Continuous Optimization **Spring 2025**

Exercise 8: Accelerated Gradient Descent

Lecturer: Aurelien Lucchi

Problem 1 (Convergence of GD on quadratic functions):

Consider the quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x}$ with $\mathbf{A}$ being symmetric and having bounded minimum and maximum eigenvalues: $\lambda_{\min}(\mathbf{A}) \geq \mu$ and $\|\mathbf{A}\| \leq L$.

1. Show that the gradient descent update can be written as

$$\mathbf{x}_{k+1} = \mathbf{V}(\mathbf{I} - \eta\mathbf{\Lambda})^k \mathbf{V}^\top \mathbf{x}_1,$$

   where $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$ (eigen-decomposition).

2. What can you conclude about the convergence rate as a function of the eigenvalues?

3. Choose $\eta = \frac{1}{L}$. What is the convergence rate of gradient descent?

Problem 2 (Convergence of AGD on quadratic functions):

Consider the quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x}$ with $\mathbf{A}$ being symmetric and having bounded minimum and maximum eigenvalues: $\lambda_{\min}(\mathbf{A}) \geq \mu$ and $\|\mathbf{A}\| \leq L$.

1. Write AGD as a recursion of the form $\mathbf{w}_k = \mathbf{G}\mathbf{w}_{k-1}$ and give the explicit form of the $\mathbf{G}$ matrix.

2. Following a similar analysis to Heavy ball (see lecture notes), write $\mathbf{G}$ as a block matrix and derive the eigenvalues of the $\mathbf{G}_i$ matrices that compose the blocks of $\mathbf{G}$.

3. Bonus (more difficult): Derive a rate of convergence for AGD based on your previous calculations. Compare this rate to the rate of GD obtained in Problem 1.

Problem 3 (Lyapunov analysis of AGD):

We optimize a function $f : \mathbb{R}^d \to \mathbb{R}$ that is convex and L-smooth using Accelerated Gradient Descent (AGD) whose update is

$$\begin{aligned}\mathbf{y}_k &= \mathbf{x}_k + \beta(\mathbf{x}_k - \mathbf{x}_{k-1})\\ \mathbf{x}_{k+1} &= \mathbf{y}_k - \eta\nabla f(\mathbf{y}_k).\end{aligned}$$

Define the following Lyapunov function $L_k := f(\mathbf{x}_k) + \frac{1}{2\eta}\|\mathbf{x}_k - \mathbf{x}_{k-1}\|^2$. Prove that for $\eta \leq \frac{1}{L}$ and $\beta \in [0, 1]$,

$$L_{k+1} - L_k \leq -\frac{1-\beta^2}{2\eta}\|\mathbf{x}_k - \mathbf{x}_{k-1}\|^2.$$

Problem 4 (Programming exercise):

Fill in the `TODOs` in the associated Jupyter Notebook.