

Optimización Tarea 3

Francisco Javier Peralta Ramírez

1. Rosembrock, n = 2

$$f(\mathbf{x}) = [100(x_2 - x_1^2)^2 + (1 - x_1)^2]$$

$$\mathbf{x}^0 = [-1.2, 1]^T$$

$$\mathbf{x}^* = [1, 1]^T$$

$$f(\mathbf{x}^*) = 0$$

Calculamos el gradiente y el Hessiano:

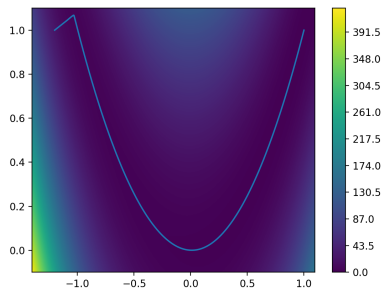
$$\nabla f(\mathbf{x}) = \begin{pmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} -400(x_2 - 3x_1^2) + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}$$

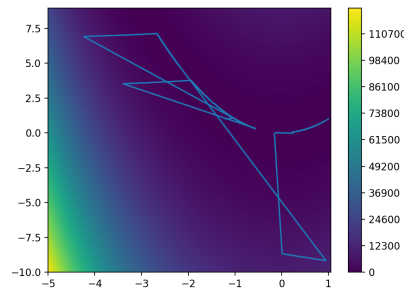
Probando el algoritmo de desenso más rápido con tres diferentes tipos de paso y la \mathbf{x} establecida previamente, podemos ver que la velocidad de convergencia varía y es mejor cuando tenemos el Hessiano para poder calcular los tamaños de paso.

Paso Fijo				Paso Hessiano				Paso Aproximado			
k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$
1	0.0745391	0.116434	6.59445	1	0.0990876	0.15478	4.56778	1	0.106279	0.166013	4.28974
5500	8.86e-05	9.38675e-05	0.032541	1500	0.000113848	0.00081548	12.8452	2500	8.0195e-05	0.000604121	13.3554
11000	1.75155e-05	2.2972e-05	0.00243845	3000	0.000159203	0.000986846	11.4966	5000	0.000124068	0.000765075	11.4513
16500	5.11592e-06	7.06653e-06	0.000243791	4500	5.12876e-05	6.99271e-05	0.000583172	7500	0.000270828	0.0011703	8.76884
22000	1.63728e-06	2.2977e-06	2.62132e-05	6000	5.11522e-06	7.20339e-06	7.964e-06	10000	0.00092844	0.00104037	0.0215924
27500	5.38665e-07	7.59848e-07	2.88247e-06	7500	7.22907e-07	1.02172e-06	1.64338e-07	12500	1.72065e-05	2.41432e-05	2.72431e-05
33000	1.78811e-07	2.52662e-07	3.19285e-07	9000	1.06039e-07	1.49948e-07	3.5523e-09	15000	9.23591e-07	1.30559e-06	8.19097e-08

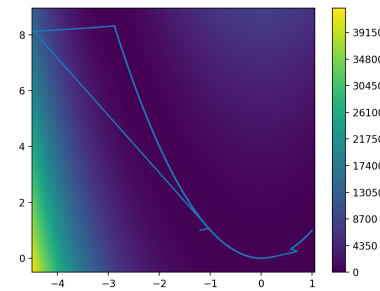
Podemos ver como va caminando nuestro algoritmo con los diferentes tipos de paso, cabe resaltar que el comportamiento con el paso fijo es muy tranquilo, mientras que es más caótico con el paso Hessiano. Aún así debemos recordar que con el paso Hessiano nuestro algoritmo convergió mucho más rápido.



(a) Paso Fijo



(b) Paso Hessiano



(c) Paso Aproximado

Si probamos con un ponto lejano al óptimo, notamos aún más los beneficios de un paso no fijo. Para el punto inicial $\mathbf{x} = [-5, 2]$ el algoritmo no converge y se va a infinito.

Paso Fijo				Paso Hessiano				Paso Aproximado			
k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$
1	4.2934	23.1207	1.02349e+07	1	0.292956	1.57761	9253.7	1	0.203989	1.09852	17329.1
400	nan	nan	nan	400	0.00076728	0.000879229	0.0179082	2000	0.000198466	0.000260206	0.00245417
800	nan	nan	nan	800	5.93796e-06	8.37716e-06	2.6149e-06	4000	1.38498e-05	1.94625e-05	1.7808e-05
1200	nan	nan	nan	1200	4.75875e-07	6.72854e-07	1.7679e-08	6000	1.33613e-06	1.8884e-06	1.71276e-07
1600	nan	nan	nan	1600	4.06735e-08	5.752e-08	1.29688e-10	8000	1.32417e-07	1.87254e-07	1.68764e-09

2. Rosembrock, n = 100

$$f(\mathbf{x}) = \sum_{i=0}^{n-1} [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$$

$$\mathbf{x}^0 = [-1.2, 1, 1, \dots, 1, -1.2, 1]^T$$

$$\mathbf{x}^* = [1, 1, \dots, 1, 1]^T$$

$$f(\mathbf{x}^*) = 0$$

Calculamos el gradiente y el Hessiano:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) - 400(x_3 - x_2^2)x_2 - 2(1 - x_2) \\ \vdots \\ 200(x_i - x_{i-1}^2) - 400(x_{i+1} - x_i^2)x_i - 2(1 - x_i) \\ \vdots \\ 200(x_n - x_{n-1}^2) \end{pmatrix}$$

Notamos que para el Hessiano se forma una matriz tridiagonal, esto se debe a que la función tiene partes que dependen de x_i y x_{i+n} $n \geq 2$

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} -400(x_2 - 3x_1^2) + 2 & -400x_1 & 0 & \dots & 0 \\ -400x_1 & 200(1 - 2x_3 + 6x_2^2) + 2 & -400x_2 & \dots & 0 \\ 0 & -400x_2 & \ddots & \dots & 0 \\ \vdots & \ddots & 0 & 200(1 - 2x_n + 6x_{n-1}^2) + 2 & -400x_{n-1} \\ 0 & 0 & \dots & -400x_{n-1} & 200 \end{pmatrix}$$

Para Rosembrock con $n = 100$ volvemos a notar un comportamiento similar a Rosembrock con $n = 2$, con un tamaño de paso fijo el algoritmo converge lentamente, sin embargo, con el paso Hessiano y el aproximado no logramos convergencia, aun que este se detiene de manera muy rápida, esto nos indica que se decrementa mucho el paso antes de acercarnos suficiente al óptimo global que buscamos. Cabe resaltar que aun cuando el algoritmo hace menos iteraciones con el paso Hessiano, este tarda considerablemente más, esto se debe a la cantidad de operaciones que se deben de realizar para calcular tamaño del paso.

Paso Fijo				Paso Hessiano				Paso Aproximado			
k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$
1	0.0616781	0.616473	296.051	1	0.0611243	0.610937	297.706	1	0.0735587	0.73522	268.14
5000	7.3087e-06	7.27748e-05	4.00401	2500	2.04781e-05	0.000204061	3.99782	2500	1.93153e-05	0.000192434	3.99933
10000	1.68916e-06	1.68706e-05	3.98771	5000	4.0394e-06	4.0358e-05	3.98712	5000	3.98534e-06	3.98143e-05	3.98724
15000	4.60143e-07	4.59937e-06	3.98671	7500	9.22464e-07	9.22176e-06	3.98665	7500	9.57036e-07	9.56707e-06	3.98666
20000	1.30289e-07	1.3026e-06	3.98663	10000	2.16997e-07	2.16959e-06	3.98663	10000	2.37312e-07	2.37268e-06	3.98663
25000	3.72826e-08	3.72767e-07	3.98662	12500	5.13937e-08	5.13864e-07	3.98662	12500	5.93097e-08	5.9301e-07	3.98662
30000	1.07005e-08	1.0699e-07	3.98662	15000	1.21916e-08	1.219e-07	3.98662	15000	1.48674e-08	1.48654e-07	3.98662

Al escoger un punto lejano a la solución, $\mathbf{x} = [-2, 5, 5, \dots, -2, 5]$, nuestro algoritmo vuelve a fallar para el tamaño de paso fijo, y para los otros se vuelve a estancar bastante cerca del óptimo global.

Paso Fijo				Paso Hessiano				Paso Aproximado			
k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$
1	3.61753	178.455	3.45954e+08	1	0.296013	14.6025	766138	1	0.283933	14.0066	831020
2500	nan	nan	nan	2500	1.30213e-05	0.000129842	3.99361	2500	2.22453e-05	0.000221668	3.9979
5000	nan	nan	nan	5000	1.95598e-06	1.95483e-05	3.9868	5000	4.09699e-06	4.09352e-05	3.98706
7500	nan	nan	nan	7500	3.35199e-07	3.35127e-06	3.98663	7500	8.78431e-07	8.78176e-06	3.98664
10000	nan	nan	nan	10000	5.86306e-08	5.8622e-07	3.98662	10000	1.93856e-07	1.93823e-06	3.98662
12500	nan	nan	nan	12500	1.02914e-08	1.029e-07	3.98662	12500	4.30562e-08	4.30501e-07	3.98662

3. Wood Function

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90(x_3^2 - x_4)^2 + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1)$$

$$x^0 = [-3, -1, -3, -1]^T$$

$$x^* = [1, 1, 1, 1]^T$$

$$f(x^*) = 0$$

Calculamos el gradiente y el Hessiano:

$$\nabla f(x) = \begin{pmatrix} 400(x_1^2 - x_2)x_1 + 2(x_1 - 1) \\ -200(x_1^2 - x_2) + 20.2(x_2 - 1) + 19.8(x_4 - 1) \\ 2(x_3 - 1) + 360(x_3^2 - x_4)x_3 \\ -180(x_3^2 - x_4) + 20.2(x_4 - 1) + 19.8(x_2 - 1) \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} 400(3x_1^2 - x_2) + 2 & -400x_1 & 0 & 0 \\ -400x_1 & 220.2 & 0 & 19.8 \\ 0 & 0 & 2 + 360(3x_3^2 - x_4) & -360x_3 \\ 0 & 19.8 & -360x_3 & 200.2 \end{pmatrix}$$

Para el punto inicial dado anteriormente, los tres tipos de paso logran converger con el Hessiano tardando más tiempo pero tomando menor cantidad de iteraciones.

Paso Fijo				Paso Hessiano				Paso Aproximado			
k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$
0	1.83325	8.19856	11186.7	0	0.340095	1.52095	4141.39	0	0.974667	4.35884	136.208
5000	4.46691e-05	9.13878e-05	7.81944	800	9.30097e-05	0.000178805	7.8757	1400	0.00189627	0.00436438	7.18138
10000	0.000377774	0.000773769	4.88003	1600	8.31637e-05	0.000161353	7.87002	2800	0.00210458	0.00499445	1.70461
15000	0.000185448	0.000407071	0.392146	2400	0.000126981	0.000251231	7.85546	4200	0.000324964	0.000655309	0.029588
20000	2.89583e-05	5.80689e-05	0.0093348	3200	0.000288759	0.000601028	7.7954	5600	4.04129e-05	8.08344e-05	0.000389454
25000	4.78733e-06	9.57534e-06	0.000254633	4000	0.00167884	0.00361621	2.70951	7000	4.64268e-06	9.28536e-06	5.0851e-06
30000	7.91924e-07	1.58385e-06	6.96743e-06	4800	0.00045081	0.000917375	0.0626358	8400	5.32143e-07	1.06429e-06	6.67238e-08
35000	1.31003e-07	2.62006e-07	1.90664e-07	5600	2.74602e-05	5.49228e-05	0.000128397	9800	6.09762e-08	1.21952e-07	8.75956e-10
40000	2.1671e-08	4.33421e-08	5.21752e-09	6400	1.97077e-07	3.94155e-07	5.6477e-09	10000	4.47454e-08	8.94908e-08	4.7169e-10

Cambiando el punto inicial a $x = [-8, -5, -8, -5]$ volvemos a ver que el paso de tamaño fijo no converge, mientras que los otros dos sí. Interesantemente en todos los casos ha convergido más rápido cuando se inicia más lejano.

Paso Fijo				Paso Hessiano				Paso Aproximado			
k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$
0	11.1561	148.841	1.72652e+10	0	0.296074	3.95012	196120	0	0.014279	0.190505	850835
500	nan	nan	nan	600	9.55289e-05	0.000188943	7.8552	800	0.000218264	0.000439633	0.0252332
1000	nan	nan	nan	1200	0.000140042	0.000281378	7.84048	1600	7.75321e-05	0.000155183	0.00266897
1500	nan	nan	nan	1800	0.000462785	0.000983471	7.75405	2400	2.47532e-05	4.95101e-05	0.000256096
2000	nan	nan	nan	2400	0.0015892	0.00330501	2.99611	3200	7.60088e-06	1.52019e-05	2.36915e-05
2500	nan	nan	nan	3000	0.000745953	0.00158663	0.242104	4000	2.30497e-06	4.60995e-06	2.16603e-06
3000	nan	nan	nan	3600	0.000156993	0.000314053	0.000598305	4800	6.96302e-07	1.3926e-06	1.97316e-07
3500	nan	nan	nan	4200	8.59582e-07	1.71916e-06	8.87778e-08	5600	2.10098e-07	4.20197e-07	1.79548e-08
4000	nan	nan	nan	4500	7.96644e-08	1.59329e-07	7.59762e-10	6400	6.33716e-08	1.26743e-07	1.63326e-09

4. Smoothing model

$$f(x) = \sum_{i=1}^n (x_i - y_i)^2 + \lambda \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$$

Para $\lambda \in 1, 100, 1000$

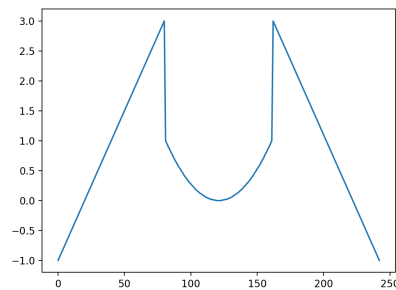
Calculamos el gradiente y el Hessiano:

$$\nabla f(x) = \begin{pmatrix} 2(x_1 - y_1) - 2\lambda(x_2 - x_1) \\ 2(x_2 - y_2) + 2\lambda(2x_2 - x_1 - x_3) \\ \vdots \\ 2(x_n - y_n) + 2\lambda(x_n - x_{n-1}) \end{pmatrix}$$

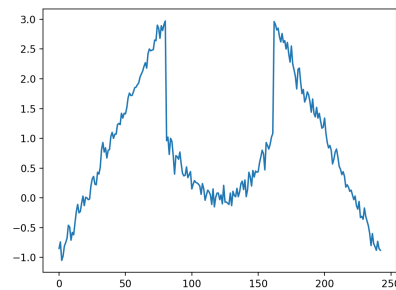
Al igual que con Rosembrock, notamos que para el Hessiano se forma una matriz tridiagonal.

$$\nabla^2 f(x) = \begin{pmatrix} 2 + 2\lambda & -2\lambda & 0 & \cdots & 0 \\ -2\lambda & 2 + 4\lambda & \ddots & \cdots & \vdots \\ 0 & \cdots & \ddots & \ddots & -2\lambda \\ 0 & \cdots & 0 & -2\lambda & 2 + 2\lambda \end{pmatrix}$$

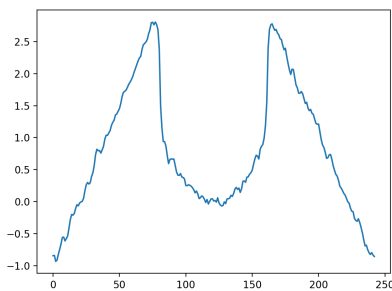
Esta función es de suavizado, donde tenemos un vector de x que queremos hacer similar a y , no queremos interpolar ya que la variable y tiene ruido, por lo que penalizamos los cambios en x con un cierto peso λ , entre más grande el peso más suave la curva.



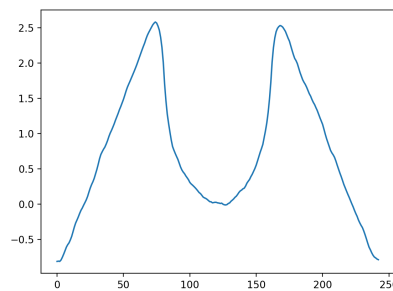
(a) x inicial



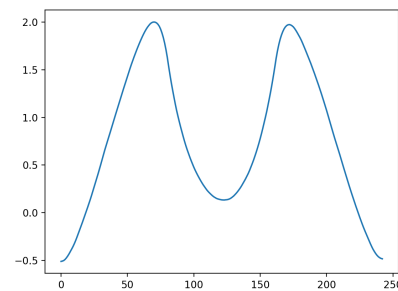
(b) y , variable a interpolar



(a) $\lambda = 1$



(b) $\lambda = 10$



(c) $\lambda = 100$

Para el caso de $\lambda = 1$ podemos ver que el tamaño de paso aproximado tiene problemas de convergencia ya que disminuye el tamaño drásticamente lo que no le permite caminar por función a la solución. Por otro lado el Hessiano tiene mucho éxito y converge muy rápido. Los resultados para las demás λ son similares, por lo que no se muestran sus tablas.

En resumen, podemos observar que el tamaño de paso es crucial para la convergencia del método. Usar un tamaño fijo pequeño no garantiza convergencia y requiere cantidades extremas de iteraciones cuando logra converger. Por otra parte una elección de tamaño de paso bien informada, como lo es el del Hessiano logra buenos resultados con convergencia en pocas iteraciones. Lamentablemente para este tamaño de paso requerimos del Hessiano de la función lo cual no siempre es posible, aun cuando lo es, este método es algo tardado en tiempo ya que cada iteración requiere de múltiples operaciones para multiplicar matriz con vector y para garantizar que el Hessiano sea definido positivo. Finalmente el tamaño de paso

Paso Fijo				Paso Hessiano				Paso Aproximado			
k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$	k	$\ x_{k+1} - x_k\ $	$\ \nabla f(x_k)\ $	$f(x_k)$
1	0.000218115	0.00436465	397.5	1	0.0577533	1.15569	392.29	0	4.29886e-10	8.60237e-09	397.531
1400	1.12102e-05	0.000221358	387.496	4	0.00370735	0.0731739	386.248	1	1.57168e-21	8.35407e-19	397.531
2800	2.0186e-06	3.97675e-05	385.794	8	0.000561169	0.0110528	385.452				
4200	4.2524e-07	8.37295e-06	385.399	12	9.53551e-05	0.00187739	385.307				
5600	9.49843e-08	1.87e-06	385.304	16	1.69221e-05	0.000333144	385.28				
7000	2.18256e-08	4.29676e-07	385.281	20	3.07357e-06	6.05083e-05	385.275				
8400	5.09909e-09	1.00384e-07	385.276	24	5.66262e-07	1.11478e-05	385.274				
9800	1.20445e-09	2.37116e-08	385.274	28	1.053e-07	2.07301e-06	385.274				
12600	6.8643e-11	1.35135e-09	385.274	32	1.97049e-08	3.87922e-07	385.274				

usando la aproximación del Hessiano parece tener lo mejor de los dos mundos, pocas iteraciones y rápidas. Sin embargo como lo vimos en este último ejercicio, no logra converger cuando el Hessiano sí.