Subgradient Optimization, Generalized Programming, and Nonconvex Duality

Robert M. Freund

May, 2004

1 Subgradient Optimization

1.1 Review of Subgradients

Recall the following facts about subgradients of convex functions. Let $S \subset \mathbb{R}^n$ be a given nonempty convex set, and let $f(\cdot) : S \to \mathbb{R}$ be a convex function. Then then $\xi \in \mathbb{R}^n$ is a subgradient of $f(\cdot)$ at $\bar{x} \in S$ if

$$f(x) \geq f(\bar{x}) + \xi^T(x - \bar{x}) \text{ for all } x \in S \ .$$

If $\bar{x} \in \text{int}S$, then there exists a subgradient of $f(\cdot)$ at $\bar{x}$. The collection of subgradients of $f(\cdot)$ at x is denoted by $\partial f(x)$, and the operator $\partial f(\cdot)$ is called the subdifferential of $f(\cdot)$. Recall that if $x \in \text{int}S$, then $\partial f(x)$ is a nonempty closed convex set.

If $f(\cdot)$ is differentiable at $x = \bar{x}$ and $\nabla f(\bar{x}) \neq 0$, then $-\nabla f(\bar{x})$ is a descent direction of $f(\cdot)$ at $\bar{x}$. However, if $f(\cdot)$ is not differentiable at $x = \bar{x}$ and ξ is a subgradient, then $-\xi$ is not necessarily a descent direction of $f(\cdot)$ at $\bar{x}$.

1.2 Computing a Subgradient

Subgradients play a very important role in algorithms for non-differentiable optimization. In these algorithms, we typically have a subroutine that receives as input a value x, and has output ξ where ξ is some subgradient of $f(x)$.

1.3 The Subgradient Method for Minimizing a Convex Function

Suppose that $f(\cdot)$ is a convex function, and that we seek to solve:

$$\begin{array}{rll} P: & z^* = \text{minimize}_x & f(x) \\ \\ & \quad \text{s.t.} & x \in \mathbb{R}^n \ . \end{array}$$

The following algorithm generalizes the steepest descent algorithm and can be used to minimize a nondifferentiable convex function $f(x)$.

Subgradient Method

> **Step 0: Initialization.** Start with any point $x^1 \in \mathbb{R}^n$. Choose an infinite sequence of positive step-size values $\{\alpha_k\}_{k=1}^{\infty}$. Set $k = 1$.
>
> **Step 1: Compute a subgradient.** Compute $\xi \in \partial f(x^k)$. If $\xi = 0$, STOP. x^k solves P.
>
> **Step 2: Compute step-size.** Compute step-size α_k from the step-size series.
>
> **Step 3: Update Iterate.** Set $x^{k+1} \leftarrow x^k - \alpha_k \frac{\xi}{\|\xi\|}$. Set $k \leftarrow k + 1$ and go to **Step 1**.

Note in this algorithm that the step-size α_k at each iteration is determined without a line-search, and in fact is predetermined in Step 0. One reason for this is that a line-search might not be worthwhile, since $-\xi$ is not necessarily a descent direction. As it turns out, the viability of the subgradient method depends critically on the sequence of step-sizes:

Theorem 1 *Suppose that $f(\cdot)$ is a convex function whose domain $D \subset \mathbb{R}^n$ satisfies $\mathrm{int} D \neq \emptyset$. Suppose that $\{\alpha_k\}_{k=1}^{\infty}$ satisfies:*

$$\lim_{k\to\infty} \alpha_k = 0 \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha_k = \infty \ .$$

Let $x^1, x^2, \ldots,$ be the iterates generated by the subgradient method. Then

$$\inf_k f(x^k) = z^* \ .$$

Proof: Suppose that the result is not true. Then there exists $\epsilon > 0$ such that $f(x^k) \geq z^* + \epsilon$ for all $k = 1, \ldots$. Let $T = \{x \in \mathbb{R}^n \mid f(x) \leq z^* + \epsilon\}$. Then there exists $\hat{x}$ and $\rho > 0$ for which $B(\hat{x}, \rho) \subset T$, where $B(\hat{x}, \rho) := \{x \in$

$\mathbb{R}^n \mid \|x - \hat{x}\| \leq \rho\}$. Let ξ^k be the subgradient chosen by the subgradient method for the iterate x^k. By the subgradient inequality we have for all $k = 1, \ldots$:

$$f(x^k) \geq z^* + \epsilon \geq f\left(\hat{x} + \rho \frac{\xi^k}{\|\xi^k\|}\right) \geq f(x^k) + (\xi^k)^T \left(\hat{x} + \rho \frac{\xi^k}{\|\xi^k\|}) - x^k\right) \ ,$$

which upon rearranging yields:

$$(\xi^k)^T(\hat{x} - x^k) \leq -\rho \frac{(\xi^k)^T \xi^k}{\|\xi^k\|} = -\rho \|\xi^k\| \ .$$

We also have for each k:

$$\begin{aligned}
\|x^{k+1} - \hat{x}\|^2 &= \|x^k - \alpha_k \tfrac{\xi^k}{\|\xi^k\|} - \hat{x}\|^2 \\
&= \|x^k - \hat{x}\|^2 + \alpha_k^2 - 2\alpha_k \tfrac{(\xi^k)^T(x^k - \hat{x})}{\|\xi^k\|} \\
&\leq \|x^k - \hat{x}\|^2 + \alpha_k^2 - 2\alpha_k \rho \\
&= \|x^k - \hat{x}\|^2 + \alpha_k(\alpha_k - 2\rho) \ .
\end{aligned}$$

For k sufficiently large, say for all $k \geq K$, we have $\alpha_k \leq \rho$, whereby:

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 - \rho \alpha_k \ .$$

However, this implies by induction that for all $j \geq 1$ we have:

$$\|x^{K+j} - \hat{x}\|^2 \leq \|x^K - \hat{x}\|^2 - \rho \sum_{k=K+1}^{K+j} \alpha_k \ .$$

Now for j sufficiently large the right-hand side expression is negative, since $\sum_{k=1}^{\infty} \alpha_k = \infty$, which yields a contradiction since the left-hand side must be nonnegative. ∎

1.4 The Subgradient Method with Projections

Problem P posed at the start of Subsection 1.3 generalizes to the following problem:

$$P_S: \quad z^* = \text{minimize}_x \quad f(x)$$

$$\text{s.t.} \qquad x \in S \ ,$$

where S is a given closed convex set. We suppose that S is a simple enough set that we can easily compute projections onto S. This means that for any point $c \in \mathbb{R}^n$, we can easily compute:

$$\Pi_S(c) := \arg \min_{x \in S} \|c - x\| \ .$$

The following algorithm is a simple extension of the subgradient method presented in Subsection 1.3, but includes a projection computation so that all iterate values x^k satisfy $x^k \in S$.

Projected Subgradient Method

Step 0: Initialization. Start with any point $x^1 \in S$. Choose an infinite sequence of positive step-size values $\{\alpha_k\}_{k=1}^{\infty}$. Set $k = 1$.

Step 1: Compute a subgradient. Compute $\xi \in \partial f(x^k)$. If $\xi = 0$, STOP. x^k solves P.

Step 2: Compute step-size. Compute step-size α_k from the step-size series.

Step 3: Update Iterate. Set $x^{k+1} \leftarrow \Pi_S\left(x^k - \alpha_k \frac{\xi}{\|\xi\|}\right)$. Set $k \leftarrow k + 1$ and go to **Step 1**.

Similar to Theorem 1, we have:

Theorem 2 *Suppose that $f(\cdot)$ is a convex function whose domain $D \subset \mathbb{R}^n$ satisfies $intD \cap S \neq \emptyset$. Suppose that $\{\alpha_k\}_{k=1}^{\infty}$ satisfies:*

$$\lim_{k \to \infty} \alpha_k = 0 \quad \textit{and} \quad \sum_{k=1}^{\infty} \alpha_k = \infty \ .$$

Let $x^1, x^2, \ldots,$ be the iterates generated by the projected subgradient method. Then

$$\inf_k f(x^k) = z^* \ .$$

The proof of Theorem 2 relies on the following "non-expansive" property of the projection operator $\Pi_S(\cdot)$:

Lemma 3 *Let S be a closed convex set and let $\Pi_S(\cdot)$ be the projection operator onto S. Then for any two vectors $c^1, c^2 \in \mathbb{R}^n$,*

$$\|\Pi_S(c^1) - \Pi_S(c^1)\| \leq \|c^1 - c^2\| \ .$$

Proof: Let $\bar{c}^1 = \Pi_S(c^1)$ and let $\bar{c}^2 = \Pi_S(c^1)$. Then from Theorem 4 of the Constrained Optimization notes (the basic separating hyperplane theorem) we have:

$$(c^1 - \bar{c}^1)^T(x - \bar{c}^1) \leq 0 \quad \text{for all } x \in S \ ,$$

and

$$(c^2 - \bar{c}^2)^T(x - \bar{c}^2) \leq 0 \quad \text{for all } x \in S \ .$$

In particular, because $\bar{c}^2 \in S$ and $\bar{c}^1 \in S$ it follows that:

$$(c^1 - \bar{c}^1)^T(\bar{c}^2 - \bar{c}^1) \leq 0 \quad \text{and} \quad (c^2 - \bar{c}^2)^T(\bar{c}^1 - \bar{c}^2) \leq 0 \ .$$

Then note that

$$\begin{aligned}
\|c^1 - c^2\|^2 &= \|\bar{c}^1 - \bar{c}^2 + (c^1 - \bar{c}^1 - c^2 + \bar{c}^2)\|^2 \\
&= \|\bar{c}^1 - \bar{c}^2\|^2 + \|c^1 - \bar{c}^1 - c^2 + \bar{c}^2\|^2 + 2(\bar{c}^1 - \bar{c}^2)^T(c^1 - \bar{c}^1 - c^2 + \bar{c}^2) \\
&\geq \|\bar{c}^1 - \bar{c}^2\|^2 + 2(\bar{c}^1 - \bar{c}^2)^T(c^1 - \bar{c}^1) + 2(\bar{c}^1 - \bar{c}^2)^T(-c^2 + \bar{c}^2) \\
&\geq \|\bar{c}^1 - \bar{c}^2\|^2 \ ,
\end{aligned}$$

from which it follows that $\|c^1 - c^2\| \geq \|\bar{c}^1 - \bar{c}^2\|$. ■

The proof of Theorem 2 can easily be constructed by using Lemma 3 and by following the logic used in the proof of Theorem 1, and is left as an Exercise.

1.5 Solving the Lagrange Dual via the Subgradient Method

We start with the primal problem:

$$
\begin{array}{llrll}
\mathrm{OP}: & z^* = \text{minimum}_x & f(x) & & \\
 & \text{s.t.} & g_i(x) & \leq & 0, \quad i = 1, \ldots, m \\
 & & x \in X \ . & &
\end{array}
$$

We create the Lagrangian

$$L(x, u) := f(x) + u^T g(x) \ .$$

The dual function is given by:

$$
\begin{array}{rll}
L^*(u) := & \text{minimum}_x & f(x) + u^T g(x) \\
 & \text{s.t.} & x \in X \ .
\end{array}
$$

The dual problem is:

$$
\begin{array}{lll}
\mathrm{D}: & v^* = \text{maximum}_u & L^*(u) \\
 & \text{s.t.} & u \geq 0 \ .
\end{array}
$$

Recall that $L^*(u)$ is a concave function. For concave functions we work with *supergradients*. If $f(\cdot)$ is a concave function whose domain is the convex set S, then $g \in \mathbb{R}^n$ is a supergradient of $f(\cdot)$ at $\bar{x} \in S$ if

$$f(x) \leq f(\bar{x}) + g^T(x - \bar{x}) \quad \text{for all } x \in S \ .$$

The premise of Lagrangian duality is that it is "easy" to compute $L^*(\bar{u})$ for any given $\bar{u}$. That is, it is easy to compute an optimal solution $\bar{x} \in X$ of

$$L^*(\bar{u}) := \ \text{minimum}_{x \in X} \ \ f(x) + \bar{u}^T g(x) \ \ = \ \ f(\bar{x}) + \bar{u}^T g(\bar{x}) \ ,$$

for any given $\bar{u}$. It turns out that computing supergradients of $L^*(\cdot)$ is then also easy. We have:

Proposition 4 *Suppose that $\bar{u}$ is given and that $\bar{x} \in X$ is an optimal solution of $L^*(\bar{u}) = \min_{x \in X} f(x) + \bar{u}^T g(x)$. Then $g := g(\bar{x})$ is a supergradient of $L^*(\cdot)$ at $u = \bar{u}$.*

Proof: For any $u \geq 0$ we have

$$\begin{aligned} L^*(u) = & \quad \min_{x \in X} f(x) + u^T g(x) \\ \leq & \quad f(\bar{x}) + u^T g(\bar{x}) \\ = & \quad f(\bar{x}) + \bar{u}^T g(\bar{x}) + (u - \bar{u})^T g(\bar{x}) \\ = & \quad \min_{x \in X} f(x) + \bar{u}^T g(x) + g(\bar{x})^T (u - \bar{u}) \\ = & \quad L^*(\bar{u}) + g^T (u - \bar{u}) \ . \end{aligned}$$

Therefore g is a supergradient of $L^*(\cdot)$ at $\bar{u}$.
q.e.d.

The Lagrange dual problem D is in the same format as problem P_S of Subsection 1.4, with $S = \mathbb{R}^m_+$. In order to apply the projected subgradient method to this problem, we need to be able to conveniently compute the projection of any vector $v \in \mathbb{R}^m$ onto $S = \mathbb{R}^m_+$. This indeed is easy. Let $u \in \mathbb{R}^n$ be given, and define u^+ to be the vector each of whose components is the positive part of the respective component of v. For example, if $u = (2, -3, 0, 1, -5)$, then $u^+ = (2, 0, 0, 1, 0)$. Then it is easy to see that $\Pi_S(u) = u^+$. We can now state the subgradient method for solving the Lagrange dual problem:

Subgradient Method for Solving the Lagrange Dual Problem

Step 0: Initialization. Start with any point $u^1 \in \mathbb{R}^n_+$. Choose an infinite sequence of positive step-size values $\{\alpha_k\}_{k=1}^{\infty}$. Set $k = 1$.

Step 1: Compute a supergradient. Solve for an optimal solution $\bar{x}$ of $L^*(u^k) = \min_{x \in X} f(x) + (u^k)^T g(x)$. Set $g := g(\bar{x})$. If $g = 0$, STOP. x^k solves D.

Step 2: Compute step-size. Compute step-size α_k from step-size series.

Step 3: Update Iterate. Set $u^{k+1} \leftarrow \left(u^k + \alpha_k \frac{g}{\|g\|}\right)^+$. Set $k \leftarrow k+1$ and go to **Step 1**.

Notice in Step 3 that the "$(u)^+$" operation is simply the projection of u onto the nonnegative orthant $\mathbb{R}^n_+$.

2 Generalized Programming and Nonconvex Duality

2.1 Geometry of nonconvex duality and the equivalence of convexification and dualization

We start with the primal problem:

$$\begin{array}{llll} \text{OP}: & z^* = \text{minimum}_x & f(x) & \\ & \text{s.t.} & g_i(x) & \leq 0, \quad i = 1, \ldots, m \\ & & x \in X \ . & \end{array}$$

We create the Lagrangian

$$L(x, u) := f(x) + u^T g(x) \ .$$

The dual function is given by:

$$\begin{array}{lll} L^*(u) := & \text{minimum}_x & f(x) + u^T g(x) \\ & \text{s.t.} & x \in X \ . \end{array}$$

The dual problem is:

$$\begin{array}{rlll} \mathrm{D}: & v^* = \mathrm{maximum}_u & L^*(u) \\ & \text{s.t.} & u \geq 0 \ . \end{array}$$

Herein we will assume that $f(\cdot)$ and $g_1(\cdot), \ldots, g_m(\cdot)$ are continuous and X is a compact set. We will *not* assume any convexity.

Recall the definition of I:

$$I := \left\{ (s, z) \in \mathbb{R}^{m+1} \mid \text{there exists } x \in X \text{ for which } s \geq g(x) \text{ and } z \geq f(x) \right\} \ .$$

Let C be the convex hull of I. That is, C is the smallest convex set containing I. From the assumption that X is compact and that $f(\cdot)$ and $g_1(\cdot), \ldots, g_m(\cdot)$ are continuous, it follows that C is a closed convex set. Let

$$\hat{z} := \min\{z : (0, z) \in C\} \ ,$$

i.e., $\hat{z}$ is the optimal value of the *convexified* primal problem, which has been convexified in the column geometry $\begin{pmatrix} s \\ z \end{pmatrix}$ of resources and costs.

Let

$$H(u) := \{(s, z) \in \mathbb{R}^{m+1} : u^T s + z = L^*(u)\} \ .$$

Then we say that $H(u)$ *supports* I (or C) if $u^T s + z \geq L^*(u)$ for every $(s, z) \in I$ (or C), and $u^T s + z = L^*(u)$ for some $(s, z) \in I$ (or C).

Lemma 5 *If $u \geq 0$, then $H(u)$ supports I and C.*

Proof: Let $(s, z) \in I$. Then $u^T s + z \geq u^T g(x) + f(x)$ for some $x \in X$. Thus $u^T s + z \geq \inf_{x \in X} u^T g(x) + f(x) = L^*(u)$. Furthermore, setting $\bar{x} = \arg\ \min_{x \in X}\{f(x) + u^T g(x)\}$ and $(s, z) = (g(\bar{x}), f(\bar{x}))$, we have $u^T s + z = L^*(u)$. Thus $H(u)$ supports I. But then $H(u)$ also supports C, since C is the convex hull of I. ∎

Lemma 6 (Weak Duality) $\hat{z} \geq v^*$.

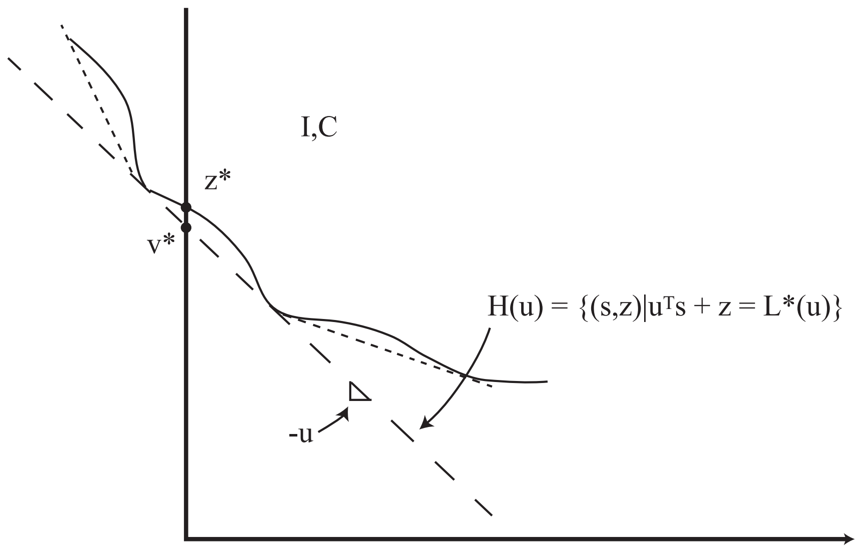

Figure 1: I, C, and $H(u)$.

Proof: If $\hat{z} = +\infty$, then we are done. Otherwise, let $(0, z) \in C$. For any $u \geq 0$, $H(u)$ supports I from Lemma 5. Therefore $L^*(u) \leq u^T 0 + z = z$. Thus

$$v^* = \sup_{u \geq 0} L^*(u) \leq \inf_{(0,z) \in C} z = \hat{z}. \blacksquare$$

Lemma 7 (Strong Duality) $\hat{z} = v^*$.

Proof: In view of Lemma 6, it suffices to show that $\hat{z} \leq v^*$. If $\hat{z} = -\infty$, we are done. If not, let $r < \hat{z}$ be given. Then $(0, r) \notin C$, and so there is a hyperplane separating $(0, r)$ from C. Thus there exists $(\bar{u}, \alpha) \neq (0, 0)$ and θ such that $\bar{u}^T 0 + \alpha r < \theta$ and $\bar{u}^T s + \alpha z \geq \theta$ for all $(s, z) \in C$. This immediately implies $\bar{u} \geq 0$, $\alpha \geq 0$. If $\alpha \neq 0$, we can re-scale so that $\alpha = 1$. Then

$$\bar{u}^T s + z \geq \theta > r$$

for all $(s, z) \in C$. In particular,

$$\bar{u}^T g(x) + f(x) \geq \theta > r \quad \text{ for all } x \in X,$$

which implies that $L^*(\bar{u}) = \inf_{x \in X} \left\{ \bar{u}^T g(x) + f(x) \right\} > r$. Since r is an arbitrary value with $r < \hat{z}$, we have $v^* \geq L^*(\bar{u}) \geq \hat{z}$.

It remains to analyze the case when $\alpha = 0$. In this case we have $\theta > 0$ and $\bar{u}^T s \geq \theta > 0$ for all $(s, z) \in C$. With $(s, z) = (g(x), f(x))$ for a given $x \in X$, we have for all $\lambda \geq 0$:

$$L^*(u) + \lambda\theta \leq f(x) + u^T g(x) + \lambda \bar{u}^T g(x) = f(x) + (u + \lambda \bar{u})^T g(x) \ .$$

Then

$$L^*(u) + \lambda\theta \leq \inf_{x \in X} \{ f(x) + (u + \lambda \bar{u})^T g(x) \} = L^*(u + \lambda \bar{u}) \ .$$

Since $\theta > 0$ and λ was any nonnegative scalar, $L^*(u + \lambda \bar{u}) \to +\infty$ as $\lambda \to \infty$, and so $v^* \geq L^*(u + \lambda \bar{u})$ implies $v^* = +\infty$. Thus, $v^* \geq \hat{z}$. ∎

2.2 The Generalized Programming Algorithm

Consider the following algorithm:

Generalized Programming Algorithm for Solving the Lagrange Dual Problem

Step 0 $E^k = \{x^1, \ldots, x^k\}$, LB $= -\infty$, UB $= +\infty$.

Step 1 Solve the following linear program (values in brackets are the dual variables):

$$\begin{array}{lll} (\mathrm{LP}^k): & \min_\lambda & \sum_{i=1}^k \lambda_i f(x^i) \\ \\ & \text{s.t.} & \sum_{i=1}^k \lambda_i g(x^i) \leq 0 \quad (u) \\ & & \sum_{i=1}^k \lambda_i = 1 \qquad\quad\ (\theta) \\ & & \lambda \geq 0 \ , \end{array}$$

for λ^k, u^k, θ^k, and also define:

$$\tilde{x}^k := \sum_{i=1}^{k} \lambda_i^k x^i, \quad \tilde{s}^k := \sum_{i=1}^{k} \lambda_i^k g(x^i), \quad \tilde{z}^k := \sum_{i=1}^{k} \lambda_i^k f(x^i) = \theta^k.$$

Step 2 (Dual function evaluation.) Solve:

$$(\mathrm{D}^k): \quad L^*(u^k) = \min_{x \in X} \{f(x) + (u^k)^T g(x)\}$$

for x^{k+1}.

Step 3 $\mathrm{UB} := \min\{\mathrm{UB}, \tilde{z}^k = \theta^k\}$, $\mathrm{LB} := \max\{\mathrm{LB}, L^*(u^k)\}$.

If $\mathrm{UB} - \mathrm{LB} \le \epsilon$, stop. Otherwise, go to Step 4.

Step 4 $E^{k+1} := E^k \cup \{x^{k+1}\}$, $k := k+1$, go to Step 1.

Notice that the linear programming dual of (LP^k) is:

$$\begin{array}{lll} (\mathrm{DP}^k): & \max_{u,\theta} & \theta \\ & & \\ & \text{s.t.} & -u^T g(x^i) + \theta \le f(x^i) \quad i = 1, \ldots, k \\ & & u \ge 0 \ , \end{array}$$

which equivalently is:

$$\begin{array}{lll} (\mathrm{DP}^k): & \max_{u,\theta} & \theta \\ & & \\ & \text{s.t.} & \theta \le f(x^i) + u^T g(x^i) \quad i = 1, \ldots, k \\ & & u \ge 0 \ . \end{array}$$

This can be re-written as:

$$(\mathrm{DP}^k): \quad \max_{u \ge 0} \ \min_{x \in E^k} \{f(x) + u^T g(x)\} \ .$$

Note that (u^{k+1}, θ^{k+1}) is always feasible in (DP^k) and $(u, \theta) = (u, L^*(u))$ is always feasible for (DP^k) for $u \ge 0$.

Geometrically, the generalized programming algorithm can be viewed as an "inner convexification" process for the primal (see Figure 2), and as an "outer convexification" process via supporting hyperplanes for the dual problem (see Figure 2).

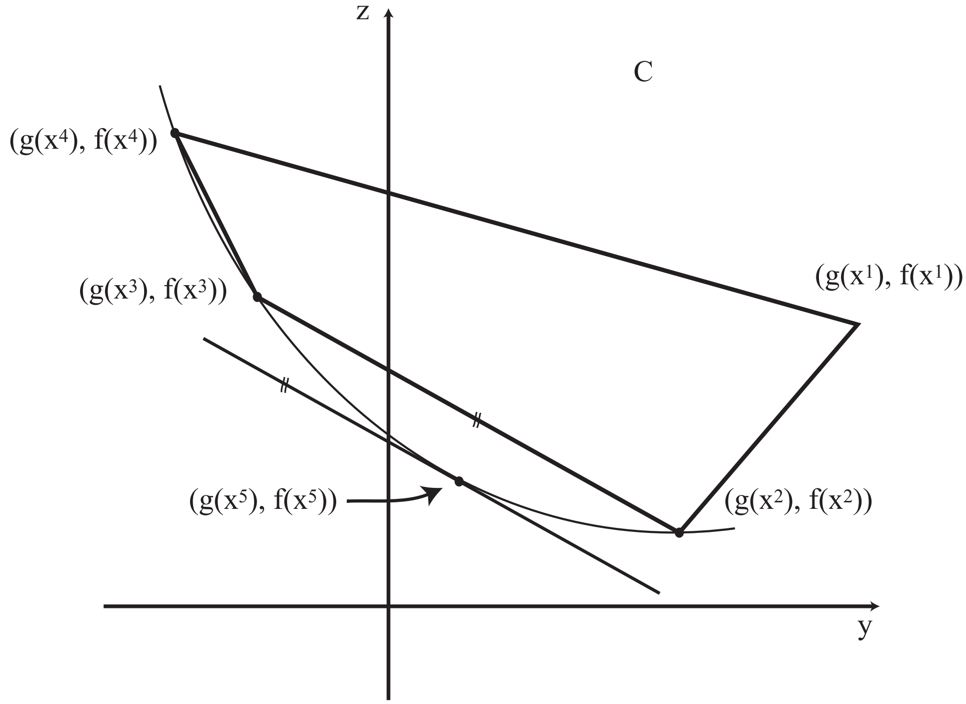

Figure 2: The primal geometry of the generalized programming algorithm.

Proposition 8 *(i)* θ^k *are decreasing in* k

(ii) $(\tilde{s}^k, \tilde{z}^k) \in C$

(iii) u^k *is feasible for (D)*

(iv) $L^*(u^k) \leq v^* = \hat{z} \leq \tilde{z}^k = \theta^k$.

(v) If $f(\cdot)$ *and* $g_1(\cdot), \ldots, g_m(\cdot)$ *are convex functions and* X *is a convex set, then* $\tilde{x}^k$ *is feasible for (P) and* $\hat{z} = z^* \leq f(\tilde{x}^k) \leq \theta^k$*, where* z^* *is the optimal value of (P).*

Proof:

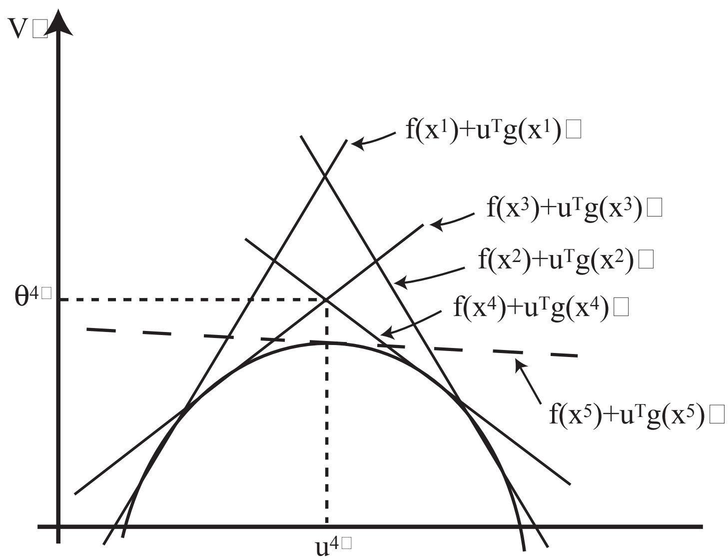

Figure 3: The dual geometry of the generalized programming algorithm.

(i) follows since (DP^{k+1}) has one extra constraint compared to (DP^k), so $\theta^{k+1} \leq \theta^k$.

(ii) follows since $(\tilde{s}^k, \tilde{z}^k)$ is in the convex hull of I, which is C.

(iii) follows since by definition, $u^k \geq 0$.

(iv) $L^*(u^k) \leq v^*$ by definition of v^*, and $v^* = \hat{z}$ by Lemma 7. Since $(\tilde{s}^k, \tilde{z}^k) \in C$, and $\tilde{s}^k \leq 0$, $(0, \tilde{z}^k) \in C$, and so $\hat{z} \leq \tilde{z}^k$. Furthermore, $z^k = \theta^k$ follows from linear programming duality.

(v) Since each $x^i \in X$, $i = 1, \ldots, k$, then $\tilde{x}^k \in X$, and $f(\tilde{x}^k) \leq \sum_{i=1}^k \lambda_i^k f(x^i) \leq \theta^k$, and $g(\tilde{x}^k) \leq \sum_{i=1}^k \lambda_i^k g(x^i) \leq 0$. Thus, $\tilde{x}^k$ is feasible for P and so $z^* \leq f(\tilde{x}^k)$. ∎

Theorem 9 (Convergence of the generalized programming algorithm) *Suppose $u^1, u^2, \ldots,$ are the iterate values of u computed by the generalized programming algorithm, and suppose that there is a convergent subsequence* *of $u^1, u^2, \ldots,$ converging to u^*. Then*

1. *u^* solves (D), and*
2. *$\lim_{k\to\infty} \theta^k = \hat{z} = v^*$.*

Proof: For any $j \le k$ we have

$$f(x^j) + (u^k)^T g(x^j) \ge \theta^k \ge v^* \ .$$

Taking limits as $k \to \infty$, we obtain:

$$f(x^j) + (u^*)^T g(x^j) \ge \bar{\theta} \ge v^*,$$

where $\bar{\theta} := \lim_{k\to\infty} \theta^k$. (This limit exists since θ^k are a monotone decreasing sequence bounded below by v^*.) Since $g_i(\cdot)$ is continuous and X is compact, there exists $B > 0$ for which $|g_i(x)| \le B$ for all $x \in X$, $i = 1, \ldots, m$. Thus

$$|L(x^{k+1}, u^k) - L(x^{k+1}, u^*)| = |(u^k - u^*)^T g(x^{k+1})| \le B \sum_{i=1}^{m} |u_i^k - u_i^*| \ .$$

For any $\epsilon > 0$ and for k sufficiently large, the RHS is bounded above by ϵ. Thus

$$L^*(u^k) = L(x^{k+1},\, u^k) \ge L(x^{k+1},\, u^*) - \epsilon = f(x^{k+1}) + (u^*)^T g(x^{k+1}) - \epsilon \ge \bar{\theta} - \epsilon \ .$$

Thus in the limit $L^*(u^*) \ge \bar{\theta} \ge v^* = \hat{z}$. Therefore since $L^*(u^*) \le v^*$, $L^*(u^*) = \bar{\theta} = v^* = \hat{z}$. ∎

Corollary 10 *If OP has a known Slater point x^0 and $x^0 \in E^k$ for some k, then the sequence of u^k are bounded and so have a convergent subsequence.*

Proof: Without loss of generality we can assume that $x^0 \in E^k$ for the initial set E^k. Then for all k we have:

$$-(u^k)^T g(x^0) + \theta^k \le f(x^0) \ ,$$

and $g(x^0) < 0$, from which it follows that

$$0 \le (u^k)_i \le \frac{-\theta^k + f(x^0)}{-g_i(x^0)} \le \frac{-v^* + f(x^0)}{-g_i(x^0)} \ , \quad i = 1, \ldots, m \ .$$

Therefore $u^1, u^2, \ldots,$ lies in a bounded set, whereby there must be a convergent subsequence of $u^1, u^2, \ldots$. ∎

3 Exercises based on Generalized Programming and Subgradient Optimization

1. Consider the primal problem:

$$\begin{array}{lll} \mathrm{OP}: & \mathrm{minimum}_x & c^T x \\ & \mathrm{s.t.} & Ax - b \le 0 \\ & & x \in \{0,1\}^n \ . \end{array}$$

Here $g(x) = Ax - b$ and $P = \{0,1\}^n = \{x \mid x_j = 0 \text{ or } 1, j = 1, \ldots, n\}$.

We create the Lagrangian:

$$L(x, u) := c^T x + u^T (Ax - b)$$

and the dual function:

$$L^*(u) := \quad \mathrm{minimum}_{x \in \{0,1\}^n} \quad c^T x + u^T (Ax - b)$$

The dual problem then is:

$$\begin{array}{lll} \mathrm{D}: & \mathrm{maximum}_u & L^*(u) \\ & \mathrm{s.t.} & u \ge 0 \end{array}$$

Now let us choose $\bar{u} \geq 0$. Notice that an optimal solution $\bar{x}$ of $L^*(\bar{u})$ is:

$$\bar{x}_j = \begin{cases} 0 & \text{if } (c - A^T\bar{u})_j \geq 0 \\ 1 & \text{if } (c - A^T\bar{u})_j \leq 0 \end{cases}$$

for $j = 1, \ldots, n$. Also,

$$L^*(\bar{u}) = c^T\bar{x} + \bar{u}^T(A\bar{x} - b) = -\bar{u}^T b - \sum_{j=1}^{n} \left[(c - A^T\bar{u})_j\right]^- \ .$$

Also

$$g := g(\bar{x}) = A\bar{x} - b$$

is a subgradient of $L^*(\bar{u})$.

Now consider the following data instance of this problem:

$$A = \begin{pmatrix} 7 & -8 \\ -2 & -2 \\ 6 & 5 \\ -5 & 6 \\ 3 & 12 \end{pmatrix} \ , \ b = \begin{pmatrix} 12 \\ -1 \\ 45 \\ 20 \\ 42 \end{pmatrix}$$

and

$$c^T = \begin{pmatrix} -4 & 1 \end{pmatrix} \ .$$

Solve the Lagrange dual problem of this instance using the subgradient algorithm starting at $u^1 = (1, 1, 1, 1, 1)^T$, with the following step-size choices:

- $\alpha_k = \frac{1}{k}$ for $k = 1, \ldots$.
- $\alpha_k = \frac{1}{\sqrt{k}}$ for $k = 1, \ldots$.
- $\alpha_k = 0.2 \times (0.75)^k$ for $k = 1, \ldots$.
- a step-size rule of your own.

2. Prove Theorem 2 of the notes by using Lemma 3 and by following the logic used in the proof of Theorem 1.

3. The generalized programming algorithm assumes that the user is given start vectors $x^1, \ldots, x^k \in X$ for which

$$\sum_{i=1}^{k} \lambda_i g(x^i) \leq 0$$

has a solution for some $\lambda_1 \geq 0, \ldots, \lambda_k \geq 0$ satisfying

$$\sum_{i=1}^{k} \lambda_i = 1 \ .$$

Here we describe an algorithm for finding such a set of start vectors.

Step 0 Choose any k vectors $x^1, \ldots, x^k \in X$.

Step 1 Solve the following linear program, where the values in brackets are the dual variables:

$$\begin{array}{lll} (\mathrm{LP}^k) & \min_{\lambda,\sigma} & \sigma \\ & \text{s.t.} & \sum_{i=1}^{k} \lambda_i g(x^i) - e\sigma \leq 0 \quad (u) \\ & & \sum_{i=1}^{k} \lambda_i = 1 \qquad\qquad\qquad (\omega) \\ & & \lambda \geq 0, \sigma \geq 0 \end{array}$$

for $\lambda^k, \sigma_k, u^k, \omega^k$.
The dual of (LP^k) is (DP^k):

$$\begin{array}{lll} (\mathrm{DP}^k) & \max_{u,\omega} & \omega \\ & \text{s.t.} & \omega \leq u^T g(x^i), \ i = 1, \ldots, k \\ & & e^T u \leq 1 \\ & & u \geq 0 \end{array}$$

Step 2 If $\sigma^k = 0$, STOP. Otherwise, solve the optimization problem:

$$\min_{x \in X} \left(u^k\right)^T g(x) \ ,$$

and let x^{k+1} be an optimal solution of this problem.

Step 3 $k \leftarrow k + 1$. Go to **Step 1**.

Prove the following:

a. LP^k and DP^k are always feasible.

b. If the algorithm stops, then it provides a feasible start for the regular generalized programming problem.

c. The sequence of σ^k values is nonincreasing.

d. The u^k vectors all lie in a closed and bounded convex set. What is this set?

e. There exists a convergent subsequence of the u^k vectors.

f. Define $D^*(u) = \min_{x \in X}\{u^T g(x)\}$. Show that $D^*(u^k) \leq \sigma^k$.