18.C06 Problem Set 7

A Julia notebook (`.ipynb`) version of this file is available on the [18.C06 Canvas site (https://canvas.mit.edu/courses/27395/assignments)](https://canvas.mit.edu/courses/27395/assignments) under "Assignments".

Due Friday Nov. 8 at **11am**; 10% penalty if it is turned in within 24 hours, and after that late psets will not be accepted. Submit in PDF format: a decent-quality scan/image of any handwritten solutions (e.g. get a scanner app on your phone or use a tablet), and a PDF printout of your Jupyter notebook showing your code and (clearly labeled) results.

TO GENERATE A PDF OF YOUR JUPYTER NOTEBOOK: From the *File* pull-down menu, select *Save and Export Notebook As*, and then select the *HTML* format (not PDF). Then open the downloaded HTML file with your favorite browser, and use the browser's *Print* function to generate the PDF file.

Problem 1 (5+5 points)

Suppose that A is a 3×3 real-symmetric matrix. (Recall from class that such a matrix has real eigenvalues and orthogonal eigenvectors.) Suppose its eigenvalues are $\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = -2$, and corresponding eigenvectors are x_1, x_2, x_3. You are given that $x_1 = [1, 0, 1]$ (denoting a column vector ala Julia).

(a) Give an approximate solution at $t = 100$ to $\frac{dx}{dt} = Ax$ for $x(0) = [1, 1, 0]$. (Give a specific quantitative vector, even if the vector is very big or very small; an answer like "≈ 0" or "$\approx \infty$" is not acceptable.)

(b) If $x_2 = [0, 1, 0]$, give a possible x_3. (You should *not* use these x_2, x_3 to solve part (a).)

Problem 2 (5+5 points)

In class, we saw that eigenvectors of distinct eigenvalues must be orthogonal for a real-symmetric (or Hermitian) matrix. In fact, there is a related orthogonality relationship for *arbitrary* diagonalizable matrices between the *left and right* eigenvectors.

(a) For an arbitrary diagonalizable real matrix $A = X\Lambda X^{-1}$, derive a relationship between eigenvectors $y_1, \ldots, y_m$ of A^T (the "left eigenvectors" of A) and the rows or columns of X^{-1}.

(b) Using the left eigenvectors y_k from (c), what must be the value of dot products $y_k^T x_j$ between these left eigenvectors and the ordinary ("right") eigenvectors x_j (the columns of X). Consider both $k = j$ and $k \neq j$.

Problem 3 (5 points)

Suppose that a 3x3 real-symmetric matrix A has orthonormal eigenvectors q_1, q_2, q_3 and corresponding eigenvalues $\lambda_1 = -2$, $\lambda_2 = 3$, $\lambda_3 = 0$. Give a compact SVD $\hat{U}\hat{\Sigma}\hat{V}^T$ of A (express the matrix entries in terms of these vectors and values).

Problem 4 (5+5)

Let A be an $m \times m$ positive semidefinite (PSD) matrix. In analyzing the "heavy-ball" accelerated gradient descent in [this course notebook (https://nbviewer.org/github/mitmath/18065/blob/main/notes/Quadratic-Gradient-Descent.ipynb)](https://nbviewer.org/github/mitmath/18065/blob/main/notes/Quadratic-Gradient-Descent.ipynb), we blueuced the problem to finding the eigenvalues of the $2m \times 2m$ matrix:

$$M = \begin{pmatrix} (1+\beta)I - \alpha A & -\beta I \\ I & 0 \end{pmatrix},$$

where I is the $m \times m$ identity matrix, and $\alpha > 0$ and $\beta \geq 0$ are real numbers. Suppose that A has orthonormal eigenvectors $q_1, \ldots, q_m$ with corresponding (real, positive) eigenvalues $\lambda_1, \ldots, \lambda_m$.

(a) For each eigenvector q_k of A with eigenvalue λ_k, consider the 2×2 matrix:

$$M_k = \begin{pmatrix} (1+\beta) - \alpha\lambda_k & -\beta \\ 1 & 0 \end{pmatrix}.$$

Let $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2$ be an eigenvector of M_k with eigenvalue λ, i.e., $M_k y = \lambda y$. Define $x \in \mathbb{R}^{2m}$ by: $x = \begin{pmatrix} y_1 q_k \\ y_2 q_k \end{pmatrix}$.

Show that $Mx = \lambda x$; that is, x is an eigenvector of M with eigenvalue λ. Since M_k has two eigenvectors y for each λ_k, you can construct two linearly independent eigenvectors x from each q_k. Hence, you can obtain all $2m$ eigenvalues of M by solving m of these 2×2 eigenvalue problems.## Problem 4 (5+5)

(b) In class, it was claimed that the optimal choices of α and β (which minimize the magnitude of the dominant eigenvalue of M) are:

$$\alpha = \left(\frac{2}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}\right)^2, \qquad \beta = \left(\frac{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}\right)^2,$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of A, respectively.

Show that for this choice of α and β, the matrix M_k has a complex conjugate pair of eigenvalues for any λ_k in the interval $\lambda_{\min} < \lambda_k < \lambda_{\max}$, and argue (hint: use the determinant) that the modulus of all these eigenvalues is exactly $\sqrt{\beta}$. For the sake of simplifying the computations, you can assume that $\lambda_{min} = 1$ and $\lambda_{max} = 9$ instead of the

---

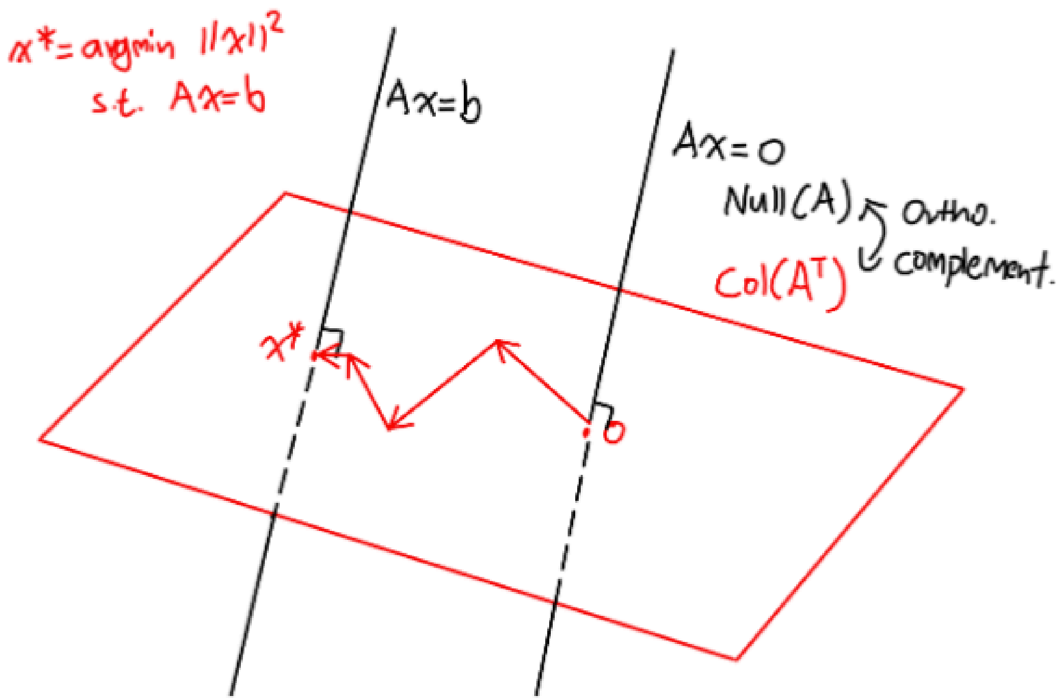
Problem 5 (5+5+5+5 points)

Suppose we have an underdetermined linear system $Ax = b$, where A is a $m \times n$ matrix, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$. We assume that A has full row rank and $m < n$, so that there are infinitely many solutions to this system. We have studied before how to compute the minimum norm solution to this problem,

$$\min_x \|x\|^2$$
$$\text{s.t. } Ax = b.$$

whose solution is $x_\star = A^+ b$ (or simply `A \ b` in Julia).

This question explores what happens when we perform gradient descent on the objective $\min_x f(x) = \|Ax - b\|^2$. The following picture is helpful to keep in mind while solving this problem.



(a) Show that for any x, the gradient step $\nabla f(x)$ lies in $C(A^T)$.

(b) Show that the minimum-norm solution $x_\star$ to the problem is achieved at the intersection of the set $\{x \in \mathbb{R}^m \mid Ax = b\}$ and the subspace $C(A^T)$, and this intersection is unique.

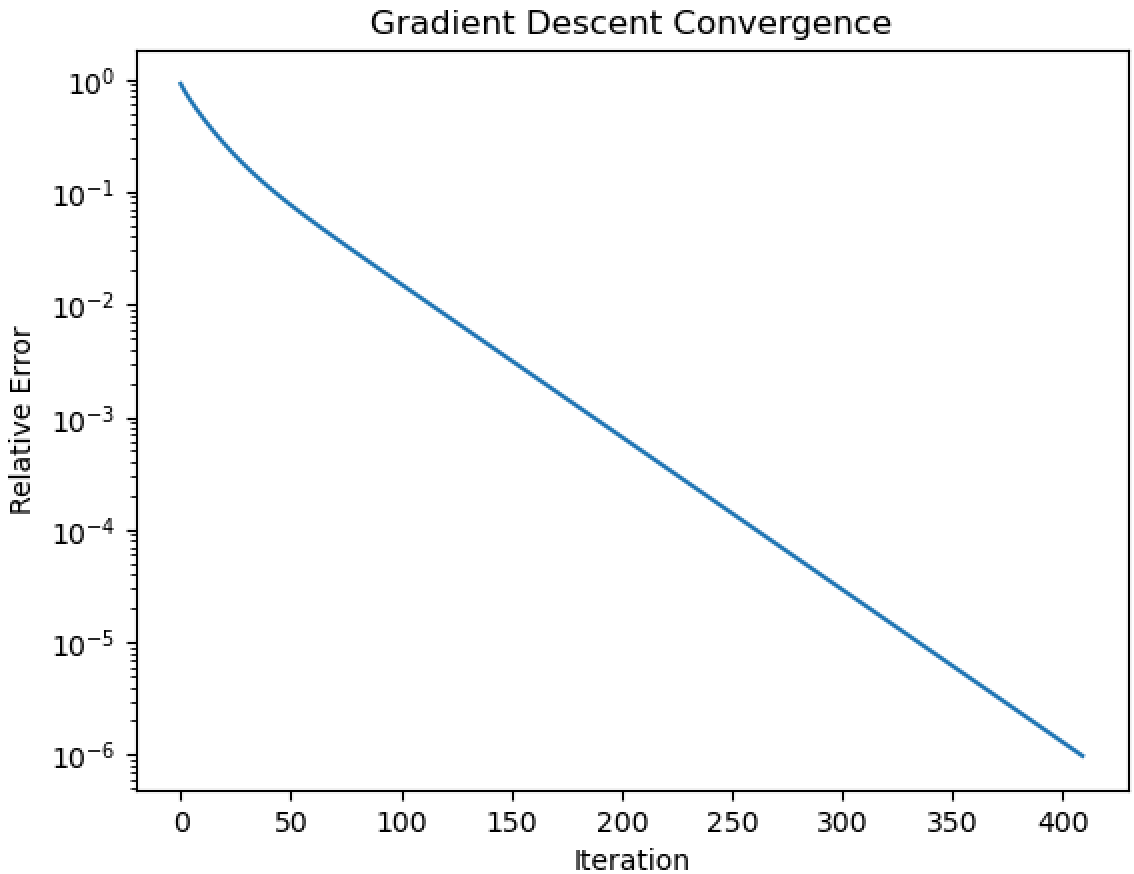
(c) Find a necessary and sufficient condition that a point $x_0 \in \mathbb{R}^m$ must satisfy so that gradient descent starting at x_0 converges to the min-norm solution.

(d) Implement gradient descent (with a fixed stepsize α is fine … doesn't need to be optimal, just small enough to converge) for this problem in Julia, try it on a random `A = randn(10,20)` matrix and a random `b = randn(10)`, starting from an initial guess `x =`

`zeros(20)`. (Why does this satisfy (c)?) Plot the error $\|x - x_\star\|/\|x_\star\|$ (`norm(x - xstar) / norm(xstar)`) on a `semilogy` scale as a function of the iteration number, where `xstar = A \ b` is the exact solution, and verify that the process is converging.

In [5]:
```
using LinearAlgebra
using Random
using PyPlot

Random.seed!(0)


A = randn(10, 20)
b = randn(10)
x = zeros(20)
xstar = A \ b

α = 0.01
max_iters = 1000
tolerance = 1e-6


errors = []


for iter in 1:max_iters

    gradient = A' * (A * x - b)

    x -= α * gradient


    relative_error = norm(x - xstar) / norm(xstar)
    push!(errors, relative_error)


    if relative_error < tolerance
        println("Converged after $iter iterations.")
        break
    end
end


PyPlot.plot(errors)
PyPlot.yscale("log")
PyPlot.xlabel("Iteration")
PyPlot.ylabel("Relative Error")
PyPlot.title("Gradient Descent Convergence")
# PyPlot.show()
```

```
Converged after 410 iterations.
```




```
Out[5]: PyObject Text(0.5, 1.0, 'Gradient Descent Convergence')
```

The initial guess x=0 satisfies the condition in c because the zero vector exists in all subspaces.