E. Herberg, M. Marić, V. Stein
Heidelberg University

Exercise 2

Date issued: 22nd April 2024
Date due: 30th April 2024

Homework Problem 2.1 (Solvability of quadratic polynomial minimization) 5 Points

Prove Lemma 3.1 of the lecture notes, i. e., the following statements for the optimization problem

$$\text{Minimize} \quad \phi(x) := \frac{1}{2} x^\top A x - b^\top x + c \quad \text{where } x \in \mathbb{R}^n \tag{3.1}$$

with symmetric $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$:

(i) If A is positive semidefinite, then the objective in (3.1) is convex. In this case, the following are equivalent:

   (a) The problem (3.1) possess at least one (global) minimizer.

   (b) The objective ϕ is bounded below.

   (c) $A x = b$ is solvable.

   The global minimizers of (3.1) are precisely the solutions of the linear system $A x = b$.

(ii) In case A is not positive semidefinite, the objective ϕ is not bounded below, thus problem (3.1) is unbounded.

Homework Problem 2.2 (M-steepest descent directions coincide with descent directions) 10 Points

(i) Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be differentiable at $x \in \mathbb{R}^n$ with $f'(x) \neq 0$. Show that the set of directions of steepest descent of f at x w.r.t. all inner product conincides with the set of descent directions, i. e., that

$$\{-M^{-1}\nabla f(x) \mid M \in \mathbb{R}^{n\times n} \text{ s.p.d.}\} = \{d \in \mathbb{R}^n \mid f'(x)d < 0\}.$$

Hint: For any $g, d \in \mathbb{R}^n$ such that $g^\intercal d > 0$ and starting with any s.p.d. matrix M, the low rank modifications

$$\tilde{M} := \left(\mathrm{Id} - \frac{gd^\intercal}{g^\intercal d}\right) M \left(\mathrm{Id} - \frac{dg^\intercal}{g^\intercal d}\right) + \frac{gg^\intercal}{g^\intercal d} \tag{DFP}$$

$$\tilde{M} := M - \frac{Mdd^\intercal M}{d^\intercal M d} + \frac{gg^\intercal}{g^\intercal d} \tag{BFGS}$$

yield s.p.d. matrices $\tilde{M}$ with $\tilde{M}d = g$.

(ii) Implement a method `compute_gradient(derivative, preconditioner)` that computes the M-gradient corresponding to a derivative $f'(x)$ and a preconditioner M and use it to visualize the result from task (i).

Homework Problem 2.3 (Implications of Termination Criteria) 4 Points

Prove Lemma 3.11 of the lecture notes, i. e., that when implementing an M-gradient descent scheme for solving s.p.d. quadratic problems with Matrix A, then

$$\|r^{(k)}\|_{M^{-1}} \le \varepsilon_{\text{rel}} \|r^{(0)}\|_{M^{-1}} \quad \Rightarrow \quad \begin{cases} \|x^{(k)} - x^*\|_A \le \sqrt{\kappa}\, \varepsilon_{\text{rel}} \|x^{(0)} - x^*\|_A \\ \|x^{(k)} - x^*\|_M \le \kappa\, \varepsilon_{\text{rel}} \|x^{(0)} - x^*\|_M \end{cases},$$

$$\|r^{(k)}\|_{M^{-1}} \le \varepsilon_{\text{abs}} \quad \Rightarrow \quad \begin{cases} \|x^{(k)} - x^*\|_A \le (1/\sqrt{\alpha})\, \varepsilon_{\text{abs}} \\ \|x^{(k)} - x^*\|_M \le (1/\alpha)\, \varepsilon_{\text{abs}} \end{cases},$$

$$\|r^{(k)}\|_{M^{-1}} \le \varepsilon_{\text{rel}} \|r^{(0)}\|_{M^{-1}} + \varepsilon_{\text{abs}} \quad \Rightarrow \quad \begin{cases} \|x^{(k)} - x^*\|_A \le \sqrt{\kappa}\, \varepsilon_{\text{rel}} \|x^{(0)} - x^*\|_A + (1/\sqrt{\alpha})\, \varepsilon_{\text{abs}} \\ \|x^{(k)} - x^*\|_M \le \kappa\, \varepsilon_{\text{rel}} \|x^{(0)} - x^*\|_M + (1/\alpha)\, \varepsilon_{\text{abs}} \end{cases},$$

$$\|r^{(k)}\|_{M^{-1}} \le \max\bigl\{\varepsilon_{\text{rel}} \|r^{(0)}\|_{M^{-1}},\ \varepsilon_{\text{abs}}\bigr\} \quad \Rightarrow \quad \begin{cases} \|x^{(k)} - x^*\|_A \le \max\bigl\{\sqrt{\kappa}\, \varepsilon_{\text{rel}} \|x^{(0)} - x^*\|_A,\ (1/\sqrt{\alpha})\, \varepsilon_{\text{abs}}\bigr\} \\ \|x^{(k)} - x^*\|_M \le \max\bigl\{\kappa\, \varepsilon_{\text{rel}} \|x^{(0)} - x^*\|_M,\ (1/\alpha)\, \varepsilon_{\text{abs}}\bigr\} \end{cases},$$

where $\alpha := \lambda_{\min}(A; M)$ and $\beta := \lambda_{\max}(A; M)$ are the extremal generalized eigenvalues of A w.r.t. M, and $\kappa := \frac{\beta}{\alpha}$ is the generalized condition number.

Homework Problem 2.4 (M-Gradient Method for Solving s.p.d. Linear Systems) 8 Points

Implement the M-gradient descent method as outlined in Algorithm 3.6 of the lecture notes. Additionally, include the option of supplying different methods for choosing the step sizes. Visualize the effects of different choices of the preconditioner, step size strategies and initial values.

Please submit your solutions as a single pdf and an archive of programs via moodle.