

### Solution 1: Lasso, Subdifferentials

- (a) The Taylor approximation of first order of a function  $f(x)$  at point  $x_0$  is

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

On the other hand, a differentiable function  $f$  is said to be convex on an interval  $\mathcal{I}$  if and only if

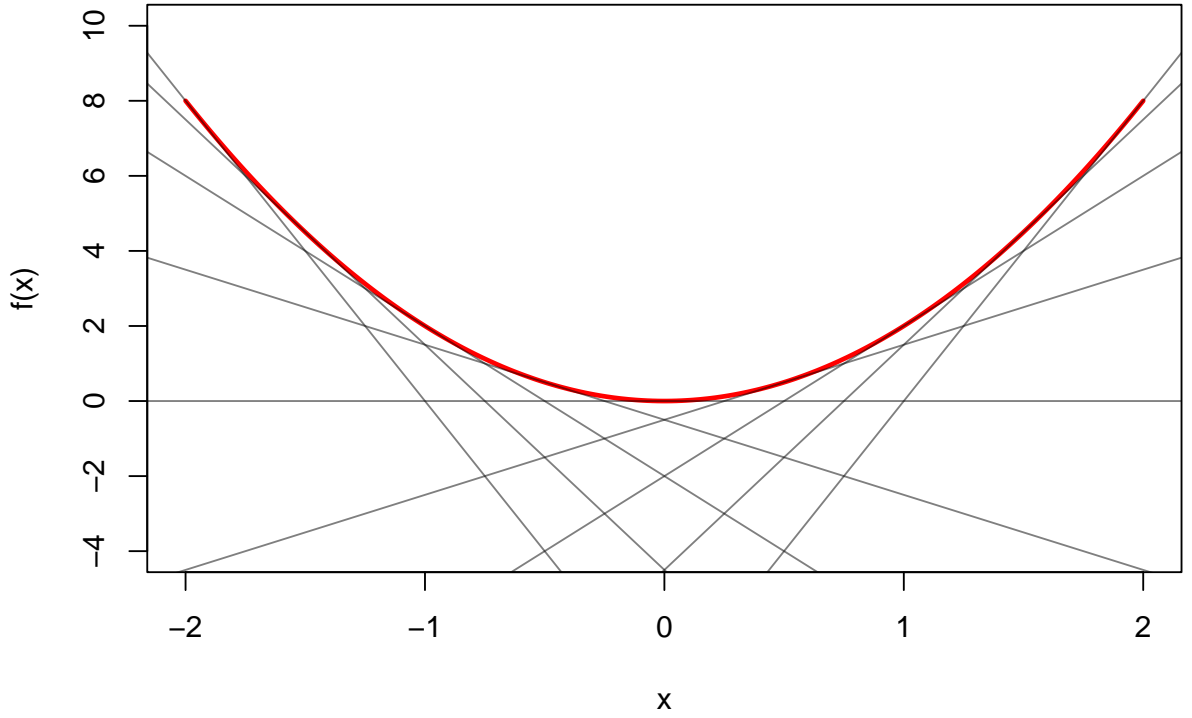
$$f(x) \geq f(x_0) + f'(x_0)(x - x_0)$$

for all points  $x, x_0 \in \mathcal{I}$ .

- (i) If we approximate a convex function with a Taylor approximation of first order, we will always get a lower bound at the given point as the second equation states.
- (ii) Visualization of such an approximation for  $2x^2$  on  $\mathcal{I} = [-2, 2]$  (we will only later see how to calculate a derivative(-like) measure for the non-differentiable functions). The approximation in this case is  $f(x) \approx 2x_0^2 + 4x_0(x - x_0) = -2x_0^2 + 4x_0x$ . We can plot this for several values of  $x$ :

```
xx <- seq(-2, 2, by = 0.01)
yy <- 2*xx^2
# this will give us the approximation function for x=0
# and what happens if we vary x (its slope)
# for given x0
approx_fun <- function(x0) c(-2*x0^2, 4*x0)

plot(xx, yy, type = "l", xlab = "x", ylab = "f(x)", ylim=c(-4,10), col = "red", lwd=2.5)
for(x0 in seq(-2,2,by=0.5))
  abline(approx_fun(x0), col = rgb(0,0,0,0.5))
```



- (b) A subdifferential of  $f$  is a set of values  $\check{\nabla}_{x_0} f$  defined as

$$\check{\nabla}_{x_0} f = \{g : f(x) \geq f(x_0) + g \cdot (x - x_0) \forall x \in \mathcal{I}\}.$$

Every scalar value  $g \in \check{\nabla}_{x_0}$  is said to be a subgradient of  $f$ . A subdifferential thus generalizes the idea of a lower approximation from before by replacing  $f'(x_0)$  with any constant  $g$  for which the approximation is still strictly below the objective function  $f$ .

- (c) We can make use of subdifferentials for convex but non-differentiable loss functions like the one induced by the Lasso, because we are now not restricted to cases where we can compute  $f'(x_0)$ . It holds that:

A point  $x_0$  is the global minimum of a convex function  $f \Leftrightarrow 0$  is contained in the subdifferential  $\check{\nabla}_{x_0} f$ .

We can define a subdifferential at point  $x_0$  also as a non-empty interval  $[x_l, x_u]$  where the lower and upper limit is defined by

$$x_l = \lim_{x \rightarrow x_0^-} \frac{f(x) - f(x_0)}{x - x_0}, \quad x_u = \lim_{x \rightarrow x_0^+} \frac{f(x) - f(x_0)}{x - x_0}.$$

These resemble the limits of the derivative  $\partial f / \partial x$  evaluated at a point very close to  $x_0$  when coming from the left or right side, respectively.

- (i) In the case for  $f(x) = |x|$ ,  $\lim_{x \rightarrow 0^\pm} |x|/x = \pm 1$  and thus  $\check{\nabla}_{x_0} f = [-1, 1]$  at  $x_0 = 0$ .
- (ii)  $x_0$  is a global minimum as  $0 \in \check{\nabla}_{x_0} f$
- (iii) The  $L1$  penalty has no derivative at  $\theta_k = 0$  for all  $\theta_k$  with  $k \in \{1, \dots, p\}$ . Thus we are particularly interested in the subdifferential at this point, which is

$$\check{\nabla}_{\theta_k} \lambda \sum_{j=1}^p |\theta_j| = \sum_{j=1}^p \check{\nabla}_{\theta_k} \lambda |\theta_j| = \check{\nabla}_{\theta_k} \lambda |\theta_k| = [-\lambda, \lambda],$$

where in the second equation we use that the subdifferential of a constant function is zero. For a (sub-) gradient at any other differentiable point, we get the conventional gradient using the given hint, which is  $-\lambda$  for  $\theta_k < 0$  and  $\lambda$  for  $\theta_k > 0$ .

- (d) The subdifferential for the Lasso w.r.t.  $\theta_2$  is then simply the combination of the standard gradient for the unregularized risk  $\nabla_{emp} := n^{-1} \sum_{i=1}^n -2x_2^{(i)}(y^{(i)} - x_1^{(i)}\theta_1 - x_2^{(i)}\theta_2)$  plus the subdifferential for the penalty:

$$\check{\nabla}_{\theta_2} \mathcal{R}_{reg} = \begin{cases} \nabla_{emp} - \lambda & \text{if } \theta_2 < 0 \\ [\nabla_{emp} - \lambda, \nabla_{emp} + \lambda] & \text{if } \theta_2 = 0 \\ \nabla_{emp} + \lambda & \text{if } \theta_2 > 0. \end{cases}$$