EE 150 - Applications of Convex Optimization in Signal Processing and Communications
Dr. Andre Tkacenko, JPL
Third Term 2011-2012

Homework Set #2

Due on **Thursday, April 19** in class.

1. (10 points) (*Existence of Moore-Penrose pseudoinverse:*) Recall that if $\mathbf{A} \in \mathbb{C}^{m \times n}$ is some matrix, then a matrix $\mathbf{A}^{\#} \in \mathbb{C}^{n \times m}$ that satisfies the four following conditions is said to be a Moore-Penrose pseudoinverse of $\mathbf{A}$.

1) $\mathbf{A}\mathbf{A}^{\#}\mathbf{A} = \mathbf{A}$.
2) $\mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#} = \mathbf{A}^{\#}$.
3) $\left(\mathbf{A}\mathbf{A}^{\#}\right)^{\dagger} = \mathbf{A}\mathbf{A}^{\#}$.
4) $\left(\mathbf{A}^{\#}\mathbf{A}\right)^{\dagger} = \mathbf{A}^{\#}\mathbf{A}$.

Suppose that $\mathbf{A}$ has rank ρ and has the following singular value decomposition (SVD):

$$\mathbf{A} = \underbrace{\begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0}_{\rho \times (n-\rho)} \\ \mathbf{0}_{(m-\rho) \times \rho} & \mathbf{0}_{(m-\rho) \times (n-\rho)} \end{bmatrix}}_{\mathbf{\Sigma}} \underbrace{\begin{bmatrix} \mathbf{V}_1^{\dagger} \\ \mathbf{V}_2^{\dagger} \end{bmatrix}}_{\mathbf{V}^{\dagger}} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^{\dagger} \,,$$

where $\mathbf{U}_1 \in \mathbb{C}^{m \times \rho}$, $\mathbf{U}_2 \in \mathbb{C}^{m \times (m-\rho)}$, $\mathbf{U} \in \mathbb{C}^{m \times m}$ is unitary, $\mathbf{\Sigma}_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_\rho)$ is a $\rho \times \rho$ diagonal matrix of singular values of $\mathbf{A}$ (where $\sigma_k > 0$ for all $1 \leq k \leq \rho$), $\mathbf{\Sigma} \in \mathbb{R}_{+}^{m \times n}$, $\mathbf{V}_1 \in \mathbb{C}^{n \times \rho}$, $\mathbf{V}_2 \in \mathbb{C}^{n \times (n-\rho)}$, and $\mathbf{V} \in \mathbb{C}^{n \times n}$ is unitary. Show that the matrix $\mathbf{B}$ defined as

$$\mathbf{B} \triangleq \underbrace{\begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix}}_{\mathbf{V}} \underbrace{\begin{bmatrix} \mathbf{\Sigma}_1^{-1} & \mathbf{0}_{\rho \times (m-\rho)} \\ \mathbf{0}_{(n-\rho) \times \rho} & \mathbf{0}_{(n-\rho) \times (m-\rho)} \end{bmatrix}}_{\mathbf{\Sigma}^{\#}} \underbrace{\begin{bmatrix} \mathbf{U}_1^{\dagger} \\ \mathbf{U}_2^{\dagger} \end{bmatrix}}_{\mathbf{U}^{\dagger}} = \mathbf{V}_1 \mathbf{\Sigma}_1^{-1} \mathbf{U}_1^{\dagger} \,,$$

satisfies the four conditions above and is thus a pseudoinverse.

2. (10 points) (*Uniqueness of Moore-Penrose pseudoinverse:*) Let $\mathbf{A}$ be some $m \times n$ matrix and suppose that $\mathbf{B}$ and $\mathbf{C}$ are any two $n \times m$ matrices that satisfy the four conditions mentioned in the previous problem defining a Moore-Penrose pseudoinverse. Show that we always have $\mathbf{B} = \mathbf{C}$.

3. (10 points) (*Maximum likelihood estimate of the covariance matrix of a Gaussian distribution:*) Suppose that we have obtained a sample of n independent, identically distributed (i.i.d.) observations, denoted $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, where $\boldsymbol{x}_k \in \mathbb{R}^{m \times 1}$, drawn from a Gaussian distribution with known mean $\boldsymbol{\mu} \in \mathbb{R}^{m \times 1}$ but unknown covariance $\mathbf{\Sigma} \in \mathbb{S}_{++}^{m}$. This distribution, parameterized by $\mathbf{\Sigma}$ and denoted here by $f(\boldsymbol{x}|\mathbf{\Sigma})$, is given by

$$f(\boldsymbol{x}|\mathbf{\Sigma}) = \frac{1}{(2\pi)^{\frac{m}{2}} \left(\det\left(\mathbf{\Sigma}\right)\right)^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})} \,.$$

Under the i.i.d. assumption, the *log-likelihood* function for the covariance matrix $\mathbf{\Sigma}$, denoted as $\mathcal{L}(\mathbf{\Sigma})$, is given by

$$\mathcal{L}(\mathbf{\Sigma}) = \sum_{k=1}^{n} \log\left(f(\boldsymbol{x}_k|\mathbf{\Sigma})\right) \,.$$

The choice of $\boldsymbol{\Sigma}$ which maximizes $\mathcal{L}(\boldsymbol{\Sigma})$ is the *maximum likelihood (ML) estimate* of $\boldsymbol{\Sigma}$ and will be denoted $\boldsymbol{\Sigma}_{\text{ML}}$.

(a) Calculate $\nabla\mathcal{L}(\boldsymbol{\Sigma})$ *assuming* $\boldsymbol{\Sigma}$ *is symmetric.*

(b) Solve the equation $\nabla\mathcal{L}(\boldsymbol{\Sigma}) = \mathbf{0}$ to find the ML estimate of $\boldsymbol{\Sigma}$. Show that we have

$$\boldsymbol{\Sigma}_{\text{ML}} = \frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^T .$$

Hint: The following identities may be useful here:

$$\begin{aligned}
\frac{d}{d\mathbf{X}}\log(\det(\mathbf{X})) &= \mathbf{X}^{-1} , \\
\frac{d}{d\mathbf{X}}\operatorname{tr}\left(\mathbf{A}\mathbf{X}^{-1}\mathbf{B}\right) &= -\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1} , \\
\frac{d}{d\mathbf{X}_s}f &= \left(\frac{d}{d\mathbf{X}_u}f\right) + \left(\frac{d}{d\mathbf{X}_u}f\right)^T - \operatorname{diag}\left(\frac{d}{d\mathbf{X}_u}f\right) .
\end{aligned}$$

For the first two identities, $\mathbf{X}$ denotes an unstructured square matrix. For the third identity, $\mathbf{X}_s$ denotes the symmetric version of some unstructured matrix $\mathbf{X}_u$.

4. (10 points) (*Complex differential of the pseudoinverse:*) In this problem, we generalize the result that

$$d\left(\mathbf{Z}^{-1}\right) = -\mathbf{Z}^{-1}(d\mathbf{Z})\,\mathbf{Z}^{-1}$$

for complex invertible square matrices to the rectangular case for the pseudoinverse. Specifically, suppose $\mathbf{Z}$ is some $m \times n$ complex matrix. Show that

$$d\left(\mathbf{Z}^{\#}\right) = -\mathbf{Z}^{\#}(d\mathbf{Z})\,\mathbf{Z}^{\#} + \mathbf{Z}^{\#}\left(\mathbf{Z}^{\#}\right)^{\dagger}\left(d\mathbf{Z}^{\dagger}\right)\left(\mathbf{I}_m - \mathbf{Z}\mathbf{Z}^{\#}\right) + \left(\mathbf{I}_n - \mathbf{Z}^{\#}\mathbf{Z}\right)\left(d\mathbf{Z}^{\dagger}\right)\left(\mathbf{Z}^{\#}\right)^{\dagger}\mathbf{Z}^{\#} .$$

Hint: Use the product rule for complex differentials, namely that

$$d(\mathbf{Z}_0\mathbf{Z}_1) = (d\mathbf{Z}_0)\,\mathbf{Z}_1 + \mathbf{Z}_0\,(d\mathbf{Z}_1) ,$$

along with the conjugate transpose rule $d\left(\mathbf{Z}^{\dagger}\right) = (d\mathbf{Z})^{\dagger}$ and the defining properties of the pseudoinverse given in the first problem.

***5.** (30 points) (*Least-squares minimization and minimum norm property of the pseudoinverse:*) In this problem, we consider a slight generalization to the traditional least-squares problem

$$\text{minimize } \xi^2 \triangleq ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 .$$

Suppose that $\mathbf{A} \in \mathbb{C}^{m\times n}$ is a modeling matrix which we would like to fit to a data matrix of observations $\mathbf{B} \in \mathbb{C}^{m\times p}$ by using a linear model of the form $\mathbf{A}\mathbf{X}$, where $\mathbf{X} \in \mathbb{C}^{n\times p}$ is a fitting matrix. To measure the quality of the fit, we will consider the *Frobenius norm* of the error or residual $(\mathbf{A}\mathbf{X} - \mathbf{B})$. In other words, to gauge the quality of the fit, we will consider the objective ξ given by

$$\xi \triangleq ||\mathbf{A}\mathbf{X} - \mathbf{B}||_F .$$

A matrix $\mathbf{X}^{\star} \in \mathbb{C}^{n\times p}$ which minimizes ξ will be called a *least-squares* solution (as it will simultaneously minimize ξ^2 as well).

(a) Show that $\mathbf{X}^\star \triangleq \mathbf{A}^\# \mathbf{B}$ is a least-squares solution by using the trick of *completing the square*. In addition, show that the optimal objective value $\xi^\star$ is given by

$$\xi^\star = \left|\left| \left(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\# \right) \mathbf{B} \right|\right|_F .$$

(b) Suppose now that we are able to get a perfect fit, i.e., that there is at least one solution to the linear system of equations $\mathbf{AX} = \mathbf{B}$. Evidently $\mathbf{X}^\star = \mathbf{A}^\# \mathbf{B}$ is one such solution. Show that any other solution $\mathbf{X}$ to $\mathbf{AX} = \mathbf{B}$ satisfies

$$||\mathbf{X}||_F \geq ||\mathbf{X}^\star||_F ,$$

with equality if and only if $\mathbf{X} = \mathbf{X}^\star$. In other words, $\mathbf{X}^\star$ is the solution to $\mathbf{AX} = \mathbf{B}$ with the smallest Frobenius norm.

Reading assignments:

1. Look over parts of *The Matrix Cookbook* as needed and continue reading the `cvx` Users' Guide.

Reminders:

Late homework policy for EE 150: Late homeworks will not be accepted. There will be no exceptions to this other than institute established emergency reasons, in which case a signed letter is required from an authorized official.

NCT Problems: Remember that problems with an asterisk, such as ***7** are no collaboration type (NCT) problems.

Texts: The abbreviation CO-BV corresponds to the textbook "*Convex Optimization*" by Stephen Boyd and Lieven Vandenberghe. In addition, CO-AE refers to the *Additional Exercises for Convex Optimization*, also by Boyd and Vandenberghe. Finally, CVX corresponds to the `cvx` *Users' Guide* by Michael Grant and Stephen Boyd.
