DS-GA 3001.03: Homework Problem Set 10

Optimization and Computational Linear Algebra for Data Science

(Fall 2016)

Afonso S. Bandeira
bandeira@cims.nyu.edu
http://www.cims.nyu.edu/~bandeira

Due on December 8, 2016 It is a Thursday!

This homework problem set is due on December 8, at the Thursday Section. Notice the unusual deadline!!

There were a couple of typos, they are fixed in blue.

This homework problem is about the Conjugate Gradient Method, I highly recommend taking a look at [1]

If you have any questions about the homework post them on the piazza page for the course, contact Vlad Kobzar at `vkobzar@cims.nyu.edu` or myself at `bandeira@cims.nyu.edu`, or stop by our office hours.

Try not to look up the answers, you'll learn much more if you try to think about the problems without looking up the solutions. If you need hints, feel free to email me or Vlad.

You can work in groups but each student must write his/her own solution based on his/her own understanding of the problem. Please list, on your submission, the students you work with for the homework (this will not affect your grade).

If you need to impose extra conditions on a problem to make it easier (or consider specific cases of the question, like taking n to be 2, e.g.), state explicitly that you have done so. Solutions where extra conditions were assume, or where only special cases where treated, will also be graded (probably scored as a partial answer).

Problems with a $(*)$ are extra credit, they will not (directly) contribute to your score of this homework. However, for every 4 (four) extra credit questions

successfully answered you get a homework "bye": your lowest homework score (or one you did not hand in) gets replaced by a perfect score.

Throughout $A \in \mathbb{R}^{n \times n}$, $A \succ 0$, $b \in \mathbb{R}^n$. Also, f is a function $f : \mathbb{R}^n \to \mathbb{R}$ defined as $\mathbf{f(x) = \frac{1}{2}x^T A x - b^T x}$ and $x, x_0, x_1, \ldots, x_k, \cdots \in \mathbb{R}^n$. $x^\natural \in \mathbb{R}^n$ is the solution to $Ax^\natural = b$. $x^\natural$ **is such that** $Ax^\natural = b$**,** $e_k = x_k - x^\natural$**, and** $r_k = b - Ax_k$**.**

Problem 10.1

- *(a) Finding* $x^\natural \in \mathbb{R}^n$ *such that* $Ax^\natural = b$ *is equivalent to finding the minimizer* $x^\natural$ *of* $f(x)$*.*
- *(b) What is the gradient of* f *at* x_k*? How does it relate to the residual* r_k*?*
- *(c) Show that* $\mathbf{r_k = -Ae_k}$*.*

Problem 10.2

- *(a) Given a search direction* d_k*, what is the solution* α_k *of the line search problem:*

$$\alpha_k = \operatorname{argmin}_\beta f\left(x_k + \beta d_k\right) \tag{1}$$

  (notice that α_k *is not necessarily asked to be positive).*

- *(b) Show that if we take* $x_{k+1} = x_k + \alpha_k d_k$ *then*

$$d_k \perp r_{k+1}.$$

- *(c) What is the point in the line* $x_k + \beta d_k$ *that is closest to* $x^\natural$*? In other words, find* τ_k *defined as*

$$\tau_k = \operatorname{argmin}_\beta \left\| x^\natural - (x_k + \beta d_k) \right\|_2. \tag{2}$$

  (the issue with this of course is that we do not know $x^\natural$*)*

- *(d) Show that if we take* $x_{k+1} = x_k + \tau_k d_k$ *then*

$$d_k \perp e_{k+1}$$

Problem 10.3

Let $d_0, \ldots, d_{n-1}$ be an orthogonal basis and $x_0 \in \mathbb{R}^n$ an initialization. For $k = 0, \ldots, n-1$ take

$$x_{k+1} = x_k + \tau_k d_k,$$

where τ_k is solution of (2).

- *(a) Show that $e_k \perp d_j$ for all $j < k$.*
- *(b) Show that $x_n = x^\natural$, meaning that the procedure converges to $x^\natural$ in n steps.*
- *(c) What is the issue with this as an algorithm to find $x^\natural$?*

Problem 10.4

Let us a define a new notion of orthogonality, where $x \perp_A y$ if $x^T Ay = 0$ (these vectors are called A conjugate or A orthogonal).

- *(a) Show that if $x_{k+1} = x_k + \alpha_k d_k$ where α_k is the solution of (1) then*

  $$e_{k+1} \perp_A d_k.$$

  (note that (1) we can easily solve).

To try to make an algorithm that converges in n steps, motivated by (b) above, the idea will then be to define a new notion of orthogonal basis as well, and take an A-orthogonal basis to be $d_0, \ldots, d_{n-1}$ non-zero vectors such that $d_i \perp_A d_j$ for all $i \neq j$.

More precisely, let $d_0, \ldots, d_{n-1}$ be an A-orthogonal basis and $x_0 \in \mathbb{R}^n$ an initialization. For $k = 0, \ldots, n-1$ take

$$x_{k+1} = x_k + \alpha_k d_k,$$

where τ_k is solution of (1).

- *(b) Show that $e_k \perp_A d_j$ for all $j < k$.*
- *(c) Show that $x_n = x^\natural$, meaning that the procedure converges to $x^\natural$ in n steps.*

You just derived the Conjugate Gradient Method to solve (certain) linear systems!

$(*)$ **Problem 10.5 (For Extra Credit)**

- *Show that if $A \succ 0$ then an A-orthogonal basis exists. How would you construct it?*

References

[1] J. R. Shewchuk, *An Introduction to the Conjugate Gradient Method without the Agonizing Pain*, 1994