

## April 18th

Why do we not look at  $\lambda$  as another Lagrange multiplier?

recall that weird thing where we often optimise with respect to both  $x$  and  $y$  while they're deterministically related

confused about the mix of asymptotics and computational concerns (like we need  $p$  not to grow fast enough relative to  $N$  in order for the optimisation to work ??)

subsetting to estimate the gradient – to analyse properties need to delve into stochastic optimisation?

what's the name of that theorem about the rate of growth being the largest in the direction specified by the gradient

Does the problem being convex exclude the existence of any extrema except the global min?

Does smoothness correspond to some property of the dgp (joint distribution of  $X$  and  $u$ )? Discreteness of covariates will probably mess this ? Is the objective function smooth in general (does the norm affect things badly)?