CPSC 406: Midterm practice questions

1. **Cholesky factorization.**

   (a) Show the steps used to compute the Cholesky factorization $A = LL^T$ of
   $$A = \begin{bmatrix} 1+\epsilon_1 & 1 \\ 1 & 1+\epsilon_2 \end{bmatrix}.$$
   Discuss what happens if either $\epsilon_1 \to 0$ or $\epsilon_2 \to 0$

   (b) Also, perform the Cholesky decomposition of
   $$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$
   Describe why this is impossible, and why that makes sense.

2. Show that $A = \begin{bmatrix} a & -a \\ -a & a \end{bmatrix}$ is positive semidefinite, but not positive definite.

3. Consider the block diagonal matrix
   $$A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}.$$
   Suppose that $B \succ 0$ and $C \prec 0$. Show that this implies A is indefinite.

4. Suppose that x and $\hat{x}$ both are optimal variables to the least squares problem.
   $$\underset{x}{\text{minimize}} \ \|Ax - b\|_2^2.$$
   Show that this implies $x - \hat{x}$ is in the null space of A.

5. Consider the matrix and vector
   $$A = \begin{bmatrix} I & I \\ I & I \end{bmatrix}, \ I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$
   Decompose $x = u + v$ where u is in the range of space of A and $v^\intercal u = 0$.

6. Beck 2.17

7. Suppose that $f(x) = x^\intercal Ax$ where $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. Suppose $t = 3$, and run gradient descent
   $$x^+ = x - t\nabla f(x).$$
   For what choice of x will this diverge? For what choice of t will this converge regardless of x?

8. **Gradient descent** Consider the minimization problem
   $$\underset{x\in\mathbf{R}^n}{\text{minimize}} \quad f(x) := \frac{1}{2}x^T Ax$$
   where A is symmetric positive semidefinite with largest eigenvalue / eigenvector pair $\lambda_{\max} > 0$, $u_{\max}$; that is,
   $$Au_{\max} = \lambda_{\max}u_{\max} \quad \text{and} \quad u_{\max}^T Au_{\max} = \max_{\|u\|_2=1} u^T Au = \lambda_{\max}.$$
   We now consider gradient descent on this objective
   $$x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$$
   where $x^{(k)}$ is the variable at iteration k, and $x^{(k+1)}$ is the variable at iteration $k+1$ (after 1 gradient step).

(a) Write the gradient of f at x.

(b) Recall that f is L-smooth if

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

What is L for $f(x) = \frac{1}{2}x^T Ax$?

(c) The *amount of descent* can be characterized as

$$f(x^{(k)}) - f(x^{(k+1)}) = \frac{1}{2}(x^{(k)})^T Ax^{(k)} - \frac{1}{2}(x^{(k)} - t\nabla f(x^{(k)}))^T A(x^{(k)} - t\nabla f(x^{(k)})).$$

Expand and simplify the right hand side. In particular, find c_1 and c_2 where

$$f(x^{(k)}) - f(x^{(k+1)}) = c_1 \nabla f(x^{(k)})^T \nabla f(x^{(k)}) + c_2 \nabla f(x^{(k)})^T A \nabla f(x^{(k)}).$$

(d) Explain why if $0 < t < 2/L$ and $x^{(k)}$ is not a stationary point, then $f(x^{(k)}) - f(x^{(k+1)}) > 0$ for any $x^{(k)}$.

(e) Now suppose $t > 2/L$. Give a direction u and show that for this choice of t and u, with $x^{(k+1)} = x^{(k)} - tu$, then

$$f(x^{(k+1)}) > f(x^{(k)}).$$

(f) Explain in 1-3 sentences why, when using gradient descent with constant step size on this objective, the recommendation is to have $t < 2/L$.