Math Problem Set # 6 *

Suleyman Gozen

Exercises from **Chapter 9** at "Humpherys, Jeffrey and Tyler J. Jarvis, Foundations of Applied Mathematics, Volume II: Algorithm Design and Optimization, forthcoming. SIAM, Philadelphia, PA."

Ex 9.1

The proof is trivial.

Ex 9.2

Proof. Note that A^TA is symmetric and positive definite. Therefore, as long as x satisfies the first order necessary condition, it is global minimizer of the objective function, $x^TA^TAx - 2b^TAx$. Then, the first order condition is

$$2A^TAx - 2A^Tb = 0$$

This results in $A^TAx = A^Tb$.

Ex 9.3

Explanation. The general description of the algorithm is problem-specific. For example, if the the Hessian matrix, $D^2f(x)$ is sparse, we can still calculate the Hessian with less numerical error, and handle this even with large dimension using lexicographic orderings in n-dimensions. Also, in multi-dimensional case, Conjugate Gradient method may show terrible computing speed depending on the sparseness of matrix. In the other case, Broyden or BFGS method does not converge in calculating the policy functions of heterogeneous model with borrowing constraints, because the approximating the gradient or Hessian near the kinked point generally produce inaccurate values which produces another massive numerical error during the entire algorithm. In this case, calculating the policy functions through value function iteration and seeing the shapes of calculated policy functions gives intuition on which point the coder should not be taking derivatives. It is hard to say which method is the dominant over the others, but in my case, analytical derivatives(Jacobian) of Euler equations were useful. So far, it works nicely with heterogenous agent models with more than 5 heterogeneities.

However, taking numerical derivatives is a bit tricky because which variables should be derivatized depend on the assumptions of the problem and algorithm. For example, if we are solving the Aiyagari model, we should not be calculating the derivative of aggregate capital such as $\partial K_t/\partial k_t(k_{t-1}, \varepsilon)$ because we assume during the algorithm that the aggregate variables , w_t and r_t(wage and interest rate) are fixed, and we are trying to solve the policy functions with these fixed values. Later, we find the joint distribution using this policy

*I thank Yung-Hsu Tsui and Jaehyung Kim for their valuable comments. The answers are heavily borrowed from Jaehyung Kim's work.

function, and update w and r. But, on the other hand, in Brock-Mirman or RBC type representative household models, the coder have to use $\partial K_t/\partial k_t(k_{t-1})$ because we assume that the amount of aggregate capital depends solely on the representative's capital tomorrow.

Automatic differentiation might help the codes to see in one page, but designing the chain rules of all the potential differentiable variables is exactly the same task with just calculating the analytical derivatives of Euler equations.

Ex 9.4

Proof.
Let D is the derivative of $f(x)^T$, and λ is the eigenvalue of Q with $Df(x_0)^T = Qx_0 - b$. Then,

$$\begin{aligned} x_1 =& x_0 - \frac{DD^T}{DQD^T}D^T \\ =& x_0 - \frac{DD^T}{D\lambda D^T}D^T \\ =& x_0 - \frac{1}{\lambda}D^T \\ =& x_0 - Q^{-1}D^T \\ =& x_0 - Q^{-1}(Qx_0 - b) \\ =& Q^{-1}b \end{aligned}$$

Ex 9.5

Proof. Note that the updating rule of the Gradient Method results in $f(x_{k+1}) = f(x_k + \alpha_k \Delta f(x_k))$. Note that x_{k+1} is obtained by minimizing $f(x_{k+1})$ w.r.t α_k. Then,

$$\frac{df(x_{k+1})}{d\alpha_k} = \Delta f(x_{k+1})\Delta f(x_k) = 0$$

Note that the updating rules of x_k indicates that

$$x_{k+1} - x_k = -\alpha_k \Delta f(x_k)$$

As long as $\Delta f(x_k)$ and $\Delta f(x_{k+1})$ are orthogonal to each other, $x_{k+1} - x_k$ and $x_{k+2} - x_{k+1}$ are orthogonal.

Ex 9.6-9.9
See the Julia code.

Ex 9.10

Proof.

Consider the quadratic function $f(x) = \frac{1}{2}x^T Qx - b^T x$ where $Q \in M_n(\mathbb{R})$ is symmetric and positive definite and $b \in \mathbb{R}^n$. Consider an initial guess $x_0 \in \mathbb{R}^n$. The derivative is $Df = Qx - b$ and the second derivative is $D^2 f = Q$. The minimum is achieved when $Df = Qx - b = 0$. Apply Newton's method and consider $x_1 = x_0 - Q^{-1}Df(x_0) = x_0 - Q^{-1}(Qx_0 - b)$. This implies that $Qx_1 = b$ and that $Qx_1 - b = 0$ which means x_1 is the minimum so the minimum is achieved after 1 iteration.

Ex 9.12

Proof.

$$\begin{aligned} Ax =& \lambda x \\ (B - \mu I)x =& \lambda x \\ Bx =& (\lambda + \mu)x \end{aligned}$$

Ex 9.15

Proof.

$$\begin{aligned} & (A + BCD)\left[A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1}\right] \\ &= \left\{I - B\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1}\right\} + \left\{BCDA^{-1} - BCDA^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1}\right\} \\ &= \left\{I + BCDA^{-1}\right\} - \left\{B\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1} + BCDA^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1}\right\} \\ &= I + BCDA^{-1} - \left(B + BCDA^{-1}B\right)\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1} \\ &= I + BCDA^{-1} - BC\left(C^{-1} + DA^{-1}B\right)\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1} \\ &= I + BCDA^{-1} - BCDA^{-1} \\ &= I. \end{aligned}$$

Ex 9.16

Proof.

Substituting

$$x = \frac{v - As}{\|s\|^2}, y = s$$

into the Sherman-Morrison-Woodbury formula yields

$$\begin{aligned} A^{-1} - \frac{A^{-1}xy^T A^{-1}}{1 + y^T A^{-1}x} =& A^{-1} - \frac{A^{-1}(v - As)/\|s\|^2 s^T A^{-1}}{1 + s^T A^{-1}(v - As)/\|s\|^2} \\ =& A^{-1} - \frac{(A^{-1}v - s)s^T A^{-1}/\|s\|^2}{1 + s^T A^{-1}v/\|s\|^2 - 1} \\ =& A^{-1} + \frac{(s - A^{-1}v)s^T A^{-1}}{s^T A^{-1}v} \end{aligned}$$

Ex 9.18

Proof.
Note that

$$\begin{aligned}\phi_k(\alpha) =& f(x_k + \alpha_k d_k)\\ =& \frac{1}{2}x_k^T Q x_k + \alpha_k (d^k)^T Q x_k + \frac{\alpha_k^2}{2}(d^k)^T Q d^k - x_k^T b - \alpha_k (d^k)^T b\end{aligned}$$

As long as Q is positive definite,i,e, $(d^k)^T Q d^k$ is positive, we can find minimizer of this function with respect to α_k. Thus, taking derivative results in;

$$\begin{aligned}0 =& \frac{\partial \phi_k(\alpha_k)}{\partial \alpha_k}\\ =& \alpha_k (d^k)^T Q d^k - (d^k)^T b + (d^k)^T Q x_k\end{aligned}$$

Then,

$$\begin{aligned}\alpha_k (d^k)^T Q d^k =& (d^k)^T b + (d^k)^T Q x_k\\ =& (d^k)^T (b - Q x_k)\\ =& (d^k)^T r_k\end{aligned}$$

Thus, dividing $(d^k)^T Q d^k$ on both sides results in the answer.

Ex 9.20

Proof.

Define the basis $W_i := \{\tilde{r}^0, \tilde{r}^0, ..., \tilde{r}^{i-1}\}$ where $\tilde{r}^k := b - Qx^k$. By applying Gram-Schmit process, we can construct

$$r^k := \tilde{r}^k - \sum_{j=0}^{k-1} \frac{(r^j, \tilde{r}^k)_Q}{\|r^j\|_A^2} r^j$$

Note that $W_i = span\{r^0, ..., r^{i-1}\}$ because every r^i can be written as a linear combination of $\tilde{r}^0, ..., \tilde{r}^i$. Also, note that with this setting we can deduce that the Conjugate Gradient method in each iteration is solving the following minimization problem;

$$\min_{x \in x^0 + W_i} f(x)$$

where the minimizer of this problem is x^i. Thus, the following problem results in $t = 0$ such that

$$\min_t f(x^i + t\tilde{r}^j)$$

with $j < i$. Thus, the first order condition of this problem is

$$Df(x^i)\tilde{r}^j = 0$$

implying the following;

$$\begin{aligned}
0 =& Df(x^i)\tilde{r}^j \\
=& (Qx^i - b)^T\tilde{r}^j \\
=& -\tilde{r}^i\tilde{r}^j \quad \forall j < i
\end{aligned}$$