Derivative Free Optimization

Solution 1: Coordinate Descent I

$$\begin{aligned}
\mathcal{R}_{\text{emp}}(\boldsymbol{\theta}) &= \frac{1}{2}\|\mathbf{X}\boldsymbol{\theta} - \mathbf{y}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2 = \frac{1}{2}\mathbf{y}^\top\mathbf{y} - \mathbf{y}^\top\mathbf{X}\boldsymbol{\theta} + \frac{1}{2}\boldsymbol{\theta}^\top\boldsymbol{\theta} + \frac{\lambda}{2}\boldsymbol{\theta}^\top\boldsymbol{\theta} \\
&= \frac{1}{2}\mathbf{y}^\top\mathbf{y} - \sum_{j=1}^{d}\mathbf{y}^\top\mathbf{x}_j\theta_j + \frac{1}{2}(1+\lambda)\boldsymbol{\theta}^\top\boldsymbol{\theta} \\
\frac{\partial\mathcal{R}_{\text{emp}}}{\theta_j} &= (1+\lambda)\theta_j - \mathbf{y}^\top\mathbf{x}_j \overset{!}{=} 0 \\
&\Rightarrow \theta_j^* = \frac{\mathbf{y}^\top\mathbf{x}_j}{1+\lambda}
\end{aligned}$$

Solution 2: Coordinate Descent II

(a) Update x_1 whle fixing x_2: We fix $x_2 = c$ (constant). The function then states as

$$g(x_1, c) = |x_1 - c| + 0.1(x_1 + c).$$

We want to choose x_1 to minimize this. Due to the absolute value, there are two cases.

(i) Case 1: $x_1 \geq c$: Then $|x_1 - c| = x_1 - c$. So $g(x_1, c) = (x_1 - c) + 0.1x_1 + 0.1c = 1.1x_1 - 0.9c$. As a function of x_1 this is strictly increasing (derivative of $1.1 > 0$). Therefore, the minimizer given $x_1 \geq c$ is at the left boundary, i.e., $x_1 = c$.

(ii) Case 2: $x_1 < c$: Then $|x_1 - c| = c - x_1$. So $g(x_1, c) = (c - x_1) + 0.1x_1 + 0.1c = 1.1c - 0.9x_1$. As a function of x_1 this is strictly decreasing (derivative of $-0.9 < 0$). Thereore, the minimizer given $x_1 < c$ is at the right boundary, i.e., $x_1 = c$.

In both cases, the best choice is $x_1^* = c$. Note that x_2 was fixed to be c, i.e., the function is minimized exactly when $x_1 = x_2$.
After updating x_1 while holding x_2 constant, we arrive at $(x_1^{[1]}, x_2^{[0]}) = (x_2^{[0]}, x_2^{[0]})$.

Update x_2 while fixing x_1: We now fix $x_1 = c$ (constant). The function then states as

$$g(c, x_2) = |c - x_2| + 0.1(c + x_2).$$

Note that g is symmetric in its arguments, therefore based on the first analysis, we conclude that again $x_2^* = c$.
After updating x_2 while holding x_1 constant, we arrive at $(x_1^{[1]}, x_2^{[1]}) = (x_1^{[1]}, x_1^{[1]}) = (x_2^{[0]}, x_2^{[0]})$.

We observe that coordinate updates will set the respective coordinate to the value of the other constant held coordinate value and once the algorithm arrives at $x_1 = x_2 = c$, neither coordinate update will move the point.

(b) Along the diagonal $x_1 = x_2 = t$, the function simplifies to

$$g(t, t) = |t - t| + 0.1(t + t) = 0.2t.$$

As $t \to -\infty$, $0.2t \to -\infty$, hence the infimum of g is $-\infty$. No finite (x_1, x_2) can achieve that inifimum, i.e., there is no global mnimizer, but the values of g can be made arbitrarly negative by letting x_1, x_2 be arbitrarly negative.