# MATH 680: Assignment 2

*Annik Gougeon, David Fleischer*

*Last Update: 23 February, 2018*

## Question 1

*Proof.* Let $C = \{x \in \mathbb{R}^n : Ax \leq b\}$ be our set of interest. Let $x, y \in C$, and let $t \in [0, 1]$ be an arbitrary real-valued scalar. Then,

$$\begin{aligned} A\left(tx + (1-t)y\right) &= tAx + (1-t)Ay \\ &\leq tb + (1 - tb) \\ &= b. \end{aligned}$$

Thus,

$$x, y \in C \implies tx + (1-t)y \in C, \quad \text{for all } 0 \leq t \leq 1.$$

That is, $C$ is a convex set, as desired. $\square$

## Question 2

### 2.1

Let $f$ be defined as

$$f(x, y) = |xy| + a(x^2 + y^2).$$

Recall that $f$ is a convex function if and only if its Hessian is positive semi-definite.

**Case 1**: $xy > 0 \implies f_+(x, y) = xy + a(x^2 + y^2)$.

We begin by finding the gradient of $f_+(x, y)$ to be

$$\nabla f_+(x, y) = (y + 2ax, x + 2ay).$$

Therefore, the Hessian is

$$\nabla^2 f_+(x, y) = \begin{bmatrix} 2a & 1 \\ 1 & 2a \end{bmatrix}.$$

For this Hessian to be positive semi-definite, we require its eigenvalues to be nonnegative. We find the eigenvalues to be $\lambda_1 = 2a - 1$ and $\lambda_2 = 2a + 1$. As a direct result, the Hessian is positive semi-definite (and therefore, convex) if and only if $a \geq \frac{1}{2}$.

Furthermore, in order for $f_+$ to be strongly convex, we must have the eigenvalues $\lambda\left(\nabla^2 f_+(x, y) - m\mathbb{I}\right) = \{\lambda_i - m\}_{i=1,2}$ be nonnegative for some $m > 0$ (i.e., $\nabla^2 f_+(x, y) - m\mathbb{I}$ must be positive semidefinite for some $m > 0$). We see that

$$\begin{aligned} \lambda_1 - m &= 2a - 1 - m \\ \lambda_2 - m &= 2a + 1 - m. \end{aligned}$$

The above eigenvalues will be nonnegative for $a > \frac{1}{2}$ and $m \leq a$. Therefore, $f_+$ is strongly convex if $a > \frac{1}{2}$.

**Case 2**: $xy < 0 \implies f_-(x,y) = -xy + a(x^2 + y^2)$.

We find $f_-$ to have gradient

$$\nabla f_-(x,y) = (-y + 2ax, -x + 2ay),$$

and Hessian

$$\nabla^2 f_-(x,y) = \begin{bmatrix} 2a & -1 \\ -1 & 2a \end{bmatrix}.$$

Once again, for the Hessian to be positive semidefinite, we must find nonnegative eigenvalues $\lambda_i \geq 0$, $i = 1, 2$. As above, we find the eigenvalues to be $\lambda_1 = 2a - 1$ and $\lambda_2 = 2a + 1$. Therfore, the Hessian must be positive semidefinite (and so $f_-$ must be convex) if and only if $a \geq \frac{1}{2}$ and strongly convex if $a > \frac{1}{2}$.

**Case 3**: $xy = 0$. For such a case we note three possible scenarios

$$xy = 0 \implies f_0(x,y) = \begin{cases} ay^2, & \text{if } x = 0 \text{ and } y \neq 0, \\ ax^2, & \text{if } x \neq 0 \text{ and } y = 0, \\ 0, & \text{else.} \end{cases}$$

If $x = 0$ and $y \neq 0$ then $f_0(x,y) = ay^2$ so

$$\frac{\partial f_0}{\partial y} = 2ay,$$

and

$$\frac{\partial^2 f_0}{\partial y^2} = 2a.$$

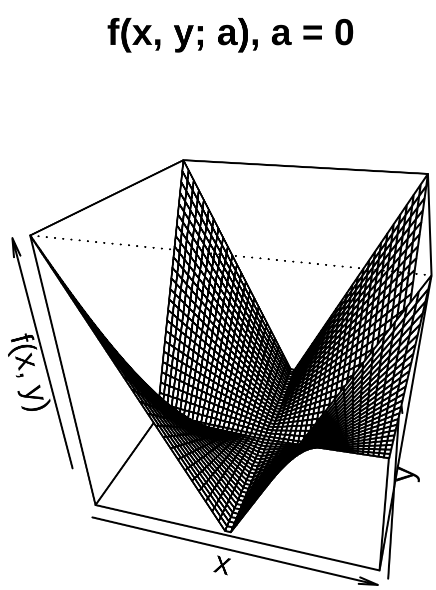
It follows that for $x = 0$ and $y \neq 0$ we find $f_0(x,y)$ to be be convex for $a \geq \frac{1}{2}$ and strongly convex for $a > \frac{1}{2}$. By symmetry, the same result is found for the case where $x \neq 0$ and $y = 0$ (corresponding to $f_0(x,y) = ax^2$), permitting us to arrive to the same conclusions.

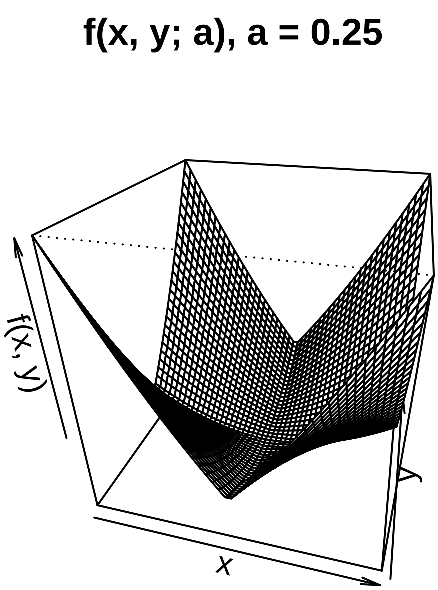
Finally, for the last scenario $x = y = 0$, we note that the domain of $f_0$ under $x = y = 0$ is defined at only a single point, trivially satisfying strong convexity

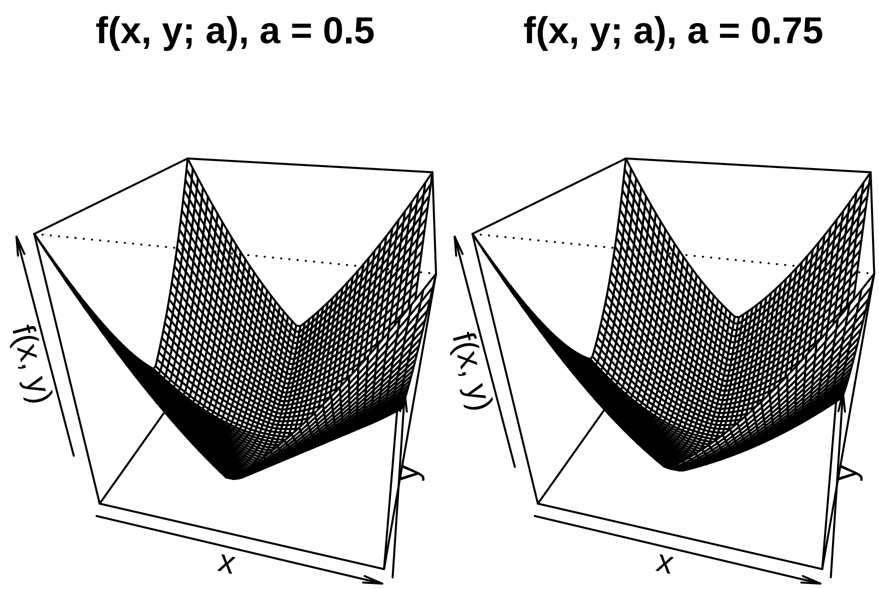
$$(\nabla f(z_1) - \nabla f(z_2))^T (z_1 - z_2) \geq m \|z_1 - z_2\|_2^2$$

on all points in its domain, as desired.

Presented below figures of $f$ evaluated on $[-1,1] \times [-1,1]$ for $a \in \{0, 0.25, 0.5, 0.75\}$.






### 2.2

#### 2.2.a

For $x \in \mathbb{R}^n_{++}$ we find gradient of $f(x) = -\sum_{i=1}^n \log x_i$ to be

$$\nabla f(x) = -\left[x_1^{-1}, x_2^{-1}, \cdots, x_n^{-1}\right],$$

with corresponding Hessian

$$\nabla^2 f(x) = \begin{bmatrix} x_1^{-2} & 0 & \cdots & 0 \\ 0 & x_2^{-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n^{-2} \end{bmatrix}.$$

Since $\nabla^2 f(x)$ is diagonal we may immediately obtain its eigenvalues $\{\lambda_i\}_{i=1}^n$ such that

$$\lambda_i = x_i^{-2}.$$

We see that, since $x \in \mathbb{R}^n_{++} \iff x_i > 0, i = 1, ..., n$, all eigenvalues must be strictly positive $\lambda_i > 0$. Therefore, $f$ must be strongly convex (since there must exist constant $m$ such that $\lambda_i - m \geq 0$), as desired.

#### 2.2.b

The entropy function, $f : \left\{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\right\} \to \mathbb{R}$ is defined as

$$f(x) = \begin{cases} -\sum_{i=1}^n x_i \log x_i & x > 0, \\ 0 & x = 0. \end{cases}$$

Now,

$$\nabla f(x) = (f_1'(x_1), f_2'(x_2), ..., f_n'(x_n))$$

where

$$f_i'(x_i) = \begin{cases} -\left(1 + \log x_i\right), & x_i > 0 \\ 0 & x_i = 0. \end{cases}$$

Likewise,

$$\nabla^2 f(x) = \begin{bmatrix} f_1''(x_1) & 0 & \cdots & 0 \\ 0 & f_2''(x_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f_n''(x_n) \end{bmatrix},$$

where

$$f_i''(x_i) = \begin{cases} -x_i^{-1}, & x_i > 0 \\ 0, & x_i = 0. \end{cases}$$

Since $\nabla^2 f(x)$ is diagonal, we immediately obtain its eigenvalues

$$\{\lambda_i\}_{i=1}^n = \{f_i''(x_i)\}_{i=1}^n$$

Clearly, for any $x \in \mathrm{dom}(f) = \mathbb{R}_+^n$ we find that all $\lambda_i \leq 0$. Since $\nabla^2 f$ always has nonpositive eigenvalues, we conclude that $\nabla^2(-f)$ must have nonnegative eigenvalues. Therefore, $f$ is nonconvex (and in particular, $f$ is concave since $-f$ is convex), as desired.

## 2.3

*Proof.* ($\implies$) Suppose $f$ is convex. Then, $\mathrm{dom}(f)$ is a convex set, and, for all $x, y \in \mathrm{dom}(f)$ and $t \in [0,1]$,

$$\begin{aligned} & f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \\ & \qquad \iff f(t(x-y)+y) - f(y) \leq t\left(f(x) - f(y)\right) \\ & \qquad \iff \frac{f(t(x-y)+y) - f(y)}{t} \leq f(x) - f(y) \\ & \qquad \iff \frac{f(t(x-y)+y) - f(y)}{t} + f(y) \leq f(x) \end{aligned}$$

Note that if we take the limit of our first term as $t \to 0$, for finite $x, y$,

$$\begin{aligned} \lim_{t\to 0} \frac{f(t(x-y)+y) - f(y)}{t} &= \frac{\partial}{\partial t} f(t(x-y)+y) \bigg]_{t=0} \\ &= \nabla f(t(x-y)+y)^T (x-y) \bigg]_{t=0} \\ &= \nabla f(y)^T (x-y). \end{aligned}$$

Therefore, taking the limit of our inequality above as $t \to 0$,

$$\begin{aligned} & \lim_{t\to 0} \left( \frac{f(t(x-y)+y) - f(y)}{t} + f(y) \right) \leq \lim_{t\to 0} f(x) \\ & \qquad \iff \nabla f(y)^T (x-y) + f(y) \leq f(x). \end{aligned}$$

By symmetry we may swap $x$ and $y$ to obtain

$$f \text{ convex} \implies \mathrm{dom}(f) \text{ convex and } f(y) \geq f(x) + \nabla f(x)^T (y-x),$$

as desired.

($\Longleftarrow$) Suppose $\operatorname{dom}(f)$ is convex and, for $x, y \in \operatorname{dom}(f)$, $x \neq y$,

$$\nabla f(x)^T(y-x) + f(x) \leq f(y).$$

Since $\operatorname{dom}(f)$ is convex we find $z = tx + (1-t)y \in \operatorname{dom}(f)$, $t \in [0,1]$. Then, for such $x, y, z$,

$$\begin{aligned}
\nabla f(z)^T(x-z) + f(z) &\leq f(x) \\
\nabla f(z)^T(y-z) + f(z) &\leq f(y).
\end{aligned}$$

Multiplying our first inequality by $t$ and the second by $(1-t)$, and then adding the two yields

$$\begin{aligned}
&t\left[\nabla f(z)^T(x-z) + f(z)\right] + (1-t)\left[\nabla f(z)^T(y-z) + f(z)\right] \leq tf(x) + (1-t)f(y) \\
&\qquad \iff t\nabla f(z)^T(x-z) + (1-t)\nabla f(z)^T(y-z) + f(z) \leq tf(x) + (1-t)f(y) \\
&\qquad \iff \nabla f(z)^T\left[t(x-z) + (1-t)(y-z)\right] + f(z) \leq tf(x) + (1-t)f(y) \\
&\qquad \iff \nabla f(z)^T\left[tx + (1-t)y - z\right] + f(z) \leq tf(x) + (1-t)f(y) \\
&\qquad \iff f(tx + (1-t)y) \leq tf(x) + (1-t)f(y),
\end{aligned}$$

where the final line was achieved by recalling that $z = tx + (1-t)y$. Therefore,

$$f \text{ convex} \Longleftarrow \operatorname{dom}(f) \text{ convex and } f(y) \geq f(x) + \nabla f(x)^T(y-x),$$

as desired. $\square$

# Question 3

## 3.1

Let $f$ be convex and twice differentiable.

($1 \implies 2$)

If $\nabla f$ is Lipschitz with constant $L > 0$ then, by definition,

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x-y\|_2.$$

By the Cauchy-Schwartz inequality

$$|\langle u, v\rangle| \leq \|u\| \cdot \|v\|,$$

substituting $u = \nabla f(x) - \nabla f(y)$ and $v = x - y$ in Cauchy-Schwartz, and by $L$-Lipschitz continuity assumed via (1), we obtain

$$\begin{aligned}
\left(\nabla f(x) - \nabla f(y)\right)^T(x-y) &\leq \|\nabla f(x) - \nabla f(y)\|_2 \cdot \|x-y\|_2 \quad \text{(Cauchy-Schwartz)} \\
&\leq L\|x-y\|_2^2 \quad (L\text{-Lipschitz}),
\end{aligned}$$

as desired.

($1 \implies 3$)

Assume (1) holds

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x-y\|_2,$$

and define $g(t) = \nabla f\left(y + t(x-y)\right)$ so that

$$\begin{aligned}
g(0) &= \nabla f(y) \\
g(1) &= \nabla f(x) \\
g'(t) &= \nabla^2 f\left(y + t(x-y)\right) \cdot (x-y).
\end{aligned}$$

Therefore,

$$\|\nabla f(x) - \nabla f(y)\|_2 = \|g(1) - g(0)\|_2.$$

However, applying the Mean Value Theorem on $g$, there must exist some $z \in [0,1]$ such that[1]

$$g'(z) = \frac{g(1) - g(0)}{1 - 0} = g(1) - g(0).$$

Thus,

$$\begin{aligned}
&\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x-y\|_2 \\
&\implies \|g(1) - g(0)\|_2 \leq L\|x-y\|_2 \\
&\implies \|g'(z)\|_2 \leq L\|x-y\|_2 \\
&\iff \|\nabla^2 f\left(z^*\right)(x-y)\|_2 \leq L\|x-y\|_2, \quad \text{for some } z^* \in [x,y] \\
&\iff \|\nabla^2 f\left(z^*\right)\|_2 \cdot \|x-y\|_2 \leq L\|x-y\|_2 \\
&\iff \|\nabla^2 f\left(z^*\right)\|_2 \leq L \\
&\iff \nabla^2 f\left(z^*\right) \preceq L\mathbb{I}.
\end{aligned}$$

Therefore, since $x$ and $y$ were arbtirary

$$\nabla^2 f(x) \preceq L \quad \text{for all } x,$$

as desired.

($3 \implies 4$)

Assume (3) holds so that $\nabla^2 f(x) \preceq L\mathbb{I}$ and define $g$ such that

$$g(z) = \frac{L}{2} z^T z - f(z) \iff f(z) = \frac{L}{2} z^T z - g(z).$$

Then,

$$\nabla^2 f(x) \preceq L\mathbb{I} \iff 0 \preceq \nabla^2 g(x),$$

[1] We can be sure that such a point exists since we have assumed $f$ is twice differentiable (and so applying the Mean Value Theorem on $g$ is permissible) and we have assumed that $f$ is convex (so such a point $z \in [0,1]$ must exist).

informing us that $g$ must be convex. From the convexity of $g$, we must have that $g$ satisfies the first order condition outlined in 2.3
$$g(y) \geq g(x) + \nabla g(x)^T (y - x).$$

Plugging in our definition of $g$ yields

$$\begin{aligned}
\frac{L}{2} y^T y - f(y) &\geq \frac{L}{2} x^T x - f(x) + (Lx - \nabla f(x))^T (y - x) \\
\iff -f(y) &\geq -f(x) - \nabla f(x)^T (y - x) + \frac{L}{2} x^T x - L x^T x + L x^T y - \frac{L}{2} y^T y \\
&= -f(x) - \nabla f(x)^T (y - x) - \frac{L}{2} \left( y^T y - 2 x^T y + x^T x \right) \\
&= -f(x) - \nabla f(x)^T (y - x) - \frac{L}{2} \|y - x\|_2^2 \\
\iff f(y) &\leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|_2^2,
\end{aligned}$$

as desired.

($4 \implies 1$)

Assume (4) holds so that
$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|_2^2.$$

Next, define $g(z) = \frac{L}{2} z^T z - f(z) \iff f(z) = \frac{L}{2} z^T z - g(z)$ and apply (4) to yield

$$\begin{aligned}
\frac{L}{2} y^T y - g(y) &\leq \frac{L}{2} x^T x - g(x) + (Lx - \nabla g(x))^T (y - x) + \frac{L}{2} \|y - x\|_2^2 \\
\iff g(x) - g(y) &\leq -\frac{L}{2} x^T x - \frac{L}{2} y^T y + L x^T y - L x^T x - \nabla g(x)^T (y - x) + \frac{L}{2} \|y - x\|_2^2 \\
&= -\frac{L}{2} \left( y^T y - 2 x^T y + x^T x \right) - \nabla g(x)^T (y - x) + \frac{L}{2} \|y - x\|_2^2 \\
&= -\frac{L}{2} \|y - x\|_2^2 - \nabla g(x)^T (y - x) + \frac{L}{2} \|y - x\|_2^2 \\
&= -\nabla g(x)^T (y - x) \\
\iff g(y) &\geq g(x) + \nabla g(x)^T (y - x).
\end{aligned}$$

Thus, recalling our first order condition for convexity from Question 2.3 we may conclude that $g$ is indeed convex. Now, if $g$ is convex then
$$\nabla^2 g(x) \succeq 0 \iff LI \succeq \nabla^2 f(x).$$

However, defining $h(t) = \nabla f(y + t(x - y))$ so that

$$\begin{aligned}
h(0) &= \nabla f(x) \\
h(1) &= \nabla f(y) \\
h'(t) &= \nabla^2 f(y + t(x - y)) \cdot (x - y).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
\|\nabla f(x) - \nabla f(y)\|_2 &= \|\nabla f(y) - \nabla f(x)\|_2 \\
&= \|h(1) - h(0)\|_2 \\
&= \| \int_0^1 h'(t)\, dt \|_2 \\
&\le \int_0^1 \|h'(t)\|_2 \, dt \quad \text{(triangle inequality)} \\
&= \int_0^1 \|\nabla^2 f\left(y + t(x-y)\right) \cdot (x-y)\|_2 \, dt \\
&\le \int_0^1 L\|x-y\|_2 \, dt \\
&= L\|x-y\|,
\end{aligned}
$$

where the final inequality was obtained through

$$
\nabla^2 f(x) \preceq L\mathbb{I} \iff \|\nabla^2 f(x)\|_2 \le L.
$$

Putting this together we find that $f(y) \le f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|_2^2$ implies that $\nabla f$ is $L$-Lipschitz, as desired. \

Since we have shown (1) $\implies$ (2), (3), (3) $\implies$ (4), and (4) $\implies$ (1) we may conclude that all the above statements are indeed equivalent.

### 3.2

(1 $\implies$ 2)

Assume (1) so that $f$ is strongly convex with parameter $m > 0$. Then, by definition, $f(x) - \frac{m}{2}\|x\|_2^2$ is convex. Defining $g(x) = f(x) - \frac{m}{2}\|x\|_2^2$ we find

$$
\nabla g(x) = \nabla f(x) - mx.
$$

However, recall that $g$ is convex if and only if

$$
0 \le \left(\nabla g(x) - \nabla g(y)\right)^T (x-y).
$$

Therefore,

$$
\begin{aligned}
0 &\le \left(\nabla f(x) - mx - \nabla f(y) + my\right)^T (x-y) \\
&= \left(\nabla f(x) - \nabla f(x)\right)^T (x-y) - m(x-y)^T(x-y) \\
&= \left(\nabla f(x) - \nabla f(x)\right)^T (x-y) - m\|x-y\|_2^2 \\
\iff m\|x-y\|_2^2 &\le \left(\nabla f(x) - \nabla f(x)\right)^T (x-y),
\end{aligned}
$$

as desired.

$(2 \implies 3)$

Assuming (2) holds we find

$$
\begin{aligned}
&(\nabla f(x) - \nabla f(y))^T (x - y) \geq m\|x - y\|_2^2 \\
&\iff \nabla f(x) + x\nabla^2 f(x) - y\nabla^2 f(x) + y\nabla f(y) \geq 2m\|x - y\| \\
&\iff \nabla^2 f(x)(x - y) + (\nabla f(x) - \nabla f(y)) \geq 2m\|x - y\| \\
&\iff \nabla^2 f(x)\|x - y\| \geq m\|x - y\|,
\end{aligned}
$$

since $\|\nabla f(x) - \nabla f(y)\| \geq m\|x - y\|$. It follows that

$$\nabla^2 f(x) \succeq mI,$$

as desired.

$(1 \implies 4)$

Once again assume (1) so that $f$ is strongly convex with parameter $m > 0$. Then, by definition, $f(x) - \frac{m}{2}\|x\|_2^2$ is convex. Defining $g(x) = f(x) - \frac{m}{2}\|x\|_2^2$ we find

$$\nabla g(x) = \nabla f(x) - mx.$$

Now, apply the first order condition from 2.3 on $g$

$$g(y) \geq g(x) + \nabla g(x)^T (y - x).$$

Hence,

$$
\begin{aligned}
f(y) - \frac{m}{2}\|y\|_2^2 &\geq f(x) - \frac{m}{2}\|x\|_2^2 + (\nabla f(x) - mx)^T (y - x) \\
\iff f(y) &\geq f(x) + \nabla f(x)^T (y - x) + \frac{m}{2} y^T y - \frac{m}{2} x^T x - m x^T y + m x^T x \\
&= f(x) + \nabla f(x)^T (y - x) + \frac{m}{2} y^T y - m x^T y + \frac{m}{2} x^T x \\
&= f(x) + \nabla f(x)^T (y - x) + \frac{m}{2}\|y - x\|_2^2 \\
\iff f(y) &\geq f(x) + \nabla f(x)^T (y - x) + \frac{m}{2}\|y - x\|_2^2,
\end{aligned}
$$

as desired.

$(4 \implies 1)$

We begin with

$$
\begin{aligned}
&f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{m}{2}\|y - x\|_2^2 \\
&\implies \nabla f(y) \geq \nabla f(x) + m\|y - x\| \\
&\implies \|\nabla f(y) - \nabla f(x)\| \geq m\|y - x\| \\
&\implies \|\nabla f(x) - \nabla f(y)\| \geq m\|x - y\|.
\end{aligned}
$$

Therefore, $f$ must be strongly convex, as desired.

# Question 4

## 4.a

For parts 1 and 2 we load `CVXR` in order to solve the (convex) 2d fused lasso problem.

```
library(CVXR)
```

### 4.a.1

We load our data `circle.csv` and define some useful constants

```
circle <- as.matrix(read.csv("../data/circle.csv", header = F))
n <- length(circle); nr <- nrow(circle); nc <- ncol(circle)
```

Next, we translate the 2d fused lasso penalty

$$\lambda \sum_{\{i,j\} \in E} |\theta_i - \theta_j|, \quad E = \text{set of edges } \{(\text{i,j})\} \text{ connecting adjascent pixels}$$

into a `CVXR` compatible function

```
fused_lasso_2d <- function(theta, lambda = 0) {
  nr <- nrow(theta); nc <- ncol(theta)
  S <- theta[1:(nr - 1),] - theta[2:nr,] # SOUTH
  N <- theta[2:nr,] - theta[1:(nr - 1),] # NORTH
  E <- theta[,1:(nc - 1)] - theta[,2:nc] # EAST
  W <- theta[,2:nc] - theta[,1:(nc - 1)] # WEST
  lambda * (sum(abs(S)) + sum(abs(N)) + sum(abs(E)) + sum(abs(W)))
}
```

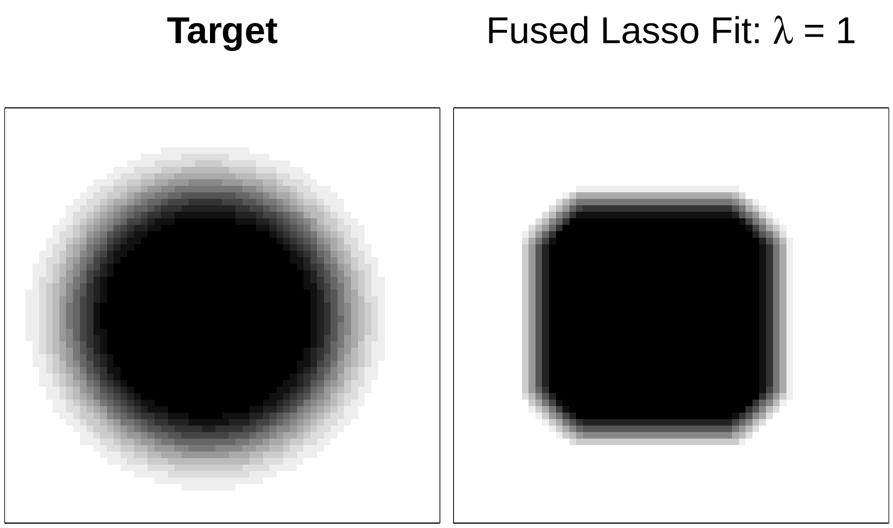
as well as defining some parameters and variables, as well as our $\ell_2$ loss $\frac{1}{2}\|Y - \theta\|_2^2$

```
lambda <- 1
theta <- Variable(nr, nc)
theta_hat <- matrix(0, nrow = nr, ncol = nc)
loss <- sum(0.5 * (circle - theta)^2)
```

Finally, we run run `CVXR` on the problem

```
obj <- loss + fused_lasso_2d(theta, lambda)
prob <- Problem(Minimize(obj))
res <- solve(prob)
theta_hat <- res$getValue(theta)
```

Comparing the original data with the fused fit we see that the fused fit of the circle is essentially a square with truncated/rounded corners, as seen below.




We can understand such behaviour by the noting the conflicting behaviour of the $\ell_2$ term and the $\ell_1$ fused term in our objective function. The $\ell_2$-loss term

$$\frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_i)^2$$

is minimized as $\hat{\theta}_i \longrightarrow y_i$, while the $\ell_1$-fused term

$$\lambda \sum_{\{i,j\}\in E} |\theta_i - \theta_j|$$

is minimized when adjascent cells are 'close' to each other (and as $\lambda \to \infty$ we obtain $\hat{\theta}_i \to \bar{y}$). Deep in the interior/far out in the extertior of the circle, the objective estimates $\hat{\theta}_i \approx y_i$ since essentially all adjascent values will correspond to 0 (black) or 1 (white), respectively. However, along the boundary of the two regions we balance the two objectives by detecting a 'changepoint' in the data (as a consequence of using the $\ell_1$ norm), inside of which most observations are close to 0 and beyond which most observations are close to 1.

### 4.a.2

We load the `lenna_64.csv` data and define some useful constants

```
lenna <- as.matrix(read.csv("../data/lenna_64.csv", header = F))
n <- length(lenna); nr <- nrow(lenna); nc <- ncol(lenna)
```

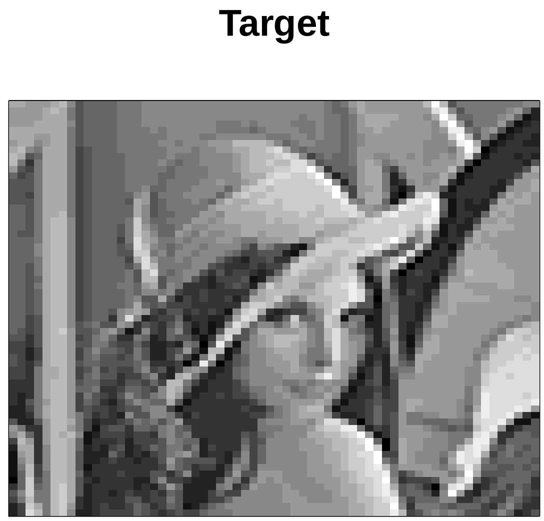
and run `CVXR` on the 2d fused lasso problem in the same way as we did for `circle.csv`, but now over a vector of tuning parameters $\{\lambda_k\}_{k=0,\ldots,8} = \left\{10^{-k/4}\right\}_{k=0,\ldots,8}$

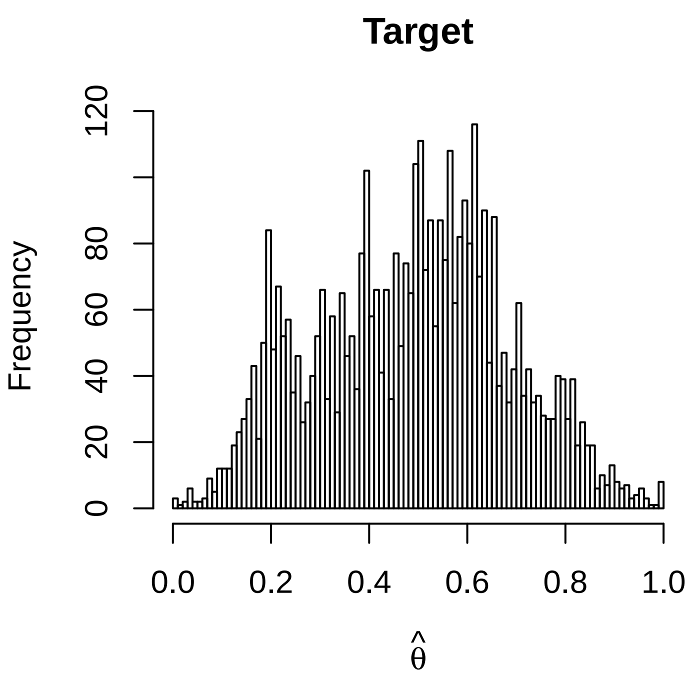
```
lambda_vals <- 10^(-(0:8)/4)
theta_vals <- vector(mode = 'list', length = length(lambda_vals))
obj_val <- vector(mode = 'numeric', length = length(lambda_vals))
theta <- Variable(nr, nc)
loss <- sum(0.5 * (lenna - theta)^2)

pt <- proc.time()
for (i in 1:length(lambda_vals)) {
  lambda <- lambda_vals[i]
  obj <- loss + fused_lasso_2d(theta, lambda)
```

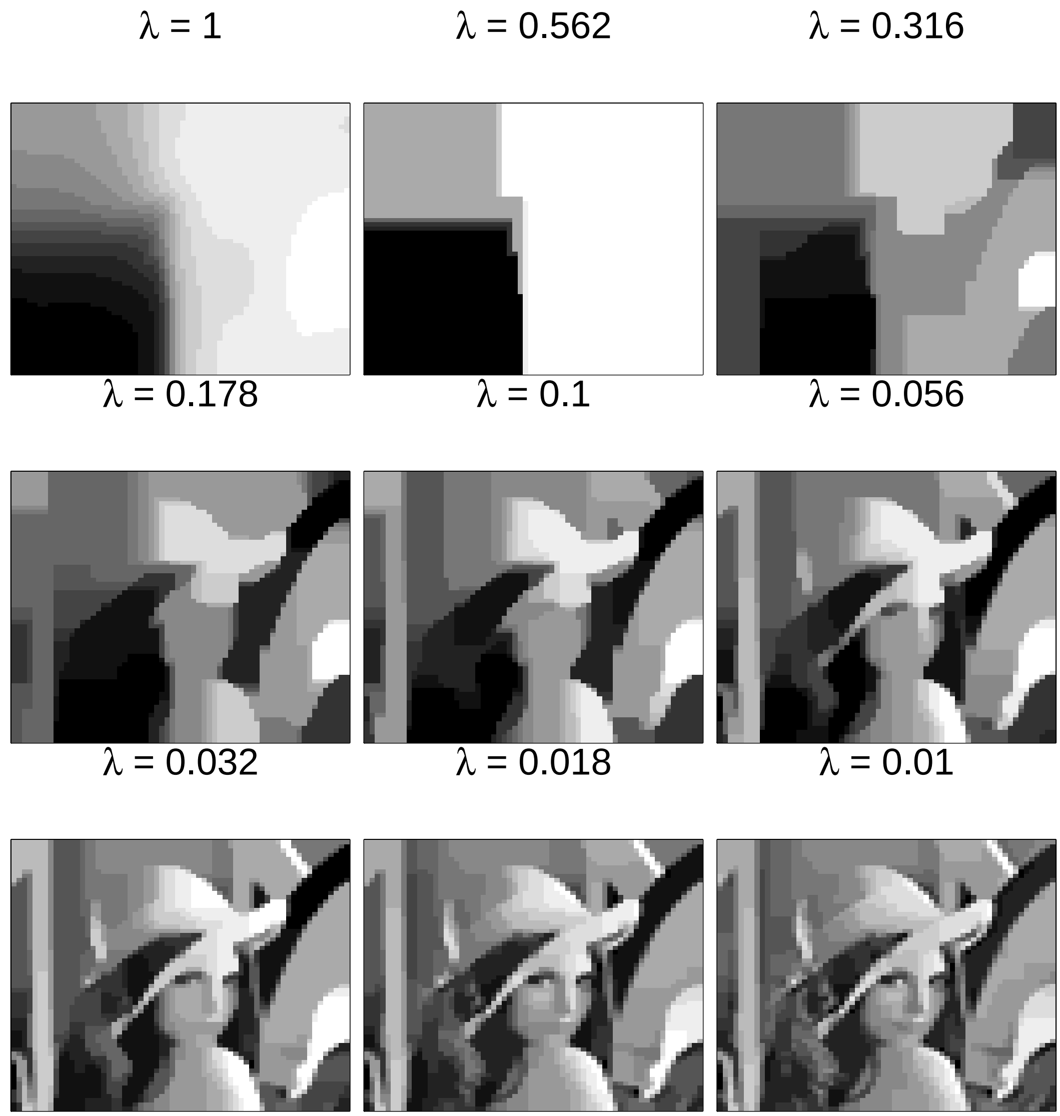
```r
  prob <- Problem(Minimize(obj))
  res <- solve(prob)
  theta_vals[[i]] <- res$getValue(theta)
  obj_val[i] <- res$value
}
print(proc.time() - pt)
```

```
##    user  system elapsed
## 229.452  68.084 311.548
```

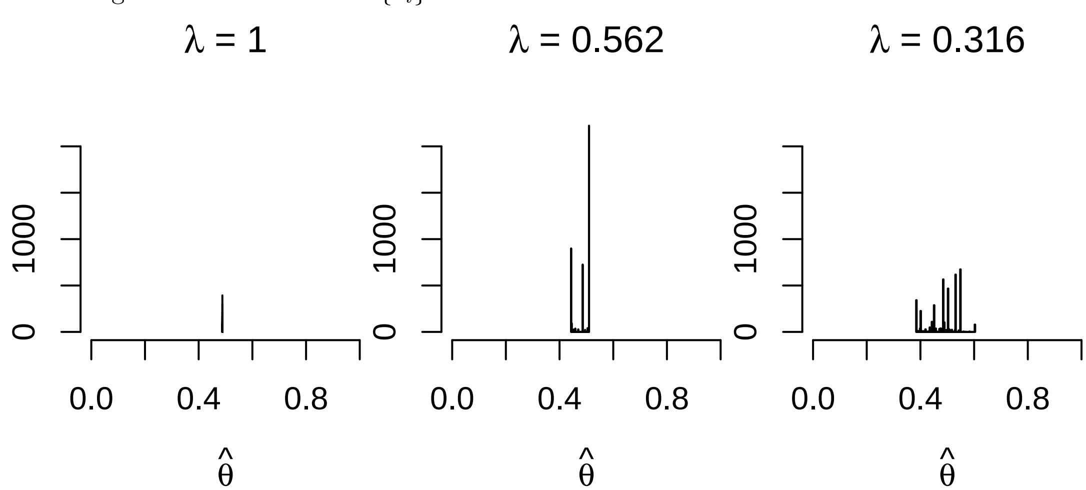
Presented below are plots of the original data and the fused fits

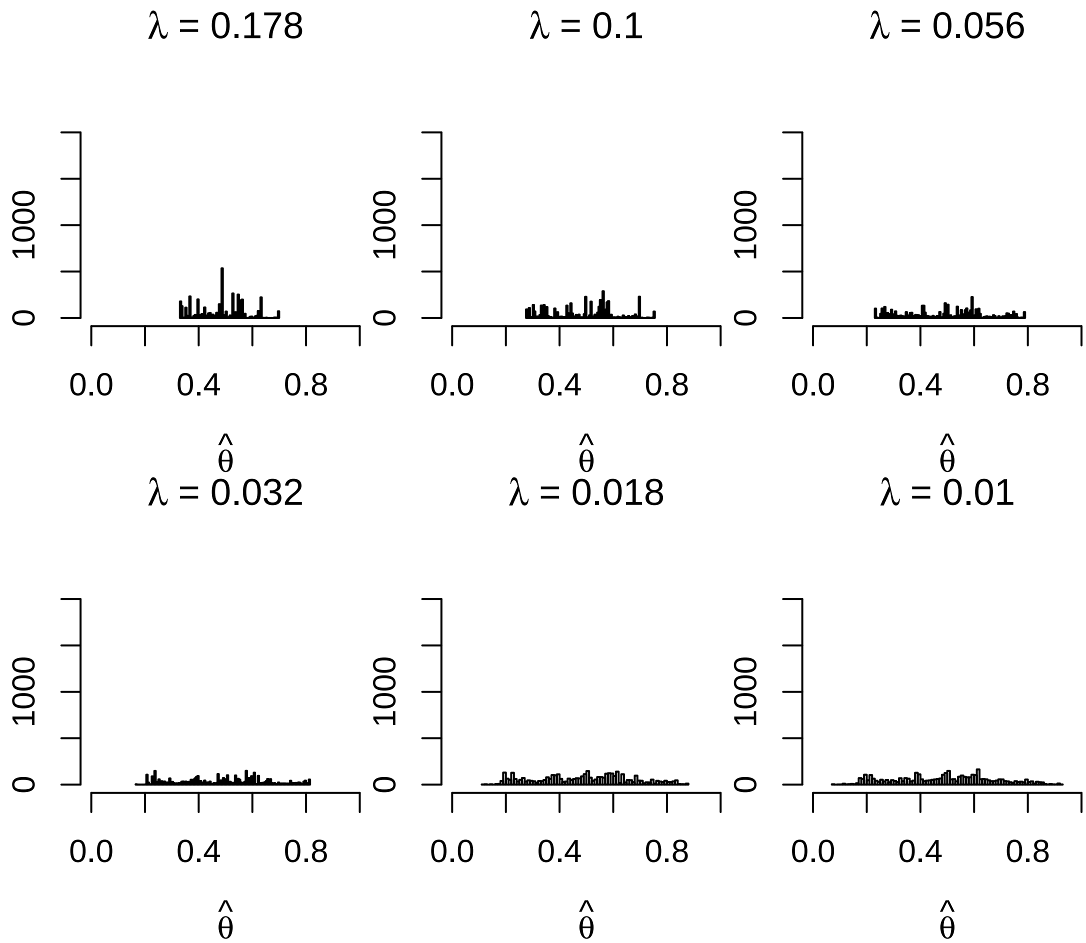





along with histograms of the fit values $\{\hat{\theta}_i\}$






Note that as $\lambda \to 0$ we find the distribution of $\hat{\theta}_i$ generally becomes less and less kurt. That is, when $\lambda$ is large we see nearly all the fits to be near the mean $\bar{y}$, and becoming more and more dispersed (towards $\hat{\theta}_i \to y_i$) as $\lambda$ shrinks.

## 4.b

### 4.b.1

Note that our expression

$$\|(x, y)\|_1^3 \le 5x + 7$$

is successfully recognized as convex without any serious manipulations. That is, in DCP we write this as

$$\|(x, y)\|_1^3 \le 5x + 7 \quad \mapsto \quad \texttt{pow(norm1(x, y), 3) <= 5 * x + 7},$$

as desired.

### 4.b.2

We now consider the expression

$$\frac{2}{x} + \frac{9}{z - y} \le 3.$$

Since DCP automatically constraining the argument to be within the function's domain handling the domain, DCP does not allow division $\frac{a}{b}$ to be input as `a/b`, as we may expect. Instead, DCP accepts `inv_pos(x)` as $\frac{1}{x}$ and restricts $x$ to $x > 0$ to enforce convexity. Therefore, we replace any instance of $\frac{1}{x}$ with `inv_pos(x)` to yield

$$\frac{2}{x} + \frac{9}{z-y} \leq 3 \quad \mapsto \quad \texttt{2 * inv\_pos(x) + 9 * inv\_pos(z - y) <= 3}$$

as desired.

**4.b.3**

We have the expression

$$\sqrt{x^2+4} + 2y \leq -5x.$$

The RHS is valid as-is, while the LHS requires some manipulation in order for the problem to be recognized as convex since $\sqrt{\cdot}$ is treated as concave, independent of its arguments. To get around this we make use of the $\ell_2$ norm (which is classified as convex)

$$\|(x_1, x_2)\|_2 = \sqrt{x_1^2 + x_2^2}.$$

Hence,

$$\|(x, 2)\|_2 = \sqrt{x^2+4}.$$

This gives us the following DCP expressions

$$\sqrt{x^2+4} + 2y \leq -5x \quad \mapsto \quad \texttt{norm2(x, 2) + 2 * y <= -5 * x}$$

as desired.

**4.b.3**

We begin with the problem

$$(x+3) \cdot z \cdot (y-5) \geq 8, \quad x \geq -3,\, z \geq 0,\, y \geq 5.$$

To translate into DCP we first note that this is equivalent to the problem

$$x \cdot z \cdot y \geq 8, \quad x \geq 0,\, z \geq 0,\, y \geq 0.$$

To enforce the (new) domain $x, y, z \geq 0$ we apply the geometric mean

$$\texttt{geo\_mean(x1, ..., xk)} = (x_1 \cdot \cdots \cdot x_k)^{\frac{1}{k}}$$

since DCP automatically restricts each argument as $x_i \geq 0$. We now wish to remove the $k^{\text{th}}$-root term. However, if we were to raise our geometric mean expression to the $k^{\text{th}}$ power then DCP would no longer treat the LHS as a concave expression since $x^k$, $k > 1$, is always considered to be convex (independent of the form $x$ takes). A solution is to instead take the $k^{\text{th}}$ root of the RHS. This gives us the DCP expression

$$(x+3) \cdot z \cdot (y-5) \geq 8, \quad x \geq -3,\, z \geq 0,\, y \geq 5 \quad \mapsto \quad \texttt{geo\_mean(x, y, z) >= 2},$$

as desired.

### 4.b.5

Our expression

$$\frac{(x+3z)^2}{\log(y-1)} + 2y^2 \le 10$$

is translated into DCP by making use of the function $\frac{s^2}{t} \mapsto$ `quad_over_lin(s, t)`. Without other major issues we translate this expression directly into DCP via

$$\frac{(x+3z)^2}{\log(y-1)} + 2y^2 \le 10 \quad \mapsto \quad \texttt{quad\_over\_lin(x + 3 * z, log(y - 1)) + 2 * square(y) <= 10},$$

as desired.

### 4.b.6

We wish to translate the following into DCP interpretable format

$$\log\left(e^{-\sqrt{x}} + e^{2z}\right) \le -e^{5y}.$$

DCP has a unique function designed to handle logarithms of sums of exponential terms. In paritcular,

$$\log\left(e^{x_1} + \cdots + e^{x_k}\right) \quad \mapsto \quad \texttt{log\_sum\_exp(x1, ..., xk)}.$$

Using this scheme we translate our expression into DCP as

$$\log\left(e^{-\sqrt{x}} + e^{2z}\right) \le -e^{5y} \quad \mapsto \quad \texttt{log\_sum\_exp(-sqrt(x), 2 * z) <= -exp(5 * y)},$$

as desired.

### 4.b.7

We begin by noting the string of equivalences of our target expression

$$\begin{aligned} \sqrt{\|(2x-3y, y+x)\|_1} = 0 &\iff \|(2x-3y, y+x)\|_1 = 0 \\ &\iff \|(2x-3y, y+x)\|_1 \le 0. \end{aligned}$$

This yields the DCP expression

$$\sqrt{\|(2x-3y, y+x)\|_1} = 0 \quad \mapsto \quad \texttt{norm1(2 * x - 3 * y, y + x) <= 0}.$$

as desired.

### 4.b.8

We wish to translate the following inequality

$$y \log\left(\frac{y}{2x}\right) \leq y + x, \quad x > 0,\, y > 0.$$

DCP handles the LHS via the Kullback–Leibler function $y \log \frac{y}{x} - y + x \mapsto$ `kl_div(y, x)`. This function automatically handles the domain restriction of $x, y > 0$. Thus,

$$y \log\left(\frac{y}{2x}\right) \leq y + x, \quad x > 0,\, y > 0 \quad \mapsto \quad \texttt{kl\_div(y, 2 * x) + x - y <= y + x - 30},$$

as desired.

# Appendix

## Question 2.1 (Alternate Solution)

Below we present an alternate (and terrible) solution to proving that $f(x,y) = |xy| + a\left(x^2+y^2\right)$ is convex $\iff a \geq \frac{1}{2}$ and strictly convex $\iff a > \frac{1}{2}$.

*Proof.* Recall that a (continuous, twice differentiable) function $f(z)$, $z \in C$, is convex on $C$ if and only if its Hessian is positive semidefinite for all $z$ on the interior of $C$,

$$\nabla^2 f(z) \in \mathbb{S}^n_+,$$

and strongly convex with parameter $m > 0$ if and only if

$$\nabla^2 f(z) - m\mathbb{I}_n \in \mathbb{S}^n_+.$$

Furthermore, a matrix $M \in \mathbb{R}^{n\times n}$ is positive semidefinite if and only if all eigenvalues of $M$ are nonnegative. Since $f$ is nondifferentiable along $x = 0, y = 0$ we first apply a differentiable approximation $f_\epsilon$

$$f_\epsilon(x,y) = \sqrt{x^2y^2+\epsilon} + a\left(x^2+y^2\right) \underset{\epsilon\to 0}{\longrightarrow} |xy| + a\left(x^2+y^2\right) = f(x,y)$$

Now, $f_\epsilon$ admits gradient

$$\nabla f_\epsilon(x,y) = \left(2ax + \frac{xy^2}{\sqrt{x^2y^2+\epsilon}}, 2ay + \frac{x^2y}{\sqrt{x^2y^2+\epsilon}}\right),$$

and Hessian

$$\nabla^2 f_\epsilon(x,y) = \begin{bmatrix} -\frac{x^2y^4}{(x^2y^2+\epsilon)^{3/2}} + \frac{y^2}{\sqrt{x^2y^2+\epsilon}} + 2a & \frac{2xy}{\sqrt{x^2y^2+\epsilon}} - \frac{x^3y^3}{(x^2y^2+\epsilon)^{3/2}} \\ \frac{2xy}{\sqrt{x^2y^2+\epsilon}} - \frac{x^3y^3}{(x^2y^2+\epsilon)^{3/2}} & -\frac{y^2x^4}{(x^2y^2+\epsilon)^{3/2}} + \frac{x^2}{\sqrt{x^2y^2+\epsilon}} + 2a \end{bmatrix}.$$

We find $\nabla^2 f_\epsilon(x,y)$ to have eigenvalues[2]

$$\begin{aligned}
\lambda_{\epsilon,1} &= \frac{x^2\left(4ay^2\sqrt{x^2y^2+\epsilon}+\epsilon\right) + 4a\epsilon\sqrt{x^2y^2+\epsilon} - \sqrt{4x^6y^6 + x^4\epsilon\left(16y^4+\epsilon\right) + 14x^2y^2\epsilon^2 + y^4\epsilon^2} + y^2\epsilon}{2\left(x^2y^2+\epsilon\right)^{3/2}} \\
\lambda_{\epsilon,2} &= \frac{x^2\left(4ay^2\sqrt{x^2y^2+\epsilon}+\epsilon\right) + 4a\epsilon\sqrt{x^2y^2+\epsilon} + \sqrt{4x^6y^6 + x^4\epsilon\left(16y^4+\epsilon\right) + 14x^2y^2\epsilon^2 + y^4\epsilon^2} + y^2\epsilon}{2\left(x^2y^2+\epsilon\right)^{3/2}}.
\end{aligned}$$

Taking the limits of $\lambda_{\epsilon,1}$ and $\lambda_{\epsilon,2}$ as $\epsilon \to 0$,

$$\begin{aligned}
\lambda_1 = \lim_{\epsilon\to 0}\lambda_{\epsilon,1} &= \frac{4ax^2y^2\sqrt{x^2y^2} - 2\sqrt{x^6y^6}}{2\left(x^2y^2\right)^{3/2}} \\
&= 2a - \frac{\left(x^2y^2\right)^{3/2}}{\sqrt{x^6y^6}} \\
&= 2a - 1, \\
\lambda_2 = \lim_{\epsilon\to 0}\lambda_{\epsilon,2} &= \frac{4ax^2y^2\sqrt{x^2y^2} + 2\sqrt{x^6y^6}}{2\left(x^2y^2\right)^{3/2}} \\
&= 2a + \frac{\left(x^2y^2\right)^{3/2}}{\sqrt{x^6y^6}} \\
&= 2a + 1.
\end{aligned}$$

[2] Details left as an exercise.

In this form we see that $\nabla^2 f(x,y)$ has nonnegative eigenvalues if and only if $a \geq \frac{1}{2}$, and so $f$ is convex for $a \geq \frac{1}{2}$. To show strong convexity, we use the result that if matrix $M$ has eigenvalues $\{\lambda_i\}_{i=1}^n$ then $M - k\mathbb{I}_n$ has eigenvalues $\{\lambda_i - k\}_{i=1}^n$. Therefore, $\nabla^2 f(x,y) - m\mathbb{I}_2$ has eigenvalues

$$\lambda_{m,1} = 2a - 1 - m$$
$$\lambda_{m,2} = 2a + 1 - m.$$

To ensure $\lambda_{m,1}, \lambda_{m,2}$ are nonnegative we set $a > \frac{1}{2}$ and $m \leq a$. Therefore, $f$ is strongly convex with parameter $m$, $a \geq m > 0$, as desired. □