. . .

MATH238 HW3

Jacob Perricone
Stanford University
jacobp2@stanford.edu

Problem 1

Recall the (local) second-order (SO) and scaled second-order (SSO) Lipschitz conditions (LC):

$$\text{SOLC} : ||\nabla f(\mathbf{x}+\mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d}|| \leq \beta ||\mathbf{d}||^2, \text{where } ||\mathbf{d}|| \leq .5$$

and

$$\text{SSOLC} : ||X(\nabla f(\mathbf{x}+\mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d})|| \leq \beta |\mathbf{d}^2 \nabla^2 f(\mathbf{x}) d|, \text{where } ||X^{-1}\mathbf{d}|| \leq .5$$

Find parameter β values (or upper bounds) of (SOLC) and (SSOLC) for each of the following scalar functions:

(a) $f(x) = \frac{1}{3}x^3 + x$, $x > 0$ We have

$$\nabla f(x) = x^2 + 1$$
$$\nabla^2 f(x) = 2x$$

Then the SOLC condition is

$$||(x+d)^2 - x^2 - 2xd|| \leq \beta ||d||^2$$
$$||d^2|| \leq \beta ||d||^2$$
$$1 \leq \beta$$

Noticing that in the scalar case, the norm can be replaced by an absolute value, so $\beta = 1$ holds.
For the SSOLC condition we have

$$||xd^2|| \leq 2\beta ||2xd^2||$$
$$\frac{1}{2} \leq \beta$$

(b) $f(x) = \log(x)$, $x > 0$.
We have

$$\nabla f(x) = \frac{1}{x}$$
$$\nabla^2 f(x) = -\frac{1}{x^2}$$

SOLC condition

$$||\frac{1}{x+d} - \frac{1}{x} + \frac{1}{x^2}|| \leq \beta||d||$$
$$||\frac{-d}{x(x+d)} + \frac{1}{x^2}|| \leq \beta||d||$$
$$||\frac{d^2}{x^2(x+d)}|| \leq \beta||d||$$
$$||\frac{1}{x^2(x+d)}|| \leq \beta$$

Thus there is no upperbound since we can set x arbitrarily close to zero, causing β to be aribtrarily large
Iin the SSOLC condition we have

$$||\frac{xd^2}{x^2(x+d)}|| \leq \beta||\frac{d^2}{x^2}||$$
$$||\frac{x}{(x+d)}|| \leq \beta$$
$$||\frac{1}{1+x^{-1}d}|| \leq \beta$$

Thus since $x^{-1}d$ is bounded above by .5, $\frac{1}{1+x^{-1}d} \geq \frac{2}{3}$, indicating that $\frac{2}{3} \leq \beta$. Now if $||x^{-1}d||$ goes to zero, then $||\frac{1}{1+x^{-1}d}|| = 1$. Therefore the tighter bound is that $1 \leq \beta$.

(c) $f(x) = \log(1+e^{-x})$, $x > 0$
We have

$$\nabla f(x) = \frac{-e^{-x}}{1+e^{-x}}$$
$$\nabla^2 f(x) = \frac{e^{-x}}{1+e^{-x}} + \frac{e^{-2x}}{(1+e^{-x})^2}$$
$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

Let $\sigma(x) = \frac{e^{-x}}{1+e^{-x}}$ Then

$$\nabla f(x) = -\sigma(x)$$
$$\nabla^2 f(x) = \sigma(x)(1-\sigma(x))$$

For the SOLC condition we have

$$||\frac{\nabla f(x+d)}{d^2} - \frac{\nabla f(x)}{d^2} - \frac{\nabla^2 f(x)}{d}|| \leq \beta$$

Now by the mean value theorem we know that $\exists \psi \in [x, x+d]$ such that

$$\nabla^2 f(\psi) = \frac{\nabla f(x+d) - \nabla f(x)}{d}$$

THus we have

$$||\frac{\nabla^2 f(\psi) - \nabla^2 f(x)}{d}|| \leq \beta$$

Applying the MVT again we can write for some $\mu \in [x, \psi]$,

$$\nabla^3 f(\mu) \approx \frac{\nabla^2 f(\psi) - \nabla^2 f(x)}{d}$$

There fore the original inequality reduces too:

$$||\nabla^3 f(\mu)|| \leq \beta$$

We have that the third derivative is equal to

$$\nabla^3 f(x) = \frac{e^x(1-e^x)}{(1+e^x)^3}$$

Plotting the third derivative in matlab, one can see that it is periodic, with a range of $[\frac{-1}{6\sqrt{3}}, \frac{1}{6\sqrt{3}}]$. Thus, we can establish the bound $\frac{1}{6\sqrt{3}} \leq \beta$.

In the SSOLC case we have

$$||\frac{-x}{1+e^{-x-d}} + \frac{x}{1+e^{-x}} - \frac{dx}{(1+e^{-x})^2} \leq \beta|\frac{-d}{x^2}|$$

$$||\frac{x^3 e^{-x}(e^{-d}-1)(1+e^{-x}) - dx^3(1+e^{-x-d})}{d^2(1+e^{-x-d})(1+e^{-x})^2}|| \leq \beta$$

$$\lim_{x\to\infty, d\neq 0} |\frac{x^3 e^{-x}(e^{-d}-1)(1+e^{-x}) - dx^3(1+e^{-x-d})}{d^2(1+e^{-x-d})(1+e^{-x})^2}| = \infty$$

$$\lim_{d\to 0} |\frac{x^3 e^{-x}(e^{-d}-1)(1+e^{-x}) - dx^3(1+e^{-x-d})}{d^2(1+e^{-x-d})(1+e^{-x})^2}| = \infty$$

Thus there exists no bound for β in both SOLC, and SSOLC.

Problem 2

In Logistic Regression, we like to determine x_0 and $\mathbf{x}$ to maximize

$$\left(\prod_{i,c_i=1} \frac{1}{1+\exp(-\mathbf{a}_i^T\mathbf{x} - x_0)}\right)\left(\prod_{i,c_i=-1} \frac{1}{1+\exp(\mathbf{a}_i^T\mathbf{x} + x_0)}\right).$$

which is equivalent to maximize the log-likelihood probability

$$-\sum_{i,c_i=1} \log\left(1+\exp(-\mathbf{a}_i^T\mathbf{x} - x_0)\right) - \sum_{i,c_i=-1} \log\left(1+\exp(\mathbf{a}_i^T\mathbf{x} + x_0)\right).$$

Or to minimize the log-logistic-loss

$$\sum_{i,c_i=1} \log\left(1+\exp(-\mathbf{a}_i^T\mathbf{x} - x_0)\right) + \sum_{i,c_i=-1} \log\left(1+\exp(\mathbf{a}_i^T\mathbf{x} + x_0)\right).$$

(a) Write down the Hessian matrix funtion of $\mathbf{x}, x_0$

Let $f(x, x_0)$ be the log-logistic loss function.

$$\nabla f(\mathbf{x}, x_0)_{x_j} = \sum_{i,c_i=1} \frac{-a_{ij}\exp[-\mathbf{a}_i^T\mathbf{x} - x_0]}{1+\exp[-\mathbf{a}_i^T\mathbf{x} - x_0]} + \sum_{i,c_i=-1} \frac{a_{ij}\exp[\mathbf{a}_i^T\mathbf{x} + x_0]}{1+\exp[\mathbf{a}_i^T\mathbf{x} + x_0]}\ \forall j$$

$$\nabla f(\mathbf{x}, x_0)_{x_0} = \sum_{i,c_i=1} \frac{-\exp[-\mathbf{a}_i^T\mathbf{x} - x_0]}{1+\exp[-\mathbf{a}_i^T\mathbf{x} - x_0]} + \sum_{i,c_i=-1} \frac{\exp[\mathbf{a}_i^T\mathbf{x} + x_0]}{1+\exp[\mathbf{a}_i^T\mathbf{x} + x_0]}$$

Let us define $\mathbf{z} = -\mathbf{a}_i^T \mathbf{x} - x_0$ and $\bar{\mathbf{z}} = \mathbf{a}_i^T \mathbf{x} + x_0$. We have that

$$\begin{aligned}
\nabla_{x_j,x_k} &= \sum_{i,c_i=1} a_{ij}a_{ik}[\frac{\exp[\mathbf{z}]}{1+\exp[\mathbf{z}]} - \frac{\exp[\mathbf{z}]}{(1+\exp[\mathbf{z}])^2}] + \sum_{i,c_i=-1} a_{ij}a_{ik}[\frac{\exp[\bar{\mathbf{z}}]}{1+\exp[\bar{\mathbf{z}}]} - \frac{\exp[\bar{\mathbf{z}}]}{(1+\exp[\bar{\mathbf{z}}])^2}] \\
&= \sum_{i,c_i=1} a_{ij}a_{ik}\frac{\exp[\mathbf{z}]}{(1+\exp[\mathbf{z}])^2} + \sum_{i,c_i=-1} a_{ij}a_{ik}\frac{\exp[\bar{\mathbf{z}}]}{(1+\exp[\bar{\mathbf{z}}])^2} \\
\nabla_{x_j,x_0} &= \sum_{i,c_i=1} a_{ij}\frac{\exp[\mathbf{z}]}{(1+\exp[\mathbf{z}])^2} + \sum_{i,c_i=-1} a_{ij}\frac{\exp[\bar{\mathbf{z}}]}{(1+\exp[\bar{\mathbf{z}}])^2} \\
\nabla_{x_0,x_0} &= \sum_{i,c_i=1} \frac{\exp[\mathbf{z}]}{(1+\exp[\mathbf{z}])^2} + \sum_{i,c_i=-1} \frac{\exp[\bar{\mathbf{z}}]}{(1+\exp[\bar{\mathbf{z}}])^2}
\end{aligned}$$

Thus the $i, j, (i, j) \in \{0, \ldots, n\}$ element of the hessian matrix is given by the equations above

---

- (b) (Computation Team Work) Apply any Quasi-Newton (e.g., slide 18 of Lecture 13 or L &Y Chapter 10) and Newton methods to solve the problem using the data in HW2 for SVM (may or may not with regulation), randomly generate data sets, and/or benchmark data sets you can find. Compare the two methods with each other and with the previous methods used in HW3.

Problem Three

Consider the LP problem

$$\begin{aligned}
\min_x f(x) = x_1 + x_2 \\
\text{Such that } : x_1 + x_2 + x_3 = 1 \\
(x_1, x_2, x_3) \geq 0
\end{aligned}$$

(a) What is the analytic center of the feasible region with the logarithmic barrier function

The analytic center is found by minimizing

$$\min_{x_i} - \log(x_1) - \log(x_2) - \log(1 - x_1 - x_2)$$

Taking the derivative with respect to x_1, x_2 we have

$$\begin{aligned}
2x_1 = 1 - x_2 \\
2x_2 = 1 - x_1
\end{aligned}$$

where the substitution $x_3 = 1 - x_1 - x_2$ was made. Solving the system of equations:

$$\begin{aligned}
x_1 = \frac{1}{3} \\
x_2 = \frac{1}{3} \\
x_3 = \frac{1}{3}
\end{aligned}$$

(b) Find the central path $\mathbf{x}(\mu) = (x_1(\mu), x_2(\mu), x_3(\mu))$.

The minimization problem is of the form

$$\min_{x_1,x_2} \quad x_1 + x_2 - \mu \log[x_1] - \mu \log[x_2] - \mu \log[1 - x_1 - x_2]$$

Differentiating with respect to x_1, x_2 we have

$$
\begin{aligned}
\nabla_{x_1} \to 1 - \frac{\mu}{x_1} + \frac{\mu}{1 - x_1 - x_2} = 0 \\
\nabla_{x_2} \to 1 - \frac{\mu}{x_2} + \frac{\mu}{1 - x_1 - x_2} = 0
\end{aligned}
$$

Adding these two equations together, we have that $x_1 = x_2$. Thus:

$$
\begin{aligned}
\nabla_x \to 1 - \frac{\mu}{x} + \frac{\mu}{1 - 2x} = 0 \\
x(1 - 2x) - \mu(1 - 2x) + \mu x = 0 \\
2x^2 - x(3\mu + 1) + \mu = 0 \\
x = \frac{3\mu + 1 \pm \sqrt{9\mu^2 - 2\mu + 1}}{4}
\end{aligned}
$$

Now noting that as $\lim_{\mu\to\infty}$ must converge to the analytic center we can eliminate the plus, so that

$$
x = \frac{3\mu + 1 - \sqrt{9\mu^2 - 2\mu + 1}}{4}
$$

To check the accuracy, lets take the limit as $\mu \to \infty$

$$
\begin{aligned}
\lim_{\mu\to\infty} \frac{3\mu + 1 - \sqrt{9\mu^2 - 2\mu + 1}}{4} &= \frac{1}{4} \lim_{\mu\to\infty} 3\mu + 1 - \sqrt{9\mu^2 - 2\mu + 1} \\
&= \frac{1}{4}[\lim_{\mu\to\infty} (3x - \sqrt{9\mu^2 - 2\mu + 1}) + 1] \\
&= \frac{1}{4}[\lim_{\mu\to\infty} \frac{2\mu - 1}{3\mu + \sqrt{9\mu^2 - 2\mu + 1}} + 1] \\
&= \frac{1}{4}[2 \lim_{\mu\to\infty} \frac{\mu}{3\mu + \sqrt{9\mu^2 - 2\mu + 1}} + 1] \\
&= \frac{1}{4}[2 \lim_{\mu\to\infty} \frac{1}{3 + \frac{\sqrt{9\mu^2 - 2\mu + 1}}{\mu}} + 1] \\
&= \frac{1}{4}[2 \lim_{\mu\to\infty} \frac{1}{3 + \frac{\sqrt{9\mu^2 - 2\mu + 1}}{\mu}} + 1] \\
&= \frac{1}{4}[2 \frac{1}{\lim_{\mu\to\infty}(3 + \frac{\sqrt{9\mu^2 - 2\mu + 1}}{\mu}}) + 1] \\
&= \frac{1}{4}[\frac{2}{\lim_{\mu\to\infty}(\sqrt{\frac{9\mu^2 - 2\mu}{\mu^2}}) + 3} + 1] \\
&= \frac{1}{4}[\frac{2}{\sqrt{\lim_{\mu\to\infty} \frac{9\mu^2 - 2\mu}{\mu^2}} + 3} + 1] \\
&= \frac{1}{4}[\frac{2}{\sqrt{\lim_{\mu\to\infty}(9 - \frac{2}{\mu})} + 3} + 1] \\
&= \frac{1}{4}[\frac{2}{\sqrt{\lim_{\mu\to\infty}(9 - \frac{2}{\mu})} + 3} + 1] \\
&= \frac{1}{4}[\frac{4}{3}] = \frac{1}{3}
\end{aligned}
$$

Thus

$$x_1(\mu) = x_2(\mu) = \frac{3\mu + 1 - \sqrt{9\mu^2 - 2\mu + 1}}{4}$$
$$x_3(\mu) = 1 - \frac{3\mu + 1 - \sqrt{9\mu^2 - 2\mu + 1}}{4}$$

(c) Whos that as μ decreases to 0, $\mathbf{x}(\mu)$ converges to the unique optimal solution. We see that the

$$\lim_{\mu \to 0} x_{1,2}(\mu) = 0$$

Thus, the optimal solution corresponds to $x_1, x_2 = 0, x_3 = 1$, which is the smallest value the objective function can take while still satisfying the constraint set.

(d) (Computational Team Work) Draw $\mathbf{x}$ part of the the primal-dual potential function level sets

$$\phi_6(\mathbf{x}, \mathbf{s}) \le 0 \quad \text{and } \phi_6(\mathbf{x}, \mathbf{s}) \le -10$$

and

$$\phi_{12}(\mathbf{x}, \mathbf{s}) \le 0 \quad \text{and } \phi_{12}(\mathbf{x}, \mathbf{s}) \le -10$$

respectively in the primal feasible region (on a plane).

Problem 4

Questions (a) and (b) of Problem 7, Section 5.9 in textbook

Hint: Use the fact that for any feasible pair $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ of LP,

$$(\mathbf{x} - \mathbf{x}(\mu))^T(\mathbf{s} - \mathbf{s}(\mu)) = 0$$

the optimality of the central path solutions.

Let $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ be the central path of 5.9. Then prove

(a) The central path point $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ is bounded for $0 < \mu \le \mu^0$ and any given $0 < \mu^0 < \infty$.
We have that $(\mathbf{x}(\mu^0) - \mathbf{x}(\mu))^\top(\mathbf{s}(\mu^0) - \mathbf{s}(\mu))^\top = 0$.
Thus

$$\sum_j^n (\mathbf{s}(\mu^0)_j \mathbf{x}(\mu)_j + \mathbf{x}(\mu^0)_j \mathbf{s}(\mu)_j) = n(\mu^0 + \mu) \le 2n\mu^0$$

Thus

$$\sum_j^n \left(\frac{\mathbf{x}(\mu)_j}{\mathbf{x}(\mu^0)_j} + \frac{\mathbf{s}(\mu)_j}{\mathbf{s}(\mu^0)_j}\right) \le 2n$$

Thus $\mathbf{x}(\mu)$, $\mathbf{s}(\mu)$ are bounded. Since the KKT condition of the barrier porblems require that $\mathbf{s} = -A^T\mathbf{y} + \nabla f(\mathbf{x})^T$, it follows that since $\mathbf{s}(\mu)$ is bounded, $\mathbf{y}(\mu)$ must be bounded as well.

(b) For $0 < \mu' < \mu$

$$\mathbf{c}^\top \mathbf{x}(\mu') \le \mathbf{c}^\top \mathbf{x}(\mu) \text{ and } \mathbf{b}^\top \mathbf{y}(\mu') \ge \mathbf{b}^\top \mathbf{y}(\mu)$$

Furthermore if $\mathbf{x}(\mu') \ne \mathbf{x}(\mu)$ and $\mathbf{y}(\mu') \ne y(\mu)$,

$$\mathbf{c}^\top \mathbf{x}(\mu') < \mathbf{c}^\top \mathbf{x}(\mu) \text{ and } \mathbf{b}^\top \mathbf{y}(\mu') > \mathbf{b}^\top \mathbf{y}(\mu)$$

Now we know that for a given (μ', μ) that

$$\mathbf{c}^\top \mathbf{x}(\mu) - \mu \sum_j \log[\mathbf{x}(\mu)_j] \leq \mathbf{c}^\top \mathbf{x}(\mu') - \mu \sum_j \log[\mathbf{x}(\mu')_j]$$

Since $\mathbf{x}(\mu)$ minimizes $\mathbf{c}^\top \mathbf{x}(\mu) - \mu \sum_j \log[\mathbf{x}(\mu)_j]$ for a given μ. Similarly we know that

$$\mathbf{c}^\top \mathbf{x}(\mu') - \mu' \sum_j \log[\mathbf{x}(\mu')_j] \leq \mathbf{c}^\top \mathbf{x}(\mu) - \mu' \sum_j \log[\mathbf{x}(\mu)_j]$$

Adding together these two equations

$$(\mu - \mu') \sum_j \log[\mathbf{x}(\mu')_j] \leq (\mu - \mu') \sum_j \log[\mathbf{x}(\mu)]$$

Thus

$$\sum_j \log[\mathbf{x}(\mu')_j] \leq \sum_j \log[\mathbf{x}(\mu)_j]$$

Therefore we have that

$$\mathbf{c}^\top \mathbf{x}(\mu') - \mathbf{c}^\top \mathbf{x}(\mu) \leq u'[\sum_j \log[\mathbf{x}(\mu')_j] - \sum_j \log[\mathbf{x}(\mu)_j]]$$

Plugging in the inequality that $\sum_j \log[\mathbf{x}(\mu')_j] - \sum_j \log[\mathbf{x}(\mu)_j] \leq 0$ we have that

$$\mathbf{c}^\top \mathbf{x}(\mu') - \mathbf{c}^\top \mathbf{x}(\mu) \leq u'[\sum_j \log[\mathbf{x}(\mu')_j] - \sum_j \log[\mathbf{x}(\mu)_j]] \leq 0$$

And thus that:

$$\mathbf{c}^\top \mathbf{x}(\mu') \leq \mathbf{c}^\top \mathbf{x}(\mu)$$

Now if $\mathbf{x}(\mu') \neq \mathbf{x}(\mu)$ the inequalities become strict so that

$$\sum_j \log[\mathbf{x}(\mu')_j] < \sum_j \log[\mathbf{x}(\mu)_j]$$

Thus that

$$\mathbf{c}^\top \mathbf{x}(\mu') < \mathbf{c}^\top \mathbf{x}(\mu)$$

In the dual case we have that for a given μ, $\mathbf{y}(\mu)$ maximizes

$$\mathbf{b}^\top \mathbf{y}(\mu) + \mu \sum_{j=1}^n \log[\mathbf{s}(\mu)_j]$$

Thus for any μ' it must hold that

$$\mathbf{b}^\top \mathbf{y}(\mu) + \mu \sum_{j=1}^n \log[\mathbf{s}(\mu)_j] \geq \mathbf{b}^\top \mathbf{y}(\mu') + \mu \sum_{j=1}^n \log[\mathbf{s}(\mu')_j]$$

$$\mathbf{b}^\top \mathbf{y}(\mu') + \mu' \sum_{j=1}^n \log[\mathbf{s}(\mu')_j] \geq \mathbf{b}^\top \mathbf{y}(\mu) + \mu' \sum_{j=1}^n \log[\mathbf{s}(\mu)_j]$$

Adding the two equation, we have

$$(\mu - \mu') \sum_{j=1}^{n} \log[\mathbf{s}(\mu)_j] \geq (\mu - \mu') \sum_{j=1}^{n} \log[\mathbf{s}(\mu')_j]$$
$$\sum_{j=1}^{n} \log[\mathbf{s}(\mu)_j] \geq \sum_{j=1}^{n} \log[\mathbf{s}(\mu')_j]$$

Now in the case $\mathbf{y}(\mu) \neq \mathbf{y}(\mu')$ then the inequalities are strict, so that

$$\mathbf{b}^\top \mathbf{y}(\mu) + \mu \sum_{j=1}^{n} \log[\mathbf{s}(\mu)_j] > \mathbf{b}^\top \mathbf{y}(\mu') + \mu \sum_{j=1}^{n} \log[\mathbf{s}(\mu')_j]$$
$$\mathbf{b}^\top \mathbf{y}(\mu') + \mu' \sum_{j=1}^{n} \log[\mathbf{s}(\mu')_j] > \mathbf{b}^\top \mathbf{y}(\mu) + \mu' \sum_{j=1}^{n} \log[\mathbf{s}(\mu)_j]$$
$$\sum_{j=1}^{n} \log[\mathbf{s}(\mu)_j] > \sum_{j=1}^{n} \log[\mathbf{s}(\mu')_j]$$

Continuing we have that

$$\mathbf{b}^\top \mathbf{y}(\mu') - \mathbf{b}^\top \mathbf{y}(\mu) \geq \mu' [\sum_{j=1}^{n} \log[\mathbf{s}(\mu')_j] - \sum_{j=1}^{n} \log[\mathbf{s}(\mu)_j]] \geq 0$$

Therefore

$$\mathbf{b}^\top \mathbf{y}(\mu') \geq \mathbf{b}^\top \mathbf{y}(\mu)$$

Or in the strict case that:

$$\mathbf{b}^\top \mathbf{y}(\mu') > \mathbf{b}^\top \mathbf{y}(\mu')$$

Problem 5

Problem 12, Section 6.8, the text book L&Y, where for any given symmetric matrix $D, |D|^2$, is the sum of all its eigenvalue squares, and $|D|_\infty$ is its largest absolute eigenvalue.

Hint: $det(I + D)$ equals the produt of the eigenvalues of $I + D$ Then the proof follows from Taylor expansion.

Prove that if $\mathbf{D} \in \S^n$ and $|D|_\infty < 1$. Then

$$\text{trace}(\mathbf{D}) \geq \log[\det(I + \mathbf{D}) \geq \text{trace}(\mathbf{D}) - \frac{|D|^2}{2(1 - |\mathbf{D}|_\infty)}]$$

We know that the trace$(\mathbf{D}) = \sum_j \lambda_j$ where λ_j is the jth eigenvalue of the matrix D. Furthermore we know that since $\mathbf{D}$ is PSD, that it can be diagonalized so that $\det(I + \mathbf{D}) = \det(X(I + \Sigma)X^{-1}) = \det(I + \Sigma) = \prod_j (1 + \lambda_j)$ where Σ is a diagonal matrix of eigenvalues λ_j. Thus we have that

$$\text{trace}(\mathbf{D}) = \sum_j \lambda_j$$
$$\log[\det(I + \mathbf{D})] = \log[\prod_j (1 + \lambda_j)] = \sum_j \log(1 + \lambda_j)$$

Since $\max \lambda_j \le 1$, we know that $\log(1+\lambda_j) \le \lambda_j \; \forall j$. Therefore

$$\text{trace}(\mathbf{D}) = \sum_j \lambda_j \ge \log[\det(I+\mathbf{D}) = \sum_j \log(1+\lambda_j)$$

Now for a given j we have the taylor expansion

$$\log(1+\lambda_j) = \lambda_j - \frac{1}{2}\frac{\lambda_j^2}{(1+c_j)^2}$$

For some $c \in (0, \lambda_j]$ since $\lambda_j < 1$. Now, we know that each c_j is in the radius of $0, \max_j |\lambda_j|$. Thus we have that

$$\frac{1}{(1+c_j)^2} \le \frac{1}{1-\max_j |\lambda_j|}$$

Thus it follows that

$$\log(1+\lambda_j) \ge \lambda_j - \frac{\lambda_j^2}{2(1-\max_j |\lambda_j|)}$$

which in matrix forms indicates that

$$\log[\det(I+\mathbf{D})] = \sum_j \log[1+\lambda_j] \ge \sum_j \lambda_j - \frac{\lambda_j^2}{2(1-\max_j |\lambda_j|)}$$
$$= \text{trace}(D) - \frac{|D|^2}{2(1-|\mathbf{D}|_\infty)}$$

Problem 6

Optimization with log-sum-exponential functions arises from smooth approximation for non-smooth optimization. Consider the non-smooth optimization problem:

$$\min_{\mathbf{x}} \max_{1\le i<m} (\mathbf{a}_i^\top \mathbf{x} + b_i)$$

given $\mathbf{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}$.

(a) Derive an equivalent LP problem and write down its dual. Let A be a matrix of vectors $\mathbf{a}_i$. and $\mathbf{b}$ be a vector of b_i's. The primal optimization problem becomes

$$\min_z \quad z$$
$$\text{such that :}$$
$$Ax + b \preceq z\mathbf{1}$$

The dual is then

$$\max_y \quad \mathbf{b}^T y$$
$$\text{such that :}$$
$$A^T y = 0$$
$$\mathbf{1}^T y = 1$$
$$y \succeq 0$$

(b) Suppose we approximate the objective function $\max_{1\leq i\leq m}(\mathbf{a}_i^\top \mathbf{x} + b_i)$ with a smooth function and consider the optimization:

$$\min_x \log\big[\sum_{i=1}^m \exp(\mathbf{a}_i^\top \mathbf{x} + b_i)\big]$$

Then z_1 and z_2 be the optimal values of the two formulations, prove that

$$0 \leq z_2 - z_1 \leq \log(m)$$

Suppose that z^* is dual optimal for the dual general approximated program

$$\max \quad b^T z - \sum_j z_i \log[z_i]$$

Such that:

$$A^T z = 0$$
$$\mathbf{1}^T z = 1$$
$$z \succeq 0$$

In this case we have that z^* is also feasible for the dual of the piecwise linear formulation, that has objective value:

$$b^T z = z_2 + \sum_j z_j^* \log[z_j^*]$$

Furthermore from the concavity of the log we have that

$$\sum_j z_j \log[\frac{1}{z_j}] \leq \log[\sum_j 1] = \log m$$

Thus we have that

$$z_1 \geq z_2 + \sum_j z_j^* \log[z_j^*] \geq z_2 + \log(m)$$

Furthermore it holds that

$$\max_i(a_i^T x + b_i) \leq log[\sum_i e^{a_i^T x + b_i}]$$

To prove this, let $r \in \mathbb{R}^m$ and let $m = \max_i r$

$$\begin{aligned}
\log[\sum_i \exp(r_i)] &= \log[\sum_i \frac{\exp(m)}{\exp(m)} \exp(r_i)] \\
&= \log[\exp(m) \sum_i \frac{1}{\exp(m)} \exp(r_i)] \\
&= m + \log(\sum_i \exp(r_i - m)) \\
\log[\sum_i \exp(r_i)] &\geq m
\end{aligned}$$

Therefore we have that $z_1 \leq z_2$. Combining the inequalities yields

$$z_2 - \log(m) \leq z_1 \leq z_2$$
$$0 \leq z_2 - z_1 \leq \log(m)$$

and the proof is complete.

(c) Suppose we use a different function for approximation:

$$\min_{\mathbf{x}} \frac{1}{\gamma} \log \big(\sum_{i=1}^{m} \exp[\gamma(\mathbf{a}_i^\top \mathbf{x} + b_i)] \big)$$

for some $\gamma > 0$. Suppose the optimal value is z_3 derivat a bound for $z_3 - z_1$ similar as above. What happens as $\gamma \to \infty$.

This problem can be reformulated as:

$$\min_{\mathbf{x}} \quad \frac{1}{\gamma} \log \big(\sum_{i=1}^{m} \exp[\gamma(y_i)] \big)$$

Such that:

$$Ax + b = y$$

Forming the Lagrangian we have

$$L(x, y, \lambda) = \frac{1}{\gamma} \log \big(\sum_{i=1}^{m} \exp[\gamma(y_i)] \big) + \lambda^T (Ax + b - y)$$

Now notice that the lagrangian is unbounded below as a function of x unless $A^T \lambda = 0$. We are aiming to minimize the lagragian, or equivalently to maximize the conjugate function:

$$c(\lambda) = \sup_{y} \{ \lambda^T y - \log \big(\sum_{i=1}^{m} \exp[y_i] \big) \}$$

Thus we see that if $\lambda_k < 0$, setting $y_k = c, y_i = 0 \forall i \neq k$ then $\lim_{c \to -\infty}$ of the expression goes $-\infty$., Thus $\lambda_k > 0$. Similarly if $\lambda \succeq 0$, and $\mathbf{1}^T \lambda \neq 1$, then, we can have $y = c\mathbf{1}$ to find that

$$\lim_{t \to \infty} \lambda^T y - \log[\sum_i \exp(y_i)] = \lim_{t \to \infty} c\mathbf{1}^T \lambda - \log(m) - c$$

which goes to infinity or negative infinity depending on $\mathbf{1}^T z - 1$.

Taking the derivative with respect to y and setting it equal to zero we have

$$\lambda_i = \frac{e^{y_i}}{\sum_j e^{x_j}}$$

Plugging this back into $g(\lambda)$ we have $g^*(\lambda) = \sum_i y_i \log[y_i]$

In summary we have that the conjugate function equals:

$$c(\lambda) = \begin{cases} \sum_i y_i \log[y_i] \text{ if } \lambda \succeq 0 \text{ and } \mathbf{1}^T \lambda = 1 \\ 0 \text{ otherwise} \end{cases}$$

The lagrangian dual function can be given for $\lambda \succeq 0, \mathbf{1}^T \lambda = 1, A^T \lambda = 0$.

$$g(\lambda) = b^T \lambda - \frac{1}{\gamma} \sum_i \lambda_i \log[\lambda_i]$$

So that the dual problem can be formulated as:

$$\max_{\lambda} \quad b^T \lambda - \frac{1}{\gamma} \sum_i \lambda_i \log[\lambda_i]$$

$$\text{such that: } A^T \lambda = 0$$

$$\mathbf{1}^T \lambda = 1$$

Let z_3 be an optimal solution to the optimization above. We then know that z_3 is also feasible for the dual of the piecewise linear formulation, which has objective value

$$b^T \lambda = z_3 + \frac{1}{\gamma} \sum_i \lambda_i^* \log(z_i^*)$$

Thus we have that

$$z_1 \geq z_3 + \frac{1}{\gamma} \sum_i \lambda_i^* \log(z_i^*) \geq z_3 - \frac{1}{\gamma} \log[m]$$

Furthermore it follows as in the previous formulation that $z_3 \geq z_1$. It thus follows that

$$z_3 - \frac{1}{\gamma} \log[m] \leq z_1 \leq z_3$$

and therefore that

$$0 \leq z_3 - z_1 \leq \frac{1}{\gamma} \log[m]$$

Now, notice, as $\gamma \to \infty$, then $z_3 \to z_1$.