HW 1

January 11, 2021

1 Problem 1

Identify which of the following sets are convex, and provide a brief justification or proof for each.

1. A set of the form $\{x \in \mathcal{R}^n | \alpha \leq a^T x \leq \beta\}$.
2. A set of the form $\{x \in \mathcal{R}^n | \alpha_i \leq x_i \leq \beta_i\}$.
3. The set of points closer to a given point than a given set, i.e., $\{x | \|x - x_0\|_2 \leq \|x - y\|_2, \forall y \in S\}$, where $S \subseteq \mathcal{R}^n$.
4. The set of points closer to a set than another, $\{x | dist(x, S) \leq dist(x, T)\}$, where $S, T \subseteq \mathcal{R}^n$.
5. A polyhedron $\{x \in \mathcal{R}^n | Ax \leq b\}$ for some $A \in \mathcal{R}^{m \times n}, b \in \mathcal{R}^m$.
6. The intersection $\cap_i S_i$, where $S_i \subset \mathcal{R}^n, i \in I$ is a collection of convex sets.

1.1 Solution: Questions 1, 2, 5, 6

For 1, 2, and 5 you can just apply the convexity definition directly. Instead of doing this directly, a faster way is to first solve question 6, which can be used to prove the others directly (life doesn't bring problems in the "correct order" :)). 6 is also (partialy) used to prove 4 (but you can again follow the definition).

Consider first two convex sets $\mathcal{A}, \mathcal{B}$. Pick $x_1, x_2 \in \mathcal{A} \cap \mathcal{B}$, which implies $x_1, x_2 \in \mathcal{A}$ and $x_1, x_2 \in \mathcal{B}$, and an arbitrary convex combination $x^\star = \lambda x_1 +$

$(1-\lambda)x_2$, where $\lambda \in [0,1]$. From the convexity of $\mathcal{A}$ (resp. $\mathcal{B}$), $x^\star \in \mathcal{A}$ (resp. $x^\star \in \mathcal{B}$), hence $x^\star \in \mathcal{A} \cap \mathcal{B}$. The proof for an arbitrary number of sets follows by induction.

We now consider the sets of type

$$\mathcal{C} = \{x \in \mathcal{R}^n \mid a^T x \leq c\}, \quad (\textit{halfspace}) \tag{1}$$

which are convex since for any $x_1, x_2 \in \mathcal{C}$, we have

$$\begin{aligned} a^T(\lambda x_1 + (1-\lambda)x_2) &= \lambda a^T x_1 + (1-\lambda)x_2 \\ &\leq \lambda c + (1-\lambda)c = c \end{aligned}$$

The convexity of the sets in 1, 2, and 5, can be now proved by noticing that those sets are nothing but intersections of sets of the type Eq.(1).

Q1: It is the intersection of halfspaces $a^T x \leq \beta$ and $-a^T x \leq -\alpha$.

Q2: These are "box" or "rectangle" constraints. $x_i \leq \beta_i$ is a halfspace with vector $a : a_i = 1$, and all other elements zero. Hence, $\alpha_i \leq x_i \leq \beta_i$ is the convex set of Q1, and the set in question is an intersection of n such sets (hence convex).

Q3: This is trivially an intersection of halfspaces.

1.2 Solution: Questions 4,5

Consider first a *specific* vector $y \in \mathcal{S}$. The set $\{x | \|x - x_0\| \leq \|x - y\|\}$ is equivalently described by

$$\begin{aligned} &\|x - x_0\| \leq \|x - y\| \iff \\ &\|x\|^2 - 2x_0^T x + \|x_0\|^2 \leq \|x\|^2 - 2y^T x + \|y\|^2 \iff \\ &2(y - x_0)^T x \leq \|y\|^2 - \|x_0\|^2, \end{aligned}$$

which is a halfspace ($2(y - x_0)$ is the supporting vector a, and scalar c is $\|y\|^2 - \|x_0\|^2$). Hence, the desired set $\{x | \|x - x_0\| \leq \|x - y\|, y \in \mathcal{S}\}$ (i.e. considering all points y in set $\mathcal{S}$ is an intersection of halfspaces.

On the other hand, the set $\{x | \mathrm{dist}(x, \mathcal{S}) \leq \mathrm{dist}(x, \mathcal{T})\}$ is not convex. For example, consider $\mathcal{S} = (-\infty, -2] \cup [2, \infty)$, and $\mathcal{T} = \{0\}$. It is easy to see that $\{x | \mathrm{dist}(x, \mathcal{S}) \leq \mathrm{dist}(x, \mathcal{T})\} = (-\infty, -1] \cup [1, \infty)$, which is clearly not convex.

2 Problem 2

Specify whether the function is strongly convex, strictly convex, convex, or non-convex, and give a brief justification for each.

1. $f(x) = x^4$.
2. $f(x_i) = \sum_i x_i log(x_i)$, for $x > 0$.
3. $f(x) = log(1 + e^x)$.
4. $f(x, y) = |x| + |y| + 2x - 2$.
5. $f(x, y) = x^2 - xy + 2y^2 + 3$.
6. $f(x, y) = 2xy + 3x^2 + y^2$.
7. $f(x, y) = e^{(x^2 - xy + y^2 + 2)} + \frac{1}{\sqrt{x} + 2y}$, for $x, y > 0$.
8. $f(x) = \frac{\|Ax - b\|^2}{1 - x^T x}$, on set $\{x : \|x\|_2 < 1\}$, where $\{x \in \mathcal{R}^n$.

2.1 Solution

Q1: $f(x) = x^4$ is decievingly tricky, despite it being simple-looking. Second derivative $12x^2 = 0$ for $x = 0$. Thus the function cannot be strongly convex. It is always non-negative though, so it is convex. The remaining question is if it is strictly convex?

In general, a function with positive definite Hessian everywhere (i.e., all eigenvalues > 0 inside the domain of the function) *are* strictly convex. Hence it is a *sufficient* condition. But note that it is *not necessary.* In fact, if the points where the function's derivative (or Hessian's determinant) goes to 0 has "finite support" or "finite measure" (i.e. are isolated points, even if infinitely many) while it is strictly positive everywhere else, then the function is strictly convex. Since the only problematic point is $x = 0$ this implies that x^4 is indeed strictly convex.

A more algebraic way to prove this, following the definition is the following: We use the definition and the fact that x^2 is strictly convex (in fact strongly convex).

We ignore the outer square and for the inner square, we use x^2 strict convexity and then we square both handsides and the $<$ does not change.

$$((1-t)x+ty)^2 < (1-t)x^2+ty^2 \Rightarrow \left[((1-t)x+ty)^2\right]^2 < \left[(1-t)x^2+ty^2\right]^2 \tag{2}$$

We set $x^2 = x', y^2 = y'$ and the rhs of the last inequality becomes

$$\left[(1-t)x'+ty'\right]^2 < \left[(1-t)x'^2+ty'^2\right] = (1-t)x^4+ty^4 \tag{3}$$

which completes the statement.

A third way would be to prove strict convexity at point $x = 0$ only (it is strictly convex for all other points), by applying the definition for $x = 0$ and any other y.

Q2: Our function is $f(x) = \sum_{i=1}^{K} x_i \log(x_i)$ with $x_i > 0$. It is enough to see what happens on one term. For this we find the second derivative which is $\frac{1}{x} > 0$ for $x > 0$, so it is convex in the strict sense definitely. However, with x unbounded the limit of the second derivative 0, so m-strong convexity does not hold for any value $m > 0$.

Q3: Our function is $f(x) = \log(1+e^x)$. For this we can compute $\frac{d^2 f}{dx^2} = \frac{e^x}{(e^x+1)^2}$, which is always > 0, hence it is strictly convex. However for very large x we have $\log(1+e^x) \approx \log(e^x) = x$ which is a straight line, thus not strongly convex. It is also evident by taking the limit of the derivative for $x \to \infty$, which goes to 0.

Q4: The first term $|x|+|y|$ is the L1 norm which is convex (not strictly) but not smooth. The second term $2x-2$ is affine, so the entire function is just convex.

Q5: Our function is $f(x,y) = x^2+2y^2-xy+3$. The Hessian is

$$H_f = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$$

If λ_1, λ_2 are the two eigenvalues, it then holds that:

$$\lambda_1 + \lambda_2 = 6 \text{ the "trace" of H}$$
$$\lambda_1 \cdot \lambda_2 = 7 \text{ the determinant of H}$$

It is easy to see that both eigenvalues have to be strictly positive, hence f is strongly convex.

Q6: Our function is $f(x, y) = 3x^2 + y^2 + 2xy$. The Hessian is

$$H_f = \begin{bmatrix} 6 & 2 \\ 2 & 2 \end{bmatrix}$$

Similar to the previous question, you can see that this should be strongly convex.

Q7:

The first term in the function is convex: it is a composition of a convex, non-decreasing function (exponential), with a quadratic function that is convex (check the eigenvalues of the quadratic in the exponent).

The second term is also convex. It's a composition of a non-increasing convex function $(1/x)$ with a concave function (the denominator is square root - which is concave - plus a linear term).

We also know that a (positive) sum of convex functions is also convex, which proves the convexity of the entire function. Notice the simplicity of this solution compared to calculating the Hessian, for example, for the entire monster.

Q8: A first easy way to solve it is through the epigraph. Namely, consider the convex sets defined by

$$\frac{\|Ax - b\|^2}{1 - x^T x} \leq t,$$

for some $t \in \mathcal{R}$.

This becomes

$$\begin{aligned} \frac{\|Ax - b\|^2}{t} &\leq 1 - x^T x \\ \Rightarrow \frac{\|Ax - b\|^2}{t} &+ x^T x - 1 \leq 0, \end{aligned}$$

which is a convex inequality: the L2 norm (first term) is convex, and it's composition with an affine function $Ax - b$ retains convexity. The second term is also trivially convex. So this inequality is convex for any t, hence the epigraph of $f(x)$ always defines a convex set, and hence the function is convex.

A second way to prove this is through composition:

The function $f(x,u) = \frac{\|Ax-b\|^2}{u}$ is convex in (x,u), since it is the quadratic over linear function (convex, see Boyd), pre-composed with an affine mapping. This function is decreasing in its second argument, so by composition rules, we can replace the second argument with any convcave function, and the result is convex. But $u = 1 - x^T x$ is concave, so we're done.

Regarding stronger statements, one can check that $\frac{y^2}{x}$ has zero determinant, so at least one eigenvalue is always 0, hence it is neither strictly nor strongly convex.

3 Problem 3

Problem 3

(a) Show that the following statements are equivalent.

- $\nabla f(x)$ is Lipschitz with constant L.
- $\nabla^2 f(x) \preceq LI$, for all x.
- $f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|_2^2$, for all x, y.

(b) Assume a strongly convex function $f(x)$ ($x \in \mathcal{R}^n$) and let $x(k+1) = x(k) - t\nabla f(x(k))$ (i.e. the standard gradient descent step). Let further $g(x) = x - t\nabla f(x)$ be such that $\|g(y) - g(x)\|_2 \leq \beta\|y-x\|_2$, for some $\beta < 1$, and any x, y. Let finally x^* denote the global minimum of $f(x)$. Show that

$$\|x(k) - x^*\|_2 \leq \beta^k \|x(0) - x^*\|_2$$

Hint: the goal is to start with $\|x(k) - x^*\|_2$, the distance from the optimal of the current point $x(k)$, and step-by-step compare it with the distance from the optimal at earlier steps $\|x(k-1) - x^*\|_2$, $\|x(k-2) - x^*\|_2$, etc.

3.1 Solution (a)

(1 to 2)

We can go from 1 to 2 as follows:

Being L-smooth (Lipschitz) means: for any x, y $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$.

Choose $y = x + dx$. Then

$$\begin{aligned}
\|\nabla f(x+dx) - \nabla f(x)\| &\leq L\|x + dx - x\| \\
\Rightarrow \|\nabla f(x+dx) - \nabla f(x)\| &\leq L\|dx\| \\
\Rightarrow \frac{\|\nabla f(x+dx) - \nabla f(x)\|}{\|dx\|} &\leq L \\
\Rightarrow \|\frac{\nabla f(x+dx) - \nabla f(x)}{dx}\| &\leq LI
\end{aligned},$$

where the last step follows from Cauchy-Schwartz inequality. Taking the limit as $dx \to 0$ makes the term inside the norm equal to the Hessian (by definition) and interpreting the last norm as some matrix norm (e.g. trace norm), since the quantity inside is a matrix, gives 2.

(2 to 3)

You can go from 2 to 3 in many ways: From the integral residual of the Taylor series for example, and then bounding it using Eq. 2.

You can also start from the fact that, for any convex function, it holds that

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

for some z between y and x (this is given in Boyd, Chapter 9, when he talks about strong convexity). Hence, you can apply 2 directly above to bound $\nabla^2 f(z)$ and get 3.

However, a more analytical way of doing it is the following:

Consider the function $g(x) = f(x) - \frac{L}{2}\|x\|^2$ with gradient and Hessian

$$\begin{aligned}
\nabla g(x) &= \nabla f(x) - Lx, \\
\nabla^2 g(x) &= \nabla^2 f(x) - LI.
\end{aligned}$$

The condition $\nabla^2 f(x) \preceq LI$ implies $\nabla^2 g(x) \preceq 0$, hence $g(x)$ is concave. We then have, by definition of concavity, $\forall x, y$

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

which implies $f(y) \leq f(x) + \nabla^T f(x)(y - x) + \frac{L}{2}\|x - y\|^2$. The converse readily follows by identical arguments.

(3 to 1)

For the second part, we consider the expression

$$f(x) \leq f(y) + \nabla^T f(y)(x - y) + \frac{L}{2}\|x - y\|^2, \ \forall x, y$$

obtained by simply switching the role of x and y. We then have

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

which implies

$$\begin{aligned}
-L\|x - y\|^2 &\leq (\nabla^T f(y) - \nabla^T f(x))(x - y) \\
&\leq \|\nabla f(y) - \nabla f(x)\|\|x - y\|,
\end{aligned}$$

where the last step is the CauchySchwarz inequality. Hence, we finally have $\|\nabla f(y) - \nabla f(x)\| \geq -L\|x - y\|$.

3.2 Solution (b)

By recalling that for convex functions $\nabla f(x^\star) = 0$ holds, we have

$$\begin{aligned}
\|x(k) - x^\star\| &= \|g(x(k-1)) - x^\star\| \\
&= \|g(x(k-1)) - x^\star + t\nabla f(x^\star)\| \\
&= \|g(x(k-1)) - g(x^\star)\| \\
&\leq \beta \|x(k-1) - x^\star\| \\
&\leq \beta(\beta\|x(k-2) - x^\star\|) \\
&\leq \ldots \leq \beta^k \|x(k-2) - x^\star\|.
\end{aligned}$$

4 problem 4

For each of these functions, answer the following questions:

1. $f(x, y) = xlog(x) + ylog(y)$, for $x, y > 1$.
2. $f(x, y) = |x| + |y| + 2x - 2y + 3$.
3. $f(x, y) = x^2 - xy + 2y^2$.

(a): which of the methods we learned (gradient descent, subgradient method, proximal gradient method) are applicable for each function.

(b): for the applicable methods, what is the convergence rate when an approximation error of ϵ is needed: $O(1/\epsilon)?O(1/\epsilon^2)?O(log(1/\epsilon))?$

(c): pick the fastest method (among the three) for each function, and show the basic update step (i.e., the gradient step, the subgradient step, or the proximal gradient step, depending on the method). Explain also how to pick the step sizes to ensure convergence.

Solution (a): Our function is $f(x, y) = x\log(x) + y\log(y)$. Referring back to Problem 2.2, we know this function is differentiable and strictly convex. It is Lipschitz with parameter L =1 for $x, y > 1$. We can thus apply gradient-descent with step size $t \leq 1$, or simply do backtracking, and it will ensure convergence.

For finite values of x and y it is also strongly convex, but one problem is that the second derivatives go to 0 as $x \to \infty$ or $y \to \infty$, which would imply no strong convexity in general, and a $O(1/\epsilon)$ convergence rate.

Yet, observe that if we start with any finite point x_0, y_0, the determinant will be strongly convex (assume so with parameter m_0). Also, picking the step as above (e.g., with backtracking) guarantees descent at every step. This means that for the entire search space of the gradient descent algorithm the function is at least m_0 strongly convex, and in fact convergence is "linear", i.e., $O(log(1/\epsilon))$.

One last point is that this is a constrained problem, since $x, y > 1$. However, these are simply "box" constraint, so after each gradient step we can project back to feasibility: i.e. if $x(k+1) < 1$, after the gradient step, set $x(k+1) = 1+\epsilon$ (same for $y(k+1)$). Projected gradient is, as we said, a proximal algorithm, where the non-smooth function $g()$ is the indicator function for a constraint. It also inherits the convergence properties of the smooth part of the objective (namely $x \log(x)$), so projected gradient descent should still converge as $O(log(1/\epsilon))$.

Solution (b): Our function in that case is $f(x, y) = 2(x - y) + 3 + (|x| + |y|)$. This is not differentiable, so one would go for subgradients, or for proximal algorithm (since the second term is the L1 norm, which we now just introduces soft thresholding.

However, observe that this function is unbouunded below. It can be reduced to $-\infty$ by increasing say y to ∞ (and keeping x constant).

It is important to remember that, gradient descent methods, while theoretically applicable in this case, make no sense. They would "never" converge to the optimal value (in finite steps). Remember that the convergence rate is $1/\epsilon$ but this divides the distance of the initial function value from the optimal (which in this case it's infinite).

Solution (c): Our function in that case is $f(x, y) = x^2 - xy + 2y^2$. This is the function of Problem 2.5, without the constant term 3.

The function is strongly convex and Lipschitz (finite and positive eigenvalues). So with a good step size we or backtracking one can achieve go $O(\log(1/\epsilon))$. Each step is $x^+ = x - t\nabla f_x$. To find the stepsize as you told me you need to be somewhere in the middle between the max and min eigenvalues of the hessian. These are 3 and 1.