The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381K: Large Scale Optimization — Fall 2015

Problem Set Two Solutions

Constantine Caramanis .

Matlab and Computational Assignments.

1. *This problem illustrates how the gradient descent algorithm behaves in different levels of strong convexity. To begin with, download the file:* `http://users.ece.utexas.edu/~cmcaram/EE381V_2012F/ps1_matlab.zip`, *which contains a matlab file that will generate the data for the problem.*

   We have a simple unconstrained optimization problem:

$$\min_{\beta \in \mathbb{R}^n} \quad f(\beta) \triangleq \frac{1}{2}\beta^T X \beta$$

   where $X \in \mathbb{R}^{n \times n}$ is a symmetric and positive definite matrix. In the matlab file, you can find three matrices for the problem, in which (a) all eigenvalues are one, (b) a half of the eigenvalues are one and the other half of them are very small, (c) all other than a few very large eigenvalues are one.

   We want to run the gradient descent algorithm which iteratively computes

$$\beta^{(n+1)} = \beta^{(n)} - \gamma \nabla f(\beta^{(n)})$$

   where γ is a constant step size. The initial $\beta^{(0)}$ is the all-one vector.

   For each matrix, find the range of γ that the solution converges to zero and the range of γ that the algorithm diverges, and explain why. Take example values of γ to illustrate the two behaviors, convergence to zero and divergence. Plot $f(\beta^{(n)})$ over n for the two of your values.

   Solution Since X is symmetric and positive definite, there always exists an eigendecomposition $X = U\Lambda U^T$ where $U \in \mathbb{R}^{m \times m}$ is a unitary matrix and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix with positive eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m > 0$.

   The gradient descent algorithm iteratively runs

$$\begin{aligned}
\beta^{(n+1)} &= \beta^{(n)} - \gamma \nabla f(\beta^{(n)}) \\
&= \beta^{(n)} - \gamma X \beta^{(n)} \\
&= \beta^{(n)} - \gamma U \Lambda U^T \beta^{(n)} \\
&= U(I - \gamma\Lambda) U^T \beta^{(n)}
\end{aligned}$$

   Let $\hat{\beta} = U^T \beta$. Then we get

$$\hat{\beta}^{(n)} = (I - \gamma\Lambda)^n \hat{\beta}^{(0)},$$

and thus for each component $i \in \{1, 2, \ldots, m\}$ we also get

$$\hat{\beta}_i^{(n)} = (1 - \gamma\lambda_i)^n \hat{\beta}_i^{(0)}. \tag{1}$$

It follows that if $|1 - \gamma\lambda_i| < 1$ for all i then the solution converges. The constant step size γ must be smaller than $2/\lambda_1$.

	convergence	divergence
(a)	$\gamma < 2$	$\gamma > 2$
(b)	$\gamma < 2$	$\gamma > 2$
(c)	$\gamma < 0.02$	$\gamma > 0.02$

This is a simple example of the convergence condition for the constant step size that we learned in class: If $\nabla^2 f(\beta) \preceq MI$, then gradient descent with constant step size $\gamma < 2/M$ converges. Since the Hessian $\nabla^2 f(\beta)$ is equal to $X \preceq \lambda_1 I$ for any β in this problem, we get the above condition.

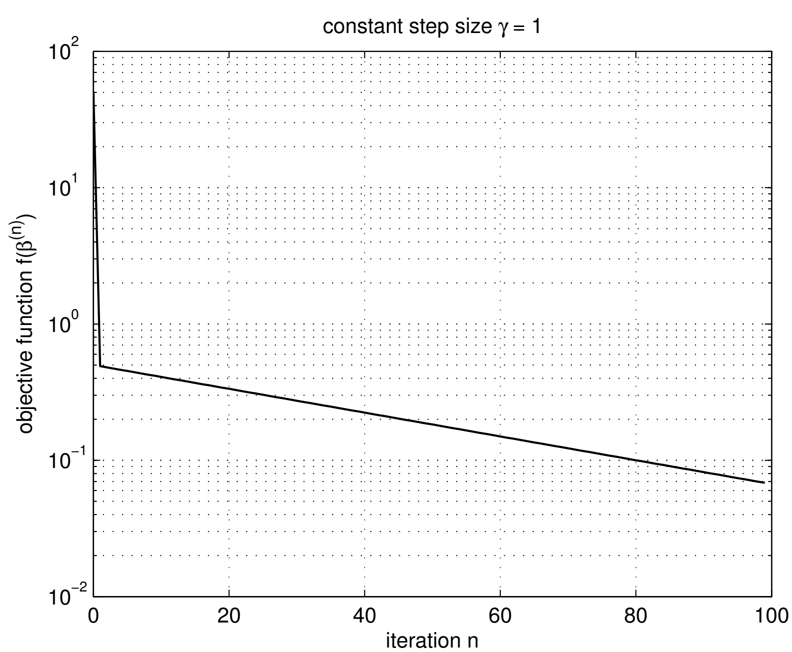
2. *Take $\gamma = 1$, and plot $f(\beta^{(n)})$ over n for the second matrix (b) of the above three. Explain the convergence behavior of the solution based on the plot.*

   Solution The objective function can also be written as $f(\beta) = \frac{1}{2}\beta^T X \beta = \frac{1}{2}\hat{\beta}^T \Lambda \hat{\beta} = \sum_{i=1}^m \lambda_i \hat{\beta}_i^2$, so the convergence behavior of the objective function also depends on $\hat{\beta}$.

   If we set $\gamma = 1$ for the second matrix (b), we get

$$\hat{\beta}_i^{(n)} = (1 - 1 \cdot 1)^n \hat{\beta}_i^{(0)} = 0 \quad \text{for } 1 \leq i \leq 50,$$
$$\hat{\beta}_i^{(n)} = (1 - 1 \cdot 0.01)^n \hat{\beta}_i^{(0)} = (0.99)^n \hat{\beta}_i^{(0)} \quad \text{for } 51 \leq i \leq 100.$$

   This behavior is shown in the figure below. There is a big drop at the first iteration because of the first 50 components of $\hat{\beta}$ vanishing at once. The following linear convergence to zero after the first iteration is due to the second 50 components of $\hat{\beta}$ decreasing geometrically with a factor of 0.99.



Written Problems

1. *Various properties of orthogonal subspaces: Let V be a finite dimensional vector space with an inner produce, and let $U \subseteq V$ be a subspace. Recall that the space $U^{\perp}$ is defined as:*

$$U^{\perp} = \{v \in V \,:\, \langle v\,,\, u\rangle = 0,\ \forall u \in U\}.$$

(a) *Show that if U is a subspace, then so is $U^{\perp}$.*

Solution It is sufficient to show that $U^{\perp}$ is closed under linear combination. Consider two different vectors $u_1^{\perp}, u_2^{\perp} \in U^{\perp}$. By definition, we have $\langle u_1^{\perp}, u\rangle = \langle u_2^{\perp}, u\rangle = 0$ for any $u \in U$. Then we get $\langle \lambda_1 u_1^{\perp} + \lambda_2 u_2^{\perp}, u\rangle = \lambda_1\langle u_1^{\perp}, u\rangle + \lambda_2\langle u_2^{\perp}, u\rangle = 0$, and so $\lambda_1 u_1^{\perp} + \lambda_2 u_2^{\perp} \in U^{\perp}$. This proves that $U^{\perp}$ is a subspace.

In fact, $U^{\perp}$ is a subspace even if U is not a subspace. Note that any properties of subspaces are not used in the proof.

(b) *Show that $(U^{\perp})^{\perp} = U$.*

Solution Let us first prove $U \subseteq (U^{\perp})^{\perp}$ by showing that every vector $u \in U$ is also in $(U^{\perp})^{\perp}$. A vector $u \in U$ satisfies that $\langle u, u^{\perp}\rangle = 0$ for every $u^{\perp} \in U^{\perp}$, otherwise such $u^{\perp}$ cannot be in $U^{\perp}$. Then it follows that $u \in (U^{\perp})^{\perp}$ by definition.

Now we prove $U \supseteq (U^{\perp})^{\perp}$ using the property in Problem 1(e). Consider a vector $v \in (U^{\perp})^{\perp}$. It can be written uniquely as $v = u + u^{\perp}$ where $u \in U$ and $u^{\perp} \in U^{\perp}$. Then we have

$$0 = \langle v, u^{\perp}\rangle = \langle u + u^{\perp}, u^{\perp}\rangle = \langle u, u^{\perp}\rangle + \langle u^{\perp}, u^{\perp}\rangle = \langle u^{\perp}, u^{\perp}\rangle$$

which implies $u^{\perp} = 0$ and $v = u$. Therefore, we get $v \in U$.

(c) *Show that if $U, W \subseteq V$ are subspaces of V, then*

$$U \subseteq W \ \Leftrightarrow U^{\perp} \supseteq W^{\perp}.$$

Solution Let us first prove $U \subseteq W \Rightarrow U^{\perp} \supseteq W^{\perp}$. Suppose $U \subseteq W$. Then any $w^{\perp} \in W^{\perp}$ satisfies $\langle w^{\perp}, u\rangle = 0$ for all $u \in U \subseteq W$ by definition, and thus it is included in $U^{\perp}$. This proves $U \subseteq W$. For the converse, we use the property in Problem 1(b) so that we get $U^{\perp} \supseteq W^{\perp} \Rightarrow (U^{\perp})^{\perp} = U \subseteq (W^{\perp})^{\perp} = W$.

(d) *Suppose now that $X \subseteq V$ is just a subset, i.e., not necessarily a subspace of V. Show that the definition $X^{\perp}$ still makes sense, and that $X^{\perp}$ is a subspace. Next show that $(X^{\perp})^{\perp} \supseteq X$, and it is defined as the smallest subspace that contains the set X.*

Solution It is sufficient to show that for any subspace V containing X also contains $(X^{\perp})^{\perp}$. Note that $U \subseteq W \Rightarrow U^{\perp} \supseteq W^{\perp}$ in Problem 1(c) holds if U and W are just subsets, not subspaces. Then for any subspace $V \supseteq X$, we have $V^{\perp} \subseteq X^{\perp}$, and also $V = (V^{\perp})^{\perp} \supseteq (X^{\perp})^{\perp}$. Since $(X^{\perp})^{\perp}$ is a subspace, it is the smallest subspace that contains X.

(e) *Show that when U is a subspace of V, then V is the direct product of U and $U^\perp$ (denoted $V = U \oplus U^\perp$). That is, show that any $v \in V$ can be written* uniquely *as*

$$v = u + u^\perp,$$

where $u \in U$, and $u^\perp \in U^\perp$.

Solution Suppose there are two different representations of $v \in V$, such that

$$v = u_1 + u_1^\perp = u_2 + u_2^\perp,$$

where $u_1, u_2 \in U$, $u_1 \neq u_2$, and $u_1^\perp, u_2^\perp \in U^\perp$, $u_1^\perp \neq u_2^\perp$. Then we have

$$u_1 - u_2 = u_2^\perp - u_1^\perp.$$

Since U and $U^\perp$ are subspaces and so closed under addition, it follows that $u_1 - u_2 = U$ and $u_2^\perp - u_1^\perp \in U^\perp$. Then $U \cap U^\perp = \{0\}$ implies that $u_1 - u_2 = u_2^\perp - u_1^\perp = 0$, which is a contradiction. Therefore, v can be written uniquely as $v = u + u^\perp$ where $u \in U$ and $u^\perp \in U^\perp$.

2. *(Boyd and Vandenberghe, Ex. 2.10) Consider the set*

$$C = \{x \in \mathbb{R}^n \,:\, x^\top A x + b^\top x + c \leq 0\},$$

where $A \in \mathbb{S}^n$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

(a) *Show that if $A \in \mathbb{S}^n_+$ (i.e., A is positive semidefinite) then the set C is convex.*

Solution Let $f(x) = x^\top A x + b^\top x + c$, and consider two different vectors $x_1, x_2 \in C$, i.e., $f(x_1) \leq 0$ and $f(x_2) \leq 0$. Then we want to show

$$f(\lambda x_1 + (1-\lambda)x_2) \leq 0$$

for any $\lambda \in [0,1]$. We first have

$$\begin{aligned}
&\lambda f(x_1) + (1-\lambda) f(x_2) - f(\lambda x_1 + (1-\lambda)x_2) \\
&= \lambda x_1^\top A x_1 + (1-\lambda) x_2^\top A x_2 - (\lambda x_1 + (1-\lambda)x_2)^\top A (\lambda x_1 + (1-\lambda)x_2) \\
&= \lambda(1-\lambda) x_1^\top A x_1 + \lambda(1-\lambda) x_2^\top A x_2 + 2\lambda(1-\lambda) x_1^\top A x_2 \\
&= \lambda(1-\lambda)(x_1 + x_2)^\top A (x_1 + x_2) \\
&\geq 0
\end{aligned}$$

where the inequality follows from $A \in \mathbb{S}^n_+$. Therefore, we get

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \leq 0.$$

This proves $\lambda x_1 + (1-\lambda)x_2 \in C$.

(b) *Consider the set obtained by intersecting C with a hyperplane:*

$$C_1 = C \cap \{x \,:\, g^\top x + h = 0\}.$$

Show that C_1 is convex if there exists $\lambda \in \mathbb{R}$ such that $(A + \lambda gg^\top) \in \mathbb{S}^n_+$.

Solution Suppose there exists $\lambda \in \mathbb{R}$ such that $(A + \lambda gg^\top) \in \mathbb{S}^n_+$. Then C_1 can be equivalently described as

$$C_1 = \{x \,:\, x^\top (A + \lambda gg)x + b^\top x + (c - \lambda h^2) \leq 0\} \cap \{x \,:\, g^\top x + h = 0\}.$$

It follows from $(A + \lambda gg^\top) \in \mathbb{S}^n_+$ that the above two sets are convex. Since the intersection of two convex sets is convex, C_1 is convex.

3. *(Boyd and Vandenberghe, Ex. 2.21) For $C, D \subseteq \mathbb{R}^n$ disjoint convex sets, let*

$$\mathcal{S} = \{(a, b) \,:\, a^\top x \leq b \;\forall x \in C,\; a^\top x \geq b \;\forall x \in D\}$$

be the set of separating hyperplanes. Show that $\mathcal{S}$ is convex.

Solution Let (a_1, b_1) and (a_2, b_2) be two different hyperplanes each of which separates C and D. Since we have for $\lambda \in [0, 1]$

$$\begin{aligned}
(\lambda a_1 + (1 - \lambda)a_2)^\top x &= \lambda a_1^\top x + (1 - \lambda)a_2^\top x \leq \lambda b_1 + (1 - \lambda)b_2, \quad \forall x \in C, \\
(\lambda a_1 + (1 - \lambda)a_2)^\top x &= \lambda a_1^\top x + (1 - \lambda)a_2^\top x \geq \lambda b_1 + (1 - \lambda)b_2, \quad \forall x \in D,
\end{aligned}$$

$(\lambda a_1 + (1 - \lambda)a_2, \lambda b_1 + (1 - \lambda)b_2)$ also separates C and D. This proves that $\mathcal{S}$ is convex.

4. *(?) In class we claimed that there are several natural operations on sets, that preserve convexity. Convince yourselves that the following all preserve convexity.*

(a) *Cartesian product: If $C_1, \ldots, C_m \subseteq \mathbb{R}^d$ are convex sets, then the set*

$$C = C_1 \times \cdots \times C_m = \{(x_1, \ldots, x_m),\; x_i \in C_i\}$$

is convex.

Solution For every $x = (x_1, \ldots, x_m), y = (y_1, \ldots, y_m) \in C$ and $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda)y = (\lambda x_1 + (1 - \lambda)y_1, \ldots, \lambda x_m + (1 - \lambda)y_m) \in C$$

because $\lambda x_i + (1 - \lambda)y_i \in C_i$ for every $1 \leq i \leq m$.

(b) *Affine and inverse maps: For $C \subseteq \mathbb{R}^n$ convex, and $A : \mathbb{R}^n \to \mathbb{R}^m$ a linear operator (i.e., an $m \times n$ matrix) then show that the following two sets are convex:*

$$\begin{aligned}
D_1 &= \{Ax \,:\, x \in C\} \\
D_2 &= \{x \,:\, Ax \in C\}.
\end{aligned}$$

Solution Affine map: For every $y_1, y_2 \in D_1$, there exist $x_1, x_2 \in C$ such that $y_1 = Ax_1$ and $y_2 = Ax_2$. Since C is convex, we have

$$\lambda x_1 + (1 - \lambda)x_2 \in C$$

for $\lambda \in [0, 1]$, and thus

$$\lambda y_1 + (1 - \lambda)y_2 = \lambda Ax_1 + (1 - \lambda)Ax_2 = A(\lambda x_1 + (1 - \lambda)x_2) \in D_1.$$

Inverse map: For every $x_1, x_2 \in D_2$, we have $Ax_1, Ax_2 \in C$. Since C is convex, we have

$$\lambda Ax_1 + (1-\lambda)Ax_2 = A(\lambda x_1 + (1-\lambda)x_2) \in C$$

for $\lambda \in [0,1]$, and thus

$$\lambda x_1 + (1-\lambda)x_2 \in D_2.$$

(c) *Minkowski sum: If $C_1, C_2 \subseteq \mathbb{R}^n$ are convex, show that*

$$C = C_1 + C_2 = \{x = x_1 + x_2 \, : \, x_1 \in C_1, x_2 \in C_2\}$$

is convex.

Solution For every $x = x_1 + x_2, y = y_1 + y_2 \in C$, $x_1, y_1 \in C_1$, and $x_2, y_2 \in C_2$, we have

$$\lambda x_1 + (1-\lambda)y_1 \in C_1, \;\; \lambda x_2 + (1-\lambda)y_2 \in C_2,$$

for $\lambda \in [0,1]$, and thus

$$\lambda x + (1-\lambda)y = (\lambda x_1 + (1-\lambda)y_1) + (x_2 + (1-\lambda)y_2) \in C.$$

5. *(?) Boyd and Vandenberghe, Ex. 2.26.*

Solution If $C = D$, their support functions are equal by definition. What is left is to show that $C = D$ if their support functions are equal.

Suppose $S_C(y) = S_D(y)$ but $C \neq D$. Let us assume that, without loss of generality, there is a point $\hat{x} \in C$ but $\hat{x} \notin D$. Since both the singleton set $\{\hat{x}\}$ and D are closed and convex, there exists a hyperplane $\{x | \hat{y}^T x = c\}$ strictly separating $\hat{x}$ and D, i.e., we can find $\hat{y}$ and c such that $\hat{y}^T \hat{x} > c$ but $\hat{y}^T x < c$ for every $x \in D$. This contradicts $S_C(\hat{y}) = S_D(\hat{y})$. C and D must be identical.

6. *(?) Boyd and Vandenberghe, Ex. 2.35.*

Solution Let K be the set of $n \times n$ copositive matrices. K is a proper cone if it satisfies the following conditions. (Read Section 2.4.1 in Boyd and Vandenberghe for the definition of a proper cone)

- K is convex : Let $X_1, X_2 \in K$. We have

$$z^T(\lambda X_1 + (1-\lambda)X_2)z = \lambda z^T X_1 z + (1-\lambda) z^T X_2 z \geq 0$$

for any $z \geq 0$ and $\lambda \in [0,1]$. Therefore, K is a convex set.

- K is closed : It is sufficient to prove that K^c is open. For this solution, we just provide a sketch of proof. To prove that K^c is open, we want to show that for every $X \in K^c$ any sufficiently small δX maintains $(X + \delta X) \in K^c$. Since $X \in K^c$, there exists $z \geq 0$ such that $z^T X z < 0$. For this z, if δX is so small that $|\delta X_{ij}| < |z^T X z| / n^2 (\max_i |z_i|)^2$, we get

$$z^T(X + \delta X)z = z^T X z + z^T \delta X z \leq z^T X z + n^2 (\max_{i,j} |\delta X_{ij}|)(\max_i |z_i|)^2 < z^T X z + |z^T X z| = 0$$

This shows that $(X + \delta X) \in K^c$. K^c is open, and equivalently K is closed.

- K has nonempty interior : Since K has the set of positive-definite matrices as a subset, K has nonempty interior.

- K is pointed, i.e., if $X, -X \in K$ then $X = 0$: If $X, -X \in K$, we have

$$z^T X z \geq 0, \quad -z^T X z \geq 0 \quad \Rightarrow \quad z^T X z = 0$$

  for any $z \geq 0$. This satisfies only if $X = 0$.

The dual cone of K is defined as

$$K^* = \{Y | \langle X, Y \rangle \geq 0 \text{ for all } X \in K\}$$

We will show that $K^* = P$ where P is the set of completely positive matrices, i.e.,

$$P = \{Y = BB^T | B_{ij} \geq 0, \ \forall i, j\}.$$

($P \subseteq K^*$) Suppose there exists a nonnegative matrix B such that $Y = BB^T$. Let N and B_i denote the number of columns of B and the ith column of B, respectively. For any $X \in K$, We have

$$\langle X, Y \rangle = \langle X, BB^T \rangle = \sum_{i=1}^{N} \langle X, B_i B_i^T \rangle = \sum_{i=1}^{N} \mathrm{Tr}(B_i B_i^T X) = \sum_{i=1}^{N} B_i^T X B_i \geq 0$$

where the inequality follows from that $z^T X z \geq 0$ for any $z \geq 0$. Therefore, $Y = BB^T \in K^*$.

($K^* \subseteq P$) Note that P is a proper cone. It is sufficient to show that $P^* \subseteq K$, i.e., every matrix X in the dual cone of P belongs to K. If it is shown, we get $K^* \subseteq (P^*)^* = P$. (Convince yourselves that the dual cone of the dual cone of a proper cone is the proper cone itself.)

Consider a matrix $X \notin K$. There exists $z \geq 0$ such that $z^T X z < 0$, so we have

$$z^T X z = \mathrm{Tr}(zz^T X) = \langle X, zz^T \rangle < 0.$$

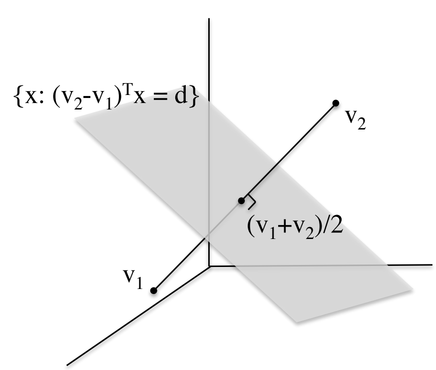
This shows that $X \notin P^*$ because $zz^T \in P$. Therefore, we have $P^* \subseteq K$, and also $K^* \subseteq P$.

7. *Consider two points, $v_1, v_2 \in \mathbb{R}^n$. Show that there exist $c \in \mathbb{R}^n$ and $d \in \mathbb{R}$ (and find them!) such that*

$$\{x : ||x - v_1|| \leq ||x - v_2||\} = \{x : c^\top x \leq d\}.$$

Thus, you are showing that the set of points in $\mathbb{R}^n$ that are closer to point v_1 than to point v_2, form a half-space.

Solution We can see geometrically that a hyperplane, perpendicular to $(v_2 - v_1)$ and lying on $(v_1+v_2)/2$, separates two half-spaces $\{x : ||x-v_1|| \leq ||x-v_2||\}$ and $\{x : ||x-v_1|| \geq ||x-v_2||\}$ (See the figure below). Then we get $c = (v_2 - v_1)$ and $d = (\|v_2\|^2 - \|v_1\|^2)/2$.



We can prove $\{x : ||x - v_1|| \le ||x - v_2||\} = \{x : (v_2 - v_1)^T x \le (\|v_2\|^2 - \|v_1\|^2)/2\}$ as follows.

$$\begin{aligned}
\{x : \|x - v_1\| \le \|x - v_2\|\} &= \{x : \|x - v_1\|^2 \le \|x - v_2\|^2\} \\
&= \{x : \|x\|^2 - 2v_1^T x + \|v_1\|^2 \le \|x\|^2 - 2v_2^T x + \|v_2\|^2\} \\
&= \{x : 2v_2^T x - 2v_1^T x \le \|v_2\|^2 - \|v_1\|^2\} \\
&= \{x : (v_2 - v_1)^T x \le (\|v_2\|^2 - \|v_1\|^2)/2\}
\end{aligned}$$

8. *Let A be an $n \times m$ real matrix, and B a $k \times m$ real matrix. Suppose that for every $x \in \mathbb{R}^m$, $Ax = 0$ only if $Bx = 0$, that is,*

$$Ax = 0 \Rightarrow Bx = 0.$$

Show that there exists a $k \times n$ real matrix C such that $CA = B$.

Solution The assumption $Ax = 0 \Rightarrow Bx = 0$ implies that $\mathrm{Null}(A) \subseteq \mathrm{Null}(B)$. Since the two null spaces are subspaces, we use the property in Problem 1(c) to get $\mathrm{Range}(A^\top) \supseteq \mathrm{Range}(B^\top)$. This means that for each $b \in \mathrm{Range}(B^\top)$ there exists $c \in \mathbb{R}^n$ such that $A^\top c = b$, so it is also true for each columns of $B^\top$. This proves that there exists a $k \times n$ real matrix C such that $A^\top C^\top = B^\top$, and equivalently $CA = B$.