AMATH 515: Homework 1
Sid Meka

1. Let $g : \mathbb{R}^m \to \mathbb{R}$ be a twice differentiable function, $A \in \mathbb{R}^{m\times n}$ be any matrix, and h be the composition $g(Ax)$, then we have two simple generalizations of the chain rule that combine linear algebra with calculus:

$$\nabla h(x) = A^T \nabla g(Ax)$$

and

$$\nabla^2 h(x) = A^T \nabla^2 g(Ax) A.$$

(a) Show what happens when you apply the above chain rules to the special case

$$h(x) = g(a^T x)$$

where a is a vector and $g : \mathbb{R} \to \mathbb{R}$ is a univariate function.
Given $h(x) = g(a^T x)$, let $z = a^T x$, where $z \in \mathbb{R}$ is a scalar.
We compute the gradient and Hessian as desired:

i. For Gradient:
Using the Chain Rule:

$$\nabla h(x) = A^T \nabla g(Ax)$$

Substituting $A = a^T$, we have:

$$a \cdot g'(a^T x)$$

ii. For Hessian:
Using the Chain Rule:

$$\nabla^2 h(x) = A^T \nabla^2 g(Ax) A$$

Since $g(Ax) = g(a^T x)$ and $\nabla^2 g(Ax)$ for a univariate g is $g''(a^T x)$, we get:

$$\nabla^2 h(x) = aa^T g''(a^T x)$$

(b) Compute the gradient and hessian of the regularized logistic regression objective:

$$\left(\sum_{i=1}^{n} \log(1 + \exp(a_i^T x)) - b^T Ax \right) + \lambda \|x\|^2$$

where a_i denote the rows of A.
The objective function for this problem is:

$$\left(\sum_{i=1}^{n} \log(1 + \exp(a_i^T x)) - b^T Ax \right) + \lambda \|x\|^2$$

i. For Gradient:
The gradient of each term:

- For $\sum_{i=1}^{n} \log(1 + \exp(a_i^T x))$, the gradient is:

$$\nabla \left(\log(1 + \exp(a_i^T x))\right) = \frac{\exp(a_i^T x)}{1 + \exp(a_i^T x)} a_i$$

Summing over i:

$$\nabla\left(\sum_{i=1}^{n}\log(1+\exp(a_i^Tx))\right)=A^T\begin{bmatrix}\frac{\exp(a_1^Tx)}{1+\exp(a_1^Tx)}\\ \frac{\exp(a_2^Tx)}{1+\exp(a_2^Tx)}\\ \vdots\\ \frac{\exp(a_n^Tx)}{1+\exp(a_n^Tx)}\end{bmatrix}$$

- For $-b^TAx$, the gradient is $-A^Tb$.
- For $\lambda\|x\|^2$, the gradient is $2\lambda x$.

Combining, we get:

$$\nabla\left(\left(\sum_{i=1}^{n}\log(1+\exp(a_i^Tx))-b^TAx\right)+\lambda\|x\|^2\right)=A^T\begin{bmatrix}\frac{\exp(a_1^Tx)}{1+\exp(a_1^Tx)}\\ \frac{\exp(a_2^Tx)}{1+\exp(a_2^Tx)}\\ \vdots\\ \frac{\exp(a_n^Tx)}{1+\exp(a_n^Tx)}\end{bmatrix}-A^Tb+2\lambda x$$

ii. For Hessian:

The Hessian of each term:

- For $\sum_{i=1}^{n}\log(1+\exp(a_i^Tx))$, the Hessian is:

$$\nabla^2\left(\log(1+\exp(a_i^Tx))\right)=\frac{\exp(a_i^Tx)}{(1+\exp(a_i^Tx))^2}a_ia_i^T$$

Summing over i:

$$\nabla^2\left(\sum_{i=1}^{n}\log(1+\exp(a_i^Tx))\right)=A^T\text{diag}\left(\begin{bmatrix}\frac{\exp(a_1^Tx)}{1+\exp(a_1^Tx)}\\ \frac{\exp(a_2^Tx)}{1+\exp(a_2^Tx)}\\ \vdots\\ \frac{\exp(a_n^Tx)}{1+\exp(a_n^Tx)}\end{bmatrix}\odot\left(1-\begin{bmatrix}\frac{\exp(a_1^Tx)}{1+\exp(a_1^Tx)}\\ \frac{\exp(a_2^Tx)}{1+\exp(a_2^Tx)}\\ \vdots\\ \frac{\exp(a_n^Tx)}{1+\exp(a_n^Tx)}\end{bmatrix}\right)\right)A$$

- For $-b^TAx$, the Hessian is 0.
- For $\lambda\|x\|^2$, the Hessian is $2\lambda I$.

Combining, we get:

$$\nabla^2\left(\left(\sum_{i=1}^{n}\log(1+\exp(a_i^Tx))-b^TAx\right)+\lambda\|x\|^2\right)= \tag{1}$$
$$A^T\text{diag}\left(\begin{bmatrix}\frac{\exp(a_1^Tx)}{1+\exp(a_1^Tx)}\\ \frac{\exp(a_2^Tx)}{1+\exp(a_2^Tx)}\\ \vdots\\ \frac{\exp(a_n^Tx)}{1+\exp(a_n^Tx)}\end{bmatrix}\odot\left(1-\begin{bmatrix}\frac{\exp(a_1^Tx)}{1+\exp(a_1^Tx)}\\ \frac{\exp(a_2^Tx)}{1+\exp(a_2^Tx)}\\ \vdots\\ \frac{\exp(a_n^Tx)}{1+\exp(a_n^Tx)}\end{bmatrix}\right)\right)A+2\lambda I$$

(c) Compute the gradient and hessian of the regularized Poisson regression objective:

$$\left(\sum_{i=1}^{n}\exp(a_i^Tx)-b^TAx\right)+\lambda\|x\|^2$$

where a_i denote the rows of A.
The objective function for this problem is:

$$\sum_{i=1}^{n} \exp(a_i^T x) - b^T Ax + \lambda\|x\|^2$$

i. For Gradient:
The gradient of each term:

- For $\sum_{i=1}^{n} \exp(a_i^T x)$, the gradient is:

$$\nabla\left(\sum_{i=1}^{n} \exp(a_i^T x)\right) = A^T \exp(Ax)\ ,$$

  where $\exp(Ax)$ is the vector with entries $\exp(a_i^T x)$.
- For $-b^T Ax$, the gradient is: $\nabla(-b^T Ax) = -A^T b$.
- For $\lambda\|x\|^2$, the gradient is: $\nabla(\lambda\|x\|^2) = 2\lambda x$.

Combining, we get:

$$\nabla\left(\sum_{i=1}^{n} \exp(a_i^T x) - b^T Ax + \lambda\|x\|^2\right) = A^T \exp(Ax) - A^T b + 2\lambda x$$

ii. For Hessian:
The Hessian of each term:

- For $\sum_{i=1}^{n} \exp(a_i^T x)$, the Hessian is:

$$\nabla^2\left(\sum_{i=1}^{n} \exp(a_i^T x)\right) = A^T \mathrm{diag}(\exp(Ax))A\ ,$$

  where $\exp(Ax)$ is the vector with entries $\exp(a_i^T x)$.
- For $-b^T Ax$, the Hessian is:

$$\nabla^2(-b^T Ax) = 0$$

- For $\lambda\|x\|^2$, the Hessian is:

$$\nabla^2(\lambda\|x\|^2) = 2\lambda I$$

Combining, we get:

$$\nabla^2\left(\sum_{i=1}^{n} \exp(a_i^T x) - b^T Ax + \lambda\|x\|^2\right) = A^T \mathrm{diag}(\exp(Ax))A + 2\lambda I$$

(d) Compute the gradient and hessian of the regularized 'concordant' regression objective

$$\|Ax - b\|_2 + \lambda\|x\|_2.$$

Give conditions that ensure that the gradient and Hessian of this objective exist at a point x.
We have that the Gradient is:

$$\nabla(\|Ax - b\|_2 + \lambda\|x\|_2) = \frac{A^T(Ax-b)}{\|Ax-b\|_2} + \lambda\frac{x}{\|x\|_2}$$

We have that the Hessian is:

$$\nabla^2(\|Ax - b\|_2 + \lambda\|x\|_2) = \frac{1}{\|Ax-b\|_2}\left(A^T A - \frac{A^T(Ax-b)(Ax-b)^T A}{\|Ax-b\|_2^2}\right) + \frac{\lambda}{\|x\|_2}\left(I - \frac{xx^T}{\|x\|_2^2}\right)$$

We also need the conditions that:

- $Ax \neq b$ so that $\|Ax - b\|_2 \neq 0$
- $x \neq 0$ so that $\|x\|_2 \neq 0$

2. Show that each of the following functions is convex.

   (a) Indicator function to a convex set: $\delta_C(x) = \begin{cases} 0 & \text{if} \quad x \in C \\ \infty & \text{if} \quad x \notin C. \end{cases}$

   The indicator function $\delta_C(x)$ is defined as:

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{if } x \notin C. \end{cases}$$

   To verify convexity, we must show that for all $\theta \in [0, 1]$ and any points $x_1, x_2 \in \mathbb{R}^n$, the following holds:

$$\delta_C(\theta x_1 + (1-\theta)x_2) \leq \theta \delta_C(x_1) + (1-\theta)\delta_C(x_2).$$

   Case 1: $x_1 \in C$ and $x_2 \in C$
   If $x_1, x_2 \in C$, then by convexity of C, $\theta x_1 + (1-\theta)x_2 \in C$. Hence, $\delta_C(\theta x_1 + (1-\theta)x_2) = 0$. Since $\delta_C(x_1) = \delta_C(x_2) = 0$, the inequality becomes $0 \leq \theta \cdot 0 + (1-\theta) \cdot 0 = 0$, which holds.

   Case 2: At least one of x_1 or x_2 is not in C
   Suppose $x_1 \notin C$. Then $\delta_C(x_1) = \infty$, and the right-hand side of the inequality becomes ∞, which is trivially satisfied. Similarly, if $x_2 \notin C$, the right-hand side again becomes ∞. Thus, the inequality holds regardless.

   Conclusion: Since the inequality holds in all cases, the indicator function $\delta_C(x)$ is convex.

   (b) Support function to any set:

$$\sigma_C(x) = \sup_{c \in C} c^T x.$$

   The support function is defined as:

$$\sigma_C(x) = \sup_{c \in C} c^T x$$

   We must verify that:

$$\sigma_C(\theta x_1 + (1-\theta)x_2) \leq \theta \sigma_C(x_1) + (1-\theta)\sigma_C(x_2)$$

   For any $x_1, x_2 \in \mathbb{R}^n$, we have that:

$$\sigma_C(\theta x_1 + (1-\theta)x_2) = \sup_{c \in C} c^T(\theta x_1 + (1-\theta)x_2)$$

   By linearity of the inner product:

$$c^T(\theta x_1 + (1-\theta)x_2) = \theta(c^T x_1) + (1-\theta)(c^T x_2)$$

   Taking the supremum over $c \in C$, we get

$$\sigma_C(\theta x_1 + (1-\theta)x_2) = \sup_{c \in C} \left[\theta(c^T x_1) + (1-\theta)(c^T x_2)\right]$$

   Since sup is subadditive and homogeneous:

$$\sup_{c \in C} \left[\theta(c^T x_1) + (1-\theta)(c^T x_2)\right] \leq \theta \sup_{c \in C}(c^T x_1) + (1-\theta) \sup_{c \in C}(c^T x_2)$$

   Thus:

$$\sigma_C(\theta x_1 + (1-\theta)x_2) \leq \theta \sigma_C(x_1) + (1-\theta)\sigma_C(x_2)$$

   proving convexity.

(c) Any norm (see Chapter 1 of Boyd and Vandenbergh for the definition of a norm).
A norm $\|\cdot\|$ satisfies:

i. $\|x\| \geq 0$, and $\|x\| = 0$ if and only if $x = 0$

ii. $\|x + y\| \leq \|x\| + \|y\|$

iii. $\|\alpha x\| = |\alpha| \, \|x\|$

We can use the Triangle Inequality and the fact that the norm is absolutely scalable, we have that:

$$\|\lambda x + (1-\lambda)y\| \leq \|\lambda x\| + \|(1-\lambda)y\| = \lambda\|x\| + (1-\lambda)\|y\|$$

where $\lambda \in [0,1]$. Thus, we have shown that any norm is convex.

3. Prove the Cauchy Schwartz inequality: For any inner product $\langle \cdot, \cdot \rangle$ and vectors x, y,

$$|\langle x, y \rangle| \leq \|x\|\|y\|$$

Proof. If $x = 0$ or $y = 0$, then both sides of our inequality equal 0 and thus our desired inequality holds. Thus, we proceed by stating $x \neq 0$ and $y \neq 0$. Consider the orthogonal decomposition

$$x = \frac{\langle x, y \rangle}{\|y\|^2} y + z$$

where z is orthogonal to y where $z = x - \frac{\langle x,y \rangle}{\|y\|^2} y$.
By the Pythagorean Theorem,

$$\begin{aligned} \|x\|^2 &= \left\| \frac{\langle x, y \rangle}{\|y\|^2} y \right\|^2 + \|z\|^2 \\ &= \frac{|\langle x, y \rangle|^2}{\|y\|^2} + \|z\|^2 \\ &\geq \frac{|\langle x, y \rangle|^2}{\|y\|^2} \end{aligned}$$

Multiplying both sides of this inequality by $\|y\|^2$ gives us: $\|x\|^2\|y\|^2 \geq |\langle x, y \rangle|^2$.
Now, we take the square root of each side to state: $\|x\| \, \|y\| \geq |\langle x, y \rangle|$.
This is the same as $|\langle x, y \rangle| \leq \|x\| \, \|y\|$.
Therefore, we have $|\langle x, y \rangle| \leq \|x\| \, \|y\|$ as desired. □

4. Prove that for any twice differentiable function f,

$$f(x + u) = f(x) + \int_0^1 \langle \nabla f(x + tu), u \rangle \, \mathrm{d}t$$

Hint: What is the analogous statement for functions of one variable?

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. Define $g(t) = f(x + tu)$, where $t \in [0, 1]$. Observe that $g(t)$ is a composition of f with the line segment parametrized by $x + tu$, and hence g is differentiable on $[0, 1]$.
By the chain rule, we have that

$$g'(t) = \langle \nabla f(x + tu), u \rangle$$

Applying the Fundamental Theorem of Calculus to $g(t)$ over $[0, 1]$, we obtain:

$$g(1) - g(0) = \int_0^1 g'(t) \, \mathrm{d}t$$

Substituting $g(1) = f(x+u)$, $g(0) = f(x)$, and $g'(t) = \langle \nabla f(x+tu), u \rangle$, we have:

$$f(x+u) - f(x) = \int_0^1 \langle \nabla f(x+tu), u \rangle \, \mathrm{d}t$$

Rearranging, we find:

$$f(x+u) = f(x) + \int_0^1 \langle \nabla f(x+tu), u \rangle \, \mathrm{d}t$$

Thus, we have proven that for any twice differentiable function f,

$$f(x+u) = f(x) + \int_0^1 \langle \nabla f(x+tu), u \rangle \, \mathrm{d}t$$

$\square$

5. Suppose that $\nabla f(x)$ is $\beta-$Lipschitz, meaning that for all x, y,

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|$$

(a) Prove that

$$f(x+u) \leq f(x) + \langle \nabla f(x), u \rangle + \frac{\beta}{2} \|u\|^2 \, \mathrm{d}t$$

Hint: Upper bound the integral above with the absolute value of the integrand, then add and subtract $\nabla f(x)$ and apply Cauchy Schwartz.

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, and assume that f is $\beta-$smooth meaning that its gradient is Lipschitz continuous with parameter β:

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|, \quad \forall x, y \in \mathbb{R}^n$$

This implies that f satisfies the inequality:

$$f(x+u) \leq f(x) + \langle \nabla f(x), u \rangle + \frac{\beta}{2} \|u\|^2, \quad \forall x, u \in \mathbb{R}^n$$

We take the steps:

i. Start with Taylor Expansion: By Taylor's Theorem, we have:

$$f(x+u) = f(x) + \langle \nabla f(x), u \rangle + \int_0^1 \langle \nabla f(x+tu) - \nabla f(x), u \rangle \, \mathrm{d}t$$

ii. Bound the Integral: Using the Lipschitz continuity of the gradient:

$$\|\nabla f(x+tu) - \nabla f(x)\| \leq \beta \|tu\| = \beta t \|u\|$$

which implies:

$$|\langle \nabla f(x+tu) - \nabla f(x), u \rangle| \leq \|\nabla f(x+tu) - \nabla f(x)\| \cdot \|u\| \leq \beta t \|u\|^2$$

iii. Integrate:

$$\int_0^1 \beta t \|u\|^2 \, dt = \frac{\beta}{2} \|u\|^2$$

iv. Combine Terms:

$$f(x+u) \leq f(x) + \langle \nabla f(x), u \rangle + \frac{\beta}{2} \|u\|^2$$

Thus, the inequality is proven.

$\square$

(b) What can you say when $u = -s\nabla f(x)$
We start by simply substituting $u = -s\nabla f(x)$ into the inequality:

$$f(x - s\nabla f(x)) \leq f(x) + \langle \nabla f(x), -s\nabla f(x) \rangle + \frac{\beta}{2} \| - s\nabla f(x)\|^2$$

This simplifies to:

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

For small s the descent term $-s\|\nabla f(x)\|^2$ dominates ensuring $f(x - s\nabla f(x)) < f(x)$. As s increases, the quadratic term $\frac{\beta}{2} s^2 \|\nabla f(x)\|^2$ may dominate potentially leading to an increase in $f(x)$. This highlights the importance of selecting an appropriate step size s when dealing with gradient descent.

6. Contraction Mapping Theorem:

(a) Let $0 < \rho < 1$. We call a function $F : \mathbb{R}^n \to \mathbb{R}^n$ a contraction with parameter ρ if for all $x, y \in \mathbb{R}^n$, $\|F(x) - F(y)\| \leq \rho \|x - y\|$. Prove that any contraction with parameter $\rho < 1$ has a unique fixed point, that is, that there exists $x \in \mathbb{R}^n$ such that $F(x) = x$.
Hint: For existence, consider the sequence $x_k = F(x_{k-1})$, starting from any initial point x_0. Prove that this is a Cauchy sequence, and then use completeness of $\mathbb{R}^n$. This proof actually shows something even stronger, that iterating the map F, starting from any initial condition, converges to the unique fixed point of F.

To prove that any contraction $F : \mathbb{R}^n \to \mathbb{R}^n$ with parameter $0 < \rho < 1$ has a unique fixed point, we will have to:

- Prove existence of a fixed point
- Prove that such a fixed point is unique

i. Step 1: Existence of a Fixed Point

A. Defining the sequence x_k:
We define the sequence x_k by choosing an arbitrary initial point $x_0 \in \mathbb{R}^n$ and iterating the contraction map F:

$$x_k = F(x_{k-1}), \quad \text{for } k \geq 1$$

B. Showing that x_k is a Cauchy sequence:

$$\begin{aligned}
\|x_k - x_{k-1}\| &= \|F(x_{k-1}) - F(x_{k-2})\| && \text{(True for any } k \geq 1) \\
\|x_k - x_{k-1}\| &\leq \rho \|x_{k-1} - x_{k-2}\| && \text{(Using the contraction property of } F) \\
\|x_k - x_{k-1}\| &\leq \rho^{k-1} \|x_1 - x_0\| && \text{(Done by iterating this inequality)}
\end{aligned}$$

So from $\|x_k - x_{k-1}\| = \|F(x_{k-1}) - F(x_{k-2})\|$, we can get to $\|x_k - x_{k-1}\| \leq \rho^{k-1} \|x_1 - x_0\|$.
Next, we consider the distance between two points x_m and x_n in the sequence, where $m > n$. Using the triangle inequality, we have:

$$\|x_m - x_n\| \leq \|x_m - x_{m-1}\| + \|x_{m-1} - x_{m-2}\| + \cdots + \|x_{n+1} - x_n\|$$

Applying the contraction inequality to each term, we have:

$$\|x_m - x_n\| \leq \rho^{m-1}\|x_1 - x_0\| + \rho^{m-2}\|x_1 - x_0\| + \cdots + \rho^n\|x_1 - x_0\|$$

Factor out $\|x_1 - x_0\|$ and simplify the geometric sum:

$$\|x_m - x_n\| \leq \|x_1 - x_0\| \sum_{k=n}^{m-1} \rho^k = \|x_1 - x_0\|\rho^n \frac{1 - \rho^{m-n}}{1 - \rho}$$

As $m \to \infty$ and $n \to \infty$, the tail of this geometric series tends to zero because $0 < \rho < 1$. Therefore, x_k is a Cauchy sequence.

C. Using completeness of $\mathbb{R}^n$:
Since x_k is Cauchy and $\mathbb{R}^n$ is complete, the sequence x_k converges to some point $x^* \in \mathbb{R}^n$. That means we have:

$$x_k \to x^* \text{ as } k \to \infty$$

D. Showing that x^* is a fixed point of F:
We proceed by showing that x^* is a fixed point of F.
Taking the limit as $k \to \infty$ in the recursive relation $x_k = F(x_{k-1})$, we get:

$$x^* = \lim_{k\to\infty} x_k$$

Note that as $x_k = F(x_{k-1})$ for $k \geq 1$, we can substitute $\lim_{k\to\infty} x_k$ as $\lim_{k\to\infty} F(x_{k-1})$. Also, note that F is continuous.
Thus:

$$x^* = \lim_{k\to\infty} F(x_{k-1})$$

Since F is continuous, this implies:

$$x^* = F(x^*)$$

Thus, x^* is a fixed point of F.

ii. Step 2: Proving that our fixed point is unique: Next, we proceed by proving our fixed point is unique.

Proof. Suppose x^* and y^* are two fixed points chosen arbitrarily of F. This means $F(x^*) = x^*$ and $F(y^*) = y^*$.
We want to show that $x^* = y^*$.
Using the definition of F being a contraction:

$$\|F(x) - F(y)\| \leq \rho\|x - y\|, \quad \text{for all } x, y \in \mathbb{R}^n \ ,$$

we substitute x^* and y^* into this inequality:

$$\|F(x^*) - F(y^*)\| \leq \rho\|x^* - y^*\|$$

Since $F(x^*) = x^*$ and $F(y^*) = y^*$, the left-hand side simplifies to:

$$\|x^* - y^*\| \leq \rho\|x^* - y^*\|$$

Rearranging this inequality gives:

$$(1 - \rho)\|x^* - y^*\| \leq 0$$

Since $1 - \rho > 0$, which we can get from $0 < \rho < 1$, the only way this inequality can hold is if:

$$\|x^* - y^*\| = 0$$

Thus, $x^* = y^*$, proving that the fixed point is unique as x^* and y^* are chosen arbitrarily. $\square$

We have shown that iterating the map F starting from any initial condition, converges to a unique fixed point of F.

(b) What is the gradient of $f(x) = \frac{1}{2}x^T Ax - x^T b$?

$$
\begin{gathered}
\nabla f(x) = \\
\nabla\left(\frac{1}{2}x^T Ax - x^T b\right) = \\
\nabla\left(\frac{1}{2}x^T Ax\right) - \nabla x^T b = \\
\nabla\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j\right) - b = \\
\frac{1}{2}\sum_{j=1}^{n} A_{kj}x_j + \frac{1}{2}\sum_{i=1}^{n} A_{ik}x_i - b = \\
\sum_{j=1}^{n} A_{kj}x_j - b = \\
Ax - b
\end{gathered}
$$

So, $\nabla f(x) = Ax - b$.

(c) Consider the map $F(x) = x - s(Ax - b)$, for symmetric positive definite matrices A. Under what conditions on s and A is F a contraction? What are the fixed points of F?
Hint: You may want to consider the eigenvalue decomposition of A.

i. Step 1:
A point x^* is a fixed point of F if: $F(x^*) = x^*$

$$
\begin{aligned}
F(x^*) &= x^* \\
x^* &= x^* - s(Ax^* - b) && \text{(Substituting } F(x^*) \text{ as } F(x) = x - s(Ax - b)) \\
s(Ax^* - b) &= 0 \\
Ax^* - b &= 0 && \text{(As } s \neq 0 \text{ as it is our step size)} \\
Ax^* &= b \\
x^* &= A^{-1}b
\end{aligned}
$$

Going into more detail why $s \neq 0$:
If $s = 0$, $F(x)$ would reduce to the identity map $F(x) = x$, which cannot contract distances and has no fixed point dynamics. By having $s \neq 0$, we thus make sure that we can have iterative behavior and thus have $F(x)$ actually be a contraction. Furthermore, if $s = 0$, $F(x)$ would not contract distances, and thus, the iterative process $F(x_k)$ would never converge.

ii. Step 2: Contraction Property

A. Applying the Definition of Contraction:
A map $F(x)$ is a contraction if there exists $0 < \rho < 1$ such that:

$$
\|F(x) - F(y)\| \leq \rho\|x - y\|, \quad \forall x, y
$$

For the given $F(x)$, let us analyze:

$$
\|F(x) - F(y)\| = \|(x - s(Ax - b)) - (y - s(Ay - b))\|
$$

Simplifying:

$$\|F(x) - F(y)\| = \|x - y - s(Ax - Ay)\|$$

Factor out $x - y$:

$$\|F(x) - F(y)\| = \|(I - sA)(x - y)\| \ ,$$

where I is the identity matrix.

B. Using the definition of the spectral norm:
We will proceed by using the key idea that Contraction depends on the spectral norm of $I - sA$.

The Euclidean norm satisfies:

$$\|F(x) - F(y)\| \leq \|(I - sA)\|\|x - y\| \ ,$$

where $\|(I - sA)\|$ is the spectral norm or largest eigenvalue in magnitude of $I - sA$.

For F to be a contraction, we require:

$$\|(I - sA)\| < 1$$

iii. Step 3: Analyze the Spectral Norm of $I - sA$

Since A is symmetric positive definite, it has an eigenvalue decomposition of $A = Q\Lambda Q^T$, where Q is an orthogonal matrix meaning that $Q^T Q = I$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ contains the eigenvalues of A, which are all positive meaning that $\lambda_i > 0$ for our Matrix A.

This means:

$$I - sA = Q(I - s\Lambda)Q^T$$

The eigenvalues of $I - sA$ are $1 - s\lambda_i$ for $i = 1, 2, \ldots, n$. The spectral norm of $I - sA$ is:

$$\|(I - sA)\| = \max_i |1 - s\lambda_i|$$

For F to be a contraction, we require:

$$\max_i |1 - s\lambda_i| < 1$$

iv. Step 4: Solving for s

For all i, it follows:

$$\begin{aligned} |1 - s\lambda_i| &< 1 && \text{(Condition that was figured out as necessary in previous step)} \\ -1 < 1 - s\lambda_i &< 1 && \text{(Definition of absolute value inequality)} \\ 0 < s\lambda_i &< 2 \end{aligned}$$

Since $\lambda_i > 0$ for all i, this implies: $0 < s < \frac{2}{\lambda_{\max}}$, where $\lambda_{\max} = \max_i \lambda_i$ is the largest eigenvalue of A.

Thus, we have that F is a contraction if $0 < s < \frac{2}{\lambda_{\max}}$, where $\lambda_{\max}$ is the largest eigenvalue of A, and the unique fixed point of F is $x^* = A^{-1}b$.

(d) Fixed step size gradient descent, defined by the iteration $x_{k+1} = x_k - s\nabla f(x)$, can be seen as a fixed point iteration algorithm, iterating the map $F(x) = x - s\nabla f(x)$. What are the fixed points of F in terms of f?
Now let $f(x) = \frac{1}{2}x^T Ax - x^T b$. What can you conclude about the convergence of gradient descent with step size s applied to f? What choice of step size s minimizes the contraction constant ρ of F?

A fixed point x^* of $F(x)$ satisfies:

$$\begin{aligned} F(x^*) &= x^* \\ x^* &= x^* - s\nabla f(x^*) && \text{(Substituting } F(x^*) \text{ as } F(x) = x - s\nabla f(x)) \\ s\nabla f(x^*) &= 0 \\ \nabla f(x^*) &= 0 && \text{(As } s \neq 0 \text{ as it is our step size)} \end{aligned}$$

Going into more detail why $s \neq 0$:
If $s = 0$, $F(x)$ would reduce to the identity map $F(x) = x$, which cannot contract distances and has no fixed point dynamics. By having $s \neq 0$, we thus make sure that we can have iterative behavior and thus have $F(x)$ actually be a contraction. Furthermore, if $s = 0$, $F(x)$ would not contract distances, and thus, the iterative process $F(x_k)$ would never converge.

Note that the fixed points of $F(x)$ correspond to the critical points of $f(x)$, or in other words, where the gradient of $f(x)$ vanishes.

The Gradient of $\frac{1}{2}x^T Ax - x^T b$:

$$\nabla\left(\frac{1}{2}x^T Ax - x^T b\right) =$$

$$\nabla\left(\frac{1}{2}x^T Ax\right) - \nabla x^T b =$$

$$\nabla\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j\right) - b =$$

$$\frac{1}{2}\sum_{j=1}^{n} A_{kj}x_j + \frac{1}{2}\sum_{i=1}^{n} A_{ik}x_i - b =$$

$$\sum_{j=1}^{n} A_{kj}x_j - b =$$

$$Ax - b$$

So, the Gradient of $\frac{1}{2}x^T Ax - x^T b$ is $Ax - b$.

Now, we substitute $Ax - b$, which is $\nabla\left(\frac{1}{2}x^T Ax - x^T b\right) =$, into $F(x)$:

$$F(x) = x - s\nabla f(x)$$

$$F(x) = x - s(Ax - b)$$

Now, by using fixed point analysis, we find that $x^* = x^* - s(Ax^* - b)$. Thus, $s(Ax^* - b) = 0$. Remember that $s \neq 0$. Thus, we have that $Ax^* - b = 0$. $Ax^* = b$. Thus, $x^* = A^{-1}b$.

Now, we aim to show that this x^* is unique:
Assume x^* and y^* are two fixed points of $F(x)$ chosen arbitrarily. This means:

$$F(x^*) = x^* \quad \text{and} \quad F(y^*) = y^*$$

Substituting $F(x) = x - s(Ax - b)$ into these definitions, we have:

$$x^* = x^* - s(Ax^* - b) \quad \text{and} \quad y^* = y^* - s(Ay^* - b)$$

We can simplify this down to state:

$$s(Ax^* - b) = 0 \quad \text{and} \quad s(Ay^* - b) = 0$$

As $s \neq 0$,

$$Ax^* = b \quad \text{and} \quad Ay^* = b$$

We can subtract equations:

$$Ax^* - Ay^* = b - b$$
$$A(x^* - y^*) = 0$$

Note that A is symmetric positive definite. This means it is invertible. Thus, we have that $x^* - y^* = 0$. This means $x^* = y^*$. Since $x^* = y^*$, and our fixed points were chosen arbitrarily, this means our fixed point x^* is unique.

The convergence of gradient descent depends on the contraction property of $F(x)$. To analyze this, consider:

$$\|F(x) - F(y)\| = \|(I - sA)(x - y)\|$$

The map $F(x)$ is a contraction if $\|(I - sA)\| < 1$, where $\|(I - sA)\|$ is the spectral norm, or largest eigenvalue in magnitude, of $I - sA$.

We now continue on with Spectral Analysis:
Since A is symmetric positive definite, it has an eigenvalue decomposition of $A = Q\Lambda Q^T$, where Q is an orthogonal matrix meaning that $Q^T Q = I$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ contains the eigenvalues of A, which are all positive meaning that $\lambda_i > 0$ for our Matrix A.

This means:

$$I - sA = Q(I - s\Lambda)Q^T$$

The eigenvalues of $I - sA$ are $1 - s\lambda_i$ for $i = 1, 2, \ldots, n$. The spectral norm of $I - sA$ is:

$$\|(I - sA)\| = \max_i |1 - s\lambda_i|$$

For F to be a contraction, we require:

$$\max_i |1 - s\lambda_i| < 1$$

For all i, it follows:

$$\begin{aligned} |1 - s\lambda_i| &< 1 && \text{(Condition that was figured out as necessary in previous step)} \\ -1 < 1 - s\lambda_i &< 1 && \text{(Definition of absolute value inequality)} \\ 0 < s\lambda_i &< 2 \end{aligned}$$

Since $\lambda_i > 0$ for all i, this implies: $0 < s < \frac{2}{\lambda_{\max}}$, where $\lambda_{\max} = \max_i \lambda_i$ is the largest eigenvalue of A.

Thus, it follows:

- Gradient descent converges for $0 < s < \frac{2}{\lambda_{\max}}$.
- The rate of convergence depends on the contraction constant $\rho = \|(I - sA)\|$.

We now consider the Optimal step size to minimize ρ. To minimize the contraction constant $\rho = \|(I - sA)\| = \max_i |1 - s\lambda_i|$, consider the eigenvalues $1 - s\lambda_i$. The goal is to minimize the largest deviation from 0. Note that the worst case eigenvalue is either $(1 - s\lambda_{\max})$ or $(1 - s\lambda_{\max})$. Thus, to balance between these, we try to minimize our contraction constant ρ. The contraction constant ρ is determined by the spectral norm of $I - sA$. It is

$$\rho = \|(I - sA)\|$$

Remember that

$$\|(I - sA)\| = \max_i |1 - s\lambda_i| \ ,$$

where λ_i are the eigenvalues of A, and $s > 0$. Thus,

$$\rho = \max_i |1 - s\lambda_i|$$

If $s < 0$, the eigenvalues $1 - s\lambda_i$ grow arbitrarily large in magnitude since $\lambda_i > 0$ leading to divergence rather than contraction. If $s = 0$, the iteration becomes stagnant meaning that there is no movement and thus, no optimization occurs. Thus, we proceed by having $s > 0$.

So, we proceed with $\rho = \max_i |1 - s\lambda_i|$, where λ_i are the eigenvalues of A, and $s > 0$. The goal is to minimize ρ, which governs the convergence rate of $F(x)$.

The map $F(x) = x - s(Ax - b)$ corresponds to a fixed point iteration:

$$F(x_k) = x_{k+1}$$

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

Thus, we have that

$$F(x_k) = x_k - s\nabla f(x_k)$$

Recall that $f(x) = \frac{1}{2}x^T Ax - x^T b$ and $\nabla f(x) = Ax - b$.
We analyze $F(x)$ in terms of its spectral properties. The eigenvalues of $I - sA$ are:

$$\mu_i = 1 - s\lambda_i, \quad i = 1, 2, \ldots, n \ ,$$

where λ_i are eigenvalues of A.
We have ρ of $F(x)$, where ρ is the spectral norm of $I - sA$. Thus, we have:

$$\rho = \max_i |1 - s\lambda_i|$$

To minimize ρ, we will prove why we want $1 - s\lambda_i = 0$ for all eigenvalues λ_i

Proof. Gradient descent converges if the map $F(x)$ satisfies the contraction condition:

$$\|F(x) - F(y)\| \leq \rho\|x - y\|, \quad \text{where } \rho < 1$$

The spectral norm of $I - sA$ determines ρ:

$$\rho = \max_i |1 - s\lambda_i|$$

The contraction constant ρ represents the largest deviation of the eigenvalues of $I - sA$ from 0. To minimize ρ, we aim to make:

$$1 - s\lambda_i = 0, \quad \text{for all } i$$

If $1 - s\lambda_i = 0$ for all i, then:

$$\mu_i = 0, \quad \forall i$$

This means the spectral norm of $I - sA$ becomes:

$$\|(I - sA)\| = \max_i |\mu_i| = 0$$

When $\|(I - sA)\| = 0$, the map $F(x)$ becomes:

$$F(x) = 0(x - y)$$

which collapses all points to the unique fixed point $x^* = A^{-1}b$ in a single step as $F(x) = 0$ when this happens. Thus, this is the fastest possible convergence.

Thus, we have proven why to minimize ρ, we will want $1 - s\lambda_i = 0$ for all eigenvalues λ_i. □

The contraction constant ρ for the gradient descent map $F(x)$ is determined by the spectral norm of $I - sA$ given by:

$$\rho = \max_i |1 - s\lambda_i|$$

This ensures that the convergence rate is governed by the step size s and the eigenvalue distribution of A. For optimal convergence, we minimize ρ by carefully choosing s to balance the contraction along the directions in which the matrix A stretches or compresses vectors.

To minimize ρ we must balance between the two extreme eigenvalues of $I - sA$:

$$|1 - s\lambda_{\max}| \quad \text{and} \quad |1 - s\lambda_{\min}|$$

To balance such extremes, we require

$$|1 - s\lambda_{\max}| = |1 - s\lambda_{\min}|$$

This equation ensures that neither extreme eigenvalue dominates ρ. Breaking the absolute values into cases, there are two possibilities:

i. $1 - s\lambda_{\max} = -(1 - s\lambda_{\min})$
ii. $1 - s\lambda_{\max} = 1 - s\lambda_{\min}$

We will briefly explain the trivial case where $1 - s\lambda_{\max} = 1 - s\lambda_{\min}$:
Here, we see that $\lambda_{\max} = \lambda_{\min}$. In this case, because $\lambda_{\max} = \lambda_{\min}$, we have that all eigenvalues of A are equal. This means $I - sA = (1 - s\lambda)I$. Note that for this case, we can say λ to mean an eigenvalue of A as all eigenvalues are equal in this case. Thus, for the trivial case, the contraction constant, ρ, is automatically minimized since all eigenvalues are equal, and no balancing is required.

We now look at the non trivial case: $1 - s\lambda_{\max} = -(1 - s\lambda_{\min})$.
In this case, we have that:

$$1 - s\lambda_{\max} = -(1 - s\lambda_{\min})$$

$$1 - s\lambda_{\max} = -1 + s\lambda_{\min}$$

$$2 = s\lambda_{\max} + s\lambda_{\min}$$

$$2 = s(\lambda_{\max} + \lambda_{\min})$$
$$s = \frac{2}{\lambda_{\max} + \lambda_{\min}}$$

Thus, in this case, we have minimized ρ appropriately as desired.

In Summary:

- We have our fixed point $x^* = A^{-1}b$.
- Gradient descent converges for $0 < s < \frac{2}{\lambda_{\max}}$.
- The rate of convergence depends on the contraction constant $\rho = \|(I - sA)\|$.
- We have that for non trivial cases: $s = \frac{2}{\lambda_{\max}+\lambda_{\min}}$