

# HW5

Yilun Zhang

November 9, 2015

## Exercise 1

**a**

Let  $S_k = \{x : \|x\|^2 = 1, e'_i x = 0, i = 1, 2, \dots, k\}$ ,  $k = 1, 2, \dots, n-2$ , and  $S_0 = \{x : \|x\|^2 = 1\}$ . Clearly  $S_0 \supseteq S_1 \cdots \supseteq S_{n-1}$ . Then

$$\lambda_1 = \min_{x \in S_0} x' Q x$$

...

$$\lambda_i = \min_{x \in S_{i-1}} x' Q x$$

...

$$\lambda_n = \min_{x \in S_{n-1}} x' Q x$$

Then obviously we have  $\lambda_1 \leq \lambda_2 \cdots \leq \lambda_n$ . Because when  $i < j$ , the constraint set  $S_j$  is subspace of  $S_i$ ,  $\lambda_i \geq \lambda_j$ .

**b**

By the definition of  $e_i$ , we have  $e'_i e_j = 0$  for  $i \neq j$  and  $\|e_i\|^2 = 1$ . If we have

$$\sum_{i=1}^n c_i e_i = 0$$

for  $c_i \in \mathbb{R}$ . For  $k = 1, 2, \dots, n$

$$e'_k \sum_{i=1}^n c_i e_i = c_k ||e_k||^2 = c_k = 0$$

This implies  $c_k = 0$ . Therefore,  $e_1, e_2, \dots, e_n$  are linearly independent.

**c**

For certain  $k$ , the optimization problem is to minimize  $x'Qx$  subject to  $x'x = 1; e'_i x = 0, i = 1, 2, \dots, k$  the Lagrangian function is

$$L(x, \lambda^*) = x'Qx + \lambda_1^*(1 - x'x) + \sum_{j=2}^k \lambda_j^* e'_j x$$

$$\nabla_x L(x, \lambda^*) = 2Qx - 2\lambda_1^* x + \sum_{j=2}^k \lambda_j^* e_j = 0$$

$$e'_k \nabla_x L(e_k, \lambda^*) = 2e'_k Q e_k - 2\lambda_1^* = 2\lambda_k - 2\lambda_1^* = 0$$

Then  $\lambda_k = \lambda_1^*$  and

$$\lambda_k = L(e_k, \lambda^*) = \lambda_k + \sum_{j=2}^k \lambda_j^* e'_j x \Rightarrow \lambda_j^* = 0 \text{ for } j = 2, 3, \dots, k$$

Then

$$\nabla_x L(e_k, \lambda^*) = 2Qe_k - 2\lambda_k e_k = 0$$

$Qe_k = \lambda_k e_k$ ,  $\lambda_k$  is the eigen value and  $e_k$  is the eigen vector.

## Exercise 2

$h(x) = 0 \Leftrightarrow ||h(x)||^2 = 0$ . If there is a Lagrangian multiplier, the gradient w.r.t  $x$  is

$$\nabla_x L(x, \lambda) = \nabla f(x) + 2\nabla h(x)h(x) = 0$$

But

$$\nabla_x L(x^*, \lambda) = \nabla f(x^*) + 2\nabla h(x^*)h(x^*) = \nabla f(x^*)$$

is not 0. So there is no Lagrangian multiplier.

## Exercise 3

**a**

The optimization problem is  
maximize(or minimize)

$$x_1 + x_2$$

subject to

$$x_1^2 + x_2^2 = 2 + u$$

Let  $x_1 = \sqrt{u+2}\cos\theta$  and  $x_2 = \sqrt{u+2}\sin\theta$ . Then our problem is to maximize(or minimize)  $\sqrt{u+2}\cos\theta + \sqrt{u+2}\sin\theta$ . It's easy to see when  $x_1 = x_2 = \frac{\sqrt{2}}{2}\sqrt{u+2}$ , the objective function has maximizer  $\sqrt{2}\sqrt{u+2}$ ; when  $x_1 = x_2 = -\frac{\sqrt{2}}{2}\sqrt{u+2}$ , the objective function has minimizer  $-\sqrt{2}\sqrt{u+2}$ , which are also primal function of maximum and minimum.

**b**

For minimization, the Lagrangian function is

$$L(x_1, x_2, \lambda) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2 - u)$$

$$\nabla_{x_1} L(x_1, x_2, \lambda) = 1 + 2\lambda x_1$$

$$\nabla_{x_2} L(x_1, x_2, \lambda) = 1 + 2\lambda x_2$$

$$\nabla_{\lambda} L(x_1, x_2, \lambda) = x_1^2 + x_2^2 - 2 - u$$

we have  $\lambda = -\frac{1}{2x_1} = -\frac{1}{2x_2}$ , plug in the optimal  $x_1, x_2$ ,  $\lambda(u) = \frac{\sqrt{2}}{2}(u+2)^{-1/2}$ . And  $p'(u) = -\frac{\sqrt{2}}{2}(u+2)^{-1/2} = -\lambda(u)$ . QED

## Exercise 4

The matlab code of this implementation is

```

function [x, it] = Alang(i, x0, lam0, rho0, quas, maxit, tol, c)
it=0;
[fc, gc, hc]=prob(i,x0,lam0,rho0);
x=x0;
lam=lam0;
rho=rho0;
while (norm(gc) > tol & it <= maxit)
x=newton(x, @prob, i, 10000, 1e-6, quas,1e-4,lam, rho );
[fc,gc,hc]=prob(i,x , lam, rho)
if(norm(c(x,i))<1e-3)
lam=lam-rho'*c(x,i);
else
rho=rho*2;
end
it=it+1;
end
end

```

All of pure newton and BFGS quasi newton converge to the true value for problem 6-9. For all problem, BFGS is a little bit faster than pure newton. Since all solver gives corrected result, a table showing result is not provided.

# 1 Proposal of Final Project: compare LASSO solvers

Tibshirani [2] introduces LASSO regression 1996, which can estimating the coefficients and select variables at the same time. Instead of ordinary least square regression, the lasso

$$\begin{aligned} & \text{minimize}_{\beta} \quad ||X\beta - Y||_2^2 \\ & \text{subject to} \quad |\beta|_1 < \tau \end{aligned}$$

This can be rewritten as a unconstrained convex optimization problem:

$$\text{minimize}_{\beta} \quad ||X\beta - Y||_2^2 + \lambda|\beta|_1$$

This can be solved by a BCL(Bound-Constrained Lagrangian Method).

But LASSO wasn't popular when it is invented. Because this method is computational intensive and the computer that time didn't have power as in today.

Long with development of computer, new methods are invented to solve lasso problem. The most famous one is proposed by Efron et al[1]: the least angle regression.

I plan to compare these methods first on linear regression and maybe implement them on generalized linear regression model(GLM).

## References

- [1] Bradley Efron et al. "Least angle regression". In: *The Annals of statistics* 32.2 (2004), pp. 407–499.
- [2] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), pp. 267–288.