

Nonlinear Optimization

Fall 2019, Rutgers University

Prof. Eckstein

Homework 2: Unconstrained Optimization

Due September 26, 2019

1. *Coordinate descent* methods are algorithms that perform line search only along coordinate directions, that is the directions $\pm e^j$, where

$$\begin{aligned}e^1 &= (1, 0, 0, 0, \dots, 0) \\e^2 &= (0, 1, 0, 0, \dots, 0) \\e^3 &= (0, 0, 1, 0, \dots, 0) \\\vdots \\e^n &= (0, 0, 0, 0, \dots, 1).\end{aligned}$$

Suppose that at each iterate $x^k \in \mathbb{R}^n$, we choose the search direction d^k as follows, where “ ∂ ” stands for partial derivative:

$$d^k = - \left(\frac{\partial f(x^k)}{\partial x_{j_k}} \right) e^{j_k}, \text{ where } j_k \text{ maximizes } \left| \frac{\partial f(x^k)}{\partial x_j} \right| \text{ over } j \in \{1, \dots, n\}.$$

In other words, we search in the negative gradient direction in a coordinate that has the largest possible absolute gradient (note that there may be “ties”).

- (a) Show that this choice of search direction is gradient-related, as defined in the class notes. *Hint:* Base your argument on showing that $\langle \nabla f(x^k), d^k \rangle = -\|\nabla f(x^k)\|_\infty^2$, where $\|\cdot\|_\infty$ denotes the “infinity norm” $\|v\|_\infty = \max_{i=1, \dots, n} |v_i|$.
(b) Argue that the gradient-related convergence theorem in the class notes implies that if in the iteration $x^{k+1} = x^k + \alpha_k d^k$ we choose α_k so that it exactly minimizes $f(x^k + \alpha d^k)$ over α , then all limit points x^∞ of $\{x^k\}$ have $\nabla f(x^\infty) = 0$.
2. *Sublinear convergence of gradient methods near “flat” local minima:* Consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = x^4$. *Instructor’s note:* while it is simplest to investigate this phenomenon in \mathbb{R}^1 , it can certainly happen in higher dimension.

Suppose we produce a sequence $\{x_k\} \subset \mathbb{R}$ by applying the steepest descent iteration (in \mathbb{R}^1) to f , using the Armijo stepsize rule with any choice of parameters $s > 0$ and $\sigma, \beta \in (0, 1)$.

- (a) Use the gradient-related convergence theorem proved in class to argue that the iterates must converge to 0 for any starting point $x_0 \in \mathbb{R}$.

- (b) Show that there is an $\bar{x} > 0$ such that if $x_{\bar{k}} \in (0, \bar{x})$ for any \bar{k} , then $x_k > 0$ for all $k \geq \bar{k}$ and $\{x^k\}$ converges sublinearly in the sense that $\lim |x_{k+1}| / |x_k| = 1$.
- (c) Show that sublinear convergence also ensues if $x_{\bar{k}} \in (-\bar{x}, 0)$ for any \bar{k} .
- (d) Argue that either $\{x_k\}$ converges to 0 in a finite number of iterations, or it must converge sublinearly.