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 $f_+(x, y)$ 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

$$\lambda_1 - m = 2a - 1 - m\lambda_2 - m \qquad\qquad = 2a + 1 - m.$$

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

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

The gradient of $f_-(x,y)$ is computed to be

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

Therefore, the Hessian is

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

Once again, the Hessian to be positive semidefinite, we need its eigenvalues to be nonnegative $\lambda_i \geq 0$, $i = 1,2$. As above, we find the eigenvalues to be $\lambda_1 = 2a - 1$ and $\lambda_2 = 2a + 1$. Therfore, we see that the Hessian is 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 $y \neq 0$,

$$\nabla f_0(x,y) = (0, 2ay)$$

with Hessian

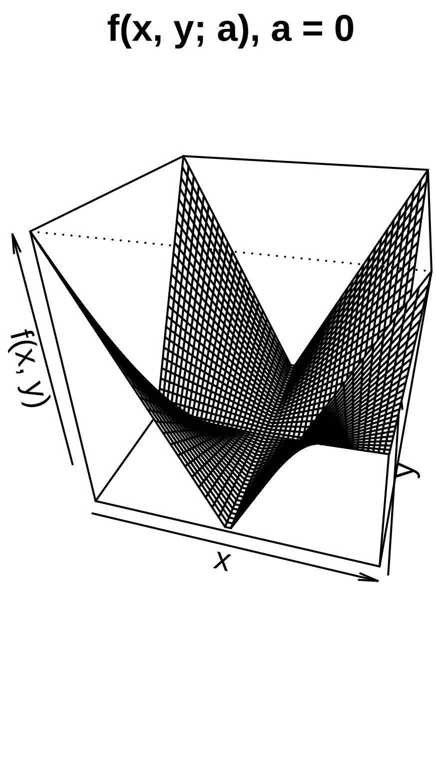
$$\nabla^2 f_0(x,y) = \begin{bmatrix} 0 & 0 \\ 0 & 2a \end{bmatrix}.$$

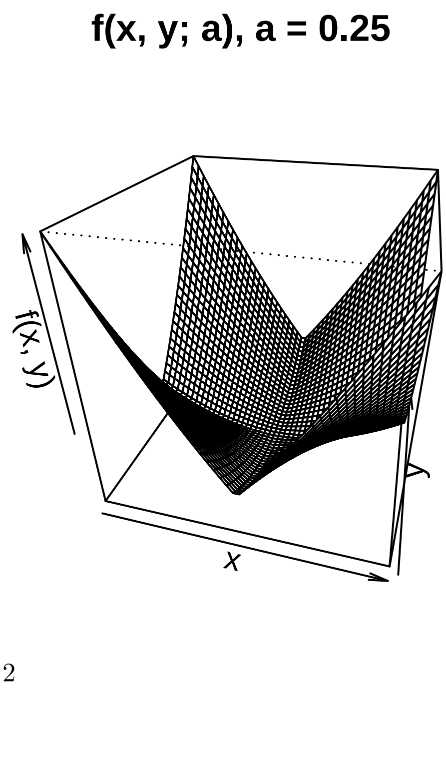
This Hessian ∇f_0^2 has only one eigenvalue $\lambda_1 = 2a$, which is nonnegative for any $a \geq 0$. It follows that $f_0(x,y)$ will also be convex for $a \geq \frac{1}{2}$ and strongly convex for $a > \frac{1}{2}$. By symmetry, the same eigenvalue is found for the case where $x \neq 0$ and $y = 0$, so we arrive to the same conclusions. Finally, for the last scenario with $x = y = 0$, we note that the domain of f_0 under $x = y = 0$ is defined at only a single point, trivially satisfying strong convexity

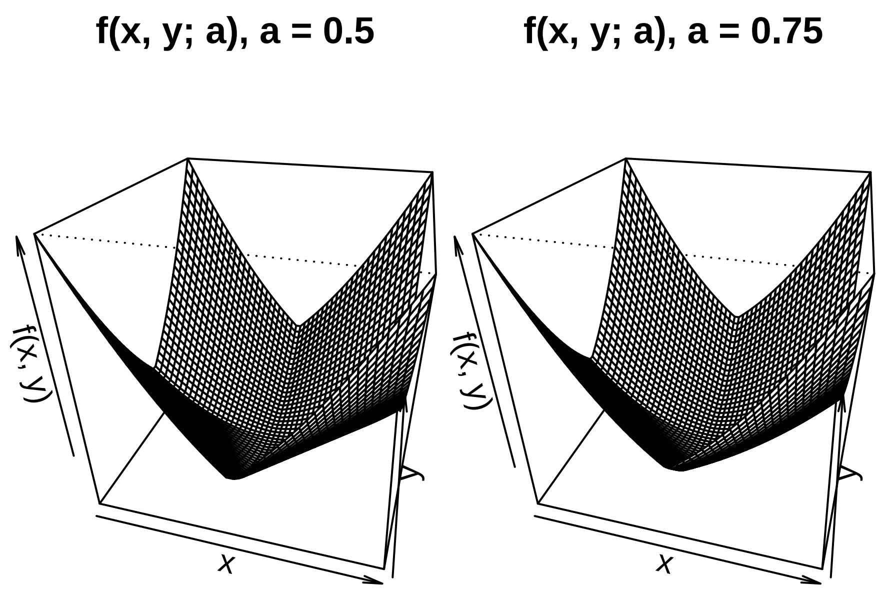
$$\left(\nabla f(z_1) - \nabla f(z_2)\right)^T (z_1 - z_2) \geq m\|z_1 - z_2\|_2^2$$

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 $\{\lambda_i\}_{i=1}^n$,

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

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

2.2 (b)

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 $t \in [0, 1]$,

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

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 $t \to 0$,

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

By symmetry we swap x and y to obtain

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

as desired.

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

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

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

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

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

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

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

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

as desired. □

Question 3

(a)

Proof. ($1 \implies 2$)

$(2 \implies 3)$

$(3 \implies 4)$

$(4 \implies 1)$

$\square$

(b)

Proof. Let f be convex and twice differentiable.

$(1 \implies 2)$ If ∇f is L-Lipschitz then, for $x, y \in \text{dom}(f)$, $L > 0$,

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2.$$

$(2 \implies 3)$

$(3 \implies 4)$

$(4 \implies 1)$

$\square$

Question 4

(a)

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

```
library(CVXR)
```

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)))
}
```

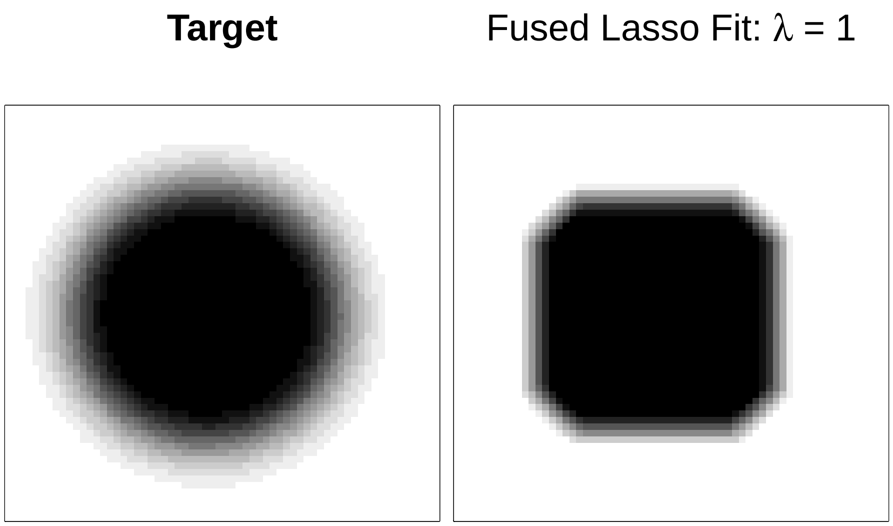
as well as defining some parameters and variables, as well as our ℓ_2 loss $\frac{1}{2}\|Y-\theta\|_2^2$

```
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 ℓ_2 term and the ℓ_1 fused term in our objective function. The ℓ_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 ℓ_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 ℓ_1 norm), inside of which most observations are close to 0 and beyond which most observations are close to 1.

2.

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

```r
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 $\{\lambda_k\}_{k=0,\ldots,8} = \left\{10^{-k/4}\right\}_{k=0,\ldots,8}$

Note that as $\lambda \to 0$ we find the distribution of $\hat{\theta}_i$ generally becomes less and less kurt. That is, when λ 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 λ shrinks.

(b)

1.

Note that our expression

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

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

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

as desired.

2.

We now consider the expression

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

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.

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 ℓ_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.

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

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

since DCP automatically restricts each argument as $x_i \geq 0$. We now wish to remove the k^{th}-root term. However, if we were to raise our geometric mean expression to the k^{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^{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.

5.

Our expression

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

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

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

as desired.

6.

We wish to translate the following into DCP interpretable format

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

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

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

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.

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.

8.

We wish to translate the following inequality

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

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,

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

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_ϵ

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

Now, f_ϵ 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[1]

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

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

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

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

[1] 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

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

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. $\square$