OSM Lab Boot Camp **Math Problem Set 6**

Harrison Beard

30 July 2018

Exercise 9.1.

Prove that an unconstrained linear objective function is either constant or has no minimum.

Solution. text. ■

Exercise 9.2.

Prove that if $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{A} \in \mathrm{M}_{m \times n}(\mathbb{R})$, then the problem of finding an $\mathbf{x}^* \in \mathbb{R}^n$ to minimize $\|\mathbf{Ax} - \mathbf{b}\|_2$ is equivalent to minimizing

$$\mathbf{x}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{Ax} - 2\mathbf{b}^{\mathsf{T}}\mathbf{Ax}. \tag{9.21}$$

In Voume 1, Chapter 3 we use projections to prove that this is equivalent to solving the normal equation

$$\mathbf{A}^{\mathsf{T}}\mathbf{Ax} = \mathbf{A}^{\mathsf{T}}\mathbf{b}.$$

Use the first- and second-order conditions to give a different proof that minimizing (9.21) is equivalent to solving the normal equation.

Solution. text. ■

Exercise 9.3.

For each of the multivariabel optimization methods we ahve discussed in this section, list the following:

(*i*). The basic idea of the method, including how it differs from the other methods in the list. Include any geometric description you can give of the method.

(*ii*). What types of optimization problems it can solve and cannot solve.

(*iii*). Relative strengths of the method.

(*iv*). Relative weaknesses of the method.

Solution.

(*i*). text

(*ii*). text

(*iii*). text

(*iv*). text

■

Exercise 9.4.

Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{Q}\mathbf{x} - \mathbf{b}^{\mathsf{T}}\mathbf{x}$, where $\mathbf{Q} \in \mathrm{M}_{m\times n}(\mathbb{R})$ satisfies $\mathbf{Q} > 0$ and $\mathbf{b} \in \mathbb{R}^n$. Show that the Method of Steepest Descent (that is, gradient descent with optimal line search), converges in one step (that is, $\mathbf{x}_1 = \mathbf{Q}^{-1}\mathbf{b}$), if and only if $\mathbf{x}_0$ is chosen such that $\mathbf{D}f(\mathbf{x}_0)^{\mathsf{T}} = \mathbf{Q}\mathbf{x}_0 - \mathbf{b}$ is an eigenvector of $\mathbf{Q}$ (and α_0 satisfies (9.2)).

Solution. text. ■

Exercise 9.5.

Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^1$. Let $\{\mathbf{x}_k\}_{k=0}^{\infty}$ be defined by the Method of Steepest Descent. Show that if $\mathbf{x}_{k+1} - \mathbf{x}_k$ is orthogonal to $\mathbf{x}_{k+2} - \mathbf{x}_{k+1}$ for each k.

Solution. text. ■

Exercise 9.6.

Write a Python/NumPy routine for implementing the steepest descent method for quadratic functions (see Example $9.2.3$).

Given a small number ε, given Numpy arrays $\mathbf{x}_0, \mathbf{b}$ of length n, and given an $n \times n$ matrix $\mathbf{Q} > 0$, your code should return a close approximation to a local minimizer $\mathbf{x}^*$ of $f = \mathbf{x}^{\mathsf{T}}\mathbf{Q}\mathbf{x} - \mathbf{b}^{\mathsf{T}}\mathbf{x} + c$.

For the stopping criterion, use the condition $\|\mathbf{D}f(\mathbf{x}_k)\|$ for some small value of ε.

Solution. text.

```
# code here
```

■

Exercise 9.7.

Write a simple Python/NumPy method for computing $\mathbf{D}f$ using forward differences and a step size of $\sqrt{\mathrm{Rerr}_f}$. It should accept a callable function $f : \mathbb{R}^n \to \mathbb{R}$, a point $\mathbf{x} \in \mathbb{R}^n$, and an estimate $\mathrm{Rerr}_f > \varepsilon$ for the maximum relative error of f near $\mathbf{x}$. It should return an estimate for $\mathbf{D}f(\mathbf{x})$.

Solution. text.

```
# code here
```

■

Exercise 9.8.

Use your differentiation method from the previous problem to construct a simple Python/NumPy method for implementing the steepest descent method for arbitrary functions, using the secant method (Exercise 6.15) for the line search.

Your method should accept a callable function f, a starting value $\mathbf{x}_0$, a small number ε, a NumPy array $\mathbf{x}_0$ of length n, and return a close approximation to a local minimizer $\mathbf{x}^*$ of f.

For the stopping criterion, use the condition $\|\mathbf{D}f(\mathbf{x}_k)\|$.

Solution. text.

```
# code here
```

■

Exercise 9.9.

Apply your code from the previous problem to the Rosenbrock function

$$f(x, y) = 100\left(y - x^2\right)^2 + (1 - x)^2$$

with an initial guess of $(x_0, y_0) = (-2, 2)$.

Solution. text. ■

Exercise 9.10.

Consider the quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x} - \mathbf{b}^\mathsf{T}\mathbf{x}$, where $\mathbf{Q} \in \mathrm{M}_n(\mathbb{R})$ is symmetric and positive definite and $\mathbf{b} \in \mathbb{R}^n$. Show that for any initial guess $\mathbf{x}_0 \in \mathbb{R}^n$, one iteration of Newton's method lands at the unique minimizer of f.

Solution. text. ■

Exercise 9.12.

Prove that if $\mathbf{A} \in \mathrm{M}_n(\mathbb{F})$ has eigenvalues $\lambda_1, \ldots, \lambda_n$ and $\mathbf{B} = \mathbf{A} + \mu\mathbf{I}$, then the eigenvectors of $\mathbf{A}$ and $\mathbf{B}$ are the same, and the eigenvalues of $\mathbf{B}$ are $\mu + \lambda_1, \mu + \lambda_2, \ldots, \mu + \lambda_n$.

Solution. text. ■

Exercise 9.15.

Prove the Sherman-Morrison-Woodbury formula (9.13).

(9.13). *Let* $\mathbf{A}$ *be a nonsingular* $n \times n$ *matrix,* $\mathbf{B}$ *an* $n \times \ell$ *matrix,* $\mathbf{C}$ *a nonsingular* $\ell \times \ell$ *matrix, and* $\mathbf{D}$ *an* $\ell \times n$ *matrix. We have*

$$(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}. \tag{9.13}$$

Solution. text. ■

Exercise 9.16.

Use (9.13) to derive (9.14).

$$\mathbf{A}_k^{-1} = \mathbf{A}_k^{-1} + \frac{(\mathbf{s}_{k-1}-\mathbf{A}_{k-1}^{-1}\mathbf{y}_{k-1})\mathbf{s}_{k-1}^{\mathsf{T}}\mathbf{A}_{k-1}^{-1}}{\mathbf{s}_{k-1}^{\mathsf{T}}\mathbf{A}_{k-1}^{-1}\mathbf{y}_{k-1}}. \tag{9.14}$$

Solution. text. ■

Exercise 9.17.

Apply (9.13) twice to derive (9.17).

$$\mathbf{A}_{k-1}^{-1} = \mathbf{A}_k^{-1} + \frac{(\mathbf{s}_k^{\mathsf{T}}\mathbf{y}_k^{\mathsf{T}}\mathbf{A}_k^{-1}\mathbf{y}_k)\mathbf{s}_k\mathbf{s}_k^{\mathsf{T}}}{(\mathbf{s}_k^{\mathsf{T}}\mathbf{y}_k)^2} - \frac{\mathbf{A}_k^{-1}\mathbf{y}_k\mathbf{s}_k^{\mathsf{T}}+\mathbf{s}_k\mathbf{y}_k^{\mathsf{T}}\mathbf{A}_k^{-1}}{\mathbf{s}_k^{\mathsf{T}}\mathbf{y}_k}. \tag{9.17}$$

Solution. text. ■

Exercise 9.18.

Let $\mathbf{Q} \in \mathrm{M}_n(\mathbb{R})$ satisfy $\mathbf{Q} > 0$, and let f be the quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{Q}\mathbf{x} - \mathbf{b}^{\mathsf{T}}\mathbf{x} + c$. Given a starting point $\mathbf{x}_0$ and $\mathbf{Q}$-conjugate directions $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{n-1}$ in $\mathbb{R}^n$, show that the optimal line search solution for $\mathbf{x}_{k-1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$ (that is, the α which minimizes $\phi_k(\alpha) = f(\mathbf{x}_k + \alpha_k\mathbf{d}_k)$) is given by $\alpha_k = \frac{\mathbf{r}_k^{\mathsf{T}}\mathbf{d}_k}{\mathbf{d}_k^{\mathsf{T}}\mathbf{Q}\mathbf{d}_k}$, where $\mathbf{r}_k = \mathbf{b} - \mathbf{Q}\mathbf{x}_k$.

Solution. text. ■

Exercise 9.20.

Prove Lemma 9.5.5.

Lemma 9.5.5. *In the Conjugate Gradient Algorithm,* $\mathbf{r}_i^{\mathsf{T}}\mathbf{r}_k = 0$ *for all* $i < k$.

Solution. text. ■