Problem 1

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be the function defined by

$$f(x,y) = x^2 + xy + y^2 = \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

(a) Find the strict global minimum point of f.

(b) Compute one iteration of the Gradient Descent method using fixed step size $t_k \equiv \frac{1}{L_{\nabla f}}$, exact line search and backtracking with $(s,\alpha,\beta) = \left(1, \frac{1}{10}, \frac{1}{4}\right)$. Assume that $(1,1)$ is the starting point for all three methods. Write your calculations in detail. Did the methods converge to the global minimum point after one iteration?

Solution

(a) f is twice continuously differentiable with a PD Hessian matrix everywhere, therefore any stationary point is a strict global minimum. Solving $\mathbf{Ax} = \mathbf{0}$ we see that $(0,0)$ is a global minimum.

(b) First we find $L_{\nabla f}$. Since $\nabla f(\mathbf{x}) = 2\mathbf{Ax}$ then

$$\|2\mathbf{Ax} - 2\mathbf{Ay}\| \le 2\|\mathbf{A}\| \|\mathbf{x}-\mathbf{y}\| = 2\lambda_{\max}\|\mathbf{x}-\mathbf{y}\|,$$

so $2\lambda_{\max}$ is a Lipschitz constant. This constant is also tight since taking $\mathbf{x}$ to be an eigenvector of $\lambda_{\max}$ and $\mathbf{y} = 2\mathbf{x}$ we have

$$\|2\mathbf{Ax} - 4\mathbf{Ax}\| = 2\|\mathbf{Ax}\| = 2\lambda_{\max}\|\mathbf{x}\| = 2\lambda_{\max}\|\mathbf{x}-\mathbf{y}\|.$$

In our case we have $2\lambda_{\max} = \frac{1}{3}$. Now we calculate GD iterations with different step-sizes:

- Constant step-size with $t_k \equiv \frac{1}{L_{\nabla f}}$ reads

$$\begin{aligned}\begin{pmatrix} x^1 \\ y^1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2\lambda_{\max}}\nabla f\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{2}{3}\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{2}{3}\begin{pmatrix} \frac{3}{2} \\ \frac{3}{2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\end{aligned}$$

  and we reached the optimal solution in one iteration. Since $\nabla f(0,0) = (0,0)$ then $\mathbf{x}^{k+1} = \mathbf{x}^k$ for all $k \ge 0$ and the method converged after one iteration (this is not always the case, for example choose the starting point $(1,2)$).

- For exact line search we need to calculate

$$\min_{t>0} f\left(\begin{pmatrix} x^0 \\ y^0 \end{pmatrix} - t\nabla f\begin{pmatrix} x^0 \\ x^0 \end{pmatrix}\right) = \min_{t>0} f\begin{pmatrix} 1-3t \\ 1-3t \end{pmatrix} = \min_{t>0} 3(1-3t)^2.$$

  The minimum of $3(1-3t)^2$ is at $t = \frac{1}{3}$ and it is feasible. Therefore

$$\begin{pmatrix} x^1 \\ y^1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{3}\nabla f\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

  and we reached the optimal solution in one iteration (this is not always the case). Alternatively, we saw in the lecture that for a quadratic function with a PD matrix we have for a descent direction $\mathbf{d}$ that

$$t = -\frac{\mathbf{d}^T\nabla f(\mathbf{x})}{\mathbf{d}^T\nabla^2 f(\mathbf{x})\,\mathbf{d}} > 0,$$

  and therefore for $\mathbf{d} = -\nabla f(\mathbf{x})$ we have

$$t = \frac{\|\nabla f(\mathbf{x})\|^2}{\nabla f(\mathbf{x})^T(2\mathbf{A})\nabla f(\mathbf{x})} = \frac{\left\|\begin{pmatrix} 3 \\ 3 \end{pmatrix}\right\|^2}{2\begin{pmatrix} 3 \\ 3 \end{pmatrix}^T\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}\begin{pmatrix} 3 \\ 3 \end{pmatrix}} = \frac{1}{3}$$

  and the result is of course the same.

- For backtracking we need to check when the following stopping criterion is satisfied:
$$f\begin{pmatrix}x^0\\y^0\end{pmatrix}-f\left(\begin{pmatrix}x^0\\y^0\end{pmatrix}-t\nabla f\begin{pmatrix}x^0\\y^0\end{pmatrix}\right)\geq \alpha t\left\|\nabla f\begin{pmatrix}x^0\\y^0\end{pmatrix}\right\|^2.$$
For $s=t=1$ we have
$$\begin{aligned}f\begin{pmatrix}x^0\\y^0\end{pmatrix}-f\left(\begin{pmatrix}x^0\\y^0\end{pmatrix}-t\nabla f\begin{pmatrix}x^0\\y^0\end{pmatrix}\right)&=f\begin{pmatrix}1\\1\end{pmatrix}-f\left(\begin{pmatrix}1\\1\end{pmatrix}-\begin{pmatrix}3\\3\end{pmatrix}\right)=-9\\&<\alpha t\left\|\nabla f\begin{pmatrix}x^0\\y^0\end{pmatrix}\right\|^2=\frac{1}{10}\left\|\begin{pmatrix}3\\3\end{pmatrix}\right\|^2=1.8,\end{aligned}$$
and therefore $t:=\beta t=\frac{1}{4}$. We have
$$f\begin{pmatrix}x^0\\y^0\end{pmatrix}-f\left(\begin{pmatrix}x^0\\y^0\end{pmatrix}-\frac{1}{4}\nabla f\begin{pmatrix}x^0\\y^0\end{pmatrix}\right)=2.8125\geq\frac{1}{10}\cdot\frac{1}{4}\left\|\begin{pmatrix}3\\3\end{pmatrix}\right\|^2=0.45$$
and therefore
$$\begin{pmatrix}x^1\\y^1\end{pmatrix}=\begin{pmatrix}1\\1\end{pmatrix}-\frac{1}{4}\begin{pmatrix}3\\3\end{pmatrix}=\begin{pmatrix}\frac{1}{4}\\\frac{1}{4}\end{pmatrix},$$
and the method did not converge after one iteration.

Cholesky Factorization

In order for the Newton's method to generate a well-defined descent sequence, we need $\nabla^2 f\left(\mathbf{x}^k\right)$ to be PD. The Cholesky factorization is a relatively numerically stable method that checks whether a matrix is PD.

Let $\mathbf{A}\in\mathbb{R}^{n\times n}$. Then $\mathbf{A}\succ 0$ if and only if there exists a lower triangular matrix $\mathbf{L}\in\mathbb{R}^{n\times n}$ such that $\mathbf{A}=\mathbf{L}\mathbf{L}^T$ (and it is called the Cholesky factorization of $\mathbf{A}$).

Problem 3

1. Given a Cholesky factorization of $\mathbf{A}$, show how to solve the system $\mathbf{A}\mathbf{x}=\mathbf{b}$.
2. Show how to attain a Cholesky factorization of a matrix $\mathbf{A}\in\mathbb{R}^{n\times n}$.

Solution

1. Find a solution $\mathbf{w}$ of the system $\mathbf{L}\mathbf{w}=\mathbf{b}$ and then find a solution $\mathbf{x}$ of the system $\mathbf{L}^T\mathbf{x}=\mathbf{w}$. Note that because of the special structure of $\mathbf{L}$, these two steps can be done via backward and forward substitution. For example:
$$\begin{pmatrix}L_{11}&0\\L_{21}&L_{22}\end{pmatrix}\begin{pmatrix}\mathbf{w}_1\\\mathbf{w}_2\end{pmatrix}=\begin{pmatrix}\mathbf{b}_1\\\mathbf{b}_2\end{pmatrix}\Longrightarrow\mathbf{w}_1=\frac{\mathbf{b}_1}{L_{11}},\mathbf{w}_2=\frac{\mathbf{b}_2-L_{21}\mathbf{w}_1}{L_{22}}=\frac{\mathbf{b}_2-L_{21}\mathbf{b}_1}{L_{11}L_{22}},$$
and we solve $\mathbf{L}^T\mathbf{x}=\mathbf{w}$ in a similar fashion.
2. We need to solve
$$\begin{aligned}\begin{pmatrix}A_{11}&\mathbf{A}_{12}\\\mathbf{A}_{12}^T&\mathbf{A}_{22}\end{pmatrix}&=\begin{pmatrix}L_{11}^2&L_{11}\mathbf{L}_{12}^T\\L_{11}\mathbf{L}_{12}&\mathbf{L}_{12}\mathbf{L}_{12}^T+\mathbf{L}_{22}\mathbf{L}_{22}^T\end{pmatrix}=\begin{pmatrix}L_{11}&\mathbf{0}_{1\times(n-1)}\\\mathbf{L}_{12}&\mathbf{L}_{22}\end{pmatrix}\begin{pmatrix}L_{11}&\mathbf{0}_{1\times(n-1)}\\\mathbf{L}_{12}&\mathbf{L}_{22}\end{pmatrix}^T.\\&=\begin{pmatrix}L_{11}&\mathbf{0}_{1\times(n-1)}\\\mathbf{L}_{12}&\mathbf{L}_{22}\end{pmatrix}\begin{pmatrix}L_{11}&\mathbf{L}_{12}^T\\\mathbf{0}_{n-1}&\mathbf{L}_{22}^T\end{pmatrix}\end{aligned}$$
We get that $L_{11}=\sqrt{A_{11}}$ and $\mathbf{L}_{12}=\frac{1}{\sqrt{A_{11}}}\mathbf{A}_{12}^T$. Therefore
$$\mathbf{L}_{22}\mathbf{L}_{22}^T=\mathbf{A}_{22}-\mathbf{L}_{12}\mathbf{L}_{12}^T=\mathbf{A}_{22}-\frac{1}{A_{11}}\mathbf{A}_{12}^T\mathbf{A}_{12},$$
and we are left with finding the Cholesky factorization of the $(n-1)\times(n-1)$ symmetric matrix $\mathbf{A}_{22}-\frac{1}{A_{11}}\mathbf{A}_{12}^T\mathbf{A}_{12}$.