

TÉCNICAS DE OPTIMIZACIÓN

I. INTRODUCCIÓN

I.1 DISEÑO ASISTIDO POR ORDENADOR

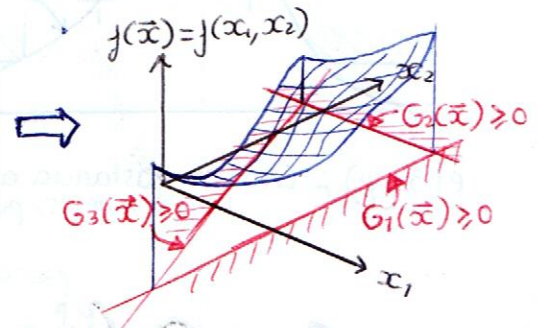
- parámetros de diseño juntos
- buen punto inicial
- simulador preciso y eficiente

- evitar la necesidad de sintonía
- modelar las imperfecciones (ej. esquinas redond.)

I.2 CONCEPTOS BÁSICOS DE OPTIMIZACIÓN

minimizar $f(\vec{x})$

sujeta a: restricciones igualdad $G_i(\vec{x}) = 0$
desigualdad $G_i(\vec{x}) \geq 0$



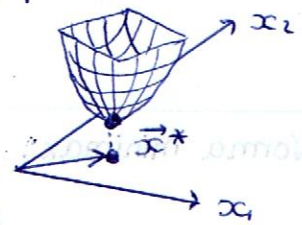
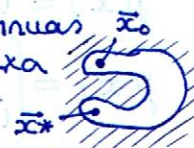
(N x 1) gradiente ~ derivada en N-dim

(N x N) hessiana ~ 2ª derivada en N-dim

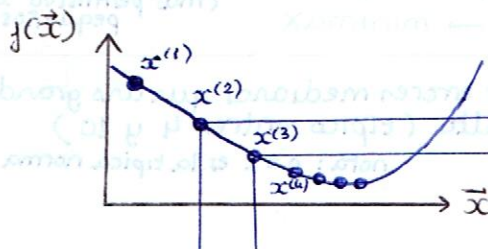
$$\vec{x}^* \text{ minimiza } f(\vec{x}) \iff \begin{cases} \nabla f(\vec{x}^*) = 0 & \text{punto estacionario} \\ \Delta \vec{x}^T \cdot \underline{\underline{H}}(\vec{x}^*) \cdot \Delta \vec{x} \geq 0 & \text{función convexa} \end{cases}$$

Problemas frecuentes:

- valles estrechos y curvados
- derivadas o función discontinuas
- región permitida no convexa



Criterios de terminación:



$$f(\vec{x}^k) - f(\vec{x}^{k-1}) \leq f_{tol}$$

otra opción $\nabla f(\vec{x}^k) \leq grad_{tol}$

$$\|\vec{x}^k - \vec{x}^{k-1}\| \leq x_{tol}$$

→ o bien dim a dim

$$x_i^k - x_i^{k-1} \leq x_{i,tol} \text{ para cada } i$$

orden de convergencia

$$\text{convergencia orden } p \iff \frac{\text{dist}(\vec{x}^k, \vec{x}^*)}{\text{dist}(\vec{x}^{k-1}, \vec{x}^*)^p} \leq cte > 0$$

Nomenclatura

$$\frac{\|h^{k+1}\|}{\|h^k\|^p} \leq a$$

- ejemplo: convergencia lineal ($p=1$) con $a=0.5 \rightarrow$ en cada paso, la distancia al mínimo es la mitad que antes
- sólo se calcula cerca del óptimo, y no es buen indicador
- mejor usar **FUNCIÓN BANANA O DE ROSENBROCK** para testear

La convergencia depende de:

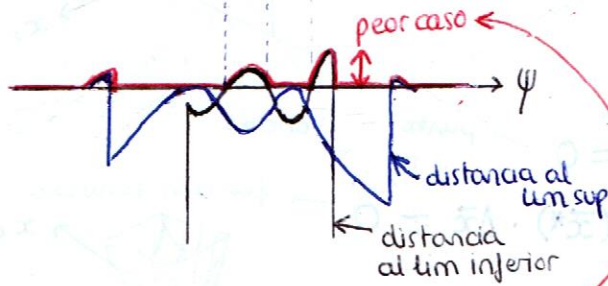
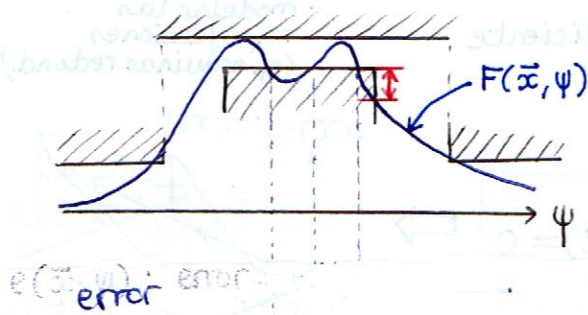
- si se usa S_{21} ó S_{11}
- si se usa lineal o logarítmico
- Punto de partida
- Norma
- Num. puntos para comparar curvas
- Tolerancias / criterios de parada

lineal + S_{21} = suave $\rightarrow$ bueno para empezar

logarítmico + S_{11} = abrupto $\rightarrow$ bueno para refinar

I.3 OPTIMIZACIÓN DE REDES DE MICROONDAS

¿qué usamos como función objetivo o función error? $f(\bar{x})$



Definimos el error en ψ

$$e(\psi, \bar{x})$$

$$e = \max(\omega_u(\psi) \cdot [F(\bar{x}, \psi) - S_u(\psi)], \omega_l(\psi) \cdot [F(\bar{x}, \psi) - S_l(\psi)], 0)$$

ponderación

límite inf

límite sup
distancia al límite superior o inferior (peor caso) ponderando frecuencias

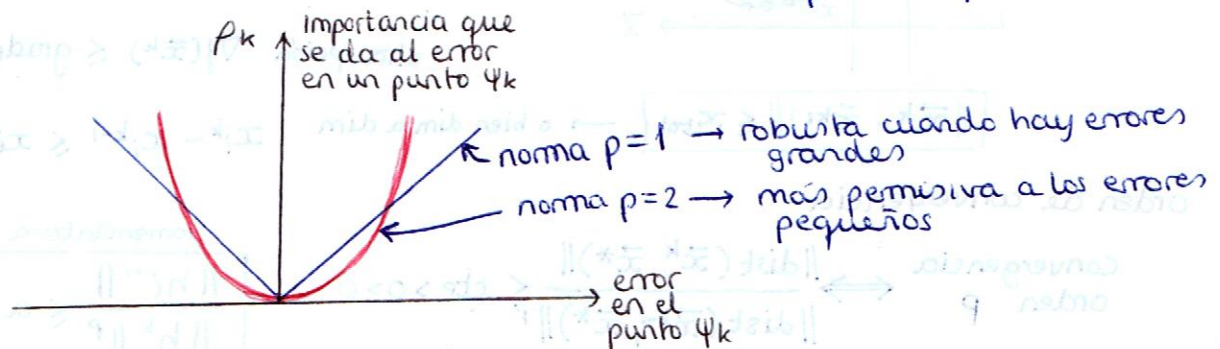
Norma minimax: $f(\bar{x}) = \max(e(\psi, \bar{x}))$ se queda con peor caso

Norma p : $f(\bar{x}) = \left[\sum_{i=1}^m |e(\psi_i, \bar{x})|^p \right]^{1/p}$ se suman los errores en diversos puntos $\psi = \psi_1, \psi_2, \dots, \psi_m$

- A mayor p , mas importancia a errores grandes (mas permisivo con errores pequeños)
- Si $p = \infty \rightarrow$ minimax

ej: si preferimos varios errores medianos que uno grande, usamos un p alto (típico entre 4 y 10)

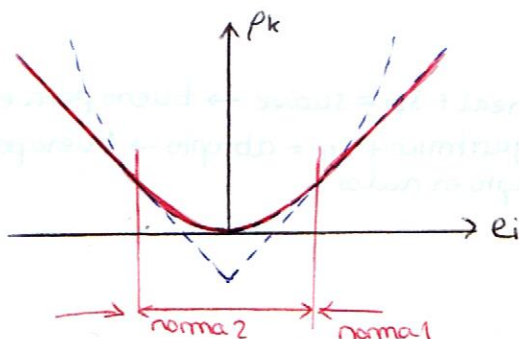
nota: $p = 2$ es la típica norma euclídea



Norma Huber $\rightarrow$ cambia de $p=1$ (errores grandes) a $p=2$ (errores pequeños)

$$f(\bar{x}) = \sum_{i=1}^m \rho_{k,i}$$

$$\text{siendo } \rho_{k,i} = \begin{cases} e_i^2/2 & \text{si } |e_i| \leq k \\ k|e_i| - \frac{k^2}{2} & \text{si } |e_i| > k \end{cases}$$



II. ALGORITMOS DE OPTIMIZACIÓN SIN RESTRICCIONES

II.1 INTRODUCCIÓN

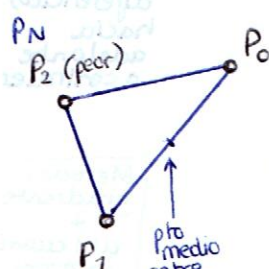
¿cómo determino $\vec{x}^{k+1}$?

- Búsqueda lineal: - Escojo una dirección de avance $\vec{d}_k$
- Hago búsqueda unidimensional $\min g(\alpha) = f(\vec{x}^k + \alpha \cdot \vec{d}_k)$
- Región de confianza: - Escojo un radio de confianza $\|\vec{p}\|_2 < \Delta$
- Aproximo $f(\vec{x}) \approx m_k(\vec{x})$ en el entorno $\vec{x} = \vec{x}^k + \vec{p}$
ej: Taylor hasta orden 2
y busco el mínimo de esa función 'simple' $m_k(\vec{x})$

II.2 MÉTODO DE DESCENSO DEL SIMPLEX

Simplex $\equiv$ Figura N dimensional de N+1 vertices

- No se avanza en la dirección del mejor vértice, sino que se elimina el peor
- Robusto, lento, evita mínimos locales



Reflexion

REFLEXIÓN



bueno
 $P_r < P_{N-1}$

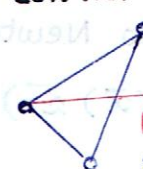
se intenta
Expansión



EXPANSIÓN

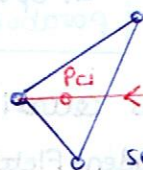
se acepta
si $f(P_e) < f(P_r)$
si no, volvemos
a P_r

CONTRACCIÓN EXTERNA



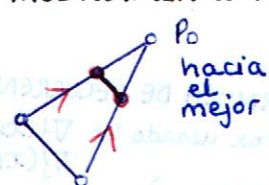
se acepta
si $f(P_{ce}) < f(P_r)$

CONTRACCIÓN INTERNA



se acepta
si $f(P_{ci}) < f(P_r)$

CONTRACCIÓN
MULTIDIMENSIONAL



hacia
el mejor

si no sirven las
contracciones

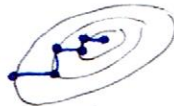
malo (habría bucle)
 $f(P_r) > f(P_{N-1})$

$f(P_r) < f(P_N)$

$f(P_r) > f(P_N)$

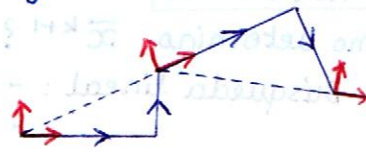
II.3 MÉTODOS DE OPTIMIZACIÓN N-DIM CON BÚSQUEDA DIRECTA

- Búsqueda directa one-at-a-time



zig zag

- Búsqueda directa con rotación de coordenadas según dirección de progreso anterior



- Método de máxima pendiente (gradiente)

$$\vec{d} = -\nabla f(\vec{x}_k)$$

$$\text{minimizar } g(\alpha) = f(\vec{x}_k + \alpha \vec{d})$$

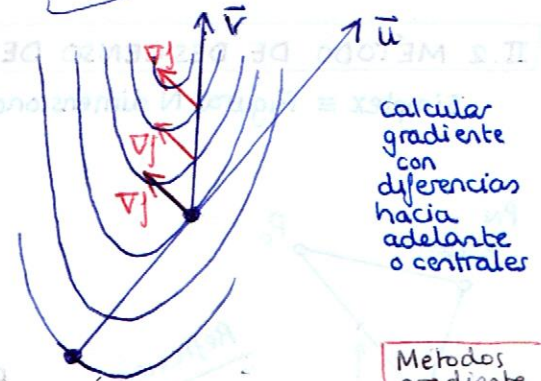
también zig-zag



- Gradientes conjugados

- Avanzar en $\vec{u} = -\nabla f$
- Avanzar en $\vec{v}$ tal que ∇f se mantiene perpendicular a $\vec{u}$ (para no entroprear lo que avanzamos en $\vec{u}$)

- $\vec{u}$ y $\vec{v}$ se llaman direcciones conjugadas
- En N dimensiones, se llega al mínimo de una $f(\vec{x})$ cuadrática tras avanzar en las N direcciones conjugadas



calcular gradiente con diferencias hacia adelante o centrales

Métodos gradiente
↓
util cuando estamos cerca

¿cómo obtenemos direcciones conjugadas?

- Algoritmo de Fletcher-Reeves (no necesita calcular Hessiana)
- Ligera modificación de Polak y Ribiere

- Método Newton (no se usa porque requiere calcular Hessiana)

$$f(\vec{x} + \Delta \vec{x}) \approx f(\vec{x}) + \underbrace{\nabla f(\vec{x})^T}_{\text{gradiente}} \cdot \Delta \vec{x} + \frac{1}{2} \Delta \vec{x}^T \cdot \underbrace{H(\vec{x})}_{\text{Hessiana}} \cdot \Delta \vec{x} \quad (\text{Taylor de orden 2})$$

paraboloide

El óptimo del paraboloide es: $\Delta \vec{x}^* = -H^{-1} \nabla f(\vec{x})$

- Métodos Quasi-Newton: aproximan la Hessiana

- Broyden-Fletcher-Goldfarb-Shanno (BFGS) $H_0 = I, H_k \approx f(H_{k-1})$
 - Davidon-Fletcher-Powell (DFP): Fórmula de recurrencia para obtener DIRECTAMENTE la inversa $H_k^{-1} \approx f(H_{k-1}^{-1})$
- Experimentalmente funciona peor

FÓRMULA DE RECURRENCIA
aprox usando los $\nabla f(x_k)$
 $\nabla f(x_{k-1})$

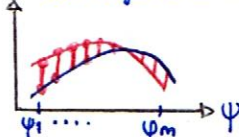
LM más rápido
Quasi-Newton más robusto

- Mínimos cuadrados (Gauss Newton)

Para $f(\vec{x})$ usa norma 2

$$f(\vec{x}) = \sum_i |k_i(\vec{x})|^2$$

vector con las diferencias



se cumple $f(\vec{x}) = \vec{K}(\vec{x})^T \vec{K}(\vec{x})$

$$\nabla f(\vec{x}) = 2 \vec{J}(\vec{x})^T \vec{K}(\vec{x})$$

Jacobiano de $K(x)$

Exigiendo $\nabla f = 0$

$$\Delta \vec{x} = -[\vec{J}(\vec{x})^T \vec{J}(\vec{x})]^{-1} \vec{J}(\vec{x})^T \vec{K}(\vec{x})$$

Búsqueda lineal tomando $\Delta \vec{x}$ como dirección de avance

Generalización Levenberg-Marquardt

$$\left[\frac{H(\vec{x}_k)}{2} + \lambda I \right] \Delta \vec{x} = -\frac{\nabla f(\vec{x}_k)}{2}$$

- si $\lambda = 0 \rightarrow$ Gauss Newton
- si $\lambda = \infty \rightarrow$ dirección de avance de máxima pendiente
- λ controla dirección; híbrida entre G-N y máxima pendiente

en realidad $\approx -H(\vec{x})^{-1} \nabla f(\vec{x})$
estamos aproximando la Hessiana.

Es como un Quasi-Newton pero sin fórmula de recurrencia (es más exacto)

II.4 MÉTODOS DE BÚSQUEDA LINEAL

cómo minimizar $g(\alpha) = f(\bar{x}_k + \alpha \bar{d}_k)$ sin gastar mucho tiempo

• Condiciones de Wolfe

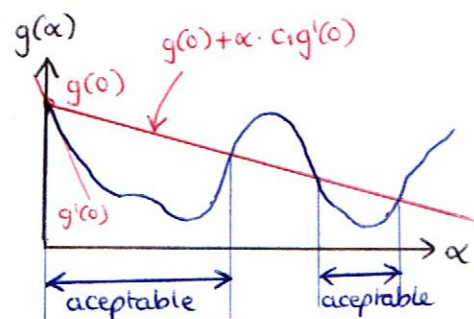
- Condición de suficiente descenso

$$f(\bar{x}_k + \alpha \bar{d}_k) \leq f(\bar{x}_k) + \alpha \cdot \underset{\substack{\uparrow \\ \text{coef} \\ [0,1]}}{C_1} \cdot \underset{\substack{\uparrow \\ \text{derivada} \\ \text{direccional}}}{\nabla f_k^T \bar{d}_k}$$

$\Downarrow$

$g(\alpha) \leq g(0) + \alpha \cdot C_1 g'(0)$

si C_1 fuera 1 sería la tangente (demasiado exigente)



- Condición de curvatura (evita avanzar poco)

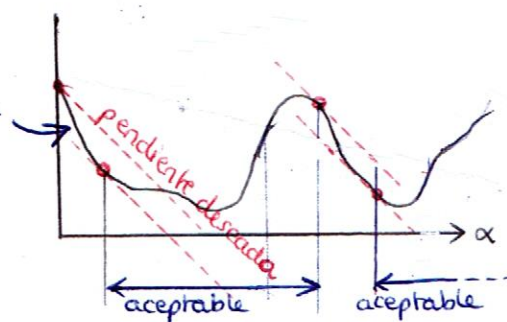
$g'(\alpha) \geq C_2 \cdot g'(0)$

o la versión estricta

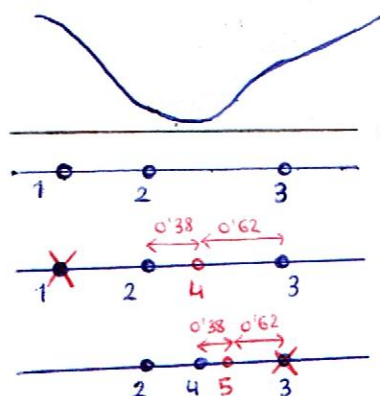
$$|g'(\alpha)| \geq C_2 g'(0)$$

No acepta pendientes tan buenas como para pararse aún

que tampoco acepta pendientes muy positivas



• Método de la sección dorada



seguro que hay un mínimo si el punto central es menor

se añade punto en intervalo grande (a 0'38 del centro)

se continúa hasta intervalo $< \sqrt{\epsilon}$

y se elimina el extremo tal que el punto central sea el menor

Utilizar la sección dorada asegura en el peor caso un intervalo 0'618 veces el anterior.

Convergencia lineal

$$\frac{\|h^{(k+1)}\|}{\|h^{(k)}\|} = 0'618...$$

(inútil buscar intervalo menor a $\sqrt{\epsilon}$ donde ϵ es la precisión en coma flotante del ordenador)

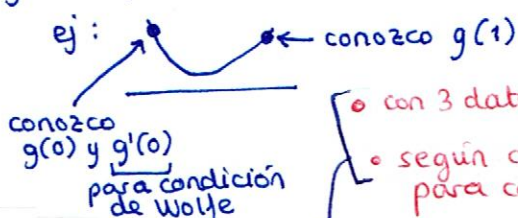
• Método de Brent

- cada iteración se aproxima $g(\alpha)$ por una parábola (3 puntos)
 - si el mínimo está en el intervalo actual, se acepta
 - si no, se usa la sección dorada para reducir el intervalo
- Convergencia más rápida y es robusto



• Interpolaciones polinómicas

$g(\alpha) \approx \text{pol. orden 2 o 3} \rightarrow$ necesitas 3 o 4 datos



con 3 datos puedo aproximar polinomio cuadrático

según calculo nuevo punto ya tengo 4 datos para calcular polinomio cúbico

$g(0), g'(0), g(\alpha_1), g(\alpha_2)$

Método LINE-SEARCH usado en quasi-Newton

- cúbica: menos iter, más error
- cuadrático-cúbica: mejor si función es suave, peor si es muy no lineal

III. ALGORITMOS DE OPTIMIZACION CON RESTRICCIONES

• Condiciones de Kuhn-Tucker

• Restricciones de igualdad

cumplir la restricción $\Leftrightarrow G_i(\bar{x}^*) = 0$ ①

seguir cumpliendo la restricción si $\bar{s}$ es una dirección de avance permitida $\Leftrightarrow G_i(\bar{x}^* + \bar{s}) = 0$
 $\approx G_i^* + \bar{s}^T \nabla G_i$

Mínimo de $f(\bar{x})$:
 En las direcciones de avance permitidas $\bar{s}$ el grad de $f(\bar{x}) > 0$ (es decir, $f(\bar{x})$ aumenta)

$$\bar{s}^T \nabla f^* = 0$$

necesario
suficiente

$$\nabla f^* = \sum_i \lambda_i^* \nabla G_i^*$$

El gradiente de la función es combinación lineal del gradiente de las funciones

$\lambda_i \equiv$ multiplicadores de Lagrange

Las condiciones ① y ② pueden escribirse:

$$\nabla L(\bar{x}^*, \bar{\lambda}^*) = 0$$

definiendo Lagrangiano

$$L(\bar{x}, \bar{\lambda}) = f(\bar{x}) - \sum \lambda_i G_i(\bar{x})$$

• Restricciones de desigualdad

Hay que añadir que para que sea un mínimo, el ∇f debe apuntar hacia dentro de la región permitida, y no hacia fuera

↓
 multiplicadores de Lagrange de las condiciones de desigualdad $\rightarrow$ positivos

Condiciones de Kuhn-Tucker

$$\nabla f^* = \sum_i \lambda_i^* \nabla G_i^*$$

$$G_i(\bar{x}) = 0 \quad i \in E \text{ (equalities)}$$

$$G_i(\bar{x}) \geq 0 \quad i \in I \text{ (inequalities)}$$

$$\lambda_i \geq 0 \quad i \in I \cap A \text{ (active inequalities)}$$

A: conjunto de restricciones activas (estamos tocando su frontera)

• Si $f(x)$ es lineal $\rightarrow$ Programación lineal

$$\min f(\bar{x}) = \bar{c}^T \bar{x} \text{ lineal}$$

solución muy sencilla
 muy utilizado en economía

$\rightarrow$ Recuerda que tenemos RESTRICCIONES

• Si $f(x)$ es cuadrática $\rightarrow$ Programación cuadrática (QP)

$$\min f(\bar{x}) = \frac{1}{2} \bar{x}^T \bar{G} \bar{x} + \bar{g}^T \bar{x}$$

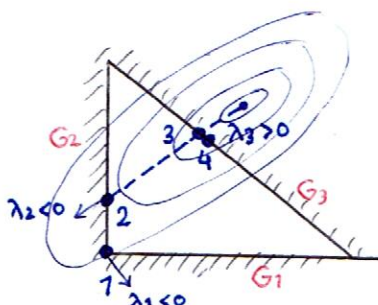
$\uparrow$ Hessiana $\uparrow$ gradiente

si existen restricciones (lineales) de desigualdad: ↓

• Método del conjunto activo

- Restricciones lineales $\bar{a}_i^T \bar{x} = 0$
- sólo de igualdad

Solución analítica con Eliminación generalizada de variables



- Busca el mínimo de la función cuadrática considerando restricciones activas como si fueran de igualdad
- Quita del conjunto activo aquellas restricciones que dan multiplicadores de Lagrange negativos
- Vuelvo a buscar el mínimo en las activas como si fueran de igualdad. Si NO QUEDAN restricciones activas busco el mínimo sin restricciones y me quedo en la frontera de la primera restricción que encuentre (haciéndola activa)
- Se continua hasta cumplir Kuhn-Tucker (multiplicador positivo)

- Para $f(x)$ cualquiera $\rightarrow$ Programación no lineal

$$\begin{array}{l} \min f(\bar{x}) \\ \text{con } G_i(\bar{x}) = 0 \quad i \in E \\ \quad \quad G_i(\bar{x}) > 0 \quad i \in I \end{array}$$

- Programación Cuadrática Secuencial (SQP)
(o método Lagrange-Newton)

- aproxima la función objetivo por función cuadrática
- aproxima las restricciones por restricciones lineales

Aplica método del conjunto activo $\rightarrow$ cada nuevo punto vuelve a aproximar la función

- Comprueba si la solución a la 'aproximación' cumple las restricciones reales o busca un punto cercano que sí lo esté
- mejora si se hace búsqueda lineal antes de aceptar un nuevo punto

$\bar{\nabla} L = 0 \iff f$ cumple restricciones
 $\uparrow$
Lagrangiano ∇f es comb. lineal del grad de las restricciones

Equivalente a exigir

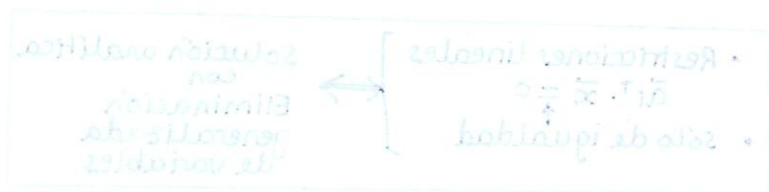
$$\min q(\bar{s}) = \frac{1}{2} \bar{s}^T \underline{W}^{(k)} \bar{s} + \bar{g}^{(k)T} \bar{s} + f^{(k)}$$

- Hessiana del Lagrangiano (se aproxima por recurrencia BFGS)

$\uparrow$ grad de la función en el punto
 $\uparrow$ valor de la función en el punto

Práctica:

- SQP cae en mínimos locales (solo bueno si empieza cerca)
- si empezamos lejos, usar la otra alternativa con restricciones: los genéticos



$$\min f(x) = \frac{1}{2} x^T Q x + g^T x$$

Restricciones lineales: $Ax = b$

Método del conjunto activo

