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

# Homework Set #1

Due on **Thursday, April 12** in class.

**1.** (10 points) Send a brief message with your name in full to the course e-mail address (`ee150.acospc@gmail.com`) in order to get on the mailing list.

**2.** (10 points) (*Young's inequality for products, Hölder's inequality, and Minkowski's inequality:*) The goal of this problem is to prove several useful inequalities for the $\ell_p$-norm. Each of the inequalities referenced above can be proved using the one previously mentioned (except, of course, for the first one).

(a) (*Young's inequality for products:*) Suppose $p, q > 0$ are such that $\frac{1}{p} + \frac{1}{q} = 1$. If $a, b \geq 0$, prove that
$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \, ,$$
with equality if and only if $a^p = b^q$.

*Hint:* Use the fact that the exponential function is *convex*, which implies that for $0 < \theta < 1$, we have $e^{\theta x + (1-\theta) y} \leq \theta e^x + (1-\theta) e^y$, with equality if and only if $x = y$.

(b) (*Hölder's inequality:*) Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ with $x_k = [\mathbf{x}]_k$ and $y_k = [\mathbf{y}]_k$ for $1 \leq k \leq n$. Show that
$$\sum_{k=1}^{n} |x_k y_k| \leq ||\mathbf{x}||_p \, ||\mathbf{y}||_q \, ,$$
with equality if and only if $\frac{|x_k|^p}{||\mathbf{x}||_p^p} = \frac{|y_k|^q}{||\mathbf{y}||_q^q}$ for all $1 \leq k \leq n$.

(c) (*Minkowski's inequality:*) Suppose that $\mathbf{x}$ and $\mathbf{y}$ are as in part (b). Prove that
$$||\mathbf{x} + \mathbf{y}||_p \leq ||\mathbf{x}||_p + ||\mathbf{y}||_p \, ,$$
with equality if and only if $x_k = \lambda_k y_k$ or $y_k = \lambda_k x_k$ for some $\lambda_k \geq 0$ for all $1 \leq k \leq n$ for $p = 1$ and $\mathbf{x} = \lambda \mathbf{y}$ or $\mathbf{y} = \lambda \mathbf{x}$ for some $\lambda \geq 0$ for $p > 1$. This is the *triangle inequality* for the $\ell_p$-norm.

*Hint:* First show that $|c + d| \leq |c| + |d|$ for $c, d \in \mathbb{C}$ with equality if and only if $c = Kd$ or $d = Kc$ for some $K \geq 0$ (this is the triangle inequality for complex scalars). Then apply Hölder's inequality (twice) and trace back all of the conditions required for equality.

**3.** (10 points) (*Dual of the $\ell_p$-norm:*) If $||\cdot||$ is a norm on $\mathbb{C}^n$, recall that, for the standard inner product, the dual norm $||\cdot||_{\circledast}$ is defined as follows:
$$||\mathbf{z}||_{\circledast} = \sup \left\{ \mathrm{Re}\left[\mathbf{z}^\dagger \mathbf{x}\right] : ||\mathbf{x}|| \leq 1 \right\} .$$

Suppose that $p, q > 0$ and satisfy $\frac{1}{p} + \frac{1}{q} = 1$. Prove that the dual of the $\ell_p$-norm $||\cdot||_p$ is the $\ell_q$-norm $||\cdot||_q$.

*Hint:* Use the equivalent formulation of the dual norm given by

$$||\mathbf{z}||_{\circledast} = \sup\left\{\left|\mathbf{z}^\dagger\mathbf{x}\right| : ||\mathbf{x}|| = 1\right\},$$

along with Hölder's inequality from the previous problem to show that $\left|\mathbf{z}^\dagger\mathbf{x}\right| \leq ||\mathbf{z}||_q$. Then find a clever choice of $\mathbf{x}$ to achieve equality.

**4.** (10 points) (*Block matrix inversion and the matrix inversion lemma:*) Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are matrices of dimensions $m \times m$, $m \times n$, $n \times m$, and $n \times n$, respectively. Suppose that $\mathbf{A}$ and $\mathbf{D}$ are invertible. This problem deals with the $(m+n) \times (m+n)$ block matrix $\mathbf{M}$ given by

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}.$$

(a) Show the following two identities:

$$\begin{aligned} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times n} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0}_{n\times m} & \mathbf{I}_n \end{bmatrix} &= \begin{bmatrix} \mathbf{A} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{bmatrix} \\ \begin{bmatrix} \mathbf{I}_m & -\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0}_{n\times m} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times n} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_n \end{bmatrix} &= \begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{D} \end{bmatrix} \end{aligned}.$$

(b) Using part (a), show that

$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times n} \\ \mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0}_{n\times m} & \mathbf{I}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_m & \mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0}_{n\times m} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times n} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_n \end{bmatrix} \end{aligned}.$$

From this result, show that we have

$$\det(\mathbf{M}) = \det(\mathbf{A})\det\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right) = \det\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)\det(\mathbf{D}).$$

*Hint:* For the first part, use the fact that

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{F} \\ \mathbf{0}_{n\times m} & \mathbf{I}_n \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}_m & -\mathbf{F} \\ \mathbf{0}_{n\times m} & \mathbf{I}_n \end{bmatrix} \text{ and } \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times n} \\ \mathbf{G} & \mathbf{I}_n \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times n} \\ -\mathbf{G} & \mathbf{I}_n \end{bmatrix},$$

for any $m \times n$ matrix $\mathbf{F}$ and $n \times m$ matrix $\mathbf{G}$. To help with the second part, use the fact that

$$\det(\mathbf{P}\mathbf{Q}) = \det(\mathbf{P})\det(\mathbf{Q}) \text{ and } \det\left(\begin{bmatrix} \mathbf{R} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{S} \end{bmatrix}\right) = \det(\mathbf{R})\det(\mathbf{S}),$$

for any $k \times k$ matrices $\mathbf{P}$ and $\mathbf{Q}$, as well as any $m \times m$ matrix $\mathbf{R}$ and $n \times n$ matrix $\mathbf{S}$.

(c) From the result given in part (b), show the following:

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{I}_m & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0}_{n\times m} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times n} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_n \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times n} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & -\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0}_{n\times m} & \mathbf{I}_n \end{bmatrix}.$$

*Hint:* Use the identities $(\mathbf{U}\mathbf{V})^{-1} = \mathbf{V}^{-1}\mathbf{U}^{-1}$ and

$$\begin{bmatrix} \mathbf{X} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{Y} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{X}^{-1} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{Y}^{-1} \end{bmatrix},$$

which hold for any $k \times k$ invertible matrices $\mathbf{U}$ and $\mathbf{V}$, as well as any $m \times m$ invertible matrix $\mathbf{X}$ and $n \times n$ invertible matrix $\mathbf{Y}$.

(d) Multiply out the expressions given in part (c) to show that

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \\ -\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} & \left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix}$$
$$= \begin{bmatrix} \left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} + \mathbf{D}^{-1} \end{bmatrix}.$$

From this, finally show that

$$\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1},$$

which is a form of the *matrix inversion lemma.*

***5.** (30 points) (*Conditional Gaussian distributions and Schur complements:*) Let $\mathbf{z} \in \mathbb{R}^k$ be a random vector with multivariate normal or Gaussian distribution with mean $\boldsymbol{\mu}_\mathbf{z} \in \mathbb{R}^k$ and covariance $\boldsymbol{\Sigma}_\mathbf{z} \in \mathbb{S}^k_{++}$. This means that the probability density function (pdf) of $\mathbf{z}$, denoted by $f_\mathbf{z}(\boldsymbol{z})$, is given by

$$f_\mathbf{z}(\boldsymbol{z}) = \frac{1}{(2\pi)^{\frac{k}{2}}\left(\det\left(\boldsymbol{\Sigma}_\mathbf{z}\right)\right)^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_\mathbf{z})^T\boldsymbol{\Sigma}_\mathbf{z}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_\mathbf{z})}.$$

For notational convenience, we will write $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{z}, \boldsymbol{\Sigma}_\mathbf{z})$. Suppose now that $\mathbf{z}$ is decomposed as $\mathbf{z} = \begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T \end{bmatrix}^T$, where $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$ (evidently $k = m + n$). Furthermore, suppose that $\boldsymbol{\mu}_\mathbf{z}$ and $\boldsymbol{\Sigma}_\mathbf{z}$ are partitioned as follows:

$$\boldsymbol{\mu}_\mathbf{z} = \begin{bmatrix} \boldsymbol{\mu}_\mathbf{x} \\ \boldsymbol{\mu}_\mathbf{y} \end{bmatrix}, \quad \boldsymbol{\Sigma}_\mathbf{z} = \begin{bmatrix} \boldsymbol{\Sigma}_\mathbf{x} & \boldsymbol{\Sigma}_\mathbf{xy} \\ \boldsymbol{\Sigma}_\mathbf{yx} & \boldsymbol{\Sigma}_\mathbf{y} \end{bmatrix}.$$

Here, $\boldsymbol{\mu}_\mathbf{x} \in \mathbb{R}^m$, $\boldsymbol{\mu}_\mathbf{y} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}_\mathbf{x} \in \mathbb{S}^m_{++}$, $\boldsymbol{\Sigma}_\mathbf{y} \in \mathbb{S}^n_{++}$, $\boldsymbol{\Sigma}_\mathbf{xy} \in \mathbb{R}^{m\times n}$, and $\boldsymbol{\Sigma}_\mathbf{yx} = \boldsymbol{\Sigma}_\mathbf{xy}^T \in \mathbb{R}^{n\times m}$. Now suppose that $\mathbf{v} \triangleq (\mathbf{x}|\mathbf{y} = \boldsymbol{y}_0)$, that is, $\mathbf{v} \in \mathbb{R}^m$ is the random variable formed from $\mathbf{x}$ *given* that $\mathbf{y} = \boldsymbol{y}_0$ was observed.

Show that $\mathbf{v} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{v}, \boldsymbol{\Sigma}_\mathbf{v})$, where $\boldsymbol{\mu}_\mathbf{v}$ and $\boldsymbol{\Sigma}_\mathbf{v}$ are given by

$$\boldsymbol{\mu}_\mathbf{v} = \boldsymbol{\mu}_\mathbf{x} + \boldsymbol{\Sigma}_\mathbf{xy}\boldsymbol{\Sigma}_\mathbf{y}^{-1}\left(\boldsymbol{y}_0 - \boldsymbol{\mu}_\mathbf{y}\right), \quad \boldsymbol{\Sigma}_\mathbf{v} = \boldsymbol{\Sigma}_\mathbf{x} - \boldsymbol{\Sigma}_\mathbf{xy}\boldsymbol{\Sigma}_\mathbf{y}^{-1}\boldsymbol{\Sigma}_\mathbf{yx}.$$

Notice that $\mathbf{\Sigma_v}$ is the *Schur complement* of $\mathbf{\Sigma_y}$ in $\mathbf{\Sigma_z}$.

*Hint:* From our knowledge of conditional probability, the pdf of $\mathbf{v}$, denoted $f_{\mathbf{v}}(\boldsymbol{v})$, is given by

$$f_{\mathbf{v}}(\boldsymbol{v}) = \frac{f_{(\mathbf{x},\mathbf{y})}(\boldsymbol{x}, \boldsymbol{y}_0)}{f_{\mathbf{y}}(\boldsymbol{y}_0)} = \frac{f_{\mathbf{z}}\Big(\left[\begin{array}{cc} \boldsymbol{v}^T & \boldsymbol{y}_0^T \end{array}\right]^T\Big)}{f_{\mathbf{y}}(\boldsymbol{y}_0)} \,,$$

where $f_{\mathbf{y}}(\boldsymbol{y}_0)$ is the pdf of $\mathbf{y}$ evaluated at $\boldsymbol{y}_0$, which can be obtained using the multivariate Gaussian distribution expression above as it can be shown that $\mathbf{y} \sim \mathcal{N}\big(\boldsymbol{\mu}_{\mathbf{y}}, \mathbf{\Sigma_y}\big)$. Also, be sure to use the results from the previous problem.

***Reading assignments:***

**1.** Look over parts of *The Matrix Cookbook* as needed and start 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.

---