Exercise sheet 8

Conjugate Gradient Method
The following problems would outline a proof of the conjugate gradient method. We want to solve the system of equations

$$Ax = b, \tag{1}$$

where A is a positive definite, symmetric matrix. The solution is $x_* = A^{-1}b$. An alternative formulation of the problem is to minimize the objective function

$$f(x) = \frac{1}{2}x^T Ax - b^T x. \tag{2}$$

The Cayley-Hamilton theorem tells us that $A^{-1} = \mathcal{P}_n(A)$ where $\mathcal{P}_k$ is a polynomial of degree less than k. And therefore, x_* can be found by searching the Krylov subspaces $\mathcal{K}_k = \text{span}\{b, Ab, \dots A^{k-1}b\}$. The Krylov sequence of points $x_0, x_1, \dots x_n$ are such that they minimize the objective function of equation 2:

$$f(x_k) \leq f(\bar{x}_k) \quad \forall x_k, \bar{x}_k \in \mathcal{K}_k. \tag{3}$$

The conjugate gradient method considers two sets of bases for $\mathcal{K}_k$:

- a conjugate orthogonal basis $\mathcal{K}_k = \text{span}\{p_o, p_1, \dots p_{k-1}\}$, with $p_i^T A p_j \propto \delta_{ij}$, of search directions, and,
- an orthogonal basis $\mathcal{K}_k = \text{span}\{r_o, r_1, \dots r_{k-1}\}$, with $r_i^T r_j \propto \delta_{ij}$, of gradients.

The gradients r_k are found at the Krylov sequence points: $r_k = b - Ax_k$.

Exercise 8 - 1

Optimal Directions

a) Given the minimal point $x_k \in \mathcal{K}_k$ and the new (A-orthogonal) direction p_k to search, we seek the minimal point in $\mathcal{K}_{k+1}$. Argue that any point $\bar{x}_{k+1}$ in $\mathcal{K}_{k+1}$ can be written as:

$$\bar{x}_{k+1} = \bar{x}_k + \bar{\alpha}_k p_k \tag{4}$$

where $\bar{x}_k$ is an arbitrary point in $\mathcal{K}_k$. (1 point)

b) Show that the loss function of equation 2 can be written as

$$f(\bar{x}_{k+1}) = f(\bar{x}_k) + \underbrace{A}_{\bar{x}_k \text{ dependent}} + \underbrace{B}_{\bar{x}_k \text{ independent}}, \tag{5}$$

and find A and B. (2 points)

c) Use the idea of A-orthogonality of the $\{p_i\}$ to show that $p_k^T A \mathcal{K}_j = 0 \; \forall j \leq k$ and thereby that A from the above equation vanishes. (2 points)

d) The objective function separates into an $\bar{x}_k$ dependent and independent part:

$$f(\bar{x}_{k+1}) = f(\bar{x}_k) + B. \tag{6}$$

Argue that x_k is the optimal value for the function in the $\bar{x}_k$ dependent part in $\mathcal{K}_k$ (remember that x_k is in the Krylov sequence). Find the value α_k of $\bar{\alpha}_k$ that minimizes B. Thereby find the optimal value x_{k+1} of $\bar{x}_{k+1}$. (3 point)

Exercise 8 - 2

Induction
We use induction to derive the conjugate gradient method algorithm. The starting conditions are:

- $x_0 = 0 \in \mathcal{K}_0$,
- $r_0 = b - Ax_0 = b \in \mathcal{K}_1$, and,
- $p_0 = b \in \mathcal{K}_1$.

a) Given the point $x_k \in \mathcal{K}_k$ and direction $p_k \in \mathcal{K}_{k+1}$ to move in, the new point is given by some $x_{k+1} = x_k + \bar{\alpha}_k p_k$. Show that

$$r_{k+1} = r_k - \bar{\alpha}_k A p_k \tag{7}$$

and that $r_{k+1} \in \mathcal{K}_{k+2}$. (2 point)

b) Given that $r_i^T r_j = \delta_{ij}$ in $\mathcal{K}_{k+1}$, show that $r_i^T r_{k+1} = 0$ for $i < k$. HINT: Use exercise c in question 1. (2 points)

c) Find the value of $\bar{\alpha}_k$ such that $r_k^T r_{k+1} = 0$. Call it α_k. (1 point)

d) We now want to find a new search direction $p_{k+1} \in \mathcal{K}_{k+2}$ such that it is A-perpendicular (and in general not perpendicular) to all the previous search directions: $p_{k+1}^T A \mathcal{K}_{k+1} = 0$. As $r_{k+1} \in \mathcal{K}_{k+2}$, this search direction can be generally written as:

$$p_{k+1} = r_{k+1} + \sum_{j \le k} \bar{\beta}_j p_j. \tag{8}$$

Requiring that $p_i^T A p_{k+1} = 0$ for $i < k$, show that $\bar{\beta}_i = 0$ for $i < k$. HINT: Use equation 8. (3 points)

e) Thus,

$$p_{k+1} = r_{k+1} + \bar{\beta}_k p_k. \tag{9}$$

Find the value of $\bar{\beta}_k$ such that $p_k^T A p_{k+1} = 0$. Call it β_k. (1 point)

Exercise 8 - 3

Cleaning Up

a) Show that $r_k^T A p_k = p_k^T A p_k$. (1 point)

b) Show that $r_{i+1}^T p_k = r_i^T p_k$ for $i < k$. (1 point)

c) Using the expression for α_k from exercise 1, show that $r_{k+1}^T p_k = 0$. What is the geometric intuition behind this? (2 points)

d) Show that the values of α_k obtained from exercises 1 and 2 match one another. Show that $\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$ amd that $\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$. (3 points)

At the end we obtain the conjugate gradient descent algorithm wikipedia.org/wiki/Conjugate_gradient_method.

$k = 0$

$x_0 = 0 \quad \in \mathcal{K}_0$

$r_0 = b \quad \in \mathcal{K}_1$

$p_0 = b \quad \in \mathcal{K}_1$

REPEAT

$\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$

$x_{k+1} = x_k + \alpha_k p_k \quad \in \mathcal{K}_{k+1}$

$r_{k+1} = r_k - \alpha_k A p_k \quad \in \mathcal{K}_{k+2}$

If $r_{k+1} = b - Ax_{k+1}$ is small enough, exit

$\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

$p_{k+1} = r_{k+1} + \beta_k p_k \quad \in \mathcal{K}_{k+2}$

$k = k + 1$

Exercise 8 - 4

a) Consider the system of equations:

$$\begin{pmatrix} 3 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 450 \\ 0 \\ 0 \end{pmatrix}. \tag{10}$$

Use your favourite linear algebra method to solve this system. (2 points)

b) Use the conjugate gradient method to solve the system of equations and compare with the system of equations before. (4 points)

c) Either draw by hand in the x-y plane or use your favourite graphing software (e.g; geogebra.org/calculator) to graph the points x_k, gradients r_k and directions p_k. (Optional)

This exercise sheet will be discussed during the exercises.
Group 01, Wednesday 18:00 - 20:00, Theresienstr. 37, A 449,
Group 02, Thursday, 10:00 - 12:00, Theresienstr. 37, A 249,

https://wwwmpa.mpa-garching.mpg.de/ ensslin/lectures/lectures.html