Análisis aplicado.
Convergencia BFGS.

28 de abril, 2015. Guillermo Santiago Novoa Pérez

1:

Broyden (algoritmo)

Tanto el algoritmo como los resultados a los problemas se encuentran en el anexo.

2:

Rayleigh (interpretación)

El cociente de Rayleigh es una función $R(x, \mathbf{M})$ que para cada vector en un espacio y para una matriz ya dada (s.p.d.) devuelve un valor en el intervalo (λ_1, λ_n) con λ_i los valores propios de $\mathbf{M}$. El cociente de Rayleigh siempre alcanza sus cotas, tanto inferior como superior, en sus respectivos vectores propios.
La fórmula del cociente de Rayleigh es:

$$R(x, \mathbf{M}) = \frac{x' \mathbf{M} x}{\|x\|^2}$$

De la hipótesis dada, es fácil obtener el cociente de Rayleigh si se divide entre $\|z\|^2$.

Interpretación.- Para toda x_k, obtenida con direcciones de descenso, su Hessiana cumplirá que su cociente de Rayleigh estará acotado entre m y M, eso quiere decir que sin importar en dónde se encuentre x_k (si se obtuvo con direcciones de descenso) su Hessiana tendrá sus valores propios dentro del intervalo (m, M). Este es el mejor resultado que se puede dar sobre las matrices Hessianas que aparecen en los puntos usados ya que no se puede dar un resultado exacto sobre el radio espectral como en los métodos pasados, sin embargo, esta cota sumada a las condiciones de descenso asegurará la convergencia del método. Más allá, esto garantiza que el problema será convexo (porque las matrices serán s.p.d.) y que existirá convergencia a un punto minimizador.

3:

Artículo Dennis Moré

El artículo abarca los siguientes temas (de manera matemática y explicativa)

- Diferencias entre métodos y familia de Broyden
- Radios de convergencia

- Actualización de las matrices de diferentes modelos
- Usos de los diferentes métodos
- El artículo se encuentra en la siguiente página: $http://www.jstor.org/stable/2029325?seq=1\#page\_scan\_tab\_contents$

4:

Problemas 6.8-6.11

6.8

$$h(t) = 1 - t + \ln(t);$$
$$h^{'}(t) = -1 + \frac{1}{t};$$
$$h''(t) = -\frac{1}{t^2};$$

Como t^2 es siempre positiva en el dominio de la función $(0, \infty)$, $-\frac{1}{t^2}$ es siempre negativa y la función $h^{'}(t)$ es estrictamente decreciente.

$$\begin{array}{rl} \forall t \in (1, \infty), & t > 1 \iff 1 > \frac{1}{t} \\ \therefore & h^{'}(t) < 0; \\ \forall t \in (0, 1), & t < 1 \iff 1 < \frac{1}{t} \\ \therefore & h^{'}(t) > 0; \end{array}$$

Por los dos hechos mencionados anteriormente, $h(t)$ tiene un único punto crítico en $t = 1$, donde $h^{'}(t) = 0$. Además, como $h^{''}(t)$ siempre es negativa, $t = 1$ es un maximizador global de la función $h(t)$, por lo que

$$0 = h(1) \geq h(t), \quad \forall t \in (0, \infty)$$

6.9

$$\begin{aligned} \psi(\mathbf{B}) = & \quad tr(\mathbf{B}) - \ln(\det(\mathbf{B})) \\ \mathbf{B} = & \quad \mathbf{P\Lambda P}^{'} \\ tr(\mathbf{B}) = & \quad tr(\mathbf{P\Lambda P}^{'}) \\ = & \quad tr(\mathbf{\Lambda P}^{'}\mathbf{P}) \end{aligned}$$

como $\mathbf{P}$ *es una matriz unitaria (*$\mathbf{B}$ *es una matriz hermitiana)*

$$\begin{aligned} \mathbf{P} = \ \mathbf{P}^{'} &= \mathbf{P^{-1}} \\ \Longrightarrow \ tr(\mathbf{B}) &= tr(\mathbf{\Lambda}) \\ \textit{Además , se tiene que: } \det(\mathbf{B}) \ &= \ \det(\mathbf{P})\det(\mathbf{\Lambda})\det(\mathbf{P}^{'}) \\ \det(\mathbf{P}) \ = \det(\mathbf{P}^{'}) \ &= \ \det(\mathbf{P^{-1}}) \ = 1, \\ \therefore \ \det(\mathbf{B}) &= \ \det(\mathbf{\Lambda}) \end{aligned}$$

$\mathbf{\Lambda}$ es una matriz diagonal, con los valores propios de $\mathbf{B}$ en la diagonal, por lo tanto es fácil calcular tanto su traza como su determinante:

$$\begin{aligned} tr(\mathbf{\Lambda}) &= \sum_{i=1}^{n} \lambda_i; \\ \det(\mathbf{\Lambda}) &= \prod_{i=1}^{n} \lambda_i; \end{aligned}$$

$$\begin{aligned} \therefore \psi(\mathbf{B}) &= \sum_{i=1}^{n} \lambda_i - \ln(\prod_{i=1}^{n} \lambda_i) \\ &= \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \ln(\lambda_i) \\ &= \sum_{i=1}^{n} \lambda_i - \ln(\lambda_i) \end{aligned}$$

Hay muchas formas de ver que λ_i siempre es mayor que $\ln(\lambda_i)$ en el dominio que estamos trabajando (como B es s.p.d. sus valores propios son reales y positivos, es decir, $\lambda_i \in (0, \infty)$), una forma de visualizarlo sería notando que:

$$x > \ln(x) \iff \exp^{x} > x$$
$$\forall x \in \mathbb{R} \backslash \{0\}$$

Por lo tanto, llegamos a la conclusión de que $\psi(\mathbf{B}) \ > \ 0 \ \forall \mathbf{B}$ *s.p.d.*

6.10

- **b)**
  Dados x,y, u, v, sea $\mathbf{Q}$ una matriz que satisface:

  $$y^{'}\mathbf{Q} = e_1^{'} \ \ , \ \ v^{'}\mathbf{Q} = e_2^{'} \ ;$$

  y sean

  $$a = \mathbf{Q^{-1}}x \ \ , \ \ b = \mathbf{Q^{-1}}u \ ;$$

  Si $\mathbf{Q}$ es de rango completo (necesario para que exista $\mathbf{Q^{-1}}$), el determinante de $\mathbf{QA}$ (A no singular) es la multiplicación de sus determinantes, entonces:

  $$\begin{aligned} \det(\mathbb{I} + xy^{'} + uv^{'}) &= \det(\ \mathbb{I} + \mathbf{Q^{-1}}xy^{'}\mathbf{Q} + \mathbf{Q{-}1}uv^{'}\mathbf{Q}) \\ &= \det(\ \mathbb{I} + ae_1^{'} + be_2^{'}) \\ &= \det([a + e_1, b + e_2, e_3, ..., e_n]) \\ \text{hacemos el determinante por menores} & \quad \text{(aprovechando la forma de la matriz)} \\ &= (1 + a_1)(1 + b_2) - a_2 b_1 \end{aligned}$$

pero como

$$\begin{aligned} a_1 &= e_1^{'}a &= e_1^{'}\mathbf{Q^{-1}}x &= y^{'}x \ ; \\ a_2 &= e_2^{'}a &= e_2^{'}\mathbf{Q^{-1}}x &= v^{'}x \ ; \\ b_2 &= e_2^{'}b &= e_2^{'}\mathbf{Q^{-1}}u &= v^{'}u \ ; \\ b_1 &= e_1^{'}b &= e_1^{'}\mathbf{Q^{-1}}u &= y^{'}u \ ; \end{aligned}$$

Por lo tanto, tenemos que $\det(\ \mathbb{I} + xy^{'} + uv^{'}) \ = (1 \ + \ y^{'}x)(1 \ + \ v^{'}u) \ - (x^{'}v)(y^{'}u)$

- **a)**
  El inciso **a)** sale inmediatamente del inciso **b)** sustituyendo $u \ = \bar{0} \ = v$.

- **c)**
  Ahora, aplicamos el inciso **b)** en la matriz $\mathbf{B}$:

$$\begin{aligned} \mathbf{B_{k+1}} &= \mathbf{B_k} - \frac{\mathbf{B_k}s_k s_k^{'}\mathbf{B_k}}{s_k^{'}\mathbf{B_k}s_k} + \frac{y_k y_k^{'}}{y_k^{'}s_k} \\ &= \mathbf{B_k}\left[\ \mathbb{I} - \frac{s_k s_k^{'}\mathbf{B_k}}{s_k^{'}\mathbf{B}s_k} + \frac{\mathbf{B_k^{-1}}y_k t_k^{'}}{y_k^{'}s_k}\right] \\ \Longrightarrow \ \det(\mathbf{B_{k+1}}) &= \det\left(\mathbf{B_k}\left[\ \mathbb{I} - \frac{s_k s_k^{'}\mathbf{B_k}}{s_k^{'}\mathbf{B}s_k} + \frac{\mathbf{B_k^{-1}}y_k t_k^{'}}{y_k^{'}s_k}\right]\right) \\ &= \det(\mathbf{B_k})\det\left(\left[\ \mathbb{I} - \frac{s_k s_k^{'}\mathbf{B_k}}{s_k^{'}\mathbf{B}s_k} + \frac{\mathbf{B_k^{-1}}y_k t_k^{'}}{y_k^{'}s_k}\right]\right) \end{aligned}$$

por el resultado anterior

$$= \ \det(\mathbf{B_k})\left[\left(1 \ - \frac{(\mathbf{B_k}s_k)^{'}}{s_k^{'}\mathbf{B_k}s_k}s_k\right)\left(1 \ + \frac{y_k^{'}}{y_k^{'}s_k}\mathbf{B_k^{-1}}y_k\right) \ - \left(-s_k^{'}\frac{y_k}{y_k^{'}s_k}\right)\left(\frac{(\mathbf{B_k}s_k)^{'}}{s_k^{'}\mathbf{B_k}s_k}\mathbf{B_k^{-1}}y_k\right)\right] \ ;$$

El primer término de la resta (dentro del corchete) se elimina y del segundo sólo queda la segunda parte, por lo que al final, nos queda la siguiente expresión:

$$\det(\mathbf{B_{k+1}}) \ = \det(\mathbf{B_k})\left[\frac{s_k^{'}y_k s_k^{'}\mathbf{B_k}\mathbf{B_k^{-1}}y_k}{y_k^{'}s_k s_k^{'}\mathbf{B_k}s_k}\right]$$

Como $s_k^{'}y_k \ = \ y_k^{'}s_k$ porque son reales, y porque dentro queda una identidad , la expresión simplificada queda de la siguiente manera:

$$\det(\mathbf{B_{k+1}}) = \det(\mathbf{B_k})\frac{y_k^{'}s_k}{s_k^{'}\mathbf{B_k}s_k} \qquad \square$$

6.11
Algunas propiedades útiles de las trazas que nos pueden servir son:

$$\begin{aligned} tr(\mathbf{A}+\mathbf{B}) &= tr(\mathbf{A})+tr(\mathbf{B}) \\ tr(\mathbf{A}\mathbf{B}^{'}) &= tr(\mathbf{B}^{'}\mathbf{A}) \\ tr(r\mathbf{A}) &= r * tr(\mathbf{A}) \quad \text{siendo r un escalar} \end{aligned}$$

Ahora, se intentará calcular la traza de **B** de una manera más econónimca:

$$\begin{aligned} tr(\mathbf{B_{k+1}}) &= tr\left(\mathbf{B_k} - \frac{\mathbf{B_k} s_k s_k^{'} \mathbf{B_k}}{s_k^{'}\mathbf{B_k} s_k} + \frac{y_k y_k^{'}}{y_k^{'} s_k}\right) \\ &= tr(\mathbf{B_k}) - tr\left(\frac{\mathbf{B_k} s_k s_k^{'} \mathbf{B_k}}{s_k^{'}\mathbf{B_k} s_k}\right) + tr\left(\frac{y_k y_k^{'}}{y_k^{'} s_k}\right) \\ &= tr(\mathbf{B_k}) - \frac{1}{s_k^{'}\mathbf{B_k} s_k} tr(\mathbf{B_k} s_k s_k^{'} \mathbf{B_k}) + \frac{y_k y_k^{'}}{y_k^{'} s_k} \\ &= tr(\mathbf{B_k}) - \frac{s_k^{'}\mathbf{B_k}\mathbf{B_k}\mathbf{s_k}}{s_k^{'}\mathbf{B_k} s_k} + \frac{y_k y_k^{'}}{y_k^{'} s_k} \\ &= tr(\mathbf{B_k}) - \frac{\|\mathbf{B_k}\|^2}{s_k^{'}\mathbf{B_k} s_k} + \frac{\|y_k\|^2}{y_k^{'} s_k} \qquad \square \end{aligned}$$

5:

Teorema

$$\|G(x) - G(x^*)\| \le L\|x - x^*\|$$

$$\tilde{s_k} = G_*^{1/2} s_k, \quad \tilde{y_k} = G_*^{-1/2} y_k, \quad \tilde{B_k} = G_*^{-1/2} B_k G_*^{1/2},$$

$$cos\tilde{\theta_k} = \frac{\tilde{s_k}^T \tilde{B_k} \tilde{s_k}}{\|\tilde{s_k}\| \|\tilde{B_k}\tilde{s_k}\|}, \quad \tilde{q_k} = \frac{\tilde{s_k}^T \tilde{B_k} \tilde{s_k}}{\|\tilde{s_k}\|^2}$$

$$\tilde{M_k} = \frac{\|\tilde{y_k}\|^2}{\tilde{y_k}^T \tilde{s_k}}, \quad \tilde{m_k} = \frac{\tilde{y_k}^T \tilde{s_k}}{\tilde{s_k}^T \tilde{s_k}}.$$

$$\tilde{B_{k+1}} = \tilde{B_k} - \frac{\tilde{B_k}\tilde{s_k}\tilde{s_k}^T\tilde{B_k}}{\tilde{s_k}^T\tilde{B_k}\tilde{s_k}} + \frac{\tilde{y_k}\tilde{y_k}^T}{\tilde{y_k}^T\tilde{s_k}}.$$

$$\begin{aligned} \psi(\tilde{B_{k+1}}) = {} & \psi(\tilde{B_k} + (\tilde{M_{k+1}} - \ln(\tilde{m_k}) - 1) \\ & + \left[1 - \frac{\tilde{q_k}}{cos^2\tilde{\theta_k}} + \ln\left(\frac{\tilde{q_k}}{cos^2\tilde{\theta_k}}\right)\right] \\ & + \ln\left(cos^2\tilde{\theta_k}\right) \end{aligned}$$

$$y_k - G_* s_k = (\tilde{G}_k - \tilde{G}_*) s_k,$$
$$\tilde{y_k} - \tilde{s_k} = G_*^{-1/2} (\tilde{G}_k - \tilde{G}_*) G_*^{-1/2} \tilde{s_k}.$$
$$\|\tilde{y_k} - \tilde{s_k}\| \leq \|G_*^{-1/2}\|^2 \|\tilde{s_k}\| \|\tilde{G}_k - \tilde{G}_*\| \leq \|G_*^{-1/2}\|^2 \|\tilde{s_k}\| L \epsilon_k,$$
$$\epsilon_k = max\{\|x_{k+1} - x^*\|, \|x_k - x^*\|\}.$$

$$\frac{\|\tilde{y_k} - \tilde{s_k}\|}{\|\tilde{s_k}\|} \leq \bar{c}\epsilon_k,$$

$$\|\tilde{y_k} - \tilde{s_k}\| \leq \bar{c}\epsilon_k \|\tilde{s_k}\|, \quad \|\tilde{s_k} - \tilde{y_k}\| \leq \bar{c}\epsilon_k \|\tilde{s_k}\|$$

$$(1 - \bar{c}\epsilon_k)\|\tilde{s_k}\| \leq \|\tilde{y_k}\| \leq (1 + \bar{c}\epsilon_k)\|\tilde{s_k}\|$$

$$(1 - \bar{c}\epsilon_k)^2 \|\tilde{s_k}\|^2 - 2\tilde{y_k}^T \tilde{s_k} + \|\tilde{s_k}\|^2 \leq \|\tilde{y_k}\|^2 - 2\tilde{y_k}^T \tilde{s_k} + \|\tilde{s_k}\|^2 \leq \bar{c}^2 \epsilon_k^2 \|\tilde{s_k}\|^2,$$
$$2\tilde{y_k}^T \tilde{s_k} \geq (1 - 2\bar{c}\epsilon_k + \bar{c}^2\epsilon_k^2 + 1 - \bar{c}^2\epsilon_k^2)\|\tilde{s_k}\|^2 = 2(1 - \bar{c}\epsilon_k)\|\tilde{s_k}\|^2$$

$$\tilde{m_k} = \frac{\tilde{y_k}^T \tilde{s_k}}{\|\tilde{s_k}\|^2} \geq 1 - \bar{c}\epsilon_k$$

$$\tilde{M_k} = \frac{\|\tilde{y_k}\|^2}{\tilde{y_k}^T \tilde{s_k}} \leq \frac{1 + \bar{c}\epsilon_k}{1 - \bar{c}\epsilon_k}$$

$$\tilde{M_k} \leq 1 + \frac{2\bar{c}}{1 - \bar{c}\epsilon_k}\epsilon_k \leq 1 + \bar{c}\epsilon_k$$

$$\frac{-x}{1-x} - \ln(1-x) = h\left(\frac{1}{1-x}\right) \leq 0$$
$$\ln(1 - \bar{c}\epsilon_k) \geq \frac{-\bar{c}\epsilon_k}{1 - \bar{c}\epsilon_k} \geq -2\bar{c}\epsilon_k$$

$$\ln(\tilde{m_k}) \geq \ln(1 - \bar{c}\epsilon_k) \geq -2\bar{c}\epsilon_k \geq -2c\epsilon_k$$

$$0 < \psi(\tilde{B_{k+1}}) \leq \psi(\tilde{B}_k) + 3c\epsilon_k + \ln(cos^2\tilde{\theta}_k) + \left[1 - \frac{\tilde{q}_k}{cos^2\tilde{\theta}_k} + \ln\left(\frac{\tilde{q}_k}{cos^2\tilde{\theta}_k}\right)\right]$$
$$\sum_{j=0}^{\infty}\left(\ln\left(\frac{1}{cos^2\tilde{\theta}_j}\right) - \left[1 - \frac{\tilde{q}_j}{cos^2\tilde{\theta}_j} + \ln\left(\frac{\tilde{q}_j}{cos^2\tilde{\theta}_j}\right)\right]\right) \leq \psi(\tilde{B}_0) + 3c\sum_{j=0}^{\infty}\epsilon_j < +\infty$$
$$\lim_{j\to\infty} \ln\left(\frac{1}{cos^2\tilde{\theta}_j}\right) = 0, \qquad \lim_{j\to\infty}\left(1 - \frac{\tilde{q}_j}{cos^2\tilde{\theta}_j} + \ln\left(\frac{\tilde{q}_j}{cos^2\tilde{\theta}_j}\right)\right) = 0$$

$$\lim_{j\to\infty} cos\tilde{\theta}_j = 1, \qquad \lim_{j\to\infty} \tilde{q}_j = 1$$

$$\begin{aligned}
\frac{\|G_*^{-1/2}(B_k - G_*)s_k\|^2}{\|G_*^{1/2}s_k\|} &= \frac{\|(\tilde{B}_k - I)\tilde{s}_k\|^2}{\|\tilde{s}_k\|^2} \\
&= \frac{\|\tilde{B}_k\tilde{s}_k\|^2 - 2\tilde{s}_k^T\tilde{B}_k\tilde{s}_k + \tilde{s}_k^T\tilde{s}_k}{\tilde{s}_k^T\tilde{s}_k} \\
&= \frac{\tilde{q}_k^2}{cos\tilde{\theta}_k^2} - 2\tilde{q}_k + 1.
\end{aligned}$$

$$\therefore \quad \lim_{k\to\infty} \frac{\|(B_k - G_*)s_k\|}{\|s_k\|} = 0 \qquad \square$$