MATH 582G Homework 5

Cade Ballew #2120804

February 24, 2023

1 Problem 1

Consider the sets

$$\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n : x_i > 0, i = 1, \dots, n\},$$
$$\mathcal{S}^n_{++} = \{X \in \mathcal{S}^n : X \succ 0\},$$

and functions $f : \mathbb{R}^n_{++} \to \mathbb{R}$, $F : \mathcal{S}^n_{++} \to \mathbb{R}$ with

$$f(x) = -\sum_{i=1}^{n} \log x_i, \quad F(X) = -\log\det X.$$

1.1 Part 1

We have that

$$\frac{\partial f(x)}{\partial x_i} = -\frac{1}{x_i},$$

so

$$\nabla f(x) = \begin{pmatrix} -1/x_1 \\ \vdots \\ -1/x_n \end{pmatrix}.$$

Then,

$$\nabla^2 f(x) = \begin{pmatrix} 1/x_1^2 & & \\ & \ddots & \\ & & 1/x_n^2 \end{pmatrix}.$$

1.2 Part 2

We have that

$$\begin{aligned}F(X+tV)-F(X) &= -\log\det(X+tV)+\log\det(X)\\ &= -\log\left(\frac{\det(X+tV)}{\det(X^{1/2})\det(X^{1/2})}\right)\\ &= -\log\left(\det(X^{-1/2})\det(X+tV)\det(X^{-1/2})\right)\\ &= -\log\det\left(X^{-1/2}(X+tV)X^{-1/2}\right)\\ &= -\log\det\left(I+tX^{-1/2}VX^{-1/2}\right).\end{aligned}$$

Also, the cyclic invariance of the trace gives

$$t\langle X^{-1},V\rangle = t\operatorname{Tr}(X^{-1}V) = t\operatorname{Tr}(X^{-1/2}X^{-1/2}V) = t\operatorname{Tr}(X^{-1/2}VX^{-1/2}).$$

Thus,

$$F(X+tV)-F(X)+t\langle X^{-1},V\rangle = -\log\det\left(I+tX^{-1/2}VX^{-1/2}\right)+t\operatorname{Tr}(X^{-1/2}VX^{-1/2}).$$

Letting μ_j denote the eigenvalues of $I+tX^{-1/2}VX^{-1/2}$,

$$\begin{aligned}&-\log\det\left(I+tX^{-1/2}VX^{-1/2}\right)+t\operatorname{Tr}(X^{-1/2}VX^{-1/2})\\ &= -\log\prod_i(1+t\mu_i)+t\sum_i\mu_i = \sum_i(-\log(1+t\mu_i)+t\mu_i)\\ &= \sum_i(-(\log(1)+t\mu_i+\mathrm{o}(t))+t\mu_i) = \mathrm{o}(t),\end{aligned}$$

where we have Taylor expanded the logarithm around 1. Thus,

$$\lim_{t\to 0}\frac{F(X+tV)-F(X)}{t} = -\langle X^{-1},V\rangle+\mathrm{o}(1),$$

so $\nabla F(x) = -X^{-1}$ by the definition of the gradient.

To compute the Hessian, we first observe that

$$(X+V)^{-1} = X^{-1/2}\left(I+X^{-1/2}VX^{-1/2}\right)^{-1}X^{-1/2}.$$

Applying a Neumann series expansion,

$$\begin{aligned}(X+V)^{-1} &= X^{-1/2}\left(I-X^{-1/2}VX^{-1/2}+\mathrm{O}(\|X^{-1/2}VX^{-1/2}\|_{\mathrm{op}}^2)\right)^{-1}X^{-1/2}\\ &= X^{-1}-X^{-1}VX^{-1},\end{aligned}$$

if we drop higher order terms. Then

$$\begin{aligned}\lim_{t\to 0}\frac{\nabla F(X+tV)-\nabla F(x)}{t} &= \frac{-(X+tV)^{-1}+X^{-1}}{t}\\ &= \frac{-X^{-1}+tX^{-1}VX^{-1}+X^{-1}}{t} = X^{-1}VX^{-1}.\end{aligned}$$

Thus, $\nabla^2F(X)[V] = X^{-1}VX^{-1}$.

1.3 Part 3

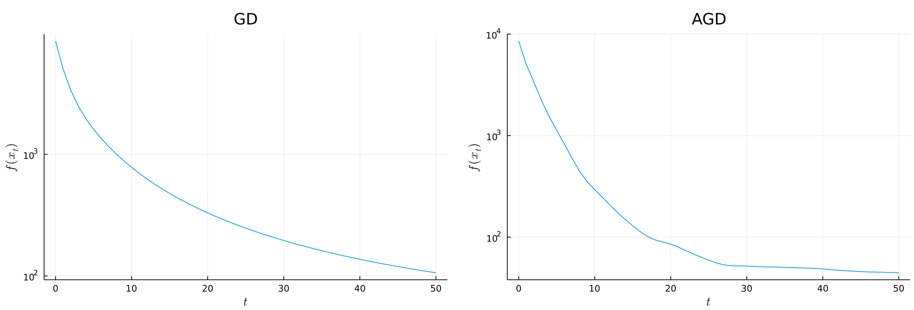
By the cylic invariance of the trace,

$$\begin{aligned}
\langle \nabla^2 F(X)[V], V \rangle &= \mathrm{Tr}(X^{-1}VX^{-1}V) = \mathrm{Tr}(X^{-1/2}X^{-1/2}VX^{-1/2}X^{-1/2}V) \\
&= \mathrm{Tr}\left((X^{-1/2}VX^{-1/2})(X^{-1/2}VX^{-1/2})\right) \\
&= \|X^{-1/2}VX^{-1/2}\|_F^2,
\end{aligned}$$

which works if $X \succ 0$ (allowing us to take the square root) and $V \in \mathcal{S}^n$. Since this is a norm, we must have that $\langle \nabla^2 F(X)[V], V \rangle > 0$ when $V \neq 0$, so $\nabla^2 F(X)$ is positive definite, and F is convex.

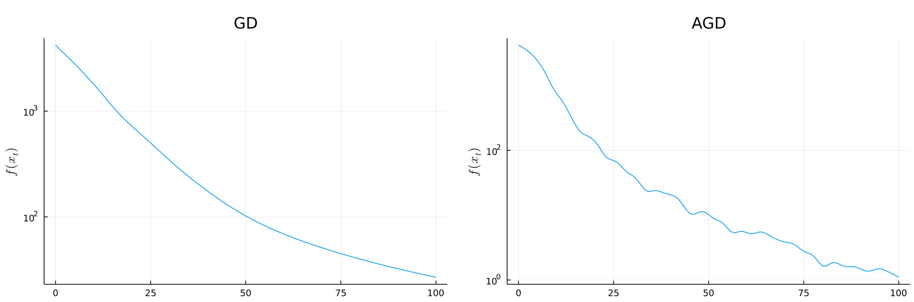
2 Problem 2

We observe the following plots from applying GD and AGD to this function with an initial guess drawn from $N(0, I)$.



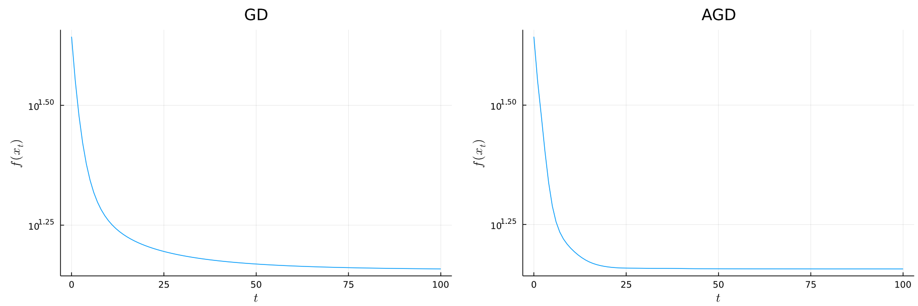
3 Problem 3

We observe the following plots from applying GD and AGD to this function with an initial guess drawn from $N(0, I)$, $\eta = 5$, and a stepsize of $2m\eta + \lambda$.



4 Problem 4

We observe the following plots from applying GD and AGD to this function.



5 Problem 5

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable convex function and let x^* be a minimizer. Consider the gradient descent iterates

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k),$$

for some sequence $\gamma_k \geq 0$.

5.1 Part 1

We have that

$$\frac{1}{2}\|x_{k+1} - x^*\|^2 = \frac{1}{2}\|x_k - x^*\|^2 + \langle x_k - x^*, x_{k+1} - x_k \rangle + \frac{1}{2}\|x_{k+1} - x_k\|^2.$$

Clearly,

$$\|x_{k+1} - x_k\|^2 = \gamma_k^2 \|\nabla f(x_k)\|^2.$$

By convexity, we have that

$$\langle x_k - x^*, x_{k+1} - x_k \rangle = -\gamma_k \langle x_k - x^*, \nabla f(x_k) \rangle \leq -\gamma_k (f(x_k) - f(x^*)),$$

so

$$\frac{1}{2}\|x_{k+1} - x^*\|^2 \leq \frac{1}{2}\|x_k - x^*\|^2 - \gamma_k (f(x_k) - f(x^*)) + \frac{\gamma_k^2}{2}\|\nabla f(x_k)\|^2.$$

5.2 Part 2

Suppose that the minimum of f is given by f^*. Then, the RHS of the inequality from part 1 can be minimized by differentiating with respect to γ_k and setting it equal to zero since it is convex in γ_k. This gives that

$$\gamma_k = \frac{f(x_k) - f^*}{\|\nabla f(x_k)\|^2}.$$

Plugging this in to the inequality and scaling by 2, we have that

$$\begin{aligned}\|x_{k+1}-x^*\|^2 &\le \|x_k-x^*\|^2 - 2\frac{f(x_k)-f^*}{\|\nabla f(x_k)\|^2}(f(x_k)-f^*) + \left(\frac{f(x_k)-f^*}{\|\nabla f(x_k)\|^2}\right)^2 \|\nabla f(x_k)\|^2 \\ &\le \|x_k-x^*\|^2 - \left(\frac{f(x_k)-f^*}{\|\nabla f(x_k)\|}\right)^2.\end{aligned}$$

5.3 Part 3

Also assume that f is β-smooth. We can rearrange and write the telescoping sum

$$\begin{aligned}\sum_{i=0}^{k-1}\left(\frac{f(x_i)-f^*}{\|\nabla f(x_i)\|}\right)^2 &\le \sum_{i=0}^{k-1}\left(\|x_i-x^*\|^2 - \|x_{i+1}-x^*\|^2\right) \\ &= \|x_0-x^*\|^2 - \|x_k-x^*\|^2 \le \|x_0-x^*\|^2.\end{aligned}$$

Letting $j \in \{0,\dots,k-1\}$ be the index for which $\|\nabla f(x_k)\|$ is maximized,

$$\begin{aligned}&\frac{1}{\beta\|x_0-x^*\|^2}\sum_{i=0}^{k-1}(f(x_i)-f^*)^2 \le \frac{1}{\beta\|x_j-x^*\|^2}\sum_{i=0}^{k-1}(f(x_i)-f^*)^2 \\ &\le \frac{1}{\|\nabla f(x_j)\|^2}\sum_{i=0}^{k-1}(f(x_i)-f^*)^2 \le \sum_{i=0}^{k-1}\left(\frac{f(x_i)-f^*}{\|\nabla f(x_i)\|}\right)^2,\end{aligned}$$

by the definition of β-smoothness and the fact that $\nabla f(x^*) = 0$. By Cauchy-Schwartz, we have that

$$\left(\sum_{i=0}^{k-1}(f(x_i)-f^*)\right)^2 \le k\sum_{i=0}^{k-1}(f(x_i)-f^*)^2,$$

so combining these,

$$\left(\sum_{i=0}^{k-1}(f(x_i)-f^*)\right)^2 \le \beta^2 k\|x_0-x^*\|^4.$$

Now, we just take the square root, divide by k, and apply Jensen's inequality on the LHS to get that

$$f\left(\frac{1}{k}\sum_{i=0}^{k-1}x_i\right) - f^* \le \frac{\beta\|x_0-x^*\|^2}{\sqrt{k}}.$$

If we additionally have that f is α-strongly convex, we begin with our original inequality and apply α-convexity in the numerator and β-smoothness in the denominator to get that

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - \left(\frac{\alpha\|x_k-x^*\|^2/2}{\beta\|x_k-x^*\|}\right)^2 = \left(1-\frac{\alpha^2}{4\beta^2}\right)\|x_k-x^*\|^2.$$