Proyecto 1 - Análisis Aplicado
Máximo descenso en caso cuadrático

José Manuel Proudinat Silva
130056

13 de febrero de 2015

1. Introducción

Este proyecto tiene como finalidad estudiar algunas propiedades del método de máximo descenso con búsqueda lineal exacta. Hacemos una exploración teórica y la complementamos con resultados prácticos que ejemplifican la teoría estudiada.

La estructura de este reporte es la siguiente: Comenzamos haciendo mención de un teorema sobre la convergencia del método mencionames, el cual demostraremos detalladamente. Posteriormente se muestran resultados de simulaciones hechas en Matlab que refuerzan nuestro conocimiento.

2. Una propiedad sobre la convergencia del método

Teorema 1. $\forall x_0 \in \mathbb{R}^n$ *el método de máximo descenso que alcanza su mínimo valor en el punto* x^* *de* f *(con* f *una cuadrática estrictamente convexa). Satisface la siguiente propiedad:*

$$||x_{k+1} - x^*||_H^2 \leq \left[\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max}}\right]^2 ||x_k - x^*||_H^2.$$

Para poder realizar la pruba de manera detallada, antes debemos enunciar dos lemas, que también demostraremos, que nos serán muy útiles para lograr nuestro objetivo.

Lema 1. *El proceso iterativo del algoritmo de máximo descenso cumple lo siguiente:*

$$||x_{k+1}-x^*||_H^2 = \left[1 - \frac{\left(\nabla f(x_k)^T \nabla f(x_k)\right)^2}{\left(\nabla f(x_k)^T H \nabla f(x_k)\right)\left(\nabla f(x_k)^T H^{-1} \nabla f(x_k)\right)}\right] ||x_k - x^*||_H^2$$

Prueba. *Haremos la prueba por cálculo directo. Sea $y_k = x_k - x^*$, entonces:*

$$\begin{aligned}
&||x_k - x^*||_H^2 - ||x_{k+1} - x^*||_H^2 = ||x_k - x^*||_H^2 - ||x_k - \alpha_k \nabla f(x_k) - x^*||_H^2 \\
&= y_k^T H y_k - \left(x_k - \alpha_k \nabla f(x_k) - x^*\right)^T H \left(x_k - \alpha_k \nabla f(x_k) - x^*\right) \\
&= y_k^T H y_k - \left(y_k - \alpha_k \nabla f(x_k)\right)^T H \left(y_k - \alpha_k \nabla f(x_k)\right) \\
&= y_k^T H y_k - \left(y_k^T H y_k - \alpha_k y_k^T H \nabla f(x_k) - \alpha_k \nabla f(x_k)^T H y_k + \alpha_k^2 \nabla f(x_k)^T H \nabla f(x_k)\right) \\
&= 2\alpha_k \nabla f(x_k)^T H y_k - \alpha_k^2 \nabla f(x_k)^T H \nabla f(x_k).
\end{aligned}$$

Además, $Hy_k = H(x_k - x^*) = Hx_k - Hx_k^* = Hx_k + H(H^{-1}g) = Hx + g = \nabla f(x_k)$. Por lo tanto, $y_k = H^{-1}\nabla f(x_k)$. Con esto y lo anterior obtenemos la siguiente igualdad:

$$\begin{aligned}
&\frac{||x_k - x^*||_H^2 - ||x_{k+1} - x^*||_H^2}{||x_k - x^*||_H^2} = \frac{2\alpha_k \nabla f(x_k)^T H y_k - \alpha_k^2 \nabla f(x_k)^T H \nabla f(x_k)}{y_k^T H y_k} \\
&= \frac{2\alpha_k \nabla f(x_k)^T \nabla f(x_k) - \alpha_k^2 \nabla f(x_k)^T H \nabla f(x_k)}{\left(H^{-1} \nabla f(x_k)\right)^T \nabla f(x_k)}
\end{aligned}$$

Como $\alpha = \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T H \nabla f(x_k)}$ se sigue que:

$$\begin{aligned}
&= \frac{\frac{2\left(\nabla f(x_k)^T \nabla f(x_k)\right)^2 - \left(\nabla f(x_k)^T \nabla f(x_k)\right)^2}{\nabla f(x_k)^T H \nabla f(x_k)}}{\nabla f(x_k)^T H^{-1} \nabla f(x_k)} \\
&= \frac{\left(\nabla f(x_k)^T \nabla f(x_k)\right)^2}{\left(\nabla f(x_k)^T H \nabla f(x_k)\right)\left(\nabla f(x_k)^T H^{-1} \nabla f(x_k)\right)}
\end{aligned}$$

Haciendo un simple dspeje llegamos a que:

$$||x_{k+1}-x^*||_H^2 = \left[1 - \frac{\left(\nabla f(x_k)^T \nabla f(x_k)\right)^2}{\left(\nabla f(x_k)^T H \nabla f(x_k)\right)\left(\nabla f(x_k)^T H^{-1} \nabla f(x_k)\right)}\right] ||x_k - x^*||_H^2.$$

□

Lema 2. ***(Desigualdad de Kantorovich)***
Sea H una matriz n x n simétrica positiva definida. $\forall x$ se tiene la siguiente desigualdad:

$$\frac{\left(x^T x\right)^2}{\left(x^T H x\right)\left(x^T H^{-1} x\right)} \geq \frac{4\lambda_{min}\lambda_{max}}{\left(\lambda_{min}+\lambda_{max}\right)^2}$$

donde λ_{min} y λ_{max} son el mínimo y máximo valor propio de H, respectivamente.

Prueba. *Factorizamos H en su descomposición espectral $UDU^T = H$ con U una matriz ortonormal y D una matriz diagonal cuyos valores son los eigenvalores de H en orden ascendente. Entonces tomamos $y = U^T x$. Así, obtenemos la siguiente igualdad:*

$$\begin{aligned}
\frac{\left(x^T x\right)^2}{\left(x^T H x\right)\left(x^T H^{-1} x\right)} &= \frac{\left(x^T\left(UU^T\right)x\right)^2}{\left(x^T\left(UDU^T\right)x\right)\left(x^T\left(UDU^T\right)^{-1}x\right)} \\
=&\frac{\left(y^T y\right)^2}{\left(y^T D y\right)\left(y^T D^{-1} y\right)} = \frac{\left(\sum y_i^2\right)^2}{\left(\sum y_i^2\lambda_i\right)\left(\sum y_i^2/\lambda_i\right)} \\
=&\frac{\left(\sum y_i^2\right)}{\left(\sum y_i^2\lambda_i\right)}\frac{\left(\sum y_i^2\right)}{\left(\sum y_i^2/\lambda_i\right)} = \frac{1/\sum(y_i^2/\sum y_i^2)\lambda_i}{\sum\left((y_i^2/\sum y_i^2)/\lambda_i\right)}
\end{aligned}$$

Lo cual es una función convexa, por lo tanto podemos buscar un mínimo. Buscando el valor que minimiza la función llegamos a que se da en:

$$\frac{4\lambda_{max}\lambda_{min}}{\left(\lambda_{max}+\lambda_{min}\right)^2}.$$

Por lo tanto:

$$\frac{\left(x^T x\right)^2}{\left(x^T H x\right)\left(x^T H^{-1} x\right)} \geq \frac{4\lambda_{min}\lambda_{max}}{\left(\lambda_{min}+\lambda_{max}\right)^2}$$

$\square$

Ya con los dos lemas demostrados, podemos continuar con la demostración del teorema.

Prueba. *Por el lema 1, tenemos:*

$$||x_{k+1} - x^*||_H^2 = \left[1 - \frac{\left(\nabla f(x_k)^T \nabla f(x_k)\right)^2}{\left(\nabla f(x_k)^T H \nabla f(x_k)\right)\left(\nabla f(x_k)^T H^{-1} \nabla f(x_k)\right)}\right] ||x_k - x^*||_H^2 = (*)$$

Por el lema 2, se sigue:

$$\begin{aligned}
(*) &\leq \left[1 - \frac{4\lambda_{max}\lambda_{min}}{(\lambda_{max} + \lambda_{min})^2}\right] ||x_k - x^*||_H^2 \\
&= \left[\frac{\lambda_{max}^2 + \lambda_{min}^2 + 2\lambda_{max}\lambda_{min} - 4\lambda_{max}\lambda_{min}}{(\lambda_{max} + \lambda_{min})^2}\right] ||x_k - x^*||_H^2 \\
&= \left[\frac{\lambda_{max} - \lambda_{min}}{(\lambda_{max} + \lambda_{min})^2}\right] ||x_k - x^*||_H^2
\end{aligned}$$

□

3. Implementación práctica

Para hacer un análisis numérico de los resultados teóricos hicimos simulaciones numéricas que refuerzan nuestro estudio. Se construyeron funciones de la forma:

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

tomando $g = -He$ para asegurar que $x^* = e$ donde e es el vector de unos.

Se hicieron pruebas con matrices de tamaño 10 y números de condición 10^1, 10^2 y 10^3. Se predijo el número de pasos que le llevaría al método converger, tomando como tolerancia 10^{-5} para la norma del gradiente. En cada paso de la iteración se calcularon ambos lados de la desigualdad de nuestro teorema y se verifica que se cumpla. También se prueba la igualdad tomando como tolerancia 10^{-3}. Se imprimen algunos resultados.

La predicción del número de pasos se hizo del siguiente modo:

Sea $\Lambda = \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max}}$, entonces, por el teorema demostrado:

$$\begin{aligned} ||x_{k+1} - x^*||_H^2 &\leq \Lambda ||x_k - x^*||_H^2 \\ \to ||x_{k+1} - x^*||_H^2 &\approx \Lambda ||x_k - x^*||_H^2 \end{aligned}$$

Sea $f_k = ||x_k - x^*||_H^2$. Entonces se tiene:

$$\begin{aligned} h_1 &= \Lambda h_0 \\ h_2 = \Lambda h_1 &= \Lambda^2 h_0 \\ \to h_n &= \Lambda^n h_0 \end{aligned}$$

Si buscamos $f_n < tol$, despjamos:

$$\begin{aligned} h_n = \Lambda^n h_0 &< tol \\ \to n \log(\Lambda) + \log(h_0) &< \log(tol) \\ \to n &< \frac{\log(tol) - \log(h_0)}{\log(\Lambda)} \end{aligned}$$

Entonces podemos tomar:

$$n = \lceil \frac{\log(tol) - \log(h_0)}{\log(\Lambda)} \rceil$$

4. Resultados computacionales

En esta sección mostramos los resultados obtenidos con la implementación en MATLAB mencionada anteriormente.

Figura 1: Resultados con número de condición 10^1

```
==============================================================
Resultados para el experimento con condicion: 1.0e+01
Predecimos que llegaremos al resultado en:   35 pasos

    iter     lado izq       lado der     desigualdad    igualdad

      1     1.214e+00     6.039e+00          1             0
      5     4.006e-02     4.366e-02          1             0
     10     5.178e-03     5.180e-03          1             1
     15     6.949e-04     6.951e-04          1             1
     20     9.325e-05     9.328e-05          1             1
     25     1.251e-05     1.252e-05          1             1
     30     1.680e-06     1.680e-06          1             1
     35     2.254e-07     2.255e-07          1             1
     40     3.025e-08     3.026e-08          1             1
     45     4.060e-09     4.061e-09          1             1
     50     5.448e-10     5.450e-10          1             1
     52     2.440e-10     2.441e-10          1             1

El valor del gradiente de la funcion fue 8.038e-06
con   52 iteraciones
==============================================================
```

Figura 2: Resultados con número de condición 10^2

```
==============================================================
Resultados para el experimento con condicion: 1.0e+02
Predecimos que llegaremos al resultado en:  332 pasos

    iter     lado izq       lado der      desigualdad     igualdad

      1      1.409e+00     5.440e+00          1              0
     15      1.233e-02     1.289e-02          1              1
     30      6.086e-03     6.090e-03          1              1
     45      3.315e-03     3.317e-03          1              1
     60      1.806e-03     1.807e-03          1              1
     75      9.840e-04     9.845e-04          1              1
     90      5.361e-04     5.364e-04          1              1
    105      2.921e-04     2.922e-04          1              1
    120      1.592e-04     1.592e-04          1              1
    135      8.671e-05     8.675e-05          1              1
    150      4.724e-05     4.727e-05          1              1
    165      2.574e-05     2.575e-05          1              1
    180      1.402e-05     1.403e-05          1              1
    195      7.641e-06     7.644e-06          1              1
    210      4.163e-06     4.165e-06          1              1
    225      2.268e-06     2.269e-06          1              1
    240      1.236e-06     1.236e-06          1              1
    255      6.733e-07     6.736e-07          1              1
    270      3.668e-07     3.670e-07          1              1
    285      1.999e-07     2.000e-07          1              1
    300      1.089e-07     1.089e-07          1              1
    315      5.933e-08     5.935e-08          1              1
    330      3.232e-08     3.234e-08          1              1
    345      1.761e-08     1.762e-08          1              1
    360      9.595e-09     9.599e-09          1              1
    375      5.228e-09     5.230e-09          1              1
    376      5.020e-09     5.023e-09          1              1

El valor del gradiente de la funcion fue 9.665e-06
con  376 iteraciones
==============================================================
```

Figura 3: Resultados con número de condición 10^3

```
==========================================================
Resultados para el experimento con condicion: 1.0e+03
Predecimos que llegaremos al resultado en: 3232 pasos

    iter    lado izq      lado der     desigualdad    igualdad

       1    1.086e+00     4.090e+00         1            0
     100    1.251e-03     1.252e-03         1            1
     200    8.208e-04     8.210e-04         1            1
     300    5.386e-04     5.387e-04         1            1
     400    3.534e-04     3.534e-04         1            1
     500    2.319e-04     2.319e-04         1            1
     600    1.521e-04     1.522e-04         1            1
     700    9.981e-05     9.983e-05         1            1
     800    6.549e-05     6.550e-05         1            1
     900    4.297e-05     4.298e-05         1            1
    1000    2.819e-05     2.820e-05         1            1
    1100    1.850e-05     1.850e-05         1            1
    1200    1.214e-05     1.214e-05         1            1
    1300    7.964e-06     7.965e-06         1            1
    1400    5.225e-06     5.226e-06         1            1
    1500    3.428e-06     3.429e-06         1            1
    1600    2.249e-06     2.250e-06         1            1
    1700    1.476e-06     1.476e-06         1            1
    1800    9.684e-07     9.686e-07         1            1
    1900    6.354e-07     6.355e-07         1            1
    2000    4.169e-07     4.170e-07         1            1
    2100    2.735e-07     2.736e-07         1            1
    2200    1.795e-07     1.795e-07         1            1
    2300    1.178e-07     1.178e-07         1            1
    2400    7.726e-08     7.728e-08         1            1
    2456    6.102e-08     6.103e-08         1            1

El valor del gradiente de la funcion fue 9.997e-06
con 2456 iteraciones
==========================================================
```

5. Conclusiones

A través de este proyecto logré desarrollar un método científico de trabajo en el cual se complementa la teoría matemático con la implementación cmputacional. De esta manera se refuerza el conocimiento que se genera.

Más específico, hablando de los resultados obtenidos, podemos notar muchos puntos importantes:

- La predicción en el número de pasos no es exacta, pero sí del mismo orden.
- La desigualdad siempre se cumple, como lo dice el teorema.
- La igualdad, dada la tolerancia especificada, se dio en la mayoría de los casos. Esto hace que la predicción que hicimos del número de casos sea muy aproximada.
- Es evidente como el número de condición de la matriz afecta el tiempo de convergencia de nuestro algoritmo.