

## Homework 2: Linear Systems

CS 205A: Mathematical Methods for Robotics, Vision, and Graphics (Fall 2013),  
Stanford University

Due Monday, October 14, at midnight

**Problem 1** (25 points). For this problem, assume that the matrix norm  $\|A\|$  for  $A \in \mathbb{R}^{n \times n}$  is induced by a vector norm  $\|\vec{v}\|$  for  $\vec{v} \in \mathbb{R}^n$  (but it may be the case that  $\|\cdot\| \neq \|\cdot\|_2$ ).

- (a) For  $A, B \in \mathbb{R}^{n \times n}$ , show  $\|A + B\| \leq \|A\| + \|B\|$ .
- (b) For  $A, B \in \mathbb{R}^{n \times n}$  and  $\vec{v} \in \mathbb{R}^n$ , show  $\|A\vec{v}\| \leq \|A\|\|\vec{v}\|$  and  $\|AB\| \leq \|A\|\|B\|$ .
- (c) For  $k > 0$  and  $A \in \mathbb{R}^{n \times n}$ , show  $\|A^k\|^{1/k} \geq |\lambda|$  for any real eigenvalue  $\lambda$  of  $A$ .
- (d) For  $A \in \mathbb{R}^{n \times n}$  and  $\|\vec{v}\|_1 \equiv \sum_i |v_i|$ , show  $\|A\|_1 = \max_j \sum_i |a_{ij}|$ .

*Extra credit.* Prove “Gelfand’s formula:”  $\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{1/k}$ , where  $\rho(A) \equiv \max\{|\lambda_i|\}$  for eigenvalues  $\lambda_1, \dots, \lambda_m$  of  $A$ . In fact, this formula holds for any matrix norm  $\|\cdot\|$ .

**Problem 2** (25 points; adapted from CS 205A 2012). In this problem you will derive facts necessary to implement LU factorization of  $A \in \mathbb{R}^{n \times n}$  with partial pivoting.

- (a) Recall that the elimination matrix corresponding to a single step of forward substitution has the form  $E \equiv I + c\vec{e}_\ell\vec{e}_k^\top$ , where  $k < \ell$ . Argue that a full iteration of forward substitution in which row  $k$  is forward-substituted to all rows  $\ell$  with  $\ell > k$  is carried out by a matrix of the form

$$M_k = I - \vec{m}_k\vec{e}_k^\top,$$

where the  $i$ -th value in  $\vec{m}_k$  is zero for all  $i \leq k$ . Also show that the inverse of  $M_k$  is  $L_k \equiv I + \vec{m}_k\vec{e}_k^\top$ .

- (b) Suppose  $P^{(ij)}$  is the permutation matrix swapping rows  $i$  and  $j$ . Show that if  $i, j > k$  then  $L_k P^{(ij)} = P^{(ij)} (I + P^{(ij)} \vec{m}_k \vec{e}_k^\top)$ .
- (c) The matrix  $L$  from running forward substitution to completion with partial pivoting is given by

$$L = P_1 L_1 \cdots P_{n-1} L_{n-1},$$

where each  $P_i$  permutes row  $i$  with row  $i'$  for some  $i' > i$ . Show that  $L$  can be rewritten as

$$L = P_1 \cdots P_{n-1} L_1^p \cdots L_{n-1}^p,$$

where  $L_k^p \equiv I + (P_{n-1} \cdots P_{k+1} \vec{m}_k) \vec{e}_k^\top$ .

- (d) Show that  $L_1^p \cdots L_{n-1}^p$  is lower triangular.

**Problem 3** (25 points; “Mini-Riesz Representation Theorem”). We will say  $\langle \cdot, \cdot \rangle$  is an inner product on  $\mathbb{R}^n$  if it satisfies:

1.  $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle \forall \vec{x}, \vec{y} \in \mathbb{R}^n$
2.  $\langle \alpha \vec{x}, \vec{y} \rangle = \alpha \langle \vec{x}, \vec{y} \rangle \forall \vec{x}, \vec{y} \in \mathbb{R}^n, \alpha \in \mathbb{R}$
3.  $\langle \vec{x} + \vec{y}, \vec{z} \rangle = \langle \vec{x}, \vec{z} \rangle + \langle \vec{y}, \vec{z} \rangle \forall \vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$
4.  $\langle \vec{x}, \vec{x} \rangle \geq 0$  with equality if and only if  $\vec{x} = \vec{0}$ .

Here we will derive a special case of an important theorem applied in geometry processing and machine learning:

- (a) Show that there exists a matrix such that  $\langle \vec{x}, \vec{y} \rangle = \vec{x}^\top A \vec{y}$ . Also show that there exists a matrix  $M$  such that  $\langle \vec{x}, \vec{y} \rangle = (M\vec{x}) \cdot (M\vec{y})$ . [This shows that all inner products are dot products after suitable rotation, stretching, and shearing of  $\mathbb{R}^n$ !]
- (b) A Mahalanobis metric on  $\mathbb{R}^n$  is a distance function of the form  $d(\vec{x}, \vec{y}) = \sqrt{\langle \vec{x} - \vec{y}, \vec{x} - \vec{y} \rangle}$  for inner product  $\langle \cdot, \cdot \rangle$ . Use the result of (a) to give an alternative characterization of Mahalanobis metrics.
- (c) Suppose we are given a series of pairs  $(\vec{x}_i, \vec{y}_i) \in \mathbb{R}^n \times \mathbb{R}^n$ . A typical “metric learning” problem might involve finding a nontrivial Mahalanobis metric such that each  $\vec{x}_i$  is close to each  $\vec{y}_i$  with respect to that metric. Propose an optimization for this problem in the form of HW0 problem 4.

Note: Make sure that your Mahalanobis distance is nonzero, but it is OK if your optimization allows pseudometrics, that is, there can exist  $\vec{x} \neq \vec{y}$  with  $d(\vec{x}, \vec{y}) = 0$ . This problem is open-ended.

**Problem 4** (25 points; “Kernel trick”). In lecture, we covered techniques for linear and nonlinear parametric regression. In this problem, we will develop one least-squares technique for nonparametric regression that is used commonly in machine learning and vision.

- (a) You can think of the least-squares problem as learning the vector  $\vec{a}$  in a function  $f(\vec{x}) = \vec{a} \cdot \vec{x}$  given a number of examples  $\vec{x}^{(1)} \mapsto y^{(1)}, \dots, \vec{x}^{(k)} \mapsto y^{(k)}$  and the assumption  $f(\vec{x}^{(i)}) \approx y^{(i)}$ . Suppose the columns of  $X$  are the vectors  $\vec{x}^{(i)}$  and that  $\vec{y}$  is the vector of values  $y^{(i)}$ . Provide the normal equations for recovering  $\vec{a}$  with Tikhonov regularization.
- (b) Show that  $\vec{a} \in \text{span} \{ \vec{x}^{(1)}, \dots, \vec{x}^{(k)} \}$  in the Tikhonov-regularized system.
- (c) Thus, we can write  $\vec{a} = c_1 \vec{x}^{(1)} + \dots + c_k \vec{x}^{(k)}$ . Give a  $k \times k$  linear system of equations satisfied by  $\vec{c}$  assuming  $X^\top X$  is invertible.
- (d) One way to do nonlinear regression might be to write a function  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and learn  $f_\phi(\vec{x}) = \vec{a} \cdot \phi(\vec{x})$ , where  $\phi$  may be nonlinear. Define  $K(\vec{x}, \vec{y}) = \phi(\vec{x}) \cdot \phi(\vec{y})$ . Assuming we continue to use regularized least squares as in (a), give an alternative form of  $f_\phi$  that can be computed by evaluating  $K$  rather than  $\phi$ . (Hint: What are the elements of  $X^\top X$ ?)

Extra credit. Consider the following formula (from the Fourier transform of the Gaussian):

$$e^{-\pi(s-t)^2} = \int_{-\infty}^{\infty} e^{-\pi x^2} (\sin(2\pi s x) \sin(2\pi t x) + \cos(2\pi s x) \cos(2\pi t x)) dx$$

Suppose we wrote  $K(x, y) = e^{-\pi(x-y)^2}$ . Explain how this “looks like”  $\phi(x) \cdot \phi(y)$  for some  $\phi$ . How does this suggest that the technique from (d) can be generalized?