

## CAAM 551 Problem set 2

Due Fri. 23 Sep. 2011

### Problem 1

#### Least Squares Regularization

Given an  $m \times n$  matrix  $\mathbf{A}$  with  $m > n$  and a positive number  $\lambda$ , a *regularized* least squares solution  $\mathbf{x}_\lambda$  may be computed by solving

$$\min \left\| \begin{pmatrix} \mathbf{A} \\ \mu \mathbf{I} \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix} \right\|_2,$$

where  $\mu = \sqrt{\lambda}$ .

- Derive the normal equations for the *regularized* least squares problem given above.
- Show  $\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I}$  is symmetric and positive definite for every positive value of  $\lambda$ . Prove that the regularized least squares solution  $x_\lambda$  is unique for every positive value of  $\lambda$ .
- Use the Singular Value Decomposition of  $A$  to express the solution  $x_\lambda$  to the problem

$$\min \left\| \begin{pmatrix} \mathbf{A} \\ \mu \mathbf{I} \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix} \right\|_2$$

where  $\mathbf{b} \in \mathbf{R}^m$  and  $\mu^2 = \lambda$ .

- Prove that  $\lim_{\lambda \rightarrow 0^+} \mathbf{x}_\lambda = \mathbf{x}_{LS}$  the minimum norm least squares solution (Regardless of the rank of  $\mathbf{A}$ ).

## Problem 2

### Implementation: Least Squares Regularization

The following exercises will lead you through a procedure for choosing a regularization parameter via the trust region subproblem

$$\min\{\|\mathbf{b} - \mathbf{A}\mathbf{x}\| : \|\mathbf{x}\| \leq \Delta\}$$

where  $\Delta > 0$  is a bound on the norm of the solution  $\mathbf{x}$ . (Here  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ ,  $\mathbf{x} \in \mathbb{R}^n$  with  $m \geq n$ ) In all cases the norm  $\|\cdot\|$  is the 2-norm.

In class we established that this problem could be solved by numerically computing the solution to

$$\text{Compute } \lambda \geq 0 \text{ s.t. } \frac{1}{\|\mathbf{x}_\lambda\|} = \frac{1}{\Delta}$$

where  $\mathbf{x}_\lambda \equiv (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}$ . (unless  $\|\mathbf{x}_0\| < \Delta$  and the solution is then  $\mathbf{x}_0$  with  $\lambda = 0$ )

We shall first work with a test problem using the following generating code

```
m = 100; n = 4;
s = [.5 , .8, 1 , 1.5];
S = diag(s);
[U,R] = qr(randn(m,n), 0); [V,R] = qr(randn(n),0);
%
% generate A with singular values
% specified by s
%
A = U*S*V';
x = ones(n,1);
b = A*x;
delta = norm(x)/(1.2);
```

- i) Write a Matlab code to graph  $\phi(\lambda) \equiv \frac{1}{\|\mathbf{x}_\lambda\|}$  over the interval  $\lambda \in [-2.5, 1]$ . On the same plot (using hold on) plot the horizontal line  $y(\lambda) = \frac{1}{\Delta}$  over the same interval and also plot the points  $(-s(j)^2, 0)$  indicated by (red) \* and the point  $(0, \frac{1}{\|\mathbf{x}_0\|})$  indicated by a (green) o .

Plot this several times and observe the graph. What do you notice? Why does the curve change? Please turn in one of the graphs and also the Matlab code with this HW.

- ii) Let  $[Q, R] = \text{qr}(A, 0)$ ; i.e.  $A = QR$  (short form), and let  $c = [b^T Q, 0]^T$ . Prove that  $x_\lambda$  is the solution to the following least squares problem

$$\min \left\| \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix} x_\lambda - \begin{bmatrix} c \\ 0 \end{bmatrix} \right\|.$$

- iii) Let

$$Q_\lambda R_\lambda = \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix}$$

be the short form QR-factorization. Give a formula for computing  $x_\lambda$  for any non-negative value of  $\lambda$  that involves  $Q_\lambda, R_\lambda$  and  $c$ . Please discuss the importance of this formula if one wishes to compute  $x_\lambda$  for many values of  $\lambda$  when  $m \gg n$  (e.g. think  $m = 100,000$  and  $n = 100$ ). Please include a (somewhat) detailed discussion of the cost of this approach vs the alternative of using the SVD.

- iv) Prove that the following Matlab code segment will implement one step of Newton's method for solving  $\phi(\lambda) = \frac{1}{\Delta}$ :

```
M = [R ; sqrt(lamda)*I];
[Q1,R1] = qr(M,0);
x = (R1)\(Q1'*c);
w = (R1')\x;
psi = (norm(x) - delta)/delta;
dlamda = psi*(x'*x)/(w'*w);
lambda = lambda + dlambda;
```

( assume  $c$  and  $R$  are as defined above and that  $I = \text{eye}(n)$ ).

- v) Please use the code segment in item *iv* above to implement Newton's method for solving  $\phi(\lambda) = \frac{1}{\Delta}$ . Your function should have the following specification:

```
function [x, lambda] = minphi(A,b,delta,tol);
```

It should accept the input  $\mathbf{A} = \mathbf{A}, \mathbf{b} = \mathbf{b}, \text{delta} = \Delta$  and `tol` (a user specified tolerance). It should return the solution  $\mathbf{x} = \mathbf{x}_\lambda$  and the value `lambda` =  $\lambda$  with

$$\left| \frac{\|\mathbf{x}_\lambda\| - \Delta}{\Delta} \right| \leq \text{tol}.$$

Please include a listing of your code with this HW.

- vi) Run your code on the test problem defined above with `tol = 10*eps` and report the values `lambda` and `dlambda` side by side for each of the Newton iterates (This should be about 5 iterates if things are working properly). Please display the numbers for `lambda` in format `long e`.

What do you notice about the iterates? Is the expected quadratic rate of convergence obtained?

- vii) Repeat the computation in items *i* and *vi* with `s = [.01 , .8, 1 , 1.5]` and then again with `s = [.001 , .8, 1 , 1.5]`. What happens to the iterates ? Explain what you observe.
- viii) In all cases, the iterates (starting with  $\lambda_1 = 0$  are monotone increasing:  $\lambda_1 < \lambda_2 < \dots$  (at least they should be). Why does this always happen? Do you ever need *safeguarding* to assure convergence and termination of the iteration?

### Problem 3

The regularization procedure can be generalized to use a weighted norm:

$$\min \|b - Ax\|_2 \quad \text{s.t.} \quad \|Wx\|_2 \leq \Delta$$

we assume  $W \in \mathbf{R}^{n \times n}$  is nonsingular. Set up the Lagrangian for this problem and derive the optimality conditions (i.e. the parameterized normal equations). Explain how to use the QR decomposition (similar to Problem 1) in finding the optimal  $\lambda$  for this formulation.

#### Problem 4

- a) Provide a numerical solution to the integral equation

$$\int_0^1 (s^2 + t^2)^{1/2} u(s) ds = \frac{(t^2 + 1)^{3/2} - t^3}{3}$$

on the interval  $[0, 1]$  using the composite Trapezoidal quadrature rule to discretize the integral. Use  $n + 1$  equally spaced points  $s_j = t_j = jh$  for  $j = 0 : n$  with  $h = 1/n$  for both the  $s$ - and  $t$ - intervals. Use “backslash” to solve the resulting linear systems  $Au = f$ . Write a code for general  $n$ . Specific values will be provided in class.

- b) Repeat part (a), this time solving the linear system using the SVD. Regularize the solution using SVD truncation (omit all singular values below a threshold). Try various truncation thresholds and compare you solution with the known true solution. Provide a graph of the comparison for the best approximate solution and give the corresponding threshold and truncation index (the last singular value retained).
- c) Repeat items a) and b) above using composite Simpson’s rule in place of the Trapezoidal rule. Explain what you see.