This document is a compilation of computational subproblems and their analytical solutions that I have encountered as a result of implementing different first-order algorithms. Some may be more detailed than others depending on how obvious the solution was to me at first.

1 Projections

In many first-order methods, a projection problem is often required to be solved per iteration in order to solve a more general optimization problem. While there exists projection free methods, sometimes the projection onto a specific set S can be done analytically and is not a computational burden. We will look at a few such instances. We seek to solve problems of the form

$$\underset{x \in S}{\operatorname{argmin}} \ ||x-u||^2$$

for some convex set $S \subset \mathbb{R}^n$ and vector $u \in \mathbb{R}^n$.

1.1 Standard Spectrahedron

Let $m^2 = n$ and notice that a projection onto the standard spectrahedron

$$\text{Spe}_m := \{X \in \mathbb{R}^{m \times m} \mid X \succeq 0, \text{tr}(X) = 1\}$$

takes the form

$$\underset{X \in \text{Spe}_n}{\operatorname{argmin}} \ ||X-U||_F^2$$

for some matrix $U \in \mathbb{R}^{m \times m}$. Since $X \succeq 0$, it has eigendecomposition $X = V^T \Lambda V$ for some orthogonal matrix V and diagonal matrix Λ. If we denote λ to be the vector satisfying $\text{diag}(\lambda) = \Lambda$, then by multiplying on the right and left by V^T and V respectively, we can consider the equivalent problem without loss of generality

$$\underset{\Lambda=\text{diag}(\lambda), \text{tr}(\Lambda)=1, \Lambda \geq 0}{\operatorname{argmin}} \ ||\Lambda - U||_F^2 = \underset{\lambda^T e = 1, \lambda \geq 0}{\operatorname{argmin}} \ ||\lambda||_2^2 - 2\langle \lambda, \text{diag}(U) \rangle = \underset{\lambda^T e = 1, \lambda \geq 1}{\operatorname{argmin}} \ ||\lambda - \text{diag}(U)||_2^2$$

after rewriting the objective function using the Frobenius product. Here, $e \in \mathbb{R}^m$ is a vector of all 1's and $\Delta_m := \{\lambda \in \mathbb{R}^m \mid \lambda^T e = 1, \lambda \geq 0\}$ is usually denoted the standard simplex. That is, projecting onto the standard spectrahedron requires projecting the diagonal of U onto the standard simplex.

2 Solving Proximal Problems of "Easy" Functions

There are many algorithms that assume the knowledge of an analytical solution to a proximal problem,

$$\text{prox}_h(\tau, u) := \min_{w \in \mathbb{R}^n} h(w) + \frac{\tau}{2} ||w-u||^2$$

for some constant $\tau \in \mathbb{R}$ and vector $u \in \mathbb{R}^n$. Here, we will solve this problem for different choices of $h(w)$.

2.1 2-Norm

Let $h(w) = ||w - b||_2$. We know that

$$h(w) = \sup_{||\xi|| \leq 1} \langle w - b, \xi \rangle,$$

so the optimization problem becomes

$$\operatorname*{argmin}_{w \in \mathbb{R}^n} \operatorname*{argmax}_{||\xi|| \leq 1} \langle w - b, \xi \rangle + \frac{\tau}{2} ||w - u||^2 = \operatorname*{argmax}_{||\xi|| \leq 1} \operatorname*{argmin}_{w \in \mathbb{R}^n} \langle w - b, \xi \rangle + \frac{\tau}{2} ||w - u||^2 .$$

by the minimax theorem. An optimal solution to the min problem is $w^* := w^*(\xi)$ such that

$$\xi + \tau(w^* - u) = 0, i.e., w^* = u - \frac{\xi}{\tau}.$$

Continuing with the optimization problem, we have

$$\begin{aligned}
\operatorname*{argmax}_{||\xi|| \leq 1} \operatorname*{argmin}_{w \in \mathbb{R}^n} \langle w - b, \xi \rangle + \frac{\tau}{2} ||w - u||^2 &= \operatorname*{argmax}_{||\xi|| \leq 1} \langle u - b, \xi \rangle - \frac{1}{\tau} ||\xi||^2 + \frac{1}{2\tau} ||\xi||^2 \\
&= \operatorname*{argmax}_{||\xi|| \leq 1} \langle u - b, \xi \rangle - \frac{1}{2\tau} ||\xi||^2 \\
&= \operatorname*{argmax}_{||\xi|| \leq 1} 2\tau \langle u - b, \xi \rangle - ||\xi||^2 \\
&= \operatorname*{argmin}_{||\xi|| \leq 1} ||\xi||^2 - 2\langle \tau(u - b), \xi \rangle \\
&= \operatorname*{argmin}_{||\xi|| \leq 1} ||\xi - \tau(u - b)||^2
\end{aligned}$$

which is to say, project $\tau(u - b)$ to the unit ball to obtain ξ, then set $w* = u - \frac{\xi}{\tau}$.

2.2 Maximum Eigenvalue

Let $n = m^2$ and let $h(w)$ denote the maximum eigenvalue of the matrix reshape$(w) := W$. Rewriting $\text{prox}_h(\tau, u)$ in its matrix equivalent representation, we have

$$\min_{w \in \mathbb{R}^{m^2}} h(w) + \frac{\tau}{2} ||w - u||^2 = \min_{w \in \mathbb{R}^{m \times m}} \Lambda(W) + \frac{\tau}{2} ||W - U||_F^2$$

where $\Lambda(W)$ denotes the largest eigenvalue of W, X is the matrix corresponding to the reshaped vector $u \in \mathbb{R}^{m^2}$, and $||\cdot||_F$ is the Frobenius norm. Since

$$\Lambda(W) = \max_{A \succeq 0, \text{tr}(A) = 1} \langle W, A \rangle,$$

we have

$$\begin{aligned}
\operatorname*{argmin}_{W \in \mathbb{R}^{m \times m}} \Lambda(W) + \frac{\tau}{2} ||W - U||_F^2 &= \operatorname*{argmin}_{W \in \mathbb{R}^{m \times m}} \operatorname*{argmax}_{A \succeq 0, \text{tr}(A) = 1} \langle W, A \rangle + \frac{\tau}{2} ||W - U||_F^2 \\
&= \operatorname*{argmax}_{A \succeq 0, \text{tr}(A) = 1} \operatorname*{argmin}_{W \in \mathbb{R}^{m \times m}} \langle W, A \rangle + \frac{\tau}{2} ||W - U||_F^2 .
\end{aligned}$$

Continuing as before, the inner minimization problem has solution W^* satisfying $A + \tau(W^* - U) = 0$. Thus, it suffices to solve

$$\begin{aligned}\underset{A\succeq 0,\text{tr}(A)=1}{\text{argmax}} \langle U - \frac{A}{\tau}, A\rangle + \frac{\tau}{2}\left|\left|U - \frac{A}{\tau} - U\right|\right|_F^2 &= \underset{A\succeq 0,\text{tr}(A)=1}{\text{argmax}} \langle U, A\rangle - \frac{1}{2\tau}\,||A||_F^2 \\ &= \underset{A\succeq 0,\text{tr}(A)=1}{\text{argmin}} ||A||_F^2 - 2\langle U\tau, A\rangle \\ &= \underset{A\succeq 0,\text{tr}(A)=1}{\text{argmin}} ||A - U\tau||_F^2 .\end{aligned}$$

Our solution w to the original problem is then obtained by projecting τU to the standard spectrahedron to get A, setting $W^* = U - \frac{A}{\tau}$ and then vectorizing W^* to get w^*.

3 Optimizing a Linear Objective

In projection free algorithms, it is often the case that we require a solution to a linear objective optimization problem over our feasible set. That is, we seek a solution to

$$x^* \in \underset{x\in S}{\text{Argmin}}\langle c, x\rangle$$

for some convex set S and cost vector c. We will discuss how to do this for various feasible sets.

3.1 Standard Spectrahedron

Reshaping the linear objective for matrix problems, linear objective over the standard spectrahedron is as follows: for any nondefective matrix $C \in \mathbb{R}^{n\times n}$ find X^* such that

$$X^* \in \underset{X\succeq 0,\text{tr}(X)=1}{\text{Argmin}} \; C \bullet X$$

where $\bullet$ represents the Frobenius inner product. Without loss of generality, we may assume that C is symmetric. Recall that

$$\min_{X\succeq 0,\text{tr}(X)=1} C \bullet X = \lambda(C)$$

where $\lambda(C)$ is the minimum eigenvalue of C. Thus, it suffices to find an $X \in \text{Spe}_n := \{X : X \succeq 0, \text{tr}(X) = 1\}$ such that $C \bullet X = \text{tr}(CX) = \lambda(C)$. Let C have eigendecomposition $U^{-1}\Lambda U$, and k be an index such that $\Lambda_{kk} = \lambda(C)$. Then choosing $X^* = U^{-1}I_kU$ where $I_k = \text{diag}(e_k)$ implies that

$$C \bullet X^* = \text{tr}(U^{-1}\Lambda UU^{-1}I_kU) = \text{tr}(U^{-1}\Lambda I_kU) = \lambda(C)$$

since $U^{-1}\Lambda I_kU$ has eigenvalues 0 with multiplicity $n - 1$ and $\lambda(C)$ with multiplicity 1. Also note that since C is symmetric, it admits an orthogonal eigendecomposition, i.e. $U^{-1} = U^T$. Thus,

$$X^* = U^{-1}I_kU = U^TI_kU = u_ku_k^T.$$

Consequently, our linear optimization only requires us to find the eigenvector corresponding to the eigenvalue of smallest magnitude.

3.2 Birkhoff Polytopes

We want to efficiently compute a solution to

$$\min_{X \in B_n} C \bullet X$$

where C is an $n \times n$ matrix, and the Birkhoff polytope, denoted B_n, is the set of doubly stochastic matrices, i.e.

$$B_n := \{P \in \mathbb{R}^{n \times n} \mid P^T e = e, Pe = e, P \geq 0\}.$$

Here, e denotes the n-dimensional vector of 1's and $P \geq 0$ implies that each entry of P is nonnegative. By the Birkhoff-von Neumann theorem, we know that $B_n = \text{conv}(S)$ where S is the set of projection matrices. Since we are minimizing a linear objective over a convex hull, we know that one of the extreme points must be an optimal solution. Thus, it suffices to find a permutation matrix P such that $C \bullet P$ is as small as possible. In particular, we must choose n indices with no overlapping rows or columns such that the sum of the entries of C of these indices is a minimum. If we view C as a cost matrix, this is a variant of the assignment problem[1]. There are different algorithms to solve this assignment problem, most notably the Hungarian algorithm which runs in $\mathcal{O}(n^3)$ time. Thankfully, MATLAB 2019a has a function, matchpairs, which solves this problem. This function requires a cost matrix and an unmatched cost as input. To avoid returning any unmatched tasks (which would result in a singular "permutation" matrix), simply set the unmatched cost larger than the largest value in C.

[1] In fact, the Birkhoff polytope is also called the assignment polytope.

References