# 555 HW2: Due February 18, 2021

## 1 Theoretical Problems

1. Consider the damped harmonic oscillator $x''(t) + \alpha x'(t) + \lambda x(t) = 0$ for $\alpha, \lambda > 0$.
   (a) In each of the above cases, give the form of the general solution.
   (b) What is the decay rate of a general solution in each of the above cases.
   (c) Determine for which conditions on $\alpha$ and $\lambda$ the system is overdamped, underdamped, or critically damped.

2. Let
$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$
   (a) Diagonalize the matrix $A$.
   (b) Give the solution to the system of differential equations $x''(t) + \alpha x'(t) + Ax(t) = 0$ for different parameter values $\alpha$.

3. Let $f \in S_{0,L}$, i.e. $f$ is a convex $L$-smooth objective and let $x^* \in \arg\min_x f(x)$.
   (a) Consider the gradient flow differential equation:
$$x'(t) = -\nabla f(x(t)). \tag{1}$$
   Using the Lyapunov function $L(t) = t(f(x(t)) - f(x^*)) + \frac{1}{2}\|x(t) - x^*\|_2^2$, show that
$$f(x(t)) - f(x^*) \leq \frac{1}{2t}\|x(0) - x^*\|_2^2. \tag{2}$$
   (Hint: Show that $L'(t) \leq 0$.)
   (b) Consider gradient descent with stepsize $s = \frac{1}{L}$, i.e.
$$x_{n+1} = x_n - \frac{1}{L}\nabla f(x_n). \tag{3}$$
   Using the Lyapunov function $L_n = n(f(x_n) - f(x^*)) + \frac{L}{2}\|x_n - x^*\|_2^2$, show that
$$f(x_n) - f(x^*) \leq \frac{L}{2n}\|x_0 - x^*\|_2^2. \tag{4}$$
   (Hint: Show that $L_{n+1} \leq L_n$.)
   (c) Consider the variably damped Hamiltonian dynamics
$$x'(t) = v(t),\ v'(t) = -\nabla f(x(t)) - \frac{3}{t}v(t). \tag{5}$$
   Using the Lyapunov function $L(t) = t^2(f(x(t)) - f(x^*)) + \frac{1}{2}\|2(x(t) - x^*) + tv(t)\|_2^2$ show that if $v(0) = 0$, then
$$f(x(t)) - f(x^*) \leq \frac{2}{t^2}\|x(0) - x^*\|_2^2. \tag{6}$$

(Hint: Show that $L'(t) \leq 0$.)

Note that an appropriate discretization of this dynamics leads to Nesterov's accelerated gradient descent

$$x_{n+\frac{1}{2}} = x_n + \beta_n(x_n - x_{n-1}),\ x_n = x_{n+\frac{1}{2}} - \frac{1}{L}\nabla f(x_{n+\frac{1}{2}}) \tag{7}$$

where $\beta_n = \frac{n}{n+3}$. For this method we have the convergence rate

$$f(x_n) - f(x^*) \leq \frac{2L}{n^2}\|x_0 - x^*\|_2^2. \tag{8}$$

The proof of this convergence uses a complicated Lyapunov function and can be found in [1], Theorem 3.3.2. For Nesterov's original argument, see [2], Section 2.2.1.

# 2 Computational Problems

1. Consider the $n \times n$ discrete Laplacian matrix $A$, for example for $n = 6$ we have

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

   (a) For any $n$, verify that all the $n$ eigenvalues, $\lambda_k$, and the corresponding eigenvectors, $\xi^k = (\xi_j^k)$, of $A$ can be obtained, for $1 \leq k \leq n$, as follows:

$$\lambda_k = 4\sin^2\frac{k\pi}{2(n+1)}, \quad \xi_j^k = \sin\frac{kj\pi}{n+1}\ (1 \leq j \leq n).$$

   Using that $\sin(x) \approx x$ for small values of $x$, estimate the condition number of $A$.

   (b) Implement gradient descent to solve this quadratic optimization problem with $n = 100$.

   (c) Implement Nesterov's accelerated gradient descent

$$x_{m+\frac{1}{2}} = x_m + \beta(x_m - x_{m-1}),\ x_{m+1} = x_{m+\frac{1}{2}} - \frac{1}{L}\nabla f(x_{m+\frac{1}{2}}), \tag{9}$$

   where $L = \lambda_{max}$, $\mu = \lambda_{min}$, and $\beta = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$, to solve the same system with $n = 100$. (Note that here $x_{-1} = x_0$).

   Note that the convergence rate of this algorithm can be proved using a Lyapunov argument, see [1], Theorem 3.3.1. For Nesterov's original argument, which uses estimate sequences, see [2], Section 2.2.1.

   (d) For $n = 100$, implement the conjugate gradient algorithm to optimize $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$, where $b = (1, 0, 0, ...)$.

   (e) Compare the convergence rates of each of these methods.

2. This is a continuation of a problem on the previous homework. Given

$$A_0 = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \in R(A_0)$$

   (a) Let $A_\epsilon = A_0 + \epsilon I$ and use gradient descent method to solve

$$x = \arg\min_{y \in \mathbb{R}^3}\left(\frac{1}{2}(A_\epsilon y, y) - (b, y)\right)$$

with the initial guess $x^0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and learning rate $\eta = \frac{1}{3}$, record the minimal number of iteration $m$ satisftying the stopping criterion that

$$\|A_\epsilon x^m - b\| \leq 10^{-8}. \tag{10}$$

| $\varepsilon$ | # of iter $= m$ |
|---|---|
| 1. | |
| $10^{-1}$ | |
| $10^{-2}$ | |
| $10^{-3}$ | |
| $10^{-4}$ | |
| 0. [singular case] | |

(b) Repeat the previous experiment with Nesterov's accelerated gradient descent (9) (use the eigenvalues of $A_\epsilon$ to determine the correct $\lambda$ and $L$).

(c) Preconditioned gradient descent: let

$$P = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

and

$$\underset{\sim}{A} = P^T A_\epsilon P = \begin{pmatrix} 1+\epsilon & -1 & 0 & \epsilon \\ -1 & 2+\epsilon & -1 & \epsilon \\ 0 & -1 & 1+\epsilon & \epsilon \\ \epsilon & \epsilon & \epsilon & 3\epsilon \end{pmatrix}, \underset{\sim}{b} = P^T b \text{ and } \underset{\sim}{D} = \mathrm{diag}(\underset{\sim}{A}) = \begin{pmatrix} 1+\epsilon & 0 & 0 & 0 \\ 0 & 2+\epsilon & 0 & 0 \\ 0 & 0 & 1+\epsilon & 0 \\ 0 & 0 & 0 & 3\epsilon \end{pmatrix}.$$

Use (preconditioned) gradient descent method to solve

$$\underset{\sim}{x} = \underset{\underset{\sim}{y} \in \mathbb{R}^4}{\arg\min} \left( \frac{1}{2}(\underset{\sim}{A}\underset{\sim}{y}, \underset{\sim}{y}) - (\underset{\sim}{b}, \underset{\sim}{y}) \right) \tag{11}$$

namely

$$\underset{\sim}{x}^{n+1} = \underset{\sim}{x}^n - \eta \underset{\sim}{D}^{-1}(\underset{\sim}{A}\underset{\sim}{x}^n - \underset{\sim}{b}), \quad n = 0, 1, 2, \cdots \tag{12}$$

with the initial guess $\underset{\sim}{x}^0 = \underset{\sim}{b}$ and learning rate $\eta = \frac{1}{3}$. Record the minimal number of iteration $m$ satisftying the stopping criterion that

$$\|A_\epsilon x^m - b\| \leq 10^{-8} \quad \text{where} \quad x^m = P\underset{\sim}{x}^m. \tag{13}$$

| $\varepsilon$ | # of iter $= m$ |
|---|---|
| 1. | |
| $10^{-1}$ | |
| $10^{-2}$ | |
| $10^{-3}$ | |
| $10^{-4}$ | |

# References


[1] Jonathan W. Siegel *Accelerated first-order optimization with orthogonality constraints.* Diss. UCLA, 2018.

[2] Yurii Nesterov *Lectures on convex optimization.* Vol. 137. Cham: Springer, 2018.