Department of Mathematics, University of Utah

Introduction to Optimization

MATH 5770/6640, ME EN 6025 – Section 001 – Fall 2021

Homework 2 Solutions

Optima and optimality conditions

Due September 21, 2021

Text: *Introduction to Nonlinear Optimization*, Amir Beck,

Exercises: # 2.1,
2.2,
2.4,
2.6,
2.11,
2.13(i, ii),
2.15(i, iii, iv, vii),
2.17(i,iii,vi,vii),
2.18

2.1. Find the global minimum and maximum points of the function $f(x,y) = x^2 + y^2 + 2x - 3y$ over the unit ball $S = B[\mathbf{0},1] = \left\{(x,y) :\ x^2+y^2 \le 1\right\}$.

Let $\boldsymbol{a}^T = (2,3)^T$, and note that $2x - 3y = \left\langle \boldsymbol{a}, (x,y)^T \right\rangle$. Therefore, by the Cauchy-Schwarz inequality,

$$f(x,y) = x^2 + y^2 + 2x - 3y \overset{r^2 = x^2+y^2}{\le} r^2 + \|\boldsymbol{a}\|_2 r,$$
$$f(x,y) = x^2 + y^2 + 2x - 3y \overset{r^2 = x^2+y^2}{\ge} r^2 - \|\boldsymbol{a}\|_2 r, \tag{1}$$

where equality is achieved in each case if and only if (x,y) is parallel to $\boldsymbol{a}$. Therefore, if we consider maximization, then we have,

$$f(x,y) \le g_+(r) \coloneqq r^2 + r\sqrt{13},$$

with equality achieved exactly when $(x,y)^T = r\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}$. Since the feasible set S is the unit ball, i.e., for all points satisfying $0 \le r \le 1$, we seek to maximize g_+ for $r \in [0,1]$. Since g_+ is monotonic increasing on $[0,1]$, then its maximum is attained at $r = 1$, i.e., we have,

$$\max_{(x,y)\in B[\mathbf{0},1]} f(x,y) = \max_{r\in[0,1]} g_+(r) = g_+(1) = 1 + \sqrt{13},$$

and furthermore, $g_+(1) = f(x,y)$ when $(x,y)^T = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}$, i.e.,

$$\operatorname*{argmax}_{(x,y)\in B[\mathbf{0},1]} f(x,y) = \left(\frac{2}{\sqrt{13}}, -\frac{3}{\sqrt{13}}\right),$$

with maximum value $1 + \sqrt{13}$.

To compute the minimum value, we return to (1), and with similar logic conclude,

$$\min_{(x,y)\in B[\mathbf{0},1]} f(x,y) = \min_{r\in[0,1]} g_-(r), \qquad\qquad g_-(r) := r^2 - r\sqrt{13}.$$

where this time we make the identification $(x,y)^T = -r\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}$ in order to achieve equality in (1) using the Cauchy-Schwarz inequality. In this case, g_- is monotonically decreasing function of r on $[0,1]$ since $g_-'(r) = 2r - \sqrt{13} < 0$. Therefore, the minimum occurs again at $r = 1$, i.e.,

$$\operatorname*{argmin}_{(x,y)\in B[\mathbf{0},1]} f(x,y) = \left(-\frac{2}{\sqrt{13}}, \frac{3}{\sqrt{13}}\right),$$

with minimum value $1 - \sqrt{13}$.

2.2. Let $\boldsymbol{a} \in \mathbb{R}^n$ be a nonzero vetor. Show that the maximum of $\boldsymbol{a}^T x$ over $B[\mathbf{0},1] = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| \le 1\}$ is attained at $\boldsymbol{x}^* = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}$ and that the maximal value if $\|\boldsymbol{a}\|$.

With $r = \|\boldsymbol{x}\|$, the Cauchy-Schwarz inequality implies,

$$\max_{\boldsymbol{x}\in B[\mathbf{0},1]} \boldsymbol{a}^T\boldsymbol{x} \le \max_{\boldsymbol{x}\in B[\mathbf{0},1]} \|\boldsymbol{a}\|\|\boldsymbol{x}\| = \max_{\|\boldsymbol{x}\|\le 1} \|\boldsymbol{a}\|\|\boldsymbol{x}\| = \|\boldsymbol{a}\|, \tag{2}$$

where equality is achieved above if and only if $\boldsymbol{x} = r\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}$. Since we also know from above that the maximum is achieved when $\|\boldsymbol{x}\| = 1$, then we require a vector $\boldsymbol{x}$ satisfying both,

$$\boldsymbol{x} = r\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}, \qquad\qquad \|\boldsymbol{x}\| = 1,$$

implying that $\boldsymbol{x} = \boldsymbol{a}/\|\boldsymbol{a}\|$, achieving maximum value $\|\boldsymbol{a}\|$ as shown by (2).

2.4. Show that if $\boldsymbol{A}, \boldsymbol{B}$ are $n \times n$ positive semidefinite matrices, then their sum $\boldsymbol{A} + \boldsymbol{B}$ is also positive semidefinite.

Let $\boldsymbol{x} \in \mathbb{R}^n$ be arbitrary. Then,

$$\boldsymbol{x}^T(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} \overset{\boldsymbol{A},\boldsymbol{B}\succeq\mathbf{0}}{\ge} 0,$$

showing that $\boldsymbol{A} + \boldsymbol{B} \succeq \mathbf{0}$.

2.6. Let $\boldsymbol{B} \in \mathbb{R}^{n\times k}$ and let $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}^T$.
(i) Prove $\boldsymbol{A}$ is positive semidefinite.
(ii) Prove that $\boldsymbol{A}$ is positive definite if and only if $\boldsymbol{B}$ as full row rank.

Let $\boldsymbol{x} \in \mathbb{R}^n$ be arbitrary. Defining $\boldsymbol{y} := \boldsymbol{B}^T\boldsymbol{x}$, then

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{y} = \|\boldsymbol{y}\|_2^2 \ge 0,$$

since $\|\cdot\|_2$ is a norm. This proves (i), that $\boldsymbol{A} \succeq \mathbf{0}$.
To prove (ii), assume $\boldsymbol{x} \neq \mathbf{0}$, and note that $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \|\boldsymbol{y}\|_2^2$ can equal zero if and only if $\boldsymbol{y} = \mathbf{0}$. Note that in order for this to happen, $\boldsymbol{x}$ must be a nonzero vector satsifying,

$$\boldsymbol{y} = \boldsymbol{B}^T\boldsymbol{x} = \mathbf{0},$$

implying in turn that $\boldsymbol{x}$ must be a non-trivial vector in the kernel of $\boldsymbol{B}^T$ in order for $\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}$ to vanish. The Fundamental Theorem of Linear Algebra guarantees that the kernel of $\boldsymbol{B}^T$ is empty (i.e., has dimension equal to 0) if and only if $\boldsymbol{B}^T$ has linearly independent columns, i.e., if and only if $\boldsymbol{B}$ as linearly independent rows, i.e., if and only if $\boldsymbol{B}$ has full row rank. Thus, $\boldsymbol{A} \succ \boldsymbol{0}$ if and only if $\boldsymbol{B}$ has full row rank.

2.11. Let $\boldsymbol{d} = \Delta_n$ (Δ_n being the unit-simplex). Show that the $n \times n$ matrix $\boldsymbol{A}$ defined by

$$A_{i,j} = \begin{cases} d_i - d_i^2, & i = j, \\ -d_i d_j, & i \neq j, \end{cases}$$

is positive semidefinite.

Note that $\boldsymbol{A}$ is a symmetric matrix: $A_{i,j} = A_{j,i}$. If $\boldsymbol{d} \in \Delta_n$, then its entries d_i all satisfy $0 \leq d_i \leq 1$ for $i = 1, \ldots, n$, and also $d_i = 1 - \sum_{j \neq i} d_j$. This implies first that

$$d_i \geq d_i^2 \quad \Longrightarrow \quad A_{i,i} \geq 0,$$

and second that

$$\sum_{j \neq i} |A_{i,j}| = d_i \left(\sum_{j \neq i} d_j \right) = d_i \left(1 - d_i\right) = d_i - d_i^2 = |A_{i,i}|.$$

Thus, $\boldsymbol{A}$ is a symmetric matrix with non-negative entries on the diagonal that is diagonally dominant. By Theorem 2.25, $\boldsymbol{A} \succeq \boldsymbol{0}$.

2.13. For each of the following matrices determine whether they are positive/negative semidefinite/definite or indefinite.

(i) $\boldsymbol{A} = \begin{pmatrix} 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 3 \end{pmatrix}$

The matrix $\boldsymbol{A}$ is positive semidefinite. To see why, first define the 2×2 matrices

$$\boldsymbol{A}_1 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}, \boldsymbol{A}_2 = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}.$$

Note that $\boldsymbol{A}_1$ is positive semidefinite (its eigenvalues are $0, 4$), and $\boldsymbol{A}_2$ is positive definite (its eigenvalues are $2, 4$). Therefore, given any $\boldsymbol{x} = (x_1, x_2, x_3, x_4)^T \in \mathbb{R}^4$, we have,

$$\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} = (x_1, x_2)\boldsymbol{A}_1(x_1, x_2)^T + (x_3, x_4)\boldsymbol{A}_2(x_3, x_4)^T \geq 0,$$

since $\boldsymbol{A}_1 \succeq \boldsymbol{0}$ and $\boldsymbol{A}_2 \succ \boldsymbol{0}$. The matrix $\boldsymbol{A}$ cannot be cannot be positive definite since choosing $\boldsymbol{x} = (1, -1, 0, 0)^T \in \mathbb{R}^4$ results in $\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} = 0$. Therefore, $\boldsymbol{A} \succeq \boldsymbol{0}$.

(ii) $\boldsymbol{B} = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 3 & 3 \\ 2 & 3 & 3 \end{pmatrix}$ By direct computation, i.e., computing the roots of $\det(\boldsymbol{B} - \lambda \boldsymbol{I}) = -\lambda(\lambda^2 - 8\lambda + 4)$, the eigenvalues of $\boldsymbol{B}$ are $0, 4 \pm 2\sqrt{3}$. Therefore, $\boldsymbol{B}$ is also positive semidefinite.

2.15. For each of the following functions, determine whether it is coercive or not:

(i) $f(x_1, x_2) = x_1^4 + x_2^4$.
This function is coercive: as $\|\boldsymbol{x}\| \to \infty$, we must have either $|x_1| \to \infty$ and/or $|x_2| \to \infty$. Thus, for large $\|\boldsymbol{x}\|$, we have either $|x_1|^4 = x_1^4 \to \infty$ and/or $|x_2|^4 = x_2^4 \to \infty$. Since both x_1^4 and x_2^4 are non-negative, then

$$\lim_{\|\boldsymbol{x}\|\to\infty} f(\boldsymbol{x}) \geq \lim_{\|\boldsymbol{x}\|\to\infty} \min\{x_1^4, x_2^4\} = \infty.$$

(iii) $f(x_1, x_2) = 2x_1^2 - 8x_1x_2 + x_2^2$.
This function is not coercive: along the line $x_1 = x_2$ we have $f(x_1, x_2) = -5x_1^2$. Then sending $\|\boldsymbol{x}\| \to \infty$ by taking $\boldsymbol{x} = (k, k)^T$ as $k \uparrow \infty$ shows that $f(\boldsymbol{x}) \to -\infty$

(iv) $f(x_1, x_2) = 4x_1^2 + 2x_1x_2 + 2x_2^2$.
This function is coercive:

$$f(x_1, x_2) = (x_1 + x_2)^2 + 3x_1^2 + x_2^2 \geq x_1^2 + x_2^2 = \|\boldsymbol{x}\|^2,$$

so that $f \uparrow \infty$ as $\|\boldsymbol{x}\| \uparrow \infty$.

(vii) $f(x_1, x_2) = \frac{\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|+1}$, where $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is positive definite.
This function is coercive: Since $\boldsymbol{A} \succ \boldsymbol{0}$, then it smallest eigenvalue $\lambda_{\min}(\boldsymbol{A})$ satisfies $\lambda_{\min}(\boldsymbol{A}) > 0$. Since the extremal eigenvalues of $\boldsymbol{A}$ bound the possible values of the Rayleigh quotient, we therefore have,

$$\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} \geq \|\boldsymbol{x}\|^2 \lambda_{\min}(\boldsymbol{A}).$$

Therefore,

$$f(\boldsymbol{x}) = \frac{\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\| + 1} \geq \lambda_{\min}(\boldsymbol{A}) \frac{\|\boldsymbol{x}\|^2}{\|\boldsymbol{x}\| + 1},$$

Since $\lambda_{\min}(\boldsymbol{A}) > 0$, we therefore have that $f \uparrow \infty$ as $\|\boldsymbol{x}\| \uparrow \infty$.

2.17. For each of the following functions, find all the stationary points and classify them according to whether they are saddle points, strict/nonstrict local/global minimum/maximum points:

(i) $f(x_1, x_2) = (4x_1^2 - x_2)^2$.
We have,

$$\nabla f(x_1, x_2) = 2(4x_1^2 - x_2) \begin{pmatrix} 8x_1 \\ -1 \end{pmatrix},$$

so that there are infinitely many stationary points lying along the curve defined by $x_2 = 4x_1^2$ (which is a parabola). Note that $f(x_1, x_2) \geq 0$ for any (x_1, x_2), and $f(x_1, x_2) = 0$ if and only if (x_1, x_2) is a stationary point satisfying $x_2 = 4x_1^2$. Therefore, the stationary points are all points of the form $(k, 4k^2)$ for any $k \in \mathbb{R}$, and they are all nonstrict global and local minima.

(iii) $f(x_1, x_2) = 2x_2^3 - 6x_2^2 + 3x_1^2 x_2$.
We have,

$$\nabla f(x_1, x_2) = 3 \begin{pmatrix} 2x_1x_2 \\ 2x_2^2 - 4x_2 + x_1^2 \end{pmatrix}.$$

The first component of the gradient vanishes if and only if $x_1 = 0$ and/or $x_2 = 0$. If $x_1 = 0$, then the second component vanishes when $x_2 = 0, 2$. Therefore, two stationary points are $(0, 0)$ and $(0, 2)$. If instead we force $x_2 = 0$, then the second component requires $x_1 = 0$, which is a point we have already recorded. The Hessian is given by

$$\nabla^2 f(x_1, x_2) = 6 \begin{pmatrix} x_2 & x_1 \\ x_1 & 2(x_2 - 1) \end{pmatrix}.$$

At the stationary point $(0, 2)$,

$$\nabla^2 f(0, 2) = 12\boldsymbol{I} \succ \mathbf{0},$$

so that $(0, 2)$ is a strict local minimum. It cannot be a global minimum since $\lim_{x_2 \to -\infty} f(0, x_2) = -\infty$. At the stationary point $(0, 0)$,

$$\nabla^2 f(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & -12 \end{pmatrix} \preceq \mathbf{0},$$

which is inconclusive. However, one can see that $(0, 0)$ is a saddle point: along $x_1 = 0$, then $f(0, x_2) = 2x_2^3 - 6x_2^2$, which is negative for small x_2. But along $x_1 = \sqrt{2x_2}$, we have $f(\sqrt{2x_2}, x_2) = 2x_2^3$, which is positive for small $x_2 > 0$. Therefore, the two stationary points are $(0, 0)$ which is a strict local minimum, and $(0, 0)$, which is a saddle point.

(vi) $f(x_1, x_2) = 2x_1^2 + 3x_2^2 - 2x_1 x_2 + 2x_1 - 3x_2$.
The function f is a quadratic function whose Hessian is

$$\nabla^2 f = \begin{pmatrix} 4 & -2 \\ -2 & 6 \end{pmatrix},$$

which is positive definite. Therefore, there is exactly 1 stationary point and it is a global minimum. The gradient is given by,

$$\nabla f(x_1, x_2) = \begin{pmatrix} 4x_1 - 2x_2 + 2 \\ -2x_1 + 6x_2 - 3 \end{pmatrix},$$

and therefore the stationary point is $(x_1, x_2) = \left(-\frac{3}{10}, \frac{2}{5}\right)$, and it is a strict global minimum.

(vii) $f(x_1, x_2) = x_1^2 + 4x_1 x_2 + x_2^2 + x_1 - x_2$.
This is again a quadratic function; the Hessian is

$$\nabla f^2 = \begin{pmatrix} 2 & 4 \\ 4 & 2 \end{pmatrix},$$

whose eigenvalues are $-2, 6$, so the Hessian is indefinite. Therefore, stationary points are saddle points. The gradient is

$$\nabla f(x_1, x_2) = \begin{pmatrix} 2x_1 + 4x_2 + 1 \\ 4x_1 + 2x_2 - 1 \end{pmatrix},$$

so the stationary point is $(x_1, x_2) = \left(\frac{1}{2}, -\frac{1}{2}\right)$, and it is a saddle point.

2.18. Let f be a twice continuously differentiable function over $\mathbb{R}^n$. Suppose that $\nabla^2 f(\boldsymbol{x}) \succ \mathbf{0}$ for any $\boldsymbol{x} \in \mathbb{R}^n$. Prove that a stationary point of f is necessarily a strict global minimum point.

Suppose $\boldsymbol{x}_*$ is a stationary point, i.e., $\nabla f(\boldsymbol{x}_*) = \mathbf{0}$. By Taylor's theorem, we can represent $f(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathbb{R}^n$ not equal to $\boldsymbol{x}_*$ by expanding around the stationary point $\boldsymbol{x}_*$:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_*) + \nabla f(\boldsymbol{x}_*)(\boldsymbol{x} - \boldsymbol{x}_*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_*)^T \nabla^2 f(\boldsymbol{z})(\boldsymbol{x} - \boldsymbol{x}_*)^T,$$

for some $\boldsymbol{z} \in [\boldsymbol{x}_*, \boldsymbol{x}]$. The above relation simplifies to

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_*) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_*)^T \nabla^2 f(\boldsymbol{z})(\boldsymbol{x} - \boldsymbol{x}_*)^T > 0,$$

where the inequality is true since $\nabla^2 f \succ \mathbf{0}$ at every point and $\boldsymbol{x} \neq \boldsymbol{x}_*$. Thus, $f(\boldsymbol{x}) > f(\boldsymbol{x}_*)$ for every $\boldsymbol{x} \neq \boldsymbol{x}_*$, showing that $\boldsymbol{x}_*$ is a strict global minimum point.

Additional problems:

P1. Define

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

Using your favorite software, visualize a plot of the Rayleigh quotient $f(\boldsymbol{x}) = R_{\boldsymbol{A}}(\boldsymbol{x})$ for $\boldsymbol{x} \in [-3, 3]^2 \backslash \{\mathbf{0}\}$, and generate a contour plot for f. Use this visualization to verify the maximum and minimum values of f, as well as the set of $\boldsymbol{x}$ that are maximizers and minimizers.

The matrix $\boldsymbol{A}$ is symmetric and has orthogonal eigenvalue decomposition,

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T, \qquad \boldsymbol{\Lambda} = \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix}, \qquad \boldsymbol{U} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}.$$

We therefore analytically know the maximum and minimum of the Rayleigh quotient $f(\boldsymbol{x})$. In particular,

$$\max_{\boldsymbol{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}} f(\boldsymbol{x}) = 5, \qquad \min \boldsymbol{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\} f(\boldsymbol{x}) = 0,$$

and,

$$\operatorname*{argmax}_{\boldsymbol{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}} f(\boldsymbol{x}) = \left\{ k(1,2)^T \;\middle|\; k \in \mathbb{R},\ k \neq 0 \right\}$$
$$\operatorname{argmin} \boldsymbol{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\} f(\boldsymbol{x}) = \left\{ k(2,-1)^T \;\middle|\; k \in \mathbb{R},\ k \neq 0 \right\}.$$

Thus, we expect f to have a maximum value of 5 along the line with equation $2x_1 - x_2 = 0$ (with $\boldsymbol{x} \neq \mathbf{0}$), and f should have a minimum value of 0 along the line with equation $x_1 + 2x_2 = 0$. These are visually verified with the plots shown in Figure 1. Code reproducing this plot is available at
`https://github.com/akilnarayan/2021Fall-Optimization-homework2`.

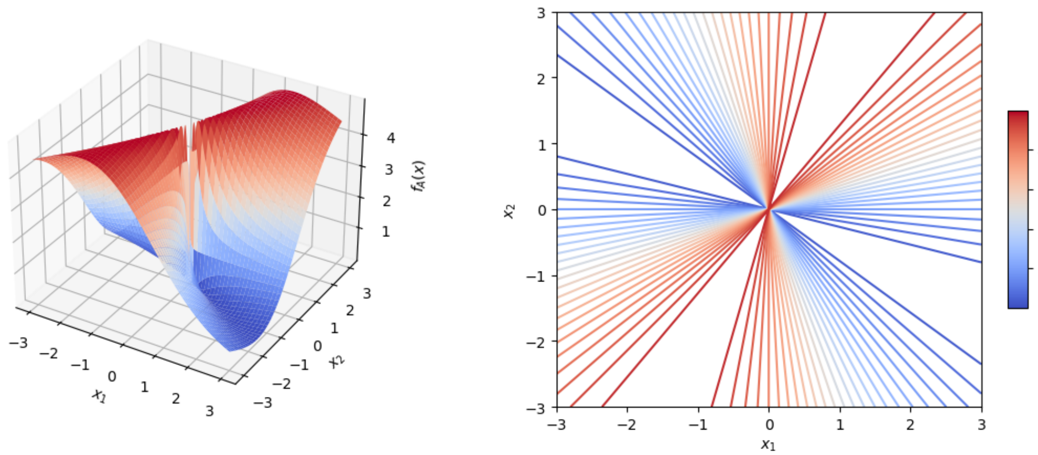

Figure 1: Visualization of the Rayleigh quotient (left) and its contour levels (right) associated to Problem P1.