Math 5467 – Final Exam (Due Wednesday May 11 by Midnight)

Instructions:

- Complete the problems below, and submit your solutions by uploading them to the Google form https://forms.gle/4t18ZzSMSqDhVt8E7.
- You must submit your exam by Wednesday May 11 @ midnight Central time zone.
- LaTeX is preferred, but you may also scan and upload legible handwritten solutions (with a good scanning app).

Problems:

1. Let a be a real number, with $|a| \neq 1$, and define $f \in L^2(\mathbb{Z}_n)$ by $f(k) = a^k$. Show that

$$\mathcal{D}f(\ell) = \frac{a^n - 1}{ae^{-2\pi i\ell/n} - 1}.$$

Hint: Write the sum defining the DFT $\mathcal{D}f(\ell)$ as a geometric series and use the formula

$$\sum_{k=0}^{n-1} r^k = \frac{r^n - 1}{r - 1}.$$

2. Assume f is convex, and let $x(t)$ satisfy the gradient descent ODE

$$x'(t) = -\nabla f(x(t))$$

with $x(0) = x_0$. Show that

$$f(x(t)) - f(x_*) \leq \frac{\|x_0 - x_*\|^2}{2t}$$

where $x_* \in \mathbb{R}^n$ is any minimizer of f. Hint: Define the energy

$$e(t) = t(f(x(t)) - f(x_*)) + \frac{1}{2}\|x(t) - x_*\|^2$$

and show that $e'(t) \leq 0$ so that $e(t) \leq e(0)$. You will find the inequality below useful, which holds for convex functions f:

$$f(x(t)) - f(x_*) \leq \nabla f(x(t))^T (x(t) - x_*). \tag{1}$$

3. Consider the version of Nesterov's accelerated gradient descent in the form

$$\begin{cases} y_{k+1} = x_k - \alpha \nabla f(x_k) \\ x_{k+1} = y_{k+1} + \dfrac{k-2}{k+1}(y_{k+1} - y_k). \end{cases} \tag{2}$$

(i) Show that Nesterov's accelerated gradient descent given in (2) satisfies

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{\alpha} + a_k \frac{x_k - x_{k-1}}{\sqrt{\alpha}} = -\nabla f(x_k) + \frac{k-2}{k+1}(\nabla f(x_{k-1}) - \nabla f(x_k)),$$

where $a_k = \frac{3}{\sqrt{\alpha}(k+1)}$.

(ii) Assume $x_k = x(\sqrt{\alpha}k)$ is the discretization of a smooth curve $x(t)$ for $t \geq 0$. Explain why it follows from part (i) that when we send $\alpha \to 0$ we obtain that x solves the ordinary differential equation (ODE)

$$x''(t) + \frac{3}{t}x'(t) = -\nabla f(x(t)). \tag{3}$$

This ODE is sometimes called *continuous time Nesterov*. From the continuum point of view, we can see that Nesterov's method is simlar to the heavy ball method with time-dependent friction $a(t) = \frac{3}{t}$. The friction starts off very large, but vanishes as $t \to \infty$, which allows for accelerated convergence for non-strongly convex functions that may be extremely flat near their minima.

(iii) Assume f is convex and let $x(t)$ satisfy (3) with $x(0) = x_0$ and $x'(0) = 0$. Show that

$$f(x(t)) - f(x_*) \leq \frac{2\|x_0 - x_*\|^2}{t^2},$$

where $x_* \in \mathbb{R}^n$ is any minimizer of f. Hint: Define the energy functional

$$e(t) = t^2(f(x(t)) - f(x_*)) + 2\left\|x(t) + \frac{t}{2}x'(t) - x_*\right\|^2.$$

Then show that

$$e'(t) = 2t(f(x(t)) - f(x_*)) + t^2\nabla f(x(t))^T x'(t) + 2t\left(x(t) + \frac{t}{2}x'(t) - x_*\right)^T\left(x''(t) + \frac{3}{t}x'(t)\right).$$

Apply the inequality (1) to the first term, and use the Nesterov ODE (3) in the last term. After simplifying you should get that $e'(t) \leq 0$ for all $t \geq 0$, so that $e(t) \leq e(0)$, from which the result follows.