Fall, 2020 CIS 515

Fundamentals of Linear Algebra and Optimization
Jean Gallier

Homework 7

November 30, 2020; Due December 14, 2020

Problem B1 (50 pts). Linear programming with box constraints is the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax = b \\ & l \leq x \leq u, \end{aligned}$$

where A is an $m \times n$ matrix, $c, u, l, x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, with $l \leq u$ (which means that $l_i \leq u_i$, for $i = 1, \ldots, n$).

(1) (20 points) Prove that the dual of the above program is the following program:

$$\begin{aligned} \text{maximize} \quad & -\nu^\top b - \lambda_1^\top u + \lambda_2^\top l \\ \text{subject to} \quad & A^\top \nu + \lambda_1 - \lambda_2 + c = 0 \\ & \lambda_1 \geq 0, \quad \lambda_2 \geq 0. \end{aligned}$$

(2) (10 points) The primal problem in (1) can be reformulated by incorporating the constraints $l \leq x \leq u$ into the objective function by defining

$$f_0(x) = \begin{cases} c^\top x & \text{if } l \leq x \leq u \\ +\infty & \text{otherwise.} \end{cases}$$

The primal is reformulated as

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Ax = b. \end{aligned}$$

Prove that the new dual function is given by

$$G(\nu) = \inf_{l \leq x \leq u} (c^\top x + \nu^\top (Ax - b)).$$

(3) (20 points) Given any real number $s \in \mathbb{R}$, let

$$s^+ = \max\{s, 0\}, \quad s^- = \max\{-s, 0\}.$$

Prove that for any fixed reals $s, \lambda, \mu \in \mathbb{R}$ with $\lambda \leq \mu$,

$$\inf_{\lambda \leq y \leq \mu} sy = \lambda s^+ - \mu s^-.$$

Hint. Consider the cases $s \geq 0$ and $s \leq 0$.

We extend the above operators to vectors $z \in \mathbb{R}^n$ componentwise by

$$z^+ = (z_1^+, \dots . z_n^+), \quad z^- = (z_1^-, \dots . z_n^-).$$

For any $w \in \mathbb{R}^n$, prove that

$$\inf_{l \leq x \leq u} x^\top w = l^\top w^+ - u^\top w^-.$$

Use the above to prove that

$$G(\nu) = -\nu^\top b + l^\top (A^\top \nu + c)^+ - u^\top (A^\top \nu + c)^-$$

and deduce that the dual program is the unconstrained problem

$$\text{maximize} \quad -\nu^\top b + l^\top (A^\top \nu + c)^+ - u^\top (A^\top \nu + c)^-$$

with respect to ν.

Problem B2 (10 pts). Verify the formula

$$(X^\top X + KI_n)^{-1} X^\top = X^\top (XX^\top + KI_m)^{-1},$$

where X is a real $m \times n$ matrix and $K > 0$. You may assume without proof that both $X^\top X + KI_n$ and $XX^\top + KI_m$ are invertible (because they are symmetric positive definite).

Problem B3 (40 pts). Recall that elastic net regression is the following optimization problem:

Program (elastic net):

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\xi^\top \xi + \frac{1}{2} K w^\top w + \tau \mathbf{1}_n^\top \epsilon \\ &\text{subject to} \\ &\qquad y - Xw - b\mathbf{1}_m = \xi \\ &\qquad w \leq \epsilon \\ &\qquad -w \leq \epsilon, \end{aligned}$$

with X an $m \times n$ matrix, $y, \xi \in \mathbb{R}^m$, $w, \epsilon \in \mathbb{R}^n$, $b \in \mathbb{R}$, where $K > 0$ and $\tau \geq 0$ are two constants controlling the influence of the ℓ^2-regularization and the ℓ^1-regularization.

The Lagrangian associated with this optimization problem is

$$\begin{aligned} L(\xi, w, \epsilon, b, \lambda, \alpha_+, \alpha_-) &= \frac{1}{2}\xi^\top \xi - \xi^\top \lambda + \lambda^\top y - b\mathbf{1}_m^\top \lambda \\ &\quad + \epsilon^\top(\tau \mathbf{1}_n - \alpha_+ - \alpha_-) + w^\top(\alpha_+ - \alpha_- - X^\top \lambda) + \frac{1}{2}K w^\top w, \end{aligned}$$

with $\lambda \in \mathbb{R}^m$ and $\alpha_+, \alpha_- \in \mathbb{R}^n_+$.

(1) (5 points) Prove that the gradient $\nabla L_{\xi, w, \epsilon, b}$ of the above Lagrangian is given by

$$\begin{pmatrix} \xi - \lambda \\ Kw + (\alpha_+ - \alpha_- - X^\top \lambda) \\ \tau \mathbf{1}_n - \alpha_+ - \alpha_- \\ -\mathbf{1}_m^\top \lambda \end{pmatrix}.$$

(2) (10 points) By setting the gradient $\nabla L_{\xi, w, \epsilon, b}$ to zero we obtain the equations

$$\begin{aligned} \xi &= \lambda \\ Kw &= -(\alpha_+ - \alpha_- - X^\top \lambda) \qquad (*_w) \\ \alpha_+ + \alpha_- - \tau \mathbf{1}_n &= 0 \\ \mathbf{1}_m^\top \lambda &= 0. \end{aligned}$$

We find that $(*_w)$ determines w.

It is more convenient to write $\lambda = \lambda_+ - \lambda_-$, with $\lambda_+, \lambda_- \in \mathbb{R}^m_+$ (recall that $\alpha_+, \alpha_- \in \mathbb{R}^n_+$), and to rescale our variables by defining $\beta_+, \beta_-, \mu_+, \mu_-$ such that

$$\alpha_+ = K\beta_+, \;\; \alpha_- = K\beta_-, \;\; \lambda_+ = K\mu_+, \;\; \lambda_- = K\mu_-.$$

We also let $\mu = \mu_+ - \mu_-$ so that $\lambda = K\mu$.

Prove that

$$\begin{aligned} w &= -(\beta_+ - \beta_- - X^\top \mu) \\ &= \begin{pmatrix} -I_n & I_n & X^\top & -X^\top \end{pmatrix} \begin{pmatrix} \beta_+ \\ \beta_- \\ \mu_+ \\ \mu_- \end{pmatrix}. \end{aligned}$$

Use the above result to prove that

$$\frac{1}{2} w^\top w = \frac{1}{2} \begin{pmatrix} \beta_+^\top & \beta_-^\top & \mu_+^\top & \mu_-^\top \end{pmatrix} Q \begin{pmatrix} \beta_+ \\ \beta_- \\ \mu_+ \\ \mu_- \end{pmatrix},$$

with Q the symmetric positive semidefinite matrix

$$Q = \begin{pmatrix} I_n & -I_n & -X^\top & X^\top \\ -I_n & I_n & X^\top & -X^\top \\ -X & X & XX^\top & -XX^\top \\ X & -X & -XX^\top & XX^\top \end{pmatrix}.$$

(3) (10 points) Prove that the dual function is given by

$$\begin{aligned} G(\mu, \beta_+, \beta_-) &= \frac{1}{2}\xi^\top \xi - \xi^\top \lambda + \lambda^\top y + w^\top(\alpha_+ - \alpha_- - X^\top \lambda) + \frac{1}{2}K w^\top w \\ &= -\frac{1}{2}K^2 \mu^\top \mu - \frac{1}{2}K w^\top w + K y^\top \mu. \end{aligned}$$

Hint. Use $(*_w)$.

(4) (15 points) Prove that

$$\frac{1}{2}\mu^\top \mu = \frac{1}{2}\begin{pmatrix} \mu_+^\top & \mu_-^\top \end{pmatrix} \begin{pmatrix} I_m & -I_m \\ -I_m & I_m \end{pmatrix} \begin{pmatrix} \mu_+ \\ \mu_- \end{pmatrix}.$$

Using (2) to rewrite $\frac{1}{2}w^\top w$, (4) to rewrite $\frac{1}{2}\mu^\top \mu$, and (3), prove that

$$G(\beta_+, \beta_-, \mu_+, \mu_-) = -\frac{1}{2}K \begin{pmatrix} \beta_+^\top & \beta_-^\top & \mu_+^\top & \mu_-^\top \end{pmatrix} P \begin{pmatrix} \beta_+ \\ \beta_- \\ \mu_+ \\ \mu_- \end{pmatrix} - K q^\top \begin{pmatrix} \beta_+ \\ \beta_- \\ \mu_+ \\ \mu_- \end{pmatrix}$$

with

$$\begin{aligned} P &= Q + K \begin{pmatrix} 0_{n,n} & 0_{n,n} & 0_{n,m} & 0_{n,m} \\ 0_{n,n} & 0_{n,n} & 0_{n,m} & 0_{n,m} \\ 0_{m,n} & 0_{m,n} & I_m & -I_m \\ 0_{m,n} & 0_{m,n} & -I_m & I_m \end{pmatrix} \\ &= \begin{pmatrix} I_n & -I_n & -X^\top & X^\top \\ -I_n & I_n & X^\top & -X^\top \\ -X & X & XX^\top + KI_m & -XX^\top - KI_m \\ X & -X & -XX^\top - KI_m & XX^\top + KI_m \end{pmatrix}, \end{aligned}$$

and

$$q = \begin{pmatrix} 0_n \\ 0_n \\ -y \\ y \end{pmatrix}.$$

Problem B4 (Extra credit 50 pts). Recall the n^2 matrices $E_{i,j}$ having the entry 1 in position (i, j) and 0 everywhere else form a basis of $\mathrm{M}_n(\mathbb{R})$.

(1) Prove that

$$d\det_A(AE_{i,j}) = \delta_{i,j}\det(A)$$

for all $A \in \mathrm{M}_n(\mathbb{R})$.

Hint. Use HW6, Problem B4(4), which states

$$d\det_A(H) = \sum_{k=1}^{n} \det(A^1, \ldots, A^{k-1}, H^k, A^{k+1}, \ldots, A^n),$$

for all $A, H \in \mathrm{M}_n(\mathbb{R})$, where $A^1, \ldots, A^n$ are the columns of A and $H^1, \ldots, H^n$ are the columns of H.

(2) Prove that for any two matrices $A, B \in \mathrm{M}_n(\mathbb{R})$,

$$d\det_A(AB) = \det(A)\mathrm{tr}(B).$$

Assuming that A is invertible, prove (again) that

$$d\det_A(H) = \det(A)\mathrm{tr}(A^{-1}H) = \mathrm{tr}(\widetilde{A}H)$$

for all $H \in \mathrm{M}_n(\mathbb{R})$.

(3) It can be shown that for any SPD matrix A, the second derivative of $f = \log\det$ is given by

$$\mathrm{D}^2 f_A(X_1, X_2) = -\mathrm{tr}(A^{-1}X_1A^{-1}X_2),$$

for all $X_1, X_2 \in \mathrm{M}_n(\mathbb{R})$. It is immediately verified that $\mathrm{D}^2 f_A$ is bilinear symmetric on $\mathrm{M}_n(\mathbb{R}) \times \mathrm{M}_n(\mathbb{R})$.

Prove that if A is SPD and X is symmetric, then $(A^{-1}X)^2$ has nonnegative eigenvalues. Conclude that if A is SPD and X is symmetric, then

$$\mathrm{D}^2 f_A(X, X) < 0$$

if $X \neq 0$.

Remark: This means that $\mathrm{D}^2 f_A$ is strictly concave on symmetric matrices (with A SPD).

TOTAL: 100 points + 50 extra cedit.