Exercises: Coordinate descent

Pierre Ablin

1 Coordinate descent

Coordinate descent is another optimization method. We let $f : \mathbb{R}^p \to \mathbb{R}$. Coordinate descent tries to minimize f alternatively with respect to individual coordinates.

We denote w^t the iterates. At iteration t, we chose an index $i \in \{1, \dots, p\}$ and try to minimize f with respect to w_i^t without changing the other coordinates w_j^t, $j \neq i$. More formally, we define $\phi_i(x, w) = f(w_1, \dots, w_{i-1}, x, w_{i+1}, w_p)$ and set at each iteration:

$$w_i^{t+1} = \arg\min_x \phi_i(x, w^t) \text{ and } w_j^{t+1} = w_j^t \text{ for } j \neq i$$

The index i is typically chosen as cyclic : $i = 1 + (t \mod p)$. Therefore , at iteration 1, the coordinate 1 is updated, at iteration 2, the coordinate 2 is updated, ... , at iteration p the coordinate p is updated and at iteartion $p + 1$ the coordinate 1 is modified again.

1.1

Assume that f is the quadratic function:

$$f(w) = \frac{1}{2}\langle w, Aw\rangle - \langle b, w\rangle$$

Compute the update rule to minimize ϕ_i.

1.2

At iteration $t + 1$, we update the coordinate i. Demonstrate that

$$f(w^{t+1}) - f(w^t) = -\frac{(Aw^t - b)_i^2}{2A_{ii}} \leq -\frac{(Aw^t - b)_i^2}{2A_{\max}}$$

where $A_{max} = \max_i A_{ii}$

1.3

At iteration t, the coordinate that is updated is i such that $(Aw^t - b)_i^2$ is maximal. Show that

$$f(w^{t+1}) - f(w^t) \leq -\frac{\|Aw^t - b\|^2}{2pA_{\max}}$$

1.4

Let $w^* = A^{-1}b$. Demonstrate that $\|Aw - b\|^2 \geq 2\sigma_{\min}(A)(f(w) - f(w^*))$.

Provide a convergence rate for the coordinate descent method. What is the difference with gradient descent ? When is it faster, or slower? Hint: what is the link between A_{max} and $\sigma_{\max}(A)$?

1.5

Now, derive a coordinate descent rule for the Lasso problem.

$$\min_w \frac{1}{2}\|Xw - y\|^2 + \lambda\|w\|_1$$