Constructing Diversification Constraints

R. Scott McIntire

Oct 1, 2024

1 Overview

Consider the function function, $f : \mathbf{R}^n \mapsto \mathbf{R}$, defined by

$$f(\mathbf{x}) \;=\; \sum_{i=1}^{k} x_{[i]} \tag{1}$$

From this we see that $f(\mathbf{x})$ is the value of the sum of the k^{th} largest values in its input $\mathbf{x}$. We are interested in optimization problems involving a vector $\mathbf{x}$ with the sum of the top values constrained by a given value; that is, such that $f(x) \leq M$ for some value M. How could we do this? One way is to write down all possible combinations of the k elements of $\mathbf{X}$ and write a constraint that bounds their sum to be less than or equal to M. But the number of constraints that one has to write are $\binom{n}{k}$. This becomes large very quickly. In the next section we seek a way to represent f to reduce the number of constraints.

2 The Upper Bound as an Optimization Problem

Another way to approach a bound on $f(\mathbf{x})$ is to bound its maximum value. Its maximum value can be described by:

$$\max_{\mathbf{y} \in \mathbf{Z}_2^n} \quad \mathbf{y}^T\mathbf{x} \tag{2}$$

$$\mathbf{y}^T\mathbf{1} \;=\; k \tag{3}$$

In English this says: "Take the maximum value of all possible sums of k values from $\mathbf{x}$." Although this optimization is succinct, there are still an exponential number of combinations to examine to find the optimal solution in this discrete setting.

claim: The solution to the above is the same as:

$$\max_{\mathbf{y}\in\mathbf{R}^n} \quad \mathbf{y}^T\mathbf{x} \tag{4}$$
$$\mathbf{0} \preceq \mathbf{y} \preceq \mathbf{1} \tag{5}$$
$$\tag{6}$$

NOTE: We do not prove this claim, asking the reader to accept the result.

This is no longer a discrete problem, it is a continuous optimization problem. The solution to this problem is the same the associated problem:

$$\min_{\mathbf{y}\in\mathbf{R}^n} \quad -\mathbf{y}^T\mathbf{x} \tag{7}$$
$$\mathbf{0} \preceq \mathbf{y} \preceq \mathbf{1} \tag{8}$$
$$\mathbf{y}^T\mathbf{1} = k \tag{9}$$

But this problem is a *convex* problem.

3 A Dual Description of the Optimization

Since the problem described by equations $(7, 8, 9)$, is a *convex* problem, the value of its solution is the same as the value of its associated *dual* problem.[1]

To form the dual problem we need the Lagrangian, which is:

$$L(\mathbf{y}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \nu) \quad = \quad -\mathbf{y}^T\mathbf{x} - \boldsymbol{\lambda}_1\mathbf{y} + \boldsymbol{\lambda}_2(\mathbf{y} - \mathbf{1}) + \nu(k - \mathbf{y}^T\mathbf{1}) \tag{10}$$

Set the function g:

$$g(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_1, \nu) \quad = \quad \inf_{\mathbf{y}} \ L(\mathbf{y}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \nu) \tag{11}$$

[1]Normally one needs to show that a convex problem satisfies Slater's condition. But this is not necessary when dealing with linear convex problems.

Substituting for L this becomes

$$g(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_1, \nu) \quad = \quad \inf_{\mathbf{y}} \; \left(\mathbf{y}^T \left(-\mathbf{x} - \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 - \nu \mathbf{1}\right) - \boldsymbol{\lambda}_2^T \mathbf{1} + \nu\, k\right) \tag{12}$$

The dual problem is then

$$\max_{\substack{\boldsymbol{\lambda}_2 \succeq \mathbf{0} \\ \boldsymbol{\lambda}_1 \succeq \mathbf{0} \\ \nu}} g(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \nu) \tag{13}$$

Which is[2]

$$\max_{\substack{\boldsymbol{\lambda}_2 \succeq \mathbf{0} \\ \boldsymbol{\lambda}_1 \succeq \mathbf{0} \\ \nu}} -\boldsymbol{\lambda}_2^T \mathbf{1} + \nu\, k \tag{14}$$

$$-\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 - \nu \mathbf{1} \quad = \quad \mathbf{x} \tag{15}$$

This is equivalent to[3]

$$\max_{\substack{\boldsymbol{\lambda} \succeq \mathbf{0} \\ \nu}} \quad -\boldsymbol{\lambda}^T \mathbf{1} + \nu\, k \tag{16}$$

$$\boldsymbol{\lambda} - \nu \mathbf{1} \succeq \mathbf{x} \tag{17}$$

Or,

$$\max_{\substack{\boldsymbol{\lambda} \succeq \mathbf{0} \\ \nu}} \qquad \nu\, k - \mathbf{1}^T \boldsymbol{\lambda} \tag{18}$$

$$\mathbf{x} \preceq \boldsymbol{\lambda} - \nu \mathbf{1} \tag{19}$$

Since maximizing over ν or $-\nu$ is the same and there are no restrictions on the sign of ν we may replace ν with $-\nu$ in the last equations giving:

$$\max_{\substack{\boldsymbol{\lambda} \succeq \mathbf{0} \\ \nu}} \qquad -\nu\, k - \mathbf{1}^T \boldsymbol{\lambda} \tag{20}$$

$$\mathbf{x} \preceq \boldsymbol{\lambda} + \nu \mathbf{1} \tag{21}$$

[2]Note that g is $-\infty$, unless the term that $\mathbf{y}$ is "dotting" is the zero vector. Consequently, the maximum must necessarily occur where the dotting vector is zero.

[3]We can remove n variables by eliminating $\boldsymbol{\lambda}_1$ from the equations while keeping the same number of inequalities. We can do this by realizing that $\boldsymbol{\lambda}_2 - \nu \mathbf{1} = \mathbf{x} + \boldsymbol{\lambda}_1$ expresses the same information as: $\boldsymbol{\lambda}_2 - \nu \mathbf{1} \succeq \mathbf{x}$. Since there is now only one $\boldsymbol{\lambda}$, we relabel $\boldsymbol{\lambda}_2$ as $\boldsymbol{\lambda}$.

But this is the same as:

$$\min_{\boldsymbol{\lambda},\nu} \quad \nu\, k + \mathbf{1}^T \boldsymbol{\lambda} \tag{22}$$

$$\mathbf{x} \preceq \boldsymbol{\lambda} + \nu \mathbf{1} \tag{23}$$

$$\boldsymbol{\lambda} \succeq \mathbf{0} \tag{24}$$

4 A Linear Number of Constraints

Therefore, if you wish to bound the top k elements of the vector $\mathbf{x}$ by M in an optimization problem; that is, if you wish to bound $f(x)$ above by M, you need to add the following constraints:

$$\nu\, k + \boldsymbol{\lambda}^T \mathbf{1} \ \leq \ M \tag{25}$$

$$\mathbf{x} \ \preceq \ \boldsymbol{\lambda} + \nu \mathbf{1} \tag{26}$$

$$\boldsymbol{\lambda} \ \succeq \ \mathbf{0} \tag{27}$$

Why? Because the expression $\nu\, k + \boldsymbol{\lambda}^T \mathbf{1}$ with the constraints (26) and (27) applied is *always* an upper bound to $f(x)$. Consequently, we have the inequality: $f(x) \leq \nu\, k + \boldsymbol{\lambda}^T \mathbf{1} \leq M$. The only concern is that there is a gap between the values of $\nu\, k + \boldsymbol{\lambda}^T \mathbf{1}$ and $f(x)$ – making (25) too restrictive. But this is not the case as the minimum over all $\boldsymbol{\lambda}$ and ν (subject to (26) and (27))is $f(x)$.

Therefore, in order to avoid a combinatorial explosion of inequality constraints, one need only add $(n+1)$ variables $(\boldsymbol{\lambda}, \nu)$ to an optimization problem to provide diversification constraints on a vector of length, n. The number of new inequality constraints added becomes $(2n+1)$.