10725/36725 Optimization
Homework 2 Solutions

1 Convexity (Kevin)

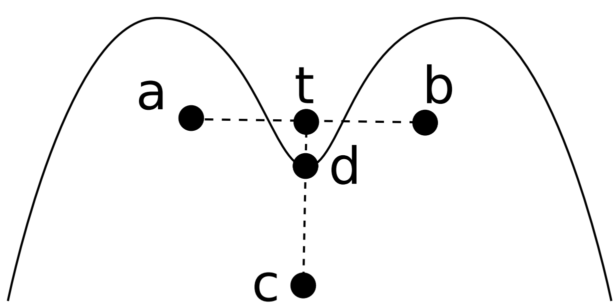
1.1 Sets

Let $A \subseteq \mathbb{R}^n$ be a closed set with non-empty interior that has a supporting hyperplane at every point on its boundary.

(a) Show that A is convex.

Solution: Assume that A is closed with non-empty interior and a supporting hyperplane at every boundary point, but it is not convex. That is, there exists $a, b \in A$ and $\alpha \in (0, 1)$ such that $t = \alpha a + (1 - \alpha)b \notin A$.

Let $c \in A^\circ$ be an interior point such that a, b and c are not co-linear. Such a point exists since there is an open ball contained in the interior. Let $d = \beta t + (1 - \beta)c \in \partial A$, for some $\beta \in (0, 1)$, be first point on the boundary when leaving c on the line towards t.



To show a contradiction, it suffices to show that any hyperplane through d that includes a and c (not strictly as it is in the interior) cannot also include b. This follows as the set of supporting hyperplanes is a subset of these. That is, for any w such that $w \cdot a \leq w \cdot d$ and $w \cdot c < w \cdot d$, we aim to show $w \cdot b > w \cdot d$.

$$b = \frac{1}{1-\alpha}\left[\frac{d-(1-\beta)c}{\beta} - \alpha a\right], \tag{1}$$

$$w \cdot b = \frac{w \cdot d}{\beta(1-\alpha)} - \frac{(1-\beta)w \cdot c}{\beta(1-\alpha)} - \frac{\alpha w \cdot a}{1-\alpha} \tag{2}$$

$$\geq \frac{(1-\beta\alpha)w \cdot d}{\beta(1-\alpha)} - \frac{(1-\beta)w \cdot c}{\beta(1-\alpha)} \tag{3}$$

$$> \frac{(\beta-\beta\alpha)w \cdot d}{\beta(1-\alpha)} = w \cdot d. \tag{4}$$

Let $X, Y \subseteq \mathbb{R}^n$ be disjoint convex sets, let $\{x \mid a^T x + b = 0\}$ be a separating hyperplane and let $f(x) = Cx + d$ be a function, where $C \in \mathbb{R}^{m\times n}, d \in \mathbb{R}^m$.

(b) [3 pts] Given that the sets $f(X)$ and $f(Y)$ are disjoint, find a hyperplane, $\{y \mid \alpha^T y + \beta = 0\}$, that separates $f(X)$ and $f(Y)$.

Here, $f(A) = \{y \mid y = f(x), x \in A\}$.

Solution: The general idea behind the solution is that, translation, rotation and scaling are straightforward to handle, and projection is the only hard case. When moving to a higher dimension, we can simply ignore the added dimensions. When projecting into a lower dimension, we can ignore the lost dimensions as we know the resulting sets are disjoint. Using the psuedoinverse does this.

Choose $\alpha^T = a^T C^\dagger$, $\beta = a^T C^\dagger d - b$, where $C^\dagger$ is the psuedoinverse of C.

Assume $\forall x \in X, a^T x + b \leq 0$. Let $x' = Cx + d$, then

$$\alpha^T x' = a^T C^\dagger C x + a^T C^\dagger d \tag{5}$$

$$= a^T I x + a^T C^\dagger d \tag{6}$$

$$\leq -b + a^T C^\dagger d = -\beta \tag{7}$$

Here, we used $C^\dagger C = I$, which is not true if we lose dimensions. Sometimes α, β is still a separator even when this does not hold, but in such cases the sets may flip sides requiring a point from each set to orient the plane. If the hyperplane only discriminates along the dimensions that are projected out, though, then α is 0, and hence it is does not define a hyperplane. Sorry about that.

1.2 Voronoi Decomposition

Let $a, b \in \mathbb{R}^n$ such that $a \neq b$.

(a) Show that the set $\{x \mid \|x-a\|_2 \leq \|x-b\|_2\}$ is a halfspace.

Solution:

$$\|x - a\|_2 \leq \|x - b\|_2 \tag{8}$$
$$\Leftrightarrow \|x - a\|_2^2 \leq \|x - b\|_2^2 \tag{9}$$
$$\Leftrightarrow x^T x - 2a^T x + a^T a \leq x^T x - 2b^T x + b^T b \tag{10}$$
$$\Leftrightarrow 2(b - a)^T x \leq b^T b - a^T a \tag{11}$$

Let $x_1, \ldots, x_k \in \mathbb{R}^n$ and let $V_i = \{x \mid \|x_k - x\|_2 \leq \|x_i - x\|_2, i \neq k\}$.

(b) Show that V_0 is a polyhedron. That is, $V_0 = \{x \mid A_0 x \leq b_0\}$.

Solution: Let the ith row of A_0 be $2(x_i - x_0)$ and the ith element of b_0 be $x_k^T x_k - x_0^T x_0$. Equivalence follows from part (a).

Let $P_i = \{x \mid A_i x \leq b_i\}$ be disjoint polyhedra that cover $\mathbb{R}^n$.

(c) Does there exist points $x_1, \ldots, x_k \in \mathbb{R}^n$ such that $V_i = P_i$ for $i = 1, \ldots, k$? If so, provide the points, otherwise construct a counterexample.

Solution: False, counterexample is $V_0 = \{(x, y) \mid x \leq 0, y \geq 0\}, V_1 = \{(x, y) \mid x \leq 0, y \leq 0\}, V_2 = \{(x, y) \mid x \geq 0\}$.

1.3 Farkas' Lemma

Let $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ such that there is some x where $Ax = b$.

(a) Show that **either** there exists $x > 0$ where $Ax = b$, **or** there exists λ such that $A^T \lambda \geq 0$, $A^T \lambda \neq 0$ and $b^T \lambda \leq 0$.

Hint: First show that $c^T x = d$ for all x such that $Ax = b$ if and only if there exists λ such that $c = A^T \lambda, d = b^T \lambda$.

Solution: Consider the sets $\{x \mid x > 0\}$ and $\{x \mid Ax = b\}$. Since the first is open, by the separating hyperplane partial converse, we know that either the sets intersect, or there is a separating hyperplane, but not both.

If the sets are disjoint, there exists $w \neq 0$ such that

$$w^T x \geq \beta \text{ if } x > 0, \text{ and} \tag{12}$$
$$w^T x \leq \beta \text{ if } Ax = b. \tag{13}$$

Let P be a basis for the null space of A. That is, $APy = 0$ for all y and thus $x = x_0 + Py$ is a solution for $Ax = b$ for any $Ax_0 = b$. We have

$$w^T x = w^T (x_0 + Py) = w^T x_0 + w^T Py. \tag{14}$$

Claim, $w^T P y = 0$ for all y. Assume there is some $w^T P y^* = z^* \neq 0$, and choose $\hat{y} = (\beta + \epsilon - w^T x_0) y^* / z^*$. Then, $\hat{x} = x_0 + P\hat{y}$ satisfies $A\hat{x} = b$, but $w^T \hat{x} > \beta$.

We are done once we prove the hint. First, given λ such that $c = A^T \lambda, d = b^T \lambda$, then for any $Ax = b$, we have

$$c^T x = \lambda^T A x = \lambda^T b = d. \tag{15}$$

Now, the harder direction. Given that $c^T x = d$ for all x such that $Ax = b$, again write $x = x_0 + Py$. Then,

$$d = c^T x = c^T (x_0 + Py) = c^T x_0 + c^T P y. \tag{16}$$

By the above argument, $c^T P y = 0$, which is true only when $P^T c = 0$. Hence, c is orthogonal to the null space of A. Therefore, by the rank-nullity theorem, it is in the row space of A. That is, there exists a lambda such that $A^T \lambda = c$.

1.4 Functions

(a) Show that the function $f(x) = (\sum_{i=1}^n x_i^p)^{1/p}$ is concave on $\mathbb{R}^n_{++}$ for all $p \in (0, 1)$.

Hint: consider the log-sum-exp and geometric mean functions in Boyd.

Solution: We will show that the Hessian is negative semi-definite.

$$\frac{\partial f(x)}{\partial x_i} = \left(\sum_{k=1}^n x_k^p \right)^{1-p/p} x_i^{p-1} \tag{17}$$

$$\frac{\partial f(x)}{\partial x_i x_j} = (1-p) \left(\sum_{k=1}^n x_k^p \right)^{1-2p/p} x_i^{p-1} x_j^{p-1} \tag{18}$$

$$\frac{\partial f(x)}{\partial x_i^2} = (1-p) \left(\sum_{k=1}^n x_k^p \right)^{1-2p/p} \left[x_i^{p-1} x_i^{p-1} - x_i^{p-2} \sum_{k=1}^n x_k^p \right] \tag{19}$$

For any v, we need to show $v^T \nabla^2 f(x) v \leq 0$, Choosing

$$K = (1-p) \left(\sum_{k=1}^n x_k^p \right)^{1-2p/p} > 0 \tag{20}$$

$$v^T \nabla^2 f(x) v = K \sum_{i=1}^n \sum_{j=1}^n v_i v_j x_i^{p-1} x_i^{p-1} - v_i^2 x_i^{p-2} x_j^p \tag{21}$$

$$= K \sum_{i=1}^n \sum_{j=1}^n a_i a_j b_i b_j - a_i^2 b_j^2, \text{ for } a_i = v_i x_i^{p/2-1}, b_i = x_i^{p/2} \tag{22}$$

$$\leq 0, \text{ by applying Cauchy-Schwarz.} \tag{23}$$

Let $G = (V, E, c, s, t)$ be an undirected graph with edge capacities $c(u, v) \geq 0$, source vertex s and sink vertex t. We say a vector $y \in \mathbb{R}^E$ is a valid *s-t-flow* if:

- $y(u, v) \leq c(u, v), \forall (u, v) \in E$, (*Capacity*)
- $y(u, v) = -y(v, u), \forall (u, v) \in E$, and (*Skew Symmetry*)
- $\sum_{v \in V} y(u, v) = 0, \forall u \in V \setminus \{s, t\}$. (*Flow Conservation*)

Let $f(x) = \min_y \|x - y\|$, subject to: y is a valid s-t-flow.

(b) Show that $f(x)$ is convex.

Solution: $\|x - y\|$ is convex in x and y and is bounded below by 0. The set of valid flows is closed and convex, as it is the intersection of linear constraints; and is non-empty, as the zero flow is valid. Therefore, $f(x)$ is convex by BV Section 3.2.5, p. 87–88.

2 The meanest nice functions

Solution courtesy of Ahmed Hefny

(a) Consider the function $g(y) = f(y - y^{(0)})$. This is the function f translated by $y^{(0)}$. Therefore $\nabla g(y) = \nabla f(y - y^{(0)})$. Also, $\arg\min_y g(y) = \arg\min_x f(x) + y_0$. Let f be a function such that when the algorithm starts at $x^{(0)} = 0$, the two conditions in question are satisfied. We will show that g satisfies the same condition when the algorithm starts at $y^{(0)}$. First, we show by induction that $y^{(k)} = x^{(k)} + y^{(0)}$: We already have $y^{(0)} = x^{(0)} + y^{(0)}$ as a base case. And $y^{(k)} = y^{(k-1)} + w^{(k)} \nabla g(y^{(k-1)}) = y^{(0)} + x^{(k-1)} + w^{(k)} \nabla f(x^{(k-1)}) = y^{(0)} + x^{(k)}$. Therefore,

$$y^{(K)} - \bar{y} = x^{(K)} - \bar{x}$$

and

$$y^{(0)} - \bar{y} = -\bar{x}$$

Therefore

$$\|x^{(K)} - \bar{x}\|^2 \geq \frac{1}{32}\| - \bar{x}\|^2 \to \|y^{(K)} - \bar{y}\|^2 \geq \frac{1}{32}\|y^{(0)} - \bar{y}\|^2 \tag{24}$$

which shows that g satisfies the first condition.

Also,

$$f(x^{(K)}) - \bar{f} = g(x^{(K)} + y^{(0)}) - g(\bar{x} + y^{(0)}) = g(y^{(K)}) + g(\bar{y}) = g(y^{(K)}) + \bar{g}$$

Therefore

$$f(x^{(K)}) - \bar{f} \geq \frac{3L\| - \bar{x}\|^2}{32(K+1)^2} \to g(y^{(K)}) - \bar{g} \geq \frac{3L\|y^{(0)} - \bar{y}\|^2}{32(K+1)^2} \tag{25}$$

which shows that g satisfies the second condition.

(b) Assume $k \geq 2$

$$\begin{aligned} \frac{\partial \zeta^{(k)}}{\partial x_1}(x) &= \frac{L}{4}\left[\frac{1}{2}\left(2x_1 + 2(x_1 - x_2)\right) - 1\right] \\ &= \frac{L}{4}\left[(2x_1 - x_2) - 1\right] \end{aligned} \tag{26}$$

$$\begin{aligned} \frac{\partial \zeta^{(k)}}{\partial x_k}(x) &= \frac{L}{4}\left[\frac{1}{2}\left(-2(x_{k-1} - x_k) + 2x_k\right)\right] \\ &= \frac{L}{4}\left(2x_k - x_{k-1}\right) \end{aligned} \tag{27}$$

$$\begin{aligned} \frac{\partial \zeta^{(k)}}{\partial x_i}(x) &= \frac{L}{4}\left[\frac{1}{2}\left(-2(x_{i-1} - x_i) + 2(x_i - x_{i+1})\right)\right] \\ &= \frac{L}{4}\left(x_{i-1} + 2x_i - x_{i+1}\right), \end{aligned} \tag{28}$$

for $1 < i < k$

$$\frac{\partial \zeta^{(k)}}{\partial x_i}(x) = 0 \tag{29}$$

for $i > k$. The second derivatives can now be computed:

$$\frac{\partial^2 \zeta^{(k)}}{\partial x_i \partial x_j}(x) = \begin{cases} 0 & i > k \text{ or } j > k \\ \frac{L}{2} & i = j \leq k \\ -\frac{L}{4} & |i - j| = 1 \text{ and } i, j \leq k \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

What remains is to show that this pattern remains for $k = 1$.

$$\frac{\partial \zeta^{(1)}}{\partial x_i}(x) = 0 \tag{31}$$

, for $i > 1$

$$\begin{aligned} \frac{\partial \zeta^{(k)}}{\partial x_1}(x) &= \frac{L}{4}\left[\frac{1}{2}\left(2x_1 + 0 + 2x_1\right) - 1\right] \\ &= \frac{L}{4}\left(2x_1 - 1\right) \end{aligned} \tag{32}$$

$$\frac{\partial^2 \zeta^{(k)}}{\partial x_1^2}(x) = \frac{L}{2} \tag{33}$$

So in general $\nabla^2 \zeta_{(k)}(x) = \frac{L}{4} A^{(k)}$ where $A^{(k)}$ is an $n \times n$ matrix with the following properties:

- $A_{ii}^{(k)} = 2$ if $1 \leq i \leq k$
- $A_{ij}^{(k)} = -1$ if $|i - j| = 1$ and $i, j \leq k$
- $A_{ij}^{(k)} = 0$ otherwise

(c) Since $\nabla^2\zeta^{(k)}(x) = \frac{L}{4}A^{(k)}$ and since $A^{(k)}$ is a symmetric matrix then,

$$\nabla\zeta^{(k)}(x) = \frac{L}{4}\left(A^{(k)}x + c\right) \tag{34}$$

to obtain c we set x to 0 and based on the derivatives obtained in part b we find that $c = \nabla\zeta_{(k)}(0) = -e_1$, where e_1 is an n-dimensional vector having the first component set to 1 and the others set 0. Therefore

$$\nabla\zeta^{(k)}(x) = \frac{L}{4}\left(A^{(k)}x - e_1\right) \tag{35}$$

To compute $\bar{x}_j^{(k)}$, we notice that setting the gradient to 0 gives $A^k\bar{x}^{(k)} - e_1 = 0$, which translates into the following recurrence relation

$$\bar{x}_j^{(k)} = 2\bar{x}_{j-1}^{(k)} - \bar{x}_{j-2}^{(k)} \tag{36}$$

for $2 \leq j \leq k$, where the dummy variable $\bar{x}_0^{(k)}$ is equal to 1 as implied by first row of $A^k\bar{x}^{(k)}$. This is a homogeneous recurrence relation with characteristic equation $(s-1)^2 = 0$ and thus its solution is of the form

$$\bar{x}_j^{(k)} = \alpha + \beta j \quad (0 \leq j \leq k) \tag{37}$$

Since $\bar{x}_0^{(k)} = 1$ then $\alpha = 1$. To find β we use the last non-zero row of A^k, which gives the equation

$$2\bar{x}_k^{(k)} - \bar{x}_{k-1}^{(k)} = 0 \tag{38}$$

$$2\beta k + 2 - \beta(k-1) - 1 = 0 \tag{39}$$

$$\beta = \frac{-1}{k+1} \tag{40}$$

Therefore

$$\bar{x}_j^{(k)} = \frac{k+1-i}{k+1} \quad (0 \leq j \leq k) \tag{41}$$

For $j > k$, $\zeta^{(k)}$ does not depend on x_j and thus any value is a minimizer (the partial derivative is always 0). The minimum value of ζ is then

$$\begin{aligned}\bar{\zeta}^{(k)} &= \tfrac{L}{4}\left[\tfrac{1}{2(k+1)^2}\left(k^2 + \textstyle\sum_{i=1}^{k-1}((k+1-i)-(k-i))^2 + 1\right) - \tfrac{k}{k+1}\right] \\ &= \tfrac{L}{4}\left[\tfrac{1}{2(k+1)^2}\left(k^2+k\right) - \tfrac{k}{k+1}\right] \\ &= -\tfrac{L}{8}\left[\tfrac{k}{k+1}\right]\end{aligned} \tag{42}$$

(d)

$$\nabla\zeta^{(k)}(x) - \nabla\zeta^{(k)}(y) = \tfrac{L}{4}A^{(k)}(x-y) = \tfrac{L}{4}(2(x-y)_* - (x-y)_- - (x-y)_+), \tag{43}$$

where $x_{*i} = x_i$ if $i \leq k$ and 0 otherwise, $x_{+i} = x_{i+1}$ if $i < k$ and 0 otherwise and similarly $x_{-i} = x_{i-1}$ if $1 < i \leq k$ and 0 otherwise. It is obvious that $\|(x-y)_*\| \leq \|x-y\|$, $\|(x-y)_-\| \leq \|x-y\|$ and $\|(x-y)_+\| \leq \|x-y\|$, since the three of them share some components of $x-y$ while setting the other components to 0. Therefore, by triangular inequality

$$\|\nabla\zeta^{(k)}(x) - \nabla\zeta^{(k)}(y)\| \leq \frac{L}{4}(2\|x-y\| + \|x-y\| + \|x-y\|) = L\|x-y\| \tag{44}$$

(e) Base case ($k=1$, $x^{(k-1)} = 0$):

$$\nabla f(x^{(k-1)}) = A^{(2K+1)}x^{(k-1)} - e_1 = -e_1, \tag{45}$$

which contains only 1 non-zero element.

Induction Step ($1 < k \leq K$):

$$\nabla f(x^{(k-1)}) = A^{(2K+1)}x^{(k-1)} - e_1 \tag{46}$$

Considering the i^{th} component ($i > k+1$):

$$\tfrac{\partial\zeta^{(k)}}{\partial x_i}(x) \quad = \tfrac{L}{4}\left(x_{i-1} + 2x_i - x_{i+1}\right), \tag{47}$$

By induction hypothesis and the fact the algorithm is gradient summing, $x_{i-1} = x_i = x_{i+1} = 0$ for $i > k+1$

(f) Knowing that $x_i^{(k)} = 0$ for $i > k$, it can be shown that $g(x^{(k)}) - f(x^{(k)}) = 0$

$$\begin{aligned} g(x^{(k)}) - f(x^{(k)}) &= \zeta^(K)(x_{(k)}) - \zeta^(2K+1)(x_{(k)}) \\ &= \tfrac{L}{4}\left[\tfrac{1}{2}\left(x_k^2 - (x_k - x_{k+1})^2 - \textstyle\sum_{i=k+1}^{2k}(x_i - x_{i+1})^2 - x_{2k+1}^2\right)\right] \\ &= 0 \end{aligned} \tag{48}$$

(g) Since, by condition (1) and gradient summing property, $x_j^{(K)} = 0$ for all $j > K$. Then,

$$\begin{aligned} \|x^{(K)} - \bar{x}\|^2 &\geq \textstyle\sum_{j=K+1}^{2K+1}\bar{x}_j^2 = \tfrac{1}{(2K+2)^2}\textstyle\sum_{j=K+1}^{2K+1}(2K+2-j)^2 \\ &= \tfrac{1}{(2K+2)^2}\left(\textstyle\sum_{j=K+1}^{2K+1}(2K+2-j)^2\right) \\ &= \tfrac{1}{(2K+2)^2}\textstyle\sum_{i=1}^{K+1}i^2 = \tfrac{1}{(2K+2)^2}\cdot\tfrac{(K+1)(K+2)(2K+3)}{6} \end{aligned} \tag{49}$$

We also have

$$\begin{aligned} \|x^{(0)} - \bar{x}\|^2 = \|\bar{x}\| &\geq \textstyle\sum_{j=0}^{2K+1} \bar{x}_j^2 = \frac{1}{(2K+2)^2} \sum_{j=0}^{2K+1} (2K+2-j)^2 \\ &= \frac{1}{(2K+2)^2} \left(\sum_{i=1}^{2K+1} i^2 \right) \\ &= \frac{1}{(2K+2)^2} \frac{(2K+1)(2K+2)(4K+3)}{6} \end{aligned} \tag{50}$$

Therefore,

$$\frac{\|x^{(K)} - \bar{x}\|^2}{\|x^{(0)} - \bar{x}\|^2} \geq \frac{(K+1)(K+2)(2K+3)}{(2K+1)(2K+2)(4K+3)} \geq \frac{1}{8} > \frac{1}{32}, \tag{51}$$

since it is decreasing in K (the denominator increases with a higher rate) and thus its lower bound is when $K \to \infty$ which is $\frac{1}{8}$.

(h) Knowing that $g = \zeta^{(K)}$ and $f = \zeta^{(2K)}$

$$\begin{aligned} f(x^{(K)}) - \bar{f} &\geq \bar{g} - \bar{f} = -\frac{L}{8}\left(\frac{K}{K+1}\right) + \frac{L}{8}\left(\frac{2K+1}{2K+2}\right) = \frac{L}{16(K+1)} \\ &= \frac{L(K+1)}{16(K+1)^2} \geq \frac{3L\|x^{(0)} - \bar{x}^{(2K+1)}\|}{32(K+1)^2} \quad \square \end{aligned} \tag{52}$$

3 Alternative formulations (Wooyoung)

Solution courtesy of Xuezhi Wang

As discussed in class, there are often more than one way to formulate an optimization problem. In this problem, we will go through examples in which you can set up your optimization in a few alternative ways.

3.1 Rank 1 approximation of matrices

In this problem, we will set up minimizing the squared error between a given $m \times n$ matrix A and a rank 1 matrix X.

$$\text{minimize}_{\text{rank}(X)=1} ||A - X||^2$$

(a) [1.5 pts] How can you reformulate the objective function and the constraint of the above optimization by writing the space of rank 1 matrices as outer products of two vectors?

Solution:

$$\text{minimize } \|A - X\|^2, \quad s.t. X = UV^T, U \in \mathbb{R}^m, V \in \mathbb{R}^n$$

(b) [1.5 pts] One possible disadvantage of writing the space of rank 1 matrices as outer products of two vectors is degeneracy of the solutions; if a pair of vectors $\mathbf{x}$ and $\mathbf{y}$ are solutions, for any non-zero scalar a, $a\mathbf{x}$ and $\frac{1}{a}\mathbf{y}$ would also be the solutions. How would you reformulate your objective functions and constraints to avoid this degeneracy?

Solution:

$$\begin{aligned} \text{minimize} \quad & \|A - \lambda UV^T\|^2, \\ s.t. \quad & U \in \mathbb{R}^m, \|U\| = 1, \\ & V \in \mathbb{R}^n, \|V\| = 1, \\ & \lambda \geq 0 \end{aligned}$$

(c) [3 pts] We started with one optimization variable X, but now we have more than one. How can you reduce the number of optimization variables in (b)? Show your work. *Hint:* try to show that one of the optimization variables is determined when others are fixed.

Solution:

$$\begin{aligned} \|A - \lambda UV^T\|^2 \quad & = tr((A - \lambda UV^T)^T(A - \lambda UV^T)) \\ & = tr(A^TA - \lambda VU^TA - \lambda A^TUV^T + \lambda^2) \\ & = tr(A^TA + \lambda^2 - 2\lambda VU^TA) \end{aligned}$$

since $tr(A^TA)$ is fixed, if U and V are fixed, the above formula will achieve the minimum if we take $\lambda = tr(VU^TA)$. Hence the original problem reduces to:

$$\min tr(A^TA) - (tr(VU^TA))^2 \quad s.t. \ \|U\| = 1, \|V\| = 1$$

(d) [2 pts] Reformulate the optimization problem in (c) into an equivalent minimization problem with a bi-linear objective function.

Solution: Since $tr(A^TA)$ is fixed, we just need to minimize $-(tr(VU^TA))^2$. And since $\lambda = tr(VU^TA) \geq 0$, we just need to minimize $-(tr(VU^TA))$, which can be written in the following bilinear form:

$$\min_{\|U\|=1, \|V\|=1} \left(-\sum_{i=1}^{m}\sum_{j=1}^{n} V_jU_ia_{ij}\right)$$

3.2 Partial minimization of the lasso problem

In classical setting of the lasso problem, the sparsity of solutions are enforced by $L1$-norm of the optimization variables ($\mathbf{y} \in \mathcal{R}^d$, $A \in \mathcal{R}^{d\times n}$, $\mathbf{x} \in \mathcal{R}^n$, x_i: i th element of $\mathbf{x}$).

$$\min ||\mathbf{y} - A\mathbf{x}||_2 + \lambda||\mathbf{x}||_1$$

In this problem, we derive the partial minimization of the lasso problem.

(a) [2 pts] How would you reformulate the optimization if we only care about the sparsity of the first k elements of $\mathbf{x}$: $x_1, x_2, \cdots, x_k$.

Solution:

$$\min \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^{k} |x_i|$$

(b) [3 pts] Show how you can eliminate $x_{k+1}, x_{k+2}, \cdots, x_n$ from the objective function.

Solution: We can reformulate the equation $Ax = y$ by

$$\begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{bmatrix} = \mathbf{y}$$

where $\mathbf{x_1}$ represents the first k elements in $\mathbf{x}$, and $\mathbf{x_2}$ represents the rest. So minimizing $||\mathbf{y} - A\mathbf{x}||$ is equivalent to minimizing $||\mathbf{y} - A_1\mathbf{x_1} - A_2\mathbf{x_2}||$. Since we don't care about the sparsity of x_2, we can fix $\mathbf{x_1}$ and get the least squared error solution of $\mathbf{x_2}$:

$$\min\ ||\mathbf{y} - A_1\mathbf{x_1} - A_2\mathbf{x_2}|| \Rightarrow f = \min\ (\mathbf{y} - A_1\mathbf{x_1} - A_2\mathbf{x_2})^T(\mathbf{y} - A_1\mathbf{x_1} - A_2\mathbf{x_2})$$

By setting the derivative of f with respect to $\mathbf{x_2}$ to zero we can get:

$$\mathbf{x_2} = (A_2^T A_2)^{-1} A_2^T (\mathbf{y} - A_1\mathbf{x_1})$$

Hence we can eliminate $x_{k+1}, x_{k+2}, \cdots, x_n$ by reformulating the problem as:

$$\min \|(I - A_2(A_2^T A_2)^{-1} A_2^T)(\mathbf{y} - A_1\mathbf{x_1})\|_2^2 + \lambda||\mathbf{x_1}||_1$$

where $\mathbf{x_1}$ represents $[x_1\ x_2\ \cdots\ x_k]$.

3.3 Equality constraint

Consider an convex optimization problem with equality constraint,

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

Assume that the feasible set is non-empty.

(a) [3 pts]Can you remove the equality constraint by reformulating x as a linear function of z. What conditions does the parameters of your linear function need to satisfy?

Solution:

$$\text{minimize } f(Fz + x_0)$$

where x_0 is a paricular solution to $Ax = b$, and F lies in the range of the nullspace of A.

3.4 Building a box-shape structure

You are asked to build a box-shape structure. The area of the wall paper provided to you for the job is A_{wall}, and the area of the floor cannot exceed A_{floor}. Your picky boss even wants to control the ratio of height (h) to width (w) and also the ratio of width (w) to depth (d) of the wall structure; the ratio between h and w should be in the range of $[\alpha,\ \beta]$, the ratio between w and d in $[\gamma,\delta]$. Your mission is to maximize the volume of the box-shape structure while satisfying all the constraints.

(a) [2 pts] Formulate your optimization problem as a maximization problem. Show your objective functions and constraints.

Solution:

$$\begin{array}{rc} \max & hwd \\ \text{subject to} & 2hw + 2hd \leq A_{wall} \\ & wd \leq A_{floor} \\ & \alpha \leq h/w \leq \beta \\ & \gamma \leq w/d \leq \delta \end{array}$$

(b) [2 pts] A geometric program is an optimization problem of the form

$$\begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 1, \quad i = 1, \cdots, m \\ & g_i(x) = 1, \quad i = 1, \cdots, p \end{array}$$

where $g_i(x) = c_i x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$, $(c_i > 0, a_j \in \mathcal{R})$ and $f_i(x) = \sum_{k=1}^{K} c_{i,k} x_1^{a_{1k}} x_2^{a_2} \cdots x_n^{a_{nk}}$, $(c_{i,k} > 0, a_{ik} \in \mathcal{R})$. Recently, efficient and robust algorithms have been developed to solve even large-scale geometric programs. Convert the optimization problem in (a) into a geometric programming problem.

Solution:

$$\begin{array}{rc} \min & 1/hwd \\ \text{subject to} & 2h(w+d)/A_{wall} \leq 1 \\ & wd/A_{floor} \leq 1 \\ & \alpha w h^{-1} \leq 1 \\ & \beta^{-1} h w^{-1} \leq 1 \\ & \delta^{-1} w d^{-1} \leq 1 \\ & \gamma d w^{-1} \leq 1 \end{array}$$