Exercise 2.

1. Program the steepest descent and Newton algorithms using the backtracking line search, Algorithm 1 (see on the next page). Use them to minimize the Rosenbrock function (see on the next page). Set the initial step length $\alpha_0 = 1$ and print the step length used by each method at each iteration. First try the initial point $x_0 = (1.2, 1.2)^T$ and then the more difficult starting point $x_0 = (-1.2, 1)^T$.

Choose $\rho = 0.5$ $\bar{\alpha} = 1.0$. $C_1 = 10^{-4}$

$x_0 = (1.2, 1.2)^T$

Steepest descent

A	B	C	D	E	F
iter	x1	x2	f(x)	||grad||	alpha
0	1.087109	1.246875	0.430975	30.98557	0.000977
1	1.114571	1.234166	0.019689	4.168659	0.000977
2	1.11082	1.235749	0.012615	0.694695	0.000977
3	1.111397	1.235392	0.012413	0.143725	0.000977
4	1.111126	1.235318	0.0124	0.172884	0.001953
5	1.111314	1.235037	0.012391	0.2146	0.001953
6	1.110895	1.23503	0.012387	0.272592	0.001953
7	1.111087	1.234846	0.012351	0.1001	0.000977
8	1.110501	1.234329	0.012335	0.35454	0.007813
9	1.11077	1.234111	0.012279	0.106231	0.000977
10	1.110428	1.233875	0.012262	0.219777	0.003906
11	1.110712	1.233553	0.012259	0.279486	0.001953
12	1.11044	1.233578	0.012222	0.100141	0.000977
13	1.110452	1.232796	0.012209	0.362773	0.007813
14	1.110103	1.232856	0.01215	0.106503	0.000977
15	1.110158	1.232444	0.012135	0.223565	0.003906
16	1.109721	1.232446	0.012132	0.284342	0.001953
17	1.109925	1.232258	0.012094	0.099939	0.000977
18	1.109331	1.231752	0.012083	0.365701	0.007813
19	1.10961	1.23153	0.012023	0.106097	0.000977
20	1.109266	1.231299	0.012008	0.222893	0.003906

Newton

A	B	C	D	E	F
iter	x1	x2	f(x)	||grad||	alpha
0	1.195918	1.430204	0.038384	0.39982	1
1	1.098284	1.196688	0.018762	4.784866	0.5
2	1.064488	1.131993	0.004289	0.656352	1
3	1.011992	1.021372	0.000903	1.265832	1
4	1.004261	1.008481	1.85E-05	0.034658	1
5	1.00005	1.000083	3.40E-08	0.00802	1
6	1	1	3.23E-14	1.45E-06	1
7	1	1	1.09E-25	1.44E-11	1
8	1	1	1.09E-25	1.44E-11	

$x_0 = (-1.2, 1)^T$

Steepest descent

A	B	C	D	E	F
iter	x1	x2	f(x)	||grad||	alpha
0	-0.98945	1.085938	5.101113	43.89852	0.000977
1	-1.06433	1.044172	5.047011	45.46014	0.001953
2	-1.02345	1.061483	4.114039	3.278977	0.000977
3	-1.02677	1.056002	4.108085	3.350914	0.001953
4	-1.02026	1.055316	4.102153	3.41296	0.001953
5	-1.02384	1.049694	4.096128	3.465561	0.001953
6	-1.01709	1.049127	4.090125	3.506376	0.001953
7	-1.02085	1.043404	4.084011	3.535751	0.001953
8	-1.01397	1.042912	4.077918	3.552267	0.001953
9	-1.01781	1.037137	4.071704	3.556104	0.001953
10	-1.01088	1.036669	4.065512	3.546935	0.001953
11	-1.01471	1.030892	4.059197	3.524745	0.001953
12	-1.00784	1.030398	4.052905	3.490355	0.001953
13	-1.01154	1.024671	4.046498	3.443559	0.001953
14	-1.00484	1.0241	4.040116	3.386256	0.001953
15	-1.00831	1.018473	4.033636	3.31809	0.001953
16	-1.00187	1.017778	4.027182	3.241852	0.001953
17	-1.00504	1.012295	4.020652	3.157105	0.001953
18	-0.99893	1.011437	4.014151	3.067242	0.001953
19	-1.00172	1.006133	4.007596	2.971827	0.001953
20	-0.99601	1.00508	4.001072	2.874517	0.001953

Newton

A	B	C	D	E	F
iter	x1	x2	f(x)	||grad||	alpha
0	-1.17528	1.38067	4.73188	4.63943	1
1	-0.93298	0.81121	4.0874	28.5501	0.125
2	-0.78254	0.58974	3.22867	11.5715	1
3	-0.46	0.10756	3.2139	30.3259	1
4	-0.39305	0.15	1.94259	3.6041	1
5	-0.20941	0.00677	1.60019	9.24842	0.25
6	-0.06572	-0.01633	1.17839	4.9198	1
7	0.14204	-0.02299	0.92241	8.66434	1
8	0.23111	0.04548	0.59749	1.77881	1
9	0.37974	0.11815	0.45263	5.8778	0.5
10	0.47959	0.22004	0.28076	2.17637	1
11	0.65341	0.39673	0.21139	9.40129	1
12	0.70262	0.49126	0.08902	0.49206	1
13	0.80279	0.63322	0.05154	3.92425	0.5
14	0.86349	0.74193	0.01999	1.24211	1
15	0.94208	0.88134	0.00717	2.53307	1
16	0.96799	0.93634	0.00107	0.23758	1
17	0.99621	0.99164	7.78E-05	0.34827	1
18	0.99948	0.99895	2.82E-07	0.00387	1
19	1	1	8.52E-12	0.00012	1
20	1	1	3.74E-21	4.47E-10	1

2. Show that if $0 < c_2 < c_1 < 1$, there may be no step lengths that satisfy the Wolfe conditions.

Pf. let $\phi(\alpha) = f(x_k + \alpha p_k)$. $\phi'(0) = \nabla f(x_k)^T p_k < 0$.

we find a counted-example. $\varphi(\alpha) = \frac{L}{2}\alpha^2 - \alpha$ $(L>0)$. then $\varphi'(0) = -1$.

Armijo: $\phi(\alpha) \le \phi(0) + c_1\alpha\phi'(0) \Rightarrow \frac{L}{2}\alpha^2 - (1-c_1)\alpha \le 0 \Leftrightarrow \alpha \le \frac{2(1-c_1)}{L}$.

Curvature: $\phi'(\alpha) \ge c_2\phi'(0) \Rightarrow L\alpha - 1 \ge -c_2 \Leftrightarrow \alpha \ge \frac{1-c_2}{L}$.

$\Rightarrow$ $c_2 \ge 2c_1 - 1$. if $c_2 < c_1$. for example. let $c_2 = 0.6$. $c_1 = 0.9$. no α s.t. Wolfe condition.

3. Show that the one-dimensional minimizer of a strongly convex quadratic function is given by formula (2) on the next page.

If f is a convex quadratic, $f(x) = \frac{1}{2}x^TQx - b^Tx$, its one-dimensional minimizer along the ray $x_k + \alpha p_k$ can be computed analytically and is given by

$$\alpha_k = -\frac{\nabla f_k^T p_k}{p_k^T Q p_k}. \tag{2}$$

Pf: $\phi(\alpha) = f(x_k + \alpha p_k) = \frac{1}{2}(x_k + \alpha p_k)^T Q (x_k + \alpha p_k) - b^T(x_k + \alpha p_k)$

$= \frac{1}{2}x_k^T Q x_k + \frac{\alpha}{2} p_k^T Q x_k + \frac{\alpha}{2} x_k^T Q p_k + \frac{\alpha^2}{2} p_k^T Q p_k - b^T x_k - \alpha b^T p_k$.

let $\phi'(\alpha) = 0 \Rightarrow \frac{1}{2} p_k^T Q x_k + \frac{1}{2} x_k^T Q p_k + \alpha p_k^T Q p_k - b^T p_k = 0$

by quadratic convex Q symmetric. $\Rightarrow (x_k^T Q - b^T) p_k + \alpha p_k^T Q p_k = 0$

(i.e. $x_k^T Q p_k = p_k^T Q x_k$. $Q^T = Q$). $\Rightarrow (Qx_k - b)^T p_k + \alpha p_k^T Q p_k = 0$

$\nabla f(x_k) = Qx_k - b$ (we have show in Ex 1. Problem 3). $\Rightarrow \alpha_k = -\frac{\nabla f(x_k)^T p_k}{p_k^T Q p_k}$.

4. Show that the one-dimensional minimizer of a strongly convex quadratic function always satisfies the Goldstein conditions (see on the next page).

The Goldstein conditions: $f(x_k) + (1-c)\alpha_k \nabla f_k^T p_k \le f(x_k + \alpha_k p_k) \le f(x_k) + c\alpha_k \nabla f_k^T p_k$,

Pf: let $\phi(\alpha) = f(x_k + \alpha p_k) = \frac{1}{2} p_k^T Q p_k \alpha^2 + \nabla f(x_k)^T p_k \alpha + f(x_k)$.

$= \frac{1}{2} p_k^T Q p_k \alpha^2 + \phi'(0)\alpha + \phi(0)$.

Goldstein $\Rightarrow$ $\phi(0) + (1-c)\alpha_k\phi'(0) \le \phi(\alpha_k) \le \phi(0) + c\alpha_k\phi'(0)$.

$\phi(\alpha_k) - \phi(0) = \alpha_k\left[\frac{\alpha_k}{2} p_k^T Q p_k + \phi'(0)\right] = \alpha_k\left[-\frac{\phi'(0)}{p_k^T Q p_k} \cdot \frac{p_k^T Q p_k}{2} + \phi'(0)\right] = \frac{1}{2} \cdot \alpha_k \phi'(0)$.

for $\phi'(0) < 0$. $c \in (0, \frac{1}{2})$. $(1-c)\alpha_k\phi'(0) \le \frac{1}{2}\alpha_k\phi'(0) \le c\alpha_k\phi'(0)$

5. Prove that $\|Bx\| \ge \|x\|/\|B^{-1}\|$ for any nonsingular matrix B.

Pf: by property of norm. $\|B\| = \sup_{x \ne 0} \frac{\|Bx\|}{\|x\|} \Rightarrow \|Bx\| \le \|B\| \cdot \|x\|$ (for any non-singular B).

$x = B^{-1}(Bx)$. $\Rightarrow \|x\| \le \|B^{-1}\| \cdot \|Bx\|$. $\Rightarrow \|Bx\| \ge \frac{\|x\|}{\|B^{-1}\|}$.

6. Consider the steepest descent method with exact line searches applied to the convex quadratic function (3) on the next page. Using the properties given in lecutre 3, show that if the initial point is such that $x_0 - x^*$ is parallel to an eigenvector of Q, then the steepest descent method will find the solution in one step.

Pf: steepest method. $p_k = -\nabla f_k$. $\alpha_0 = -\dfrac{\nabla f_0^T p_0}{p_0^T Q p_0}$ $\nabla f_0 = Q x_0 - b$

$x_0 - x^*$ parallel to some eigenvector. $Q(x_0 - x^*) = \lambda(x_0 - x^*)$

since $\nabla f(x^*) = Qx^* - b = 0$. we have $b = Qx^*$. thus. $\nabla f_0 = Q(x_0 - x^*) = \lambda(x_0 - x^*)$

$$x_1 = x_0 + \alpha_0 p_0 = x_0 - \frac{\nabla f_0^T \nabla f_0}{\nabla f_0^T Q \nabla f_0} \cdot \nabla f_0 = x_0 - \frac{\lambda^2 \|x_0 - x^*\|^2}{\lambda^3 \|x_0 - x^*\|^2} \cdot \lambda(x_0 - x)^* = x_0 - (x_0 - x^*) = x^*$$ $\Rightarrow$ find solution in one step

7. Let Q be a positive definite symmetric matrix. Prove that for any vector x, we have

$$\frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} \geq \frac{4\lambda_n \lambda_1}{(\lambda_n + \lambda_1)^2},$$

where λ_n and λ_1 are, respectively, the largest and smallest eigenvalues of Q.

Pf: $Q = U \Lambda U^{\perp}$ spectral decomposition denote. $y = U^{\perp} x$. $w_i = \dfrac{y_i}{\Sigma y_i}$ $\Lambda = \mathrm{diag}(\lambda_1, \dots \lambda_n)$ $0 < \lambda_1 \leq \dots \leq \lambda_n$

$$LHS = \frac{(\Sigma y_i^2)^2}{(\Sigma \lambda_i y_i^2)(\Sigma \frac{1}{\lambda_i} y_i^2)} = \frac{1}{(\Sigma \lambda_i w_i)(\Sigma \frac{1}{\lambda_i} w_i)}$$

by the convexity of the exoression, $\exists \theta \in [0,1]$. $\sum_i w_i \lambda_i = \theta \lambda_1 + (1-\theta)\lambda_n$ $\Sigma \frac{1}{\lambda_i} w_i = \theta \frac{1}{\lambda_1} + (1-\theta)\frac{1}{\lambda_n}$.

$$LHS(\theta) = \frac{1}{(\theta\lambda_1 + (1-\theta)\lambda_n)(\theta\frac{1}{\lambda_1} + (1-\theta)\frac{1}{\lambda_n})}$$

denote $\varphi(\theta) = (2 - \frac{\lambda_n}{\lambda_1} - \frac{\lambda_1}{\lambda_n})\theta^2 + (\frac{\lambda_1}{\lambda_n} + \frac{\lambda_n}{\lambda_1} - 2)\theta + 1$. $\varphi'(\theta) = 2(2 - \frac{\lambda_n}{\lambda_1} - \frac{\lambda_1}{\lambda_n})\theta + (\frac{\lambda_n}{\lambda_1} + \frac{\lambda_1}{\lambda_n} - 2) = 0 \Rightarrow \theta = \frac{1}{2}$

$$LHS_{min} = LHS|_{\theta = \frac{1}{2}} = \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}$$ i.e. $LHS \geq RHS$

8. Compute the eigenvalues of the 2 diagonal blocks of (4) on the next page and verify that each block has a positive and a negative eigenvalue. Then compute the eigenvalues of Aand verify that its inertia is the same as that of B.

The matrix

$$A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ \cancel{1} & \cancel{2} & \cancel{2} & \cancel{2} \\ 2 & 2 & 3 & 3 \\ 3 & 2 & 3 & 4 \end{bmatrix} \begin{matrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 3 & 4 \\ 2 & 2 & 3 & 3 \\ 1 & 2 & 2 & 2 \end{matrix}$$

can be written in the form (3.51) with $P = [e_1, e_4, e_3, e_2]$,

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{9} & \frac{2}{3} & 1 & 0 \\ \frac{2}{9} & \frac{1}{3} & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 3 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ 0 & 0 & \frac{7}{9} & \frac{5}{9} \\ 0 & 0 & \frac{5}{9} & \frac{10}{9} \end{bmatrix}. \qquad (4)$$

Sol: For block-diag B.

$$(\lambda(\lambda-4) - 9)[(\lambda - \tfrac{7}{9})(\lambda - \tfrac{10}{9}) - (\tfrac{5}{9})^2] = 0$$

$\Rightarrow$ $\lambda = 2 \pm \sqrt{13}$, $\dfrac{17 \pm \sqrt{109}}{18}$ block 1 has a positive + a negative ev
block 2 has 2 positive ev.

For A:

$$PAP^T = \begin{bmatrix} 0 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 3 & 2 \\ 1 & 2 & 2 & 2 \end{bmatrix} =: \begin{bmatrix} E & C^T \\ C & H \end{bmatrix} \qquad E^{-1} = -\frac{1}{9}\begin{bmatrix} 4 & -3 \\ -3 & 0 \end{bmatrix}$$

$$PAP^T = \begin{bmatrix} I & 0 \\ CE^{-1} & I \end{bmatrix} \begin{bmatrix} E & 0 \\ 0 & H - CE^{-1}C^T \end{bmatrix} \begin{bmatrix} I & E^{-1}C^T \\ 0 & I \end{bmatrix}$$

$$H - CE^{-1}C^T = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} + \frac{1}{9}\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 4 & -3 \\ -3 & 0 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} + \frac{1}{9}\begin{bmatrix} -1 & -6 \\ -2 & -3 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} + \frac{1}{9}\begin{bmatrix} -20 & -13 \\ -13 & -8 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{7}{9} & \frac{5}{9} \\ \frac{5}{9} & \frac{10}{9} \end{bmatrix}$$ thus. A.B must have same ineria.

If we compute characteristic equation of A directly.

$$\begin{vmatrix} -\lambda & 1 & 2 & 3 \\ 1 & 2-\lambda & 2 & 2 \\ 2 & 2 & 3-\lambda & 3 \\ 3 & 2 & 3 & 4-\lambda \end{vmatrix} = -\lambda \begin{vmatrix} 2-\lambda & 2 & 2 \\ 2 & 3-\lambda & 3 \\ 2 & 3 & 4-\lambda \end{vmatrix} - 1 \cdot \begin{vmatrix} 1 & 2 & 2 \\ 2 & 3-\lambda & 3 \\ 3 & 3 & 4-\lambda \end{vmatrix} + 2 \begin{vmatrix} 1 & 2-\lambda & 2 \\ 2 & 2 & 3 \\ 3 & 2 & 4-\lambda \end{vmatrix} - 3 \begin{vmatrix} 1 & 2-\lambda & 2 \\ 2 & 2 & 3-\lambda \\ 3 & 2 & 3 \end{vmatrix}$$

$$= \lambda^4 - 9\lambda^3 + 2\lambda^2 + 7\lambda - 5 = (\lambda+1)(\lambda^3 - 10\lambda^2 + 12\lambda - 5).$$

$\lambda_1 = -1.$ $\quad \lambda_2 \lambda_3 \lambda_4 = 5.$

we can make approximation $\lambda_2 \approx 9.65$ $\lambda_3 \approx 0.175$ $\lambda_4 \approx 0.175$. we have inertia $(n_+, n_-, n_0) = (3, 1, 0)$ same as B.

9. Describe the effect that the modified Cholesky factorization (5) on the next page would have on the Hessian $\nabla^2 f(x_k) = \mathrm{diag}(-2, 12, 4)$.

The Cholesky factorization $PAP^T + E = LDL^T = MM^T$ (5)

where E is a nonnegative diagonal matrix that is zero if A is sufficiently positive definite.

let $P = [e_2 \ e_3 \ e_1]$ $\quad PAP^T = \mathrm{diag}(12, 4, -2)$.

then set some δ. denote $E = \mathrm{diag}(\delta_1, \delta_2, \delta_3)$. $\delta_1, \delta_2 = 0$ $\quad \delta_3 > 2 + \delta$.

which makes. $PAP^T + E \succ 0$.

then Newton step will be descent direction surely.

10. Consider a block diagonal matrix B with 1×1 and 2×2 blocks. Show that the eigenvalues and eigenvectors of B can be obtained by computing the spectral decomposition of each diagonal block separately.

let $B = \mathrm{block}(B_{i1}, \dots, B_{ik})$. $\quad B_{ij} \in \{\mathbb{R}^{1\times 1}, \mathbb{R}^{2\times 2}\}$.

$|B - \lambda I| = \prod_j (B_{ij} - \lambda I)$. we can solve the equation of 1st / 2nd order. to find eigenvalues and eigenvectors.

with other entries on that column / row keeps 0.