minimizador local del problema* (2.1)*, entonces*

$$-f'(\bar{x}) \in \mathcal{N}_{\mathcal{D}}(\bar{x}).$$

Demostración. Tomo $\{x^k\} \subset \mathcal{D} \setminus \{\bar{x}\}$ arbitraria tal que $x^k \to \bar{x}$. Como $f(\bar{x}) \leq f(x^k)$ para k suficientemente grande y

$$f(x^k) = f(\bar{x}) + \langle f'(\bar{x}), x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|),$$

obtenemos que

$$\frac{\langle -f'(\bar{x}), x^k - \bar{x}\rangle}{\|x^k - \bar{x}\|} \leq \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\|}.$$

Por lo tanto

$$\limsup_{k\to+\infty} \frac{\langle -f'(\bar{x}), x^k - \bar{x}\rangle}{\|x^k - \bar{x}\|} \leq 0.$$

Concluyendo que $-f'(\bar{x}) \in \mathcal{N}_{\mathcal{D}}(\bar{x})$. ■

Definición 5 *Decimos que $\bar{x} \in \mathcal{D}$ es un punto estacionario del problema* (2.1) *si $-f'(\bar{x}) \in \mathcal{N}_{\mathcal{D}}(\bar{x})$.*

Note que para el caso $\mathcal{D} = \mathbb{R}^n$ tendremos $\bar{x} \in \mathrm{int}(\mathcal{D})$ y por lo tanto $\mathcal{N}_{\mathcal{D}}(\bar{x}) = \{0\}$, obteniendo en este caso que $\bar{x}$ es estacionario si $f'(\bar{x}) = 0$.

2.4. Optimización con restricciones de igualdad

Consideremos un caso particular del problema (2.1) con $\mathcal{D} = \{x \in \mathbb{R}^n \mid h(x) = 0\}$ donde $h : \mathbb{R}^n \to \mathbb{R}^m$, o sea

$$\begin{array}{rl} \text{minimizar} & f(x) \\ \text{sujeto a} & h(x) = 0. \end{array} \tag{2.6}$$

Ejemplo 6 *El Ejemplo 5 solo tiene sentido si disponemos de dinero infinito para realizar las inversiones, algo bastante alejado de la realidad. Por lo general, al momento de realizar las inversiones uno cuenta solo con un cierto capital* κ *y por lo tanto uno debe buscar una solución donde el dinero a invertir sea igual al capital disponible, o sea*

$$\kappa = \sum_{i=1}^{n} x_i = e^\top x,$$

donde $e = (1, \dots, 1) \in \mathbb{R}^n$. *En este caso, el problema a resolver es*

$$\begin{array}{rl} \text{minimizar} & \frac{1}{2}\langle Sx, x\rangle - \alpha\langle \mu, x\rangle \\ \text{sujeto a} & e^\top x = \kappa. \end{array}$$

Los problemas con este tipo de estructura son denominados problemas cuadráticos (QP, por sus siglas en inglés) y pueden resolverse usando la rutina `qp` de `Octave`.

Supongamos por un momento que las restricciones del problema (2.6) son afines, o sea, $h(x) = Ax - b$ con $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^m$. En este caso, un punto estacionario $\bar{x}$ del problema (2.6) debe cumplir $\bar{x} \in \mathcal{D}$ y

$$-f'(\bar{x}) \in \mathcal{N}_{\mathcal{D}}(\bar{x}) = \mathrm{Im}(A^\top).$$

Por lo tanto, existe $\bar{\lambda} \in \mathbb{R}^m$ tal que $0 = f'(\bar{x}) + A^\top\bar{\lambda}$. En otras palabras, $(\bar{x}, \bar{\lambda})$ es solución del sistema no lineal

$$0 = \begin{bmatrix} f'(x) + A^\top\lambda \\ Ax - b \end{bmatrix} = \begin{bmatrix} f'(x) + h'(x)^\top\lambda \\ h(x) \end{bmatrix}.$$

Recordemos que como $h : \mathbb{R}^n \to \mathbb{R}^m$ entonces $h'(x) \in \mathbb{R}^{m\times n}$ con

$$h'(x) = \begin{bmatrix} h_1'(x)^\top \\ \vdots \\ h_m'(x)^\top \end{bmatrix}, \quad h_i'(x) \in \mathbb{R}^n.$$

Definición 6 *La función Lagrangiana asociada al problema* (2.6) *es* $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *tal que*

$$L(x, \lambda) = f(x) + \langle \lambda, h(x)\rangle.$$

El conjunto de mutiplicadores de Lagrange asociado a $\bar{x}$ *factible para el problema* (2.6) *es*

$$\mathcal{M}(\bar{x}) = \{\lambda \in \mathbb{R}^m \mid 0 = f'(\bar{x}) + h'(\bar{x})^\top\lambda\} = \left\{\lambda \in \mathbb{R}^m \,\middle|\, 0 = \frac{\partial L}{\partial x}(\bar{x}, \lambda)\right\}.$$

Los multiplicadores de Lagrange son importantes al momento de resolver (2.6), pues nos permite encontrar puntos estacionarios mediante la resolución de un sistema de ecuaciones no lineales, sin necesidad de calcular el cono normal $\mathcal{N}_{\mathcal{D}}$.

Teorema 13 *Sea* $f : \mathbb{R}^n \to \mathbb{R}$ *y* $\mathcal{D} = \{x \in \mathbb{R}^n \mid h(x) = 0\}$ *con* $h : \mathbb{R}^n \to \mathbb{R}^m$ *diferenciable en* $\bar{x}$. *Si* $\bar{x} \in \mathcal{D}$ *y* $\bar{\lambda} \in \mathcal{M}(\bar{x})$, *o sea*

$$0 = \begin{bmatrix} f'(\bar{x}) + h'(\bar{x})^\top\bar{\lambda} \\ h(\bar{x}) \end{bmatrix}, \tag{2.7}$$

entonces $\bar{x}$ *es un punto estacionario del problema* (2.6), *o sea* $-f'(\bar{x}) \in \mathcal{N}_{\mathcal{D}}(\bar{x})$.

Demostración. Por definición de $\mathcal{N}_{\mathcal{D}}(\bar{x})$, tomemos $\{x^k\} \subset \mathcal{D} \setminus \{\bar{x}\}$ arbitraria tal que $x^k \to \bar{x}$. Como h es diferenciable en $\bar{x}$ con $x^k, \bar{x} \in \mathcal{D}$ tenemos

$$0 = h(x^k) = h(\bar{x}) + h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) = h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|).$$

Entonces, como $0 = f'(\bar{x}) + h'(\bar{x})^\top \bar{\lambda}$,

$$\begin{aligned}\langle -f'(\bar{x}), x^k - \bar{x}\rangle &= \langle h'(\bar{x})^\top \bar{\lambda}, x^k - \bar{x}\rangle \\ &= \langle \bar{\lambda}, h'(\bar{x})(x^k - \bar{x})\rangle \\ &= -\langle \bar{\lambda}, o(\|x^k - \bar{x}\|)\rangle = o(\|x^k - \bar{x}\|).\end{aligned}$$

Por lo tanto

$$\limsup_{k \to +\infty} \frac{\langle -f'(\bar{x}), x^k - \bar{x}\rangle}{\|x^k - \bar{x}\|} \leq 0.$$

O sea $-f'(\bar{x}) \in \mathcal{N}_{\mathcal{D}}(\bar{x})$. ■

En la literatura, la ecuación (2.7) se conoce como *condición de Lagrange*. Hay que remarcar que uno debe ser cuidadoso al usar el resutado anterior, ya que la afirmación recíproca no es cierta. O sea, puede ocurrir que no exista ningún multiplicador de Lagrange asociado a una solución $\bar{x}$.

Ejemplo 7 *Considere el problema*

$$\begin{array}{rl}\text{minimizar} & x \\ \text{sujeto a} & x^2 = 0.\end{array}$$

Claramente $\bar{x} = 0$ es la solución del problema. Si existe $\lambda \in \mathcal{M}(\bar{x})$ entonces

$$0 = f'(\bar{x}) + h'(\bar{x})^\top \lambda = 1 + 2\lambda\bar{x} = 1.$$

Por lo tanto $\mathcal{M}(\bar{x}) = \emptyset$.

Encontrar la solución de un sistema de ecuaciones no lineales es claramente mas simple que hallar $\bar{x}$ tal que $-f'(\bar{x}) \in \mathcal{N}_{\mathcal{D}}(\bar{x})$. Por lo tanto necesitaremos imponer cierta condición de regularidad sobre h en $\bar{x}$, dicha condición se conoce en la literatura como *Constraint Qualification* (QP). Una de ellas es la siguiente.

Definición 7 *Diremos que la condición de regularidad de independencia lineal (LICQ, Linear Independence Constraint Qualification) vale en $\bar{x} \in \mathcal{D}$ si $\{h_i'(\bar{x})\}_{i=1}^m$ son linealmente independientes.*

Por lo tanto, si LICQ vale en $\bar{x}$ entonces $h'(\bar{x})^\top$ tiene rango m, o equivalentemente

$$h'(\bar{x})^\top \lambda = 0 \quad \Rightarrow \quad \lambda = 0.$$

Con esta condición podemos garantizar la existencia de multiplicadores de Lagrange.

Teorema 14 *Sean $f : \mathbb{R}^n \to \mathbb{R}$ y $h : \mathbb{R}^n \to \mathbb{R}^m$ contínuamente diferenciables en $\bar{x}$. Si $\bar{x}$ es minimizador local de f en $\mathcal{D} = \{x \in \mathbb{R}^n \mid h(x) = 0\}$ y vale LICQ en $\bar{x}$, entonces existe $\bar{\lambda} \in \mathcal{M}(\bar{x})$, o sea $(\bar{x}, \bar{\lambda})$ es solución de (2.7). Más aún, $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$.*

Demostración. Por definición de minimizador local, existe $\varepsilon > 0$ tal que $f(\bar{x}) \leq f(x)$ para todo $x \in \mathcal{D} \cap \mathcal{B}_\varepsilon(\bar{x})$.

Tomamos una sucesión arbitraria de números reales $r_k \to +\infty$ y generamos $\{x^k\}$ tal que x^k es solución global de

$$\begin{array}{rl}\text{minimizar} & f(x) + \frac{1}{2}\|x - \bar{x}\|^2 + \frac{r_k}{2}\|h(x)\|^2 \\ \text{sujeto a} & x \in \mathcal{B}_\varepsilon(\bar{x}).\end{array}$$

Note que $x^k \in \mathcal{B}_\varepsilon(\bar{x})$ existe por el Teorema 5. Además, como $\bar{x} \in \mathcal{B}_\varepsilon(\bar{x})$, tenemos que

$$f(\bar{x}) \geq f(x^k) + \frac{1}{2}\|x^k - \bar{x}\|^2 + \frac{r_k}{2}\|h(x^k)\|^2. \tag{2.8}$$

Como $\{x^k\}$ es acotada, tomando subsucesiones si fuera necesario, existe $\hat{x} \in \mathcal{B}_\varepsilon(\bar{x})$ tal que $x^k \to \hat{x}$. Veamos que $\hat{x} = \bar{x}$. Por un lado, de (2.8) obtenemos que

$$\|h(x^k)\|^2 \leq \frac{2}{r_k}(f(\bar{x}) - f(x^k).$$

Tomando límite concluimos que $\|h(\hat{x})\| = 0$ y por lo tanto $\hat{x} \in \mathcal{D} \cap \mathcal{B}_\varepsilon(\bar{x})$. Con esto, obtenemos que $f(\bar{x}) \leq f(\hat{x})$. Por otro lado, de (2.8) tenemos que

$$\frac{1}{2}\|x^k - \bar{x}\|^2 \leq f(\bar{x}) - f(x^k).$$

Tomando límite obtenemos que $\frac{1}{2}\|\hat{x} - \bar{x}\|^2 \leq f(\bar{x}) - f(\hat{x}) \leq 0$. Conclyendo que $\hat{x} = \bar{x}$.

Ahora, como $x^k \to \bar{x}$ entonces para k suficientemente grande tendremos $x^k \in \operatorname{int}(\mathcal{B}_\varepsilon(\bar{x}))$. Como x^k es un minimizador, por Teorema 12 y Ejercicio 3, tenemos que

$$0 = f'(x^k) + x^k - \bar{x} + h'(x^k)^\top \lambda^k, \quad \text{con} \quad \lambda^k = r_k h(x^k). \tag{2.9}$$

Veamos que $\{\lambda^k\}$ es acotada. Si no lo es definimos $\eta_k = \|\lambda^k\|$ y, tomando subsucesiones si fuera necesario, tendremos $\eta_k \to +\infty$ tal que $\frac{1}{\eta_k}\lambda^k \to \hat{\lambda}$ con $\|\hat{\lambda}\| = 1$. Por lo tanto, dividiendo por η_k en (2.9) y tomando límite, obtendremos que

$$h'(\bar{x})^\top \hat{\lambda} = 0$$

y por lo tanto $\hat{\lambda} = 0$ por LICQ, contradiciendo que $\|\hat{\lambda}\| = 1$.

Como $\{\lambda^k\}$ es acotada, tomando subsucesiones, existe $\bar{\lambda}$ tal que $\lambda^k \to \bar{\lambda}$. Tomando límite en (2.9) obtenemos que

$$0 = f'(\bar{x}) + h'(\bar{x})^\top \bar{\lambda}.$$

Veamos que $\bar{\lambda}$ es único. Si existiera $\tilde{\lambda} \in \mathcal{M}(\bar{x})$ tendríamos

$$0 = f'(\bar{x}) + h'(\bar{x})^\top \tilde{\lambda} = h'(\bar{x})^\top (\tilde{\lambda} - \bar{\lambda})$$

y por LICQ obtenemos que $\tilde{\lambda} = \bar{\lambda}$. ■

Como ya vimos, cuando $h(x) = Ax - b$ podemos garantizar que $\mathcal{M}(\bar{x}) \neq \emptyset$. Más aún, podemos tener infinitos multiplicadores de Lagrange asociados a $\bar{x}$ si $\operatorname{Ker}(A^\top) \neq \emptyset$, pues si $\lambda \in \mathcal{M}(\bar{x})$ entonces $\lambda + \nu \in \mathcal{M}(\bar{x})$ para todo $\nu \in \operatorname{Ker}(A^\top)$. Claramente LICQ no puede valer en este caso. La existencia de multiplicadores se garantiza usando otra condición de regularidad.

Definición 8 *Diremos que la condición de regularidad del rango constante (CRCQ, Constant Rank Constraint Qualification) vale en $\bar{x} \in \mathcal{D}$ si existe $\varepsilon > 0$ tal que para cualquier conjunto de índices $J \subset \{1, \ldots, m\}$ el rango de $\{h'_j(x)\}_{j \in J}$ es constante para todo $x \in \mathcal{B}_\varepsilon(\bar{x})$.*

Por lo tanto, si CRCQ vale en $\bar{x}$ y existe $J \subset \{1, \ldots, m\}$ tal que $\{h'_j(\bar{x})\}_{j \in J}$ son linealmente dependientes, entonces $\{h'_j(x)\}_{j \in J}$ deben ser linealmente dependientes para x suficientemente cerca de $\bar{x}$. Con esta condición también podemos garantizar la existencia de multiplicadores de Lagrange.

Teorema 15 *Sean $f : \mathbb{R}^n \to \mathbb{R}$ y $h : \mathbb{R}^n \to \mathbb{R}^m$ contínuamente diferenciables en $\bar{x}$. Si $\bar{x}$ es minimizador local de f en $\mathcal{D} = \{x \in \mathbb{R}^n \mid h(x) = 0\}$ y vale CRCQ en $\bar{x}$, entonces existe $\bar{\lambda} \in \mathcal{M}(\bar{x})$.*

Demostración. La demostración es igual que la del Teorema 14 hasta (2.9). Por lo tanto, tenemos $\{x^k\}$ tal que $x^k \to \bar{x}$ y vale (2.9).

Ahora, usaremos el hecho de que toda combinación lineal de vectores puede escribirse como combinación lineal de un subconjunto de ellos linealmente independientes. En nuestro caso, para cada k existe $J_k \subset \{1, \dots, m\}$ y $\{\tilde{\lambda}_j^k\}_{j \in J_k}$ tal que (2.9) puede escribirse como

$$0 = f'(x^k) + x^k - \bar{x} + \sum_{j \in J_k} \tilde{\lambda}_j^k h_j'(x^k),$$

donde $\{h_j'(x^k)\}_{j \in J_k}$ son linealmente independientes.

Como $\{1, \dots, m\}$ es finito, tiene que existir $J \subset \{1, \dots, m\}$ tal que $J_k = J$ para infinitos valores de k. Consideremos entonces aquella subsucesión con k tal que $J_k = J$. Por lo tanto, sin perdida de generalidad consideremos que

$$0 = f'(x^k) + x^k - \bar{x} + \sum_{j \in J} \tilde{\lambda}_j^k h_j'(x^k). \tag{2.10}$$

Definamos $\eta_k = \sqrt{\sum_{j \in J} (\tilde{\lambda}_j^k)^2}$. Si $\eta_k \to +\infty$, tomando subsucesiones podemos considerar que $\frac{1}{\eta_k} \tilde{\lambda}_j^k \to \hat{\lambda}_j$ para cada $j \in J$ y por lo tanto $\sum_{j \in J} \hat{\lambda}_j^2 = 1$. Dividiendo por η_k en (2.10) y tomando límite, obtendremos que

$$0 = \sum_{j \in J} \hat{\lambda}_j h_j'(\bar{x}).$$

Por lo tanto $\{h_j'(\bar{x})\}_{j \in J}$ son linealmente dependientes y por CRCQ también debemos tener $\{h_j'(x^k)\}_{j \in J}$ linealmente dependientes para k suficientemente grande, contradiciendo el hecho de que estos vectores eran linealmente independientes.

Por lo tanto $\{\eta_k\}$ es acotada y tomando subsucesiones si fuera necesario, podemos garantizar que $\tilde{\lambda}_j^k \to \bar{\lambda}_j$ para cada $j \in J$. Tomando límite en (2.10) obtenemos

$$0 = f'(\bar{x}) + \sum_{j \in J} \bar{\lambda}_j h_j'(\bar{x}).$$

Tomando $\bar{\lambda}_j = 0$ para $j \notin J$, concluímos que

$$0 = f'(\bar{x}) + h'(\bar{x})^\top \bar{\lambda}.$$

■

Como se puede ver, la existencia de multiplicadores de Lagrange puede garantizarse si uno consigue definir $\mathcal{D}$ con una función h suficientemente regular.

Ejemplo 8 *Sea $Q \in \mathbb{R}^{n \times n}$ simétrica definida positiva. Considere el problema*

$$\begin{array}{rl} \text{minimizar} & \frac{1}{2}\langle Qx, x\rangle \\ \text{sujeto a} & x \in \mathcal{S}^1 = \{x \in \mathbb{R}^n \mid \|x\| = 1\}. \end{array}$$

1. *Tomo $h(x) = \frac{1}{2}(1 - \|x\|^2)$, entonces $\mathcal{S}^1 = \{x \in \mathbb{R}^n \mid h(x) = 0\}$ con h continuamente diferenciable con $h'(x) = -x$. Por Teorema 5 existe $\bar{x} \in \mathcal{S}^1$ minimizador y como $h'(\bar{x}) \neq 0$ (o sea, vale LICQ), existe $\bar{\lambda} \in \mathcal{M}(\bar{x})$ por Teorema 14. Entonces $Q\bar{x} = \bar{\lambda}\bar{x}$ con $\|\bar{x}\| = 1$.*

2. *Tomo $h(x) = \frac{1}{4}(1 - \|x\|^2)^2$, entonces $\mathcal{S}^1 = \{x \in \mathbb{R}^n \mid h(x) = 0\}$ con h continuamente diferenciable con $h'(x) = -(1 - \|x\|^2)x$. Por Teorema 5 existe $\bar{x} \in \mathcal{S}^1$ minimizador. No obstante $\mathcal{M}(\bar{x}) = \emptyset$, pues si existiese $\bar{\lambda} \in \mathcal{M}(\bar{x})$ entonces*

$$0 = Q\bar{x} - \bar{\lambda}(1 - \|\bar{x}\|^2)\bar{x} = Q\bar{x} \quad \text{con} \quad \|\bar{x}\| = 1,$$

contradiciendo el hecho de que Q es definida positiva.

2.4.1. Usando el método de Newton.

Para hallar $\bar{w} = (\bar{x}, \bar{\lambda})$ tal que $F(\bar{w}) = 0$ con

$$F(x, \lambda) = \begin{bmatrix} \frac{\partial L}{\partial x}(x, \lambda) \\ h(x) \end{bmatrix},$$

el método de Newton genera $\{w^k\}$ con $w^k = (x^k, \lambda^k)$ tal que

$$\begin{aligned} 0 &= F(w^k) + F'(w^k)(w^{k+1} - w^k) \\ &= \begin{bmatrix} \frac{\partial L}{\partial x}(x^k, \lambda^k) \\ h(x^k) \end{bmatrix} + \begin{bmatrix} \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) & h'(x^k)^\top \\ h'(x^k) & 0 \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix} \\ &= \begin{bmatrix} f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x^{k+1} - x^k) + h'(x^k)^\top \lambda^{k+1} \\ h(x^k) + h'(x^k)(x^{k+1} - x^k) \end{bmatrix}, \end{aligned}$$

donde $\frac{\partial^2 L}{\partial x^2}(x, \lambda) = f''(x) + \sum_{i=1}^m \lambda_i h_i''(x)$.

Desde el punto de vista de optimización, podemos decir que en la k-ésima iteración el método de Newton halla (x^{k+1}, λ^{k+1}) donde λ^{k+1} es un multiplicador de Lagrange asociado a un punto estacionario x^{k+1} del problema

$$\begin{array}{rl} \text{minimizar} & f(x^k) + \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x - x^k), x - x^k \right\rangle \\ \text{sujeto a} & h(x^k) + h'(x^k)(x - x^k) = 0. \end{array} \tag{2.11}$$

Note que la existencia de λ^{k+1} puede deducirse usando el Teorema 15 o simplemente la Proposición 10.

El proceso iterativo que genera $\{(x^k, \lambda^k)\}$ mediante la resolución de (2.11) es conocido en la literatura como método de programación cuadrática secuencial (SQP, Sequential Quadratic Programming).

Para garantizar la convergencia de $\{w^k\}$ generada por el método de Newton (o por SQP) a $\bar{w}$, necesitamos que $F(\bar{w}) = 0$ con $F'(\bar{w})$ sea no singular. Estas condiciones equivalen a decir que $w = \bar{w}$ es la única solución de $0 = F(\bar{w}) + F'(\bar{w})(w - \bar{w})$. Desde el punto de vista de optimización, esto es equivalente a decir que $\bar{\lambda}$ es el único multiplicador de Lagrange asociado a la única solución $\bar{x}$ de

$$\begin{array}{rl} \text{minimizar} & f(\bar{x}) + \langle f'(\bar{x}), x - \bar{x} \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})(x - \bar{x}), x - \bar{x} \right\rangle \\ \text{sujeto a} & h(\bar{x}) + h'(\bar{x})(x - \bar{x}) = 0. \end{array}$$

Como el valor optimal de este problema es $f(\bar{x})$, si existiera otra solución $\tilde{x}$ debería satisfacer

$$\begin{aligned} f(\bar{x}) &= f(\bar{x}) + \langle f'(\bar{x}), \tilde{x} - \bar{x} \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})(\tilde{x} - \bar{x}), \tilde{x} - \bar{x} \right\rangle, \\ 0 &= h(\bar{x}) + h'(\bar{x})(\tilde{x} - \bar{x}) = h'(\bar{x})(\tilde{x} - \bar{x}). \end{aligned}$$

Como $f'(\bar{x}) + h'(\bar{x})^\top \bar{\lambda} = 0$ entonces

$$\langle f'(\bar{x}), \tilde{x} - \bar{x} \rangle = -\langle h'(\bar{x})^\top \bar{\lambda}, \tilde{x} - \bar{x} \rangle = -\langle \bar{\lambda}, h'(\bar{x})(\tilde{x} - \bar{x}) \rangle = 0.$$

Obteniendo así que

$$0 = \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})(\tilde{x} - \bar{x}), \tilde{x} - \bar{x} \right\rangle \quad \text{con} \quad h'(\bar{x})(\tilde{x} - \bar{x}) = 0.$$

Por lo tanto, para garantizar que no exista otra solución distinta a $\bar{x}$ debemos tener

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})(x - \bar{x}), x - \bar{x} \right\rangle > 0$$

para $x \neq \bar{x}$ con $h'(\bar{x})(x - \bar{x}) = 0$.

Veamos que esta condición nos garantiza existencia de minimizador

Teorema 16 (Condición suficiente de optimalidad) *Sean $f : \mathbb{R}^n \to \mathbb{R}$ y $h : \mathbb{R}^n \to \mathbb{R}^m$ dos veces diferenciables en $\bar{x}$. Si $h(\bar{x}) = 0$ y existe $\bar{\lambda} \in \mathcal{M}(\bar{x})$ tal que*

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})d, d \right\rangle > 0 \quad \forall d \neq 0 \quad \textit{con} \quad h'(\bar{x})d = 0.$$

Entonces $\bar{x}$ es un minimizador local de (2.6). Más aún, existen $\gamma > 0$ y $\varepsilon > 0$ tales que

$$f(x) \geq f(\bar{x}) + \frac{\gamma}{2}\|x - \bar{x}\|^2 \quad \textit{para todo} \quad x \in \mathcal{B}_\varepsilon(\bar{x}) \quad \textit{con} \quad h(x) = 0. \tag{2.12}$$

2.5. Optimización con restricciones de igualdad y desigualdad

Consideremos un caso particular del problema (2.1) con $\mathcal{D} = \{x \in \mathbb{R}^n \mid h(x) = 0\, g(x) \leq 0\}$ donde $h : \mathbb{R}^n \to \mathbb{R}^l$ y $g : \mathbb{R}^n \to \mathbb{R}^m$, o sea

$$\begin{array}{rl} \text{minimizar} & f(x) \\ \text{sujeto a} & h(x) = 0, \\ & g(x) \leq 0, \end{array} \tag{2.13}$$

donde la desigualdad $g(x) \leq 0$ es tomada componente a componente, i.e., $g_i(x) \leq 0$ para $i = 1, \ldots, m$.

Ejemplo 9 *Por mas que agreguemos una restricción $e^\top x = \kappa$ en el Ejemplo 5 para forzar a que solo se pueda invertir un cierto capital κ, puede ocurrir que obtengamos $x_i < 0$ para algún i. Este tipo de solución significaría que se deben vender acciones de la empresa i. Si nuestra intención es solo comprar acciones, deberíamos restringirnos solo al conjunto con $x_i \geq 0$. Por lo tanto, el problema a resolver es*

$$\begin{array}{rl} \text{minimizar} & \frac{1}{2}\langle Sx, x\rangle - \alpha\langle \mu, x\rangle \\ \text{sujeto a} & e^\top x = \kappa, \\ & x \geq 0. \end{array}$$

Definición 9 *La función Lagrangiana asociada al problema (2.13) es $L : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ tal que*

$$L(x, \lambda) = f(x) + \langle \lambda, h(x)\rangle + \langle \mu, g(x)\rangle.$$

El conjunto de mutiplicadores de Lagrange asociado a $\bar{x}$ factible para el problema (2.13) es

$$\mathcal{M}(\bar{x}) = \left\{ (\lambda, \mu) \in \mathbb{R}^l \times \mathbb{R}^m \;\middle|\; \begin{array}{l} 0 = f'(\bar{x}) + h'(\bar{x})^\top \lambda + g'(\bar{x})^\top \mu, \\ \mu_i \geq 0\, i \in I(\bar{x}),\ \mu_i = 0\, i \notin I(\bar{x}) \end{array} \right\},$$

donde $I(\bar{x}) = \{i \in \{1, \ldots, m\} \mid g_i(\bar{x}) = 0\}$ es el conjunto de índices de desigualdades activas en $\bar{x}$.

Hay que resaltar que el problema con restricciones de igualdad es un caso particular de (2.13) considerando $g(x) = 0$ para todo $x \in \mathbb{R}^n$. Veamos que la existencia de mutiplicador nos garantiza que estamos en un punto estacionario.

Teorema 17 *Sea $f : \mathbb{R}^n \to \mathbb{R}$ y $\mathcal{D} = \{x \in \mathbb{R}^n \mid h(x) = 0, g(x) \leq 0\}$ con $h : \mathbb{R}^n \to \mathbb{R}^l$ y $g : \mathbb{R}^n \to \mathbb{R}^m$ diferenciables en $\bar{x}$. Si $\bar{x} \in \mathcal{D}$ y $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, o sea*

$$\begin{array}{l} 0 = f'(\bar{x}) + h'(\bar{x})^\top \bar{\lambda} + g'(\bar{x})^\top \bar{\mu}, \\ 0 = h(\bar{x}), \\ 0 \geq g(\bar{x}),\ \bar{\mu} \geq 0,\ \bar{\mu}_i g_i(\bar{x}) = 0\ i = 1, \ldots, m. \end{array} \tag{2.14}$$

entonces $\bar{x}$ es un punto estacionario del problema (2.13), o sea $-f'(\bar{x}) \in \mathcal{N}_\mathcal{D}(\bar{x})$.

Demostración. Por definición de $\mathcal{N}_{\mathcal{D}}(\bar{x})$, tomemos $\{x^k\} \subset \mathcal{D} \setminus \{\bar{x}\}$ arbitraria tal que $x^k \to \bar{x}$. Como h y g son diferenciables en $\bar{x}$ con $x^k, \bar{x} \in \mathcal{D}$ tenemos

$$\begin{aligned} 0 = h(x^k) = h(\bar{x}) + h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \\ 0 \geq g(x^k) = g(\bar{x}) + g'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|). \end{aligned}$$

Entonces, como $0 = f'(\bar{x}) + h'(\bar{x})^\top \bar{\lambda} + g'(\bar{x})^\top \bar{\mu}$,

$$\begin{aligned} \langle -f'(\bar{x}), x^k - \bar{x} \rangle &= \langle h'(\bar{x})^\top \bar{\lambda} + g'(\bar{x})^\top \bar{\mu}, x^k - \bar{x} \rangle \\ &= \langle \bar{\lambda}, h'(\bar{x})(x^k - \bar{x}) \rangle + \langle \bar{\mu}, g'(\bar{x})(x^k - \bar{x}) \rangle \\ &\leq -\langle \bar{\lambda}, h(\bar{x}) \rangle - \langle \bar{\mu}, g(\bar{x}) \rangle + o(\|x^k - \bar{x}\|) \\ &= o(\|x^k - \bar{x}\|), \end{aligned}$$

donde usamos que $h(\bar{x}) = 0$, $\bar{\mu}_i \geq 0$, $g_i(\bar{x}) \leq 0$ y $\bar{\mu}_i g_i(\bar{x}) = 0$ para todo $i \in \{1, \ldots, m\}$. Por lo tanto

$$\limsup_{k \to +\infty} \frac{\langle -f'(\bar{x}), x^k - \bar{x} \rangle}{\|x^k - \bar{x}\|} \leq 0.$$

O sea $-f'(\bar{x}) \in \mathcal{N}_{\mathcal{D}}(\bar{x})$. ∎

En la literatura, la ecuación (2.14) se conoce como *condición de Karush-Kuhn-Tucker*.

Al igual que para el caso de restricciones de igualdad, puede ocurrir que no exista ningún multiplicador de Lagrange asociado a una solución $\bar{x}$.

Ejemplo 10 *Considere el problema*

$$\begin{array}{rl} \text{minimizar} & x_1 \\ \text{sujeto a} & x_1 - 1 \leq 0, \\ & -x_2 \leq 0, \\ & x_2 - x_1^3 \leq 0. \end{array}$$

Para cualquier x factible tendremos $x_1 \geq (x_2)^{1/3} \geq 0$. Por lo tanto el mínimo se alcanza en $x_1 = 0$, que por la segunda y tercera restricción vale si $x_2 = 0$. Entonces $\bar{x} = 0$ es minimizador global. Ahora, si existe $\mu \in \mathcal{M}(\bar{x})$ entonces $I(\bar{x}) = \{2, 3\}$ y

$$0 = f'(\bar{x}) + g'(\bar{x})^\top \mu = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \mu_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \mu_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

con $\mu_1 = 0$, $\mu_2 \geq 0$ y $\mu_3 \geq 0$. Obteniendo así algo contradictorio. Por lo tanto $\mathcal{M}(\bar{x}) = \emptyset$.

Para garantizar la existencia de multiplicadores de Lagrange tendremos que pedir alguna condición de regularidad sobre las restriccioes.

Definición 10 *Diremos que la condición de regularidad de independencia lineal (LICQ, Linear Independence Constraint Qualification) vale en $\bar{x} \in \mathcal{D}$ si para $I = I(\bar{x})$, $\{h_j'(\bar{x})\}_{j=1}^l \cup \{g_i'(\bar{x})\}_{i \in I}$ son linealmente independientes.*

Por lo tanto, si LICQ vale en $\bar{x}$ entonces

$$0 = h'(\bar{x})^\top \lambda + g_I'(\bar{x})^\top \mu_I \quad \Rightarrow \quad (\lambda, \mu_I) = 0,$$

donde $g_I'(\bar{x})$ es una submatriz de $g'(\bar{x})$ con filas indexadas por I y μ_I es un subvector de μ con componentes indexadas por I.

Ya que los multiplicadores asociados a las desigualdades son nonegativos, podemos usar esta propiedad para obtener una condición de regularidad más débil.

Definición 11 *Diremos que la condición de regularidad de Mangasarian Fromovitz (MFCQ) vale en $\bar{x} \in \mathcal{D}$ si para $I = I(\bar{x})$,*

$$\left.\begin{array}{r} 0 = h'(\bar{x})^\top \lambda + g_I'(\bar{x})^\top \mu_I \\ \mu_I \geq 0 \end{array}\right\} \Rightarrow \quad (\lambda, \mu_I) = 0,$$

Claramente, si vale LICQ entonces vale MFCQ. La reciproca no es cierta.

Ejercicio 4 *Sea $\phi : \mathbb{R}^n \to \mathbb{R}$ tal que $\phi(x) = \frac{1}{2}\|\max\{0, g(x)\}\|^2$ con $g : \mathbb{R}^n \to \mathbb{R}^m$ contínuamente diferenciable en $\mathbb{R}^n$. Demuestre que ϕ es contínuamente diferenciable en $\mathbb{R}^n$ con $\phi'(x) = g'(x)^\top(\max\{0, g(x)\})$ y que $\phi(x) = 0$ si y solo si $g(x) \leq 0$.*

Teorema 18 *Sean $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ y $g : \mathbb{R}^n \to \mathbb{R}^m$ contínuamente diferenciables en $\bar{x}$. Si $\bar{x}$ es minimizador local de f en $\mathcal{D} = \{x \in \mathbb{R}^n \mid h(x) = 0,\, g(x) \leq 0\}$ y vale MFCQ en $\bar{x}$, entonces existe $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, o sea $(\bar{x}, \bar{\lambda}, \bar{\mu})$ es solución de (2.14). Más aún, $\mathcal{M}(\bar{x})$ es compacto.*

Demostración. Por definición de minimizador local, existe $\varepsilon > 0$ tal que $f(\bar{x}) \leq f(x)$ para todo $x \in \mathcal{D} \cap \mathcal{B}_\varepsilon(\bar{x})$.

Tomamos una sucesión arbitraria de números reales $r_k \to +\infty$ y generamos $\{x^k\}$ tal que x^k es solución global de

$$\begin{array}{rl} \text{minimizar} & f(x) + \frac{1}{2}\|x - \bar{x}\|^2 + \frac{r_k}{2}(\|h(x)\|^2 + \|\max\{0, g(x)\}\|^2 \\ \text{sujeto a} & x \in \mathcal{B}_\varepsilon(\bar{x}). \end{array}$$

Note que $x^k \in \mathcal{B}_\varepsilon(\bar{x})$ existe por el Teorema 5. Además, como $\bar{x} \in \mathcal{B}_\varepsilon(\bar{x})$, tenemos que

$$f(\bar{x}) \geq f(x^k) + \frac{1}{2}\|x^k - \bar{x}\|^2 + \frac{r_k}{2}(\|h(x^k)\|^2 + \|\max\{0, g(x)\}\|^2). \tag{2.15}$$

Como $\{x^k\}$ es acotada, tomando subsucesiones si fuera necesario, existe $\hat{x} \in \mathcal{B}_\varepsilon(\bar{x})$ tal que $x^k \to \hat{x}$. Veamos que $\hat{x} = \bar{x}$. Por un lado, de (2.15) obtenemos que

$$\|h(x^k)\|^2 + \|\max\{0, g(x)\}\|^2 \leq \frac{2}{r_k}(f(\bar{x}) - f(x^k).$$

Tomando límite concluimos que $\|h(\hat{x})\| = 0$ y $\|\max\{0, g(\hat{x})\}\| = 0$, por lo tanto $\hat{x} \in \mathcal{D} \cap \mathcal{B}_\varepsilon(\bar{x})$. Con esto, obtenemos que $f(\bar{x}) \leq f(\hat{x})$. Por otro lado, de (2.15) tenemos que

$$\frac{1}{2}\|x^k - \bar{x}\|^2 \leq f(\bar{x}) - f(x^k).$$

Tomando límite obtenemos que $\frac{1}{2}\|\hat{x} - \bar{x}\|^2 \leq f(\bar{x}) - f(\hat{x}) \leq 0$. Conclyendo que $\hat{x} = \bar{x}$.

Ahora, como $x^k \to \bar{x}$ entonces para k suficientemente grande tendremos $x^k \in \operatorname{int}(\mathcal{B}_\varepsilon(\bar{x}))$. Como x^k es un minimizador, por Teorema 12 y Ejercicio 3, tenemos que

$$0 = f'(x^k) + x^k - \bar{x} + h'(x^k)^\top \lambda^k + g'(x^k)^\top \mu^k, \tag{2.16}$$

con $\lambda^k = r_k h(x^k)$ y $\mu^k = r_k \max\{0, g(x^k)\}$. Note que si $i \notin I$ entonces $g_i(\bar{x}) < 0$ y por lo tanto $g_i(x^k) < 0$ para k suficientemente grande. Usando la definición de μ^k y tomando k grande, tendremos entonces $\mu_i^k \geq 0$ para $i \in I$ y $\mu_i^k = 0$ para $i \notin I$.

Veamos que $\{(\lambda^k, \mu^k)\}$ es acotada. Si no lo es definimos $\eta_k = \|(\lambda^k, \mu^k)\|$ y, tomando subsucesiones si fuera necesario, tendremos $\eta_k \to +\infty$ tal que $\frac{1}{\eta_k}(\lambda^k, \mu^k) \to (\hat{\lambda}, \hat{\mu})$ con $\|(\hat{\lambda}, \hat{\mu})\| = 1$, $\hat{\mu}_i \geq 0$ para $i \in I$ y $\hat{\mu}_i = 0$ para $i \notin I$. Por otro lado, dividiendo por η_k en (2.16) y tomando límite, obtendremos que

$$0 = h'(\bar{x})^\top \hat{\lambda} + g'(\bar{x})^\top \hat{\mu} = h'(\bar{x})^\top \hat{\lambda} + g_I'(\bar{x})^\top \hat{\mu}_I,$$

con $\hat{\mu}_I \geq 0$. Por lo tanto $(\hat{\lambda}, \hat{\mu}) = 0$ por MFCQ, contradiciendo que $\|(\hat{\lambda}, \hat{\mu})\| = 1$.

Como $\{(\lambda^k, \mu^k)\}$ es acotada, tomando subsucesiones, existe $(\bar{\lambda}, \bar{\mu})$ con $\bar{\mu}_i \geq 0$ para $i \in I$ y $\bar{\mu}_i = 0$ para $i \notin I$, tal que $(\lambda^k, \mu^k) \to (\bar{\lambda}, \bar{\mu})$. Tomando límite en (2.16) obtenemos que

$$0 = f'(\bar{x}) + h'(\bar{x})^\top \bar{\lambda} + g'(\bar{x})^\top \bar{\mu},$$

con $h(\bar{x}) = 0$, $g(\bar{x}) \leq 0$ y $\bar{\mu}_i g_i(\bar{x}) = 0$ $i = 1, \dots, m$.

Veamos que $\mathcal{M}(\bar{x})$ es compacto. Por definición de $\mathcal{M}(\bar{x})$ es claro que este conjunto es cerrado, por lo tanto faltaría ver que es acotado. Si existiera $\{(\bar{\lambda}^k, \bar{\mu}^k)\}$ no acotada con $(\bar{\lambda}^k, \bar{\mu}^k) \in \mathcal{M}(\bar{x})$ tendríamos

$$0 = f'(\bar{x}) + h'(\bar{x})^\top \bar{\lambda}^k + g'(\bar{x})^\top \bar{\mu}^k, \quad \bar{\mu}_i^k g_i(\bar{x}) = 0 \, i = 1, \dots, m \tag{2.17}$$

y $\eta_k = \|(\bar{\lambda}^k, \bar{\mu}^k)\| \to +\infty$. Tomando subsucesiones si fuera necesario, tendremos $\frac{1}{\eta_k}(\bar{\lambda}^k, \bar{\mu}^k) \to (\bar{\lambda}, \bar{\mu})$ con $\|(\bar{\lambda}, \bar{\mu})\| = 1$, $\bar{\mu}_i \geq 0$ para $i \in I$ y $\bar{\mu}_i = 0$ para $i \notin I$. Dividiendo por η_k en (2.17) y tomando límite, obtendremos que

$$0 = h'(\bar{x})^\top \bar{\lambda} + g'(\bar{x})^\top \bar{\mu}, \quad \bar{\mu}_i g_i(\bar{x}) = 0 \, i = 1, \dots, m,$$

y por MFCQ deberemos tener $(\bar{\lambda}, \bar{\mu}) = 0$, contradiciendo que $\|(\bar{\lambda}, \bar{\mu})\| = 1$. Por lo tanto $\mathcal{M}(\bar{x})$ es acotado. ■

Similar al caso con restricciones de igualdad, si $\bar{x}$ es un punto estacionario del problema de minimizar f en $\mathcal{D} = \{x \in \mathbb{R}^n \mid Cx = d, \; Ax \leq b\}$ podemos garantizar que $\mathcal{M}(\bar{x}) \neq \emptyset$. Más aún, podemos tener infinitos multiplicadores de Lagrange asociados a $\bar{x}$ si $\mathrm{Ker}(C^\top) \neq \emptyset$ o $\mathrm{Ker}(A^\top) \neq \emptyset$, pues si $(\lambda, \mu) \in \mathcal{M}(\bar{x})$ entonces $(\lambda + \zeta, \mu + \nu) \in \mathcal{M}(\bar{x})$ para todo $\zeta \in \mathrm{Ker}(C^\top)$ y $\nu \in \mathrm{Ker}(A^\top)$ con $\nu_i \geq 0$ si $i \in I(\bar{x})$ y $\nu_i = 0$ si $i \notin I(\bar{x})$. Claramente MFCQ no puede valer en este caso. La existencia de multiplicadores se garantiza usando CRCQ.

Definición 12 *Diremos que la condición de regularidad del rango constante (CRCQ, Constant Rank Constraint Qualification) vale en $\bar{x} \in \mathcal{D}$ si existe $\varepsilon > 0$ tal que para cualquier conjunto de índices $J_0 \subset \{1, \dots, m\}$, $J_1 \subset I(\bar{x})$ el rango de $\{h_j'(x)\}_{j \in J_0} \cup \{g_j'(x)\}_{j \in J_1}$ es constante para todo $x \in \mathcal{B}_\varepsilon(\bar{x})$.*

De manera análoga al Teorema 15 se puede demostrar el siguiente resultado de existencia de multiplicadores para problemas con restricciones de igualdad y desigualdad.

Teorema 19 *Sean $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ y $g : \mathbb{R}^n \to \mathbb{R}^m$ contínuamente diferenciables en $\bar{x}$. Si $\bar{x}$ es minimizador local de f en $\mathcal{D} = \{x \in \mathbb{R}^n \mid h(x) = 0, \, g(x) \leq 0\}$ y vale CRCQ en $\bar{x}$, entonces existe $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$.*

2.5.1. Usando el método de Newton.

Análogo al caso con restricciones de igualdad, el método de Newton para resolver el problema 2.13 puede ser escrito como una sucesión de problemas de optimización, en este caso es conocido en la literatura como método SQP (Sequential Quadratic Programming) y puede describirse de la siguiente manera: en la k-ésima iteración el método SQP halla $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ donde $(\lambda^{k+1}, \mu^{k+1})$ es un multiplicador de Lagrange asociado a un punto estacionario x^{k+1} del problema

$$\begin{array}{rl} \text{minimizar} & f(x^k) + \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k), x - x^k \right\rangle \\ \text{sujeto a} & h(x^k) + h'(x^k)(x - x^k) = 0, \\ & g(x^k) + g'(x^k)(x - x^k) \leq 0. \end{array} \tag{2.18}$$

Note que la existencia de $(\lambda^{k+1}, \mu^{k+1})$ puede deducirse usando el Teorema 19 y recuerde que la función Lagrangiana es $L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle$.

Para garantizar la convergencia de $\{(x^k, \lambda^k, \mu^k)\}$ generada por el método SQP a $(\bar{x}, \bar{\lambda}, \bar{\mu})$, necesitamos que $(\bar{\lambda}, \bar{\mu})$ sea el único multiplicador de Lagrange asociado a la única solución $\bar{x}$ de

$$\begin{array}{rl} \text{minimizar} & f(\bar{x}) + \langle f'(\bar{x}), x - \bar{x}\rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x - \bar{x}), x - \bar{x}\right\rangle \\ \text{sujeto a} & h(\bar{x}) + h'(\bar{x})(x - \bar{x}) = 0, \\ & g(\bar{x}) + g'(\bar{x})(x - \bar{x}) \leq 0. \end{array}$$

Como el valor optimal de este problema es $f(\bar{x})$, para garantizar que no exista otra solución distinta a $\bar{x}$ debemos tener

$$\langle f'(\bar{x}), x - \bar{x}\rangle + \tfrac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x - \bar{x}), x - \bar{x}\right\rangle > 0 \tag{2.19}$$

para $x \neq \bar{x}$ con $h'(\bar{x})(x - \bar{x}) = 0$ y $g_I'(\bar{x})(x - \bar{x}) \leq 0$ con $I = I(\bar{x})$. Como $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ dicho x cumple que

$$\begin{array}{rcl} \langle f'(\bar{x}), x - \bar{x}\rangle & = & \langle -h'(\bar{x})^\top \bar{\lambda} - g'(\bar{x})^\top \bar{\mu}, x - \bar{x}\rangle \\ & = & -\langle \bar{\lambda}, h'(\bar{x})(x - \bar{x})\rangle - \langle \bar{\mu}, g'(\bar{x})(x - \bar{x})\rangle \\ & = & 0 - \langle \bar{\mu}_I, g_I'(\bar{x})(x - \bar{x})\rangle \\ & \geq & 0. \end{array}$$

Por lo tanto, si $\langle f'(\bar{x}), x - \bar{x}\rangle > 0$ para todo x próximo a $\bar{x}$, claramente vale (2.19). En caso de que $\langle f'(\bar{x}), x - \bar{x}\rangle = 0$, deberemos pedir

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x - \bar{x}), x - \bar{x}\right\rangle > 0.$$

Definición 13 *El cono crítico asociado al problema* (2.13) *en* $\bar{x} \in \mathcal{D}$ *es*

$$\mathcal{C}(\bar{x}) = \{d \in \mathbb{R}^n \mid \langle f'(\bar{x}), d\rangle = 0,\ h'(\bar{x})d = 0,\ g_I'(\bar{x})d \leq 0\}$$

Veamos que esta condición nos garantiza existencia de minimizador

Teorema 20 (Condición suficiente de optimalidad) *Sean* $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ *y* $g : \mathbb{R}^n \to \mathbb{R}^m$ *dos veces diferenciables en* $\bar{x}$. *Si* $\bar{x} \in \mathcal{D}$ *y existe* $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ *tal que*

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})d, d\right\rangle > 0 \quad \forall d \in \mathcal{C}(\bar{x}) \setminus \{0\}.$$

Entonces $\bar{x}$ *es un minimizador local de* (2.13). *Más aún, existen* $\gamma > 0$ *y* $\varepsilon > 0$ *tales que*

$$f(x) \geq f(\bar{x}) + \frac{\gamma}{2}\|x - \bar{x}\|^2 \quad \textit{para todo} \quad x \in \mathcal{B}_\varepsilon(\bar{x}) \cap \mathcal{D}. \tag{2.20}$$

Demostración. Supongamos que (2.20) no vale. Entonces podemos generar $\{x^k\} \in \mathcal{D}$ tal que $x^k \to \bar{x}$ y

$$f(x^k) < f(\bar{x}) + \frac{1}{2k}\|x^k - \bar{x}\|^2. \tag{2.21}$$

Tomando subsucesiones si fuera necesario, sabemos que existe $d \in \mathbb{R}^n$ con $\|d\| = 1$ tal que $\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to d$. Para simplificar notación, tomo $I = I(\bar{x})$. Usando (2.21) y el hecho que $x^k \in \mathcal{D}$ entonces

$$\begin{array}{rcl} \frac{1}{2k}\|x^k - \bar{x}\|^2 & > & f(x^k) - f(\bar{x}) = \langle f'(\bar{x}), x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|), \\ 0 & = & h(x^k) - h(\bar{x}) = h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \\ 0 & \geq & g_I(x^k) - g_I(\bar{x}) = g_I'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|). \end{array}$$

Dividiendo por $\|x^k - \bar{x}\|$ y tomando límite obtenemos que

$$\langle f'(\bar{x}), d\rangle \leq 0, \quad h'(\bar{x})d = 0, \quad g_I'(\bar{x})d \leq 0.$$

Usando que $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, sabemos que

$$\langle f'(\bar{x}), d\rangle = \langle -h'(\bar{x})^\top \bar{\lambda} - g'(\bar{x})^\top \bar{\mu}, d\rangle = -\langle \bar{\lambda}, h'(\bar{x})d\rangle - \langle \bar{\mu}_I, g_I'(\bar{x})d\rangle \geq 0.$$

Por lo tanto $\langle f'(\bar{x}), d\rangle = 0$, concluyendo así que $d \in \mathcal{C}(\bar{x})$.

Por otro lado, usando que $L(x^k, \bar{\lambda}, \bar{\mu}) = f(x^k) + \langle \bar{\mu}_I, g_I(x^k)\rangle \leq f(x^k)$, $L(\bar{x}, \bar{\lambda}, \bar{\mu}) = f(\bar{x})$ y $\frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0$, de (2.21) obtenemos que

$$\begin{aligned}
\frac{1}{2k}\|x^k - \bar{x}\|^2 &> f(x^k) - f(\bar{x}) \\
&\geq L(x^k, \bar{\lambda}, \bar{\mu}) - L(\bar{x}, \bar{\lambda}, \bar{\mu}) \\
&= \frac{1}{2}\langle \frac{\partial^2 L}{\partial x^2}(x^k - \bar{x}), x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|^2).
\end{aligned}$$

Dividiendo por $\frac{1}{2}\|x^k - \bar{x}\|^2$ y tomando límite, obtenemos que

$$0 \geq \langle \frac{\partial^2 L}{\partial x^2} d, d\rangle > 0,$$

que solo puede ocurrir para $d = 0$, contradiciendo el hecho que $\|d\| = 1$. ■